Quantum theory from the perspective of general probabilistic theories

Sabri Walid Al-Safi
Clare College
University of Cambridge

July 2014

This dissertation is submitted for the degree of
Doctor of Philosophy

Acknowledgements

The period during which this work was carried out would have been considerably less enjoyable without the intervention of many people. I am inestimably grateful for the unflinching love, support and tolerance of my parents, Linda and Walid, and my brother Khalid. I owe special thanks to Jassy, whose passion, awareness and integrity I will always aspire to, and whose counsel has always proved judicious. Thank you all for your patience.

I couldn't have undertaken this research without the energy and geniality of my supervisor, Tony Short, just as I couldn't have made it intelligible without the enthusiasm of my colleagues at the Centre for Quantum Information and Foundations in Cambridge, as well as many like-minded researchers whom I've met along the way. Thank you for helping me to see the way forwards.

I would also like to thank - in no particular order, not even a partial one - friends who have given me a great deal of pleasure during this time: Jack, Matt, Danny & Ellie, Zach, Laurence, John, Alex, Ilan, Sam, David, Dan, Sarah, Pippa, Will, Cat & Tom, Terry F. and Matt L., who brightened Pav. F basement; and tea, for inducing just the right kind of daze.

To Jassy

Abstract

This thesis explores various perspectives on quantum phenomena, and how our understanding of these phenomena is informed by the study of general probabilistic theories. Particular attention is given to quantum non-locality, and its interaction with areas of physical and mathematical interest such as entropy, reversible dynamics, information-based games and the idea of negative probability. We begin with a review of non-signaling distributions and convex operational theories, including "black box" descriptions of experiments and the mathematics of convex vector spaces.

In Chapter 3 we derive various classical and quantum-like quasiprobabilistic representations of arbitrary non-signaling distributions. Previously, results in which the density operator is allowed to become non-positive [1] have proved useful in derivations of quantum theory from physical requirements [2]; we derive a dual result in which the measurement operators instead are allowed to become non-positive, and show that the generation of any non-signaling distribution is possible using a fixed separable state with negligible correlation. We also derive two distinct "quasi-local" models of non-signaling correlations.

Chapter 4 investigates non-local games, in particular the game known as Information Causality. By analysing the probability of success in this game, we prove the conjectured tightness of a bound given in [3] concerning how well entanglement allows us to perform the task of random access coding, and introduce a quadratic bias bound which seems to capture a great deal of information about the set of quantum-achievable correlations. By reformulating Information Causality in terms of entropies, we find that a sensible measure of entropy precludes many general probabilistic theories whose non-locality is stronger than that of quantum theory.

Chapter 5 explores the role that reversible transitivity (the principle that any two pure states are joined by a reversible transformation) plays as a characteristic feature of quantum theory. It has previously been shown that in Boxworld, the theory allowing for the full set of non-signaling correlations, any reversible transformation on a restricted class of composite systems is merely a composition of relabellings of measurement choices and outcomes, and permutations of subsystems [4]. We develop a tabular description of Boxworld states and effects first introduced in [5], and use this to extend this reversibility result to any composite Boxworld system in which none of the subsystems are classical.

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

Contents

List of Figures

List of Notation

General

$\mathcal{H}$	Finite-dimensional, complex Hilbert space.
$\mathcal{L}(\mathcal{H})$	Vector space of linear operators on $\mathcal{H}$.
$\|\psi\rangle$	Unit vector in $\mathcal{H}$ representing a pure quantum state.
ρ	Hermitian, positive semi-definite, unit-trace (density) operator on $\mathcal{H}$.
$\tilde{\rho}$	Hermitian, unit-trace operator on $\mathcal{H}$.
E_i	Hermitian, positive semi-definite (POVM) operator on $\mathcal{H}$, satisfying $\mathrm{Tr}(E) \leq 1$
$M_{a\|x}$	POVM operator corresponding to measuring x for and obtaining outcome a.
$\tilde{M}_{a\|x}$	Non-positive measurement operator corresponding to measuring x and obtaining outcome a.
$[N]$	The set $\{1, \ldots, N\}$.
$H_c(\{\mathbf{p}\})$	Classical entropy of a probability distribution $\{\mathbf{p}\} = \{p_1, \ldots p_n)\}$.
$I_c(\{\mathbf{p}\} : \{\mathbf{q}\})$	Classical mutual information between distributions $\{\mathbf{p}\}$ and $\{\mathbf{q}\}$.

Outcome distributions

$P\left(a_1,\ldots,a_N\middle\vert x_1,\ldots,x_N\right)$	Multipartite probability of outcome a_i being obtained on subsystem i, conditional on having locally measured x_i
Ω	A subset of the set of subsystems $[N]$ over which the outcome distribution is defined.
$P\left(\mathbf{a}_\Omega\middle\vert\mathbf{x}_\Omega\right)$	Marginal outcome distribution over the subsystems belonging to Ω.
$M^{(i)}$	The number of fiducial measurement choices at subsystem i.
$K_{x_i}^{(i)}$	The number of possible outcomes for of fiducial measurement x_i on subsystem i.
$P\left(\mathbf{a}\middle\vert\mathbf{x}\right)$	Abbreviated form of the above, where $\mathbf{a}=(a_1,\ldots,a_N)$ and $\mathbf{x}=(x_1,\ldots,x_N)$.
$E_{x_1x_2}$	Bias of a g-bit system, conditional on measurements x_1 and x_2 being performed on subsystems 1 and 2.
$\mathcal{C}$ / $\mathcal{C}_1$	The CHSH value of an outcome distribution.
$P_\Lambda(\boldsymbol{\lambda})$	Joint probability distribution over local variables.
$\tilde{P}_\Lambda(\boldsymbol{\lambda})$	Joint quasiprobability distribution over local variables.
$P^{(i)}(a_i\vert x_i,\lambda_i)$	Local outcome function, determining probability of obtaining outcome a_i on subsystem i, conditional on having locally measured x_i and λ_i being the value of the local variable.
$\tilde{P}^{(i)}(a_i\vert x_i,\lambda_i)$	Quasiprobabilistic local outcome function.

V	Vector space representing the theory.
$\mathcal{S}$	Subset of V representing normalised states.
$\mathcal{S}_+$	Cone of V generated by $\mathcal{S}$.
$\mathcal{E}$	Subset of V representing effects.
$\mathcal{E}_+$	Cone of V generated by $\mathcal{E}$.
$\mathcal{U}$	Vector in V representing the unit effect.
$\mathcal{U}_N$	Unit effect vector for a composite system comprising N subsystems.
$\mathcal{S}_{[N]}$	The set of sub-unit effects of a composite system.
$\mathcal{S}_{\{i\}}$	The set of i-sub-unit effects, for subsystem i of a composite system.
$X_{a\|x}$	Fiducial effect corresponding to measuring x and obtaining outcome a.
$X^{(i)}_{a_i\|x_i}$	Local fiducial effect on subsystem i.
T	Transformation of states in V.
$T^\dagger$	Transformation of effects in V (adjoint of T).
$\mathcal{M}$	Set of effects corresponding to a measurement.
$\mathbb{M}^*$	The set of fine-grained measurements of a system in the theory.
$H^{me}(s)$	Measurement entropy of a state s.
$H^{mi}(s)$	Mixing entropy of a state s.

Chapter 1

Introduction

Physics is mathematical not because we know so much about the physical world, but because we know so little; it is only its mathematical properties that we can discover.

"An Outline of Philosophy"
Bertrand Russell

The history of quantum theory is replete with results, commonly referred to as "no-go theorems", which assert that a certain physical state of affairs is impossible, as long as one assumes that the mathematical formalism of quantum theory accurately describes the workings of Nature. No-go theorems often have ramifications for the practical application of quantum theory: the No-Cloning Theorem for example implies that it is impossible to construct a machine which can make copies of arbitrary quantum states [6]. On the other hand, many no-go theorems also have ramifications for our understanding of Nature: the Kochen-Specker theorem for example states that quantum mechanical properties cannot be embedded into a more fundamental set of properties of reality which, unlike quantum observables, are mutually compatible [7]. Invariably however, no-go theorems offer crucial insight into phenomena which are central to a full understanding of quantum theory, such as indeterminacy, entanglement, uncertainty relations, and the measurement process.

Arguably the most enduring and provocative phenomenon in the study of quantum foundations, with profound implications both practically and philosophically, is that of non-locality, encapsulated by Bell in 1964 with a no-go theorem known today as Bell's Theorem [8]. Basing his ideas on a 1930's thought experiment by Einstein *et al.* [9] later refined by Bohm [10], Bell proved that it is impossible for any physical theory obeying a well-defined notion of local causality to accurately reproduce all the experimental predictions of quantum theory. In particular, correlations in the local measurement statistics of a separated, but entangled, pair of electrons cannot be accurately reproduced by a physical theory which assumes certain - apparently reasonable - locality relations between the involved systems. Such "local" theories obey what are now known as *Bell inequalities*: restrictions on experimental outcome statistics which can be significantly violated by quantum measurements. Assuming that quantum theoretic predictions are accurate for Bell's proposed experiment (as verified by numerous experimental realisations to date, e.g. [11-19]), then local theories cannot be representative of Nature; the universe itself is non-local.

Bell's arguments opened up more fascinating questions than they resolved, and paved the way for extensive research into non-locality that continues to this day. Philosophical debate has raged over how Bell's original assumptions are to be interpreted, and whether there exist hidden assumptions in his argument that one might prefer to drop rather than locality [20-24]. Repeated realisations of Bell's experiment have attempted to close down various loopholes proposed over the years as a way for local theories to "get around" tests done so far, for example due to instrumental inefficiencies [14, 15, 25, 26] or sub-luminal influences between sites [11, 12, 19]. The practical utilization of non-locality has also been explored, with the development of a key distribution protocol that relies on violation of Bell inequalities as proof against eavesdropping [27].

Much of this thesis focuses on the technical study of non-locality as a quantifiable feature of measurement outcomes on quantum systems, and the intuition that it gives us about the mathematical structure of quantum theory. One of the greatest insights of Bell's Theorem is a focus on the measurement *results*, rather than the underlying physical state of affairs (or the mathematical language used to codify that state of affairs). Unlike the related phenomenon of entanglement, which admits only an algebraic characterization in terms of Hilbert spaces, non-locality is defined in terms of the statistics of experimental outcomes, i.e. the results which are noted in the observer's logbook, and how they correlate with another logbook some distance away. The purity of this approach is useful in multiple ways. Firstly, it opens a direct link to information theory, which ultimately cares about the storage and transmission of information that is accessible to humans - even quantum information theory in the end is concerned with the eventual classical readout of results. Secondly, it grants us a language with which to discuss hypothetical models of nature in terms of the measurement results they generate; these models are of special interest if quantum theory cannot replicate those results. Lastly, it is a potential means to putting quantum theory onto a more intuitive footing, and stripping it of the opaque Hilbert-space axiomatisation which pervades any introductory course in quantum theory.

The study of *general probabilistic theories* [2, 28-40] has been developed in recent years in order to capture and explore the idea of adopting a purely operational description of experiments. By "operational", we mean that the objects of the theory are meaningful only insofar as they influence experimental outcomes. For example, if two states of a system cannot be distinguished via the statistics of any measurement on that system, then they are to be regarded as the same state. By making a sparse and reasonable set of operational assumptions, a mathematical framework may be developed in which any model of nature may be represented, as long as that model admits a basic notion of systems, whose states inform the probabilities of measurement outcomes. Thus, quantum theory and classical theory (the meaning of which we will be more precise about in Chapter 2) each have their place in the set of general probabilistic theories, as do a whole host of theories which are capable of producing a greater amount of non-locality than is achievable in quantum theory.

A surprising number of features of quantum theory, which have previously been regarded as especially quantum, turn out to be quite generic within this larger set of general probabilistic theories. Features such as mixed states, entanglement, non-locality, and the use of the tensor product for composite systems in fact apply to all general probabilistic theories except classical theory, and can be seen as following from natural assumptions about measurement outcomes [29, 32, 41]. The No-Cloning and No-Broadcasting Theorems have analogues which also hold in all non-classical general probabilistic theories [31], for which the proofs are as intuitive as the quantum theoretic proofs (if not more so) and from which the quantum theoretic versions easily follow. Even quantum information theoretic protocols are fair game: the result that two copies of a Werner state cannot be deterministically purified follows from a more general result that holds in all general probabilistic theories satisfying a few reasonable conditions. [42]. Clearly, general probabilistic theories provide a valuable shift in perspective which allows for novel and illuminating proofs. In this thesis we hope to convince the reader that it is also helpful in shaping our understanding of fundamental concepts in quantum theory.

1.1 Overview of the Thesis

This thesis is divided into a further five chapters. In Chapter 2 we introduce and develop the framework of general probabilistic theories. Chapter 2 can be further divided into two parts: firstly, the characterization of a general probabilistic theory as a set of constraints on probability distributions over experimental outcomes, and secondly, an explanation of how such a theory can be represented using the mathematics of convex subsets of vector spaces. In both cases, we provide archetypal examples in order to clarify the meaning and importance of the concepts introduced. No new results are included in this chapter, however the author is not aware of a similarly comprehensive presentation of the mathematical framework of general probabilistic theories in the literature, which has an overview of the relevant convex and conic geometry, and encompasses individual and composite systems.

In Chapter 3, we derive a set of results that apply when one relaxes the positivity of outcome probabilities, and demonstrate that this allows for a "quasi-local" simulation of arbitrarily strong non-local states. In many situations, a local model is more convenient, since it bears similarity with the well-explored and intuitive territory of classical probability theory. We also discuss how quasiprobabilities, despite their counter-intuitive nature, have been been useful in the history of quantum mechanics, and may be used in future. The content of Chapter 3 is based mostly on a collaboration with Anthony J. Short [43], with additional discussion and some extended results.

In 2005,Wim van Dam demonstrated that superstrong non-locality is sufficient to reduce the communication complexity of any distributed computation to a triviality, revealing a close relationship between non-locality and information theory. From the perspective of general probabilistic thoeries, this result gives a neat physical principle which allows one to "rule out" correlations which maximally violate Bell's inequality, although not those which violate it to a non-maximal, yet post-quantum - degree. Ruling out this latter class of correlations has been achieved by a recent principle known as Information Causality; in Chapter 4, we examine

Information Causality from two distinct perspectives. Firstly, by analysing the probability of success in the Information Causality game and a closely related "inner product game", we prove that a bound on useful entanglement for the task of random access coding given in [3] is tight, as conjectured, and derive a quadratic bias bound on measurement probabilities which seems to capture a great deal of information about the structure of the set of quantum-achievable correlations. Secondly, we show that Information Causality might be more easily interpreted when reformulated in terms of entropies rather than mutual information, and show that the existence of a sensible measure of entropy is just as powerful in helping to "rule out" very non-local correlations. Chapter 4 features a detailed review of the main result of [1]; the remaining results are based on a collaboration with Anthony J. Short [44], with an extended discussion of the results.

Quantum phenomena such as entanglement and the tensor product are now seen to be generic features of general probabilistic theories. To understand what makes quantum theory stand out from other general probabilistic theories, it is beneficial to focus on the those features of quantum theory that appear to be rare amongst this set. One such feature is the ability to transform any pure state to any other via a reversible transformation. In Chapter 5, we develop the study of reversible dynamics in the general probabilistic theory known as Boxworld. In order to develop the intuition behind our results, we make use of a matrix-like formulation of Boxworld states and effects which was introduced in [5] to study measurements in Boxworld. We show that the reversible dynamics of composite Boxworld systems are composed merely of relabellings of measurement choices and outcomes, and permutations of subsystems. This extends a remarkable result by Gross *et al.* [4], which applies to Boxworld systems made up of identical subsystems. We then discuss the potential for, and the potential limits of, applying the techniques we develop to large classes of general probabilistic theories. This chapter features a review of the original result [4], a development of the above mentioned matrix-like formalism for general probabilistic theories introduced in [5], and the results obtained in [45], in collaboration with Anthony J. Short.

In the final chapter, we summarise the results obtained in Chapters 3 - 5, draw out parallels and contrasts between their approaches, and discuss how future work might build upon this research.

Chapter 2

General Probabilistic Theories

In other contexts, physicists have been able to take words from ordinary language and use them as technical terms with no great harm done. Take for example the "strangeness", "charm", and "beauty" of elementary particle physics. No one is taken in by this "baby talk"…. Would that it were so with "measurement". But in fact the word has had such a damaging effect on the discussion, that I think it should now be banned altogether in quantum mechanics.

"Against "Measurement""
John S. Bell

As a model of nature, quantum theory has met with unrivaled experimental verification and widespread acceptance. Part of its appeal lies in its unified, coherent account of various phenomena which are inexplicable in the classic Newtonian world-view. Whether talking about the position of a particle or the polarisation of a light-wave, experimental observations in quantum theory are predicted according to a standardised set of operations, involving a standardised set of mathematical objects. Whatever else it is, quantum theory is unarguably a useful and powerful tool for calculating the probabilistic outcomes of experiments.

Yet the underlying theory as it is usually presented - the quantum theory of Hilbert spaces and positive definite operators - is plagued with a multitude of interpretational issues. It is natural to wonder whether many of these these issues are due simply to the language in which quantum theory is currently written, or whether they are somehow inherent in the probabilities that quantum theory predicts, and thereby intrinsic to the universe. Recent years have seen the development of *general probabilistic theories* [2, 28-40], providing a framework in which the outcome statistics of experiments become the main focus of attention rather than the underlying physical theory which actually generates those statistics, be it Newtonian physics, quantum theory, or some post-quantum theory that has yet to be discovered. A major benefit of this abstract perspective is that one can make statements not just about quantum theory, but about any conceivable physical theory that makes probabilistic predictions about experimental outcomes, including those that may eventually supersede quantum theory. This has profound implications for the study of nature: as discussed in Chapter 1, Bell's famous theorem concerning local theories tells us not only that quantum theory is non-local, but that any theory which generates the same predictions is non-local, and hence nature itself is non-local.

In this chapter we introduce the framework of general probabilistic theories in two stages. Sections 2.1-2.4 make up the first stage, in which hypothetical physical theories are characterised by the outcome distributions which they predict are achievable in many-party experiments. Section 2.5 makes up the second stage, in

which we borrow various results from the mathematics of convex vector spaces in order to construct the framework of convex vector space representations. This framework provides a neat geometric picture of outcome distributions, rendering general probabilistic theories in a language that permits a much more general treatment of states, measurements and transformations. This allows for many features of quantum theory to be generalised to the full set of general probablistic theories, and consequently allows for the comparison of quantum theory to other theories.

2.1 Bell's Theorem and the CHSH game

In this section we outline a proof of Bell's Theorem and provide a more rigorous footing to non-locality. Although most readers will be familiar with the arguments, it is useful to introduce the language and notation that will be employed throughout this thesis. One of the simplest reformulations of Bell's Theorem, introduced by Clauser *et al.* in 1969 [46], introduces a "CHSH value", a measure of the correlations on the inputs and outputs of a very simple 2-player game, the "CHSH game". The basic set-up for this game is as follows: suppose that Alice and Bob choose, independently and uniformly at random, one of two inputs, and without communicating, generate in some way one of two possible outputs. It is convenient to label both the inputs and outputs using binary digits 0 and 1. Denoting Alice and Bob's inputs by $x, y \in \{0, 1\}$ respectively, and their outputs $a, b \in \{0, 1\}$, the statistics of the game are characterised by the conditional probability distribution:

$$P\left(a, b|x, y\right). \tag{2.1}$$

The CHSH value is a linear function of this binary, bipartite conditional probability distribution, and plays the role of the figure of merit in the game: the value which Alice and Bob are both attempting to maximise. In order to define it, it is first convenient to introduce a measure of the correlation of the outputs for a

specific choice of inputs x and y:

$$\begin{aligned} E_{xy} &= P\left(0,0|x,y\right) + P\left(1,1|x,y\right) - P\left(0,1|x,y\right) - P\left(1,0|x,y\right) \\ &= P\left(a=b|x,y\right) - P\left(a\neq b|x,y\right). \end{aligned} \tag{2.2}$$

(Note that if we had labelled the outputs as 1 and -1 instead of 0 and 1 respectively, then E_{xy} would be the expected value of the product $a \cdot b$, given the choices x and y). The CHSH value is the following linear function of correlators:

$$\mathcal{C} = E_{00} + E_{01} + E_{10} - E_{11}. \tag{2.3}$$

Thus, in order to maximise the CHSH value, Alice and Bob should attempt to produce correlated outputs whenever either x or y is equal to 0, but produce *anti*-correlated outputs whenever $x = y = 1$. Writing $E_{xy} = 2P\left(a=b|x,y\right) - 1 = 1 - 2P\left(a\neq b|x,y\right)$, and using the fact that Alice and Bob each choose their inputs independently and uniformly randomly, the CHSH expression may be simplified:

$$\begin{aligned} &E_{00} + E_{01} + E_{10} - E_{11} \\ &\quad = 2\left(P\left(a=b|0,0\right) + P\left(a=b|0,1\right) + P\left(a=b|1,0\right) + P\left(a\neq b|1,1\right)\right) - 4 \\ &\quad = 8P(a \oplus b = x \cdot y) - 4, \end{aligned} \tag{2.4}$$

where "$\oplus$" and "$\cdot$" denote addition and multiplication modulo 2.

Suppose that in order to process their inputs and generate their outputs, Alice and Bob perform measurements on a bipartite system, such that the particular measurement performed is determined by their inputs, and the outcome of the measurement determines their outputs. Despite having used the word "system" with all its connotations, we are making no claims about what *kind* of system is being used. For example, we are not committed to the measurement of a quantum system, although the reader is welcome to imagine that this is the case. We must also give heed to Bell's quotation at the beginning of this chapter, and be careful with what we mean by the word "measurement". It is not assumed that a pre-existing

property of the system is being revealed, merely that each subsystem interacts with some classical system representing the input, and some classical system representing the output. For the moment, as long as a single, definite value is assigned to the input and output for each run of the experiment, then what happens inbetween is none of our concern.

Now, if Alice and Bob's subsystems were to operate completely independently of one another, then we should expect that outcome probability depends solely on the local input, and that the joint output probability is given by the product of the local probabilities,

$$P\left(a,b|x,y\right)=P\left(a|x\right)\cdot P\left(b|y\right). \tag{2.5}$$

If the two subsystems are physically separated at the time of measurement, and we have no reason to believe they are related by some common cause, then we would expect the statistics to behave as a *product distribution*, obeying (2.5).

Now suppose that we want to allow for the fact that a common cause is influencing the statistics of both subsystems, despite them being separate at the time of measurement. Perhaps, for example, they interacted in the past in such a way that the outputs are random but perfectly correlated. We can codify this by introducing a random variable λ, which takes values in some set Λ with probability $P_{\Lambda}(\lambda)$, and such that *given* any individual value of λ, the subsystems behave independently, i.e. $P\left(a,b|x,y\right)=P\left(a|x,\lambda\right)\cdot P\left(b|y,\lambda\right)$. The actual experiment statistics are now obtained by averaging over λ,

$$P\left(a,b|x,y\right)=\sum_{\lambda\in\Lambda}P_{\Lambda}(\lambda)\left[P\left(a|x,\lambda\right)\cdot P\left(b|y,\lambda\right)\right]. \tag{2.6}$$

Naively, we might expect that this state of affairs completely characterises the outcome statistics that are possible on physically separated subsystems. If neither can be influenced directly by the other then, up to some common cause, we expect them to behave independently. In short, we expect the outcome distribution to be *local*, one that can be written in the form 2.6 for some set of local variables Λ. However, it can be shown that any local outcome distribution is itself merely an

average over the set of *deterministic* product distributions (see Section 4.1 for a rigorous proof of this) - there are only 16 such distributions, and a straightforward check demonstrates that all of these obey $P(a \oplus b = x \cdot y) \in \left\{\frac{1}{4}, \frac{3}{4}\right\}$, therefore have CHSH values $\mathcal{C} = \pm 2$. The maximal CHSH value obtainable by local distributions is therefore 2.

We now show that Alice and Bob may achieve a CHSH value greater than 2 by employing quantum measurements on a shared entangled state. Let $|\psi\rangle$ be the maximally entangled state $\left(\frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}}\right)$ on a bipartite system composed of two qubits, and suppose that Alice and Bob's inputs x and y determine which measurement they make on their individual qubit. We will consider measurements of the form $\mathbf{n} \cdot \boldsymbol{\sigma}$ for some unit vector $\mathbf{n}$, where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is a vector whose components are the 2-dimensional Pauli operators, and employ the following expectation values:

$$\begin{aligned}
\langle\psi|\sigma_x \otimes \sigma_x|\psi\rangle &= 1 \\
\langle\psi|\sigma_x \otimes \sigma_z|\psi\rangle &= 0 \\
\langle\psi|\sigma_z \otimes \sigma_x|\psi\rangle &= 0 \\
\langle\psi|\sigma_z \otimes \sigma_z|\psi\rangle &= 1
\end{aligned}$$

Note that $\mathbf{n} \cdot \boldsymbol{\sigma}$ is itself a Hermitian operator with eigenvalues ± 1: suppose that Alice chooses from two possible measurement operators $\{A_0, A_1\}$, and B chooses from $\{B_0, B_1\}$, and each measurement operator is of the form $\mathbf{n} \cdot \boldsymbol{\sigma}$. Suppose further that their outputs a and b are 0 or 1 depending on whether the outcome of the measurement is $+1$ or -1 respectively on Alice's or Bob's subsystem. Recall that in this case the correlator E_{xy} is the expectation of the product of the outcomes, $E_{xy} = (-1)^{a+b}$, which can also be written as $\langle\psi|A_x \otimes B_y|\psi\rangle$.

Suppose that Alice and Bob make the following measurement choices:

Input	0	1
Alice	σ_x	σ_z
Bob	$\left(\frac{\sigma_x+\sigma_z}{\sqrt{2}}\right)$	$\left(\frac{\sigma_x-\sigma_z}{\sqrt{2}}\right)$

(2.7)

Observe that the unit vector **n** is always zero in the y-coordinate, and the operators A_x and B_y are represented by unit vectors in the two-dimensional plane spanned by σ_x and σ_z.

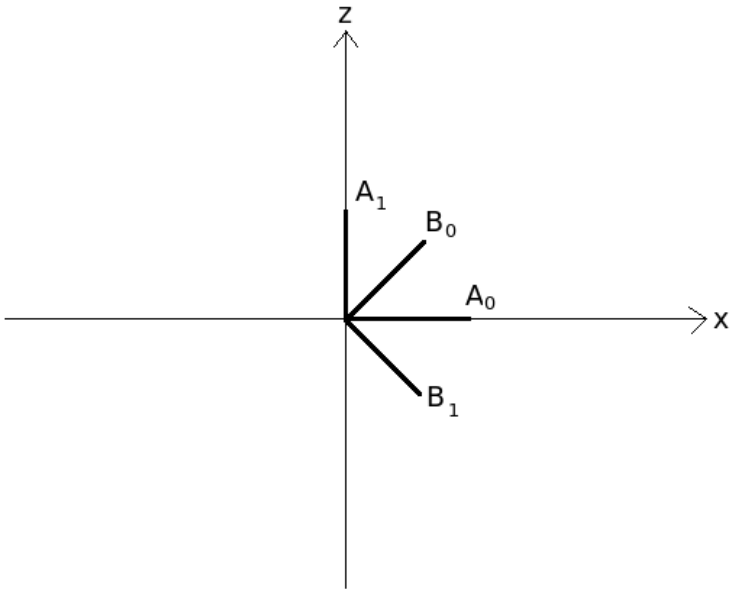

Figure 2.1: Unit vectors representing Alice and Bob's measurement choices.

The correlators can be calculated using the above expectation values and lin-

earity of the inner product:

$$E_{00} = \langle\psi|A_0 \otimes B_0|\psi\rangle = \frac{1}{\sqrt{2}}$$
$$E_{01} = \langle\psi|A_0 \otimes B_1|\psi\rangle = \frac{1}{\sqrt{2}}$$
$$E_{10} = \langle\psi|A_1 \otimes B_0|\psi\rangle = \frac{1}{\sqrt{2}}$$
$$E_{11} = \langle\psi|A_1 \otimes B_1|\psi\rangle = -\frac{1}{\sqrt{2}}$$

giving $\mathcal{C} = E_{00} + E_{01} + E_{10} - E_{11} = \frac{4}{\sqrt{2}} = 2\sqrt{2} > 2$. This simple quantum mechanical operation on entangled qubits generates a correlation between two distant parties that is impossible under the assumption that measurements made at physically separate locations are governed by local distributions.

The beauty of the CHSH value is that it is an easily calculated quantity which provides a succinct demonstration of how quantum theory departs from the classical regime, as well as a rough measure of the strength of non-locality which is achievable in any general probabilistic theory. By exploring the maximum obtainable value that a theory allows for quantities such as the CHSH value, we can set up a hierarchy of general probabilistic theories, based on the degree of non-locality they allow for. Local theories cannot exceed a CHSH value of 2, whereas this may be surpassed by quantum theory, up to a maximum of $2\sqrt{2}$ [47]. These values lie well below the algebraic maximum of 4, which is achieved in the maximally non-local theory known as "Boxworld" [48].

2.2 Black boxes and non-signaling distributions

In this section we attempt to extract some of the insights of Bell's Theorem, and begin to explore what can be said about Nature purely in terms of experimental outcomes. Recall our discussion in the previous Section of the word "measurement", in particular that we are not too concerned with what is really going on between

the inputs and the outputs. In fact, it is useful at this stage to think of a system as a "black box" into which the inputs go, and out of which come the outputs. When we later introduce general probabilistic theories, it may be helpful to think of each system as just such a black box; it may also be surprising, from this perspective, to see just how much of the seemingly internal workings of quantum theory admits a faithful *operational* analogue, i.e. how much can be described purely in terms of input/output probabilities.

A basic set of physical notions, or operational primitives, is clearly required in order to say anything about experimental outcomes: we assume that there exist various types of system, that each system may be prepared in one of various states, and that one of various measurements may be performed on a system. It goes without saying that the possible states and measurements depend on the type of system in play, and that the likelihood of a given outcome occurring is governed by the state.

In adopting an operational attitude, we are interested not in the *intrinsic* qualities of states of systems, such as their Hilbert-space dimension, but merely its *observable* qualities, i.e. how the state informs the outcomes of measurements. It is therefore useful to introduce the concept of an *effect*, which is no more than the occurrence of a particular outcome. An effect must belong to the set of possible outcomes for at least one measurement, but in principle multiple measurements may share a single effect. The state of a system is completely and uniquely characterised by the probabilities it assigns to individual effects; any two states which assign identical probabilities to the set of effects are considered to be the same state. This assumption neatly encapsulates the operational stance of general probabilistic theories: we should not suppose there to exist distinct states which nevertheless have exactly the same observable qualities.

It is common at this point to make a number of finiteness assumptions; these bring convenient simplifications, and may generally be justified on physical grounds. Firstly, a measurement comprises a finite set of effects, so that states assign discrete (rather than continuous) *outcome distributions* - probability distributions over

the outcome-sets of measurements. This restriction makes sense in the context of real-world quantum measurements: the result of an experiment, even after a large number of repeats with extremely refined instruments, can only be recorded to a certain number of significant figures. Although the underlying theory may predict a continuum of outcomes, for example in infinite dimensional Hilbert spaces, the outcome set is in practice always restricted to a finite set. Even if we did have access to infinitesimally accurate instruments, it would still be of interest to understand measurement statistics resulting from finite dimensional quantum systems, and this restriction would still be a reasonable one to make.

A second finiteness assumption is that for any system there exists a finite set of *fiducial measurements* whose outcome distributions suffice to uniquely identify the state of that system. The effects belonging to fiducial measurements are known as *fiducial effects*; a state is therefore completely characterised by the probabilities it assigns to the finite set of fiducial effects. This means that we restrict ourselves to considering only the statistics generated in finite dimensional quantum theory. Although this is not fully general, many of the quantum phenomena we wish to illuminate, such as entanglement, non-locality and non-contextuality, occur just as prominently in finite dimensions. In this thesis we are interested in seeing what can be demonstrated in this simpler setting, and leave the infinite-dimensional considerations to others.

We now make rigorous the above descriptions of measurement and outcome. An individual system, in which an experiment is modelled, allows for a certain number of measurement choices, indexed by a variable x which takes values in the finite set $\{1, \ldots, M\}$. For each measurement choice x, a certain number of outcomes are permitted, indexed by a variable a which takes values in the finite set $\{1, \ldots, K_x\}$ (note that the number of possible outcomes may depend on which measurement is performed). For convenience we will make an exception to this rule in the case that $M = K = 2$, and will instead label the measurement choices and outcomes using the set $\{0, 1\}$. An experiment involving this system is characterised by a set of real numbers $P(a|x)$, giving the probability that a is output,

conditioned on x having been input (see Figure 2.2).

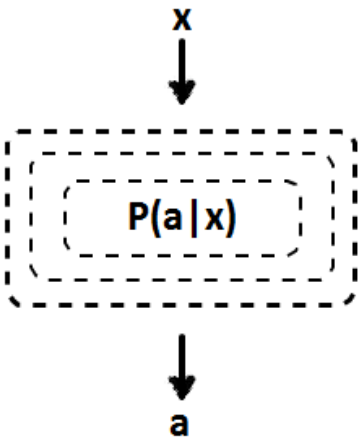

Figure 2.2: "Black box" description of experimental statistics.

Now consider a multipartite system comprising N subsystems, each of which can be regarded as an individual system in its own right, with its own set of measurements. Suppose that the measurements at subsystem i are indexed by some variable x_i which takes values in the finite set $\{1, \ldots, M^{(i)}\}$. Each measurement choice allows for one of a finite number of outcomes, indexed by the variable a_i which takes values in the finite set $\{1, \ldots, K_{x_i}^{(i)}\}$. Imagine that a measurement is performed individually on each subsystem, simultaneously and without any communication between subsystems. This global measurement on the multipartite system is described by a set of real numbers $P(a_1, \ldots, a_N|x_1, \ldots, x_N)$, giving the probability that outcomes $a_1, \ldots, a_N$ occur at systems $1, \ldots, N$ respectively, conditioned on measurement choices $x_1, \ldots, x_N$ having been performed on those systems (see Figure 2.3).

Just as a specification of the outcome distributions $P(a|x)$ uniquely characterises the experiment on a single system, we assume that a specification of the values $P(a_1, \ldots, a_N|x_1, \ldots, x_N)$ is sufficient to characterise the multipartite experiment. This assumption, known variously as local tomography [2, 34], local distinguishability [33], the Local Observability Principle [41] or the Global State

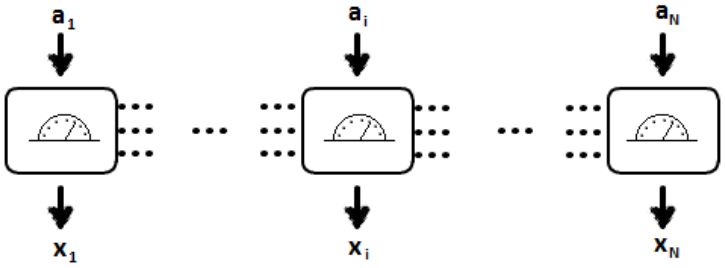

Figure 2.3: Multipartite experiment composed of N black boxes.

Assumption [29], asserts that there are no *additional* degrees of freedom in a composite system which are not captured by the differing measurement choices for the individual subsystems.

The values $P(a_1, \ldots, a_N|x_1, \ldots, x_N)$ correspond to a conditional probability distribution, and in order for them to be physically meaningful they must obey the normal laws of probability. Since each individual value is a probability, it must be positive:

$$0 \leq P(a_1, \ldots, a_N|x_1, \ldots, x_N). \tag{2.8}$$

For any fixed choice of measurements $x_1, \ldots, x_N$ on all subsystems, the probabilities for all possible values of outcomes $a_1, \ldots, a_N$ must sum to 1, i.e. the outcome distribution for that choice of local measurements is normalised:

$$\sum_{a_1,\ldots,a_N} P(a_1, \ldots, a_N|x_1, \ldots, x_N) = 1. \tag{2.9}$$

(This also implies that each value $P(a_1, \ldots, a_N|x_1, \ldots, x_N)$ is at most 1).

We have stated that the conditional probability $P(a_1, \ldots, a_N|x_1, \ldots, x_N)$ results from individual measurements made on separate subsystems, although we have not defined what we mean by "separate". This operational framework concerns only experimental outcomes, and has no *a priori* notion of distance or simultaneity. Nevertheless, our goal is to discuss the outcomes of experiments made in the real world, and so we make the assumption that information cannot be trans-

mitted from one set of subsystems to another through the process of measurement. A justification for this is that the subsystems may represent physically separated real-world structures; since the local measurements in principle may be carried out simultaneously, there should be no influence from one set of subsystems to another during the measurement procedure.

To ensure that information cannot be sent between the subsystems, we demand that a *no-signaling condition* is satisfied by the outcome distribution P. This condition asserts that the outcome statistics for any subset Ω of the N subsystems must not be affected by measurement choices made on those subsystems not belonging to Ω, i.e. there is no way for one subsystem to signal information to any other subsystems. In mathematical terms, we demand that for all $\Omega \subset [N]$, the reduced value,

$$\sum_{a_i : i \notin \Omega} P\left(a_1, \ldots, a_N | x_1, \ldots, x_N\right), \tag{2.10}$$

is well-defined and independent of the value of each x_i for systems $i \notin \Omega$. If the no-signaling condition is obeyed by P, then we say that P is a *non-signaling distribution*. Non-signaling distributions have the following property: whatever the choice of measurement made by systems *outside* of any subset Ω, the marginal outcome distribution on the systems *inside* Ω is invariant. In fact, the no-signaling condition is equivalent to the existence of a well-defined reduced outcome distribution for any subset of N; writing $\Omega = \{i_1, \ldots, i_r\} \subset N$, the reduced outcome distribution is defined:

$$P\left(a_{i_1}, \ldots, a_{i_r} | x_{i_1}, \ldots, x_{i_r}\right) = \sum_{a_i : i \notin \Omega} P\left(a_1, \ldots, a_N | x_1, \ldots, x_N\right). \tag{2.11}$$

It should be stressed that the no-signaling condition is distinct from the impossibility of superluminal propagation of information or matter, which holds in the special theory of relativity. The no-signaling condition is a reasoned assumption stated in terms of outcome statistics, and does not rely on any concept of space-time. On the other hand, the impossibility of superluminal signaling is deduced

from the principles of special relativity, and is therefore true in our best model of spacetime.

It is difficult to imagine reasonable theories of nature which allow for states that violate the no-signaling condition. The possibility that the reduced state of a subsystem might immediately change according to choices made in a distant location violates any intuitive feeling one has about what the reduced state of a subsystem even means. Conversely, it is not difficult to conceive of theories in which superluminal propagation is in principle possible: non-relativistic quantum mechanics provides one such example, in which the no-signaling condition holds but there is no reason that the movement of a particle should be hampered by the speed of light.

In order to simplify notation, we write $\mathbf{a}$ and $\mathbf{x}$ for the strings $(a_1, \ldots, a_N)$ and $(x_1, \ldots, x_N)$. When discussing non-signaling distributions we use the notation $\mathbf{a}_\Omega$ and $\mathbf{x}_\Omega$ for the "reduced strings" $(a_{i_1}, \ldots, a_{i_r})$ and $(x_{i_1}, \ldots, x_{i_r})$, and write $P(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ for the reduced outcome distribution defined by (2.11). The no-signaling condition can then be expressed in a compact form: for any $\Omega \subset [N]$, measurement-choice strings $\mathbf{x}$, $\mathbf{x}'$ satisfying $\mathbf{x}_\Omega = \mathbf{x}'_\Omega$, and fixed choice of outcomes $\mathbf{a}_\Omega$,

$$\sum_{a_i : i \notin \Omega} P(\mathbf{a}|\mathbf{x}) = \sum_{a'_i : i \notin \Omega} P(\mathbf{a}'|\mathbf{x}') = P(\mathbf{a}_\Omega|\mathbf{x}_\Omega), \tag{2.12}$$

where $\mathbf{a}'$ denotes the string whose components are a_i for $i \in \Omega$ and a'_i for $i \notin \Omega$.

The following Proposition demonstrates that in order to check whether a known outcome distribution is non-signaling, it suffices to verify that the non-signaling condition holds for all Ω with $|\Omega| = 1$, i.e. that it is not possible for any *single* subsystem to signal to the rest via the choice of measurement on that subsystem.

Proposition 1. *Suppose that the outcome distribution $P(\mathbf{a}|\mathbf{x})$ satisfies the condition that for all subsystems i and distinct measurement choices $x_i \neq x'_i$ on subsys-*

tem i,

$$\sum_{a_i=1}^{K_{x_i}} P\left(a_1,\ldots,a_i,\ldots,a_N|x_1,\ldots,x_i,\ldots x_N\right) = \sum_{a_i'=1}^{K_{x_i'}} P\left(a_1,\ldots,a_i',\ldots,a_N|x_1,\ldots,x_i',\ldots x_N\right). \tag{2.13}$$

Then P is a non-signaling outcome distribution.

Proof. Suppose that $\mathbf{x}$ and $\mathbf{x}'$ are measurement-choice strings, satisfying $\mathbf{x}_\Omega = \mathbf{x}'_\Omega$ for a given subset Ω of the N subsystems. Fix a choice of outcomes $\mathbf{a}_\Omega$ for the subsystems belonging to Ω. We wish to show that summing over the set of possible outcomes for all subsystems not in Ω leads to the same value, whether the outcome distribution is taken with respect to $\mathbf{x}$ or with respect to $\mathbf{x}'$, i.e.

$$\sum_{a_i:i\notin\Omega} P\left(\mathbf{a}|\mathbf{x}\right) = \sum_{a_i':i\notin\Omega} P\left(\mathbf{a}'|\mathbf{x}'\right). \tag{2.14}$$

As strings of numbers, $\mathbf{x}$ and $\mathbf{x}'$ differ in at most $s = N - |\Omega|$ components (those not belonging to Ω). Writing $[N] \setminus \Omega = \{i_1,\ldots,i_s\}$, it is possible to change s components of $\mathbf{x}$ one at a time, terminating at $\mathbf{x}'$:

$$\mathbf{x} = \mathbf{x}^0 \to \mathbf{x}^{i_1} \to \cdots \to \mathbf{x}^{i_s} = \mathbf{x}'. \tag{2.15}$$

However, by varying the order of summation on the LHS of (2.14) and using

property (2.13), we have that,

$$\begin{aligned}
\sum_{a_i : i \notin \Omega} P(\mathbf{a}|\mathbf{x}) &= \sum_{a_i : i \notin \Omega \cup \{i_1\}} \sum_{a_{i_1}} P(\mathbf{a}|\mathbf{x}) \\
&= \sum_{a_i : i \notin \Omega \cup \{i_1\}} \sum_{a_{i_1}} P\left(\mathbf{a}|\mathbf{x}^{i_1}\right) \\
&= \sum_{a_i : i \notin \Omega \cup \{i_2\}} \sum_{a_{i_2}} P\left(\mathbf{a}|\mathbf{x}^{i_2}\right) \\
&\quad \vdots \qquad\qquad\qquad \vdots \\
&= \sum_{a_i : i \notin \Omega \cup \{i_s\}} \sum_{a_{i_s}} P\left(\mathbf{a}|\mathbf{x}^{i_s}\right) \\
&= \sum_{a'_i : i \notin \Omega} P(\mathbf{a}'|\mathbf{x}') \,. \qquad (2.16)
\end{aligned}$$

□

2.3 Examples of probabilistic theories

Any theory of nature must provide some means for predicting experimental outcomes. By virtue of this fact, any experiment described by that theory may consequently be described by a "black box" scenario as outlined in the previous Section, in which we ignore the internal workings of the box (i.e. the mathematical operations which generate predictions in that theory). A *state* in the theory must have some mechanism for specifying outcome probabilities for any measurement allowed by the theory; thus it assigns probabilities to *effects*, or possible measurement outcomes. Furthermore, we assume that any two distinct states must be distinguishable via their outcome statistics for at least one measurement; thus distinct states assign distinct probabilities to at least one effect. As we have previously discussed, we will assume that there exists a finite set of fiducial measurements, each consisting of a finite set of fiducial effects, such that distinct states assign distinct probabilities to at least one fiducial effect.

Note that the positivity (2.8), normalisation (2.9) and non-signaling (2.10) conditions are all clearly preserved under convex combinations. That is to say, if the number of subsystems, measurements and outcomes is fixed, the outcome distributions $P_1, \ldots, P_r$ satisfy all three conditions, and $\{p_1, \ldots, p_r\}$ is any probability distribution, then the outcome distribution defined by:

$$P\left(\mathbf{a}|\mathbf{x}\right) = \sum_{i=1}^{r} p_i P_i\left(\mathbf{a}|\mathbf{x}\right), \tag{2.17}$$

also satisfies all three conditions. Hence the set of non-signaling outcome distributions forms a convex set, which we will explore in more detail in Section 2.4.

It is often convenient to specify a physical theory by explicitly delineating the set of outcome distributions which are obtainable within that theory; generally speaking, this involves a threefold specification: firstly, which types of individual system are described by the theory, including their fiducial measurements; secondly, for each individual system a specification of which outcome distributions are obtainable on the fiducial measurements; and finally a specification of which joint distributions are permitted in composite systems. A general probabilistic theory is just such a specification of systems and outcome distributions. We will demand that the set of distributions in a general probabilistic theory is always convex: the reason for this is that for any set of states -- with outcome distributions $P_1, \ldots, P_r$ -- that can be prepared in a system, it should be possible to prepare the state whose outcome distribution is P_i with probability p_i, so that (2.17) gives the outcome distribution after such a preparation, and hence also corresponds to a state in the theory. By analogy to the pure states of quantum theory, the *extreme states* of a general physical theory are those that give rise to extremal points of the convex set of outcome distributions: they are those states which cannot be written as a convex sum of distinct states of the theory (see Section 2.5.1 for a rigorous treatment of these notions).

It should finally be noted that there is a distinction between "black boxes" and full general probabilistic theories. The black box scenario is simply a descrip-

tion of inputs, outputs, and the subsequent outcome distributions. This is useful, for example, for defining the non-signaling principle and the CHSH value purely in terms of such distributions. General probabilistic theories, on the other hand, whilst often described in terms of outcome distributions, include a characterisation (usually based on physical principles) of which local and composite outcome distributions are possible, as well as concepts like states, effects, fiducial measurements and so on. As in quantum theory, general probabilistic theories often allow for measurements of effects *other* than fiducial effects.

2.3.1 Classical theory

In classical theory, individual systems have a single fiducial measurement, and any outcome distribution on that measurement constitutes a permitted state. The fiducial measurement captures full information about the system. An example of a classical system is a die: the fiducial measurement corresponds to observing the upturned face and a fair die (once rolled) represents the uniformly distributed (or completely mixed) state. A general state of the system is given by a probability distribution $(p_1, \ldots, p_6)$. The extreme states are the *deterministic* states s_i, which give outcome i with certainty, e.g. the state $s_1 = (1, 0, \ldots, 0)$: it is clear that these cannot be written as a convex combination of distinct states, and conversely that any state is a convex combination of the extreme states.

The joint states of composite classical systems are convex combinations of product states. Product states are given by a product of outcome distributions at each site:

$$P\left(\mathbf{a}|\mathbf{1}\right) = P^{(1)}\left(a_1|1\right) \cdots P^{(N)}\left(a_N|1\right) = \prod_{i=1}^{N} P^{(i)}\left(a_i|1\right), \tag{2.18}$$

where $x_i = 1$ for all i since we only have one fiducial measurement. Clearly, this distribution can be achieved simply by preparing the state $P^{(i)}$ at each subsystem

i. A convex combination of states of the form (2.18) can be described by a random variable $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$ and an associated probability distribution $P_\Lambda(\boldsymbol{\lambda})$. Each λ_i is a local variable taking values from some finite set Λ_i, and each value of λ_i corresponds to some outcome distribution $P^{(i)}\left(a_i|1, \lambda_i\right)$ that is allowed at subsystem i. The joint states of multipartite classical systems are exactly those of the form:

$$P\left(\mathbf{a}|\mathbf{1}\right) = \sum_{\boldsymbol{\lambda}} P_\Lambda(\boldsymbol{\lambda}) \prod_{i=1}^{N} P^{(i)}\left(a_i|1, \boldsymbol{\lambda}_i\right). \tag{2.19}$$

2.3.2 Quantum theory

In quantum theory, outcome distributions are generated by measurements on quantum states. A quantum system, with M fiducial measurements and K_x outcomes for measurement x, consists of a finite-dimensional Hilbert space $\mathcal{H}$ and for each x a set of POVM elements $\{M_{a|x}\}$ satisfying $M_{a|x} \geq 0$ and $\sum_{a=1}^{K_x} M_{a|x} = \mathbb{I}$. The outcome distributions $P\left(a|x\right)$ allowed in the system are generated from density operators $\rho \in \mathcal{D}\left(\mathcal{H}\right)$ according to the Born trace rule:

$$P\left(a|x\right) = \text{Tr}\left(M_{a|x}\rho\right) \tag{2.20}$$

On any finite-dimensional Hilbert space there always exists a finite set of fiducial POVM measurements, such that distinct density operators will generate distinct probabilities for at least one fiducial POVM element [49]. In general, at least D^2 independent POVM matrices are required to specify a single density operator in a D-dimensional Hilbert space - however, the choice of POVM elements and how they are split up into distinct measurements is not determined by the dimension D. In order to define a quantum system in terms of general probabilistic theories, therefore, one needs to specify both the dimension of the Hilbert space, and the particular fiducial measurements being employed.

The simplest example of a quantum system is a qubit, with one possible set of fiducial measurements being the three projective measurements onto the eigenbases of the Pauli operators σ_x, σ_y and σ_z. Listing these measurements as $x = 1, 2$ and 3 respectively, and listing their two possible outcomes as $a = 1$ or 2, the state of the qubit is completely specified by the vector:

$$\Big(\; P\left(1|1\right), \quad P\left(2|1\right) \; \Big| \; P\left(1|2\right), \quad P\left(2|2\right) \; \Big| \; P\left(1|3\right), \quad P\left(2|3\right) \; \Big) . \tag{2.21}$$

A qubit that is in the "up" state of the x-direction is represented by the vector $(1, 0|\frac{1}{2}, \frac{1}{2}|\frac{1}{2}, \frac{1}{2})$, whereas the vector $(1, 0|1, 0|1, 0)$ is not achievable by any qubit state.

Multipartite quantum systems consist of a collection of N quantum systems with Hilbert spaces $\mathcal{H}_1, \ldots, \mathcal{H}_N$, and with the fiducial measurements at system i given by the POVM operators $\{M^{(i)}_{a_i|x_i}\}$. The set of multipartite outcome distributions on the composite system is generated by local POVM measurements on a joint quantum state $\rho \in \mathcal{D}\left(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N\right)$:

$$P\left(\mathbf{a}|\mathbf{x}\right) = \mathrm{Tr}\left(\left(M^{(1)}_{a_1|x_1} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}\right)\rho\right). \tag{2.22}$$

Any outcome distribution resulting from a multipartite quantum measurement is non-signaling: this is due to the fact that $\sum_{a_i} M^{(i)}_{a_i|x_i} = \mathbb{I}^{(i)}$, regardless of i and the choice of measurement x_i. Indeed, fixing i and x_i, we see that:

$$\begin{aligned}
&\sum_{a_i=1}^{K_{x_i}} \mathrm{Tr}\left(\left(M^{(1)}_{a_1|x_1} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}\right)\rho\right) \\
&\qquad = \mathrm{Tr}\left(\left(M^{(1)}_{a_1|x_1} \otimes \cdots \otimes \left[\sum_{a_i=1}^{K_{x_i}} M^{(i)}_{a_i|x_i}\right] \otimes \cdots \otimes M^{(N)}_{a_N|x_N}\right)\rho\right) \\
&\qquad = \mathrm{Tr}\left(\left(M^{(1)}_{a_1|x_1} \otimes \cdots \otimes \mathbb{I}^{(i)} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}\right)\rho\right),
\end{aligned} \tag{2.23}$$

which is clearly independent of the vaule of x_i. It follows from Proposition 1 that

$P(\mathbf{a}|\mathbf{x})$ is non-signaling.

The set of outcome distributions generated by a fixed composite quantum system is subtly different from the set of non-signaling outcome distributions (with a set number of measurement choices and outcomes) which can be achieved by measurements of arbitrary quantum systems. In the former case, we have already explicitly defined each subsystem (i.e. their Hilbert spaces and fiducial measurements); on the other hand, we may say that an outcome distribution $P(\mathbf{a}|\mathbf{x})$ for some pre-defined number of subsystems, measurement choices and outcomes is *quantum achievable* if there exist *some* set of finite-dimensional Hilbert spaces $\mathcal{H}_1, \ldots, \mathcal{H}_N$, POVM operators $\{M^{(i)}_{a_i|x_i}\}$ and some state $\rho \in \mathcal{D}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N)$ such that (2.22) holds.

2.3.3 Boxworld

In an individual Boxworld system, any set of distributions over the outcomes of the fiducial measurements is permitted: the state space is constrained only by the positivity and normalisation conditions: any conditional probability distribution $P(a|x)$ corresponds to an allowed state. Thus the outcome distribution $(1,0|1,0|1,0)$ which was not allowed in the qubit system, is a perfectly valid Boxworld distribution. The set of Boxworld distributions for a single system is a convex polytope, whose extremal vertices are the *deterministic* states, i.e. those which, conditional on the measurement choice, produce one output with certainty. Any Boxworld system with only one fiducial measurement is equivalent to a classical system. The Boxworld system with two fiducial measurements, each of which has two possible outcomes is referred to as a *g-bit* system, which stands for generalised bit. This system may loosely be seen as a Boxworld analogue of the classical bit and quantum qubit systems.

As with classical theory, the joint state space of composite Boxworld systems should include at least those states which are convex combinations of products of local states. A multipartite Boxworld state with outcome distribution $P(\mathbf{a}|\mathbf{x})$ is

local if there exists a random variable $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$ taking values in a product of finite sets $\Lambda_1 \times \cdots \times \Lambda_N$, an associated probability distribution $P_\Lambda(\boldsymbol{\lambda})$, and for each value of $\lambda_i \in \Lambda_i$ a local outcome distribution $P^{(i)}\left(a_i|x_1, \lambda_i\right)$ such that:

$$P\left(\mathbf{a}|\mathbf{x}\right) = \sum_{\boldsymbol{\lambda}} P_\Lambda(\boldsymbol{\lambda}) \prod_{i=1}^{N} P^{(i)}(a_i|x_i, \lambda_i). \tag{2.24}$$

The set of local Boxworld outcome distributions also forms a convex polytope, whose extremal vertices are given by products of deterministic states; such states are known as *pure product states*. Any outcome distribution which cannot be written as a convex combination of pure product states (2.24) is said to be *non-local*. A multipartite density operator ρ is said to be a *local quantum state* if any local POVM measurement on ρ results in a local outcome distribution being generated: thus, separable states are local quantum states, whereas the Bell states are non-local quantum states (since there exist local POVM measurements on them which violate the CHSH inequality).

Not all valid non-signaling distributions in composite Boxworld systems admit a local description (2.24). For example, the outcome distribution generated by CHSH-violating measurements on a Bell state is non-signaling, but non-local. Boxworld is the general probabilistic theory whose systems are as described above, and whose multipartite states are all those outcome distributions which obey the positivity (2.8), normalisation (2.9) and non-signaling (2.10) conditions. Since these three conditions are represented by sets of linear inequalities, the set of outcome distributions in a composite Boxworld system forms a convex polytope. Any composite quantum system with has the same number of fiducial measurements on each local subsystem is capable of generating a subset of the Boxworld outcome distributions (as we will see, this turns out to be a strict subset).

The first known example of a non-signaling, non-quantum-achievable outcome distribution is the celebrated PR-box state on a bipartite system comprising two g-bit subsystems [48]. It is convenient to label the measurement choices and outcome sets of a g-bit as binary numbers $\{0, 1\}$ rather than $\{1, 2\}$, though this

change in labeling does not affect the outcome probabilities, and therefore has no operational significance. In binary notation, the PR-box is defined by:

$$P\left(a_1, a_2|x_1, x_2\right) = \frac{1}{2}\delta(a_1 \oplus a_2 = x_1 \cdot x_2), \tag{2.25}$$

where "$\oplus$" and "$\cdot$" denote addition and multiplication modulo 2. This outcome distribution is clearly seen to be non-signaling, as the outcome on a single system is always uniformly random, regardless of the measurement choice on that or the other system. Note also that a_1 and a_2 are perfectly correlated for every choice of fiducial measurement except for $x_1 = x_2 = 1$, in which case they are perfectly anti-correlated. Hence the CHSH value takes its algebraic maximum of 4 on the PR-box state:

$$\mathcal{C}_1 = E_{00} + E_{01} + E_{10} - E_{11} = 1 + 1 + 1 - (-1) = 4. \tag{2.26}$$

where E_{xy} was defined in Section 2.1 as $P\left(a = b|x, y\right) - P\left(a \neq b|x, y\right)$

The PR-box state is in fact one of the extremal vertices of the convex polytope of states described by Boxworld for two g-bit systems [30]. This is the simplest example of a composite system in which the sets of locally achievable, quantum-achievable and non-signaling outcome distributions are distinct; we explore the geometry of these sets in more detail in Section 2.4. The state of a system of two g-bits may be depicted by means of a table or matrix, whose rows correspond to outcomes of fiducial measurements on system 1, and whose columns correspond to outcomes of fiducial measurements on system 2. The entry in the row corresponding to (a_1, x_1) and column corresponding to (a_2, x_2) is the value $P\left(a_1, a_2|x_1, x_2\right)$. The PR-box state is then represented by the following table:

		$x_2 = 0$		$x_2 = 1$	
		$a_2 = 0$	$a_2 = 1$	$a_2 = 0$	$a_2 = 1$
$x_1 = 0$	$a_1 = 0$	1/2	0	1/2	0
	$a_1 = 1$	0	1/2	0	1/2
$x_1 = 1$	$a_1 = 0$	1/2	0	0	1/2
	$a_1 = 1$	0	1/2	1/2	0

(2.27)

Increasing the number of measurements on a subsystem will generate more blocks of the table, whereas increasing the number of outcomes specific to a measurement will increase the size of the blocks corresponding to that measurement. An arbitrary bipartite Boxworld system need not have the same number of blocks across as it has down, and those blocks need not be square. Certain properties of the table must remain invariant however: all entries must be positive, and the sum of the entries in an individual block must be equal to 1 (due to the positivity and normalisation conditions). Each row of the table is subdivided into a set of smaller "block-rows"; in the above example the third row is subdivided into two block-rows, $(1/2, 0)$ and $(0, 1/2)$. Due to the non-signaling condition the sum of entries across each of the block-rows is invariant across a row (and is equal to $\frac{1}{2}$ for all rows in the above PR-box example). The same is true for columns and block-columns.

Whilst 2-dimensional tables are sufficient to describe all bipartite Boxworld distributions, an n-partite Boxworld system will generally be described by an n-dimensional array. This representation thus becomes infeasible for obvious reasons when the number of subsystems exceeds 3; however, since many of the interesting Boxworld phenomena occur with 3 or fewer subsystems, this does not pose too great a problem.

2.4 The non-signaling polytope

The full set of outcome distributions $P(\mathbf{a}|\mathbf{x})$ for a composite Boxworld system can be naturally represented by vectors, whose components are indexed by all possible strings $(\mathbf{a}|\mathbf{x})$ of inputs and outputs at each subsystem. The allowed set of non-signaling vectors then forms a bounded polytope, due to the linear constraints imposed by the positivity, normalisation and non-signaling conditions, which is often known as the Non-signaling Polytope. In this naive representation of the Non-signaling Polytope, the vectors are of length $\prod_{i=1}^{N}\left(\sum_{x_i=1}^{M^{(i)}} K_{x_i}\right)$, but generally belong to a lower-dimensional vector subspace. Even for the bipartite g-bit system in which the PR-box lives, these vectors have 16 components, but lie in an 8-dimensional vector subspace [30], which is, for example, orthogonal to the vector $e_1 + e_2 - e_3 - e_4$, since the non-signaling condition demands that $P(0,0|0,0) + P(0,1|0,0) = P(0,0|0,1) + P(0,1|0,1)$.

The extremal local states of the Non-signaling Polytope, i.e. the pure product states, are represented by vectors whose only non-zero entries are equal to 1. In a system of two g-bits, exactly four of the entries of a pure product state are equal to 1, since each of the four possible global measurement choices (x_1, x_2) results deterministically in an outcome (a_1, a_2). The number of extremal local states is equal to the product of the number of deterministic states on each individual g-bit system. There are four deterministic states of a g-bit system, corresponding to the four possible maps from $\{0,1\}$ to itself; this gives 16 extremal local states:

$$P_L^{\mu\nu\sigma\tau}(a_1,a_2|x_1,x_2) = \delta\big(a_1 = [\mu x_1] \oplus [\nu]\big)\cdot\delta\big(a_2 = [\sigma x_2] \oplus [\tau]\big) \qquad (2.28)$$
$$\mu,\nu,\sigma,\tau \in \{0,1\}.$$

Such states, being $\{0,1\}$-valued, are clearly extremal in the set of non-signaling states; however, they do not constitute the entire set of extremal non-signaling states. The PR-box state described in Section 2.3.3 is an example of an extremal *non-local* vertex of the Non-signaling Polytope [30]. By performing a simple local operation - for example, flipping the bit x_1 at subsystem 1 - the PR-box is con-

verted into a new outcome distribution - in this case defined by $P(a_1, a_2|x_1, x_2) = \frac{1}{2}\delta(a_1 \oplus a_2 = [x_1 \cdot x_2] \oplus x_2)$. This is also an extremal state, in which the outputs are anti-correlated for all choices of measurement except $x_1 = 0, x_2 = 1$. Via this type of local operation on inputs and outputs, 8 distinct extremal non-local distributions can be obtained:

$$P_{NL}^{\mu\nu\sigma}(a_1, a_2|x_1, x_2) = \frac{1}{2}\delta\big(a_1 \oplus a_2 = [x_1 \cdot x_2] \oplus [\mu x_1] \oplus [\nu x_2] \oplus [\sigma]\big) \quad (2.29)$$
$$\mu, \nu, \sigma \in \{0, 1\}.$$

The CHSH value $\mathcal{C}_1$ takes a maximal value of 2 in the set of local outcome distributions, a maximal value of $2\sqrt{2}$ in the set of quantum-achievable outcome distributions, and a maximal value of 4 in the full set of non-signaling outcome distributions. Note that for each value of (x_1, x_2) the bias,

$$E_{x_1x_2} = P(0, 0|x_1, x_2) + P(1, 1|x_1, x_2) - P(0, 1|x_1, x_2) - P(1, 0|x_1, x_2), \quad (2.30)$$

is a linear functional on the set of vectors, hence the CHSH value is itself a linear functional. So too is the similarly defined figure of merit:

$$\mathcal{C}_2 = E_{00} - E_{01} + E_{10} + E_{11}, \quad (2.31)$$

for which the maximal values taken by local, quantum and non-signaling outcome distributions are the same as for $\mathcal{C}_1$ (but are not necessarily achieved on the same outcome distributions). For example, $\mathcal{C}_1$ achieves its algebraic maximum of 4 and minimum of -4 on the non-local states P_{NL}^{000} and P_{NL}^{001} respectively, whereas $\mathcal{C}_1$ achieves the same values on the non-local states P_{NL}^{010} and P_{NL}^{011} respectively. All other extremal non-local states achieve 0 for both $\mathcal{C}_1$ and $\mathcal{C}_2$, whereas all extremal local states achieve ± 2 for both $\mathcal{C}_1$ and $\mathcal{C}_2$. By plotting $\mathcal{C}_1$ and $\mathcal{C}_2$ on x- and y- axes, one obtains a 2-dimensional projection, or slice, of the Non-signaling Polytope (see figure 2.4).

The set of quantum-achievable states in this projection is described by the for-

mula $\mathcal{C}_1^2 + \mathcal{C}_2^2 \leq 8$. A derivation of this inequality from Information Causality is given in Section 4.2.2, although it can also be derived from the fact that the set of biases E_{xy} is quantum achievable if and only if:

$$|E_{00}E_{01} - E_{10}E_{11}| \leq \sum_{y=0,1} \sqrt{(1 - E_{0y}^2)(1 - E_{1y}^2)}. \tag{2.32}$$

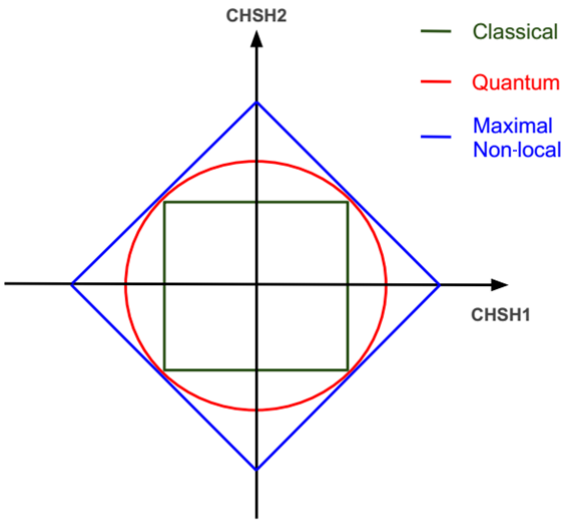

Figure 2.4: 2-dimensional projection of the non-signaling polytope for two g-bits.

Many of the beguiling aspects of quantum non-locality are captured by the simple diagram illustrated in Figure 2.4. The CHSH and Tsirelson inequalities demonstrate that for some figures of merit, the optimal quantum value lies strictly between the optimal local value and the maximum algebraic value. Figure 2.4 reveals a more complex story: even in this simple two g-bit case, the quantum achievability of correlations describes a convex set whose boundary traces a mysterious path between the local boundary and the non-signaling boundary; touching both at various points, but for the most part lying somewhere in the middle.

What governs the shape traced out by the quantum boundary? What aspects

of nature can elucidate its form, not just in the simple 2-dimensional slice of two g-bit systems illustrated above, but for any number of systems and measurement choices? A definitive answer to this question is not yet known, though various attempts have led to partial answers which shed light on the issue, notably the principles known as Macroscopic Locality [50] and Information Causality [51], and arguments based on the non-local computation of binary-valued functions [52]. The author's own contribution to this research is described in Chapter 4, in which an investigation of the Information Causality principle yields two distinct perspectives on why the quantum boundary takes the shape that it does.

2.5 Convex vector space representations

In Section 2.4 we described a natural embedding of the states of a general probabilistic theory into a vector space. This method is just one possible (and usually inefficient) vector space representation of a system, although it does serve the useful purpose of illustrating that such a representation always exists, and can be realised in a real, finite-dimensional vector space as long as the set of fiducial measurements is finite. Note that a very elegant description of this form of finite-dimensional quantum theory has long been established: the well-known Bloch-sphere representation of a qubit system [53], in which a 2-dimensional density operator corresponds to a vector with modulus less than or equal to one. The Bloch-sphere representation differs from the embedding discussed in Section 2.4, in which each component is equal to a fiducial outcome probability; for example, the completely mixed state is represented by a three-dimensional zero vector in the Bloch-sphere, and by a six-dimensional vector with every entry equal to $\frac{1}{2}$ in the representation given in Section 2.4. Nevertheless, these representations are closely related, and describe the same basic set of quantum states.

In order to understand the structure of a given general probabilistic theory, and to compare it with quantum theory, it is beneficial to study vector space representations of that theory, thus describing it in a language closely resembling that of

quantum theory, about which we already have a good deal of intuition. By examining theories from this perspective, we can highlight those features which are central to a good description of real life experiments, and those features which differ markedly from quantum theory. It is often the case, for example, that quantum phenomena which appear special when contrasted with classical physics are partly or entirely generic in the set of general probabilistic theories; teleportation, entanglement-swapping and the impossibility of cloning are but a few such phenomena [31, 54].

On the other hand, many properties are particular to a select group of theories. One such property is the existence of a rich class of reversible transformations, such that a pure state of a system may be mapped to any other pure state by a reversible transformation [2, 55] (in quantum theory, by a unitary operator). Another such property is self-duality, in which the set of vectors representing states and the set of vectors representing effects are essentially the same [56, 57] (in quantum theory, these are the positive operators of the underlying Hilbert space). These properties are not at all obvious from the description of systems in terms of fiducial measurements, but become much more approachable using vector space representations. By skipping the black box scenario and constructing probabilistic models directly from their vector space representation, it is straightforward to find theories which share some but not all of the properties of quantum theory, so that conjectures concerning these properties can be more easily generated and tested (the so-called polygon models, which link non-locality and the property of strong self-duality, are a good example [36]). Exploring which features of general probabilistic theories are possessed by quantum theory helps to build our intuition about its capabilities, which is an invaluable step towards understanding its information processing potential.

Various constructions of general probabilistic theories and their vector space representations exist, each differing slightly from the others due to different operational assumptions, conventions or mathematical convenience. The treatment given here tries to be as general as possible, whilst retaining the ability to describe

classical, quantum and Boxworld theories on an equal footing so that their respective features can be highlighted and contrasted. In Section 2.5.1 we introduce the mathematical prerequisites for this framework, most of which belong to the realm of convex geometry (a similar and very readable introduction can be found in [37]). In Sections 2.5.2 and 2.5.3 we describe how one or more systems may be described via a real, finite-dimensional vector space. In Section 2.5.4 we then discuss how the notion of transformations can be introduced in this setting and demonstrate that unlike quantum theory, Boxworld does not permit entanglement to be reversibly generated. We then provide some concrete examples of these representations in Section 2.5.5.

It is worth mentioning that the convex vector space approach to general probabilistic theories is but one node in a network of efforts to put quantum theory on a more fundamental footing than is provided by the complex Hilbert space formalism. Other notable instances include the quantum logic approach pioneered by Piron, Foulis and Randall [58, 59], and the categorical approach pioneered by Abramsky and Coecke [60]. The former of these sets a precedent for analysing Hilbert-space quantum mechanics as a probability calculus with axioms differing from standard probability but of interest in its own right. The latter formulates a "categorical semantics" for quantum theory, in which quantum theory is interpreted as a dagger compact category, whose morphisms correspond to physical processes. Both these approaches are novel, and present a fascinating vantage point, but it can be argued that they lack the conservatism and intuitive physical appeal of the approach we now outline.

2.5.1 Convex geometry

Much of the discussion of general probabilistic theories relies on fundamental concepts arising in the geometry of convex subsets of real vector spaces. Our interest in convexity stems from a simple argument that any set of vectors corresponding to preparable states of a system must be convex: if the vectors s_1 and s_2 correspond to two distinct states in which the system may be prepared, then $ps_1 + (1-p)s_2$ is

the state corresponding to having prepared the system in state s_1 with probability p, and in state s_2 with probability $1-p$.

Definition. *Let C be a subset of $\mathbb{R}^n$.*

- *The **convex hull** of C is $conv(C) = \{\sum_{i=1}^{m} \lambda_i c_i : \lambda_i \geq 0, \sum \lambda_i = 1, c_i \in C\}$.*
- *The **conic hull** of C is $cone(C) = \{\sum_{i=1}^{m} \lambda_i c_i : \lambda_i \geq 0, c_i \in C\}$.*
- *C is **convex** if $C = conv(C)$, and C is a **cone** if $C = cone(C)$*

For example, the positive quadrant of $\mathbb{R}^n$ is convex and a cone. All cones are convex, however bounded convex sets, such as the unit ball, are convex without being a cone. In the Bloch-sphere, pure quantum states are assigned vectors on the surface of the sphere: these are the extreme points of the sphere. In order to generalise this notion to all general probabilistic theories it will be important to define extreme points of convex sets, and extreme rays of cones.

Definition. *Let $C \subset \mathbb{R}^n$ be convex;*

- *$x \in C$ is an **extreme point** if, whenever $c_1, c_2 \in C$ and $0 < \lambda < 1$ are such that $x = \lambda c_1 + (1-\lambda)c_2$, then $c_1 = c_2 = x$.*
- *Denote by $ext(C)$ the set of extreme points of C.*

Definition. *Let $C \subset \mathbb{R}^n$ be a cone;*

- *For $x \in \mathbb{R}^n$, the **ray** generated by x is defined $R_x = \{\lambda x : \lambda \geq 0\}$. For $x \in C$, R_x is an **extreme ray** if, whenever $c_1, c_2 \in cone(C)$ and $\lambda_1, \lambda_2 > 0$ are such that $\lambda_1 c_1 + \lambda_2 c_2 = x$, then c_1 and c_2 both belong to R_x. We say that x generates an extreme ray of C.*
- *Denote by $extray(C)$ the set of extreme rays of C.*

The following well-known Theorems concerning extreme points are very useful when discussing convex sets. Recall that in $\mathbb{R}^n$, C is compact iff it is closed and bounded.

Theorem 1 (originally due to Minkowski, see [61])**.** *Let $C \subset \mathbb{R}^n$ be compact and convex. Then,*

$$C = conv(ext(C))).$$

in particular, $ext(C)$ is non-empty, and its convex hull is the whole of C.

Theorem 2 (Carathéodory, [62])**.** *Let $C \subset \mathbb{R}^n$ be compact. Then $conv(C)$ is also compact.*

Theorem 3 (Carathéodory, [62])**.** *Let $P \subset \mathbb{R}^n$ and let $x \in conv(P)$. Then there exists a subset $P_x \subseteq P$ with $|P_x| \leq n+1$ such that $x \in conv(P_x)$.*

Clearly the unit sphere has infinitely many extreme points, whereas a tetrahedron or a square has only finitely many. Theories in which sets of states possess only finitely many extreme points are of interest since they are easy to define, and yet their non-locality properties are often non-trivial [36].

Definition. *A subset $C \subset \mathbb{R}^n$ is a **polytope** (resp. **polytopic cone**) if it is the convex (resp. conic) hull of finitely many points.*

Theorems 1 and 2 tell us that equivalently, a compact, convex subset $C \subset \mathbb{R}^n$ is a polytope if $ext(C)$ is finite. Note also that if P is a finite set of points, then $ext(conv(P)) \subset P$. Hence one way of defining a polytope is to give a finite set of points which makes up its extreme points - this is known as the V-representation of a polytope. Another standard way to define a polytope is by giving a finite set of half-spaces of $\mathbb{R}^n$ whose intersection is bounded - this is known as the H-representation of a polytope [63].

Definition. *A **half-space** of $\mathbb{R}^n$ is a set of the form $\mathcal{H} = \{v \in \mathbb{R}^n : \langle x, v\rangle \leq b\}$ for some $x \in \mathbb{R}^n$ and $b \in \mathbb{R}$. If $b = 0$ then $\mathcal{H}$ is a **linear half-space**.*

Theorem 4 (see [63]). *A compact subset $C \subset \mathbb{R}^n$ is convex if and only if it can be written as an intersection of half-spaces. It is a polytope if and only if it can be written as the intersection of a finite set of half-spaces.*

Definition. *Let C be any subset of $\mathbb{R}^n$. The **dual cone** C^* is the cone defined by*

$$C^* = \{x \in V : \langle x, y\rangle \geq 0 \; \forall y \in C\} \tag{2.33}$$

For example, the dual cones of rays are half-spaces, and vice versa. The concept of a dual cone becomes useful when talking about effects, and particularly when constructing states and effects of composite systems. Recall that $\mathbb{R}^n$ comes equiped with the standard, Euclidean inner product, and the corresponding Euclidean norm: this allows us to define a topology on $\mathbb{R}^n$ via the metric $d(x, y) = \|x - y\|^2$. In this context, C^* is a closed cone for any $C \subset V$ (even if C itself is neither a cone nor closed), and hence $(C^*)^*$ is a closed cone. Moreover, since every vector in C has non-negative inner product with every vector in C^*, we also have that $C \subseteq (C^*)^*$. If C itself is a closed cone then it turns out that $C = (C^*)^*$; this will be useful later in establishing duality relations between sets of states and effects.

Proposition 2. *Let C be a closed cone in $\mathbb{R}^n$. Then $(C^*)^* = C$.*

Proof. As argued above, clearly $C \subseteq (C^*)^*$. In order to prove that $(C^*)^* \subseteq C$ it suffices to show that $x \notin C \Rightarrow x \notin (C^*)^*$; equivalently, for $x \notin C$ there exists $z \in C^*$ such that $\langle z, x\rangle < 0$. Given $x \notin C$, the function

$$f : V \to \mathbb{R}$$
$$f(v) = \|x - v\|,$$

is continuous and bounded below by 0. Moreover, picking an arbitrary $c \in C$, the set $D = C \cap \{d \in \mathbb{R}^n : f(d) \leq f(c)\}$ is the intersection of two closed sets, and is bounded. Being closed and bounded in a finite-dimensional space, D is therefore compact. Hence f attains its minimum value in D at a point y: clearly $f(y)$ is the

minimum value of f over C also. Let $z = y - x$: either $y = 0$ or it must be that y is the closest non-zero point to z on the ray R_y, hence $\langle z, y \rangle = 0$. Moreover, since,

$$0 < \|y - x\|^2 = \langle y - x, y \rangle - \langle y - x, x \rangle = -\langle z, x \rangle, \tag{2.34}$$

it must also be that $\langle z, x \rangle < 0$. It remains to show that z is indeed a member of C^*, i.e. for all $v \in C$, $\langle z, v \rangle > 0$. Suppose for contradiction that $y' \in C$ satisfies $\langle z, y' \rangle < 0$. Define the vector,

$$y_\varepsilon = (1 - \varepsilon)y + \varepsilon y', \tag{2.35}$$

noting that for all $0 \leq \varepsilon \leq 1$, y_ε is a member of C. We will show that for sufficiently small ε, $\|y_\varepsilon - x\|^2 < \|y - x\|^2$, i.e. $f(y_\varepsilon) < f(y)$, contradicting the definition of y. More precisely, we demand

$$\varepsilon < \frac{2\langle z, y - y' \rangle}{\|y - y'\|^2}, \tag{2.36}$$

noting that both numerator and denominator are necessarily positive. By the Law of Cosines,

$$\begin{aligned} \|y_\varepsilon - x\|^2 &= \|y - x\|^2 + \|y - y_\varepsilon\|^2 - 2\langle y - x, y - y_\varepsilon \rangle \\ &= \|y - x\|^2 + \varepsilon^2 \|y - y'\|^2 - 2\varepsilon \langle z, y - y' \rangle \\ &< \|y - x\|^2, \end{aligned} \tag{2.37}$$

as desired. □

Definition. *Let C be a cone in $\mathbb{R}^n$. The **generalised inequality** associated with C is the partial order $\leq_C$ on the vectors in $\mathbb{R}^n$ given by $x \leq_C y$ if $\exists z \in C$ such that $x + z = y$.*

For example, if C is the cone defined by the positive quadrant of $\mathbb{R}^n$, then $\mathbf{x} \leq_c \mathbf{y}$ iff every component of $\mathbf{x}$ has a lesser or equal value than the same component of $\mathbf{y}$. The partial order induced by any cone C is transitive (in the sense that $x \leq_C y$ and

$y \leq_C z$ implies $x \leq_C z$), reflexive (in the sense that $x \leq_C x$) and antisymmetric (in the sense that $x \leq_C y$ and $y \leq_C x$ implies $x = y$). Conversely, any such partial ordering in $\mathbb{R}^n$ induces a cone $C = \{v \in \mathbb{R}^n : v \geq 0\}$, however this is not relevant to our discussion. Once a partial order is established, an analogue of an interval on the real line may be defined, *viz.*

Definition. *Let $C \subset \mathbb{R}^n$ be a cone. For any $x, y \in \mathbb{R}^n$, the **order interval** $[x, y]_C$ is the set $\{z \in \mathbb{R}^n : x \leq_C z \leq_C y\}$.*

For example, if C is the cone defined by the positive quadrant in $\mathbb{R}^n$, then the interval $[0, \mathbf{1}]$ between the zero vector and the vector with 1 in every component is the hypercube of vectors for which every component has a value between 0 and 1. On the other hand, the interval $[0, \mathbf{e}_1]$, where $\mathbf{e}_1$ is the first standard basis vector, is simply the line segment between 0 and $\mathbf{e}_1$. Clearly, some order intervals have a richer structure than others - the following definition characterises those vectors u which generate interesting order intervals $[0, u]$.

Definition. *Let $C \subset \mathbb{R}^n$ be a cone. $u \in C$ is an **order unit** of C if, for any $v \in \mathbb{R}^n$, there exists a $\lambda > 0$ such that $v \leq_C \lambda u$.*

In some sense an order unit provides a benchmark by which the "order magnitude" of any vector can be evaluated, in much the same way as the unit element 1 does on the real line. If u is an order unit in C then for all other $v \in C$, there exists a $\lambda > 0$ such that $\frac{1}{\lambda} v$ is contained in the interval $[0, u]$, hence this interval inherits something of the structure of the cone C. Not all cones possess an order unit: for example, a ray contains an order unit only in 1-dimensional space.

Proposition 3. *Let $C \subset \mathbb{R}^n$ be a closed cone. If u is an order unit in C^*, then $D = \{c \in C : \langle u, c \rangle = 1\}$ is a compact, convex subset of C.*

Proof. Since D is an intersection of closed and convex sets, it is closed and convex itself. It remains to show that D is bounded: suppose for contradiction that there exists a sequence $\{d_i\} \subset D$ such that $\|d_i\| \to \infty$. For any vector v, there exists a λ such that $v \leq_{C^*} \lambda u$, i.e. $\langle v, d \rangle \leq \lambda$ for all $d \in D$. The sequence of real numbers

$\frac{\langle d_1, d_i \rangle}{\|d_1\|\|d_i\|}$ therefore tends to zero as i tends to infinity, hence, defining $i_1 = 1$, for any $\varepsilon > 0$ there exists a positive integer i_2 such that $\frac{\langle d_{i_1}, d_j \rangle}{\|d_{i_1}\|\|d_j\|} < \varepsilon$ for all $j > i_2$.

Continuing iteratively, for any $\epsilon > 0$ it is possible to find a set of $n+1$ vectors $\{d'_1, \ldots, d'_{n+1}\}$ such that $\frac{\langle d'_i, d'_j \rangle}{\|d'_i\|\|d'_j\|} < \varepsilon$ for all $1 \leq i, j \leq n+1$. The angle between d'_i and d'_j is given by:

$$\theta_{ij} = \cos^{-1}\left(\frac{\langle d'_i, d'_j \rangle}{\|d'_i\|\|d'_j\|}\right) \tag{2.38}$$

Thus there exist sets of $n+1$ vectors whose pairwise angles can be taken to be arbitrarily close to $\frac{\pi}{2}$, contradicting the fact that we are in an n-dimensional Euclidean space. □

2.5.2 Abstract state spaces

We established in Section 2.4 that the states of any system with finitely many fiducial measurements may be represented by vectors in a real, finite-dimensional vector space. Abstract state spaces provide a broad framework for describing systems in general probabilistic theories in terms of real, finite-dimensional vector spaces, without first having to introduce canonical sets of fiducial measurements and the outcome distributions which correspond to allowed states. Whilst any system defined in terms of fiducial measurements admits at least one representation in this manner, the appeal of the abstract state space is that there is no preferred set of fiducial measurements. Rather, fiducial sets of measurements can be inferred from the description of state and effect vectors within the real vector space. Before defining an abstract state space, it is helpful to discuss how the mathematical machinery of Section 2.5.1 corresponds to operational assumptions about the system.

To begin with, some convex subset $\mathcal{S} \subset V$ corresponds to the set of preparable states of the system. By restricting attention to a vector subspace if necessary, the set of states can be assumed to span the entire vector space. We will also assume for convenience that the set of states is closed: a physical justification for this is that whenever $\{s_i\}_{i \geq 0}$ is a sequence of states such that $s_i \to s$ for some vector s, then

the physical properties of the states s can be implemented to arbitrary accuracy by preparing the states s_i, hence we may as well assume that s is a state.

The *state cone* $\mathcal{S}_+ = cone(\mathcal{S})$ is therefore a closed cone in V: this has a convex subset the *sub-normalised* states $\mathcal{S}_{\leq 1} = \{\lambda s : 0 \leq \lambda \leq 1, s \in \mathcal{S}\} = conv(\mathcal{S} \cup \{0\})$. Sub-normalised states play an important role in the discussion of measurement outcomes, and we wish them to be distinct from the set of allowed states, hence we demand that for all $s \in \mathcal{S}$ and $0 \leq \lambda < 1$, $\lambda s \notin \mathcal{S}$. This prohibits the zero vector from being a member of $\mathcal{S}$. This in turn implies that $\lambda\mathcal{S} \cap \mathcal{S} = \emptyset$ for all $\lambda \neq 1$, and that $\mathcal{S}_+ \neq V$. In fact the condition $\lambda s \notin \mathcal{S}$ does not lead to a loss of generality: if it is not satisfied, it is possible to add an extra dimension to the vector space in order to make it so, whilst keeping the property that $\mathcal{S}$ spans the whole of V.

An effect takes the form of a function $\hat{e}$ from the set of states to the interval $[0, 1]$, such that the value $\hat{e}(s)$ is the probability of the outcome corresponding to e occurring, given that the system is in state s. An effect function must respect the convexity of the set of states, i.e. for any states $s_i \in \mathcal{S}$ and $p_i \in [0, 1]$ satisfying $\sum_i p_i = 1$, $\hat{e}\left(\sum_i p_i s_i\right) = \sum_i p_i \hat{e}(s_i)$. Such a function is said to be *convex-linear* on $\mathcal{S}$. The physical motivation for convex-linearity is as follows: if a system is prepared in state s_i with probability p_i, then for any measurement involving e, the probability of the outcome corresponding to $\hat{e}$ must be the associated average over the probabilities that would be obtained for each preparation. It turns out that this condition is enough to extend $\hat{e}$ to a linear function on the whole of V.

Proposition 4. *Let $\mathcal{S} \subset V$ be a convex subset of a real vector space, such that span($\mathcal{S}$) $= V$ and $\lambda\mathcal{S} \cap \mathcal{S} = \emptyset$ for all $\lambda \neq 1$, and let W be a real vector space. Any convex-linear function $\hat{e} : \mathcal{S} \to W$ is uniquely extensible to a linear function $\hat{e} : V \to W$.*

Proof. Any linear function which is an extension of $\hat{e}$ must respect linear combi-

nations involving members of $\mathcal{S}$:

$$\hat{e}\left(\sum_i \lambda_i s_i\right) = \sum_i \lambda_i \hat{e}(s_i). \tag{2.39}$$

Since $span(\mathcal{S}) = V$, this defines a linear function on the whole of V, assuming that the function is in fact well-defined. To show that it is well-defined, we must show that if a vector v is decomposed in distinct ways as a linear combination of members of $\mathcal{S}$, then the above definition of $\hat{e}$ is independent of which combination is used. Suppose then that $v = \sum_i \lambda_i s_i = \sum_j \lambda'_j s'_j$, and firstly assume that the λ_i and λ'_j are all positive, i.e. $v \in cone(\mathcal{S})$. Letting $\sum_i \lambda_i = \Lambda$, we have $\frac{1}{\Lambda} v = \sum_i \frac{\lambda_i}{\Lambda} s_i \in \mathcal{S}$, since $\mathcal{S}$ is convex. By the same reasoning, $\frac{1}{\Lambda'} v \in \mathcal{S}$ where $\Lambda' = \sum_j \lambda'_j$, thus $\Lambda = \Lambda'$. Since

$$\hat{e}\left(\frac{1}{\Lambda} v\right) = \hat{e}\left(\sum_i \frac{\lambda_i}{\Lambda} s_i\right) = \hat{e}\left(\sum_j \frac{\lambda'_j}{\Lambda} s'_j\right), \tag{2.40}$$

is uniquely defined, so is $\hat{e}(v) = \Lambda \hat{e}\left(\frac{1}{\Lambda} v\right)$, hence $\hat{e}$ extends in a unique and well-defined way to the whole of $cone(\mathcal{S})$. Now suppose that some of the λ_i and λ'_j are negative; we may re-arrange the equality so that all coefficients are positive:

$$\sum_{\lambda_i > 0} \lambda_i s_i - \sum_{\lambda'_j < 0} \lambda'_j s_j = \sum_{\lambda'_j > 0} \lambda'_j s_j - \sum_{\lambda_i < 0} \lambda_i s_i. \tag{2.41}$$

Since the LHS and RHS are both elements of $cone(\mathcal{S})$, we have from our previous argument that,

$$\sum_{\lambda_i > 0} \lambda_i \hat{e}(s_i) - \sum_{\lambda'_j < 0} \lambda'_j \hat{e}(s_j) = \sum_{\lambda'_j > 0} \lambda'_j \hat{e}(s_j) - \sum_{\lambda_i < 0} \lambda_i \hat{e}(s_i). \tag{2.42}$$

Re-arranging again gives that $\sum_i \lambda_i \hat{e}(s_i) = \sum_j \lambda'_j \hat{e}(s'_j)$, which verifies that our linearly extended function is indeed well-defined on the whole of V. □

Therefore an effect function $\hat{e} : \mathcal{S} \to [0, 1]$ is associated with a unique vector e such that $\langle e, s\rangle = \hat{e}(s)$ for all states s, i.e. the set of effects generates an *effect cone* $\mathcal{E}_+ = \mathcal{S}_+^*$ which is dual to the state cone. It follows from Proposition 2 that also $\mathcal{S}_+ = \mathcal{E}_+^*$. From here on we will use the term "effect" both for a non-negative, convex-linear functional $\hat{e}$ on $\mathcal{S}$ and for the effect vector e which corresponds to that functional. We make an assumption at this point that any such effect vector e corresponds to a physically realisable effect. That is to say, there exists a measurement outcome whose probability of occurring is equal to $\langle e, s\rangle$ whenever that measurement is physically performed on a system in the state represented by s. The duality of the cones $\mathcal{E}_+$ and $\mathcal{S}_+$ is a direct result of this assumption. This duality property holds in finite-dimensional quantum theory, as we will see, but is difficult to justify on physical grounds; if the assumption is not made, then $\mathcal{S}_+$ and $\mathcal{E}_+$ must be defined separately, and many of the results concerning multipartite systems will not follow, or would require modification.

A measurement in an abstract state space consists of a set of effects $\{e_1, \ldots, e_n\}$ such that $\sum_{i=1}^n \langle e_i, s\rangle = 1$ for all $s \in \mathcal{S}$. This implies that the vector $\mathcal{U} = \sum_i e_i$ is the vector uniquely defined by linear extension of the function which maps all vectors in $\mathcal{S}$ to 1. Just as we assume that all valid effect vectors are physically realisable, we assume that any set of effects which sum to give $\mathcal{U}$ corresponds to a possible measurement that can physically be performed (again, this is in principle true in finite-dimensional quantum theory). The vector $\mathcal{U}$ corresponds to the *unit effect*, which itself can be seen as a trivial measurement with only one outcome. Note that all physically realisable effects must therefore obey $e \leq_{\mathcal{E}_+} \mathcal{U}$, where "$\leq_{\mathcal{E}_+}$" is the generalised inequality associated with $\mathcal{E}_+$.

Definition. *An **abstract state space** is a 3-tuple $(V, \mathcal{S}_+, \mathcal{U})$, where V is a real, finite-dimensional vector space, $\mathcal{S}_+ \subset V$ is a closed cone which spans V, and $\mathcal{U}$ is an order unit in $\mathcal{S}_+^*$.*

Note that the set of normalised states is not an explicit part of the definition, rather it is defined as the set $\mathcal{S} = \{s \in \mathcal{S}_+ : \langle \mathcal{U}, s\rangle = 1\}$, which by Proposition 3 is compact and convex. The effect cone is also not explicitly mentioned, but is

implicitly the dual cone to $\mathcal{S}_+$. By analogy with normalised states, $e \in \mathcal{E}_+$ is said to be a *proper* effect if $\langle e, s \rangle \leq 1 \ \forall s \in \mathcal{S}$, and we denote by $\mathcal{E}$ the set of proper effects. Note that $\mathcal{E}$ can equivalently be defined as the order interval $[0, \mathcal{U}]_{\mathcal{E}_+}$. The elements of $\mathcal{E}_+ \setminus \mathcal{E}$ are said to be *improper* effects.

Definition. *Let $(V, \mathcal{S}_+, \mathcal{U})$ be an abstract state space. Then,*

- *The **pure states** are those that belong to $ext(\mathcal{S})$;*
- *The **pure effects** are those that belong to $ext([0, \mathcal{U}]_{\mathcal{E}_+})$;*
- *The **extreme ray effects** are those that generate rays in $extray(\mathcal{E}_+)$;*
- *A **pure measurement** is one for which the outcomes correspond to distinct extreme ray effects.*

The condition $\lambda \mathcal{S} \cap \mathcal{S} = \emptyset$ for $\lambda \neq 1$ ensures that the pure states are exactly those states which generate the extreme rays of $\mathcal{S}_+$, hence there is no need to distinguish between the extreme vectors of $\mathcal{S}$ and extreme rays of $\mathcal{S}_+$. As we will see in Section 2.5.5 however, there often exist pure effects which are not extreme ray effects. Pure measurements are those that provide maximal information about the system, in the sense that the outcome statistics of any measurement may be deduced from the outcome statistics of a related pure measurement. To see this, note that any set of effects lies in the conic hull of the extreme ray effects, so it is always possible to "split up" an outcome into a disjoint union of outcomes corresponding to extreme ray effects. We can define this "splitting up" procedure more rigorously, and equivalently define the pure measurements as those which admit no non-trivial refinements.

Definition. *Let $\mathcal{M}_1 = \{e_1, \ldots, e_n\}$ and $\mathcal{M}_2 = \{f_1, \ldots, f_m\}$ be two measurements on an abstract state space. $\mathcal{M}_1$ is a **refinement** of $\mathcal{M}_2$ if there exists a surjective mapping $\phi : [n] \to [m]$ such that,*

$$f_i = \sum_{j : \phi(j) = i} e_j \quad \forall i. \tag{2.43}$$

*Conversely, $\mathcal{M}_2$ is a **coarse-graining** of $\mathcal{M}_1$. A **trivial refinement** is one for which,*

$$\phi(j) = i \iff e_i = \lambda f_j. \tag{2.44}$$

An abstract state space may be defined in two equivalent ways: firstly, by giving the vector space V, the state cone $\mathcal{S}_+$ and an order unit $\mathcal{U}$ of $\mathcal{S}_+^*$, from which the sets $\mathcal{E}_+$ and $\mathcal{E}$ can be deduced. Secondly, by giving the vector space V, the effect cone $\mathcal{E}_+$ and order unit $\mathcal{U} \in \mathcal{E}_+$, from which $\mathcal{S}_+$ is the dual cone to $\mathcal{E}_+$ and $\mathcal{S}$ is the compact, convex subset of $\mathcal{S}_+$ having unit inner product with $\mathcal{U}$. These complementary methods allow us to construct lower and upper bounds on the set of states of composite systems.

2.5.3 Composite Systems

In order to construct a composite system in terms of its subsystems, we must define the vector space, state (or effect) cone and unit effect that characterise its abstract state space. Recall that the principle of local tomography tells us that any joint state is characterised by the outcome statistics of joint fiducial measurements, i.e. the normalised set of conditional probabilities $P(a_1, \ldots, a_N | x_1, \ldots, x_N)$. However, if the systems are not Boxworld or classical systems, we must be careful that the reduced states of an individual system are in fact allowed on that system, i.e. $\sum_{a_j : j \neq i} P(\mathbf{a}|\mathbf{x})$ is an allowed state of system i, whatever the choice of $\{x_j : j \neq i\}$. As we will shortly see, it turns out that all possible joint states can be neatly represented in the tensor product of the vector spaces representing each individual system.

Definition. *Let $C^{(1)}, \ldots, C^{(N)}$ be cones in the real, finite-dimensional vector spaces $V^{(1)}, \ldots, V^{(N)}$ respectively. The **tensor cone** $C^{(1)} \tilde{\otimes} \cdots \tilde{\otimes} C^{(N)}$ is the cone in $V^{(1)} \otimes \cdots \otimes V^{(N)}$ defined by $cone(\{c^{(1)} \otimes \cdots \otimes c^{(N)} : c^{(i)} \in C^{(i)}\})$.*

Let $\{(V^{(i)}, \mathcal{S}_+^{(i)}, \mathcal{U}^{(i)})\}_{i=1}^N$ be a set of abstract state spaces representing systems in some general probabilistic theory, and consider the N-fold tensor product space

$V_N = \bigotimes_{i=1}^N V^{(i)}$. Suppose that on each of the systems, the state $s^{(i)}$ is prepared: then it is natural to define the global state corresponding to this joint preparation as the product state $s^{(1)} \otimes \cdots \otimes s^{(N)}$. Moreover, instead of being prepared independently, the systems may agree to prepare states according to some shared classical correlation, i.e. there is a shared probability distribution $\{p_j\}$ such that with probability p_j, the state $s_j^{(i)}$ is prepared on subsystem i. The global state after this preparation is then a convex combination of product states:

$$\sum_j p_j s^{(1)} \otimes \cdots \otimes s_j^{(N)} \tag{2.45}$$

The cone generated by all convex combinations of product states is the tensor cone $\mathcal{S}_+^{min} = \mathcal{S}_+^{(1)} \tilde{\otimes} \cdots \tilde{\otimes} \mathcal{S}_+^{(N)}$. Whatever construction we employ for the set of states of the composite system, it is physically reasonable to demand that the composite state at least contains $\mathcal{S}_+^{min}$. Define $\mathcal{U}_N = \mathcal{U}^{(1)} \otimes \cdots \otimes \mathcal{U}^{(N)}$, and note that this has unit inner product with the convex hull of product states. Note that the set of states $\mathcal{S}^{(i)}$ on each system i span the vector space $V^{(i)}$, therefore the set of product states spans V_N, and $\mathcal{U}_N$ is the unique vector with unit inner product on all product states. We therefore take $\mathcal{U}_N$ to be the N-fold unit effect.

Designating the set of allowed composite states to be $\mathcal{S}^{min} = \{s \in \mathcal{S}_+^{min} : \langle \mathcal{U}_N, s \rangle = 1\}$ defines a perfectly valid composite system. Measurements on the composite system then correspond to sets of effects belonging to the effect cone $\mathcal{E}_+^{min} = \left(\mathcal{S}_+^{min}\right)^*$, and the proper effects are those belonging to the order interval $[0, \mathcal{U}_N]_{\mathcal{E}_+^{min}}$.

Definition. *Let $\{(V^{(i)}, \mathcal{S}_+^{(i)}, \mathcal{U}^{(i)})\}_{i=1}^N$ be a set of abstract state spaces. The **minimal tensor product** or (min tensor product) of systems $1, \ldots, N$ is represented by the abstract state space*

$$V^{min} = \left(\bigotimes_{i=1}^N V^{(i)}, \mathcal{S}_+^{min}, \mathcal{U}_N \right) \tag{2.46}$$

Notice that convex combinations of product states do not allow for any notion

of entanglement or non-locality, both of which are present in quantum theory and Boxworld, so there must in general be larger consistent sets of states that can be prepared on composite systems. Clearly, any composite state s should obey the following properties:

$$\langle e^{(1)} \otimes \cdots \otimes e^{(N)}, s\rangle \geq 0 \ \ \forall e^{(i)} \in \mathcal{E}_+^{(i)} \tag{2.47}$$

$$\langle \mathcal{U}_N, s\rangle = 1. \tag{2.48}$$

In quantum and classical theory, composite states have well-defined reduced states, which determine the local outcome statistics of local measurements, regardless of operations performed on the remaining subsystems. Suppose that a joint measurement is made on a state s, and the outcomes on systems $2, \ldots, N$ correspond to the effects $e^{(2)}, \ldots, e^{(N)}$. In principle, the choice of measurement at system 1 may be made after these measurements occur; the reduced state $\tilde{s}^{(1)} \in \mathcal{S}_{\leq 1}^{(1)}$ *relative* to the fixed outcomes $e^{(2)}, \ldots, e^{(N)}$ is such that for any $e^{(1)} \in \mathcal{E}_+^{(1)}$,

$$\langle e^{(1)}, \tilde{s}^{(1)}\rangle = \langle e^{(1)} \otimes e^{(2)} \otimes \cdots \otimes e^{(N)}, s\rangle. \tag{2.49}$$

This defines a unique vector $\tilde{s}^{(1)} \in V$ which, since the RHS is strictly between 0 and 1, must lie in $\mathcal{S}_{\leq 1}^{(1)}$. $\tilde{s}^{(1)}$ can be interpreted as the state of subsystem 1 post-measurement, conditioned on the outcomes corresponding to effects $e^{(2)}, \ldots, e^{(N)}$ being recorded at subsystems $2, \ldots, N$.

In order to normalise the state $\tilde{s}^{(1)}$, it must be divided by the positive real number $\langle \mathcal{U}^{(1)}, \tilde{s}^{(1)}\rangle$: this value can be interpreted as the marginal probability of obtaining the outcomes $e^{(2)}, \ldots, e^{(N)}$ for any global measurement which includes the outcomes corresponding to these effects on each subsystem. Suppose that we complete this global measurement using an additional set of effects on each subsystem $i \geq 2$ which, when summed together with $e^{(i)}$, give the unit effect. Then the average over the resulting (normalised) reduced states, relative to each possible outcome on subsystems $2, \ldots N$, and weighted according to their marginal

probabilities, is the unique state $s^{(1)} \in \mathcal{S}^{(1)}$ such that for any $e^{(1)} \in \mathcal{E}_+^{(1)}$,

$$\langle e^{(1)}, s^{(1)} \rangle = \langle e^{(1)} \otimes u^{(2)} \otimes \cdots \otimes u^{(N)}, s \rangle, \tag{2.50}$$

known as the *reduced state* on subsystem 1. This reduced state provides a full description of the outcome statistics for a local measurement at subsystem 1, given that the state of the global system is s. As a consequence of our tensor product construction, any allowed state satisfying (2.47) and (2.48) also leads to reduced states which are genuine states on each individual subsystem. This set of states may be explicitly constructed by defining the cones of effects and states in the following way,

$$\begin{aligned} \mathcal{E}_+^{max} &= \mathcal{E}_+^{(1)} \hat{\otimes} \cdots \hat{\otimes} \mathcal{E}_+^{(N)} \\ \mathcal{S}_+^{max} &= \left(\mathcal{E}_+^{max} \right)^*, \end{aligned} \tag{2.51}$$

thus defining $\mathcal{S}^{max} = \{ s \in \mathcal{S}_+^{max} : \langle u_N, s \rangle = 1 \}$ as the largest possible set of states for the composite system.

Definition. *Let $\{(V^{(i)}, \mathcal{S}_+^{(i)}, u^{(i)})\}_{i=1}^N$ be a set of abstract state spaces. The **maximal tensor product** or (max tensor product) of systems $1, \ldots, N$ is represented by the abstract state space*

$$V^{max} = \left(\bigotimes_{i=1}^{N} V^{(i)}, \mathcal{S}_+^{max}, u_N \right) \tag{2.52}$$

As well as describing an abstract state space for each type of individual system, a general probabilistic theory must provide a description of how those systems may combine into composite systems. For any set $\{(V^{(i)}, \mathcal{S}_+^{(i)}, u^{(i)})\}_{i=1}^N$ of permitted abstract state spaces, the theory must offer a composite state cone $\mathcal{S}_+^{min} \subseteq (\mathcal{S}_+)_N \subseteq \mathcal{S}_+^{max}$, or a composite effect cone $\mathcal{E}_+^{max} \subseteq (\mathcal{E}_+)_N \subseteq \mathcal{E}_+^{min}$.

2.5.4 Transformations

Quantum theory is not just a static description of what kinds of measurements can be performed, and what kind of outcome statistics we can expect from those measurements. It also provides an explicit model of how one state may transform into another. These transformations are described by unitary operators on the Hilbert space of the system, or at least of a larger system containing it. Any transformation, indeed any evolution of part of the universe as portrayed by quantum mechanics, is described by a reversible transformation from the perspective of anyone who regards that part of the universe as a closed system.

Transformations also fit naturally into the abstract state space framework described in Sections 2.5.2 and 2.5.3. Suppose that $(V, \mathcal{S}_+, \mathcal{U})$ is an abstract state space, and consider a mapping T from $\mathcal{S}$ into itself (i.e. a transformation from allowed states to allowed states). Recall from Section 2.5.2 that effects must be convex-linear on $\mathcal{S}$ in order to respect probabilistic preparations of states. Much the same reasoning can be applied here: the transformation acting on the state $ps_1 + (1-p)s_2$ is expected to have the operational properties of the corresponding average of transformed states, i.e. $T(ps_1 + (1-p)s_2) = pT(s_1) + (1-p)T(s_2)$. It follows from Proposition 4 that T can be seen as a linear map from V to itself, which also maps $\mathcal{S}$ into itself. An *allowed transformation* is any such linear map: just as we assume that any linear functional which maps states to the interval $[0, 1]$ corresponds to a physically realisable effect, we assume that any allowed transformation may be realised by some physical operation on the system.

Suppose now that a system is comprised of N subsystems, with vector spaces $V^{(1)}, \ldots, V^{(N)}$ admitting state spaces $\mathcal{S}^{(1)}, \ldots, \mathcal{S}^{(N)}$, and that an allowed transformation $T^{(i)}$ is locally performed on each subsystem i. We would expect that there is an allowed global transformation T of the composite system which occurs as a result of these local transformations being performed. Consider the effect of this global transformation on a product state $s^{(1)} \otimes \cdots \otimes s^{(N)}$, and then performing a measurement locally on each subsystem. The state which gives the correct outcome probabilities for any local measurement is $T^{(1)}(s^{(1)}) \otimes \cdots \otimes T^{(N)}(s^{(N)})$.

$T = T^{(1)} \otimes \cdots \otimes T^{(N)}$ is the unique linear extension to $\bigotimes V^{(i)}$ of the map which acts in this manner on the set of product states. "Product transformations" of this kind must therefore be allowed transformations of the composite system: observe that this will always hold in min- and max- tensor product spaces, since product transformations map the sets $\mathcal{S}^{min}$ and $\mathcal{S}^{max}$ into themselves. Note, if a local transformation is performed on only a subset of the N subsystems, then we consider the identity transformation $\mathbb{I}^{(i)}$ to have been performed on the remaining subsystems.

In quantum theory, one may view transformations as acting on measurement operators rather than states, by moving to the Heisenberg picture. An analogous trick may be performed here. Since, in the operational perspective, a state is defined by its outcome statistics, the action of a transformation is defined by how it influences the inner product between states and effects, i.e. T is determined by the set of values $\langle e, T(s)\rangle$. Recall that the adjoint $T^\dagger$ of a linear map T is the linear map for which satisfies $\langle T^\dagger(x), y\rangle = \langle x, T(y)\rangle$ for all $x, y \in V$. Thus T is determined equivalently by its adjoint transformation on effects: note that if T is an allowed transformation then $T^\dagger$ maps the effect cone $\mathcal{E}_+$ to itself.

Proposition 5. *Let $(V, \mathcal{S}_+, \mathcal{U})$ be an abstract state space, and T an allowed transformation. Then $T^\dagger(\mathcal{U}) = \mathcal{U}$.*

Proof. By definition of the adjoint, we have that for all states $s \in \mathcal{S}$,

$$\langle T^\dagger(\mathcal{U}), s\rangle = \langle \mathcal{U}, T(s)\rangle = 1. \tag{2.53}$$

Since $\mathcal{S}$ spans V, any two linear functionals which are equivalent on $\mathcal{S}$ are equivalent on the whole of V. $T^\dagger(\mathcal{U})$ has the same inner product as $\mathcal{U}$ does for all $s \in \mathcal{S}$, thus $T^\dagger(\mathcal{U}) = \mathcal{U}$. □

An allowed transformation is *reversible* if there exists an allowed transformation S such that $TS(s) = s$ for all states s in $\mathcal{S}$. This implies that $S = T^{-1}$ when viewed as linear maps on the whole of V. If a transformation is reversible, then so is its adjoint, which satisfies $\left(T^\dagger\right)^{-1} = (T^{-1})^\dagger$. The property of being

reversible imposes strong restrictions on T. Clearly, for example, it must be a bijective mapping from $\mathcal{S}$ to itself, otherwise T^{-1} will not be a well-defined allowed transformation. As the following Proposition demonstrates, this means that the set of pure states must be permuted by T.

Proposition 6. *Let $(V, \mathcal{S}_+, \mathcal{U})$ be an abstract state space, and T an allowed, reversible transformation. Then T maps pure states to pure states, $T^\dagger$ maps pure effects to pure effects and $T^\dagger$ maps extreme ray effects to extreme ray effects.*

Proof. Suppose that $T(s) = ps_1 + (1-p)s_2$, where $0 < p < 1$ and $s_1, s_2 \in \mathcal{S}$. T has a linear inverse T^{-1}, so we have that $s = pT^{-1}(s_1) + (1-p)T^{-1}(s_2)$. Since s is a pure state, this implies $T^{-1}(s_1) = T^{-1}(s_2) = s$, hence $s_1 = s_2 = T(s)$. Therefore $T(s)$ is also a pure state. The proof that $T^\dagger$ maps pure effects to pure effects is identical, and the proof that $T^\dagger$ maps extreme ray effects to extreme effects is very similar (it simply involves switching the convex combination for a conic combination). □

2.5.5 Examples

So far we have introduced machinery for dealing with abstract state space representations of general probabilistic theories, and derived results about their structure in this abstract setting; we now supplement this with concrete examples. We begin by recapping and constructing the abstract state spaces for quantum theory, classical theory and Boxworld, then describe polygon models introduced in [36], whose state spaces take the shape of regular polygons.

Quantum Theory

In the standard presentation, quantum theory is not introduced by means of a real vector space, but instead by means of a complex Hilbert space $\mathcal{H}$, in which the states are given by density operators and measurements by sets of POVM elements. However, the set of density operators form a convex subset of the *real* vector space $herm\,(\mathcal{H})$ of Hermitian operators over $\mathcal{H}$. Defining $V = herm\,(\mathcal{H})$ to be our

vector space, and $\mathcal{S}_+ = \mathcal{P}(herm(\mathcal{H}))$ to be the cone of positive operators on $\mathcal{H}$ (i.e. Hermitian operators P for which $\langle x|P|x\rangle > 0$ for all non-zero $x \in \mathcal{H}$), we almost have a full abstract state space. Before defining $\mathcal{U}$, we need to have an inner product on V: this is provided by the Hilbert-Schmidt inner product on linear operators, $\langle A, B\rangle = \mathrm{Tr}\left(A^\dagger B\right)$. With this inner product, the unit effect is simply the identity $\mathbb{I}$, and the cone of effects is identical to the cone of states. The proper effects are positive operators E for which $E \leq \mathbb{I}$ in the operator norm.

The convex set of normalised states $\mathcal{S}$ is the set $\{\rho \in \mathcal{P}(\mathcal{H}) : \mathrm{Tr}(\rho) = 1\}$ (i.e. those that have unit inner product with $\mathbb{I}$). Pure states are given by the set of projectors onto 1-dimensional subspaces. Pure effects are given by projectors of any rank, however extreme ray effects are again given by the set of projectors onto 1-dimensional subspaces (this is in agreement with the fact that $\mathcal{E}_+ = \mathcal{S}_+$, and our previous comment that the pure states generate the extreme rays of $\mathcal{S}_+$). Observe that whenever $\mathcal{H}$ is of dimension larger than 2, there will exist infinitely many pure effects which are not extreme ray effects. Reversible transformations are maps $\rho \to U\rho U^\dagger$, where U is a unitary operator. As expected, these map the cone of positive operators bijectively to itself, and pure states onto pure states. The adjoint mapping also takes pure effects to pure effects, and extreme ray effects to extreme ray effects.

The set of states of composite quantum systems, the unit-trace positive operators in the tensor product of the Hilbert spaces, lies strictly between the sets of states described by the min and max tensor products. The states of the min tensor product are the separable states, i.e. those which can be written as a convex combination of pure product states: this also permits effects that are not allowed in standard quantum theory, for example entanglement witnesses, which give positive outcomes for any product state, but are negative for some entangled states. On the other hand, composite quantum systems with the max tensor product include states not allowed in the standard formulation, for example the partial transpose of many entangled states generates an operator which is positive for tensor products of POVM elements, but has global negative eigenvalues.

Classical Theory

Recall that classical systems are those for which there is a single fiducial measurement, over which any outcome distribution corresponds to an allowed state. For a fiducial measurement with K outcomes, this can be represented in $\mathbb{R}^K$ by setting the fiducial effects to be the standard basis vectors $\{e_1, \ldots, e_K\}$. These are the extreme ray effects, hence the effect cone is the positive quadrant. This implies that the ith component of a vector representing a state is the probability of obtaining outcome i when the fiducial measurement is performed on the system in that state. Thus all state vectors have unit inner product with the unit effect, which is the vector containing 1 in every component. The set of normalised states forms a simplex, whose vertices are also the standard basis vectors. The pure state e_i with a 1 in the ith component (and 0 elsewhere) is the state which deterministically assigns outcome i when the measurement is performed. Like quantum theory, the state and effect cones are identical.

Reversible transformations of a classical system must permute the fiducial effects, which can be interpreted as a "relabelling" of the measurement outcome once the measurement has been performed. Note that any such transformation maps the standard basis to itself, and hence extends to a valid linear map on $\mathbb{R}^K$.

The min and max tensor products of composite classical systems coincide, and thus there is only one reasonable construction of composite states and effects: the only allowed states are classically correlated, and there is no entanglement. To see this, it is enough to observe that the only composite product effect with which the pure product state $e_{i_1}^{(1)} \otimes \cdots \otimes e_{i_N}^{(N)}$ has non-zero product, is in fact the effect given by the same vector. Hence any linear combination of product effects must have only positive coefficients in order to have positive inner product with all pure product states. Since the product effects span the tensor product space, the only possible composite effects are given by $\mathcal{E}_+^{max}$, which must therefore be equal to $\mathcal{E}_+^{min}$.

Boxworld

In a Boxworld system it is again simplest to construct the effect cone first, then obtain the state space via duality. For a system with M measurements, for which measurement x has K_x outcomes, recall that any set of distributions on the outcomes for each measurement is an allowed state. This can be encapsulated in a real vector space V of dimension $d = 1 + \sum_x (K_x - 1)$. Choose a linearly independent set of vectors $\{\mathcal{U}\} \cup \{X_{a|x}\}$ for $1 \leq x \leq M$ and $1 \leq a \leq K_x - 1$, where the vectors of the form $X_{a|x}$ are fiducial effects, corresponding to outcome a occurring when fiducial measurement x is performed. These vectors form a basis for V. The remaining fiducial effect vectors are then restricted by the fact that summing the effects of a fiducial measurement must give the unit effect:

$$X_{K_x|x} = \mathcal{U} - \sum_{a=1}^{K_x-1} X_{a|x}. \tag{2.54}$$

For example, a g-bit system can be represented in a 3-dimensional real vector space, where $\mathcal{U}, X_{0|0}$ and $X_{0|1}$ take the place of the standard basis vectors, so that e.g. $X_{1|1} = (1, 0, -1)$. Any vector s which has a positive inner product with all fiducial effect vectors and unit inner product with the unit effect $\mathcal{U}$ corresponds to the state with outcome distribution $P(a|x) = \langle X_{a|x}, s\rangle$. These inner products always have a value at most one since $\langle \mathcal{U}, s\rangle = 1$. Pure states are those which have inner product only 0 or 1 with all fiducial effects: these deterministically assign a single outcome to each choice of measurement (for example, the state which always produces outcome 1 is a pure state).

Min tensor product Boxworld states are convex combinations of tensor products of these pure states. There exist effects in $\mathcal{E}_+^{min} \setminus \mathcal{E}_+^{max}$ which are positive for all such states: for example, the effect,

$$X_{1|0} \otimes X_{1|0} + X_{0|0} \otimes X_{1|1} + X_{1|1} \otimes X_{0|0} - X_{1|1} \otimes X_{1|1}, \tag{2.55}$$

gives inner product 0 or 1 for all pure product states, but is not a positive linear

combination of tensor products of fiducial effects. However, for a variant of the PR-box state, in which the outcomes are exactly correlated (with probability $1/2$ each for both being 0 or both being 1) whenever both inputs are equal to 0, but anti-correlated (again with probability $1/2$ for the two possibilities) otherwise, the probability of obtaining the above effect is $3/2$, which is clearly not permitted.

Min tensor product Boxworld outcome distributions clearly cannot violate the CHSH inequality, and are not often the subject of much interest. When referring to composite Boxworld systems, it is usually assumed that the max tensor product is being used; this is the convention that we will follow in this thesis. A Boxworld state on a composite system - whose N subsystems have abstract state spaces $(V^{(i)}, \mathcal{S}_+^{(i)}, \mathcal{U}^{(i)})$ - is defined by a non-signaling outcome distribution $P(\mathbf{a}|\mathbf{x})$, and corresponds to a unique vector $s \in \mathcal{S}^{max}$, such that $\langle \mathcal{U}, s\rangle = 1$ (i.e. the state is normalized) and $P(a_1, \ldots, a_N|x_1, \ldots, x_N) = \langle X^{(1)}_{a_1|x_1} \otimes \cdots \otimes X^{(N)}_{a_N|x_N}, s\rangle$ [4]. Pure product states are those of the form $s^{(1)} \otimes \cdots \otimes s^{(N)}$ where $s^{(i)}$ is a deterministic state on subsystem i. For a system of two g-bits, all pure states are either pure product states or variants of the PR-box; we do not yet have a full understanding of the pure states of composite Boxworld systems, although computational enumerations have been carried out for some cases [64, 65].

In quantum theory, the Pauli measurements are not the only measurements that can be made on a qubit system, despite constituting a fiducial set of measurements; any positive semi-definite operator $P \leq \mathbb{I}$ corresponds to an outcome of some carefully constructed measurement. Similarly, in Boxworld the fiducial measurements are not the only measurements that can be performed, and the fiducial effects are not the only vectors which correspond to measurement outcomes. However, since the fiducial measurements fully characterise a state s, the probability of obtaining any non-fiducial effect must be a function of the values $P(a_1, \ldots, a_N|x_1, \ldots, x_N)$ associated with s. The set of allowed N-partite effects in Boxworld is the set of vectors $e \in \bigotimes V^{(i)}$ such that $\langle e, s\rangle \in [0, 1]$ for all $s \in \mathcal{S}$. This includes the N-partite fiducial effects, as well as all linear combinations of the fiducial effects that are positive and sub-normalised for all states. These non-fiducial effects can in

principle be measured by ignoring particular measurement outcomes, forgetting which of several outcomes occurred, or performing measurements probabilistically.

The same "relabelling"' idea introduced for classical systems can be employed in Boxworld to generate reversible transformations. For example, on a single system, relabelling by switching the measurement choices x and x' corresponds to the permutation of fiducial effects $X_{a|x} \leftrightarrow X_{a|x'}$ for all a (we must have $K_{x'} = K_x$ for this to work). Relabelling by switching the measurement outcomes a and a' for the same measurement corresponds to a single switch of the fiducial effect $X_{a|x}$ with $X_{a'|x}$. Since these transformations can be validly defined on the basis made up of $\mathcal{U}$ and a subset of the fiducial effects, they extend to linear maps on the whole vector space. Chapter 5 is concerned with proving that all allowed, reversible Boxworld transformations are composed of these local relabellings, as well as permutations of subsystems, so long as no subsystem is classical. This is no easy task; however, it is possible to prove relatively easily that in Boxworld, reversible transformations map pure product states to pure product states.

Proposition 7. *Let $\{(V^{(i)}, \mathcal{S}_+^{(i)}, \mathcal{U}^{(i)})_{i=1}^N\}$ be subsystems of a composite Boxworld system. Let T be a reversible transformation on the composite system. Then for any pure product state s, $T(s)$ is also a pure product state.*

Proof. This proof closely follows that given in [4]. Recall that the adjoint $T^\dagger$ must permute the extreme rays of the effect cone: for a Boxworld system these are the product fiducial effects on each system. Hence $\langle e, T(s)\rangle = \langle T^\dagger(e), s\rangle \in \{0, 1\}$ for any product fiducial effects e.

A joint local measurement involving measurement choices x_i on subsystem i comprises all product fiducial effects of the form $X_{a_1|x_1} \otimes \cdots \otimes X_{a_N|x_N}$, where each a_i ranges from 1 to K_{x_i}. Since the sum of these vectors is $\mathcal{U}_N = \mathcal{U}^{(1)} \otimes \cdots \otimes \mathcal{U}^{(N)}$, $T(s)$ must have inner product 1 with exactly one such vector, and zero with the others. Thus for each string of measurement choices $(x_1, \ldots, x_N)$, $T(s)$ deterministically assigns some string of outcomes $(a_1, \ldots, a_N)$.

Suppose now that for two measurement strings **x** and **x**′ with $x_i = x_i'$ for any fixed i, $T(s)$ assigns the outcome strings **a** and **a**′ for which $a_i \neq a_i'$. This implies that $T(s)$ has inner product 1 with both the effects,

$$\mathcal{U}^{(1)} \otimes \cdots \otimes X^{(i)}_{a_i|x_i} \otimes \cdots \otimes \mathcal{U}^{(N)}$$
$$\mathcal{U}^{(1)} \otimes \cdots \otimes X^{(i)}_{a_i'|x_i} \otimes \cdots \otimes \mathcal{U}^{(N)}.$$

However, these distinct effects are both present in the same decomposition of $\mathcal{U}_N$:

$$\mathcal{U}_N = \sum_a \mathcal{U}^{(1)} \otimes \cdots \otimes X^{(i)}_{a|x_i} \otimes \cdots \otimes \mathcal{U}^{(N)},$$

thus $T(s)$ has inner product at least 2 with $\mathcal{U}_N$, contradicting the assumption that T takes normalised states to normalised states. □

Polygon Models

Polygon models, first introduced in [36], are specifically constructed so that the normalized state space of a single system takes the shape of a regular n-gon (or n-sided polygon) in 3-dimensional real Euclidean space. In contrast to the above theories, whose state spaces *result* from mathematical restrictions on the outcome distributions, polygon systems are *defined* by their state space, and their outcome distributions follow accordingly. Note that as n becomes large, the state space begins to resemble a circle, which itself represents a "2-dimensional slice" of the Bloch sphere of a qubit, in which only the σ_x and σ_z measurements are considered. An analysis of polygon models for varying n potentially gives us insight into the significance of the shape of the quantum local state space for phenomena such as non-locality.

For fixed n, the normalized state space $\mathcal{S}$ of the n-gon system is the convex hull of the extremal states $\omega_1, \ldots, \omega_n$, defined by:

$$\omega_n = \begin{pmatrix} r_n \cos\left(\frac{2i\pi}{n}\right) \\ r_n \cos\left(\frac{2i\pi}{n}\right) \\ 1 \end{pmatrix}, \tag{2.56}$$

where $r_n = \sqrt{\sec\left(\frac{\pi}{n}\right)}$. Recall that the unit effect is the unique vector whose inner product with all states is equal to 1. Clearly, this is the vector:

$$\mathcal{U} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{2.57}$$

The extremal effect vectors must be defined separately in the cases where n is odd or even. For even n, they are the vectors $e_1, \ldots, e_n$ defined by:

$$e_i = \frac{1}{2} \begin{pmatrix} r_n \cos\left(\frac{(2i-1)\pi}{n}\right) \\ r_n \cos\left(\frac{(2i-1)\pi}{n}\right) \\ 1 \end{pmatrix}. \tag{2.58}$$

For odd n, there are $2n$ extremal effect vectors. The first n are defined by:

$$e_i = \frac{1}{1+r_n^2} \begin{pmatrix} r_n \cos\left(\frac{2i\pi}{n}\right) \\ r_n \cos\left(\frac{2i\pi}{n}\right) \\ 1 \end{pmatrix} (1 \leq i \leq n), \tag{2.59}$$

and the remaining n are defined by $e_{n+i} = \mathcal{U} - e_i$ for $1 \leq i \leq n$. Note that in the case of even n, $\mathcal{U} - e_i$ is equal to $e_{(i+n/2) \bmod n}$, whereas in the case of odd n, this process generates $2n$ distinct effect vectors. Compare this with the features of odd- and even- sided polygons: the first n vectors correspond to normal planes which are parallel to and touch the sides of the polygon. The second set of n vectors correspond to normal planes which touch the polygon diametrically opposite to its sides. The sides of *even* polygons are arranged in diametrically opposing pairs, so that the first and second sets of effect vectors correspond to the same st of planes.

On the other hand, the sides of *odd* polygons are diametrically opposite to their vertices, so that the second set of effect vectors corresponds to an entirely new set of planes.

Figure 2.5 (taken from [36]) depicts the state spaces of the polygon models for some values of n, with state and effect vectors projected onto the 2-dimensional polygon. Note that setting $n = 3$ gives a classical "triangle" system whose single fiducial measurement has 3 outcomes, and setting $n = 4$ gives a Boxworld system with binary inputs and outputs. For $n = 5$ and beyond, however, this set of systems do not overlap with classical, quantum or Boxworld systems.

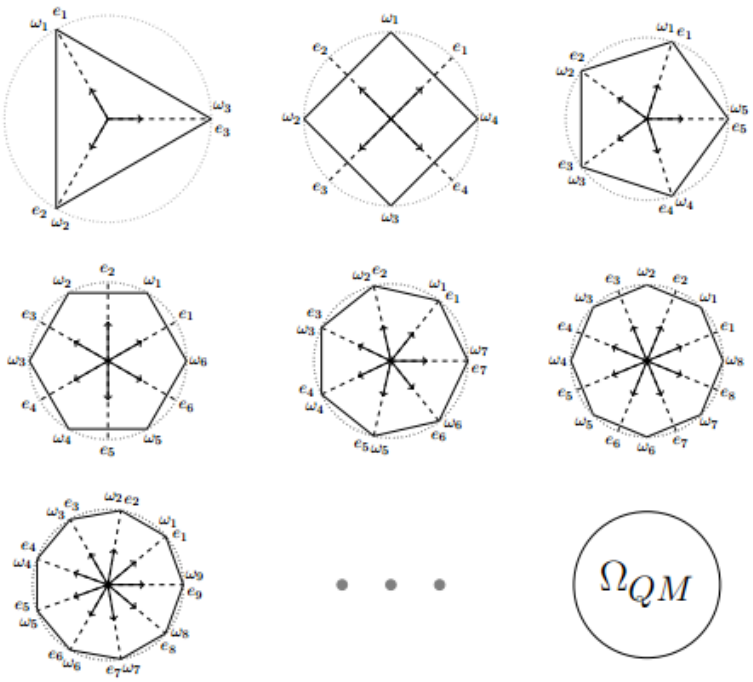

Figure 2.5: Illustration of polygon state spaces and the first n extremal effects (Ω_{QM} denotes the quantum state space). Taken from [36].

Composite states of bipartite polygon systems may be described by a 3x3 matrix S, such that $e^T S f \geq 0$ whenever e and f are local extremal effects or unit vectors of the first and second systems respectively. The value $e^T S f$ gives the

probability of obtaining the local outcomes corresponding to e and f, whenever a local measurement is performed which contains these effects on each system. It can be shown that the states defined by the following matrices are entangled pure states (except for the case $n = 3$):

$$\text{even } n : \phi = \begin{pmatrix} \cos\left(\frac{\pi}{n}\right) & \sin\left(\frac{\pi}{n}\right) & 0 \\ -\sin\left(\frac{\pi}{n}\right) & \cos\left(\frac{\pi}{n}\right) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.60}$$

$$\text{odd } n : \phi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.61}$$

By maximising over local binary measurements of the form $\{e_i, \mathcal{U} - e_i\}$, it is possible to violate Tsirelson's bound on the CHSH value for arbitrarily large, even values of n. For odd values of n on the other hand, Tsirelson's bound is always obeyed by these measurements. Moreover, as n increases, the optimal values for both odd and even n appear to approach Tsirelson's bound from below and above respectively. Intriguingly, Tsirelson's bound provides a kind of dividing line between the optimal CHSH values obtained by odd and even polygons. Yet, even when n is so large that the polygon approximates a circular "slice" of the Bloch sphere, it is still possible to violate quantum predictions.

Chapter 3

Quasiprobability representations

We must by all means stick to our sensations, that is, simply to the present impressions whether of the mind or of any criterion whatever, and similarly to our actual feelings, in order that we may have the means of determining that which needs confirmation and that which is obscure.

"Letter to Herodotus"
Epicurus

The notion of negative probabilities has a history that is both surprisingly long, and surprisingly closely intertwined with the history of quantum mechanics. Ideas about the relation of negative probabilities to quantum theory arose not long after its own development: in his Bakerian Lecture to the Royal Society in 1942 [66], Dirac remarked that negative energies and negative probabilities arose naturally in his calculations regarding relativistic extensions of quantum mechanics. However, when calculating final measurement probabilities, the negative values were invariably "averaged out" to produce positive quantities and probabilities of observed events, and were physically significant only in their contribution to these observed values. He commented, "*Negative energies and probabilities should not be considered as nonsense, ... [but] simply as things which do not appear in experimental results*" [66].

Dirac conceived that those particles associated with negative quantities could be seen as belonging to some hypothetical world differing markedly from our own, but in which analogous transition laws applied, and in which the "averaged out" final probabilities were in complete agreement with the probabilities we observe. In the paper's introduction, in which he discussed Heisenberg's operational take on quantum mechanics, Dirac appears to propose the following viewpoint: as long as we are still ultimately talking about natural observations, it is no great sin if negative quantities are encountered along the way. Even earlier, in 1932, Wigner gave a striking precedent to this attitude in his paper introducing the well-known Wigner phase-space distribution [67]. Whilst a classical joint probability distribution is generally incapable of describing simultaneously a wave function's position and momentum, the Wigner distribution nearly accomplishes this by allowing the values to be negative for some choices of position and momentum. Averaging over all choices of position gives the correct formula for the momentum, and *vice versa*. Wigner noted, "*negative values ... must not hinder the use of [the Wigner distribution] in calculations, as an auxiliary function*" [67].

In the context of quantum information theory, interest in quasiprobability distributions has seen a marked revival, including several applications and results in

quantum information and quantum foundations (a comprehensive review of developments in this topic can be found in [68]). These have tended to focus on quasiprobabilistic representations of quantum states, with a view to progressing our calculational abilities and intuition regarding them. In this chapter we take a different perspective, and are concerned with simulating *all* non-signaling theories via quasiprobabilistic extensions of *both* quantum and classical theory. The work herein is largely based on [1] and [43]; Subsection 3.2.2 and Section 3.4 are exceptions to this, and constitute original, unpublished work. Section 3.3 is the result of collaboration with Anthony Short.

In Section 3.1 we introduce and discuss quasiprobability distributions in a manner that is directly applicable to the discussion of outcome statistics of multipartite experiments. In Section 3.2 we give a detailed discussion of a recent result by Acín *et al* [1] which demonstrates that a specific quasiprobabilistic extension of quantum theory allows the generation of arbitrary non-signaling distributions. In Section 3.3 we show that the same can be achieved for local, classical probability theory in two distinct ways. This results in four (two quantum-like, two classical-like) distinct quasiprobabilistic models for non-signaling correlations; we discuss several interesting features of these models as well as their relevance to the study of quantum theory. In Section 3.4 we then discuss further quasiprobabilistic extensions of quantum theory, focusing on one that employs zero entanglement and an arbitrarily small degree of classical correlation.

3.1 Quasiprobability distributions

Definition. *A function $\tilde{P} : A_1 \times \cdots \times A_N \to \mathbb{R}$, for finite sets $A_1, \ldots, A_N$, is said to be a* (multipartite) quasiprobability distribution *if:*

$$\sum_{a_1,\ldots a_N} \tilde{P}(a_1, \ldots, a_N) = 1. \tag{3.1}$$

Since we will generally be studying the multipartite case, the word “multi-

partite" preceding the phrase "quasiprobability distribution" will be dropped. For convenience we will also use bold symbols to denote strings of values, for example **a** denotes $(a_1, \ldots, a_N)$, and consequently $\tilde{P}(\mathbf{a})$ denotes $\tilde{P}(a_1, \ldots, a_N)$. Like ordinary probability distributions, quasiprobabilities may be conditioned on some other set of variables:

Definition. *A function* $\tilde{P} : A_1 \times \cdots \times A_N \times X_1 \times \cdots X_N \to \mathbb{R}$ *is said to be a* conditional quasiprobability distribution *if, for all fixed choices of* $\mathbf{x} \in X_1 \times \cdots \times X_N$*:*

$$\sum_{a_1, \ldots a_N} \tilde{P}(\mathbf{a}|\mathbf{x}) = 1, \tag{3.2}$$

where the notation $\tilde{P}(\mathbf{a}|\mathbf{x})$ *is used in place of* $\tilde{P}(a_1, \ldots, a_N, x_1, \ldots, x_N)$*.*

Conditional quasiprobability distributions form a direct analogue of outcome distributions. Indeed, we can define a more general form of the non-signaling condition which applies to quasiprobability distributions. Due to the difficulty in finding a convincing physical interpretation for negative probabilities, it is difficult to physically motivate this non-signaling condition in terms of communication between distant parties. Nevertheless, this form of the non-signaling condition forms an important mathematical step in the derivation of our results.

Definition. *The conditional quasiprobability distribution* $\tilde{P}(\mathbf{a}|\mathbf{x})$ *is* non-signaling *if, for all* i *and choices of* $x_1, \ldots, x_N$*, the expression*

$$\sum_{a_i=1}^{K_{x_i}} \tilde{P}(a_1, \ldots, a_N | x_1, \ldots, x_N) \tag{3.3}$$

is independent of the particular value of x_i*.*

The only assumption we have made about the outcome sets $A_1, \cdots, A_N$ so far is that they are finite. Without loss of generality, we may as well assume that $A_i = \{1, \ldots, K_{x_i}^{(i)}\}$, where $K_{x_i}^{(i)}$ is the number of possible local outcomes when measurement x_i is performed locally at subsystem i. For a string of outcome values **a** and a subset $\Omega = \{i_1, \ldots, i_k\} \subseteq [N]$ with $i_1 < \cdots < i_k$, we will use the

notation $\mathbf{a}_\Omega, \mathbf{x}_\Omega$ to denote the reduced strings $(a_{i_1}, \ldots, a_{i_k}), (x_{i_1}, \ldots, x_{i_k})$. Just as with conventional outcome distributions, a non-signaling conditional quasiprobability distribution has well-defined marginal distributions $\tilde{P}(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ for any choice of Ω. The following technical definition will be helpful in characterizing conditional quasiprobability distributions according to a restricted subset of its marginal values.

Definition. *The outcome a_i on system i is* maximal with respect to x_i *if it takes on the greatest value permitted by measurement choice x_i, i.e. $a_i = K_{x_i}^{(i)}$. For a subset $\Omega \subseteq [N]$, the string $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ is* nowhere-maximal *if, for all $i \in \Omega$, a_i is* not *maximal with respect to x_i.*

We now state and prove a lemma concerning non-signaling quasiprobability distributions, of which great use will be made throughout this chapter. The moral of the lemma is as follows: if one wants to generate a non-signaling quasiprobability distribution P, then essentially one needs only focus on the set of nowhere-maximal, marginal values of P. As long as one's generated distribution is itself non-signaling and agrees with P for these inputs, then it is guaranteed to agree with P for *all* inputs. Ideas very similar to this have been utilised implicitly in the literature of non-signaling correlations [1,69], but the author does not know of any explicit statement of the result.

Lemma 1. *A non-signaling, conditional quasiprobability distribution $\tilde{P}(\mathbf{a}|\mathbf{x})$ is uniquely characterized by the set of non-maximal marginal values $\left\{\tilde{P}(\mathbf{a}_\Omega|\mathbf{x}_\Omega)\right\}$ as $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ runs over the set of nowhere-maximal strings and all subsets $\emptyset \neq \Omega \subseteq [N]$.*

Proof. Suppose that $\tilde{P}(\mathbf{a}|\mathbf{x})$ and $\tilde{P}'(\mathbf{a}|\mathbf{x})$ are non-signaling conditional quasiprobability distributions, whose marginals $\tilde{P}(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ and $\tilde{P}'(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ agree for all nowhere-maximal choices of $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$. We argue that $\tilde{P}(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ and $\tilde{P}'(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ actually agree for *all* choices of $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$, by induction on $n = \#\{i \in \Omega : a_i \text{ maximal}\}$.

The case $n = 0$ holds by assumption. Consider a string $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ with $n > 0$ maximal a_i; let $\Omega = \{i_1, \ldots, i_M\}$ and without loss of generality suppose that

$a_{i_1} = K_{x_{i_1}}^{(i_1)}$ (the proof is no different otherwise). Since $\tilde{P}\left(a_1, \ldots, a_N | x_1, \ldots, x_N\right)$ is non-signaling, so are all of its marginal distributions, therefore applying the non-signaling condition to system i_1,

$$\sum_{a_{i_1}=1}^{K_{x_{i_1}}^{(i_1)}} \tilde{P}\left(\mathbf{a}_\Omega | \mathbf{x}_\Omega\right) = \tilde{P}\left(a_{i_2}, \ldots, a_{i_M} | x_{i_2}, \ldots, x_{i_M}\right). \tag{3.4}$$

Rewriting this, for our particular choice of $(\mathbf{a}_\Omega | \mathbf{x}_\Omega)$, the marginal $\tilde{P}\left(\mathbf{a}_\Omega | \mathbf{x}_\Omega\right)$ can be expressed in terms of quantities where fewer of the a_{i_j} are maximal:

$$\begin{aligned}\tilde{P}\left(K_{i_1}^{(i_1)}, \ldots, a_{i_M} | x_{i_1}, \ldots, x_{i_M}\right) &= \tilde{P}\left(a_{i_2}, \ldots, a_{i_M} | x_{i_2}, \ldots, x_{i_M}\right) \\ &\quad - \sum_{\substack{a_{i_1} \text{ not} \\ \text{maximal}}} \tilde{P}\left(a_{i_1}, \ldots, a_{i_M} | x_{i_1}, \ldots, x_{i_M}\right). \end{aligned} \tag{3.5}$$

The number of maximal a_i is now smaller than n for each term on the RHS, so by the induction hypothesis we can switch $\tilde{P}$ with $\tilde{P}'$ for each term individually:

$$\begin{aligned}\tilde{P}\left(K_{i_1}^{(i_1)}, \ldots, a_{i_M} | x_{i_1}, \ldots, x_{i_M}\right) &= \tilde{P}'\left(a_{i_2}, \ldots, a_{i_M} | x_{i_2}, \ldots, x_{i_M}\right) \\ &\quad - \sum_{\substack{a_{i_1} \text{ not} \\ \text{maximal}}} \tilde{P}'\left(a_{i_1}, \ldots, a_{i_M} | x_{i_1}, \ldots, x_{i_M}\right). \end{aligned} \tag{3.6}$$

$\tilde{P}'$ is also non-signaling, hence equation (3.4) is still valid after replacing $\tilde{P}$ with $\tilde{P}'$. Therefore the reverse of the procedure used to obtain equation (3.5) can be applied to $\tilde{P}'$ on the RHS, giving

$$\begin{aligned}\tilde{P}\left(K_{i_1}^{(i_1)}, \ldots, a_{i_M} | x_{i_1}, \ldots, x_{i_M}\right) &= \tilde{P}'\left(K_{i_1}^{(i_1)}, \ldots, a_{i_M} | x_{i_1}, \ldots, x_{i_M}\right) \\ \Rightarrow \tilde{P}\left(\mathbf{a}_\Omega | \mathbf{x}_\Omega\right) &= \tilde{P}'\left(\mathbf{a}_\Omega | \mathbf{x}_\Omega\right). \end{aligned} \tag{3.7}$$

□

Example. A simple example may help to illustrate the central idea of the proof,

which is that we can exploit the non-signaling condition in an iterative fashion in order to express a particular value $P(\mathbf{a}|\mathbf{x})$ in terms of marginals for which the a_i are nowhere maximal. Suppose that P is the outcome distribution for a 2-party, 2-input and 2-output experiment, i.e. one possible state of a g-bit system. Recall that for a g-bit system, we will prefer to use binary notation $\{0,1\}$ to label inputs and outputs, rather than the set $\{1,2\}$. The probability $P(1,1|0,0)$ has a maximal outcome value for both parties, but admits the following reduction using the non-signaling and normalisation conditions:

$$\begin{aligned} P(1,1|0,0) &= P(a_2=1|x_2=0) - P(0,1|0,0) \\ &= \big[1 - P(a_2=0|x_2=0)\big] - \big[P(a_1=0|x_1=0) - P(0,0|0,0)\big] \\ &= 1 + P(0,0|0,0) - \big[P(a_1=0|x_1=0) + P(a_2=0|x_2=0)\big]. \end{aligned}$$

3.2 Quantum theory with quasiprobabilities

An outcome distribution P is quantum-achievable if there exist Hilbert spaces (of any dimension) $\mathcal{H}_1, \ldots, \mathcal{H}_N$, positive operators $M^{(i)}_{a_i|x_i} \in \mathcal{P}(\mathcal{H}_i)$ satisfying $\sum_{a_i} M^{(i)}_{a_i|x_i} = \mathbb{I}^{(i)}$, and a density matrix $\rho \in \mathcal{D}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N)$ such that, for all choices of a_i and x_i,

$$P(a_1,\ldots,a_N|x_1,\ldots,x_N) = \mathrm{Tr}\left(M^{(1)}_{a_1|x_1} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}\ \rho\right). \tag{3.8}$$

The set of quantum-achievable outcome distributions $\mathcal{Q}$ inhabits a specific portion of the full set of non-signaling distributions. If quantum theory - or some (perhaps relativistic) extension of it - provides an accurate description of natural phenomena, then the correlations belonging to $\mathcal{Q}$ are exactly those which, in principle, it is possible to generate between spatially separated parties in the real world. It therefore seems a worthwhile endeavour for two reasons to explore what defining characteristics single out $\mathcal{Q}$ as a subset of non-signaling distributions.

Firstly, from a practical vantage, it is beneficial to have a good understanding

of exactly which correlations it is possible to generate between distant parties, and which correlations are unattainable. Any additional intuition gained from developing a broad treatment of correlations (rather than the formal paradigm of quantum theory) is also a useful aid in developing algorithms that solve information-theoretic problems with real-world applications [70, 71]. Secondly, on a more foundational level, analysing the principles or rules that delimit $\mathcal{Q}$ from other non-signaling distributions [50, 52, 72, 73] gives us a greater appreciation of what it is that makes quantum theory so uniquely successful at describing Nature. Even if quantum theory is superseded by a more complete or more accurate physical theory, abstract notions concerning which correlations are allowable will persist.

One approach to this issue that has met with some success is to ask the question "*given a non-signaling distribution P(**a**|**x**), are there simple criteria by which one can determine whether P is quantum-achievable or not?*". Even when the number of parties is small, this turns out to be a more difficult task than might be anticipated: in 2007 Navascués *et al* introduced a hierarchy of conditions which are necessarily satisfied by bipartite quantum-achievable correlations [74], and later demonstrated that any non-signaling distribution which lies outside of $\mathcal{Q}$ violates one of these conditions (and therefore all subsequent conditions) [75]. Specifically, they introduce a sequence $\{\mathcal{Q}^n\}_{n\geq 1}$ of sets of outcome distributions which satisfies $\mathcal{Q}^{n+1} \subseteq \mathcal{Q}^n$ and $\cap_{n=1}^{\infty}\mathcal{Q}^n = \mathcal{Q}$ (see Figure 3.1). Determining whether a given outcome distribution belongs to the set $\mathcal{Q}^n$ can be formulated as a semi-definite program; if the distribution fails this test then it is not quantum-achievable. However, in general it cannot be confirmed that a distribution *is* quantum-achievable after finitely many tests.

Another perspective is to instead ask "*in what ways does the set $\mathcal{Q}$ stand out?*". Acín *et al* [1] have shown that the standard Born trace rule can be extended to generate any non-signaling correlation, by allowing the quantum state (usually represented by a positive density operator) to be a non-positive operator. Explicitly, given any non-signaling distribution $P(\mathbf{a}|\mathbf{x})$ on N parties, there exist Hilbert spaces $\mathcal{H}_i$, POVM elements $M^{(i)}_{a_i|x_i} \in \mathcal{L}(\mathcal{H}_i)$ and a Hermitian, unit-trace but not

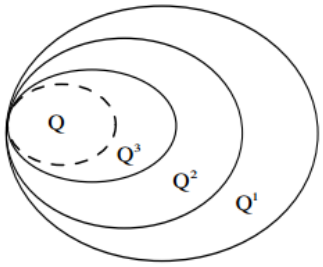

Figure 3.1: Hierarchy of sets which approximates the set of quantum-achievable outcome distributions. Taken from [75].

necessarily positive operator $\tilde{\rho}$ such that:

$$P\left(\mathbf{a}|\mathbf{x}\right) = \mathrm{Tr}\left(M^{(1)}_{a_1|x_1} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}\tilde{\rho}\right). \tag{3.9}$$

This result provides a universal, quantum-like manner in which all non-signaling correlations can be written. In other words, the set of non-signaling correlations is exactly that which can be generated using the formula (3.9). Moreover, the set of quantum-achievable correlations can be neatly recovered by demanding a single additional constraint: that the operator $\tilde{\rho}$ be positive. The set of classical or local correlations can also be neatly recovered, by demanding that $\tilde{\rho}$ be positive and separable.

As well as being of standalone interest, this formulation has recently been used to prove the following result: any physical theory for which the local structure is identical to that of qubits, and which admits at least one continuous, reversible interaction, must have the global structure specified by quantum theory [2]. The ability to represent the correlations of a broad class of theories in a similar way to quantum or classical correlations may provide a powerful tool in analysing such theories, and identifying which properties of quantum theory are special within this class of theories.

In Section 3.2.1 we reproduce the proof of the result of Acín *et al* in detail, generalising the published version to allow for outcome sets that vary according

to the subsystem i and measurement choice x_i. We then make the original and interesting remark that with some simple modifications, the operator $\tilde{\rho}$ can be made so that it and the set of POVM operators $M^{(1)}_{a_1|x_1} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}$ pairwise commute. This will lead us naturally into Section 3.3, in which we study quasiprobabilistic extensions of the local classical formalism, rather than the quantum formalism.

3.2.1 Non-positive density matrices

We now give a detailed exposition of the procedure described by Acín *et al* [1] for representing arbitrary non-signaling outcome distributions in terms of local POVM operators acting on a Hermitian, unit-trace, but not necessarily positive density operator $\tilde{\rho}$. In fact, the published proof assumes that all measurements have the same number of outcomes and that all subsystems have the same number of measurements, and gives the explicit form of the constructed operators only in the case $N = 2$ (but sketches the multipartite generalisation). The proof given here provides a concise and rigorous construction in the multipartite case and allows for differing numbers of measurement outcomes.

Lemma 2. *Let $\mathcal{H}$ be a complex, d-dimensional Hilbert space, and let V be the real, d^2-dimensional vector space of Hermitian matrices acting on $\mathcal{H}$. Then there exists a basis $\{E_\alpha\}$ of V satisfying the following properties:*

(i) *E_α is positive semidefinite for all α;*

(ii) *$E_1 = \mathbb{I}$;*

(iii) *$\mathbb{I} - \sum_{\alpha>1} E_\alpha$ is a positive semi-definite operator.*

Proof. Let $\{|e_i\rangle\}$ be a basis for $\mathcal{H}$, and define the $d \cdot (d-1)/2$ vectors

$$|f_{jk}\rangle = |e_j\rangle + |e_k\rangle \qquad k < l \tag{3.10}$$

and the $d \cdot (d-1)/2$ vectors

$$|g_{mn}\rangle = |e_m\rangle + i|e_n\rangle \qquad m < n. \tag{3.11}$$

The set of d^2 rank-1 projectors $\{|e_i\rangle\langle e_i|, |f_{jk}\rangle\langle f_{jk}|, |g_{mn}\rangle\langle g_{mn}|\}$ forms a linearly independent subset of the real vector space V. To see this, consider the decompositions of these projectors into matrices of the form $|e_j\rangle\langle e_k|$. Note that $|f_{jk}\rangle\langle f_{jk}|$ is the only projector to produce an $|e_j\rangle\langle e_k|$ cross-term with a real coefficient, whilst $|g_{mn}\rangle\langle g_{mn}|$ is the only projector to produce an $|e_m\rangle\langle e_n|$ cross-term with a complex coefficient. Since V is of dimension d^2, this set of projectors forms a basis for V.

Suppose that $\{E_\alpha\}$ denotes a relabelling of the above basis for V such that $E_1 = |e_d\rangle\langle e_d|$. Note that the identity can be decomposed as $\mathbb{I} = \sum_{i=1}^{d} |e_i\rangle\langle e_i|$; therefore $E_1 = |e_d\rangle\langle e_d|$ lies in the span of the set $\{\mathbb{I}, E_\alpha\}_{\alpha>1}$. Setting E_1 equal to $\mathbb{I}$ instead of $|e_d\rangle\langle e_d|$ forms a new set which therefore also spans V. Hence we may assume without loss of generality that $E_1 = \mathbb{I}$. Since this new set also has d^2 elements, it again forms a basis for V, thus conditions (i) and (ii) are satisfied.

To additionally satisfy condition (iii), note that the matrix $E = \sum_{\alpha>1} E_\alpha$ is also positive semi-definite, and so by compactness the real number

$$\omega = \max_{|||\psi\rangle||=1} \langle\psi|E|\psi\rangle \tag{3.12}$$

is both positive and finite. After scaling every E_α by $\frac{1}{\omega}$ for $\alpha > 1$, we have that $\mathbb{I} - \sum_{\alpha>1} E_\alpha$ is also a positive operator. □

Theorem 5. *An outcome distribution P is non-signaling if and only if there exist Hilbert spaces (of any dimension) $\mathcal{H}_1, \ldots, \mathcal{H}_N$, operators $M^{(i)}_{a_i|x_i} \in \mathcal{P}(\mathcal{H}_i)$ satisfying $\sum_{a_i} M^{(i)}_{a_i|x_i} = \mathbb{I}^{(i)}$, and a Hermitian, unit-trace operator $\tilde{\rho} \in \mathcal{L}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N)$ such that*

$$P(a_1, \ldots, a_N|x_1, \ldots, x_N) = Tr\left(M^{(1)}_{a_1|x_1} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}\, \tilde{\rho}\right). \tag{3.13}$$

Proof. The "if" direction of the proof is little more than an algebraic substitution of the condition $\sum_{a_i} M^{(i)}_{a_i|x_i} = \mathbb{I}^{(i)}$. Indeed, for an arbitrary choice of x_1, summing

over a_1 and using linearity of the trace operation gives:

$$\begin{aligned}\sum_{a_1=1}^{K_{x_1}^{(1)}} P\left(a_1,\ldots,a_N|x_1,\ldots,x_N\right) &= \sum_{a_1=1}^{K_{x_1}^{(1)}} \mathrm{Tr}\left(M_{a_1|x_1}^{(1)}\otimes\cdots\otimes M_{a_N|x_N}^{(N)}\,\tilde{\rho}\right)\\ &= \mathrm{Tr}\left(\left(\sum_{a_1=1}^{K_{x_1}^{(1)}} M_{a_1|x_1}^{(1)}\right)\otimes\cdots\otimes M_{a_N|x_N}^{(N)}\,\tilde{\rho}\right)\\ &= \mathrm{Tr}\left(\mathbb{I}^{(1)}\otimes\cdots\otimes M_{a_N|x_N}^{(N)}\,\tilde{\rho}\right)\end{aligned} \tag{3.14}$$

Equation (3.14) clearly has no dependence on the value of x_1, hence any outcome distribution P that can be represented in this way is non-signaling.

To prove the converse, we construct local measurement operators $M_{a_i|x_i}^{(i)}$ and a Hermitian operator $\tilde{\rho}$ of unit trace which generates a non-signaling quasi-distribution $\tilde{P}'$, then show that $\tilde{P}'$ agrees with P for all nowhere-maximal marginal strings $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$. Using Lemma 1, this is sufficient to conclude that $\tilde{P}'$ is identical to the outcome distribution P, and therefore that P is generated according to (3.13).

Recall that system i has $M^{(i)}$ measurement choices, with measurement x allowing for $K_x^{(i)}$ possible outcomes. Let $\mathcal{H}_i \cong \mathbb{C}^{d_i}$, where d_i is large enough that

$$d_i^2 \geq 1 + \sum_{x=1}^{M^{(i)}} \left(K_x^{(i)} - 1\right). \tag{3.15}$$

Let $\{E_\alpha\}$ be the basis of positive semi-definite matrices defined in Lemma 2 for the vector space V of Hermitian matrices acting on $\mathcal{H}_i$. Note that V has a real inner product $\langle A, B\rangle = \mathrm{Tr}\,(AB)$. Let $\{\tilde{E}_\alpha\}$ be the dual basis to $\{E_\alpha\}$, defined by $\langle E_\alpha, \tilde{E}_\beta\rangle = \mathrm{Tr}\left(E_\alpha\tilde{E}_\beta\right) = \delta_{\alpha\beta}$. For $1 \leq x_i \leq M^{(i)}$ and $1 \leq a_i \leq K_x^{(i)} - 1$, define $\{M_{a_i|x_i}^{(i)}\}$ (the set of non-maximal POVM elements) to be some subset of $\{E_\alpha\}_{\alpha>1}$; note that this is possible as long as d_i is large enough to satisfy (3.15). For each matrix $M_{a_i|x_i}^{(i)}$, let $\tilde{M}_{a_i|x_i}^{(i)}$ denote the corresponding dual matrix, which is

Hermitian but not necessarily positive semi-definite. To complete the POVM set for each measurement x_i, define $M^{(i)}_{K^{(i)}_{x_i}|x_i} = \mathbb{I}^{(i)} - \sum_{a_i=1}^{K^{(i)}_{x_i}-1} M^{(i)}_{a_i|x_i}$ which, due to condition (iii) of Lemma 2, is also a positive semi-definite operator. Note that the POVM elements have been defined purely in terms of the number of measurement choices and outcomes, and do not depend on the specific outcome distribution P.

Our non-positive density operator $\tilde{\rho} \in \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N$, on the other hand, does depend on P. In giving a concise formula for $\tilde{\rho}$, a slight abuse of notation is needed: for a subset $\Omega \subset [N]$ of the N systems, $\hat{\bigotimes}_{i\in\Omega} \tilde{M}^{(i)}_{a_i|x_i}$ will denote an N-fold tensor product, in which the matrix $\tilde{\mathbb{I}}^{(i)}$ is inserted for components i not belonging to Ω. For example,

$$\hat{\bigotimes}_{i\in\{2,3\}} \tilde{M}^{(i)}_{a_i|x_i} = \tilde{\mathbb{I}}^{(1)} \otimes \tilde{M}^{(2)}_{a_2|x_2} \otimes \tilde{M}^{(3)}_{a_3|x_3}. \tag{3.16}$$

Note that since P is assumed to be non-signaling, the marginal distributions $P(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ are well-defined for all $\emptyset \neq \Omega \subseteq [N]$. For the empty set $\Omega = \emptyset$, we will also define $P(\mathbf{a}_\emptyset|\mathbf{x}_\emptyset) = 1$. The operator $\tilde{\rho}$ is then given by the formula

$$\tilde{\rho} = \sum_{\Omega\subseteq[N]} \sum_{\substack{\text{nowhere-maximal} \\ \text{strings } (\mathbf{a}_\Omega|\mathbf{x}_\Omega)}} P(\mathbf{a}_\Omega|\mathbf{x}_\Omega) \bigotimes_{i\in\Omega} \tilde{M}^{(i)}_{a_i|x_i}. \tag{3.17}$$

Since each $\tilde{M}^{(i)}_{a_i|x_i}$ is not guaranteed to be positive semi-definite, neither is the operator $\tilde{\rho}$. However, by construction they are all Hermitian, hence so is $\tilde{\rho}$. Suppose now that $\tilde{P}'$ is the (non-signaling) quasiprobability distribution generated by the POVM elements $M^{(i)}_{a_i|x_i}$ acting on the state $\tilde{\rho}$, according to the Born trace rule (3.13). Given a subset $\Omega \subseteq [N]$ and nowhere-maximal string $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$, the marginal value $\tilde{P}'(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ is obtained by taking the trace of the product of $\tilde{\rho}$ with the matrix $\bigotimes_{i\in\Omega} M^{(i)}_{a_i|x_i}$.

Using the linearity of the trace, the fact that $\mathrm{Tr}(A \otimes B) = \mathrm{Tr}(A) \cdot \mathrm{Tr}(B)$, and the duality of the constructed basis sets, it is clear that these marginal values are exactly $P(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$. Therefore by Lemma 1, $\tilde{P}'$ is in fact the outcome distribution P, as desired. By post-multiplying $\tilde{\rho}$ by the N-partite identity matrix $\bigotimes_{i=1}^{N} \mathbb{I}^{(i)}$, a

similar procedure verifies that $\tilde{\rho}$ is of unit trace:

$$\mathrm{Tr}\left(\tilde{\rho}\right) = \mathrm{Tr}\left(\tilde{\rho}\left(\bigotimes_{i=1}^{N}\mathbb{I}^{(i)}\right)\right) = 1. \tag{3.18}$$

□

Example The construction outlined in the previous proof will now be applied to the canonical PR-box state. Recall that the PR-box is a special state of the system comprised of two g-bit subsystems, i.e. each subsystem has two fiducial measurement choices and two possible outcomes for each fiducial measurement. Hence it is sufficient to set $d_1 = d_2 = 2$ in order to satisfy (3.15). By following the procedure in Lemma 2, before performing the rescaling, we obtain the following basis of positive semi-definite matrices for the vector space of 2x2 complex Hermitian matrices:

$$\begin{aligned} E_1 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \\ E_3 &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad E_4 = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}. \end{aligned} \tag{3.19}$$

In order to satisfy property (iii) of Lemma 2, it is straightforward to check that the following matrix is positive semi-definite,

$$E_1 - \frac{1}{\omega}\left(E_2 + E_3 + E_4\right), \tag{3.20}$$

as long as $\omega \geq \frac{5}{2}$. For convenience, we will set $\omega = 4$, generating the following set of POVM operators (recall that for the PR-box, measurement choices and

outcomes belong to the set $\{0,1\}$):

$$M_{0|0} = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 0 \end{pmatrix}, \quad M_{1|0} = \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & 1 \end{pmatrix},$$
$$M_{0|1} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix}, \quad M_{1|1} = \begin{pmatrix} \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} \end{pmatrix}. \tag{3.21}$$

Since the POVM elements are taken to be identical for both systems, the superscript (i) is unnecessary. and has been omitted. The relevant dual elements to the POVM operators are then,

$$\tilde{\mathbb{I}} = \begin{pmatrix} 0 & \frac{-1+i}{2} \\ \frac{-1-i}{2} & 1 \end{pmatrix}, \quad \tilde{M}_{0|0} = \begin{pmatrix} 4 & 0 \\ 0 & -4 \end{pmatrix}, \quad \tilde{M}_{0|1} = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}. \tag{3.22}$$

Recall that the PR-box outcome statistics are described by the formula $P\left(a_1, a_2 | x_1, x_2\right) = \frac{1}{2}\delta(a_1 \oplus a_2 = x_1 \cdot x_2 \text{ (}\textit{mod 2}\text{)})$. In this case, the marginal distributions are uniformly distributed for both systems, and the formula for $\tilde{\rho}$ reduces to,

$$\begin{aligned} \tilde{\rho} = & \sum_{x_1,x_2=0}^{1} \frac{1}{2}\delta\big(x_1 \cdot x_2 = 0\big)\tilde{M}^{(1)}_{0|x_1} \otimes \tilde{M}^{(2)}_{0|x_2} \\ & + \frac{1}{2}\left(\tilde{M}^{(1)}_{0|0} \otimes \tilde{\mathbb{I}}^{(2)} + \tilde{M}^{(1)}_{0|1} \otimes \tilde{\mathbb{I}}^{(2)}\right) \\ & + \frac{1}{2}\left(\tilde{\mathbb{I}}^{(1)} \otimes \tilde{M}^{(2)}_{0|0} + \tilde{\mathbb{I}}^{(1)} \otimes \tilde{M}^{(2)}_{0|1}\right) \\ & + \tilde{\mathbb{I}}^{(1)} \otimes \tilde{\mathbb{I}}^{(2)}. \end{aligned} \tag{3.23}$$

Substituting the matrices obtained in (3.22) gives,

$$\tilde{\rho} = \begin{pmatrix} 8 & 3+i & 3+i & -1+\frac{i}{2} \\ 3-i & -6 & -\frac{1}{2} & \frac{-5-i}{2} \\ 3-i & -\frac{1}{2} & 10 & \frac{-5-i}{2} \\ -1-\frac{i}{2} & \frac{-5+i}{2} & \frac{-5+i}{2} & -11 \end{pmatrix} \tag{3.24}$$

Note that $\tilde{\rho}$ is Hermitian, of unit trace, and non-positive (for example $\langle e_4|\tilde{\rho}|e_4\rangle = -11$). It can be checked that the POVM elements (3.21) do indeed generate the correct probabilities when acting on the matrix $\tilde{\rho}$ even in the case where the string $(\mathbf{a}|\mathbf{x})$ is not nowhere-maximal, for example,

$$\text{Tr}\left(M_{0|1} \otimes M_{1|1}\ \tilde{\rho}\right) = \frac{1}{2} \tag{3.25}$$

$$\text{Tr}\left(M_{1|1} \otimes M_{1|1}\ \tilde{\rho}\right) = 0. \tag{3.26}$$

This concludes the example of generating PR-box statistics using a non-positive density operator.

3.2.2 Commuting operators

It turns out that with a simple modification of the proof of Theorem 5, the same result can be achieved but with the additional constraint that all operators (both the POVM elements and $\tilde{\rho}$) commute. The cost of such a modification is an increase in the dimension of the Hilbert space at each system - in fact we require that $d_i = 1+\sum_{x=1}^{M^{(i)}}\left(K_x^{(i)}-1\right)$, so that dimension of the Hilbert space is roughly the square of that in the proof of Theorem 5.

The construction of the POVM elements is then particularly simple. Let

$$\{|e_0\rangle\} \cup \{|e_{a_i|x_i}\rangle\ |1 \leq x_i \leq M^{(i)}, 1 \leq a \leq K_{x_i}^{(i)} - 1\}, \tag{3.27}$$

be a basis for the Hilbert space $\mathcal{C}^{d_i}$, and for non-maximal pairs (a_i, x_i), define $M_{a_i|x_i}^{(i)}$ to be the projector $|e_{a_i|x_i}\rangle\langle e_{a_i|x_i}|$. Writing the identity as $\mathbb{I}^{(i)} = |e_0\rangle\langle e_0| + \sum |e_{a_i|x_i}\rangle\langle e_{a_i|x_i}|$, it is clear for each x_i that the matrix given by,

$$M_{K_{x_i}^{(i)}|x_i}^{(i)} = \mathbb{I}^{(i)} - \sum_{a_i < K_{x_i}^{(i)}} M_{a_i|x_i}^{(i)}, \tag{3.28}$$

is itself a diagonal matrix all entries either 1 or 0. The dual basis of matrices (in the

same sense as in the proof of Theorem 5) is given by the set of diagonal matrices:

$$\tilde{M}^{(i)}_{a_i|x_i} = |e_{a_i|x_i}\rangle\langle e_{a_i|x_i}| - |e_0\rangle\langle e_0| \quad \text{(non-maximal } (a_i|x_i) \tag{3.29}$$

$$\tilde{I}^{(i)} = |e_0\rangle\langle e_0|. \tag{3.30}$$

Again, $\tilde{\rho}$ is given by the formula:

$$\tilde{\rho} = \sum_{\Omega\subseteq[N]} \sum_{\substack{\text{nowhere-maximal} \\ \text{strings } (\mathbf{a}_\Omega|\mathbf{x}_\Omega)}} P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right) \bigotimes_{i\in\Omega} \tilde{M}^{(i)}_{a_i|x_i}, \tag{3.31}$$

which, for exactly the same reasons as before, is of unit trace and reproduces the probability distribution P for all nowhere-maximal strings $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$, hence exactly reproduces P. Notice that, as a linear combination of tensor products of diagonal matrices, $\tilde{\rho}$ is also diagonal in the standard basis representation. Since diagonal matrices commute, all the operators thus far defined are pairwise commuting.

Example Let us again turn to the case of the system comprised of two g-bit subsystems. In this case we require $d_1 = d_2 = 3$, and for all systems the POVM elements take the form:

$$M_{0|0} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad M_{1|0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$M_{0|1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad M_{1|1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{3.32}$$

and the dual set of matrices takes the form:

$$\tilde{\mathbb{I}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \tilde{M}_{0|0} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \tilde{M}_{0|1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{3.33}$$

This construction allows for slightly easier calculation of the non-positive density operator; in fact it is not too hard to find the operator for an arbitrary state $P(a_1, a_2|x_1x_2)$ of two g-bit systems. Then the operator $\tilde{\rho}$ which generates this probability distribution when acted on by the POVM elements (3.32) is the 9x9 diagonal matrix whose entries in order are:

$$\begin{aligned}
&\{P(0,0|0,0)\,,\\
&\ P(0,0|0,1)\,,\\
&\ P(0,1|0,1) - P(0,0|0,0)\\
&\ P(0,0|1,0)\,,\\
&\ P(0,0|1,1)\,,\\
&\ P(0,1|1,0) - P(0,0|1,1)\\
&\ P(1,0|1,0) - P(0,0|0,0)\,,\\
&\ P(1,0|0,1) - P(0,0|1,1)\,,\\
&\ 1 - \sum_{x_1} P(a_1 = 0|x_1) - \sum_{x_2} P(a_2 = 0|x_2) + \sum_{x_1,x_2} P(0,0|x_1,x_2)\}
\end{aligned} \tag{3.34}$$

3.3 Classical theory with quasiprobabilities

As discussed in Chapter 2, an outcome distribution is said to be *locally achievable* or *classical* if it can be written in such a way that, up to some shared statistical randomness, measurement outcomes on an individual system are determined solely by the value of some (physical) variable which is unique to that system. Specifically, $P(\mathbf{a}|\mathbf{x})$ is local if there exists for each system i some set of values Λ_i which determine the output at that system, with a joint probability distribution $P_\Lambda(\boldsymbol{\lambda}) = P_\Lambda(\lambda_1, \ldots, \lambda_N)$ for the local variable $\lambda_i \in \Lambda_i$, plus a set of conditional probability distributions $P^{(i)}(a_i|x_i, \lambda_i)$ on system i, such that for each choice of x_i

and λ_i, $\sum_{a_i} P^{(i)}(a_i|x_i,\lambda_i) = 1$, and,

$$P\left(\mathbf{a}|\mathbf{x}\right) = \sum_{\lambda_1,\ldots,\lambda_N} \left(P^{(1)}(a_1|x_1,\lambda_1)\cdots P^{(N)}(a_N|x_N,\lambda_N)\right) P_\Lambda(\boldsymbol{\lambda}). \tag{3.35}$$

There are two distinct positivity conditions which one could relax in (3.35): either the joint distribution $P_\Lambda(\lambda)$, or the local distributions $P^{(i)}(a_i|x_i,\lambda_i)$ may be permitted to take on negative values. In this section, we demonstrate that each of these modifications expands the set of distributions of the form given in (3.35) so that it encapsulates all non-signaling distributions.

The simulation of quantum states and measurements with local quasiprobability distributions has an established history, with the Wigner phase-space distribution being a notable early example [67]. Let $\psi(x)$ be the position-space wave function of a quantum system with one dimension, in some pure state ψ. The Wigner phase-space distribution is the following quasiprobability distribution:

$$\tilde{P}(x,p) = \frac{1}{\pi\hbar}\int_{-\infty}^{\infty} \psi^*(x+y)\psi(x-y)e^{2ipy/\hbar}dy. \tag{3.36}$$

For some choices of position x and momentum p, $\tilde{P}(x,p)$ may well be negative. However, the marginal distribution for x when averaging over p gives the exact probability distribution for x, and vice versa:

$$\int_{-\infty}^{\infty} \tilde{P}(x,p)dp = |\psi(x)|^2, \tag{3.37}$$

$$\int_{-\infty}^{\infty} \tilde{P}(x,p)dx = |\psi(p)|^2. \tag{3.38}$$

Wigner was able to generalise his function to mixed states of several quantum particles, and used it to calculate quantum correction terms to the thermodynamic equilibrium of many-particle systems. More recently, and more relevant to quantum information theory, Franco & Penna [76] analysed a finite-dimensional version of the Wigner function, known as the discrete Wigner function, for the

quantum state of two qubit systems. They discovered that this function was intimately tied to the entanglement properties of the state: if at least one of values taken on by the Wigner function falls below a critical negative threshold, then the state is entangled.

Local quasiprobabilistic representations of quantum states have provided a useful calculative tool in what may otherwise have been intractable problems. However, the notion of general non-signaling correlations has not been around for as long as quantum theory, and not as much attention has been devoted to local quasiprobabilistic representations of arbitrary non-signaling outcome distributions. In Section 3.3.1 we demonstrate how to generate any non-signaling outcome distribution by relaxing the positivity of P_Λ; in Section 3.3.2 we do the same for $P^{(i)}(a_i|x_i, \lambda_i)$. Then in Section 3.3.3, we show how these results lead to quantum corollaries which extend the results of Section 3.1.

3.3.1 Non-positive mixtures

The local function $P^{(i)}(a_i|x_i, \lambda_i)$ is conditioned on the parameter λ_i, which so far we have treated as simply some random variable taking values from some finite set Λ_i. λ_i can be interpreted as providing a complete specification of all physical properties pertaining to the system, even if such a full specification is beyond the full scope of whatever scientific theory is currently available to describe that system. If $\boldsymbol{\lambda} = (\lambda_1, \ldots \lambda_N)$ takes on a single value with probability 1, the multipartite system behaves as if in a product state, i.e. $P(\mathbf{a}|\mathbf{x})$ is the product of the local distributions $P^{(i)}(a_i|x_i, \lambda_i)$.

However, it is conceivable that $\boldsymbol{\lambda}$ is not deterministically a single value, but takes on different values with different probabilities; in this case P becomes a convex combination of the product states corresponding to each value of $\boldsymbol{\lambda}$. The joint probability distribution P_Λ then describes a kind of statistical mixing over the joint physical states of the multipartite system.

In this section we show that arbitrary non-signaling distributions can be generated if we relax the positivity requirement on P_Λ: this is analogous not to the

local measurements generating outcomes in an unphysical manner, but rather to an unphysical mixing of otherwise physically meaningful objects. Interestingly, in our construction the product distributions for fixed choices of λ_i are completely deterministic: for each measurement choice x_i and λ_i there is a specific outcome a_i which occurs with probability 1. Obviously, not much information about the distribution P is contained in the local distributions; instead it is P_Λ which carries this information.

Theorem 6. *An N-partite conditional probability distribution $P(\mathbf{a}|\mathbf{x})$ is non-signaling if and only if it can be represented in the form*

$$P(\mathbf{a}|\mathbf{x}) = \sum_{\lambda_1,\ldots,\lambda_N} \left(P^{(1)}(a_1|x_1,\lambda_1)\cdots P^{(N)}(a_N|x_N,\lambda_N)\right)\tilde{P}_\Lambda(\boldsymbol{\lambda}), \tag{3.39}$$

where $\tilde{P}_\Lambda(\boldsymbol{\lambda})$ is a quasiprobability distribution, and $P^{(i)}(a_i|x_i,\lambda_i)$ is a conditional probability distribution for each i.

Proof. The "if" direction of the proof is not difficult: each $P^{(i)}(a_i|x_i,\lambda_i)$ is a conditional probability distribution for each choice of λ_i, and hence normalized. Fixing the string $(\mathbf{a}|\mathbf{x})$ and some system i, then summing over all possible values of a_i gives:

$$\begin{aligned}\sum_{a_i=1}^{K_{x_i}} P(\mathbf{a}|\mathbf{x}) &= \sum_{a_i=1}^{K_{x_i}}\sum_{\boldsymbol{\lambda}} \left(P^{(1)}(a_1|x_1,\lambda_1)\cdots P^{(N)}(a_N|x_N,\lambda_N)\right)\tilde{P}_\Lambda(\boldsymbol{\lambda}) \\ &= \sum_{\boldsymbol{\lambda}}\left(\left(\sum_{a_i} P^{(i)}(a_i|x_i,\lambda_i)\right)\prod_{j\neq i} P^{(j)}(a_j|x_j,\lambda_j)\right)\tilde{P}_\Lambda(\boldsymbol{\lambda}).\end{aligned} \tag{3.40}$$

For each λ_i, by normalization $\left(\sum_{a_i} P^{(i)}(a_i|x_i,\lambda_i)\right) = 1$, so that the RHS has no dependence on x_i. Hence the distribution $P(\mathbf{a}|\mathbf{x})$ is non-signaling, and has the following well-defined marginals:

$$P(\mathbf{a}_\Omega|\mathbf{x}_\Omega) = \sum_{\boldsymbol{\lambda}}\left(\prod_{i\in\Omega} P^{(i)}(a_i|x_i,\lambda_i)\right)\tilde{P}_\Lambda(\boldsymbol{\lambda}) \tag{3.41}$$

To prove the converse, we must show that given any non-signaling distribution $P\left(a_1,\ldots,a_N|x_1,\ldots,x_N\right)$, there exists a set Λ_i of local variables at each system i, a joint distribution $\tilde{P}_\Lambda$ on the strings $(\lambda_1,\ldots,\lambda_N)$, and local outcome functions $P^{(i)}(a_i|x_i,\lambda_i)$ such that (3.39) holds.

With this in mind, let Λ_i be the set of all possible ordered measurement choice/ outcome pairs $[a_i',x_i']$, along with one additional value for each system, which we will refer to as ξ_i. The size of the set Λ_i is then $\sum_{x_i} K_{x_i}+1$. For a string $(\lambda_1,\ldots,\lambda_N)\in\Lambda_1\times\cdots\times\Lambda_N$, let $\Omega=\{i\in[N]\,:\,\lambda_i\neq\xi_i\}$ (i.e. the set of indices i for which $\lambda_i\neq\xi_i$) and define

$$\tilde{P}_\Lambda(\boldsymbol{\lambda})=\left[\prod_{i\notin\Omega}(1-M^{(i)})\right]P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right) \tag{3.42}$$

so that, for example,

$$\begin{aligned}\tilde{P}_\Lambda([a_1',x_1'],[a_2',x_2'],\ldots,[a_N',x_N']) &= P\left(a_1',\ldots,a_N'|x_1',\ldots,x_N'\right)\\ \tilde{P}_\Lambda(\xi_1,[a_2',x_2'],\ldots,[a_N',x_N']) &= (1-M^{(1)})P\left(a_2',\ldots,a_N'|x_2',\ldots,x_N'\right)\\ &\vdots=\vdots\\ \tilde{P}_\Lambda(\xi_1,\xi_2,\ldots,\xi_N) &= (1-M^{(1)})\cdots(1-M^{(N)}).\end{aligned} \tag{3.43}$$

To show that $\tilde{P}_\Lambda$ is indeed a quasiprobability distribution, it must be checked that the sum of $P_\Lambda(\boldsymbol{\lambda})$ over all possible strings $\boldsymbol{\lambda}$ is equal to 1. By separating out this sum according to the subset Ω of components $\boldsymbol{\lambda}$ which do not equal ξ_i, we obtain:

$$\sum_{\lambda_1,\ldots,\lambda_N}\tilde{P}_\Lambda(\boldsymbol{\lambda})=\sum_{\Omega\subseteq[N]}\sum_{\mathbf{a}_\Omega,\mathbf{x}_\Omega}\left[\prod_{i\notin\Omega}(1-M^{(i)})\right]P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right). \tag{3.44}$$

Note that since $P\left(\mathbf{a}|\mathbf{x}\right)$ is a normalized outcome distribution, the sum over $\mathbf{a}_\Omega$ of

$P(\mathbf{a}_\Omega|\mathbf{a}_\Omega)$ is equal to 1. Summing also over $\mathbf{x}_\Omega$ gives:

$$\sum_{\lambda_1,\ldots,\lambda_N} \tilde{P}_\Lambda(\boldsymbol{\lambda}) = \sum_{\Omega\subseteq[N]} \left[\prod_{i\notin\Omega}(1-M^{(i)})\right]\left[\prod_{j\in\Omega} M^{(j)}\right]. \tag{3.45}$$

The RHS is now the expansion of the product of terms of $(M^{(i)} + (1 - M^{(i)}))$ as i ranges from to N, each term equaling 1;

$$\sum_{\lambda_1,\ldots,\lambda_N} \tilde{P}_\Lambda(\boldsymbol{\lambda}) = \prod_{i\in[N]} \left(M^{(i)} + (1-M^{(i)})\right) = 1. \tag{3.46}$$

The $P^{(i)}(a_i|x_i,\lambda_i)$ are defined to be the following conditional probability distributions, which treat the maximal outcome values as a special case:

$$P^{(i)}(a_i|x_i,\lambda_i) = \begin{cases} \delta_{\lambda_i,[a_i,x_i]} & \text{if } a_i < K_{x_i} \\ 1 - \sum_{a_i'<K_{x_i}} \delta_{\lambda_i,[a_i',x_i]} & \text{if } a_i = K_{x_i} \end{cases} \tag{3.47}$$

Suppose that $\tilde{P}'(\mathbf{a}|\mathbf{x})$ is the conditional quasiprobability distribution generated by the local functions $P^{(i)}(a_i|x_i,\lambda_i)$ according to the quasiprobabilistic mixture $\tilde{P}_\Lambda$, i.e.

$$\tilde{P}'(\mathbf{a}|\mathbf{x}) = \sum_{\boldsymbol{\lambda}} \left(P^{(1)}(a_1|x_1,\lambda_1)\cdots P^{(N)}(a_N|x_N,\lambda_N)\right)\tilde{P}_\Lambda(\boldsymbol{\lambda}). \tag{3.48}$$

By exactly the same argument as in the "if" direction at the beginning of this proof, $\tilde{P}'$ is non-signaling. Let Ω be any subset of $[N]$ and let $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ be any nowhere-maximal string. The marginal distribution $\tilde{P}'(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$ is then given by:

$$\tilde{P}'(\mathbf{a}_\Omega|\mathbf{x}_\Omega) = \sum_{\boldsymbol{\lambda}} \left(\prod_{i\in\Omega} P^{(i)}(a_i|x_i,\lambda_i)\right)\tilde{P}_\Lambda(\boldsymbol{\lambda}). \tag{3.49}$$

Note that for non-maximal a_i, $P^{(i)}(a_i|x_i,\lambda_i) = \delta_{\lambda_i,[a_i,x_i]}$, thus the only strings $\boldsymbol{\lambda}$ contributing to the sum are those which, for all $i\in\Omega$, have $\lambda_i = [a_i,x_i]$. For any

such $\boldsymbol{\lambda}$, let $\Omega' \subseteq [N]$ be the subset of components not equal to ξ_i, and note that $\Omega \subseteq \Omega'$. By collecting summands for which Ω' is the same, we have,

$$\tilde{P}'\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right) = \sum_{\substack{\Omega': \\ \Omega\subseteq\Omega'\subseteq[N]}} \left\{ \left[\prod_{i\notin\Omega'} \left(1 - M^{(i)}\right)\right] \sum_{\substack{a_j,x_j: \\ j\in\Omega'\setminus\Omega}} P\left(\mathbf{a}_{\Omega'}|\mathbf{x}_{\Omega'}\right) \right\}. \tag{3.50}$$

For each Ω', summing $P\left(\mathbf{a}_{\Omega'}|\mathbf{x}_{\Omega'}\right)$ over a_j for $j \in \Omega' \setminus \Omega$ gives the marginal value $P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right)$. Summing also over x_j for $j \in \Omega' \setminus \Omega$ gives a factor of $M^{(i)}$, so that,

$$\tilde{P}'\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right) = \sum_{\substack{\Omega': \\ \Omega\subseteq\Omega'\subseteq[N]}} \left\{ \left[\prod_{i\notin\Omega'} \left(1 - M^{(i)}\right)\right] \left[\prod_{j\in\Omega'\setminus\Omega} M^{(j)}\right] P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right) \right\}. \tag{3.51}$$

By the same reasoning as used for showing that $\tilde{P}_\Lambda$ is normalized, the sum over Ω' can be rewritten as a product of terms $\left(M^{(i)} + (1 - M^{(i)})\right)$ for $i \in \Omega' \setminus \Omega$, which (after factoring out $P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right)$) is therefore equal to 1. Hence $\tilde{P}'\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right) = P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right)$.

$\tilde{P}'\left(\mathbf{a}|\mathbf{x}\right)$ is a non-signaling conditional quasiprobability distribution which agrees with $P\left(\mathbf{a}|\mathbf{x}\right)$ for all non-maximal marginal values. It follows from Lemma 1 that $\tilde{P}$ is in fact the outcome distribution P. □

We mentioned before proving the Theorem that the value of λ_i can be seen as a codification of the local physical state, or configuration, which completely determines the *local* outcome probabilities via the distribution $P^{(i)}(a_i|x_i, \lambda_i)$. The size of the set of strings of local variables $\Lambda_1 \times \cdots \times \Lambda_N$ can loosely be seen as an indicator of how efficient that particular representation is; note that in the above proof, the size of this set is is $\prod_{k=1}^{N} \left(1 + \sum_{x_i} K_{x_i}\right)$.

It is possible to reduce this number by noting that certain values of λ_i give rise to exactly the same local functions. In particular, when λ_k is equal to either ξ_i or $[K_{x_i}, x_i]$ it will always be the case that $\delta_{[a_i,x_i],\lambda_i} = 0$ for $a_i < K_{x_i}$. This implies that each local conditional probability distribution deterministically assigns the great-

est value outcome K_{x_i} for all measurements x_i: $P^{(i)}(a_i|x_i,\lambda_i) = \delta_{a_i,K_{x_i}}$. Combining this set of local states into a single state η_i, say, for which $P^{(i)}(a_i|x_i,\eta_i) = \delta_{a_i,K_{x_i}}$, leaves the generated probability distribution intact, as long as the statistical weighting of any given product of local distributions $\prod_{i=1}^{N} P^{(i)}(a_i|x_i,\lambda_i)$ is the same.

For clarity, this compression of the local variable space is not used in the main presentation of the theorem. However, it brings the size of $\Lambda_1 \times \cdots \times \Lambda_N$ down to $\prod_{i=1}^{N}(1 + \sum_{x_i}(K_{x_i} - 1))$. This has a striking comparison with Lemma 1, in which we deduced that a conditional quasiprobability distribution is characterized by its set of non-maximal, marginal values. The number of such values is:

$$\sum_{\emptyset \neq \Omega \subseteq [N]} \left(\prod_{i \in \Omega} \left[\sum_{x_i} (K_{x_i} - 1) \right] \right) = \prod_{i=1}^{N} \left(1 + \sum_{x_i} (K_{x_i} - 1) \right) - 1. \qquad (3.52)$$

The extra -1 on the RHS of (3.52) occurs because when $\Omega = \emptyset$, we demand that $P(\emptyset|\emptyset) = 1$ for normalization. This roughly corresponds to the case $\boldsymbol{\lambda} = (\eta_1, \ldots, \eta_N)$, whose weighting can be interpreted as fixed by normalization of P_Λ.

3.3.2 Non-positive effects

In this section we show that a dual result to Theorem 6 holds in which the local functions become conditional quasiprobability distributions, whilst the joint distribution P_Λ remains a genuine probability distribution. Recall that P_Λ can be regarded as a statistical mixture over the different possible products of local states, where each such product is represented by a string $\boldsymbol{\lambda}$ specifying the local states on each system. In this section we explore what can be generated when those local states are allowed to produce certain outcomes with negative probabilities. Despite taking a statistical mixture of completely local objects, one is able to generate the entire set of non-signaling outcome distributions.

Theorem 7. *An N-partite conditional probability distribution* $P(\mathbf{a}|\mathbf{x})$ *is non-signaling*

if and only if it can be represented in the form

$$P\left(\mathbf{a}|\mathbf{x}\right) = \sum_{\lambda_1,\ldots,\lambda_N} \left(\tilde{P}^{(1)}(a_1|x_1,\lambda_1)\cdots\tilde{P}^{(N)}(a_N|x_N,\lambda_N)\right) P_\Lambda(\boldsymbol{\lambda}) \qquad (3.53)$$

where $P_\Lambda(\boldsymbol{\lambda})$ is a probability distribution, and $\tilde{P}^{(i)}(a_i|x_i,\lambda_i)$ is a conditional quasiprobability distribution for each i.

Proof. As in Theorem 6, the "if" direction is not difficult; since the $\tilde{P}^{(i)}(a_i|x_i,\lambda_i)$ are conditional quasiprobability distributions, they satisfy $\sum_{a_i}\tilde{P}^{(i)}(a_i|x_i,\lambda_i) = 1$ (independent of the choice of x_i). It follows from a simple algebraic substitution that the distribution given by (3.53) is non-signaling.

Conversely, to prove that all non-signaling distributions $P\left(\mathbf{a}|\mathbf{x}\right)$ can be written in the form of (3.53), we again have to make a choice of local conditional quasiprobability distributions and strings $\boldsymbol{\lambda}$ with a joint probability distribution P_Λ. Let Λ_i be the set of ordered pairs $[a_i', x_i']$ consisting of the allowed measurement choices and outputs for system i (note that we do not require the extra variable ξ_i this time). The local variables are then given by:

$$P_\Lambda([a_1',x_1'],\ldots,[a_N',x_N']) = \frac{P\left(a_1',\ldots,a_N'|x_1',\ldots,x_N'\right)}{M^{(1)}M^{(2)}\cdots M^{(N)}} \qquad (3.54)$$

Clearly, each value of $P_\Lambda(\boldsymbol{\lambda})$ is positive and lies in $[0,1]$. Moreover, summing $P\left(\mathbf{a}'|\mathbf{x}'\right)$ over $\mathbf{a}'$ gives 1 by normalization; the number of values of $\mathbf{x}'$ to sum over is then the product of the number of measurement choices at each site. Thus P_Λ is a normalized probability distribution.

The local distributions are defined as follows :

$$\tilde{P}^{(i)}(a_i|x_i,\lambda_i) = \begin{cases} M^{(i)}\delta_{\lambda_i,[a_i,x_i]} & \text{if } a_i < K_{x_i} \\ 1 - \sum_{a_i'<K_{x_i}} M^{(i)}\delta_{\lambda_i,[a_i',x_i]} & \text{if } a_i = K_{x_i} \end{cases} \qquad (3.55)$$

For each choice of x_i and λ_i, $\tilde{P}^{(i)}(K_{x_i}|x_i,\lambda_i)$ is specifically constructed so that the sum over a_i of $\tilde{P}^{(i)}(a_i|x_i,\lambda_i)$ is equal to 1. Therefore $\tilde{P}^{(i)}(a_i|x_i,\lambda_i)$ is a conditional quasiprobability distribution.

It remains to show that the given values for $P_\Lambda(\boldsymbol{\lambda})$ and $\tilde{P}^{(i)}(a_i|x_i,\lambda_i)$ satisfy (3.53). Consider the quasiprobability distribution $\tilde{P}'(\mathbf{a}|\mathbf{x})$ given by

$$\tilde{P}'(\mathbf{a}|\mathbf{x}) = \sum_{\lambda_1,\ldots,\lambda_N} \left(\tilde{P}^{(i)}(a_i|x_i,\lambda_i)\cdots\tilde{P}^{(N)}(a_N|x_N,\lambda_N)\right) P_\Lambda(\boldsymbol{\lambda}). \tag{3.56}$$

Again, a simple algebraic substitution demonstrates that $\tilde{P}'(\mathbf{a}|\mathbf{x})$ is non-signaling. For any $\Omega \subseteq [N]$, and nowhere-maximal string $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$,

$$\begin{aligned}
\tilde{P}'(\mathbf{a}_\Omega|\mathbf{x}_\Omega) &= \sum_{\boldsymbol{\lambda}} \left(\prod_{i\in\Omega} \tilde{P}^{(i)}(a_i|x_i,\lambda_i)\right) P_\Lambda(\boldsymbol{\lambda}) \\
&= \sum_{(\mathbf{a}'|\mathbf{x}')} \left(\prod_{i\in\Omega} M^{(i)}\delta_{a_i',a_i}\delta_{x_i',x_i}\right) \frac{P(\mathbf{a}'|\mathbf{x}')}{\prod_{j\in[N]} M^{(j)}} \\
&= \left(\prod_{i\in\Omega} M^{(i)}\right) \frac{\left(\prod_{k\in[N]\setminus\Omega} M^{(k)}\right) P(\mathbf{a}_\Omega|\mathbf{x}_\Omega)}{\prod_{j\in[N]} M^{(j)}} \\
&= P(\mathbf{a}_\Omega|\mathbf{x}_\Omega).
\end{aligned} \tag{3.57}$$

For example, when $\Omega = \{1,2\}$,

$$\begin{aligned}
P'(a_1,a_2|x_1,x_2) &= \sum_{\lambda_1,\ldots,\lambda_N} M^{(1)}\delta_{\lambda_1,[a_1,x_1]} M^{(2)}\delta_{\lambda_2,[a_2,x_2]} P_\Lambda(\lambda_1,\ldots,\lambda_N). \\
&= M^{(1)}M^{(2)} \frac{P(a_1,a_2|x_1,x_2)}{M^{(1)}M^{(2)}} \\
&= P(a_1,a_2|x_1,x_2).
\end{aligned} \tag{3.58}$$

Since P and $\tilde{P}'$ are quasiprobability distributions whose non-maximal marginals agree, it follows again from Lemma 1 that they are identical and that (3.53) holds. □

Again, it is interesting to look at the size of the set of strings of local variables; in this case, $\prod_{i=1}^N (\sum_{x_i} K_{x_i})$. A very similar reduction can be made as for Theorem 6, since whenever $\lambda_i = [K_{x_i}, x_i]$ for any x_i, the local outcome function again takes

the form $P^{(i)}(a_i|x_i,\lambda_i) = \delta_{a_i,K_{x_i}}$. Hence it is possible to collect these $M^{(i)}$ local variables at each system i into a single local variable η_i, again adjusting the joint probability distribution P_Λ in order to generate the same outcome statistics and preserve normalization. Just as before, this brings the total size of the set of strings of local variables to $\prod_{i=1}^{N}\left(\sum_{x_i}(K_{x_i}-1)+1\right)$.

3.3.3 Quantum corollaries

In this section we show how quasiprobabilistic quantum representations may be explicitly constructed from quasiprobabilistic local representations, thus recovering the results of Section 3.1 as a corollary of Theorem 6, and obtaining a dual quantum representation as a corollary of Theorem 7. This involves a careful construction of density matrix ρ and measurements operators $M^{(i)}_{a_i|x_i}$ which, upon application of the standard Born trace rule, reduce to the local quasidistributions constructed in Section 3.3.

Corollary 1. *An N-partite conditional probability distribution $P(\mathbf{a}|\mathbf{x})$ is nonsignaling if and only if it can be represented in the form*

$$P(\mathbf{a}|\mathbf{x}) = Tr\left(\left(M^{(1)}_{a_1|x_1}\otimes\cdots\otimes M^{(N)}_{a_N|x_N}\right)\tilde{\rho}\right), \tag{3.59}$$

where $\tilde{M}^{(i)}_{a_i|x_i}$ are genuine POVM elements (positive operators satisfying $\sum_{a_i} M^{(i)}_{a_i|x_i} = \mathbb{I}^{(i)}$) for each i, and $\tilde{\rho}$ is a Hermitian operator satisfying $Tr(\rho) = 1$. Furthermore this representation can be chosen such that the operators $M^{(1)}_{a_1|x_1}\otimes\cdots\otimes M^{(N)}_{a_N|x_N}$ and $\tilde{\rho}$ all commute.

Proof. To prove the "if" direction, sum the right-hand side of (3.59) over a_i using the fact that $\sum_{a_i} M^{(i)}_{a_i|x_i} = \mathbb{I}^{(i)}$. To prove the converse, we assign a Hilbert space with orthonormal basis $\{|\lambda_i\rangle|\lambda_i\in\Lambda_i\}$ to each system i, where Λ_i is the set of all ordered pairs $[a_i,x_i]$ and the extra element ξ_i (as in the proof of Theorem 2). We

then take

$$\tilde{\rho} = \sum_{\lambda_1,\ldots,\lambda_N} \tilde{P}_\Lambda(\boldsymbol{\lambda})\ |\lambda_1\rangle\langle\lambda_1| \otimes \cdots \otimes |\lambda_N\rangle\langle\lambda_N|, \tag{3.60}$$

$$M^{(i)}_{a_i|x_i} = \sum_{\lambda_i} P^{(i)}(a_i|x_i,\lambda_i)\,|\lambda_i\rangle\langle\lambda_i|, \tag{3.61}$$

where $\tilde{P}_\Lambda(\boldsymbol{\lambda})$ and $P^{(i)}(a_i|x_i,\lambda_i)$ are given by (3.42) and (3.47) respectively. Substituting these into (3.59) gives:

$$\mathrm{Tr}\left(\left(M^{(1)}_{a_1|x_1} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}\right)\tilde{\rho}\right) = \sum_{\boldsymbol{\lambda}} \left[\prod_{i=1}^{N} P^{(i)}(a_i|x_i,\lambda_i)\right] \tilde{P}_\Lambda(\boldsymbol{\lambda}). \tag{3.62}$$

Thus we recover the local quasiprobability distribution constructed in the proof of Theorem 6, and proven there to agree exactly with $P\left(\mathbf{a}|\mathbf{x}\right)$. □

A similar corollary to Theorem 7 holds, which provides a kind of dual result to Theorem 5.

Corollary 2. *An N-partite conditional probability distribution $P\left(\mathbf{a}|\mathbf{x}\right)$ is non-signaling if and only if it can be represented in the form*

$$P\left(\mathbf{a}|\mathbf{x}\right) = Tr\left(\left(\tilde{M}^{(1)}_{a_1|x_1} \otimes \cdots \otimes \tilde{M}^{(N)}_{a_N|x_N}\right)\rho\right), \tag{3.63}$$

where $\tilde{M}^{(i)}_{a_i|x_i} \in \mathcal{L}\left(\mathcal{H}_i\right)$ are Hermitian operators satisfying $\sum_{a_i} \tilde{M}^{(i)}_{a_i|x_i} = \mathbb{I}^{(i)}$ for each i, and ρ is a genuine density operator. Furthermore, this representation can be chosen such that the operators $\tilde{M}^{(1)}_{a_1|x_1} \otimes \cdots \otimes \tilde{M}^{(N)}_{a_N|x_N}$ and ρ all commute.

Proof. Again, the `if" direction follows directly from substitution of the fact that $\sum_{a_i} \tilde{M}^{(i)}_{a_i|x_i} = \mathbb{I}^{(i)}$. To prove the converse, we use the results of Theorem 7. To each system i, we assign a Hilbert space with orthonormal basis $\{|\lambda_i\rangle|\lambda_i \in \Lambda_i\}$,

where Λ_i is the set of all ordered pairs $[a_i, x_i]$. We then take

$$\rho = \sum_{\lambda_1,\ldots,\lambda_N} P_\Lambda(\boldsymbol{\lambda})\ |\lambda_1\rangle\langle\lambda_1| \otimes \cdots \otimes |\lambda_N\rangle\langle\lambda_N|, \tag{3.64}$$

$$\tilde{M}^{(i)}_{a_i|x_i} = \sum_{\lambda_i} \tilde{P}^{(i)}(a_i|x_i,\lambda_i)\ |\lambda_i\rangle\langle\lambda_i|, \tag{3.65}$$

where $P_\Lambda(\lambda_1,\ldots,\lambda_N)$ and $\tilde{P}^{(i)}(a_i|x_i,\lambda_i)$ are given by (3.54) and (3.55) respectively. Substituting this choice of operators into (3.63) leads to exactly the same equation as (3.53): therefore the result follows directly from Theorem 7. □

Note that, from the discussions directly following the proofs of these Theorems 6 and 7, the dimension of the density operator in both Corollaries can be taken to be $d = \prod_{i=1}^{N} \left(1 + \sum_{x_i}(K_{x_i} - 1)\right)$.

3.4 Further quasiprobabilistic quantum models

In Section 3.3 it was shown that there are two distinct modifications of classical probability theory which allow one to generate arbitrary non-signaling distributions. These are dual to one another: one way is to allow the local distributions to take on negative values, the other way is to allow the joint probability distribution to take on negative values. Each of these classical representations admits a quantum-like corollary, in which the POVM elements and the density operators are all diagonal with respect to some basis, but one or the other are no longer positive operators.

In this Section we discuss further quantum-like representations in which the observables are represented by non-positive operators. In contrast to previous sections, the density matrix does not depend on the specific probability distribution $P\left(a_1,\ldots,a_N|x_1,\ldots,x_N\right)$ that is being generated. In fact, the density matrix that we use will be not only separable, but fixed by the number of measurements and measurement outcomes on each subsystem. This provides a result which is

stronger than Corollary 2, in which the density matrix varies according to the outcome distribution being generated. We then demonstrate the interesting result that it is possible to reduce the correlation present in our fixed density matrix so that it is arbitrarily close to the set of product states, whilst still being able to generate arbitrary non-signaling distributions by acting on it with commuting, non-positive observables. The cost of this manoeuvre is that the negativity of the measurement operators becomes much larger.

3.4.1 Non-positive observables acting on $\rho^{N,d}$

From here on we restrict ourselves to the scenario in which each local measurement has the same number of possible outcomes, given by the integer K, and each party has the same number of measurement choices, an integer M. This is intended to simplify the presentation, and (as will be seen) allows for an intriguing generalization in the following section.

Once M, K, and the number of parties is fixed, the density matrix we will use is also fixed, and only the measurement operators vary in order to generate different distributions. In an N-partite quantum system where each local Hilbert space has dimension d and a chosen basis $\{|i\rangle\}_{i=1}^{d}$, define the separable state:

$$\rho^{N,d} = \frac{1}{d}\left(|1\rangle^{\otimes N}\langle 1|^{\otimes N} + \ldots + |d\rangle^{\otimes N}\langle d|^{\otimes N}\right). \tag{3.66}$$

Note that this matrix depends only on the two parameters N and d: we will use different values of N and and d and hence different density matrices at various stages in the proof, before giving the final density matrix (which is also defined as $\rho^{N,d}$ for specified N and d).

The following Lemma hints at the usefulness of this particular separable state for our purposes. If, for the moment, we drop not only the positivity of measurement operators, but also the requirement that summing over a measurement choice x_i gives the identity, then it is possible to generate *any* set of real (or, indeed, complex) numbers via the Born trace rule, not only those that are normalized

or non-signaling.

Lemma 3. *Let $Q(a_1,\ldots,a_N|x_1,\ldots,x_N) = Q(\mathbf{a}|\mathbf{x})$ be any set of real numbers, not necessarily positive, normalized or non-signaling. Then there exists an integer d and operators $\tilde{M}^{(i)}_{a_i|x_i}$ on a Hilbert space of dimension d, such that all the N-fold tensor products of these operators commute with $\rho^{N,d}$, and*

$$Q(\mathbf{a}|\mathbf{x}) = Tr\left(\left(\tilde{M}^{(1)}_{a_1|x_1}\otimes\cdots\otimes\tilde{M}^{(N)}_{a_N|x_N}\right)\rho^{N,d}\right) \tag{3.67}$$

Proof. Let $d=(M\cdot K)^{N-1}$ and let $\{|e_{a|x}\rangle : 1\le a\le K, 1\le x\le M\}$ be an orthonormal basis of $\mathbb{C}^{M\cdot K}$. In the (N-1)-fold tensor product space $\mathbb{C}^{M\cdot K}\otimes\ldots\otimes\mathbb{C}^{M\cdot K}$, for $1\le i\le N-1$ define the matrices:

$$\tilde{M}^{(i)}_{a_i|x_i} = \mathbb{I}^{(1)}\otimes\cdots\otimes\mathbb{I}^{(i-1)}\otimes|e_{a_i|x_i}\rangle\langle e_{a_i|x_i}|\otimes\mathbb{I}^{(i+1)}\otimes\cdots\otimes\mathbb{I}^{(N-1)}, \tag{3.68}$$

and for $i=N$ define

$$\tilde{M}^{(N)}_{a_N|x_N} = \sum_{\substack{(\mathbf{a}'|\mathbf{x}'):\\ a'_N=a_N,\\ x'_N=x_N}}\left(d\cdot Q(\mathbf{a}'|\mathbf{x}')\bigotimes_{i=1}^{N-1}|e_{a'_i|x'_i}\rangle\langle e_{a'_i|x'_i}|\right), \tag{3.69}$$

Note that all the above operators are defined on a d-dimensional local vector space $\mathbb{C}^d$. The vectors $|e_{a_1,\ldots,a_{N-1}|x_1,\ldots,x_{N-1}}\rangle = |e_{a_1|x_1}\rangle\otimes\cdots\otimes|e_{a_{N-1}|x_{N-1}}\rangle$ form an orthonormal basis of $\mathbb{C}^d$, with respect to which all the above operators are diagonal. This is the basis which we will use to define the density matrix $\rho^{N,d}$ as in (3.66); that is to say,

$$\rho^{N,d} = \frac{1}{d}\sum_{\substack{a_1,\ldots,a_{N-1}\\ x_1,\ldots,x_{N-1}}}|e_{a_1,\ldots,a_{N-1}|x_1,\ldots,x_{N-1}}\rangle^{\otimes N}\langle e_{a_1,\ldots,a_{N-1}|x_1,\ldots,x_{N-1}}|^{\otimes N} \tag{3.70}$$

Taking the trace of tensor products of the above operators acting on the state $\rho^{N,d}$

yields the following simplification:

$$\begin{aligned}&\mathrm{Tr}\left(\left(\tilde{M}^{(1)}_{a_1|x_1}\otimes\cdots\otimes\tilde{M}^{(N)}_{a_N|x_N}\right)\rho^{N,d}\right)\\ =&\frac{1}{d}\sum_{\substack{a'_1,\ldots,a'_{N-1},\\ x'_1,\ldots,x'_{N-1}}}\prod_{i=1}^{N}\langle e_{a'_1,\ldots,a'_{N-1}|x'_1,\ldots,x'_{N-1}}|\tilde{M}^{(i)}_{a_i|x_i}|e_{a'_1,\ldots,a'_{N-1}|x'_1,\ldots,x'_{N-1}}\rangle.\end{aligned}\tag{3.71}$$

For $i<N$, by construction each term in the product is equal to zero, except in the case where $a'_i=a_i$, $x'_i=x_i$, in which case the term is equal to 1. Thus,

$$\begin{aligned}&\mathrm{Tr}\left(\left(\tilde{M}^{(1)}_{a_1|x_1}\otimes\cdots\otimes\tilde{M}^{(N)}_{a_N|x_N}\right)\rho^{N,d}\right)\\ &=\frac{1}{d}\langle e_{a_1,\ldots,a_{N-1}|x_1,\ldots,x_{N-1}}|M^{(N)}_{a_N|x_N}|e_{a_1,\ldots,a_{N-1}|x_1,\ldots,x_{N-1}}\rangle\\ &=Q\left(a_1,\ldots,a_N|x_1,\ldots,x_N\right).\end{aligned}\tag{3.72}$$

□

This result is used repeatedly in the argument of the main proof of this section. More precisely, it is applied once for each subset $\Omega\subseteq[N]$ in order to generate the non-maximal marginals of the distribution P on that subset. Note that these non-maximal marginal values $P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right)$ are simply a set of real numbers between 0 and 1; without the maximal outcome values a_i, they do not in general obey the normalisation or non-signaling conditions.

Theorem 8. *Let $P\left(a_1,\ldots,a_N|x_1,\ldots,x_N\right))$ be any outcome distribution on N parties, such that each party has M measurement choices, and K possible outcomes for each measurement. The distribution is non-signaling if and only if there exist (not necessarily positive) local operators $\tilde{M}^{(i)}_{a_i|x_i}$ on a Hilbert space of dimension d such that for each i and choice of x_i, $\sum_{a_i=1}^{K}\tilde{M}^{(i)}_{a_i|x_i}=\mathbb{I}^{(i)}$ (the identity on system i) and*

$$P\left(a_1,\ldots,a_N|x_1,\ldots,x_N\right)=\mathit{Tr}\left(\left(\tilde{M}^{(1)}_{a_1|x_1}\otimes\cdots\otimes\tilde{M}^{(N)}_{a_N|x_N}\right)\rho^{N,d}\right).\tag{3.73}$$

Proof. As with many of the proofs in this chapter, the "if" direction is little more than an algebraic substitution of the condition $\sum_{a_i=1}^{K} \tilde{M}^{(i)}_{a_i|x_i} = \mathbb{I}^{(i)}$. The converse direction proceeds by applying the result of Lemma 3 to each subset $\Omega \subseteq [N]$ in a recursive manner, to obtain a set of matrices $\{\tilde{M}^{(i)}_{a_i|x_i}(\Omega)\}_{a_i<K}$. The final measurement matrices will be the direct summands $\tilde{M}^{(i)}_{a_i|x_i} = \bigoplus_{\Omega\subseteq[N]} \tilde{M}^{(i)}_{a_i|x_i}(\Omega)$, and by construction at each stage we will ensure that these final matrices will generate the correct marginal outcome statistics on Ω, for all nowhere-maximal marginal strings $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$.

Let $d = \sum_{i=1}^{N}(M\cdot(K-1))^{i-1}\binom{N}{i}$ (the reason for this will be clear later). Applying Lemma 3 with $\Omega = [N]$, there exist operators $\{\tilde{M}^{(i)}_{a_i|x_i}([N])\}_{a_i<K}$ of dimension $d([N]) = (M\cdot(K-1))^{N-1}$ such that for all nowhere-maximal strings $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$,

$$\begin{aligned}&\mathrm{Tr}\left(\left(\tilde{M}^{(1)}_{a_1|x_1}([N])\otimes\cdots\otimes\tilde{M}^{(N)}_{a_N|x_N}([N])\right)\rho^{N,d([N])}\right)\\&=\frac{d}{d([N])}\cdot P\left(a_1,\ldots,a_N|x_1,\ldots,x_N\right).\end{aligned}\tag{3.74}$$

Let $\Omega = \{i_1,\ldots,i_k\} \subset [N]$. By Lemma 3 it is possible to find operators $\tilde{M}^{(i)}_{a_i|x_i}(\Omega)$ such that

$$\mathrm{Tr}\left(\left(\hat{\bigotimes}_{i\in\Omega}\tilde{M}^{(i)}_{a_i|x_i}(\Omega)\right)\rho^{k,d(\Omega)}\right) = Q\left(a_1,\ldots,a_k|x_1,\ldots,x_k\right)\tag{3.75}$$

for any set of numbers $Q\left(a_1,\ldots,a_k|x_1,\ldots,x_k\right)$, as long as $d(\Omega)$ equals the number of possible strings $(a_1,\ldots,a_k|x_1,\ldots,x_k)$.

Assuming that the operators $\tilde{M}^{(i)}_{a_i|x_i}(\Omega')$ have been fixed for all $\Omega\subset\Omega'\subseteq[N]$, apply Lemma 3 to obtain operators $\tilde{M}^{(i)}_{a_i|x_i}(\Omega)\in\mathcal{L}\left(\mathbb{C}^{d(\Omega)}\right)$ for all $|\Omega| = k$ parties, where $d(\Omega) = (M\cdot(K-1))^{|\Omega|-1}$, such that for all nowhere-maximal strings

$(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$,

$$\begin{aligned} & d(\Omega) \cdot \mathrm{Tr}\left(\left(\tilde{M}^{(i_1)}_{a_{i_1}|x_{i_1}}(\Omega) \otimes \cdots \otimes \tilde{M}^{(i_k)}_{a_{i_k}|x_{i_k}}(\Omega)\right)\rho^{k,d(\Omega)}\right) \\ & = d \cdot P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right) \\ & - \sum_{\Omega \subset \Omega' \subseteq [N]} d(\Omega') \cdot \mathrm{Tr}\left(\left(\tilde{M}^{(i_1)}_{a_{i_1}|x_{i_1}}(\Omega') \otimes \cdots \otimes \tilde{M}^{(i_k)}_{a_{i_k}|x_{i_k}}(\Omega')\right)\rho^{k,d(\Omega')}\right). \quad (3.76) \end{aligned}$$

For $i \notin \Omega$, define $\tilde{M}^{(i)}_{a_i|x_i}\Omega$ to be the zero matrix, and define the overall measurement operators to be the direct sums of all the above operators for that choice of i, a_i and x_i:

$$\tilde{M}^{(i)}_{a_i|x_i} = \bigoplus_{\Omega \subseteq [N]} \tilde{M}^{(i)}_{a_i|x_i}(\Omega) \in \mathcal{L}\left(\bigoplus_{\Omega \subseteq [N]} \mathbb{C}^{d(\Omega)}\right). \quad (3.77)$$

The dimension of these operators is $\sum_{\Omega \subseteq [N]} d(\Omega)$, agreeing with the initial definition of d. Let $\rho^{N,d}$ be defined according to the orthonormal basis of $\bigoplus_{\Omega \subseteq [N]} \mathbb{C}^{d(\Omega)}$ which is the union of the orthonormal bases obtained in each application of Lemma 3 on the vector space $\mathbb{C}^{d(\Omega)}$. For any $\Omega \subseteq [N]$, let $\hat{\bigotimes}_{i\in\Omega}\tilde{M}^{(i)}_{a_i|x_i}(\Omega)$ again denote the N-fold tensor product whose ith components for $i \notin \Omega$ are equal to the identity $\mathbb{I}^{(i)}$. By expanding $\rho^{N,d}$ in terms of the chosen basis vectors, we see that for all nowhere-maximal strings $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$,

$$\mathrm{Tr}\left(\left(\hat{\bigotimes}_{i\in\Omega}\tilde{M}^{(i)}_{a_i|x_i}\right)\rho^{N,d}\right) = \frac{1}{d}\sum_{j=1}^{d}\prod_{i=1}^{N}\langle j|\tilde{M}^{(i)}_{a_i|x_i}|j\rangle. \quad (3.78)$$

Splitting up the sum on the RHS according to $d = \sum_{\Omega \subseteq [N]} d(\Omega)$, substituting $\langle j|\mathbb{I}^{(i)}|j\rangle = 1$, and noting that $\tilde{M}^{(i)}_{a_i|x_i}(\Omega)$ is the zero matrix for $i \notin \Omega$:

$$\mathrm{Tr}\left(\left(\hat{\bigotimes}_{i\in\Omega}\tilde{M}^{(i)}_{a_i|x_i}\right)\rho^{N,d}\right) = \frac{1}{d}\sum_{\substack{\Omega':\\ \Omega \subseteq \Omega' \subseteq [N]}}\sum_{j=1}^{d(\Omega')}\prod_{i=1}^{|\Omega'|}\langle j|\tilde{M}^{(i)}_{a_i|x_i}(\Omega')|j\rangle. \quad (3.79)$$

Note that the summand corresponding to each subset Ω' is in fact a scalar multiple of the trace of the product of $\hat{\bigotimes}_{i\in\Omega'} \tilde{M}^{(i)}_{a_i|x_i}(\Omega')$ with the density operator $\rho^{N,d(\Omega')}$. Hence, separating out the term corresponding to Ω in the sum:

$$\begin{aligned}\mathrm{Tr}\left(\left(\hat{\bigotimes}_{i\in\Omega}\tilde{M}^{(i)}_{a_i|x_i}\right)\rho^{N,d}\right) = \frac{d(\Omega)}{d}\mathrm{Tr}\left(\left(\hat{\bigotimes}_{i\in\Omega}\tilde{M}^{(i)}_{a_i|x_i}(\Omega)\right)\rho^{N,d(\Omega)}\right) \\ + \sum_{\substack{\Omega':\\ \Omega\subset\Omega'\subseteq[N]}} \frac{d(\Omega')}{d}\mathrm{Tr}\left(\left(\hat{\bigotimes}_{i\in\Omega}\tilde{M}^{(i)}_{a_i|x_i}(\Omega')\right)\rho^{N,d(\Omega')}\right).\end{aligned} \tag{3.80}$$

Substituting equation (3.76) for the first term on the RHS gives

$$\mathrm{Tr}\left(\left(\hat{\bigotimes}_{i\in\Omega}\tilde{M}^{(i)}_{a_i|x_i}\right)\rho^{N,d}\right) = P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right). \tag{3.81}$$

So far, we have only defined the measurement operators for non-maximal pairs (a_i, x_i) on each system i. However, as long as the remaining operators are defined so that $\sum_{a_i}\tilde{M}^{(i)}_{a_i|x_i} = \mathbb{I}^{(i)}$ for each i and choice of x_i, then the quasiprobability distribution generated by these measurement operators acting on the state $\rho^{N,d}$ will be non-signaling and normalized. By (3.81), the non-maximal marginals of this quasiprobability distribution agree with those of P; from Lemma 1 it then follows that this distribution is exactly $P(\mathbf{a}|\mathbf{x})$. To complete the proof, we need simply define $\tilde{M}^{(i)}_{K|x_i} = \mathbb{I}^{(i)} - \sum_{j=1}^{R-1}\tilde{M}^{(i)}_{j|x_i}$ for each i and x_i. □

Example Let $P(a_1, a_2|x_1, x_2)$ be any state of a g-bit system. In this case we require $d = 4$, $d(\{1,2\}) = 2$ and $d(\{1\}) = d(\{2\}) = 1$. Starting with $\Omega = \{1,2\}$,

following the technique in the proof of Lemma 3 we obtain

$$M^{(1)}_{0|0}(\{1,2\}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{3.82}$$

$$M^{(1)}_{0|1}(\{1,2\}) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{3.83}$$

$$M^{(2)}_{0|0}(\{1,2\}) = \begin{pmatrix} 4P\left(0,0|0,0\right) & 0 \\ 0 & 4P\left(0,0|0,1\right) \end{pmatrix}, \tag{3.84}$$

$$M^{(2)}_{0|1}(\{1,2\}) = \begin{pmatrix} 4P\left(0,0|1,0\right) & 0 \\ 0 & 4P\left(0,0|1,1\right) \end{pmatrix}. \tag{3.85}$$

Similarly, following the technique with $\Omega = \{1\}$ and $\Omega = \{2\}$ generates the following 1-dimensional matrices:

$$M^{(1)}_{0|0}(\{1\}) = (4P\left(a_1 = 0|x_1 = 0\right) - 1), \tag{3.86}$$

$$M^{(1)}_{0|1}(\{1\}) = (4P\left(a_1 = 0|x_1 = 1\right) - 1), \tag{3.87}$$

$$M^{(2)}_{0|0}(\{2\}) = (4[P\left(a_2 = 0|x_2 = 0\right) - P\left(0,0|0,0\right) - P\left(0,0|1,0\right)]), \tag{3.88}$$

$$M^{(2}_{0|1}(\{2\}) = (4[P\left(a_2 = 0|x_2 = 0\right) - P\left(0,0|0,1\right) - P\left(0,0|1,1\right)]). \tag{3.89}$$

The measurement operators for A are therefore:

$$M^{(1)}_{0|0} = Diag(1, 0, 4P\left(a_1 = 0|x_1 = 0\right) - 1, 0) \tag{3.90}$$

$$M^{(1)}_{0|1} = Diag(0, 1, 4P\left(a_1 = 0|x_1 = 1\right) - 1, 0) \tag{3.91}$$

$$M^{(1)}_{1|0} = \mathbb{I} - M^{(1)}_{1|1}, \qquad M^{(1)}_{1|1} = \mathbb{I} - M^{(2)}_{1|2}, \tag{3.92}$$

and the operators for B are:

$$M_{0|0}^{(2)} = Diag(4P\left(0,0|0,0\right),$$
$$4P\left(0,0|1,0\right),$$
$$0,$$
$$4[P\left(a_2=0|x_2=0\right) - P\left(0,0|0,0\right) - P\left(0,0|0,0\right)]) \quad (3.93)$$

$$M_{0|1}^{(2)} = Diag(4P\left(0,0|0,1\right),$$
$$4P\left(0,0|1,1\right),$$
$$0,$$
$$4[P\left(a_2=0|x_2=1\right) - P\left(0,0|0,1\right) - P\left(0,0|1,1\right)]) \quad (3.94)$$

$$M_{1|0}^{(2)} = \mathbb{I} - M_{0|0}^{(2)}, \qquad M_{1|1}^{(2)} = \mathbb{I} - M_{0|1}^{(2)}. \quad (3.95)$$

The density operator $\rho^{N,d}$ is represented by a 16x16 matrix with a 1 in the (i,i)th position for $i \in \{1,5,9,13\}$ and 0s elsewhere.

3.4.2 Approximate product states

In Section 3.2.1, the choice of POVM elements depended only on the number of parties, measurements and measurement outcomes; once these were fixed, it was by modifying the non-positive density operator that any given non-signaling outcome distribution could be generated. The procedure in Section 3.4.1 is in some sense dual to this: the density operator $\rho^{N,d}$ depends only on N (the number of parties) and the local dimension d, which itself depends on the number of measurement choices and outcomes. $\rho^{N,d}$ is specially chosen to simplify the proof of Lemma 3: specifically, multiplying $\rho^{N,d}$ by a tensor-product of local matrices and taking the trace leads to the simplification (3.71), enabling the construction of local matrices which generate an arbitrary set of real numbers $Q\left(\mathbf{a}|\mathbf{x}\right)$.

As we demonstrate in this section, it is possible to choose other initial density matrices as long as some form of Lemma 3 still applies. Perhaps surprisingly, it is possible to use any of a class of density operators which are of one greater dimension than $\rho^{N,d}$, but which come arbitrarily close to the set of product states. For $\varepsilon \in \mathbb{R}$, define:

$$\rho_{\varepsilon}^{N,d} = (1-\varepsilon)\,|0\rangle^{\otimes N}\langle 0|^{\otimes N} + \varepsilon \sum_{i=1}^{d} |i\rangle^{\otimes N}\langle i|^{\otimes N} \quad . \tag{3.96}$$

This is a valid density matrix for $0 < \epsilon < 1/\sqrt{d}$, and as ε approaches zero, it approaches the product state $|0\rangle^{\otimes N}\langle 0|^{\otimes N}$. $\rho_{\varepsilon}^{N,d}$ is defined on the vector space $\left(\mathbb{C} \oplus \mathbb{C}^d\right)^{\otimes N}$, where d takes the same value as in the proof of Theorem 8. Recall the construction of operators $\tilde{M}_{a_i|x_i}^{(i)}(\Omega)$ for each $\Omega \subseteq [N]$. For non-maximal pairs (a_i, x_i), consider the new operators:

$$\tilde{N}_{a_i|x_i}^{(i)}(\Omega) = \begin{cases} \tilde{M}_{a_i|x_i}^{(i)}(\Omega) & \text{if } i \neq \max(\Omega) \\ \frac{1}{d\cdot\varepsilon}\tilde{M}_{a_N|x_N}^{(N)}(\Omega) & \text{if } i = \max(\Omega), \end{cases} \tag{3.97}$$

and define the non-maximal overall measurement operators as:

$$\tilde{N}_{a_i|x_i}^{(i)} = \mathbf{0} \oplus \bigoplus_{\Omega \subseteq [N]} \tilde{N}_{a_i|x_i}^{(i)}(\Omega). \tag{3.98}$$

For general subsets Ω and nowhere-maximal strings $(\mathbf{a}_\Omega|\mathbf{x}_\Omega)$,

$$\begin{aligned} \mathrm{Tr}\left(\hat{\bigotimes}_{i\in\Omega} \tilde{N}_{a_i|x_i}^{(i)} \rho_{\varepsilon}^{N,d}\right) &= \varepsilon \sum_{j=1}^{d} \prod_{i=1}^{N} \langle j|\tilde{N}_{a_i|x_i}^{(i)}(\Omega)|j\rangle \\ &= \frac{1}{d} \sum_{\Omega\subseteq\Omega'\subseteq[N]} \sum_{j=1}^{d(\Omega')} \prod_{i=1}^{|\Omega'|} \langle j|\tilde{M}_{a_i|x_i}^{(i)}(\Omega')\,|j\rangle \end{aligned} \tag{3.99}$$

where the same arguments as when obtaining (3.79) are used to deduce the second line, plus the fact that for each subset Ω, for exactly one value of i does $\tilde{N}_{a_i|x_i}^{(i)}(\Omega)$

have a factor of $\frac{1}{d\cdot\varepsilon}$. Equation (3.99) agrees precisely with equation (3.79), thus recovering the non-maximal marginals $P\left(\mathbf{a}_\Omega|\mathbf{x}_\Omega\right)$. Defining the remaining measurement operators $\tilde{N}^{(i)}_{Kx_i|x_i} = \mathbb{I}^{(i)} - \sum_{a_i=1}^{K_{x_i}-1} \tilde{N}^{(i)}_{a_i|x_i}$ is again sufficient to generate the full outcome distribution $P\left(\mathbf{a}|\mathbf{x}\right)$.

3.5 Discussion

Classical theory and quantum theory each admit two distinct quasiprobabilistic extensions, all of which allow for the generation of arbitrary non-signaling correlations. Moreover, the quantum case can be seen as following immediately from the classical version. This suggests two things: firstly, it does not appear to be a special property of quantum theory, or quantum-like formalisms, that allows for non-signaling distributions to be generated in this manner. Secondly, it seems clear that any general probabilistic theory which, like quantum theory and Boxworld, "contains" classical probability theory, can also be modified to generate arbitrary non-signaling distributions in this manner (note that this does not detract from the importance of results such as that of Acín *et al* [1], which is still immensely useful in providing a quantum-like framework for describing non-quantum correlations). Here we use "contain" in the sense that local classical systems can be simulated by some subset of states and measurements on local systems in that theory, and any convex mixtures of product states are allowed in composite systems (in the language of Chapter 2, the composite state space contains the min-tensor product of the local state spaces). It would be interesting to pursue this line of investigation, which concerns the relationship between classical theory and other non-signaling probabilistic theories.

Recall that the joint probability distribution $P_\Lambda(\boldsymbol{\lambda})$ of a locally achievable outcome distribution (3.35) is defined over a set of strings of the form $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_N)$, where each λ_i corresponds to some local variable at system i. This local variable is often taken to codify the local, physical state of affairs at system i. Before Theorem 6 it was remarked that P_Λ can be interpreted as a statistical mixture of

product states, each corresponding to a possible value of $\boldsymbol{\lambda}$. This can be taken a step further: in "epistemic" views of mechanics, such a non-deterministic probability distribution over states may well correspond to an observer's ignorance of an underlying (perhaps deterministic) model, whose values codify the *real* state of affairs. For various reasons (for example, instrumental imprecision) the observer cannot access this real state of affairs, and must instead regard the system as being in a convex mixture of these states. From this perspective, Theorem 7 is particularly surprising: P_Λ is a genuine probability distribution representing one's ignorance about completely local states of affairs which, albeit non-physical due to their negative values, cannot generate any non-locality of their own accord. Despite this, the mixture of those states generates the strongest non-locality which is algebraically possible. This is particularly striking in the case of Section 3.4.2, in which even the classical correlation shared between the parties becomes negligible. When quasiprobabilities are introduced, it is no longer sensible to think of the joint probability distribution on local variables as an ignorance or mixture of allowed local states, but rather as a non-local object in itself.

In the case of Theorem 6, it is the joint probability distribution P_Λ that is allowed to take on negative values, rather than the local outcome functions. Instead of defining a convex combination of product states, P_Λ now defines an "affine" combination of product states. In our proof of Theorem 6, we showed that these product states could in fact be taken to be products of pure states: states which are completely deterministic on the local system. It is a reasonably well-known fact that in Boxworld, the states of a joint system can be written as affine combinations of pure product states [29, 77]; similar results have also been obtained using a measure-theory approach [78, 79]. Theorem 6 is therefore an alternative and constructive proof of this result. On the other hand, to the best of the author's knowledge, no result along the lines of Theorems 7 or 8 exists in the published literature; these are important and intriguing results both because of their potential novel applications in quantum information and foundations, and because of the surprising and counterintuitive aspects highlighted in the previous paragraph.

It is interesting to examine the extent to which the local quasiprobability distributions constructed in this chapter actually violate the positivity condition obeyed by conventional probabilities. Since in some sense we are replacing the "weirdness" of non-locality with the "weirdness" of negative probabilities, one might assume that more strongly non-local distributions lead to more negative quasiprobabilities. This does not seem to be the case in our constructions; for example, in Theorem 6, with an odd number of systems the largest negative value taken on by P_Λ will in fact be $\prod_{i=1}^{N}(1 - M^{(i)})$. This value depends purely on the number of measurement choices at each system; even locally-achievable distributions will result in highly negative quasiprobability distributions being constructed, so long as the number of possible measurements is high.

It would nevertheless be interesting to explore whether other constructions exist which do align negativity with non-locality in some way. From the results of Navascués *et al* [74, 75], in which a convergent hierarchy of conditions are given for characterizing the set of quantum correlations, it seems unlikely that one could determine that the initial outcome distribution $P(\mathbf{a}|\mathbf{x})$ is quantum-achievable simply by looking at the negativity of the local quasiprobability distributions simulating it. However, a converse result may well be possible, in which quantum-achievable distributions can always be represented using negative values no less than a given bound (even if some non-quantum outcome distributions also do not violate this bound).

Although in this chapter we have tended to refrain from discussing the physical meaning of negative probabilities, some considerations along these lines do suggest themselves at this point. Given the accuracy with which local quasiprobabilistic representations may simulate any quantum - or indeed non-signaling - outcome distribution, it seems natural to ask if we *should* be looking for some special physical meaning associated to negative probabilities. Have we veered far off the mark by investing so much research into non-locality, when perhaps negativity is somehow more real? On the other hand, what if these quasiprobabilistic models are no more than a purely mathematical exercise in equation-solving, simply made

much easier by the removal of linear positivity constraints? A clue to the answer to these questions may lie in the variety of approaches described in this chapter. Rather than there being a single, canonical method for deriving local or quantum quasiprobabilistic representations, there appears to be a multitude, of which we have likely only scratched the surface! The sheer redundancy of quasiprobabilistic representations lends credibility to the opinion that although they might be useful for the purpose of calculation (*a la* Wigner), it might be futile to search for anything special they have to tell us about the world.

Chapter 4

Information Causality

All reality is a game. Physics at its most fundamental, the very fabric of our universe, results directly from the interaction of certain fairly simple rules, and chance; the same description may be applied to the best, most elegant and both intellectually and aesthetically satisfying games. By ... resulting from events which, at the sub-atomic level, cannot be fully predicted, the future ... retains the possibility of change, the hope of coming to prevail…. In this, the future is a game; time is one of its rules…. The very first-rank games acknowledge the element of chance, even if they rightly restrict raw luck.

"Player of Games"
Iain M. Banks

4.1 Non-local games

Non-local games provide a convenient and intuitive setting in which to explore which correlations are more useful than others for being exploited in information-based tasks [7,80-83]. In a non-local game, two parties (whom, following convention, we will refer to as Alice and Bob in this chapter) each receive some separate input, and, without communicating, return an output. They win the game if, according to some predicate, their outputs are "correct" with respect to their inputs. This predicate generally refers to whether the *joint* output is correct with respect to the *joint* input, rather than Alice or Bob individually being correct or not. Their goal is to attempt to maximise their average rate of success, given the probabilities with which their inputs are distributed.

Related to the notion of a non-local game is that of a distributed computation. The setup is very similar: Alice and Bob are each given an input, and must produce outputs which are jointly correct with respect to the joint input. The difference is that Alice and Bob *are* allowed to communicate after they know their inputs, but must keep their communication to a minimum. Further restrictions may be introduced, for example the communication might only be one-way from Alice to Bob. Note that there are two distinct ways to optimize a distributed computation: firstly, the probability of success must be maximised for a *fixed* amount of communication (for example, one bit); and secondly, the amount of communication must be minimised whilst achieving at least some fixed probability of success (for example, winning every time).

To put this on a rigorous footing, suppose that Alice's inputs belong to some finite set X_1, and Bob's inputs belong to a finite set X_2. A probability distribution $\pi(x_1, x_2)$ determines the likelihood of Alice and Bob receiving the pair of inputs $x_1 \in X_1$ and $x_2 \in X_2$. Alice and Bob then output a pair of values a_1 and a_2 from the finite sets A_1 and A_2 respectively, and their outputs are judged according to some predicate V_{a_1,a_2,x_1,x_2} taking values in $\{0, 1\}$ (one could conceive of an extension in which V instead assigns a weighting in $[0, 1]$, however the case where answers are either "right" or "wrong" is interesting enough for our purposes). The

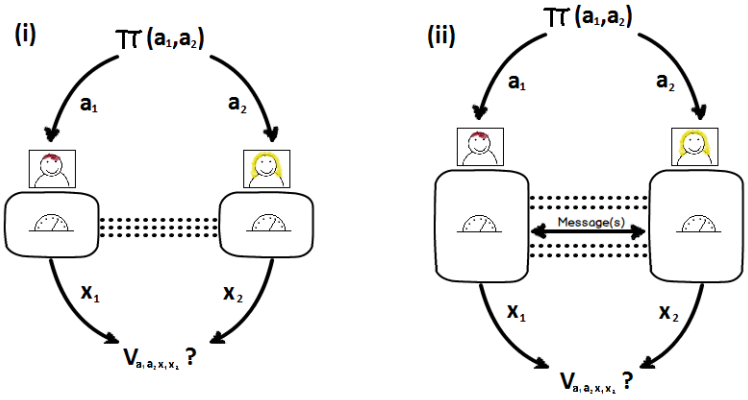

Figure 4.1: (i) Non-local game; (ii) Distributed computation.

functions π and V then define a nonlocal game $G = G(\pi, V)$.

Whatever strategy Alice and Bob employ, the probability of their outputs are given by a conditional probability distribution $P\left(a_1, a_2|x_1, x_2\right)$. If no communication is allowed, then P will be a non-signaling outcome distribution. Their average success probability is then given by the formula:

$$P_G = \sum_{x_1,x_2} \pi(x_1, x_2) \sum_{a_1,a_2} P\left(a_1, a_2|x_1, x_2\right) V_{a_1,a_2,x_1,x_2}. \tag{4.1}$$

If Alice and Bob are restricted to classical strategies, i.e. the set of locally achievable outcome distributions, then P is a convex combination of product states:

$$P\left(a_1, a_2|x_1, x_2\right) = \sum_{\boldsymbol{\lambda}} P^{(1)}\left(a_1|x_1, \lambda_1\right) P^{(2)}\left(a_2|x_2, \lambda_2\right) P_\Lambda(\boldsymbol{\lambda}) \tag{4.2}$$

Substituting this into (4.1) gives:

$$P_G = \sum_{x_1,x_2} \pi(x_1,x_2) \sum_{a_1,a_2} \left(\sum_{\boldsymbol{\lambda}} P^{(1)}\left(a_1|x_1,\lambda_1\right) P^{(2)}\left(a_2|x_2,\lambda_2\right) P_\Lambda(\boldsymbol{\lambda}) \right) V_{a_1,a_2,x_1,x_2}$$
$$= \sum_{\boldsymbol{\lambda}} P_\Lambda(\boldsymbol{\lambda}) \left(\sum_{x_1,x_2} \pi(x_1,x_2) \sum_{a_1,a_2} P^{(1)}\left(a_1|x_1,\lambda_1\right) P^{(2)}\left(a_2|x_2,\lambda_2\right) V_{a_1,a_2,x_1,x_2} \right). \tag{4.3}$$

Hence the probability of success at a classically achievable strategy will be an average of the probabilities of success for some set of product strategies. The *optimal* classical probability of success is therefore attained via some product strategy. This argument can be taken further: for any convex set of non-signaling distributions, the optimal success probability will always be attained via a distribution which cannot be written as a convex combination of other distributions (i.e. one which is an extreme point of the set).

Any local strategy $P^{(1)}\left(a_1|x_1\right)$ employed at Alice's end is in fact a convex combination of deterministic strategies $\tilde{a}_1 : X_1 \to A_1$ which assign a definite output for each input. To see this, let each such strategy $\tilde{a}_1$ be weighted by the following product:

$$q_{\tilde{a}_1} = \prod_{x_1'} P^{(1)}\left(\tilde{a}_1(x_1')|x_1'\right). \tag{4.4}$$

Let Q be the strategy which is the convex combination of deterministic strategies $\hat{a}$ according to the above weighting. Then for any pair (a_1, x_1), we have:

$$\begin{aligned} Q\left(a_1|x_1\right) &= \sum_{\tilde{a}_1} \delta(\tilde{a}_1(x_1), a_1) q_{\tilde{a}_1} \\ &= \sum_{\tilde{a}_1} \delta(\tilde{a}_1(x_1), a_1) \prod_{x_1'} P^{(1)}\left(\tilde{a}_1(x_1')|x_1'\right) \\ &= P^{(1)}\left(a_1|x_1\right) \sum_{\tilde{a}_1 : \tilde{a}_1(x_1)=a_1} \prod_{x_1' \neq x_1} P^{(1)}\left(\tilde{a}_1(x_1')|x_1'\right) \\ &= P^{(1)}\left(a_1|x_1\right) \end{aligned} \tag{4.5}$$

The same is true for any local strategy that Bob employs, hence any product strategy (and any classical strategy) is a convex combination of deterministic product strategies. The optimal probability of success for classical strategies is therefore the maximum probability of success over the (finite) set of deterministic product strategies:

$$P_G(C) = \max_{\tilde{a}_1, \tilde{a}_2} \sum_{x_1, x_2} \pi(x_1, x_2) V_{\tilde{a}_1(x_1), \tilde{a}_2(x_2), x_1, x_2}. \tag{4.6}$$

If Alice and Bob have access to a shared, entangled quantum state, then P takes the form of a quantum-achievable outcome distribution. Alice and Bob perform quantum measurements which depend on their inputs $x_1 \in X_1$, $x_2 \in X_2$, and the outcome of the measurement will determine their individual outputs. Specifically, suppose that they share one part each of a density matrix ρ belonging to a bipartite quantum system $\mathcal{H}_1 \otimes \mathcal{H}_2$. Each of Alice's possible inputs $x_1 \in X_1$ will correspond to a set of positive operators $\{M^{(1)}_{a_1|x_1}\}$ which sum to the identity in $\mathcal{H}_1$: $\sum_{a_1} M^{(1)}_{a_1|x_1} = \mathbb{I}^{(1)}$. Similarly, each each of Bob's inputs $x_2 \in X_2$ corresponds to a set of positive operators $\{M^{(2)}_{a_2|x_2}\}$ satisfying $\sum_{a_2} M^{(2)}_{a_2|x_2} = \mathbb{I}^{(2)}$. The outcome distribution can then be written:

$$P\left(a_1, a_2 | x_1, x_2\right) = \mathrm{Tr}\left(M^{(1)}_{a_1|x_1} \otimes M^{(2)}_{a_2|x_2} \rho\right) \tag{4.7}$$

By purifying the state ρ in larger Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$, the outcome distribution can also be realised by quantum measurements on a pure state $|\psi\rangle$, so that,

$$P\left(a_1, a_2 | x_1, x_2\right) = \langle\psi| M^{(1)}_{a_1|x_1} \otimes M^{(2)}_{a_2|x_2} |\psi\rangle, \tag{4.8}$$

and the probability of winning the game is given by,

$$P_G = \sum_{x_1, x_2} \pi(x_1, x_2) \sum_{a_1, a_2} \langle\psi| M^{(1)}_{a_1|x_1} \otimes M^{(2)}_{a_2|x_2} |\psi\rangle V_{a_1, a_2, x_1, x_2}. \tag{4.9}$$

The optimal quantum strategy $P_G(Q)$ is therefore the supremum of (4.9) over all Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$, and density matrices and POVMs operators defined on

those spaces. Unlike classical strategies, there is no simple reduction to a maximization problem over a finite set of strategies; indeed, it is not generally known whether the optimal quantum strategy can always be achieved, or if instead there exist games where, as the dimensions of $\mathcal{H}_1$ and $\mathcal{H}_2$ increase, the optimal value is approached but is never quite reached [80].

In this section we review some of the literature on non-local games, and discuss in detail a particular game known as the Information Causality game. In the following sections we then study the Information Causality game from two intriguing perspectives. Firstly, in Section 4.2, we analyse the players' probability of success in the Information Causality game and thus derive a bound on a class of non-local games known as Random Access Codes, proving a conjecture given in [3]. The results of this section were obtained in collaboration with Anthony Short and published in [44]. Secondly, in Section 4.3, we reformulate the standard description of Information Causality in terms of entropies, arriving at an elegant derivation of Tsirelson's bound from entropic considerations. The results of Section 4.3, also published in [44] are substantially the author's own work.

4.1.1 The CHSH game revisited

There are several non-local games for which the optimal quantum strategy is known to be achieved in low dimensions [47,80,83]. A good example of this is the *CHSH game* introduced in Section 2.1, in which quantum strategies clearly outperform their classical counterparts. Recall that in the CHSH game the input and output sets are all of size 2, and that for convenience we will take these sets to be $\{0, 1\}$. The input distribution π is uniform over the 4 possible input pairs, and the predicate V takes the following form:

$$V_{a_1,a_2,x_1,x_2} = \delta(a_1 \oplus a_2 = x_1 \cdot x_2), \tag{4.10}$$

where "$\oplus$" and "$\cdot$" denote addition and multiplication modulo 2. Note that $a_1 \oplus a_2 = 0$ if and only if the values of a_1 and a_2 are *correlated*, whereas it equals 1 if and only if they are *anti-correlated*. The expression $x_1 \cdot x_2$ is equal to 1 if and only if both x_1 and x_2 are equal to 1, i.e. for just one of the possible values of the pair of inputs. Hence the predicate can be summarised in words as: "*Alice and Bob's outputs must be correlated for all but one choice of the pair of inputs, otherwise they must be anti-correlated*". Since $\pi(x_1, x_2)$ is always equal to $\frac{1}{4}$, the probability of success for any strategy with outcome distribution P can be written:

$$P_G = \frac{1}{4} \sum_{x_1,x_2,a_1,a_2} P\left(a_1, a_2 | x_1, x_2\right) \delta(a_1 \oplus a_2 = x_1 \cdot x_2). \tag{4.11}$$

As we have argued, the optimal classical success probability for the CHSH game is found by considering the deterministic product strategies. It is easy to check that for every choice of local deterministic strategies $\hat{a}_1$ and $\hat{a}_2$, the equation $\hat{a}_1(x_1) \oplus \hat{a}_2(x_2) = x_1 \cdot x_2$ is satisfied either for exactly 3 out of the four possible input choices (e.g. $\hat{a}_1(x_1) = 0 = \hat{a}_2(x_2)$ fails only when $x_1 = x_2 = 1$), or for exactly 1 out of the four inputs (e.g. $\hat{a}_i(x_i) = x_i$ succeeds only when $x_1 = x_2 = 0$). Therefore $P_G(C) = \frac{3}{4}$.

The optimal quantum probability of success for the CHSH game is $\frac{2+\sqrt{2}}{4}$, commonly known as Tsirelson's bound since his proof in [47]. Observe that the probability of success P_G at the CHSH game is distinct from the CHSH value $\mathcal{C}$ introduced in Section 2.1. However, if one assumes that Alice and Bob are given their inputs uniformly at random, then it is easily checked that they are related in the following manner:

$$P_G = \frac{\mathcal{C} + 4}{8}. \tag{4.12}$$

Therefore from Tsirelson's bound it can be deduced that the optimal CHSH value achievable in quantum theory is $2\sqrt{2}$, as stated in Section 2.1.

As well as providing a concise indicator of how quantum theory deviates from any local, classical theory, the CHSH game provides a stepping stone for exploring non-signaling correlations that cannot be achieved in quantum theory. The *PR-*

box [48] is an example of a non-signaling outcome distribution which performs perfectly at the CHSH game, and thus outperforms quantum theory. Recall that in order to succeed, Alice and Bob's outputs must anti-correlate if and only if both their inputs are equal to 1. The PR-box is defined in such a way that this clearly occurs:

$$P(a_1, a_2|x_1, x_2) = \frac{1}{2}\delta(a_1 \oplus a_2 = x_1 \cdot x_2). \tag{4.13}$$

Note that whatever the input on Alice's side, her outcome will always be uniformly random; the same is true for Bob, and hence this distribution is non-signaling.

The PR-box's violation of Tsirelson's bound demonstrates that the non-signaling condition is not by itself sufficient to single out the set of quantum achievable outcome distributions. This raises several natural questions. Why does Tsirelson's bound take on this slightly peculiar value, and what is it about nature that forbids us from going above it? A related question is: what physical principles, when imposed in *addition* to the non-signaling condition, allow us to recover Tsirelson's bound purely through information-theoretic arguments, rather than through the paradigm of Hilbert spaces and density operators? In the following two sections, we review some attempts to answer these questions. In Section 4.1.2 we discuss the well-known result of W. van Dam, that if nature were to allow the existence of PR-boxes, then all communication complexity tasks would become trivial [84]. In Section 4.1.3 we introduce the Information Causality game, which describes an information-theoretic principle from which one can rule out not just the PR-box, but any non-signaling distribution whose probability of success at the CHSH game is greater than Tsirelson's bound.

4.1.2 Inner product games

In the *Non-Local Inner Product game* [84], Alice is given an n-bit string $\mathbf{x}$ and Bob is given an n-bit string $\mathbf{y}$. Then, after performing local operations but without any communication, Alice must output a bit a and Bob must attempt output a bit b such that $a \oplus b = \mathbf{x} \cdot \mathbf{y} = \oplus_{i=1}^{n}(x_i \cdot y_i)$. Note that if $n = 1$ and $\mathbf{x}$ and $\mathbf{y}$ are

chosen uniformly at random, then the Non-Local Inner Product game reduces to the CHSH game.

Closely related to the Non-Local Inner Product game is the *Distributed Inner Product Computation* [84]. In this task, Alice and Bob are again given bit strings **x** and **y**; however, instead of both producing individual outputs, Alice is permitted to transmit any number of bits to Bob, after which Bob attempts to output the value $\mathbf{x} \cdot \mathbf{y}$. As well as maximising the probability of success, the aim is to minimise the number of bits which are transmitted from Alice to Bob. Note that a strategy for the Non-Local Inner Product Game can easily be converted into a strategy for the Distributed Inner Product Computation with one bit of communication and an identical probability of success: instead of outputting her bit a, Alice transmits it to Bob, who then calculates and outputs the sum (modulo 2) of a and his output b of the Non-Local Inner Product game. This conversion of strategies will turn out to be very important in Section 4.2.

Recall that if Alice and Bob share a single g-bit in the PR-box state, then they can succeed with probability 1 at the CHSH game. Following [84], it can easily be seen that with n shared PR-box states, they can also succeed with probability 1 at the Non-Local Inner Product game (and therefore perform all Distributed Inner Product Computations perfectly with the transmission of a single bit). For each value of i with $1 \leq i \leq n$, suppose that Alice and Bob input the measurement choices corresponding to x_i and y_i respectively into their half of the same shared PR-box. The output of that PR-box will be two bits - a_i for Alice and b_i for Bob - satisfying $a_i \oplus b_i = x_i \cdot y_i$. After doing this for all n components of **x** and **y**, Alice's

overall output bit is $a = \oplus_{i=1}^{n} a_i$, and Bob's overall output bit is $b = \oplus_{i=1}^{n} b_i$. Then,

$$\begin{aligned} a \oplus b &= \left(\bigoplus_{i=1}^{n} a_i \right) \oplus \left(\bigoplus_{j=1}^{n} b_j \right) \\ &= \bigoplus_{i=1}^{n} (a_i \oplus b_i) \\ &= \bigoplus_{i=1}^{n} (x_i \cdot y_i) \\ &= \mathbf{x} \cdot \mathbf{y}. \end{aligned} \tag{4.14}$$

As van Dam points out [84], not much work is required to extend this idea, so that Alice and Bob can compute more sophisticated functions of their input bit strings. It is a well-known fact that any Boolean function $f(\mathbf{x}, \mathbf{y})$ may be written as a multi-variable polynomial in the components of $\mathbf{x}$ and $\mathbf{y}$, and by collecting terms in $\mathbf{y}$ it is possible to decompose f as:

$$f(\mathbf{x}, \mathbf{y}) = \bigoplus_{i=1}^{2^n} \left(P_i(\mathbf{x}) \cdot Q_i(\mathbf{y}) \right). \tag{4.15}$$

where each P_i is a multi-variable polynomial and each Q_i is a monomial, i.e. of the form $y_{i_1} \cdot \ldots \cdot y_{i_m}$ for some subset $\{i_1, \ldots, i_m\} \subset [n]$ (note that there are 2^n such subsets). For example, the function $f(\mathbf{x}, \mathbf{y}) = y_1$ AND $(x_1$ OR $y_2)$ can be written as:

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) &= y_1 \cdot (x_1 \oplus y_2 \oplus x_1 \cdot y_2) \\ &= [x_1 \cdot y_1] \oplus [(x_1 \oplus 1) \cdot y_1 \cdot y_2] \end{aligned} \tag{4.16}$$

The distributed computation of a general function of two n-bit strings thus reduces to the distributed computation of the inner product of 2^n binary variables. By using the bit $P_i(\mathbf{x})$ as the input for Alice's half of each PR-box, and $Q_i(\mathbf{y})$ for Bob's half, it is thus possible to compute any Boolean function in this way, with at

most 2^n uses of the PR-box and one bit of communication from Alice to Bob.

If Nature allowed the existence of PR-box correlations, then the above protocol would permit the distributed computation of arbitrary functions using trivial communication complexity. This does not seem to immediately conflict with physical intuition - after all, it's not as if the speed of computation is independent of the spatial separation between Alice and Bob. However, as van Dam points out, it renders the very notion of communication complexity defunct, since all functions could be evaluated in this trivial fashion. This consequence of strongly nonlocal correlations is "*squarely against the world-view and experience of probably all researchers in the field of complexity theory*" [84]; the existence of non-trivial communication complexity can be regarded as an intuitive physical principle which rules out the existence of PR-box correlations.

Ruling out just the existence of PR-boxes, however, does not even come close to ruling out the existence of all non-signaling correlations which are not quantum-achievable. It would be nice if, for example, a similar protocol could be designed which rules out all those correlations that violate the CHSH inequality beyond Tsirelson's bound. Though this has been done for correlations whose maximal CHSH violation lies in a defined interval above Tsireloson's bound [72], it is an open question whether Tsirelson's bound provides a critical threshold value for the non-triviality of communication complexity. Since the publication of van Dam's protocol, various other natural principles have been shown to exactly recover Tsirelson's bound [50-52], or at least come close to it [72]; in the next section we describe the Information Causality principle, which is one of the better known examples of this.

4.1.3 The information causality game

Like van Dam's protocol, Information Causality is formulated in terms of a game. The *Information Causality* game [51] again involves two parties; a bit string **a** of length n is chosen uniformly at random and given to Alice, whilst Bob is given a number k, $1 \leq k \leq n$. Alice may then transmit an m-bit message **x** to Bob,

after which Bob makes a guess at a_k, the kth bit of Alice's original message. The parties may decide on a shared strategy but play separately, and no communication other than the message **x** is permitted. Note that the Information Causality game can be formulated as an inner product game: if Bob's number k is represented as a bit string **y** with a 1 in the kth place and zeros elsewhere, then Bob's is required to make a guess at the output of the inner product function $f(\mathbf{a}, \mathbf{y}) = \mathbf{a} \cdot \mathbf{y}$.

However, there is a crucial difference between the Information Causality game and other non-local games; rather than attempting to maximise the average probability that Bob's guess is correct, Alice and Bob must try to maximise a slightly unusual figure of merit. For any given strategy that Alice and Bob agree upon, suppose that $\beta(k)$ is the classical random variable encoding Bob's guess at the value of a_k (i.e. assuming he was given k). Then their degree of success is measured by the function:

$$I = \sum_{k=1}^{N} I_c(a_k : \beta(k)), \tag{4.17}$$

where $I_c(A : B)$ denotes the classical mutual information of the joint probability distribution A, B. Classical intuition suggests that since **a** was chosen at random and independently of Bob's system, the total information Bob has about **a** must be no greater than whatever can be extracted from m classical bits; accordingly the Information Causality principle demands that for any strategy that Alice and Bob employ,

$$I = \sum_{k=1}^{N} I_c(a_k : \beta(k)) \leq m. \tag{4.18}$$

The formula (4.18) can be seen as a generalisation of the non-signaling condition as applied to Bob's system in the game, with the non-signaling condition being the special case $m = 0$. To see this, let $P(\emptyset, \beta(k)|\mathbf{a}, k)$ be the outcome distribution associated with the game (where the outcome at Alice's system plays no role and so is denoted by $\emptyset$), and recall that the non-signaling condition is equivalent to the existence of a well-defined marginal distribution $P(\beta(k)|k)$ which is independent of **a**; since $\beta(k)$ is in fact Bob's ouput *given* the number k, $P(\beta(k)|k)$

is simply the distribution of the variable $\beta(k)$: this distribution being independent of **a** is in turn equivalent to (4.18) holding with $m = 0$.

Two further points are worth mentioning, concerning (4.18). Firstly, the inequality can be very simply saturated with a classical strategy in which Alice simply sends to Bob the first m bits of **a**, so that $I_c(a_k : \beta(k)) = 1$ if $1 \leq k \leq m$, and 0 otherwise. Secondly, in (4.18) we have avoided defining measures of information for states of systems which may not be classical systems. Although Alice and Bob's strategy may involve bipartite states from general physical theories, Bob's guess must ultimately take the form of a classical bit, and the figure of merit I is a sum of mutual informations between purely classical random variables.

However, classical theory is not the only general physical theory which possesses a measure of mutual information; namely, in quantum theory the quantum mutual information $I_q(A : B)$ between systems A and B is defined as $H_q(A) + H_q(B) - H_q(A, B)$, where H_q is the von Neumann entropy. I_q is positive, symmetric, and generally behaves in accordance with our intuition about information shared between two states. Equation (4.18) holds in any general physical theory in which, for any two systems X and Y, there exists some measure of mutual information $\tilde{I}(X : Y)$ which satisfies four basic properties [51]:

(i) Consistency: Whenever X and Y are both classical systems, $\tilde{I}$ reduces to the classical mutual information, $\tilde{I}(X : Y) = I_c(X : Y)$

(ii) Data Processing: For any allowed transformation $\tau : Y \to Y'$, $\tilde{I}(X : Y) \geq \tilde{I}(X : Y')$

(iii) Chain Rule: For all tripartite systems X, Y, Z, $\tilde{I}(X, Y : Z) - \tilde{I}(X : Z) = \tilde{I}(Y : X, Z) - \tilde{I}(X : Y)$

(iv) Symmetry: For all bipartite systems X, Y, $\tilde{I}(X, Y) = \tilde{I}(Y, X)$

It is well known that all four properties hold both for the classical mutual information I_c and the quantum mutual information I_q in their respective general

physical theories [53]; hence, if Alice and Bob are restricted to making measurements on shared classical or quantum bipartite states, they will never be able to violate the Information Causality bound (4.18). Note however, that whilst properties (i), (ii) and (iv) seem very reasonable, property (iii) appears to be without immediate physical significance. This will become important later in Section 4.3, where it is shown that the Information Causality principle can be framed so that requirement (iii) essentially disappears.

Given the existence of a mutual information which satisfies properties (i)-(iv), we will demonstrate that (4.18) holds using the techniques presented in [51]. Firstly, observe that any two systems X and Y can be individually mapped to independent classical systems (for example the map which simply traces out the system, represented by $\tau : X \to \emptyset$). By Symmetry and Data Processing, one obtains $\tilde{I}(X : Y) \geq \tilde{I}(\emptyset : \emptyset)$. It then follows from Consistency that $\tilde{I}(X : Y) \geq 0$, hence $\tilde{I}$ is always non-negative.

Let $\mathbf{X}$ be the classical system comprising the m bits which Alice transmits to Bob; the equation $m = \tilde{I}(\mathbf{X} : \mathbf{X})$ follows immediately from Consistency. Consider the tripartite system $\mathbf{X}, \mathbf{A}, B$, where $\mathbf{A}$ is the classical system containing Alice's input $\mathbf{a}$ and B is Bob's system before Alice transmits her message $\mathbf{x}$. The state of system $\mathbf{X}$ is described by a set of probabilities $p_{\mathbf{x}}$ for each bitstring $\mathbf{x}$, and corresponds to a particular reduced state $\psi_{\mathbf{x}}$ of system $\mathbf{a}, B$, such that the joint state of $\mathbf{X}, \mathbf{A}, B$ is the convex combination of the product states $(\mathbf{x}, \psi_{\mathbf{x}})$ according to the distribution $p_{\mathbf{x}}$. The transformation $\tau : \mathbf{x} \to \psi_{\mathbf{x}}$ therefore maps system $\mathbf{X}$ to system $\mathbf{A}, B$, so that by Data Processing,

$$m = \tilde{I}(\mathbf{X} : \mathbf{X}) \geq \tilde{I}(\mathbf{X} : \mathbf{A}, B). \tag{4.19}$$

Applying the Chain Rule to the RHS of (4.19) with $X = B, Y = \mathbf{X}$ and $Z = \mathbf{A}$ gives:

$$\tilde{I}(\mathbf{X} : \mathbf{A}, B) = \tilde{I}(\mathbf{X}, B : \mathbf{A}) + \tilde{I}(\mathbf{X} : B) - \tilde{I}(\mathbf{A} : B) \tag{4.20}$$

Since $\mathbf{a}$ is chosen uniformly at random, it is independent of system B. There-

fore the state of $\mathbf{A}, B$ can be constructed from $\mathbf{A}, \emptyset$ by simply mapping $\emptyset$ to the reduced state at system B. By Data Processing and Consistency, $\tilde{I}(\mathbf{A}, B) \leq I_c(\mathbf{A}, \emptyset) = 0$. Since $\tilde{I}$ is non-negative, $\tilde{I}(\mathbf{A} : B) = 0$. Also, $\tilde{I}(\mathbf{X} : B) \geq 0$, so that:

$$\tilde{I}(\mathbf{X} : \mathbf{A}, B) \geq \tilde{I}(\mathbf{X}, B : \mathbf{A}) = \tilde{I}(\mathbf{A} : \mathbf{X}, B) \tag{4.21}$$

where the right-hand equality follows from Symmetry. Writing the system $\mathbf{A}$ as $\mathbf{A}_1, \mathbf{A}_{\{2,\ldots,n\}}$, and applying the Chain rule with $X = \mathbf{A}_1$, $Y = \mathbf{A}_{\{2,\ldots,n\}}$ and $Z = \mathbf{X}, B$ to the RHS of (4.21) gives:

$$\begin{aligned}\tilde{I}(\mathbf{A}_1, \mathbf{A}_{\{2,\ldots,n\}} : \mathbf{X}, B) =& \tilde{I}(\mathbf{A}_{\{2,\ldots,n\}} : \mathbf{A}_1, \mathbf{X}, B) \\ &+ \tilde{I}(\mathbf{A}_1 : \mathbf{X}, B) - \tilde{I}(\mathbf{A}_1 : \mathbf{A}_{\{2,\ldots,n\}}).\end{aligned} \tag{4.22}$$

Since $\mathbf{A}_1$ and $\mathbf{A}_{\{2,\ldots,n\}}$ are independent classical systems, the rightmost term is equal to 0. Applying the trace-out map to system $\mathbf{A}_1$ in the tripartite system $\mathbf{A}_1, \mathbf{X}, B$ then gives:

$$\tilde{I}(\mathbf{A}_1, \mathbf{A}_{\{2,\ldots,n\}} : \mathbf{X}, B) \geq \tilde{I}(\mathbf{A}_{\{2,\ldots,n\}} : \mathbf{X}, B) + \tilde{I}(\mathbf{A}_1 : \mathbf{X}, B). \tag{4.23}$$

This inequality can be iterated $n - 1$ times, writing $\mathbf{A}_{\{2,\ldots,n\}}$ as $\mathbf{A}_2, \mathbf{A}_{\{3,\ldots,n\}}$ etc... , resulting in the inequality:

$$\tilde{I}(\mathbf{A} : \mathbf{X}, B) \geq \sum_{k=1}^{n} \tilde{I}(\mathbf{A}_k : \mathbf{X}, B). \tag{4.24}$$

Bob's output $\beta(k)$ is derived ultimately from $\mathbf{X}$ and B, hence there must exist some map from the state of $\mathbf{X}, B$ to $\beta(k)$. Hence by Data Processing and Consistency,

$$\sum_{k=1}^{n} \tilde{I}(\mathbf{A}_k : \mathbf{X}, B) \geq \sum_{k=1}^{n} I_c(\mathbf{A}_k : \beta(k)) = I. \tag{4.25}$$

Putting together the equations (4.19), (4.21), (4.24) and (4.25) gives the desired result $m \geq I$.

So far, we have demonstrated that the Information Causality principle holds in any general physical theory which, like classical or quantum theory, admits a sensible measure of mutual information. We now present the derivation of the fact that, for general physical theories allowing violations of Tsirelson's bound, Information Causality does *not* hold (again, this is based on the derivation presented in [51]). We first describe the protocol in the case where Alice and Bob have access to perfect PR-boxes, and then move onto the case of noisy PR-boxes, which give correct outputs only with a specified probability.

Recall that n is the number of bits given to Alice, and suppose that $n = 2^r$ for some integer r. We will demonstrate that if Alice and Bob have access to an initial supply of many shared PR-boxes, it is possible to perfectly evaluate the function $f(\mathbf{a}, k) = a_k$ using only r chosen outputs of the PR-boxes and one bit of communication. This is easy to demonstrate in the case $r = 1$: Alice inputs $a_1 \oplus a_2$ and obtains output c, whilst Bob inputs $(k-1)$ and obtains output b. By definition of the PR-box, $b \oplus c = (k-1) \cdot (a_1 \oplus a_2)$, hence $a_1 \oplus b \oplus c = a_k$. Alice needs only send Bob the value $x = a_1 \oplus c$, which he can then sum with his output b to obtain his desired bit a_k.

For $r > 1$, write $\mathbf{a} = \mathbf{a}^{(0)}\mathbf{a}^{(1)}$ as a concatenation of two bit-strings of length $n/2$, and let $(k_1, \ldots, k_r)$ be a bit-string representation of $(k-1)$, so that the value of the rth bit k_r indicates whether a_k belongs to either $\mathbf{a}^{(0)}$ or $\mathbf{a}^{(1)}$. Assume that the protocol has been constructed for the case $n = 2^{r-1}$, and suppose that Alice and Bob perform the protocol for both of Alice's strings $\mathbf{a}^{(0)}$ and $\mathbf{a}^{(1)}$ - with Bob using the input represented by $(k_1, \ldots, k_{r-1})$ in both cases - but that Alice does not transmit any messages to Bob. Suppose that Alice would normally communicate the single bit $x^{(0)}$ to Bob had they been performing the protocol only for the string $\mathbf{a}^{(0)}$, and similarly the bit $x^{(1)}$ had they been performing the protocol only for $\mathbf{a}^{(1)}$. Note that Bob now wishes to know the value of exactly one of $x^{(0)}$ or $x^{(1)}$, depending on whether k_r is 0 or 1. This reduces the task to the $r = 1$ case, hence by the previous paragraph, using *one* extra PR-box Bob may obtain the value of bit $x^{(k_r)}$, and complete the protocol for the $(n/2)$-bit string $\mathbf{a}^{(k_r)}$. Figure 4.2 depicts this

protocol in the case where $r = 2$ and $k = 3$; note that whilst the general protocol requires three PR-box states, Bob only makes use of two of the outputs.

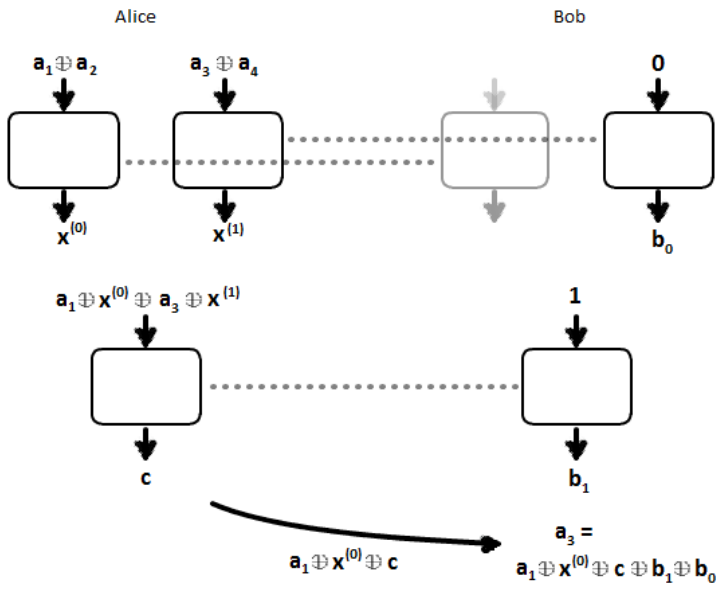

Figure 4.2: How to violate the Information Causality protocol using perfect PR-box states ($r = 2, k = 3$). Adapted from [51].

Now suppose that Alice and Bob use a noisy PR-box, which on inputs x, y outputs bits a and b, whose individual values are uniformly random, such that with probability p, $a \oplus b = x \cdot y$ and with probability $1 - p$, $a \oplus b = (x \cdot y) \oplus 1$ (in fact, any non-signaling outcome distribution can be brought to this form by means of local operations, without affecting its overall CHSH value [85]). Tsirelson's bound is violated by such a box if and only if $p > \frac{2+\sqrt{2}}{4}$ [47]. In the protocol with perfect PR-boxes, Alice and Bob compute their final answer by a modulo 2 summation of the outputs of r PR-boxes. Therefore in the noisy case, the correct answer is obtained not just when all the PR-boxes function correctly, but if and only if an even number of them output the wrong answer. Thus the probability of

Bob having the correct answer at the end of the protocol is given by:

$$\sum_{j=0}^{\lfloor \frac{r}{2} \rfloor} \binom{r}{2j} p^{r-2j}(1-p)^{2j} = \frac{1}{2}[1+(2p-1)^r] = \frac{1+E^r}{2}, \tag{4.26}$$

where $E = p - (1-p)$ is the *bias* of the PR-box. The figure of merit I can be manipulated into a form into which we can subsitute (4.26). Expanding it in terms of the classical entropy,

$$\begin{aligned} I &= \sum_{k=1}^{n} I_c(a_k : \beta(k)) \\ &= \sum_{k=1}^{n} H(a_k) - H_c(a_k|\beta(k)). \end{aligned} \tag{4.27}$$

Since **a** is uniformly distributed, $H_c(a_k) = 1$ for all k. If one has access to $\beta(k)$, then a_k has the same distribution as $a_k \oplus \beta(k)$, hence $H_c(a_k|\beta(k)) = H_c(a_k \oplus \beta(k)|\beta(k))$, so:

$$\begin{aligned} I &\geq \sum_{k=1}^{n} 1 - H_c(a_k \oplus \beta(k)) \\ &= \sum_{k=1}^{n} 1 - h(\text{Prob}(a_k = \beta(k))), \end{aligned} \tag{4.28}$$

$$\tag{4.29}$$

where $h(p)$ is the binary entropy of the distribution $\{p, 1-p\}$. Now substituting

the success probability (4.26) and using the inequality $1 - h(\frac{1+y}{2}) \geq \frac{y^2}{2\ln 2}$:

$$I \geq \sum_{k=1}^{n} 1 - h\left(\frac{1+E^r}{2}\right) \tag{4.30}$$

$$\geq \frac{1}{2\ln 2}\sum_{k=1}^{n} E^{2r} \tag{4.31}$$

$$= \frac{(2E^2)^r}{2\ln 2} \tag{4.32}$$

In order for this last value not to become arbitrarily large with r, and exceed 1 (the value of m in this protocol), we require that $E^2 < \frac{1}{2}$, i.e. $p < \frac{2+\sqrt{2}}{4}$. Hence, Tsirelson's bound gives the precise threshold beyond which the violation of Information Causality is possible.

4.2 Probability of success in the information causality game

As mentioned at the beginning of the previous Section, the Information Causality game differs from other non-local games in that one attempts to maximise a strange function I on the inputs and outputs, rather than the probability of success. On the other hand, Tsirelson's bound is commonly formulated as an upper limit on the probability of success at the CHSH game. This contrast raises some natural questions about Information Causality, and how one uses it to derive Tsirelson's bound. If quantum theory cannot improve on the classical bound on I, does this also imply that quantum theory also cannot improve on the optimal probability of success? If not, how are probability of success and I (defined as in (4.18)) related?

In fact, it can easily be seen that the optimal quantum probability of success in the Information Causality game is sometimes greater than the optimal classical probability of success. For example, in a simple version of the game in which $n = 2$ and $m = 1$, the optimal classical probability of success is $\frac{3}{4}$ (e.g. when

Alice sends Bob $x = a_1$ and he guesses $\beta(k) = x$ regardless of k, they always win when $k = 1$ and win half the time when $k = 2$). However, by exploiting quantum measurements which saturate Tsirelson's bound, Alice and Bob can achieve a success probability of $\frac{2+\sqrt{2}}{4}$. To do this, Alice and Bob first generate bits a' and b' satisfying $a' \oplus b' = (a_1 \oplus a_2) \cdot (k-1)$ with probability $\frac{2+\sqrt{2}}{4}$. Then Alice sends Bob $x = a' \oplus a_1$ and Bob outputs $\beta(k) = b' \oplus x$.

It is also possible to obtain very different values of I for strategies with the same probabilities of success. As above, Alice can send Bob her first bit to obtain $I = 1$ and probability of success $\frac{3}{4}$; alternatively, Alice and Bob can uniformly "mix" this strategy with one where Alice sends Bob her second bit and he outputs that, so that the overall probability of success is the same but they win with probability $\frac{3}{4}$ if either $k = 1$ or $k = 2$. For this strategy,

$$\begin{aligned} I &= I_c(a_1 : \beta(1)) + I_c(a_2 : \beta(2)) \\ &= 2\left(H_c(\beta(1)) - H_c(\beta(1)|a_1)\right) \\ &= 2\left(1 - h\left(\frac{3}{4}\right)\right) \\ &\approx 0.38. \end{aligned} \tag{4.33}$$

The figure I seems to take heavily into consideration whether the likelihood of success is "shifted" into a few of Alice's components, or is evenly distributed across them. Consider mixing the first-bit strategy with a small amount of noise: since I is a continuous function, the value of I will be greater than 0.38. However, the probability of success will now be slightly weighted toward a half, i.e. strictly smaller. Thus, I and the probability of success are not even related monotonically.

4.2.1 Random Access Codes

With probability of success as the figure of merit, the Information Causality game becomes very similar to the task of Random Access Coding [71]. In an (n, m, p) Random Access Code (RAC), Alice has a random bit string **a** of length n, which

she codes into a bit string of length $m < n$ which is transmitted to Bob. Bob must then be able to recreate any one of Alice's bits a_k with probability at least p. Much of the research into Random Access Coding so far concerns the case where $m = 1$ and where the bit string **a** is uniformly chosen, and attempts to find the optimal success probability p as a function of n.

When Alice and Bob are restricted to classical strategies, the optimal success probability in the case $m = 1$ is known; it is attained by using a "majority-vote" strategy, in which Alice simply sends Bob the bit that most frequently occurs in her string [71]. This gives an analytical success probability which yields a simple approximation for large n using Stirling's formula:

$$P^C_{success} = \frac{1}{2}\left(1 + \frac{1}{2^{n-1}}\binom{n-1}{\lfloor \frac{n-1}{2} \rfloor}\right) \approx \frac{1}{2}\left(1 + \sqrt{\frac{2}{\pi n}}\right). \tag{4.34}$$

If Alice and Bob are given access to measurements on quantum states, there are two further classes of RAC that can be considered. In an Entanglement-Assisted RAC (EARAC) [3], Alice and Bob may have an initial supply of entangled states upon which to perform measurements. In a Quantum RAC (QRAC), Alice and Bob do not share any prior entanglement, however instead of transmitting m classical bits to Bob, Alice transmits m qubits to Bob. Both scenarios have been investigated in the $m = 1$ case; for both EARACs and QRACs, the following formula has been shown to be an upper bound on the probability of success [3, 71].

$$P^Q_{success} \leq \frac{1}{2}\left(1 + \sqrt{\frac{1}{n}}\right). \tag{4.35}$$

Pawloski *et al* construct an EARAC protocol in [3] which attains the bound (4.35) exactly when n is of the form $2^j 3^k$, for integers j and k. We show in Section 4.2.2 that this bound is achievable for all n, using strategies converted from a related inner product game. For QRACs it is still not known for which n (4.35) is achievable, however the optimal success probability is known to be lower-bounded

by the following formula [71]:

$$P^{Q}_{success} \geq \frac{1}{2}\left(1 + \sqrt{\frac{1}{6\pi n}}\right). \tag{4.36}$$

4.2.2 Optimal success probability of EARACs

Since the Information Causality game is not itself judged in terms of probability of success, it is interesting to explore in more depth how it is used to derive a bound on the probability of success at winning the CHSH game. With this in mind, a crucial step in the argument appears to be the relation of entropy to PR-box probabilities, in the approximated bound

$$I \geq \sum_{k=1}^{N} 1 - H(P_k) \geq \frac{1}{2\ln 2}\sum_{k=1}^{N} E_k^2, \tag{4.37}$$

where E_k is the success bias of the game, conditioned on Bob being given the number k:

$$\begin{aligned} E_k &= \text{Prob}(a_k = \beta(k)) - (1 - \text{Prob}(a_k = \beta(k))) \\ &= 2\text{Prob}(a_k = \beta(k)) - 1. \end{aligned} \tag{4.38}$$

(Note that E_k is not the same as the bias E of a noisy PR-box state, as defined in Section 4.1.3.) The protocol designed in Section 4.1.3 can also be seen as a derivation of a quadratic bound for quantum strategies in the Information Causality game:

$$\sum_{k=1}^{N} E_k^2 \leq 2\ln 2. \tag{4.39}$$

It turns out to be very useful to look at a more direct way of deriving a similar bound via the rules of quantum theory, rather than using the properties of mutual information. Recall the Inner Product game defined in Section 4.1.2, in particular the fact that any strategy can be converted to a strategy for the Distributed Inner

Product Computation with $m = 1$. Note also that the setup for the Information Causality game is in fact an example of a Distributed Inner Product Computation. Bob attempts to determine the value of $\mathbf{a} \cdot \mathbf{y}$, with the following additional promise on the distribution of Bob's string $\mathbf{y}$: with uniform probability, exactly one of the bits (i.e. the kth bit) is 1, whilst the rest are non-zero.

In this Section we derive a quantum upper bound on the Inner Product game which is very similar in form to the quadratic bias bound (4.39), and show that there exist strategies that achieve this bound. By the above reasoning, such strategies can be converted to strategies for Information Causality: we then show that one of these converted strategies attains the optimal probability of success for EARACs, given by (4.35), and hence gives the optimal probability of success for Information Causality with $m = 1$. The following Lemma, proved in [86], is crucial in deriving our results about the Inner Product Game.

Lemma 4. *For any sub-normalized vectors $u_1, \ldots, u_s$ and $v_1, \ldots, v_t$ in a real Euclidean space of dimension* $\min(s, t)$*, there exists a density matrix ρ on a finite dimensional Hilbert space $H = H_1 \otimes H_2$, with ± 1-valued Hermitian operators $A_1, \ldots, A_s$ on H_1 and $B_1, \ldots, B_t$ on H_2 such that $Tr((A_i \otimes B_j)\rho) = \langle u_i, v_j \rangle$ for all $1 \leq i \leq s$, $1 \leq j \leq t$.* □

Theorem 9. *Suppose that Alice and Bob play the Non-local Inner Product Game, where Alice's input* $\mathbf{x}$ *is chosen uniformly from bit strings of length n, and Bob's input* $\mathbf{y}$ *is chosen according to any random distribution from bit strings of length n. Let $E_{\mathbf{y}}$ be the bias of the game, conditioned on the value of* $\mathbf{y}$*. Then the set of biases $\{E_{\mathbf{y}}\}$ is achievable if, and only if, $\sum_{\mathbf{y}} E_{\mathbf{y}}^2 \leq 1$.*

Proof. We first prove the "only if" direction. Since Alice and Bob do not communicate in the game, the outcome distribution $P(a, b|\mathbf{x}, \mathbf{y})$ of their outputs given their inputs must result from a local POVM measurement on a shared density matrix ρ. That is to say, for each value of $\mathbf{x}$ Bob receives, there is a pair of positive operators $\{M_{0|\mathbf{x}}^{(1)}, M_{1|\mathbf{x}}^{(1)}\}$ which sum to the identity $\mathbb{I}^{(1)}$, and for each value of $\mathbf{y}$ a

similar pair $\{M^{(2)}_{0|\mathbf{y}}, M^{(2)}_{1|\mathbf{y}}\}$ for Alice, such that:

$$P\left(a,b|\mathbf{x},\mathbf{y}\right) = \mathrm{Tr}\left(M^{(1)}_{a|\mathbf{x}} \otimes M^{(2)}_{b|\mathbf{y}}\ \rho\right) \tag{4.40}$$

The POVM measurement (4.40) can be realised as a standard (projective) quantum measurement on a density matrix $\tilde{\rho}$ belonging to a Hilbert space of larger dimension; moreover, the state $\tilde{\rho}$ can be purified into a state $|\psi\rangle$ of even larger dimension. Thus we can assume that Alice and Bob begin with the entangled pure state $|\psi\rangle$, and their outputs are obtained by measurement of Hermitian operators $\hat{a}_{\mathbf{x}}$ and $\hat{b}_{\mathbf{y}}$ respectively with eigenvalues in $\{0,1\}$.

For this set of measurements, let $P_{\mathbf{xy}}$ denote the probability that $a \oplus b = \mathbf{x}\cdot\mathbf{y}$ given that Alice and Bob are given the particular strings $\mathbf{x}$ and $\mathbf{y}$. Suppose that a $+1$ value is associated to the game if $a \oplus b = \mathbf{x}\cdot\mathbf{y}$ (i.e. they succeed), and a -1 value is associated to it otherwise. Then the bias $E_{\mathbf{xy}} = P_{\mathbf{xy}} - (1 - P_{\mathbf{xy}})$ can be seen as the expected value of the game, conditioned on the inputs $\mathbf{x}$ and $\mathbf{y}$. In terms of the state $|\psi\rangle$ and operators $\hat{a}_{\mathbf{x}}$ and $\hat{b}_{\mathbf{y}}$, the bias has a simple form as the expected value of the operator $(-1)^{\hat{a}_{\mathbf{x}}+\hat{b}_{\mathbf{y}}+\mathbf{x}\cdot\mathbf{y}}$.

$$E_{\mathbf{xy}} = \langle\psi|(-1)^{\hat{a}_{\mathbf{x}}+\hat{b}_{\mathbf{y}}+\mathbf{x}\cdot\mathbf{y}}|\psi\rangle. \tag{4.41}$$

The bias of the game conditioned only on Bob's input $\mathbf{y}$ is given by the average over Alice's inputs. Since Alice's bit string is chosen uniformly at random, this is $E_{\mathbf{y}} = \frac{1}{2^n}\sum_{\mathbf{x}} E_{\mathbf{xy}}$. To derive the quadratic bound on Bob's biases, define the normalized states:

$$|A\rangle = \frac{1}{\sqrt{2^n}}\sum_{\mathbf{x}}(-1)^{\hat{a}_{\mathbf{x}}}|\psi\rangle \otimes |\mathbf{x}\rangle \tag{4.42}$$

$$|B_{\mathbf{y}}\rangle = \frac{1}{\sqrt{2^n}}\sum_{\mathbf{x}}(-1)^{\hat{b}_{\mathbf{y}}+\mathbf{x}\cdot\mathbf{y}}|\psi\rangle \otimes |\mathbf{x}\rangle \tag{4.43}$$

so that $\{|B_{\mathbf{y}}\rangle\}$ forms an orthonormal set satisfying $\langle B_{\mathbf{y}}|B_{\mathbf{y}'}\rangle = \delta_{\mathbf{y}\mathbf{y}'}$. It follows that

$$\begin{aligned}
\langle A| \left(\sum_{\mathbf{y}} |B_{\mathbf{y}}\rangle\langle B_{\mathbf{y}}| \right) |A\rangle &= \sum_{\mathbf{y}} \langle A|B_{\mathbf{y}}\rangle^2 \\
&= \sum_{\mathbf{y}} \left(\frac{1}{2^n} \sum_{\mathbf{x}} \sum_{\mathbf{x}'} \langle x|x'\rangle \langle \psi|(-1)^{\hat{a}_{\mathbf{x}}} (-1)^{\hat{b}_{\mathbf{y}}+\mathbf{x}'\cdot\mathbf{y}} |\psi\rangle \right)^2 \\
&= \sum_{\mathbf{y}} \left(\frac{1}{2^n} \sum_{\mathbf{x}} \langle \psi|(-1)^{\hat{a}_{\mathbf{x}}+\hat{b}_{\mathbf{y}}+\mathbf{x}\cdot\mathbf{y}} |\psi\rangle \right)^2 \\
&= \sum_{\mathbf{y}} E_{\mathbf{y}}^2. \qquad (4.44)
\end{aligned}$$

Since $\sum_{\mathbf{y}} |B_{\mathbf{y}}\rangle\langle B_{\mathbf{y}}|$ is a projector, and $|A\rangle$ is normalised,

$$\sum_{\mathbf{y}} E_{\mathbf{y}}^2 = \langle A| \left(\sum_{\mathbf{y}} |B_{\mathbf{y}}\rangle\langle B_{\mathbf{y}}| \right) |A\rangle \leq 1. \qquad (4.45)$$

We now prove the "if" direction: suppose that the real numbers $B_{\mathbf{y}}$ satisfy $\sum_{\mathbf{y}} B_{\mathbf{y}}^2 \leq 1$, and let the vectors $e_{\mathbf{y}}$ be the standard basis vectors of some real vector space. Define the sub-normalised vectors $u_{\mathbf{x}} = \sum_{\mathbf{y}} (-1)^{\mathbf{x}\cdot\mathbf{y}} B_{\mathbf{y}} e_{\mathbf{y}}$ and $v_{\mathbf{y}} = e_{\mathbf{y}}$. By Lemma 4 (after purifying the state ρ), there exists a state $|\psi\rangle$ and $0,1$-valued operators $\hat{a}_{\mathbf{x}}$ and $\hat{b}_{\mathbf{y}}$ (where e.g. $(-1)^{\hat{a}_{\mathbf{x}}} = A_{\mathbf{x}}$) such that $\langle \psi|(-1)^{\hat{a}_{\mathbf{x}}+\hat{b}_{\mathbf{y}}}|\psi\rangle = \langle u_{\mathbf{x}}, v_{\mathbf{y}}\rangle = (-1)^{\mathbf{x}\cdot\mathbf{y}} B_{\mathbf{y}}$, and hence

$$\begin{aligned}
E_{\mathbf{y}} &= \frac{1}{2^n} \sum_{\mathbf{x}} E_{\mathbf{x}\mathbf{y}} \\
&= \frac{1}{2^n} \sum_{\mathbf{x}} (-1)^{\mathbf{x}\cdot\mathbf{y}} \langle \psi|(-1)^{\hat{a}_{\mathbf{x}}+\hat{b}_{\mathbf{y}}}|\psi\rangle \\
&= \frac{1}{2^n} \sum_{\mathbf{x}} B_{\mathbf{y}} \\
&= B_{\mathbf{y}}
\end{aligned}$$

□

Corollary 3. *There exists a quantum strategy for the Information Causality game with $m = 1$ which succeeds with probability $\frac{1}{2}\left(1 + \frac{1}{\sqrt{n}}\right)$.*

Proof. Suppose that Alice and Bob play the Non-local Inner Product game where Alice's input string is chosen uniformly at random, but Bob's bit string is chosen uniformly from the set of n-bit strings which have exactly one non-zero bit. By the "if" direction of Theorem 9, it is possible to find a quantum strategy such that $E_{\mathbf{y}} = \frac{1}{\sqrt{n}}$ for all such $\mathbf{y}$ (since this choice of $E_{\mathbf{y}}$ satisfies $\sum_{\mathbf{y}} E_{\mathbf{y}}^2 = 1$). Consequently, for each string $\mathbf{y}$ with one non-zero bit their probability of success conditioned on $\mathbf{y}$ is

$$P_{\mathbf{y}} = \frac{1 + E_{\mathbf{y}}}{2} = \frac{1}{2}\left(1 + \frac{1}{\sqrt{n}}\right). \tag{4.46}$$

The total probability of success is given by averaging $P_{\mathbf{y}}$ over the distribution on Bob's string $\mathbf{y}$. However, since $P_{\mathbf{y}}$ is the same for all $\mathbf{y}$, the probability of success is simply $\frac{1}{2}\left(1 + \frac{1}{\sqrt{n}}\right)$. Recall that any strategy for this version of the Non-local Inner Product game can be transferred to a strategy for $m = 1$ Information Causality with the same probability of success. □

The probability of success in Corollary 3 is equal to the upper bound (4.35) on probability of success for EARACs, hence this is the optimal probability of success and (4.35) can be saturated for all integers n; moreover, any optimal strategy for the modified inner product game converts to an optimal strategy in $m = 1$ Information Causality. For larger values of m, the optimal success probability for EARACs is not generally known; however, it's worth noting that in deriving Tsirelson's bound from Information Causality, the protocol given in Section 4.1.3 only involves one bit being transmitted from Alice to Bob, who then computes the direct product with his own output. If in fact the Information Causality game specified that at most one bit is transmitted to Bob, and the corresponding condition $I \leq 1$ was used instead of $I \leq m$, then the same results follow. Although it is interesting to consider the case of general m, the crux of the argument lies in the $m = 1$ scenario.

The quadratic bias bound $\sum_{\mathbf{y}} E_{\mathbf{y}}^2 \leq 1$ given in Theorem 9 for the Non-local Inner Product game appears to capture a great deal about the set of quantum correlations. In the case where Bob's input $\mathbf{y}$ is chosen uniformly from all bit strings of length n, the probability of success conditioned on $\mathbf{y}$ obeys the bound:

$$\begin{aligned} P_{\mathbf{y}} &= \frac{1}{2}\left(1 + \frac{1}{2^n}\sum_{\mathbf{y}} E_{\mathbf{y}}\right) \\ &\leq \frac{1}{2}\left(1 + \sqrt{\frac{1}{2^n}\sum_{\mathbf{y}} E_{\mathbf{y}}^2}\right) \\ &\leq \frac{1}{2}\left(1 + \frac{1}{\sqrt{2^n}}\right). \end{aligned} \tag{4.47}$$

Again, since this value is constant, the total probability of success also obeys this bound. In the case $n = 1$, the Non-local Inner Product game is the CHSH game, and the formula (4.47) reduces to Tsirelson's bound. In fact, in this case the inequality $E_0^2 + E_1^2 \leq 1$ describes the exact quantum boundary in a particular 2-D slice of the non-signaling polytope. However, it is not known whether either the quadratic bias bound or the Information Causality principle is sufficient to recover the complete boundary on the set of quantum correlations.

Just as with the Information Causality principle, note that it is possible to saturate the quadratic bias bound using only classical correlations. If Alice and Bob pick some fixed value $\mathbf{y} = \mathbf{y}'$, and Alice outputs $\mathbf{x} \cdot \mathbf{y}'$ whilst Bob outputs 0, then they win every time Bob's input is $\mathbf{y}'$, hence $E_{\mathbf{y}'} = 1$; clearly this implies $E_{\mathbf{y}} = 0$ for $\mathbf{y} \neq \mathbf{y}'$. Unfortunately, although the quadratic bias bound is appealing for its relatively simple proof and obvious connection to success probabilities, it lacks the direct physical intuition of the Information Causality principle.

4.3 The role of entropy

In Section 4.1 we derived the Information Causality principle from the existence of a measure of mutual information $\tilde{I}$ which satisfies four conditions: Consistency with classical mutual information, a Data Processing inequality, the Chain Rule and Symmetry. This in itself suggests an interesting perspective: any physical theory which admits a "reasonable" measure of mutual information (where "reasonable" here means "satisfying at least the above four conditions") must obey the Information Causality principle, and therefore must obey any bounds on non-signaling correlations which follow from the Information Causality principle (including Tsirelson's bound). In Boxworld, where strongly non-local states such as the PR-box are permitted, clearly *no* such measure of mutual information can be attributed to pairs of systems.

The usual classical and quantum mutual informations, denoted I_c and I_q respectively, are known to obey all four of the above conditions [53]. However, neither of these represent the most basic informational primitive of their respective theory; rather, they are often both defined as the same function of the classical or quantum entropy:

$$I(X:Y) = H(X) + H(Y) - H(X,Y). \tag{4.48}$$

This suggests that it is worth exploring the principle of Information Causality for measures of mutual information which are defined in a similar way, i.e. $\tilde{I}(X:Y) = \tilde{H}(X) + \tilde{H}(Y) - \tilde{H}(X,Y)$, where $\tilde{H}$ is some measure of entropy in a general probabilistic theory. Measures of entropy in the framework of general probabilistic theories have already received attention in the literature, including their relation to mutual information and Information Causality [38-40]. Unsurprisingly, given the above comments, the measures of entropy considered have some intuitive properties, but tend to disobey one or more natural principles.

Suppose then that a mutual information $\tilde{I}$ between two systems is defined in terms of an entropy function $\tilde{H}$ on individual systems as in (4.48). What properties

must the entropy satisfy in order that $\tilde{I}$ satisfies the four assumptions required in the derivation of Information Causality? Since (4.48) is symmetric in X and Y, $\tilde{I}$ will automatically satisfy Symmetry. If $\tilde{H}$ reduces to the classical entropy H_c when X is classical, then $\tilde{I}(X:Y)$ will also satisfy Consistency. The Chain Rule, which we earlier commented was without immediate physical significance, also follows straight from (4.48):

$$\tilde{I}(X,Y:Z) - \tilde{I}(X:Z) = \tilde{I}(Y:X,Z) - \tilde{I}(X:Y)$$

$$\begin{aligned} \iff & \Big[\tilde{H}(X,Y) + \tilde{H}(Z) - \tilde{H}(X,Y,Z)\Big] - \Big[\tilde{H}(X) + \tilde{H}(Z) - \tilde{H}(X,Z)\Big] \\ & = \Big[\tilde{H}(Y) + \tilde{H}(X,Z) - \tilde{H}(Y,X,Z)\Big] - \Big[\tilde{H}(X) + \tilde{H}(Y) - \tilde{H}(X,Y)\Big] \end{aligned}$$

$$\begin{aligned} \iff & \tilde{H}(X) + \tilde{H}(X,Y) + \tilde{H}(X,Z) - \tilde{H}(X,Y,Z) \\ & = \tilde{H}(X) + \tilde{H}(X,Y) + \tilde{H}(X,Z) - \tilde{H}(X,Y,Z) \end{aligned} \tag{4.49}$$

Thus the awkward-looking Chain Rule simply disappears, when everything is formulated in terms of entropies rather than mutual information. We are left with the Data Processing inequality; for any two systems X and Y, and an allowed transformation τ on system Y alone:

$$\begin{aligned} & \tilde{I}(X:Y) \geq \tilde{I}(X:\tau(Y)) \\ \iff & \tilde{H}(X) + \tilde{H}(Y) - \tilde{H}(X,Y) \geq \tilde{H}(X) + \tilde{H}(\tau(Y)) - \tilde{H}(X,\tau(Y)) \\ \iff & \tilde{H}(X,\tau(Y)) - \tilde{H}(\tau(Y)) \geq \tilde{H}(X,Y) - \tilde{H}(Y) \\ \iff & \tilde{H}(X|\tau(Y)) \geq \tilde{H}(X|Y) \end{aligned} \tag{4.50}$$

where we identify the generalised conditional entropy,

$$\tilde{H}(X|Y) = \tilde{H}(X,Y) - \tilde{H}(Y), \tag{4.51}$$

and interpret it as the uncertainty about the state of system X, given knowledge of the state of system Y. The inequality (4.50) has an intuitive physical interpretation in this sense: one cannot learn more about system X by performing a transformation on system Y alone, hence one's uncertainty about X, given Y, should not decrease just because of a local transformation performed at Y.

The constraint (4.50) can also be written in the following way:

$$\begin{aligned} &\tilde{H}(X,\tau(Y)) - \tilde{H}(X,Y) \geq \tilde{H}(\tau(Y)) - \tilde{H}(Y) \\ &\quad \iff \Delta_\tau \tilde{H}(X,Y) \geq \Delta_\tau \tilde{H}(Y), \end{aligned} \tag{4.52}$$

where τ is a local operation at Y, and $\Delta_\tau \tilde{H}$ represents the change in entropy due to τ. In this form, the inequality is a statement about changes in entropy rather than about the conditional entropy, thus has the advantage that one does not need to introduce the idea of conditional entropy. The inequality (4.52) also has an intuitive physical interpretation: local transformations may generate uncertainty on a bipartite level which is greater than the uncertainty generated only locally (for example, the transformation may destroy correlations between the systems).

We have thus demonstrated that the Information Causality principle holds in any theory that permits an entropy $\tilde{H}$ satisfying two very reasonable properties:

(i) Consistency If system X is classical, $\tilde{H}(X)$ reduces to the classical entropy of the probability distribution on X.

(ii) Ancillary Evolution For any local transformation τ on system Y,

$$\tilde{H}(X|\tau(Y)) \geq \tilde{H}(X|Y). \tag{4.53}$$

For the Information Causality principle to follow from these principles, it is in fact also necessary for the theory to contain certain local transformations. In Section 2.5.4 we assumed that any transformation which maps the set of allowed states to itself in a convex-linear fashion is part of the theory. However, there is no overriding reason why the theory should not contain a strict subset of all such

transformations. Indeed, one might conceive of a physical theory whose states are the full set of Boxworld states, yet for which no transformations exist, either from a system to itself or from one system to another. In such a theory, Ancillary Evolution holds for every measure of entropy. Therefore, even though Information Causality is violated by this theory, both Consistency and Ancillary Evolution hold for the measure of entropy which assigns the classical entropy to states of classical systems, and 0 to states of all other systems.

It is interesting to determine the minimal set of transformations which a theory must allow, alongside admitting an entropy that obeys Consistency and Ancillary Evolution, in order for the Information Causality principle to hold. This is achieved by examining where the Data Processing inequality is used in the proof of the Information Causality principle in terms of the mutual information. There are three types of transformation to which the Data Processing inequality is applied: ``tracing out'' of a system; ``constructing'' a system, i.e. a transformation from a classical system (possibly an ``empty'' system denoted by $\emptyset$) to any other type of system; and ``measuring'' a system, i.e. a transformation from any type of system to a classical system. So long as all permissible transformations belonging to these classes are contained in the theory, then Information Causality will indeed follow from Consistency and Ancillary Evolution.

Nevertheless, this reduction in the number and complexity of properties motivates an attempt to reformulate - and derive - the Information Causality principle purely in terms of entropies; this is the topic of Section 4.3.1. In Section 4.3.2 we then give a brief overview of some attempts at introducing measures of entropy into general physical theories, and discuss how they relate to our own perspective on the Information Causality principle.

4.3.1 An entropic information causality

Before deriving an entropic version of the Information Causality principle, we deduce a number of standard inequalities which must hold for any measure of entropy satisfying the Consistency and Ancillary Evolution properties from the

previous section:

Subadditivity For systems X and Y, if $\tau : Y \to \emptyset$ is the "tracing out" map, then by Ancillary Evolution:

$$\tilde{H}(X|\emptyset) \geq \tilde{H}(X|Y). \tag{4.54}$$

The entropy of X given a traced-out system must be the same as the entropy of X, hence:

$$\begin{aligned} \tilde{H}(X) &\geq \tilde{H}(X,Y) - \tilde{H}(Y) \\ \Rightarrow \tilde{H}(X,Y) &\leq \tilde{H}(X) + \tilde{H}(Y). \end{aligned} \tag{4.55}$$

If X and Y are independent systems, then there is a local map corresponding to the preparation of the state of Y independent of X, $\omega : \emptyset \to Y$. Ancillary Evolution then implies that $H(X|Y) \geq H(X|\emptyset)$; combining this with (4.55) gives $\tilde{H}(X,Y) = \tilde{H}(X) + \tilde{H}(Y)$.

Strong Subadditivity For three systems X_1, X_2 and Y, if $\tau : (X_2, Y) \to Y$ is the marginalization map, then it follows from Ancillary Evolution that $\tilde{H}(X_1|X_2,Y) \leq \tilde{H}(X_1|Y)$. Hence,

$$\begin{aligned} \tilde{H}(X_1,X_2|Y) &= \tilde{H}(X_1,X_2,Y) - \tilde{H}(Y) \\ &= \tilde{H}(X_1,X_2,Y) - \tilde{H}(X_2,Y) + \tilde{H}(X_2,Y) - \tilde{H}(Y) \\ &= \tilde{H}(X_1|X_2,Y) + \tilde{H}(X_2|Y) \\ &\leq \tilde{H}(X_1|Y) + \tilde{H}(X_2|Y). \end{aligned} \tag{4.56}$$

The process may be iterated for a larger number of systems:

$$\tilde{H}(X_1,\ldots,X_n|Y) \leq \tilde{H}(X_1|Y) + \ldots + \tilde{H}(X_n|Y). \tag{4.57}$$

Classical Positivity One's uncertainty about the state of a classical system X is

never negative, even when one conditions on another system Y (which may or may not be independent of X, or classical itself):

$$\text{System X is classical} \Rightarrow \tilde{H}(X|Y) \geq 0. \tag{4.58}$$

To prove this, suppose that ψ is the state of system X,. Note that the state of X is described by a probability distribution on a finite set E of outcomes for a single fiducial measurement. For each outcome $e \in E$, there is an associated probability p_e, a deterministic state x_e on system X which outputs e with certainty, and a corresponding reduced state ψ_e of Y. The state ψ is therefore the convex combination of the product states $x_e \cdot \psi_e$ (i.e. the product state where X is in state x_e and Y is in state ψ_e) weighted according to p_e.

Suppose that the system X' is identical to X, and that the joint system X, X' is initially in a convex combination of the product states $x_e \cdot x'_e$, weighted according to p_e. On system X', let $\tau : X' \to Y$ be the transformation described by performing the fiducial measurement E on system X', and preparing system Y in state ψ_e if outcome e is obtained. The state of system X, Y is now the convex combination of the product states $x_e \cdot \psi_e$ weighted according to p_e, i.e. after the transformation τ, the state of system X, Y is ψ.

By the Consistency postulate, before the transformation τ is performed, the conditional entropy $\tilde{H}(X|X')$ reduces to the classical conditional entropy:

$$\tilde{H}(X|X') = H_c(X|X') = H_c(X, X') - H_c(X'). \tag{4.59}$$

Since X' is perfectly correlated with X, $H_c(X, X') = H_c(X')$ therefore $\tilde{H}(X|X') = 0$. Then by the Ancillary Evolution postulate:

$$\tilde{H}(X|Y) = \tilde{H}(X|\tau(X')) \geq \tilde{H}(X|X') = 0. \tag{4.60}$$

The Subadditivity, Strong Subadditivity and Classical Positivity properties must be obeyed by any measure of entropy which satisfies the Consistency and Ancillary Evolution postulates. Using these three natural properties, it is quite easy to derive an Entropic Information Causality principle. Recall that in the Information Causality game, Alice is given a bit-string **a** of length n and Bob is given a number $1 \leq k \leq n$. After Alice transmits an m-bit message **x** to Bob, Bob outputs $\beta(k)$, his best guess at the value of a_k, the kth bit of Alice's string. Unlike in Section 4.1.3, we will no longer assume that Alice's input **a** is uniformly chosen.

Suppose that Alice and Bob share a state of some system A, B, in a general probabilistic theory that admits some measure of entropy $\tilde{H}$ satisfying the Consistency and Ancillary Evolution postulates. Whatever their strategy, Bob's final guess $\beta(k)$ is ultimately deduced from his system B and the classical system **X** which holds the message **x**; therefore there exists a transformation from **X**, B to $\beta(k)$, and so by Consistency and Ancillary Evolution:

$$\sum_k H_c(a_k|\beta(k)) = \sum_k \tilde{H}(a_k|\beta(k)) \geq \sum_k \tilde{H}(a_k|\mathbf{x}, B). \tag{4.61}$$

Writing **a** as a multipartite state $a_1, \ldots, a_n$ and using the iterated Strong Subadditivity property (4.57),

$$\begin{aligned} \sum_k \tilde{H}(a_k|\mathbf{x}, B) &\geq \tilde{H}(\mathbf{a}|\mathbf{x}, B) \\ &= \tilde{H}(\mathbf{a}, \mathbf{x}, B) - \tilde{H}(\mathbf{x}, B). \end{aligned} \tag{4.62}$$

Then by Subadditivity on the bipartite system **x**, B,

$$\tilde{H}(\mathbf{a}, \mathbf{x}, B) - \tilde{H}(\mathbf{x}, B) \geq \tilde{H}(\mathbf{a}, \mathbf{x}, B) - \tilde{H}(B) - \tilde{H}(\mathbf{x}). \tag{4.63}$$

Introducing a term $\tilde{H}(\mathbf{a}) - \tilde{H}(\mathbf{a})$ and using the fact that $\tilde{H}(\mathbf{a}, B) = \tilde{H}(\mathbf{a}) + \tilde{H}(B)$ since $\mathbf{a}$ and B are independent,

$$\begin{aligned}\tilde{H}(\mathbf{a}, \mathbf{x}, B) - \tilde{H}(B) - \tilde{H}(\mathbf{x}) &= \tilde{H}(\mathbf{a}, \mathbf{x}, B) - \tilde{H}(B) - \tilde{H}(\mathbf{a}) + \tilde{H}(\mathbf{a}) - \tilde{H}(\mathbf{x}) \\ &= \tilde{H}(\mathbf{x}, \mathbf{a}, B) - \tilde{H}(\mathbf{a}, B) + \tilde{H}(\mathbf{a}) - \tilde{H}(\mathbf{x}) \\ &= \tilde{H}(\mathbf{x}|\mathbf{a}, B) + \tilde{H}(\mathbf{a}) - \tilde{H}(\mathbf{x}). \end{aligned} \tag{4.64}$$

Since $\mathbf{a}$ and $\mathbf{x}$ are classical, and using the positivity of $\tilde{H}$ on classical systems,

$$\tilde{H}(\mathbf{x}|\mathbf{a}, B) + \tilde{H}(\mathbf{a}) - \tilde{H}(\mathbf{x}) \geq H_c(\mathbf{a}) - H_c(\mathbf{x}). \tag{4.65}$$

Putting this all together gives an entropic formulation of the Information Causality principle:

$$\sum_k H_c(a_k|\beta(k)) \geq H_c(\mathbf{a}) - H_c(\mathbf{x}). \tag{4.66}$$

The standard Information Causality principle $\sum_{k=1}^{N} I_c(a_k : \beta(k)) \leq m$ can be deduced from (4.66) as follows: since $\mathbf{x}$ is an m-bit string

$$\begin{aligned} m &\geq H_c(\mathbf{x}) \\ &\geq H_c(\mathbf{a}) - \sum_k H_c(a_k|\beta(k)). \end{aligned} \tag{4.67}$$

If the string $\mathbf{a}$ is uniformly random, then $H_c(\mathbf{a}) = \sum_k H_c(a_k)$:

$$\begin{aligned} m &\geq \sum_k [H_c(a_k) - H_c(ak|\beta(k))] \\ &= \sum_k I_c(a_k : \beta(k)). \end{aligned} \tag{4.68}$$

This entropic formulation of Information Causality has several advantages: firstly, it deals in entropy, which is a more basic informational quantity than mutual information, and has a more immediate interpretation as the uncertainty inherent in the state of a system. Secondly, it leads to an even more pleasing statement re-

garding measures of information in general physical theories. If a general physical theory admits some measure of entropy satisfying Consistency and Ancillary Evolution (which are both very reasonable demands), then the Information Causality principle (and subsequent bounds on non-locality) must hold in that theory. To put it more strongly, Tsirelson's bound cannot be violated in any physical theory which admits an extension of classical entropy that behaves reasonably under dynamics. Though this fact already follows from the proofs contained in the original Information Causality paper [51], it is striking to see it presented in this way.

Both the entropic Information Causality principle (4.66) and the quadratic bias bound (4.39) can be seen as fundamental restrictions on information processing which, when applied to the full set of non-signaling correlations, recovers many of the bounds on quantum non-locality. However, whereas the quadratic bias bound suffers from a lack of a simple, physical interpretation, it is relatively easy to motivate equation (4.66) from physical considerations. After his guess $\beta(k)$, the remaining uncertainty that Bob has about Alice's inputs should be no less than his original uncertainty $H_c(\mathbf{a})$, minus the information in the message $\mathbf{x}$.

4.3.2 Entropy in general physical theories

We now describe some measures of entropy which have previously been introduced for general probabilistic theories such as Boxworld. These measures are invariably seen as a generalisation of the Shannon entropy for classical systems, and the Consistency property will hold. Hence, in strongly non-local theories such as Boxworld, it is expected that Ancillary Evolution will not hold, and that this measure of entropy will behave poorly under some dynamical transformation.

It is most convenient to introduce these entropies using the convex set representation of general probabilistic theories, described in Chapter 2. Recall that states and effects are assigned vectors s and e respectively, such that the probability of obtaining the outcome corresponding to e, for a system in state s, is given by the inner product $\langle e, s\rangle$. A measurement $\mathcal{M}$ is described by a set of effects $\{e_1, \ldots e_r\}$ which sum to the unit effect: $\sum_i e_i = u$.

Measurement entropy

For any state s, a measurement $\mathcal{M}$ defines a classical probability distribution $\mathcal{M}(s) = \{\langle e_i, s\rangle\}$; it is tempting to define a measure of entropy which is the minimum of the classical Shannon entropy of this distribution over all measurements $\mathcal{M}$. However, this poses a problem: an entropy defined in this way could always be decreased by simply "merging" effects together and considering the corresponding outcomes to be the same - at the extreme, the unit measurement $\{\mathcal{U}\}$ consisting of just the unit effect, will always give zero entropy.

In order to counter this problem, recall the notion of a refinement of a measurement given in Section 2.5.2, and note that "merging" effects as described above leads to a coarse-graining of the measurement. Therefore we wish to consider Shannon entropies of outcome distribution only over pure measurements, i.e. those which are made up of (non-parallel) extreme-ray effects.

We will recap the notion of refinement for convenience: measurement $\mathcal{M}_2 = \{f_1, \ldots, f_s\}$ is a *refinement* of $\mathcal{M}_1$ if there exists a function $\phi : \mathcal{M}_2 \to \mathcal{M}_1$ such that each e_i is the sum of those f_j which map to it:

$$e_i = \sum_{j:\phi(f_j)=e_i} f_j. \tag{4.69}$$

Thus, every measurement is a refinement of the unit measurement, which is completely non-informative; in general, the refined measurement $\mathcal{M}_1$ will be strictly more informative than $\mathcal{M}_2$. However, simply splitting up e_1 into two effects $f_1 = f_2 = \frac{1}{2}e_1$ will result in a refinement which is not more informative (this could be achieved in practice by a post-processing of results: if outcome 1 is obtained, the actual output is uniformly assigned one of two distinct, new values). To avoid this issue, the refinement is said to be *trivial* if $\phi(f_j) = e_i \Rightarrow f_j = \lambda_j e_i$ for real numbers λ_j. A measurement is *fine-grained* if it admits no non-trivial fine-grainings. Let $\mathbb{M}^*$ denote the set of all fine-grained measurements.

The *measurement entropy* H^{me} of a state is the infimum of the Shannon en-

tropies of the outcome distributions for all fine-grained measurements [38, 40]:

$$H^{me}(s) = \inf_{\mathcal{M}\in\mathbb{M}^*} H_c(\mathcal{M}(s)). \tag{4.70}$$

In a classical system, a fine-grained measurement consists only of vectors which are scalar multiples of the effects belonging to the single fiducial measurement. It is easily checked that such a measurement is a trivial refinement of the fiducial measurement itself, and that the outcome entropy is therefore minimized for the fiducial measurement. Thus the measurement entropy reduces to the Shannon entropy for classical systems. In a quantum system, a fine-grained measurement consists of measurement operators which are scalar multiples of rank-one projectors; it turns out also that the measurement entropy reduces to the von Neumann entropy [38, 40], however this is not immediately obvious. In Boxworld, a fine-grained measurement is one in which all effects are tensor products of fiducial effect vectors, i.e. $e_i = X^{(1)}_{a_1|x_1} \otimes \cdots \otimes X^{(n)}_{a_n|x_n}$.

All three of the above theories share the property that whenever $e^{(1)}$ is an extreme-ray effect on system 1 and $e^{(2)}$ is an extreme-ray effect on system 2, then $e^{(1)} \otimes e^{(2)}$ is an extreme-ray effect of the joint system (in Boxworld, as with all max-tensor product theories, these are the only joint extreme-ray effects). Suppose that $\mathcal{M}^{(1)} = \{e^{(1)}_i\}$ and $\mathcal{M}^{(2)} = \{e^{(2)}_j\}$ are fine-grained measurements on systems 1 and 2 respectively. For any joint state s, the outcome distribution $\mathcal{M}^{(1)} \otimes \mathcal{M}^{(2)}(S) = \{\langle e^{(1)}_i \otimes e^{(2)}_j, s\rangle\}$ has as its marginal on system 1 the outcome distribution $\mathcal{M}^{(1)}(s)$ (where $s^{(1)}$ is the reduced state on system 1), and similarly for system 2. Therefore, by subadditivity of the Shannon entropy,

$$H_c(\mathcal{M}^{(1)} \otimes \mathcal{M}^{(2)}(s)) \leq H_c(\mathcal{M}^{(1)}(s^{(1)})) + H_c(\mathcal{M}^{(2)}(s^{(2)})). \tag{4.71}$$

By considering tensor products of fine-grained measurements which approach the measurement entropy to arbitrary precision on the RHS, it is clear that the measurement entropy is indeed subadditive in theories for which the aforementioned property holds.

On the other hand, the measurement entropy fails to obey strong subadditivity for some tripartite Boxworld systems [38, 40]. This is demonstrated in [38] using a similar procedure as employed in Section 4.1.3 in order for Bob to perfectly guess one of a pair of random bits given to Alice. Suppose that the random bits given to Alice are represented by two classical systems A_1 and A_2, whilst Bob is in possession of a g-bit system B. In order to obtain a random bit string (a_1, a_2), Alice simply measures her systems using the single fiducial measurements x_1 and x_2. Depending on whether Bob wishes to obtain the value of the first bit or the second, his measurement choice is given by $x_3 = 1$ or $x_3 = 2$ respectively. For Bob's output a_3, suppose that the tripartite system is in the state represented by:

$$P\left(a_1, a_2, a_3 | x_1, x_2, x_3\right) = \frac{1}{4}\delta(a_3 = a_{x_3}). \tag{4.72}$$

The only system with more than one fiducial measurement choice is B, hence to show that this is non-signaling, it suffices to observe that the reduced system A_1A_2 is in a completely uniform state, no matter the value of x_3. The reduced state on B is the completely mixed state of a g-bit, hence $H^{me}(B) = 1$. Since Bob can choose his measurement in order to correlate exactly with either a_1 or a_2, then $H^{me}(A_1B) = H^{me}(A_2B) = 1$. Any measurement on the total system A_1A_2B generates 4 possible outcomes with uniform probability, hence $H^{me}(A_1A_2B) = 2$. Therefore,

$$H^{me}(A_1A_2B) + H^{me}(B) = 3 > 2 = H^{me}(A_1B) + H^{me}(A_2B). \tag{4.73}$$

The measurement entropy does not obey Classical Positivity in Boxworld, either. Consider a bipartite system composed of a classical bit system A, and a system B with two measurement choices, each of which admits three outputs, which is in the state represented by the following matrix:

	$x_2 = 1$			$x_2 = 2$		
	$a_2 = 1$	$a_2 = 2$	$a_2 = 3$	$a_2 = 1$	$a_2 = 2$	$a_2 = 3$
$a_1 = 1$	1/2	0	0	0	1/4	1/4
$a_1 = 2$	0	1/4	1/4	1/2	0	0

Now consider a bipartite measurement which takes the following form: first measure system A to obtain a_1, then measure system B using the measurement choice $x_1 = a_1$. This guarantees the output $a_2 = 1$ on the second system, and a_1 takes each of its possible values with probability $\frac{1}{2}$. Hence $H^{me}(AB) \leq H_c(\{\frac{1}{2}, \frac{1}{2}\}) = 1$. However, each of the fine-grained measurements on the reduced state on system 2 results in an outcome distribution $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\}$, which has entropy 1.5. Thus $H^{me}(A|B) = H^{me}(A|B) - H^{me}(B) \leq -0.5$, in violation of the Classical Positivity principle.

Mixing Entropy

A second measure of entropy for general probabilistic theories which reduces to the Shannon and von Neumann entropies on classical and quantum systems respectively, is the *mixing entropy* [38, 40]. Every decomposition of a state s into pure states:

$$s = \sum_{i=1}^{r} p_i s_i \tag{4.74}$$

corresponds to a probability distribution $\mathbf{p} = \{p_i\}$. Suppose that $\mathbb{P}(s)$ is the set of probability distributions that result from decompositions of s in this manner. The mixing entropy H^{mi} of s is the infimum of the classical entropy over this set:

$$H^{mi}(s) = \inf_{\mathbf{p} \in \mathbb{P}(s)} H_c(\mathbf{p}). \tag{4.75}$$

As shown in [38], the mixing entropy is not subadditive (let alone strongly subadditive) in Boxworld. An explicit counterexample is the following state s of a system comprising two g-bits:

		$x_2 = 1$		$x_2 = 2$	
		$a_2 = 1$	$a_2 = 2$	$a_2 = 1$	$a_2 = 2$
$x_1 = 1$	$a_1 = 1$	1/4	3/8	3/8	1/4
	$a_1 = 2$	3/8	0	1/4	1/8
$x_1 = 2$	$a_1 = 1$	1/4	3/8	1/4	3/8
	$a_1 = 2$	3/8	0	3/8	0

The reduced state on either system is the state which outputs 1 with probability $\frac{5}{8}$ and 2 with probability $\frac{3}{8}$, regardless of the measurement choice. This confirms that the state is non-signaling, and demonstrates that

$$H^{mi}(A) = H^{mi}(B) = 0.95... \,. \tag{4.76}$$

Recall that there are 16 local deterministic and 8 non-local PR-box variants making up the extreme states of the bipartite system AB (the PR-box variants are obtained from the PR-box by local relabellings of the inputs and outputs). Out of these 24 extreme states, only 4 of the local deterministic states and 1 PR-box variant have positive entries only where s also has positive entries. For each of these extreme states s_i, it can then be checked that for $p_i s_i$ to have all its positive entries less than or equal to those of s, we must have $p_i \leq \frac{1}{4}$ for every i. This implies that any decomposition of s consists of at least 4 extreme states, none of whose coefficients is greater than $\frac{1}{4}$. Hence

$$H^{mi}(AB) \geq 2 > H^{mi}(A) + H^{mi}(B). \tag{4.77}$$

However, Classical Positivity *is* satisfied by the mixing entropy [40]. If system A is classical, then for any system B, the extreme states of system AB are of the form $\omega^A \otimes \omega^B$, where ω^A are ω^B are reduced states on systems A and B respectively. Then any decomposition of a state s of system AB takes the form

$$s = \sum_{i,j} p_{ij} \omega_i^A \otimes \omega_j^B = \sum_j \left(\sum_i p_{ij} \omega_i^A \right) \otimes \omega_j^B, \tag{4.78}$$

where ω_i^A correspond to the outcomes of the single fiducial measurement on system A. Writing $q_j = \sum_i p_{ij}$, one decomposition of the reduced state on system B is

$$s^B = \sum_j q_j \omega_j^B. \tag{4.79}$$

Any decomposition of s generates an entropy $H_c(\mathbf{p}) \geq H_c(\mathbf{q}) \geq H^{mi}(B)$, therefore $H^{mi}(AB) \geq H^{mi}(B)$ and so $H^{mi}(A|B) \geq 0$.

Relation to Information Causality

Both the measurement and mixing entropies are extensions of the von Neumann entropy for quantum systems, and therefore obey all three of Subadditivity, Strong Subadditivity and Classical Positivity when confined to only classical or quantum theory. In Boxworld however, they both violate two of these natural properties; since they both obey Consistency, they both must certainly violate Ancillary Evolution. The entropic formulation of Information Causality gives two interesting extra perspectives on this: firstly, because Boxworld allows violations of Tsirelson's bound, there simply can be no measure of entropy that will be "perfect" in the sense of obeying both Ancillary Evolution and Consistency. Even if there exists some measure of entropy in Boxworld which satisfies Subadditivity, Strong Subadditivity and Classical Positivity, it will inevitably violate some other natural consequence of Ancillary Evolution and Consistency. Secondly, despite already possessing many natural properties, both the measurement and mixing entropies will always disobey Ancillary Evolution in *any* theory allowing a violation

of Tsirelson's bound.

On one hand, this is slightly disappointing: without natural extensions of entropies which obey natural laws, we may find it more difficult to explore information processing in more general physical paradigms. Ordinarily, counter-intuitive phenomena provide us with something to learn from rather than to be scared of, however in the case of Boxworld it seems that even the most fundamental tenets of information processing run into intractable obstacles. On the other hand, this provides exciting evidence that entropy plays a deep role in limiting the amount of non-locality which is achievable in nature. The existence of a reasonable measure of entropy may even be a more fundamental physical requirement than the standard Hilbert-space quantum postulates.

4.4 Discussion

We have explored two different perspectives on the Information Causality game presented in [51]. The first perspective is to consider the probability of success as the figure of merit: we see that quantum theory gives an advantage which is not reflected in the original figure of merit I, for which quantum and classical theory perform equally well. Investigating how these probabilities are involved in deriving Tsirelson's bound from Information Causality leads us to a quadratic quantum bound on the biases achieved, given the different inputs for Bob in the non-local inner product game:

$$\sum_{\mathbf{y}} E_{\mathbf{y}}^2 \leq 1. \tag{4.80}$$

This inequality holds for any probability distribution over Bob's inputs, and hence applies to the version of Information Causality without any communication from Alice to Bob. This is another example of a bound which quantum and classical correlations can both saturate, but stronger non-local correlations can violate, and from which one can recover the same sections of quantum boundary within

two-dimensional slices of the non-signaling polytope. Furthermore, the fact that quantum correlations allow one to achieve any set of biases satisfying this rule suggests that it captures a significant amount of information about the set of quantum correlations. The question remains whether quadratic inequalities such as this are sufficient to completely recover the quantum boundary, and whether other quadratic inequalities can tell us useful things about quantum capabilities in other non-local tasks.

The second perspective is to view the Information Causality inequality not just as a constraint on possible physical theories but also as a consequence of the existence of a 'good' measure of entropy. In this direction, we have shown that Information Causality can be derived given any extension of the entropy from classical to more general systems which satisfies $\tilde{H}(X|\tau(Y)) \geq \tilde{H}(X|Y)$ under local transformations on system Y. Conversely, any theory which violates Information Causality (such as Boxworld) cannot have any measure of entropy which obeys the above evolution law and agrees with the Shannon entropy for classical systems. In particular, any general probabilistic theory which violates Tsirelson's bound cannot possess a measure of entropy which obeys these conditions; this strange link between non-locality and entropy is definitely worth exploring further.

Given the results of Section 4.3, as well as those of [38-40], it seems that the existence of a "good" entropy for quantum theory, which shares so many of the properties of the classical entropy, is very special within the class of general probabilistic theories. Are there other theories for which one can define a "reasonable" measure of entropy satisfying Consistency and Ancillary Evolution, or is this a defining feature of quantum theory? Of course, we could consider a general probabilistic theory in which states and measurements are represented by the same Hilbert-space objects as quantum theory, but which allows only a strict subset of quantum states. Trivially, the restriction of the von Neumann entropy would satisfy the same properties as for full quantum theory.

A more interesting question is whether a reasonable entropy is definable in theories which *cannot* be simulated by quantum theory, for example those which have

non-local correlations that are unattainable via measurements of quantum states. Since Information Causality may be deduced from simple entropic properties, the existence of a reasonable entropy potentially places stronger bounds on quantum theory than Information Causality does alone. It would be interesting to look for other games in which possessing a reasonable measure of entropy precludes one from performing better than is possible classically.

Chapter 5

Reversible Boxworld dynamics

To do mathematics is to engage in an act of discovery and conjecture, intuition and inspiration; to be in a state of confusion— not because it makes no sense to you, but because you gave it sense and you still don't understand what your creation is up to; to have a breakthrough idea; to be frustrated as an artist; to be awed and overwhelmed by an almost painful beauty; to be alive, damn it.

"A Mathematician's Lament"
Paul Lockhart

In this chapter we explore the class of reversible dynamics in the general probabilistic theory known as Boxworld. Recall from Section 2.5.4 that the allowed transformations of a general probabilistic theory are determined once its state space has been defined, and that reversible transformations are those for which there is an allowed inverse transformation. Thus this chapter is concerned with the set of reversible transformations that are permitted by the state space structure of Boxworld. As this state space is constrained only by the conditions of non-signaling and local tomography for composite systems (as well as the probabilistic laws of positivity and normalization), Boxworld represents something of a canonical example of a general probabilistic theory, and is a natural benchmark against which to analyse various features of quantum theory. One such feature is the principle of reversible transitivity, which states that any pure state of either an individual or composite system may be reversibly transformed to any other. This holds trivially in quantum theory; indeed, any set of n orthonormal states in n-dimensional Hilbert space may be linearly mapped to any other such set, and this mapping extends naturally to a unitary transformation on the Hilbert space.

The importance of reversible transitivity as a defining feature of quantum theory is indicated by its ubiquitous use (or the use of stronger versions of it) in various so-called "derivations" of quantum theory from reasonable axioms [2, 28, 33-35, 87]. Roughly speaking, in these derivations the authors consider reasonable (or physically motivated) constraints on the systems of general probabilistic theories, much the same as the types of systems we described in Chapter 2. It is then shown that some small set of these constraints is sufficient to single out quantum theory. By "single out" we mean that in any theory which conforms to these constraints, the systems represent finite-dimensional quantum systems, and individual systems combine to form composite systems in exactly the manner proscribed by the tensor product of their respective Hilbert spaces.

Reversible transitivity is one such constraint that applies to systems of general probabilistic theories, and tends to play a crucial role in the process of singling out quantum theory. For example, if reversible transitivity is assumed, then every

system has a well-defined maximally mixed state analogous to the quantum maximally mixed state. To show this, it is often first argued on physical grounds that the set of reversible transformations on any system forms a compact, topological group $\mathcal{T}$, which therefore admits a Haar measure. This allows one to define on any system a unique state,

$$\omega = \int_{\mathcal{T}} T(s)\, \mathrm{d}\mu(T), \tag{5.1}$$

for an arbitrary choice of pure state s (note that ω is independent from s, again due to reversible transitivity). The maximally mixed state ω is equivalent in various ways to the well-known quantum maximally mixed state $\frac{1}{n}\mathbb{I}$ in a Hilbert space of dimension n: ω is invariant under reversible transformations, ω gives a non-zero outcome probability for every non-zero effect, and no state s is perfectly distinguishable from ω in a single experiment.

We have stated that reversible transitivity is a useful constraint, but have not yet shown why it might be a reasonable one. To motivate reversible transitivity on physical grounds, one might appeal to the notion of conservation of information [28, 33], i.e. that information is ultimately neither lost nor gained in the course of any physical process. If a system A is in a pure state s_A, then it is always possible to transform A into some other pure state t_A: if necessary, one can perform a "throw away and replace" operation, in which every state in A is irreversibly mapped to t_A. However, conservation of information stipulates that -- in order for information not to be lost -- the system A must be coupled with the environment E in such a way that the global transformation on AE is reversible. Before the transformation, the global system AE is in some state s_{AE} whose marginal on system A is s_A; after the transformation, the global system AE is in state t_{AE}, whose marginal on system A is t_A. Therefore the state of system A may be transformed from an arbitrary pure state s_A to any other pure state t_A in a manner which is ultimately reversible.

It must be conceded that this does not quite grant us a reason to expect reversible transitivity in the form we have defined it, since it does not guarantee that

a reversible transformation exists on system A *alone* which maps state s to state t. In fact, the author is not aware of a conclusive, natural reason to *expect* that the principle of reversible transitivity is obeyed by the universe; despite this, the principle of reversible transitivity holds a far greater immediate appeal than the standard Hilbert-space quantum postulates. Moreover, the conservation of information does at least suggest that it is interesting to study *which* reversible transformations are permitted in general probabilistic theories, since it implies that all transformations must be reversible as long as the system is taken to be "large enough". A demand for reversibility *on some level* would also open up links to the field of thermodynamics, in which closed systems undergo reversible processes (the links between quantum information theory, quantum foundations and thermodynamics is a recent area of investigation, for example see [88-90]).

As demonstrated by Proposition 7 in Section 2.5.5, the property of reversible transitivity does not hold in Boxworld. Proposition 7 shows that reversible transformations map pure product states to pure product states: therefore, a pure product state cannot be reversibly mapped to a pure entangled state (such as the PR-box state). However, this result does not give us a full *characterisation* of reversible Boxworld transformations: entanglement cannot be reversibly generated, but do there exist interesting reversible dynamics that begin and end with entangled states? Recently, it has been shown that, for a restricted class of composite Boxworld systems, the set of reversible transformations does not allow for *any* interesting dynamics [4]. Specifically, for composite systems in which all subsystems have the same number of local fiducial measurements, and the same number of outcomes for each measurement, the only allowed reversible transformations are compositions of relabellings of measurement choices and outcomes, and permutations of subsystems. These transformations can be described easily via the outcome distribution representing the state. For example, the transformation of a system comprising two g-bit subsystems, which permutes the two subsystems

then switches the outputs of subsystem 2, is described by:

$$P\left(a_1, a_2|x_1, x_2\right) \rightarrow P\left(a_2, a_1 \oplus 1|x_2, x_1\right). \tag{5.2}$$

These transformations are trivially convex-linear, i.e. they respect probabilistic mixtures of states, and they map allowed states to allowed states in a reversible manner. That they are the *only* possible transformations in the Boxworld systems considered is surprising, and suggests a powerful link between a theory's state space and its set of reversible transformations. What is it about these systems that reduces the reversible transformations to trivial relabellings? Is it related to the polytopic nature of the state space, its symmetries, or simply the fact that so many states are allowed? By developing techniques to extend this result to the case of non-identical subsystems, we can hope to shed some light on the answers to these questions, and to learn something about the significance of the reversible transitivity which quantum theory enjoys.

We begin in Section 5.1 by recapping the method of proof used in [4] to demonstrate that all reversible dynamics are trivial in the case that all subsystems have the same number of measurements, each with the same number of outcomes. In Section 5.2 we introduce and develop a "tabular" representation of Boxworld states and effects, which has previously been used in [5] to explore the structure of Boxworld measurements. By considering the action of reversible transformations in this tabular regime, it is not difficult to recover the bipartite version of the generalisation of [4] to non-identical subsystems. By considering these tabular representations, one builds up an intuition with which the full generalisation to multipartite systems becomes much more tractable. In Section 5.3 we use this intuition to set up some important results about decompositions of Boxworld effects, then in Section 5.4 we employ these results to obtain Theorem 13, the main result of the chapter. In Section 5.5 we explore the limits of extending this result to other general probabilistic theories: namely, we answer in the negative the question of whether a lack of interesting reversible dynamics is due solely to the polytopic structure of the state space.

Much of the work in this chapter was undertaken collaboratively with Anthony Short and published in [43], although the recognition of the importance of multi-form effects, and the final proof of the result in Section 5.4, are due to the author. The probabilistic model constructed in Section 5.5 is a result of the author's original, unpublished work.

5.1 Identical subsystems: review

Recall from Chapter 2 that an individual Boxworld system is characterised by a finite number of fiducial measurements, each of which has a fixed, finite number of possible outcomes. Boxworld systems combine under the max-tensor product, so that the joint states of composite Boxworld systems are all those non-signaling states which are normalised and produce positive probabilities for all local fiducial effects. In the case where all subsystems are identical, and all fiducial measurements have the same number of outcomes, we can assign constants M and K to be the number of fiducial measurements per subsystem and the number of outcomes per measurement respectively. For each subsystem we will assign a set of $M \cdot K$ fiducial effect vectors $X^{(i)}_{a|x} = X_{a|x}$ of the abstract state space for each subsystem, which in our construction will not depend on i. This set of vectors will correspond to a valid representation of Boxworld as long as the set $\mathcal{U} \cup \{X_{a|x}\}$ (where $1 \leq x \leq M$ and $1 \leq a \leq K-1$) is linearly independent, and for each x, $\sum_a X_{a|x} = \mathcal{U}$.

In Boxworld, the composite effect cone is generated by tensor products of the extreme rays of the local effect cones. These extreme rays are the local fiducial effect vectors, hence the composite fiducial effects e are the tensor products of local fiducial effects:

$$e = X_{a_1|x_1} \otimes \cdots \otimes X_{a_N|x_N}. \tag{5.3}$$

Recall from Proposition 6 of Section 2.5.4 that reversible adjoint transformations must map extreme rays of the effect cone to other extreme rays. Each reversible transformation is therefore a permutation of the set of composite fiducial effects.

Since each fiducial effect may be described as a "string" of local fiducial effects, it is convenient to consider a one-to-one correspondence between composite fiducial effects and strings of fiducial effects, such that the above composite effect corresponds to the *fiducial effect string*,

$$\tilde{e} = (X_{a_1|x_1}, \ldots, X_{a_n|x_N}). \tag{5.4}$$

The ith component of $\tilde{e}$ is therefore $\tilde{e}_i = X_{a_i|x_i}$. This notation allows for a convenient notion of distance between fiducial effect strings.

Definition. *For fiducial effect strings $\tilde{e}$ and $\tilde{f}$, the **Hamming distance** between $\tilde{e}$ and $\tilde{f}$ is defined to be the number of components in which they differ, i.e.*

$$d_H(\tilde{e}, \tilde{f}) = \#\{i \in [N] : \tilde{e}_i \neq \tilde{f}_i\}. \tag{5.5}$$

With these comments in hand, the proof proceeds via the following steps:

- The local fiducial effect vectors are explicitly constructed in a vector space V so that any reversible adjoint transformation $T^\dagger$ corresponds to an orthogonal matrix on the composite tensor product space.
- The inner product between *composite* fiducial effects is then the product of the component-wise inner products between *local* fiducial effects, and is preserved by $T^\dagger$.
- For any pair of composite fiducial effects who differ in exactly one component, their images under $T^\dagger$ similarly differ in exactly one component.
- The permutation induced by $T^\dagger$ on the fiducial effect strings thus preserves Hamming distance 1. A combinatorial theorem implies that any such permutation must be composed of permutations of components, and local permutations.

- $T^\dagger$ therefore acts on the composite fiducial effects in an identical manner, thus it is trivial, in the sense of being composed of permutations of subsystems and local permutations. If $T^\dagger$ is trivial for this particular choice of the fiducial effect vectors, then $T^\dagger$ is trivial as a mapping of the composite outcome distributions $P\left(a_1, \ldots, a_N | x_1, \ldots, x_N\right)$.

We now proceed through these steps in more detail, closely following the argument given in [4].

5.1.1 Fiducial effect vectors

We now construct explicit real vectors for the fiducial effects $X_{a_i|x_i}$ of a Box-world system that has M fiducial measurement choices and K outcomes for each fiducial measurement. We first construct, for each fiducial measurement choice, a $(K-1)$-simplex whose K vertices loosely represent the K outcomes of that measurement, then form a direct sum of the spaces containing these simplices in order to represent the full outcome set. We add to this direct sum a one-dimensional space to represent the unit effect, and ensure that each fiducial effect overlaps symmetrically with this extra dimension so that summing over the effects of any one measurement gives the unit effect.

A $(K-1)$-simplex is formed of K vertices in $\mathbb{R}^{K-1}$, whose positions may be defined as a set of vectors $\{v_1, \ldots, v_K\}$ equidistant from the origin, with constant pairwise inner products. To construct such a set of vectors, consider the standard basis vectors e_i of $\mathbb{R}^K$, and define $c = \frac{1}{K}\sum_1^K e_i$ to be their barycentre. Then using the fact that $\langle e_i, c\rangle = \langle c, c\rangle = \frac{1}{K}$ for all i it is straightforward to check that the K

vectors defined by $v_i = (e_i - c)$ satisfy the following equalities,

$$\langle v_i, v_j \rangle = \begin{cases} \frac{K-1}{K} & (i = j) \\ -\frac{1}{K} & (i \neq j) \end{cases} \tag{5.6}$$

$$\langle v_i, c \rangle = 0 \tag{5.7}$$

$$\sum_{i=1}^{K} v_i = 0. \tag{5.8}$$

The hyperplane orthogonal to c contains the v_i and is isometrically isomorphic to $\mathbb{R}^{K-1}$, hence our desired set of vectors exists also in a space of one less dimension, and moreover $\{v_1, \ldots, v_{K-1}\}$ is a basis for $\mathbb{R}^{K-1}$. Note that for all $0 \leq j, k \leq K - 1$,

$$\begin{aligned} \langle v_j | \left(\sum_{i=1}^{K} |v_i\rangle\langle v_i| \right) |v_k\rangle &= \begin{cases} \frac{K-1}{K} & (j = k) \\ -\frac{1}{K} & (j \neq k) \end{cases} \\ &= \langle v_j, v_k \rangle. \end{aligned} \tag{5.9}$$

Since the v_i span $\mathbb{R}^{K-1}$, it follows that

$$\sum_{i=1}^{K} |v_i\rangle\langle v_i| = \mathbb{I}. \tag{5.10}$$

Let $\{e_i\}$ be the standard basis of the vector space $\mathbb{R}^M$. We associate each measurement x with the basis vector e_x, so that the simplex corresponding to that measurement is embedded in a separate vector space $e_x \otimes \mathbb{R}^{K-1}$. We wish to construct the local fiducial effect vectors in the space $\left(\mathbb{R}^M \otimes \mathbb{R}^{K-1}\right) \oplus \mathbb{R}$, and associate the rightmost $\mathbb{R}$ with the unit effect by setting $\mathcal{U} = (0 \otimes 0) \oplus 1$. The local fiducial effect vectors are:

$$X_{a|x} = \left(\sqrt{\frac{M}{K}}\, e_x \otimes v_a \right) + \left(\frac{1}{K} \mathcal{U} \right). \tag{5.11}$$

From (5.8) it can be seen that for each x, we have $\sum_a X_{a|x} = \mathcal{U}$ as required. Moreover, since $\{e_i\}$ and $\{v_1, \ldots, v_{K-1}\}$ are linearly independent sets, so is the set $\mathcal{U} \cup \{X_{a|x}\}$ for $1 \leq x \leq M$ and $1 \leq a \leq K-1$. Therefore this choice of fiducial effect vectors is a valid representation of a single Boxworld system. From (5.10) it also holds that,

$$\begin{aligned}\sum_{a,x} |X_{a|x}\rangle\langle X_{a|x}| &= \frac{M}{K} \sum_x \left(|e_x\rangle\langle e_x| \otimes \left(\sum_a |v_a\rangle\langle v_a| \right) \right) + \frac{1}{K^2} \sum_{a,x} |\mathcal{U}\rangle\langle\mathcal{U}| \\ &= \frac{M}{K} \left(\sum_{a,x} |e_x\rangle\langle e_x| \otimes |v_a\rangle\langle v_a| + |\mathcal{U}\rangle\langle\mathcal{U}| \right) \\ &= \frac{M}{K} \mathbb{I}. \end{aligned} \tag{5.12}$$

Similarly, for the composite fiducial effect vectors,

$$\sum_{\mathbf{a},\mathbf{x}} |X_{a_1|x_1}\rangle\langle X_{a_1|x_1}| \otimes \cdots \otimes |X_{a_N|x_N}\rangle\langle X_{a_N|x_N}| = \left(\frac{M}{K} \right)^N \mathbb{I} \tag{5.13}$$

This immediately implies that any reversible adjoint transformation $T^\dagger$ is orthogonal, since it acts to permute the composite fiducial effects,

$$\begin{aligned} T^\dagger T &= \left(\frac{K}{M} \right)^N T^\dagger \left(\sum_{\mathbf{a},\mathbf{x}} |X_{a_1|x_1}\rangle\langle X_{a_1|x_1}| \otimes \cdots \otimes |X_{a_N|x_N}\rangle\langle X_{a_N|x_N}| \right) T \\ &= \left(\frac{K}{M} \right)^N \left(\sum_{\mathbf{a},\mathbf{x}} |X_{a_1|x_1}\rangle\langle X_{a_1|x_1}| \otimes \cdots \otimes |X_{a_N|x_N}\rangle\langle X_{a_N|x_N}| \right) \\ &= \mathbb{I}. \end{aligned} \tag{5.14}$$

Therefore $T^\dagger$ preserves the inner product between any pair of composite fiducial effects, which is determined by the component-wise inner products between

local fiducial effects:

$$\langle X_{a|x}, X_{a'|x'} \rangle = \frac{1}{K^2} \begin{cases} 1 & (x \neq x') \\ 1 - M & (x = x', a \neq a') \\ 1 + M \cdot (K-1) & (x = x', a = a') \end{cases} \tag{5.15}$$

5.1.2 Induced fiducial effect string permutations

Recall that any reversible adjoint transformation $T^\dagger$ induces a permutation on the fiducial effect strings, i.e. strings whose entries are the local fiducial effects. When M and K are constant across subsystems, and in particular when $M \geq 2$ (ensuring that the subsystems are non-classical), then we can use the vector representation of Section 5.1.1 to show that these permutations are highly restricted in form. The assumption that $M \geq 2$ guarantees that the inner product between any two local fiducial effect vectors is non-zero, and is strictly negative exactly when they correspond to different outcomes of the same measurement.

Lemma 5. *On a composite Boxworld system with $M \geq 2$ local measurements and K local outcomes, suppose that T is a reversible transformation, with an induced permutation $\tilde{T^\dagger}$ on fiducial effect strings. Then $\tilde{T^\dagger}$ preserves a Hamming distance of 1 between pairs of fiducial effect strings, i.e., for any composite fiducial effects e and f with corresponding fiducial effect strings $\tilde{e}$, $\tilde{f}$,*

$$d_H(\tilde{e}, \tilde{f}) = 1 \implies d_H(\tilde{T^\dagger}(\tilde{e}), \tilde{T^\dagger}(\tilde{f})) = 1 \tag{5.16}$$

Proof. Suppose $\tilde{e} = \left(X_{a_1|x_1}, \ldots, X_{a_N|x_N}\right)$ and $\tilde{f} = \left(X_{a'_1|x'_1}, \ldots, X_{a'_N|x'_N}\right)$ are two fiducial effect strings that differ only in component j. The inner product between the composite fiducial effects e and f is the product of the component-wise inner products between local fiducial effects:

$$\langle e, f \rangle = \prod_{i=1}^{N} \langle X_{a_i|x_i}, X_{a'_i|x'_i} \rangle. \tag{5.17}$$

From (5.15), the greatest value that $K^{2n}\langle e, f\rangle$ can take is $(1 + M(K - 1))^N$, which occurs when $a_i = a'_i$ and $x_i = x'_i$ for all i. Since the only negative term in (5.15) is $1 - M$ (when $x = x', a \neq a'$), the least value that can be taken is $(1 - M)(1 + M(K - 1))^{N-1}$. As $M \geq 2$, this occurs if and only if e and f differ in one component, on which they represent different outcomes of the same measurement. In this case, since T is orthogonal,

$$K^{2n}\langle T^\dagger(e), T^\dagger(f)\rangle = (1 - M)(1 + M(K - 1))^{N-1}, \tag{5.18}$$

therefore $\tilde{T^\dagger}(\tilde{e})$ and $\tilde{T^\dagger}(\tilde{f})$ must also differ in exactly one component.

It remains to show that if $\tilde{e}$ and $\tilde{f}$ differ in one component, on which they represent outcomes of *different* measurements, i.e. $x_j \neq x'_j$, then again $\tilde{T^\dagger}(\tilde{e})$ and $\tilde{T^\dagger}(\tilde{f})$ differ in exactly one component. In this case

$$K^{2n}\langle e, f\rangle = K^{2n}\langle T^\dagger(e), T^\dagger(f)\rangle = (1 + M \cdot (K - 1))^{N-1}. \tag{5.19}$$

This alone does not suffice to show that $T^\dagger(e)$ differs in only one component from $T^\dagger(f)$, since for example it may be that $(1 - M)^2 = (1 + M \cdot (K - 1))$. However, consider any local pure state $s^{(j)}$ on subsystem j which gives outcome a_j for measurement x_j and outcome a'_j for measurement x'_j (outcomes for the remaining measurement choices may be assigned arbitrarily). For the remaining subsystems with $i \neq j$, let $s^{(i)}$ be any pure state which gives outcome a_i for measurement x_i. Then the pure product state $s = s^{(1)} \otimes \cdots \otimes s^{(N)}$ is constructed such that $\langle s, e\rangle = \langle s, f\rangle = 1$, hence $e + f \not\leq_{\mathcal{E}_+} \mathcal{U}_N$.

Recall from Section 2.5.4 that a reversible adjoint transformation $T^\dagger$ maps $\mathcal{E}_+^{max}$ linearly and bijectively to itself such that $T^\dagger(\mathcal{U}_N) = \mathcal{U}_N$; hence $T^\dagger(e) + T^\dagger(f) \not\leq_{\mathcal{E}_+} \mathcal{U}_N$. It follows that for at least one component k, the local fiducial effects $\tilde{T^\dagger}(\tilde{e})_k$ and $\tilde{T^\dagger}(\tilde{f})_k$ correspond to outcomes of different measurements. This fact, combined with (5.19), implies that $\tilde{T^\dagger}(\tilde{e})$ differs from $\tilde{T^\dagger}(\tilde{f})$ in exactly one component. □

We now apply a general combinatorial theorem concerning permutations of

strings. We will consider strings belonging to some Cartesian product of finite alphabets $\mathcal{A}_1 \times \cdots \times \mathcal{A}_N$; in this Section each alphabet will comprise the $M \cdot K$ distinct local fiducial effect vectors and will therefore be identical, however we state and prove the theorem without the assumption of identical alphabets, as this will become useful when we apply the theorem in the case that not all subsystems are identical. In order to prove the theorem, it is convenient to make rigorous our notion of local operations and permutations of components, at least in regards to fiducial effect strings.

A permutation Q of $\mathcal{A}_1 \times \cdots \times \mathcal{A}_N$ is a *local operation* if there is some permutation Q_i of the set $\mathcal{A}_i$ such that

$$Q : (a_1, \ldots, a_i, \ldots, a_N) \mapsto (a_1, \ldots, Q_i(a_i), \ldots, a_N). \tag{5.20}$$

Q is a *component permutation* if there is some permutation σ of the set $[N]$ such that

$$Q : (a_1, \ldots, a_N) \mapsto (a_{\sigma(1)}, \ldots, a_{\sigma(N)}). \tag{5.21}$$

Note that if $i, j \in [N]$ such that $\sigma : i \mapsto j$, then Q is a valid permutation only if $\mathcal{A}_i = \mathcal{A}_j$.

We may as well assume for the time being that an alphabet of size n is simply the set $\{0, \ldots, n-1\}$. Denote by $\mathbf{0}$ the string $(0, \ldots, 0)$, and define the *Hamming weight* of a string $\mathbf{a}$ to be the number of non-zero components, i.e. $W_H(\mathbf{a}) = d_H(\mathbf{a}, \mathbf{0})$.

Theorem 10. *Let $\mathcal{A}_1, \ldots, \mathcal{A}_N$ be finite alphabets, and Q be a permutation of the set $\mathcal{A}_1 \times \cdots \times \mathcal{A}_N$. If Q preserves a Hamming distance of 1 between pairs of strings, then it is a composition of operations which permute components, and operations which act independently on each component (local operations).*

Proof. Let $\mathcal{A}_i = \{0, \ldots, n_i - 1\}$. We may assume (by pre-composing Q with local operations if necessary) that Q maps the string $\mathbf{0}$ to itself. By assumption, Q acts as a permutation on the set of strings with exactly one non-zero component. Let L_i denote the set of strings which are non-zero only in the ith component: the

components of each L_i are at Hamming distance 1 from each other, so Q maps each L_i to some L_j, and moreover maps L_i to L_j only if $n_i = n_j$.

Suppose that we pre-compose Q with the component permutation which maps component j to component i just when Q maps L_i to L_j, to obtain a new map Q_1 which is maps each set L_i to itself. Suppose further that we pre-compose Q_1 with a local operation for every component i, which acts as the inverse of Q_1 on the set L_i, to obtain a new map Q_2 which therefore fixes each string of Hamming weight 1. Note that if Q_2 can be written as a combination of local operations and component permutations then so can Q, hence we may assume without loss of generality that Q fixes strings of Hamming weight 1.

Given that Q fixes strings of Hamming weight 1, we show by induction that it fixes all strings. Suppose that it fixes strings of weight less than W, and let **a** be a string of weight W. However, **a** is uniquely specified by all those strings of Hamming distance 1 from it, that have weight $W-1$. Since those strings are fixed by Q, so too must **a** be. Hence, Q fixes strings of weight W and by induction, all strings. □

5.1.3 Characterisation of reversible transformations

It follows from Sections 5.1.1 and 5.1.2 that reversible adjoint transformations permute the composite fiducial effect vectors in a way which is a composition of local subsystem operations and permutations of subsystems. Thus the transformation itself acts in a similar manner on states: it is a composition of local subsystem operations and permutations of subsystems. Since the systems are identical, we will see that any permutation of subsystems is a valid reversible transformation. However, local operations are further restricted by the following Lemma which, importantly, applies in the general Boxworld scenario, without assuming that all measurements have the same number of outcomes.

Lemma 6. *The only allowed reversible transformations of a single boxworld system are relabellings of measurement choices and measurement outcomes.*

Proof. Let T be a reversible transformation, so that the adjoint $T^\dagger$ permutes the local fiducial effect vectors $X_{a|x}$. Suppose that the system has M measurements, and that measurement x has K_x outcomes. If $x \neq x'$, then for any choice of $1 \leq a \leq K_x$ and $1 \leq a' \leq K_{x'}$ there exists a pure state s which gives outcome a when x is measured and outcome a' when x' is measured. This state obeys $\langle X_{a|x}, s\rangle = \langle X_{a'|x'}, s\rangle = 1$, hence $X_{a|x} + X_{a'|x'} \not\leq_{\mathcal{E}_+} \mathcal{U}_N$. However, for any $1 \leq a, a' \leq K_x$ we have that $X_{a|x} + X_{a'|x} \leq_{\mathcal{E}_+} \mathcal{U}_N$, since $\sum_a X_{a|x} = \mathcal{U}$, and any conic combination of the fiducial effect vectors is a member of $\mathcal{E}_+$. Now,

$$X_{a|x} + X_{a'|x} \leq_{\mathcal{E}_+} \mathcal{U}_N \iff T^\dagger(X_{a|x}) + T^\dagger(X_{a'|x}) \leq_{\mathcal{E}_+} \mathcal{U}_N, \tag{5.22}$$

hence two fiducial effect vectors correspond to outcomes of the same measurement x only if their images under $T^\dagger$ also correspond to outcomes of the same measurement x'. Therefore the permutation is a composition of permutations of the form:

$$X_{a|x} \mapsto X_{a'|x} \tag{5.23}$$

corresponding to relabelling of outcomes, and permutations of the form:

$$X_{a|x} \mapsto X_{a|x'} \tag{5.24}$$

corresponding to relabelling of measurement choices. Note that one measurement choice may be mapped onto another only if they have the same number of outcomes. □

Combining the results of the previous Sections leads to the main result of [4].

Theorem 11. *In a composite Boxworld system, with $M \geq 2$ local measurements for each subsystem and K local outcomes for each measurement, the only reversible transformations are permutations of subsystems, and local relabellings of measurement choices and outcomes. Furthermore, all such transformations are allowed.*

Proof. The fact that these are the only possible reversible transformations follows from Sections 5.1.1 and 5.1.2, and Lemma 6. It remains to show that any such transformation is allowed. This follows easily from viewing these transformations acting on the outcome distributions representing states. For example, permuting subsystems 1 and 2 is given by the mapping:

$$P\left(a_1, a_2, \ldots, a_N | x_1, x_2, \ldots, x_N\right) \mapsto P\left(a_2, a_1, \ldots, a_N | x_2, x_1, \ldots, x_N\right), \quad (5.25)$$

and relabelling the outcomes of measurement x_1 on subsystem 1 via the permutation σ of the set $[K]$ is given by the mapping:

$$P\left(a_1, \ldots, a_N | x_1, \ldots, x_N\right) \mapsto P\left(\sigma(a_1), \ldots, a_N | x_1, \ldots, x_N\right). \quad (5.26)$$

All of these mappings map allowed states to allowed states in a reversible and convex-linear fashion, hence correspond to allowed reversible transformations. □

We conclude this Section with some remarks about attempting a direct generalization of this result to composite systems with differing numbers of measurements on subsystems, and differing numbers of outcomes per measurement. A natural first step would be to relax the demand that M and K are constant across subsystems, but retain the demand that each local fiducial measurement has the same number of outcomes for a given subsystem. The local fiducial effect vectors for subsystem i can still be constructed as in Section 5.1.1, given that there are $M^{(i)}$ fiducial measurement choices, each of which has $K^{(i)}$ possible outcomes. The inner product between any pair of vectors is still given by (5.15) -- but with M and K replaced by $M^{(i)}$ and $K^{(i)}$ -- and composite fiducial effects are still given by tensor products of the local fiducial effects. The proof that reversible transformations are orthogonal carries through, and so it seems worth examining whether a consideration of inner products will again demonstrate that a Hamming distance of 1 is preserved. If so, due to the fact that Theorem 10 applies to arbitrarily-sized alphabets, it follows that all reversible transformations are compositions of local

operations and subsystem permutations.

Unfortunately however, the proof of Lemma 5 does not carry through so nicely if M and K are allowed to vary across subsystems. A simple counterexample is provided by a composite system comprising 3 subsystems where the number of measurement outcomes $K^{(1)} = K^{(2)} = K^{(3)} = 2$ is constant but $M^{(1)} = 2$, $M^{(2)} = 2$ and $M^{(3)} = 10$. In this case, for any two effects whose components agree except for the third subsystem, where they correspond to outcomes of different measurements, the inner product will be:

$$\frac{(1 + M^{(1)} \cdot (K^{(1)} - 1))(1 + M^{(2)} \cdot (K^{(2)} - 1))}{(K^{(1)} \cdot K^{(2)} \cdot K^{(3)})^2} = \frac{3 \times 3}{12} = \frac{3}{4}. \tag{5.27}$$

On the other hand, a pair of effects with different outcomes of the same measurement on components 1 and 3, and outcomes of different measurements on component 2, will generate the same inner product:

$$\frac{(1 - M^{(1)})(1 - M^{(3)})}{(K^{(1)} \cdot K^{(2)} \cdot K^{(3)})^2} = \frac{(-1) \times (-9)}{12} = \frac{3}{4}. \tag{5.28}$$

Any reversible transformation which maps a pair of effects of the first kind to a pair of the second kind would necessarily be non-trivial, since it would not preserve the Hamming distance. However, as we prove later in this chapter, it is impossible to extend this transformation in a consistent, linear manner to a full permutation of the composite fiducial effect vectors; this system does not in fact allow for non-trivial reversible transformations. It may be observed that there is a lot of play in the variables $M^{(i)}$ and $K^{(i)}$ in attempting to generate these "pathological" systems, and that as the number of subsystems increases there will certainly be infinitely many pathological systems (for example, combining any number of copies of the above system will result in a system which will have many more overlaps in the inner product). It does not appear to be a promising task, therefore, to simply go through and check each case where Lemma 5 might go wrong.

A natural second attempt to extend the proof is to embed an arbitrary Box-

world system into one which is of the identical-subsystem form considered in this Section, in the hope that the smaller system will inherit its reversible transformations from the larger. Suppose, for example, that it was possible to "add in" a new measurement with just two outcomes on any one subsystem, in such a way that the reversible transformations of the old system are induced by those of the new. By repeatedly performing this procedure, one can arrive at a system for which subsystems have the same number of measurements. Suppose also that it was possible to also "add in" outcomes to measurements in a similar way, so that one may arrive at a system for which M and K are both constant across subsystems. Having characterised the reversible transformations of this larger system, we would perhaps be able to infer from this procedure that the reversible transformations of the smaller system are trivial, since they are inherited from the larger system.

However, this embedding procedure does not proceed as straightforwardly as hoped. Consider for example a system with any number of subsystems, measurements and outcomes, and suppose we added in an extra outcome to the first fiducial measurement of the first subsystem (which previously had K outcomes, say). This generates a set of new composite fiducial effects, whose first component is $X^{(1)}_{K+1|1}$. For any reversible transformation on the larger system including this set of effects, we might attempt to induce a transformation on the smaller system by simply restricting the permutation of fiducial effect strings to those which do not have $X^{(1)}_{K+1|1}$ in the first component. However, two problems are encountered: firstly, the transformation may map fiducial effects *not* in the above set to those which are; secondly, the induced permutation is unlikely to respect the linearities of the smaller set of fiducial effects, such as

$$\sum_{a=1}^{K_1^{(1)}} X^{(1)}_{a|1} \otimes \cdots \otimes X^{(N)}_{a_N|x_N} = \sum_{a=1}^{K_2^{(1)}} X^{(1)}_{a|2} \otimes \cdots \otimes X^{(N)}_{a_N|x_N} \tag{5.29}$$

Given these difficulties, it seems that a fresh approach is needed to demonstrate that reversible transformations preserve Hamming distance 1 between pairs of effects. Note that proving this in the general case is sufficient to generalise the fact

that all reversible transformations are trivial, since the remaining steps of the proof do not rely on the number of measurements and outcomes being constant across subsystems. In the rest of this Chapter we develop a proof of this property of reversible transformations, which is not based on considering specific orthogonal representations of Boxworld systems as before, but on the basic convex structure of the state and effect spaces.

It should be noted that our demands concerning classical systems must be adapted somewhat: in the case where M is constant, simply demanding $M \geq 2$ ensures that no subsystem is classical. However, if M is allowed to vary then we also need to rule out the possibility that some subsystems are classical and others are not. In the remainder of this Chapter we will demand that $M^{(i)} \geq 2$ for all i, i.e. that all subsystems are non-classical. If this were not the case, then non-trivial reversible transformations *do* in general exist. As a simple example, consider a bipartite system with $M^{(1)} = 1, K_1^{(1)} = 2$ so that subsystem 1 is classical, and $M^{(2)} = 2, K_1^{(2)} = K_2^{(2)} = 2$ so that subsystem 2 is a g-bit system. The reversible "C-NOT" operation on this system is defined by:

$$X_{1|1}^{(1)} \otimes X_{a|x}^{(2)} \rightarrow X_{1|1}^{(1)} \otimes X_{a|x}^{(2)} \tag{5.30}$$

$$X_{2|1}^{(1)} \otimes X_{a|x}^{(2)} \rightarrow X_{2|1}^{(1)} \otimes X_{a\oplus 1|x}^{(2)}. \tag{5.31}$$

To check that this transformation is indeed linear, one need only verify that the unit effect $\mathcal{U}^{(1)} \otimes \mathcal{U}^{(2)}$ is mapped to itself, and that effects of the form $\mathcal{U}^{(1)} \otimes X_{a|x}^{(2)}$ and $X_{a|x}^{(1)} \otimes \mathcal{U}^{(2)}$ have a well-defined image, independent of how the local unit effect is decomposed on each subsystem. This turns out to be true because there is only a single decomposition of the unit effect of the classical subsystem into local fiducial effects. Since the operation is a linear permutation of composite fiducial effects, it is an allowed transformation. A similar example can be constructed for any system with at least one classical subsystem, in which a local operation is performed, conditioned on the outcome obtained on the classical subsystem.

5.2 Effect tables

Recall from Section 2.3.3, in which we first introduced Boxworld, that under the assumption of local tomography, a composite Boxworld state can be represented as a multi-dimensional tabular array. This idea was first introduced and discussed in [5] in the context of Boxworld measurements; we will develop our own investigation of this idea, as it provides a neat diagrammatic portrayal of many of the results we will prove before achieving the main result. Our development of these tabular representations is original work that has not been published. The number of dimensions or axes is the number of subsystems, the rows in a given axis correspond to outcomes of measurements on that subsystem, and each entry gives the probability of obtaining the outcomes corresponding to the row containing that entry for each respective subsystem. For example, the PR-box state of two g-bits, in which outcomes are randomly correlated as long as the local fiducial measurement choices are not both equal to 1, and randomly anti-correlated otherwise, has the following tabular representation:

		$x_2 = 1$		$x_2 = 2$	
		$a_2 = 1$	$a_2 = 2$	$a_2 = 1$	$a_2 = 2$
$x_1 = 1$	$a_1 = 1$	1/2	0	1/2	0
	$a_1 = 2$	0	1/2	0	1/2
$x_1 = 2$	$a_1 = 1$	1/2	0	0	1/2
	$a_1 = 2$	0	1/2	1/2	0

(5.32)

We will refer to the subset of entries corresponding to the outcomes of only a single measurement choice from each subsystem as a *block* of the tabular array. Diagrammatically, a block is a set of entries enclosed by lines that separate local measurement choices. For example, by specifying the second measurement on both subsystems in (5.32), we get the bottom-right block:

0	1/2
1/2	0

(5.33)

From hereon in we will omit for convenience the first rows and columns which label the local measurements and outcomes, since these can be deduced from the number of blocks and rows per block, and include only the outcome probabilities, *viz.*

1/2	0	1/2	0
0	1/2	0	1/2
1/2	0	0	1/2
0	1/2	1/2	0

(5.34)

A pure product state is represented by a table whose only non-zero entries are 1. For each measurement choice on each axis, a particular row is selected (the outcome determined by the measurement choice) -- the 1s in the table occur exactly where these selected rows overlap. For example, the following table represents a state of two subsystems with three fiducial measurements each, where in the first subsystem the state assigns outcome 1 for the first measurement and 2 for the second and third measurements, and in the second subsystem the state assigns the outcome equal in value to the measurement choice.

1	0	0	0	1	0	0	1	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	1	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	1	0

(5.35)

Composite effects can also be represented using tables, which we will refer to as *effect tables*, so that the probability of obtaining that effect is given by the entry-wise dot product between its table and the table representing the state of

the system. The composite fiducial effects, which are tensor products of local fiducial effects, have only a single 1 in the entry corresponding to the measurement outcome specified on each subsystem. General composite effects, which are conic combinations of the composite fiducial effects, have effect tables that are non-negative in every entry (a table can correspond to a valid effect even if some entries are negative, but in this case there always exists another table with all positive entries, which corresponds to the same effect).

The tabular representation of effects is over-complete, in the sense that some effects may correspond to more than one possible table. This over-completeness is codified by a set of linearities which are imposed on the set of effect tables, due to the set of non-signaling conditions on the system. For example, in a two g-bit system, the condition that the reduced outcome probability of obtaining outcome 0 for measurement 0 on subsystem 2 is independent of the measurement choice on subsystem 1 is given by the equation:

$$P\left(0,0|0,0\right)+P\left(0,1|0,0\right)=P\left(0,0|0,1\right)+P\left(0,1|0,1\right). \tag{5.36}$$

This equation is reflected by demanding the following equivalence between effect tables:

$$\left[\begin{array}{cc|cc} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \equiv \left[\begin{array}{cc|cc} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{5.37}$$

Both of these tables are representations of the effect $\mathcal{U}^{(1)}\otimes X_{0|0}^{(2)}$; demanding that they are equivalent is no more than demanding consistency with the equation,

$$X_{0|0}^{(1)}\otimes X_{0|0}^{(2)}+X_{1|0}^{(1)}\otimes X_{0|0}^{(2)}=X_{0|01}^{(1)}\otimes X_{0|0}^{(2)}+X_{1|1}^{(1)}\otimes X_{0|0}^{(2)}, \tag{5.38}$$

where each side corresponds to a distinct decomposition of the local unit effect $\mathcal{U}^{(1)}$ into local fiducial effects. In this manner, the linearities inherent in the local effect cone of the abstract state space automatically engender the single-system non-signaling conditions which are respected by the set of composite states $\mathcal{S}^{max}$. By Proposition 1, we know that the full set of non-signaling conditions is generated by those non-signaling conditions involving single subsystems. Hence, the linearities of the local effect cone give rise to the full set of non-signaling conditions on the composite state space.

The effect tables which interest us the most are the *binary effect tables*, those consisting of 0s and 1s, which represent effects that are sums of composite fiducial effects (with all coefficients equal to 1). A binary effect table is said to be *multiform* if it is equivalent to some other binary effect table, as in (5.37). We will use the term *sub-row* to refer to binary effect tables which are similar to the effect tables in (5.37) in that their non-zero entries lie only along one row of their constituent blocks. As long as every subsystem is non-classical, then every axis splits up into at least 2 blocks, so that sub-rows (and any tables covering a sub-row) are always multiform. Moreover, each equivalence between sub-rows is due to a decomposition of the local unit effect on the corresponding subsystem, as in (5.38). We will say that a binary effect table *covers* another binary effect table if the non-zero entries of the latter are a subset of the non-zero entries of the former: clearly, any effect table which covers a sub-row is again multiform, for example:

$$\left[\begin{array}{cc|cc} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \equiv \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right] \tag{5.39}$$

In order to check whether two effect tables are equivalent, it suffices to show that they have the same dot product with all pure product states. This is because the pure product states span the vector space of the composite abstract state space [29], hence any two effects which the same dot product with all of them must

be identical. In cases where both tables cover sub-rows, it may be possible to determine an equivalence between binary effect tables by applying various sub-row equivalences. This is achieved by "sliding" a sub-row across to another sub-row contained in the same row. For example, in (5.40) below, the two horizontal sub-rows of the top left block can be slid across to the right, and then the two vertical sub-rows slid downwards to form the effect table on the RHS.

$$\left[\begin{array}{cc|cc} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \equiv \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right] \tag{5.40}$$

However this sliding process will not always be a useful indicator of equivalence, as demonstrated in the following example which does not cover any sub-rows:

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right] \equiv \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \tag{5.41}$$

The equivalence (5.41) can be verified by dot product with pure product states. However, a similar consideration demonstrates that this is not in fact an equivalence between *valid* effects. The binary table on the LHS of (5.41) has dot product 2 with the pure product state s that deterministically outputs 0 for every fiducial measurement choice on both subsystems, hence this effect table represents an improper effect. This is illustrated in (5.42), where arrows are used to highlight which rows and columns correspond to the outcomes determined by s; the entries of the effect table which are picked out by s are those lying in the overlap of these rows and columns, and their sum is the dot product of the effect with s, which is 2 in this case. The results described in this chapter will rely on the multiform tables which do correspond to proper effects and give genuine probabilities for all states, and thus examples like (5.41) will not generally present a major obstacle.

$$
\begin{array}{cc}
 & \begin{array}{cc} \downarrow & \end{array} \quad \begin{array}{cc} \downarrow & \end{array} \\
\rightarrow & \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \\
\rightarrow & \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right]
\end{array}
\quad ; \quad 1+0+0+1=2 \tag{5.42}
$$

Some further observations will help to familiarise the reader with effect tables and their properties. As mentioned above, sub-rows are always multiform. In the bipartite g-bit examples given, all sub-rows have the same total number of 1s, however this will not be true if any two local fiducial measurements have differing numbers of outcomes. If in fact there are two fiducial measurements with differing numbers of outcomes K_1 and K_2 on the *same* subsystem, then there will exist a sub-row containing K_1 1s which is equivalent to a sub-row containing K_2 1s, hence equivalent binary effect tables need not have the same number of non-zero entries.

Calculating dot products between binary effect tables and pure product states is an invaluable tool; as we have mentioned, it determines whether two distinct tables correspond to the same effect or not. In bipartite systems, it also allows for a useful characterisation of exactly when a binary effect table corresponds to a proper effect. In (5.42) we see that the presence of non-zero entries in diagonal blocks allows for the construction of a pure product state, which selects the rows and columns necessary to "hit" those two entries, and, if necessary, completes the lattice by selecting arbitrary rows and columns from remaining measurements on each subsystem. Note that the blocks do not have to be on the same diagonal line, merely unaligned (i.e. not in the same block-row or block-column). Below is an example of this kind of improper binary effect table, and one of the six pure product states for which it gives dot product greater than 1.

$$
\begin{array}{cc|cc|cc}
 & \downarrow & & \downarrow & & \downarrow \\
\rightarrow & 1 & 1 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 0 & 1 & 0 & 0 \\
\hline
 & 0 & 0 & 0 & 0 & 0 & 0 \\
\rightarrow & 0 & 0 & 0 & 0 & 1 & 0
\end{array}
\tag{5.43}
$$

A binary effect table also corresponds to an improper effect if it has distinct non-zero entries in the same row, but not within the same block of that row. If this is the case, then a similar pure product state construction can be used to hit both non-zero entries, again with an arbitrary completion of the lattice for that state if necessary. Below is an example of this kind of improper binary effect table, and one of the eight pure product states for which it gives dot product greater than 1.

$$
\begin{array}{cc|cc|cc}
 & \downarrow & & \downarrow & & \downarrow \\
\rightarrow & 1 & 1 & 0 & 0 & 1 & 0 \\
 & 0 & 0 & 0 & 1 & 0 & 0 \\
\hline
 & 0 & 0 & 0 & 0 & 0 & 0 \\
\rightarrow & 0 & 0 & 0 & 0 & 0 & 0
\end{array}
\tag{5.44}
$$

These two constraints are the only obstacles to a binary, bipartite effect table being a proper effect. Indeed, as long as all non-zero entries are contained in the same *row of blocks* (or block-row), but never in *distinct blocks of the same row*, then the effect is proper. To see this, suppose that an effect table corresponding to an effect $E \in \mathcal{E}_+$ satisfies these conditions, and suppose without loss of generality that it has non-zero entries only in the block row corresponding to the first fiducial measurement on subsystem 1 (i.e. the leftmost column of blocks). In each of these blocks, fill in extra 1s if necessary to get an effect $F \geq_{\mathcal{E}_+} E$ such that every non-zero entry lies in a vertical sub-row. Then slide all sub-rows up to the topmost block (the subrows will not clash since no two non-zero entries lie in distinct blocks of the same row) to obtain an effect which is either the unit effect, or a strict subset of it. Now $E \leq_{\mathcal{E}_+} F \leq_{\mathcal{E}_+} \mathcal{U}$, hence E must be a proper effect. This process is

illustrated by the following operations on a binary, bipartite effect table satisfying these conditions:

$$\left(\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccc|cc} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Thus, a binary, bipartite effect table is proper if and only if all its non-zero entries lie in a single block-row, without any two entries lying in distinct blocks of a single row. It is tempting to try to extend these conditions in order to characterise proper binary effect tables of multipartite systems.

Unfortunately however, it is not the case in multipartite Boxworld systems that an improper effect can be "detected" using pure product states in the same way; there exist improper effects E for which $\langle E, s\rangle \in \{0, 1\}$ for all pure product states s. In fact, this occurs even in tripartite systems, as demonstrated by the tripartite effect E depicted in Figure 5.1. In this representation of E, each of the smallest cubes represents an entry of the effect table; shaded cubes represent entries with a 1 in them, whereas unshaded cubes represent zero entries. Assigning subsystems 1, 2 and 3 to the *x*-, *y*- and *z*- axes respectively, it can be seen that the algebraic form of E is,

$$\begin{aligned} E =& X^{(1)}_{1|1} \otimes X^{(2)}_{1|1} \otimes X^{(3)}_{1|1} + X^{(1)}_{2|1} \otimes X^{(2)}_{1|2} \otimes X^{(3)}_{2|2} \\ &+ X^{(1)}_{1|2} \otimes X^{(2)}_{2|1} \otimes X^{(3)}_{1|2} + X^{(1)}_{2|2} \otimes X^{(2)}_{2|2} \otimes X^{(3)}_{2|1}, \end{aligned} \tag{5.45}$$

where the ordering of the summands agrees with the alphabetical ordering of the cubes' labels in Figure 5.1.

Similarly to bipartite systems, specifying a tripartite pure product state corresponds to picking out a (three-dimensional) grid of entries in the effect table. This grid is obtained by choosing one particular outcome for each local fiducial mea-

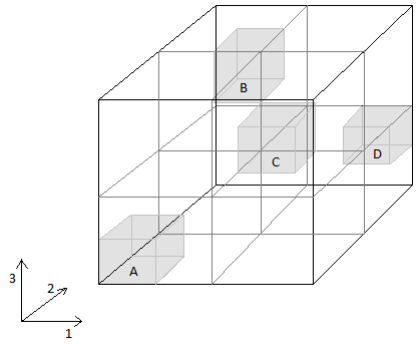

Figure 5.1: Improper tripartite binary effect table which is not detected by pure product states.

surement on each subsystem, and selecting the entries that lie in the intersection of the subsequent rows. For example, the pure product state which gives outcome 1 for every local fiducial measurement will result in a grid that contains the shaded cube **A** in Figure 5.1, but none of the other shaded cubes. By inspection it can be verified that any pure product state will pick out at most one shaded cube, hence $\langle E, s\rangle \in \{0, 1\}$ for all pure product states s: in short, we would not be able to deduce that E is improper by looking at its inner products with pure product states.

We must resort to other methods to prove that E is actually improper. We first argue that if any further cube were to be shaded in Figure 5.1, there would be some pure product state that picks out two of the shaded cubes. To see this, note that whenever two cubes are completely unaligned (i.e. correspond to different measurement choices in *all three* subsystems), or are in the same plane but diagonally opposing blocks (i.e. correspond to the same local effect on *one* subsystem, but different measurement choices on the *remaining* subsystems), then it is possible to construct a pure product state which hits both of them.

In fact, a similar observation shows that if any entry corresponding to an unshaded cube is assigned a positive, non-zero entry, then there will be a pure product state whose inner product with the resulting effect is larger than 1. Moreover, for any shaded cube, setting the corresponding entry of the effect table to be larger

than 1 will generate in an improper effect (one need only consider a pure product state which hits the corresponding fiducial effect). It follows that for any fiducial effect F and positive number ε, the effect $E + \varepsilon F$ is an improper effect. Therefore, either E is actually the unit effect, or else it does not lie in the order interval $[0, \mathcal{U}]_{\mathcal{E}_+}$, as no conic combination of fiducial effects can be added to it to obtain $\mathcal{U}$.

It remains to show simply that E is not the unit effect. This follows from the existence of a pure product state s for which $\langle E, s \rangle = 0$. Let $s^{(1)}$ and $s^{(2)}$ both be equal to the g-bit state which deterministically assigns outcome 2 to both fiducial measurements, and let $s^{(3)}$ be the g-bit state which assigns outcome 1 for the first fiducial measurement and outcome 2 for the second fiducial measurement. It can then be straightforwardly verified that $\langle E, s^{(1)} \otimes s^{(2)} \otimes s^{(3)} \rangle = 0$, either by directly using the decomposition (5.45) or by observing that the grid of entries specified by $s^{(1)} \otimes s^{(2)} \otimes s^{(3)}$ does not hit any shaded cubes in Figure 5.1. Hence E is not the unit effect and is therefore improper, despite the fact that $\langle E, s \rangle \in \{0, 1\}$ for all pure product states s.

5.2.1 Reversible transformations on effect tables

Since reversible adjoint transformations permute the set of composite fiducial effects, they induce a permutation on the entries of binary effect tables, such that entry a is mapped to entry b if and only if the fiducial effect which is represented by a single 1 in entry a is mapped to the fiducial effect which is represented by a single 1 in entry b. Effect tables provide a compact visualisation of the reversible transformations which correspond to relabellings of measurements and outcomes, and permutations of subsystems. These are given by block permutations (along an axis), row permutations and transpose-like operations on tables respectively. We give some examples to illustrate this procedure.

A relabelling of the measurement inputs on subsystem 1 will permute the blocks in line with the first axis, but leave the ordering of the rows within each block invariant. For effects in a bipartite system of two g-bits, switching the measurement inputs on subsystem 1 is pictured below.

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{5.46}$$

A relabelling of the outcomes of a measurement in a subsystem are given by permuting the rows along the axis corresponding to the subsystem, but doing so only within the block that corresponds to the measurement in question. For example, switching the measurement outcomes for the first measurement on the second subsystem of the table on the RHS of (5.46) is given below.

$$\left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{5.47}$$

Permuting the subsystems corresponds to permuting the axes corresponding to those subsystems (equivalently, permuting the coordinates of each entry). In the case of bipartite systems, this is the same as taking the transpose of the table (imagining it were a matrix). Note that the table need not be symmetric, and no two blocks need be the same size, so in general one must be careful to transpose the lines delineating the blocks as well as the entries themselves. If we apply a permutation of the two subsystems to the table on the RHS of (5.47), we obtain the mapping below.

$$\left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{5.48}$$

5.2.2 Diagrammatic proof for general bipartite systems

We now give a proof that the only reversible transformations allowed in a bipartite Boxworld system are composed of local operations and permutations of the subsystems, so long as none of the subsystems are classical. This proof applies even to the case where the subsystems are non-identical, and is therefore already a novel result, not implied by the proof outlined in Section 5.1. Bipartite Boxworld effects are represented by 2-dimensional effect tables, for which local operations correspond to permutations of rows along a single axis, and a permutation of the subsystems corresponds to transposing the table (note that this transpose is only valid if the two subsystems are identical). Assuming that every measurement is non-trivial and has more than one outcome, each block in the effect table has height and width both at least 2 cells. As long as none of the subsystems are classical, then along both the horizontal and vertical axes there are at least 2 blocks.

Recall from Section 5.1.3 that we need only demonstrate that reversible transformations preserve a Hamming distance of 1 between pairs of fiducial effect strings. We will obtain this result for binary, bipartite Boxworld systems in a stepwise manner, supplementing many of the steps with an effect table illustration which provides a visualisation of the argument. In the bipartite setting a satisfactory proof of each step can be given by reference to effect table diagrams, since they take a much simpler form. In the general multipartite setting the same cannot be said, however the stepwise process used will be very similar, hence this proof of the bipartite case can be seen as setting up the necessary ideas and intuition that will then be applied to the multipartite case.

Central to the following argument are the notions of sub-rows and multiform effect tables. Recall that a sub-row is a binary effect table where the entries along a single block of a single row are exactly the non-zero entries. A binary effect table is multiform if there exists a distinct binary effect table which gives the same inner product for all states. Binary effect table E (strictly) *covers* binary effect table F if the non-zero components of F are a strict subset of the non-zero components of E.

Step 1. The only proper binary effect tables equivalent to a sub-row are other sub-rows in the same row.

Suppose that E is a sub-row, and that F is an effect table which is not a sub-row in the same row. Without loss of generality, suppose that E is a sub-row for some *column* of the table, i.e. E represents an effect of the form $X^{(1)}_{a|x} \otimes \mathcal{U}^{(2)}$. To demonstrate that E and F cannot represent equivalent effects, it suffices to construct a pure product state which has differing inner products with the effects represented by E and F. Recall that a pure, bipartite product state is uniquely specified by selecting one column from each block along the horizontal axis, and one row from each block along the vertical axis.

If F has a non-zero entry which is in a distinct row, then it is possible to select outcomes for each measurement of subsystem 1 so that the corresponding set of columns contains non-zero entries of F but not of E. To complete the specification of the pure product state, select a row from each block along the second axis, such that the previous non-zero entry of F is contained in at least one of them. The resulting state has inner product 1 with F and 0 with E. In the below table, the relevant non-zero entry of F is underlined; by selecting the second and third columns, and e.g. the second and third rows, we get a pure product state which distinguishes E from F.

$$\begin{array}{cc|cc|c} & \downarrow & \downarrow & & \\ \hline 1 & 0 & 0 & 0 & \\ \rightarrow\ 1 & 0 & \underline{1} & 0 & \\ \hline \rightarrow\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & \end{array} \tag{5.49}$$

If instead F has non-zero entries *only* in the same column that E does, then either F strictly covers a sub-row, in which case it is improper, or F has non-zero entries only for a strict subset of each block of that column. In the latter case, it is possible to select a set of rows of the table, one for each block along the

vertical axis, such that the set of rows contains non-zero entries of E but not of F. Selecting a set of columns which includes the one in which both E and F have non-zero entries will generate a pure product state which has inner product 1 with E and 0 with F. This process is again demonstrated in the below diagram, where the underlined entries are the non-zero entries of F.

$$\begin{array}{cc|cc|cc} & & \downarrow & \downarrow & & \\ \rightarrow & 0 & 1 & 0 & 0 & \\ & 0 & 1 & 0 & 0 & \\ \hline \rightarrow & 0 & 0 & 0 & 0 & \\ & 0 & \underline{1} & 0 & 0 & \\ \hline & 0 & \underline{1} & 0 & 0 & \\ \rightarrow & 0 & 0 & 0 & 0 & \end{array} \tag{5.50}$$

This concludes the proof of Step 1.

Step 2. A proper, binary effect table which does not cover any sub-rows is not multiform.

Let E be a binary, bipartite effect table which does not cover any sub-rows, and suppose for contradiction that it is equivalent to a distinct, binary, bipartite effect table F. Then there must exist some component (i, j) of the table for which E is zero and F is non-zero (here i represents the column and j the row, unlike conventional matrix notation). We aim to construct a pure product state which "hits" this (i, j), but none of the entries of E. Recall that the effect E can be proper only if all its non-zero entries lie in some block-row - without loss of generality suppose this is the first column of blocks, i.e. every non-zero entry of E is associated with the first measurement choice on subsystem 1. If the non-zero entry of F does not lie in this column of blocks, then by moving leftwards from this entry we must arrive at a cell (i', j) in the first column of blocks for which E is zero (otherwise E would cover a horizontal sub-row). Select a set of columns which includes i and i', and a

set of rows which includes j and a row from each remaining block such that E is zero on the i'th component of that row (this is possible since otherwise E would cover a vertical sub-row). The resulting pure product state has inner product 0 with E and 1 with F. In the below diagram, the underlined entry is the non-zero entry of F, so that $i = 3$ and $j = 5$; thus we obtain $i' = 2$ and also select row 3.

$$\begin{array}{cc|cc|cc} & & \downarrow & & \downarrow & \\ & 0 & 1 & 0 & 0 \\ & 0 & 1 & 0 & 0 \\ \rightarrow & 0 & 0 & 0 & 0 \\ \hline & 0 & 0 & 0 & 0 \\ \rightarrow & 1 & 0 & \underline{1} & 0 \end{array} \tag{5.51}$$

Suppose instead that entry (i, j) belongs to the first column of blocks. In this case it is similarly possible to select a set of rows which includes j and a row from each remaining block such that E is zero on the ith component of that row. Any set of columns which includes i will now result in a pure product state which has inner product 0 with E and 1 with F.

Step 3. Reversible transformations map sub-rows to sub-rows.

Let E be a sub-row; E is multiform, being equivalent to all other sub-rows contained in the same row as E (there is at least one other if all systems are non-classical). Let F be another sub-row contained in the same row: thus E and F are binary effect tables representing the same proper effect, but whose non-zero components are disjoint. Let T be a permutation of the effect table components, which represents an allowed reversible transformation on the Boxworld system. The image $T(E)$ of E under this transformation is a proper binary effect table which is equivalent to $T(F)$, hence is multiform. It follows from Step 2 that $T(E)$ covers a sub-row. The diagram below is an example of sub-rows E and F (underlined)

transforming under an arbitrary permutation of components - in this diagram the permutation does not in fact correspond to an allowed transformation, since it is ill-defined on effects which have multiple effect table representations.

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline \underline{1} & 0 & 0 & 0 \\ \underline{1} & 0 & 0 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & \underline{1} \\ 0 & \underline{1} & 0 & 0 \end{array}\right] \tag{5.52}$$

Suppose that $T(E)$ is not itself a sub-row, i.e. that it *strictly* covers a sub-row. Let G denote this sub-row, and let G' be another sub-row in the same row as G. The image $T^{-1}(G)$ of G under the inverse permutation T^{-1} is a binary effect which is strictly covered by E, and which is equivalent to the effect $T^{-1}(G')$. Hence $T^{-1}(G)$ is multiform. However, since $T^{-1}(G)$ is a strict subset of a sub-row, this contradicts Step 2. Therefore $T(E)$ is itself a sub-row, and T maps sub-rows to sub-rows.

Step 4. Reversible transformations preserve Hamming distance 1 between pairs of fiducial effect strings.

Suppose that e and f are composite fiducial effects, which have Hamming distance 1 between their respective fiducial effect strings. e and f are uniquely represented by binary effect tables E and F which have only a single non-zero entry, such that these two non-zero entries lie in some row or column of the table. E and F may be extended uniquely to sub-rows E' and F' which lie in this row and cover E and F respectively.

If the E' and F' are in fact the same sub-row, then by Step 3 their image under an allowed reversible transformation T is some sub-row $T(E')$. Thus the images $T(E)$ and $T(F)$ of E and F under T also belong to the same sub-row, therefore their corresponding fiducial effect strings also have a Hamming distance of 1. If E' and F' are not the same sub-row, they are nevertheless equivalent sub-rows,

and by Step 3 must be transformed to equivalent sub-rows under T. However, according to Step 1, sub-rows can only be equivalent if they belong to the same row, hence again $T(E)$ and $T(F)$ correspond to fiducial effect strings which have a Hamming distance of 1.

Step 5. The only reversible transformations allowed in bipartite boxworld systems, where none of the subsystems are classical, are compositions of relabellings of measurement choices and outcomes, and permutations of subsystems (assuming they are identical).

From Step 4 and Theorem 10 (which applies to arbitrary alphabets in each component), we have that any reversible transformation must be a composition of local operations and permutation of the subsystems. By Lemma 6, which also applies in the case where different measurements may have different numbers of outcomes, we have that the only local operations allowed are relabellings of measurement choices and outcomes. Note that permuting the states of the two subsystems corresponds to a valid operation only if the measurement choices on each subsystem can be matched up, so that the paired measurements have the same number of outcomes. It is not possible to permute the state of the subsystems if, for example, they have a different number of fiducial measurements or if the first subsystem has a measurement with M outcomes, but none of the measurements on the second subsystem have M outcomes. However, if the first subsystem has two fiducial measurements with 2 and 3 outcomes, and the second subsystem has two fiducial measurements measurements with 3 and 2 outcomes, then the states of these subsystems may be permuted as long as the first measurement of the first subsystem is matched with the second measurement of the second subsystem.

5.3 Decompositions of effects

In Section 5.2.2 we demonstrated, by means of diagrammatic representations, that the allowed reversible transformations of a bipartite Boxworld system are trivial, so long as none of the subsystems are classical. In generalising this result to general multipartite systems, it will be helpful to shift our primary focus to the algebraic, rather than diagrammatic, representations of effects, but to keep these diagrammatic notions at the back of one's mind as an aid to understanding the method of proof. The proof proceeds in a stepwise fashion largely similar to the bipartite case, though some care has to be taken to make rigorous some facts which are less evidently true in tabular arrays of larger dimension.

We begin by recapping the necessary features of Boxworld. An N-partite system is made up of N subsystems, for which $M^{(i)}$ is the number of local fiducial measurement choices on subsystem i, and $K_{x_i}^{(i)}$ is the number of outcomes for measurement x_i on subsystem i. We will always assume that these values are at least equal to 2, so that all subsystems are non-classical and all measurements are non-trivial. To each subsystem i belongs a set of effects $\mathcal{E}^{(i)}$, whose extreme rays are the fiducial effects $X_{a_i|x_i}^{(i)}$ which correspond to outcome a_i being obtained when measurement x_i is performed. The composite fiducial effects of an N-partite system are those of the form $e = X_{a_1|x_1}^{(1)} \otimes \cdots \otimes X_{a_N|x_N}^{(N)}$: a general N-partite effect is a conic combination of the composite fiducial effect vectors, and a general N-partite state is a vector which has non-negative inner product with all the composite fiducial effect vectors.

The pure product states of an N-partite system are those of the form $s = s^{(1)} \otimes \cdots \otimes s^{(N)}$, where $s^{(i)}$ is a pure state of subsystem i. Pure states of an individual subsystem assign specific outcomes for each fiducial measurement choice; we will use the notation $s_{x_i}^{(i)}$ to index this outcome, with $1 \leq s_{x_i}^{(i)} \leq K_{x_i}^{(i)}$. If the state s and effect e satisfy $\langle e, s \rangle = 1$, we will say that s *hits* e. Note that if $s = s^{(1)} \otimes \cdots \otimes s^{(N)}$ is a pure product state and $e = X_{a_1|x_1}^{(1)} \otimes \cdots \otimes X_{a_N|x_N}^{(N)}$ is a composite fiducial effect, then s hits e if, and only if, $s_{x_i}^{(i)} = a_i$ for all i.

Recall that reversible transformations permute the set of composite fiducial

effects, hence induce permutations on the components of the effect table representing the system. In particular, binary effect tables, whose non-zero entries are all equal to 1, are mapped to other binary effect tables. Binary effect tables represent effects which are sums of composite fiducial effects, i.e. which are of the form $E = \sum_\alpha e_\alpha$, where each e_α is a composite fiducial effect. This type of effect merits special attention in our treatment.

Definition. *Suppose that for a set of composite fiducial effects* $\{e_\alpha\}$ *and an effect* E *we have that* $E = \sum_\alpha e_\alpha$. *Then* $\{e_\alpha\}$ *is a* decomposition *of* E.

We may also use the terminology: E *admits* the decomposition $\{e_\alpha\}$. We will tend to use lowercase letters for composite fiducial effects, and uppercase for sums of extreme ray effects.

Definition. *An effect* E *is* multiform *if it can be written* $E = \sum_\alpha e_\alpha = \sum_\beta f_\beta$ *where* $\{e_\alpha\}$ *and* $\{f_\beta\}$ *are distinct sets of composite fiducial effects.*

In other words, the distinct binary effect tables which have 1s in the entries corresponding to effects e_α and f_β respectively, are equivalent, both representing the effect E.

Definition. *A* sub-unit effect *is an effect of the form* $X^{(1)}_{a_1|x_1} \otimes \cdots \otimes \mathcal{U}^{(i)} \otimes \cdots \otimes X^{(N)}_{a_N|x_N}$, *where exactly one component of the tensor product is the unit effect, and the remainder are local fiducial effects. A sub-unit effect is an* i-sub-unit *effect if its* i*th component is the unit effect.*

Sub-unit effects are represented by the binary effect tables previously referred to as sub-rows; for bipartite systems, 1-sub-unit effects are represented by horizontal sub-rows and 2-sub-unit effects are represented by vertical sub-rows. Just as sub-rows are multiform, so are sub-unit effects, with the number of decompositions of an i-sub-unit effect being at least $M^{(i)}$, the number of fiducial measurement choices on subsystem i.

Fiducial and sub-unit effects can uniquely be written as a tensor products of vectors, such that each component i of the tensor product is equal either to some

$X^{(i)}_{a_i|x_i}$ or to $\mathcal{U}^{(i)}$. Therefore we may refer to their ith component in a well-defined sense; for example, if

$$E = X^{(1)}_{a_1|x_1} \otimes \mathcal{U}^{(2)} \otimes X^{(3)}_{a_3|x_3}, \tag{5.53}$$

then $E^{(1)} = X^{(1)}_{a_1|x_1}$ and $E^{(2)} = \mathcal{U}^{(2)}$. For a subset $\Omega \subseteq [N]$ we will write $E^{\Omega} = \bigotimes_{i\in\Omega} E^{(i)}$, so that in the above example $E^{\{2,3\}} = \mathcal{U}^{(2)} \otimes X^{(3)}_{a_3|x_3}$. Finally, in analogy to one binary effect table covering another if the non-zero components of the latter are a subset of the non-zero components of the former, we have the following definition.

Definition. *A set of fiducial effects $\{e_\alpha\}_{\alpha\in A}$ (strictly)* covers *the effect E if there is some (strict) subset $B \subseteq A$ such that $\sum_{\alpha\in B} e_\alpha = E$.*

The following two lemmas (and their corollaries) establish useful facts about effects and the ways in which they may be decomposed into sums of composite fiducial effects. We are primarily concerned with effects that are sums of composite fiducial effects with coefficient 1, and hence which may in principle be represented by binary effect tables (although these may well exist in many-dimensional spaces). In terms of binary effect tables, Lemma 7 is the algebraic version of Step 1 of Section 5.2.2, demonstrating that the only binary effect tables equivalent to a sub-row are other sub-rows belonging to the same row.

Being rather technical in nature, the proof of Lemma 7 merits some discussion (the following proofs are no less technical, but the discussion here will serve equally well to illuminate them also). In essence, the linear relations (or lack of) between the fiducial effect vectors are exploited in order to derive a categorisation of the ways in which sub-unit effects may be decomposed as sums of fiducial effects. However, instead of working directly with the notion of linear independence, we employ the trick of considering the inner products between fiducial effects and pure product states. This strategy is hinted at in [4] and, after some thought, is unsurprising: we know that the pure states on a single subsystem are the deterministic states, and we know also that distinct effects must have distinct

inner products for at least one pure product state. Indeed, one useful perspective to bear in mind is the following: demanding that every valid conditional distribution over the measurements corresponds to an allowed state is equivalent to demanding the precise linear dependencies that exist between the fiducial effect vectors.

Consider a pair of distinct fiducial effects in an individual non-classical Box-world system. If this pair of effects belongs to the same measurement, for example the pair $X_{1|1}$ and $X_{2|1}$, then the state which deterministically outputs 1 for every fiducial measurement will hit the first effect, but not the second. If the pair belongs to distinct measurements, for example the pair $X_{1|1}$ and $X_{1|2}$, then any state which outputs 1 for the first measurement, 2 for the second measurement, and anything for the remaining measurements will again hit the first effect but not the second. It will be noticed that the ability to distinguish distinct effects with pure states is not specific to the examples chosen; this is central to proving the first part of the Lemma.

Suppose again that a pair of fiducial effects belongs to different measurements, for example $X_{1|1}$ and $X_{1|2}$. Then it is just as easy to find a pure product state which hits *both* effects: the state s which always outputs 1 will again suffice. Therefore $\langle X_{1|1} + X_{1|2}, s\rangle = 2$, implying that $X_{1|1} + X_{1|2}$ is an improper effect. Thus, $X_{1|1}$ and $X_{1|2}$ *cannot* belong to the same decomposition of the unit effect. Again, this is not specific to the example chosen: it follows almost immediately that the only fiducial effect decompositions of the unit effect are given by the set of outcomes belonging to a single measurement.

Lemma 7. *Let $E = \sum_\alpha e_\alpha$ be an i-sub-unit effect. Then each composite fiducial effect e_α satisfies $e_\alpha^{(j)} = E^{(j)}$ for all components $j \neq i$. Moreover, the set of ith components $\{e_\alpha^{(i)}\}$ forms a local fiducial measurement on subsystem i.*

Proof. We will prove the lemma by contradiction. Suppose first that $e_{\alpha'}^{(j)} \neq E^{(j)}$ for some α' and some $j \neq i$. Let $E^{(j)} = X^{(j)}_{a_j|x_j}$ and $e^{(j)}_{\alpha'} = X^{(j)}_{a'_j|x'_j}$. Either $x_j \neq x'_j$, or $x_j = x'_j$ but $a_j \neq a'_j$, so we can construct a pure product state $s^{(j)}$ on system j such that $s^{(j)}_{x_j} \neq a_j$ and $s^{(j)}_{x'_j} = a'_j$, so that $s^{(j)}$ hits $e^{(j)}_{\alpha'}$ but not $E^{(j)}$. Then there

exists a pure product state s whose jth component is $s^{(j)}$, so that $\langle E, s\rangle = 0$, but for which $\langle e_{\alpha'}, s\rangle = 1$, contradicting the fact that $\langle e_{\alpha'}, s\rangle \leq \langle E, s\rangle$.

$\{e_\alpha^{(i)}\}$ is a set of fiducial effects satisfying $\sum_\alpha e_\alpha^{(i)} = \mathcal{U}^{(i)}$, hence any pure state $s^{(i)}$ on system i must hit exactly one of the $e_\alpha^{(i)}$. If any two of the $e_\alpha^{(i)}$ are effects corresponding to different measurements, then there is a pure state $s^{(i)}$ which hits both of them. Hence the effects all belong to the same fiducial measurement x; if $\{e_\alpha^{(i)}\}$ is not the *full* set of outcomes of measurement x, then there is a pure state $s^{(i)}$ which hits none of them. It follows that $\{e_\alpha^{(i)}\}$ forms a fiducial measurement on subsystem i. □

Corollary 4. *Let $E = \sum_{\alpha=1}^r e_\alpha$ be a sub-unit effect. Then $\{e_\alpha\}$ does not strictly cover a multiform effect.*

Proof. From Lemma 1 it follows that $e_\alpha^{(j)} = E^{(j)}$ for all components $j \neq i$, and that the set $\{e_\alpha^{(i)}\}$ corresponds to the full set of outcomes for a local fiducial measurement on subsystem i. In other words, the decompositions of E are in a one-to-one correspondence with the *local* fiducial effect decompositions of $\mathcal{U}^{(i)}$, each of which is obtained by fixing a measurement choice x_i, and summing all the fiducial effect vectors which correspond to an outcome for that measurement:

$$\mathcal{U}^{(i)} = \sum_{x_i} X^{(i)}_{a_i|x_i}. \tag{5.54}$$

In particular, there are only a finite number of such local decompositions of $\mathcal{U}^{(i)}$, and the sets of local fiducial effects making up these decompositions are pairwise disjoint. Likewise, there are only a finite number of decompositions of E, and the sets of *composite* fiducial effects making up these decompositions are pairwise disjoint.

Suppose that $\{e_\alpha\}$ covers a multiform effect, i.e. (after relabelling the e_α if necessary) for some integer $s < r$ and some integer t, there exist distinct sets $\{e_\alpha\}_{\alpha=1}^s$, $\{f_\beta\}_{\beta=1}^t$ of fiducial effects such that $\sum_{\alpha=1}^s e_\alpha = \sum_{\beta=1}^t f_\beta$. Then $\{f_1, \ldots, f_t, e_{s+1}, \ldots e_r\}$ is a decomposition of E distinct from $\{e_1, \ldots, e_r\}$. How-

ever, both these sets contain e_r, contradicting the fact that the decompositions of E are pairwise disjoint. Therefore $\{e_\alpha\}$ does not cover a multiform effect. □

Just as Lemma 7 was the algebraic version of Step 1 of 5.2.2, Lemma 8 is the algebraic version of Step 2; however, since it is difficult to categorise which effects are proper when the number of subsystems is large, a slightly different approach must be used. Specifically, let r be the least number of non-zero entries over all possible sub-rows of the system. Then Lemma 8 states that any binary effect table which has at most r non-zero entries, and which does not cover any sub-rows, is not multiform. It turns out that, along with Lemma 7, this is enough to prove the desired result in the case that all subsystems have at least one measurement with r outcomes - for the general case, however, we will require a more sophisticated approach.

For convenience we will assume from here on that the subsystems are arranged in order of increasing numbers of measurement outcomes, i.e. $K_j^{(i)} \leq K_{j+1}^{(i)}$ and $K_1^{(i)} \leq K_1^{(i+1)}$; this amounts to a relabelling of subsystems and measurement choices. $K_1^{(1)}$ is therefore the smallest number of outcomes possible for any fiducial measurement.

Lemma 8. *For $r \leq K_1^{(1)}$ suppose that $\{e_\alpha\}_{\alpha=1}^r$ does not cover any sub-unit effects. Then for any fiducial effect $f \notin \{e_\alpha\}$, there is a pure product state which hits f but none of the e_α.*

Proof. Let $f = X_{a_1|x_1}^{(1)} \otimes \cdots \otimes X_{a_N|x_N}^{(N)}$. We proceed by induction on the number of subsystems N. When $N = 1$ set $s_{x_1} = a_1$ to ensure that s hits $X_{a_1|x_1}^{(1)}$. The conditions imply that no partial sum of $\{e_\alpha\}$ equals the unit effect, hence for each other choice of measurement $x' \neq x_1$, it must be possible to choose $s_{x'}$ such that $X_{s_{x'}|x'} \notin \{e_\alpha\}$. By construction s hits $X_{a_1|x_1}^{(1)}$ but none of the e_α.

When $N > 1$, note that for any fiducial effect g on system 1, the set $\{e_\alpha^{\{2,\ldots,N\}} : e_\alpha^{(1)} = g\}$ is a decomposition of some effect on the remaining $N-1$ subsystems. This decomposition has size most $K_1^{(1)} \leq K_1^{(2)}$, and also cannot cover any sub-unit effects, hence, by the induction hypothesis applied to the case $N-1$, there

exists a pure product state $s^{(2)} \otimes \cdots \otimes s^{(N)}$ which hits $f^{\{2,\ldots,N\}}$ but none of the set $\{e_\alpha^{\{2,\ldots,N\}} : e_\alpha^{(1)} = g\}$.

Again, it is necessary to set $s_{x_1}^{(1)} = a_1$. Consider the set $\{e_\alpha^{(1)}\}$, and the outcomes for fiducial measurement choices other than x_1 on system 1. One of two cases must occur:

(a) The set $\{e_\alpha^{(1)}\}$ fills none of the other measurements, i.e. for every $x' \neq x_1$, there is an $a_{x'}$ such that $X_{a_{x'}|x'}^{(1)} \notin \{e_\alpha^{(1)}\}$. For each such x' set $s_{x'}^{(1)} = a_{x'}$ so that s can hit e_α only if $e_\alpha^{(1)} = f^{(1)}$. However, using the inductive hypothesis, there exists a pure product state $s^{(2)} \otimes \cdots \otimes s^{(N)}$ which hits $f^{\{2,\ldots,N\}}$ but none of the set $\{e_\alpha^{\{2,\ldots,N\}} : e_\alpha^{(1)} = f^{(1)}\}$.

(b) There exists a measurement $x' \neq x_1$ on system 1 with $K_{x'}^{(1)} = r$ which is filled by the set $\{e_\alpha^{(1)}\}$, i.e. (after reordering) $e_\alpha^{(1)} = X_{\alpha|x'}^{(1)}$ for $1 \leq \alpha \leq r$. $\{e_\alpha\}$ covers no sub-unit effects, so there must be some α' and some system $i \neq 1$ such that $e_{\alpha'}^{(i)} \neq f^{(i)}$. Set $s_{x'}^{(1)} = \alpha'$ so that s does not hit any e_α with $\alpha \neq \alpha'$; the remaining components of $s^{(1)}$ may be chosen arbitrarily. By the inductive hypothesis, there exists a pure product state $s^{(2)} \otimes \cdots \otimes s^{(N)}$ which hits $f^{\{2,\ldots,N\}}$ but not the single effect $e_{\alpha'}^{\{2,\ldots,N\}}$.

In both cases, by construction $s = s^{(1)} \otimes \cdots \otimes s^{(N)}$ hits f but none of the e_α. □

Corollary 5. *The only multiform effects which have a decomposition with exactly $K_1^{(1)}$ elements are sub-unit effects.*

Proof. Suppose $E = \sum_{\alpha=1}^{r} e_\alpha = \sum_{\beta=1}^{s} f_\beta$ are distinct decompositions, with $r = K_1^{(1)}$, and suppose without loss of generality that $f_1 \notin \{e_\alpha\}_{\alpha=1}^{r}$. Every pure product state which hits f_1 must also hit one of the e_α, so it follows from Lemma 8 that $\{e_\alpha\}$ covers a sub-unit effect. By Lemma 7, every decomposition of a sub-unit effect has at least $K_1^{(1)}$ elements, hence E is itself a sub-unit effect. □

5.3.1 Identical subsystems revisited

We now demonstrate that Lemmas 7 and 8 are sufficient to prove that all reversible Boxworld transformations are trivial, in the case that all subsystems are non-classical, and that there is a fixed positive integer r such that the least number of outcomes for any local fiducial measurement on each subsystem i is r. This is a weaker condition than demanding that all subsystems are identical, which is weaker still than demanding in addition that all fiducial measurements on all subsystems have the same number of outcomes. This last condition is exactly that assumed in [4], hence the result we recover in this section, whilst not fully general, is already stronger than has previously been demonstrated.

Assuming that the least number of fiducial measurement outcomes is fixed across subsystems, consider an i-sub-unit effect E for arbitrary i. Note that E is a multiform effect with at least one decomposition with exactly $r = K_1^{(1)}$ elements (corresponding to the measurement on subsystem i which has r outcomes). It follows that $T^\dagger(E)$ is also a multiform effect with at least one decomposition with exactly $K_1^{(1)}$ elements, hence by Corollary 5, $T^\dagger(E)$ is a sub-unit effect. Thus, $T^\dagger$ maps sub-unit effects to sub-unit effects.

Now let $e = X^{(1)}_{a_1|x_1} \otimes \cdots \otimes X^{(N)}_{a_N|x_N}$ and $e' = X^{(1)}_{a'_1|x'_1} \otimes \cdots \otimes X^{(N)}_{a'_N|x'_N}$ be composite fiducial effects whose fiducial effect strings differ only in component i. Observe that both effects belong to a decomposition of the i-sub-unit effect

$$E = X^{(1)}_{a_1|x_1} \otimes \cdots \otimes \mathcal{U}^{(i)} \otimes \cdots \otimes X^{(N)}_{a_N|x_N}. \tag{5.55}$$

Suppose that E is mapped to a j-sub-unit effect, for some j (not necessarily equal to i). It follows that $T^\dagger(e)$ and $T^\dagger(e')$ are distinct composite fiducial effects which belong to some decomposition of $T^\dagger(E)$. Lemma 7 then implies that $T^\dagger(e)$ and $T^\dagger(e')$ differ only in component j. Therefore $T^\dagger$ preserves a Hamming distance of 1 between fiducial effect strings, and we may apply Theorem 10 and Lemma 6 to deduce that $T^\dagger$ must be some relabelling of subsystems, measurement choices and measurement outcomes.

Theorem 12. *The allowed reversible transformations of a Boxworld system with non-classical subsystems, such that the least possible number of measurement outcomes is constant across subsystems, are relabellings of subsystems, and local relabellings of measurements and measurement outcomes.*

Proof. We have already argued that these are the only possible reversible transformations. It remains to show that any such transformation is allowed, as long as the number of outcomes of each local fiducial measurement is respected. In particular, a relabelling of measurements on a single subsystem is valid only if whenever measurement x is mapped to measurement x', then both x and x' have the same number of outcomes. A relabelling of subsystems is similarly valid only if whenever subsystem i is mapped to subsystem j, then the measurement choices in i are individually mapped to measurement choices in j which have the same number of outcomes. Modulo these considerations, it is clear that any such relabelling is a well-defined mapping on outcome distributions of the form $P(a_1, \ldots, a_N | x_1, \ldots, x_N)$. Thus valid relabellings are reversible and convex-linear mappings of allowed states to allowed states. □

5.4 Main result

We now relax the condition that $K_1^{(i)} = K_1^{(1)}$ for all subsystems i, and demand merely that each subsystem is non-classical (has more than one fiducial measurement), and each local fiducial measurement is non-trivial (has more than one outcome). This makes the task more complicated; not all sub-unit effects have decompositions with $K_1^{(1)}$ elements, and so Corollary 5 cannot universally be applied in the same way.

However, Corollary 5 *can* still be applied for each subsystem i that *does* obey $K_1^{(i)} = K_1^{(1)}$, so that $T^\dagger$ permutes the set of sub-unit effects *within* this set of subsystems. Now consider a system j for which $K_1^{(j)}$ is the next possible greater value than $K_1^{(1)}$. A j-sub-unit effect E must be transformed to something which is multiform with some decomposition $\{e_\alpha\}$ involving $K_1^{(j)}$ elements. We will

demonstrate that if (a) $\sum_\alpha e_\alpha$ is not a sub-unit effect, and (b) $T^\dagger$ permutes sub-unit effects on subsystems with $K_1^{(i)} = K_1^{(1)}$, then for any other decomposition $\{f_\beta\}$ of $T^\dagger(E)$ there exists a pure product state s which hits f_1 but none of the effects $\{e_\alpha\}$. This sets up an iterative process, at each stage assuming that $T^\dagger$ permutes the sub-unit effects on subsystems with smaller numbers of outcomes. The iteration terminates when $K_1^{(j)}$ takes on its maximal value, and $T^\dagger$ thus permutes the full set of sub-unit effects in the system.

The following definition is convenient for discussing the set of sub-units which belong to one of several subsystems:

Definition. *In a multipartite Boxworld system comprising N subsystems, $\mathcal{S}_{\{i\}}$ is the set of sub-unit effects at system $i \in [N]$. If $\Omega \subseteq [N]$ is a subset of the N subsystems, then $\mathcal{S}_\Omega = \cup_{i\in\Omega}\mathcal{S}_{\{i\}}$.*

In order to prove that sub-unit effects are mapped to sub-unit effects, the following lemmas will prove useful.

Lemma 9. *Let $\Omega \subseteq [N]$ and suppose that $T^\dagger$ is an allowed reversible transformation which permutes the set $\mathcal{S}_\Omega$. Then the images of two composite fiducial effects will be identical for all components outside Ω if and only if the original effects were. i.e.:*

$$e_1^{\bar{\Omega}} = e_2^{\bar{\Omega}} \iff \left(T^\dagger(e_1)\right)^{\bar{\Omega}} = \left(T^\dagger(e_2)\right)^{\bar{\Omega}}$$

where $\bar{\Omega} = [N] \setminus \Omega$.

Proof. Suppose firstly that the composite fiducial effects e_1 and e_2 differ only in one component $i \in \Omega$. Observe that e_1 and e_2 belong to (possibly different) decompositions of a unique sub-unit effect $E \in \mathcal{S}_{\{i\}}$. By assumption $T^\dagger(E)$ is a j-sub-unit effect for some $j \in \Omega$; $T^\dagger(e_1)$ and $T^\dagger(e_2)$ belong to decompositions of $T^\dagger(E)$, hence by Lemma 7 can only differ in component j.

Suppose now that e_1, e_2 satisfy $e_1^{(k)} = e_2^{(k)}$ for all $k \notin \Omega$, but that they differ in any number of components belonging to Ω. Then it is possible to move from e_1

to e_2 by changing one component at a time (each component belonging to Ω). At each step, $T^\dagger$ maps the corresponding pair of effects to a pair which differ only in components belonging to Ω. Hence $T^\dagger(e_1)^{(k)} = T^\dagger(e_2)^{(k)}$ for all $k \notin \Omega$.

To prove the converse direction, note that if $T^\dagger$ is an allowed reversible transformation which permutes the set $\mathcal{S}_\Omega$, then so is $(T^\dagger)^{-1}$. □

Lemma 10. *Suppose that $E = \sum_{\alpha=1}^{r} e_\alpha$ is a sub-unit effect, and that $\{T^\dagger(e_\alpha)\}$ covers a sub-unit effect F. Then $T^\dagger(E) = F$.*

Proof. Without loss of generality let $\sum_{\alpha=1}^{s} T^\dagger(e_\alpha) = F$ for $s \leq r$, and let $\sum_\beta f_\beta$ be a distinct decomposition of F. Then E covers the multiform effect $(T^\dagger)^{-1}(F) = \sum_{\alpha=1}^{s} e_\alpha = \sum_{\beta=1}(T^\dagger)^{-1}(f_\beta)$. It follows from Corollary 4 that $s = r$ and that $T^\dagger(E) = F$. □

Lemma 11. *Suppose that $\{e_\alpha\}$ does not cover any sub-unit effects, but that there exists some subsystem i for which $\sum_\alpha e_\alpha^{(i)} = \mathcal{U}^{(i)}$. Then for any fiducial effect $f \notin \{e_\alpha\}$, there exists a pure product state which hits f but none of the e_α.*

Proof. Let $f^{(i)} = X^{(i)}_{a|x}$ and $\Omega_i = [N] \setminus \{i\}$. Note that $\{e_\alpha^{(i)}\}$ is the complete set of outcomes for some fiducial measurement x' on subsystem i: without loss of generality, $e_\alpha^{(i)} = X^{(i)}_{\alpha|x'}$.

If $x' = x$, then $f^{(i)} = e_a^{(i)}$. Set $s_x^{(i)} = a$ and choose the remaining components of $s^{(i)}$ arbitrarily, so that $s^{(i)}$ hits $e_a^{(i)}$ but none of the other $e_\alpha^{(i)}$. Note that f^{Ω_i} and $e_a^{\Omega_i}$ must be distinct fiducial effects, so by Lemma 8 there exists a pure product state s^{Ω_i} which hits f^{Ω_i} but not $e_a^{\Omega_i}$.

If $x' \neq x$, then since $\sum e_\alpha$ is not a sub-unit effect, there exists α' and $i' \neq i$ such that $e_{\alpha'}^{(i')} \neq f^{(i')}$. Set $s_x^{(i)} = a$ and $s_{x'}^{(i)} = \alpha'$, and choose the remaining components of $s^{(i)}$ arbitrarily. Again, by Lemma 8 there is a pure product state s^{Ω_i} which hits f^{Ω_i} but not the single fiducial effect $e_{\alpha'}^{\Omega_i}$.

In both cases, combining $s^{(i)}$ with s^{Ω_i} gives a pure product state s which hits f but none of the e_α. □

Lemma 12. *Reversible Boxworld transformations map sub-unit effects to sub-unit effects, so long as none of the subsystems are classical.*

Proof. We begin by considering the action of $T^\dagger$ on a 1-sub-unit effect E. $T^\dagger(E)$ is a multiform effect with a decomposition containing $K_1^{(1)}$ elements, hence by Corollary 5 it is a j-sub-unit effect for some subsystem j with $K_1^{(j)} = K_1^{(1)}$. By the same reasoning $T^\dagger$ permutes the set $\mathcal{S}_\Omega$, where $\Omega = \{j : K_1^{(j)} = K_1^{(1)}\}$.

We now show iteratively for each positive integer $r > K_1^{(1)}$ that $T^\dagger$ permutes the set $\mathcal{S}_{\Omega_r}$, where $\Omega_r = \{i : K_1^{(i)} = r\}$. Let $i \in \Omega_r$, let $\sum_{\alpha=1}^r e_\alpha = \sum_{\beta=1}^s e'_\beta$ be distinct decompositions of an i-sub-unit effect E and assume that $T^\dagger$ permutes the set $\mathcal{S}_\Omega$, where $\Omega = \{j : K_1^{(j)} < r\}$. Note that $T^\dagger(E)$ is also multiform, since $T^\dagger(E) = \sum_{\alpha=1}^r T^\dagger(e_\alpha) = \sum_\beta T^\dagger(e'_\beta)$. Write $f_\alpha = T^\dagger(e_\alpha)$ and $g = T^\dagger(e'_1)$, noting that $g \notin \{f_\alpha\}$.

Assuming that $\{f_\alpha\}$ does not cover a sub-unit effect, our aim is to construct a pure product state s that hits g but none of the f_α, giving a contradiction. Hence $\{T^\dagger(e_\alpha)\}$ must cover a sub-unit effect. It then follows from Lemma 10 that $T^\dagger(E)$ is itself an i'-sub-unit effect for some $i' \in \Omega_r$. By continuing the iteration, we complete the proof of the lemma.

To obtain the desired contradiction, assume that $\{f_\alpha\}$ does not cover a sub-unit effect. Let $\bar{\Omega} = [N] \setminus \Omega$ and consider the set $\{f_\alpha^{\bar{\Omega}}\}_{\alpha=1}^r$ (recall that for a fiducial effect f, $f^{\bar{\Omega}}$ is the tensor product of all those components of f belonging to $\bar{\Omega}$). Since $e_1'^{(i)}$ is distinct from $e_\alpha^{(i)}$ for all α, and $i \in \bar{\Omega}$, we have that $e_1'^{\bar{\Omega}} \notin \{e_\alpha^{\bar{\Omega}}\}$. It follows from Lemma 9 that $g^{\bar{\Omega}} \notin \{f_\alpha^{\bar{\Omega}}\}$.

Suppose that there exists a subsystem $i' \in \bar{\Omega}$ such that $\{f_\alpha^{(i')}\}_{\alpha=1}^r$ covers the local unit effect $\mathcal{U}^{(i')}$, i.e. $\sum_\alpha f_\alpha^{(i')} \geq_{\mathcal{E}_+^{(i')}} \mathcal{U}^{(i')}$. Recall that the fiducial effect decompositions of $\mathcal{U}^{(i')}$ are obtained by fixing a fiducial measurement on subsystem i' and taking the set of local fiducial effect vectors which correspond to an outcome of that measurement. Since all fiducial measurements on subsystem i' have at least r outcomes, it follows that $\sum_\alpha f_\alpha^{(i)} = \mathcal{U}^{(i)}$, and by Lemma 11 there exists a state which hits g but none of the f_α.

Suppose instead that there is no subsystem $i' \in \bar{\Omega}$ for which $\{f_\alpha^{(i')}\}_{\alpha=1}^r$ covers $\mathcal{U}^{(i')}$. Then the set $\{f_\alpha^{\bar{\Omega}}\}$ -- considered as a collection of fiducial effects over the maximal tensor product of all subsystems belonging to $\bar{\Omega}$ -- does not cover any

sub-unit effects. By Lemma 8 applied to the subsystems belonging to $\bar{\Omega}$, there exists a pure product state $s^{\bar{\Omega}}$ which hits $g^{\bar{\Omega}}$ but none of the $f_\alpha^{\bar{\Omega}}$. Combining $s^{\bar{\Omega}}$ with any pure state s^{Ω} which hits g^{Ω} gives a pure product state s which hits g but none of the f_α. □

Having proved that reversible transformations of Boxworld systems map sub-unit effects to sub-unit effects, it is straightforward once again to show that they must be trivial.

Theorem 13. *The only reversible transformations of non-classical systems allowed in Boxworld are relabellings of subsystems, and local relabellings of measurement choices and measurement outcomes.*

Proof. To complete the proof, we again need only check that all relabellings of this form are allowed transformations, as long as subsystems are only permuted only if they are identical, and measurement choices are permuted only if they have the same number of outcomes. Again, this is obvious from considering the action of such transformations on distributions in the form $P\left(a_1, \ldots, a_N | x_1, \ldots, x_N\right)$. □

5.5 Polytopic models with non-trivial reversible dynamics

In this Section we introduce and discuss a new probabilistic model which does not belong to Boxworld, but shares some similarities with it. At several points in the proof of Theorem 13 above, we have relied on two essential features of Boxworld which do not hold in quantum theory; that the number of extreme-ray vectors of the local state and effect cones is finite (that is to say, the model is polytopic), and that local systems combine under the maximal tensor product. The combination of these features implies that the extreme-ray vectors of the composite effect cone are exactly the tensor products of the local extreme-ray effect vectors. Consequently, the adjoint of any reversible transformation of the composite space must map composite fiducial effects to composite fiducial effects.

The obvious question is then: are these features sufficient to derive the conclusion of Theorem 13? Does there exist any probabilistic model whose local state-cones are polytopic cones, and whose systems combine under the maximal tensor product, yet which has reversible transformations on its composite systems which are not made up of local operations and relabelling of subsystems? In this Section we demonstrate that such models do exist, by way of an explicit example.

Before introducing this example, it is worth discussing polytopic models in more detail. The local systems of a polytopic model will still be characterised by a finite set of fiducial measurements, however it will not generally be the case that *any* outcome distribution on those measurements is permitted. Instead, some further set of restrictions are imposed on the space of possible outcome distributions, so that they form a strict subset of the full set of outcome distributions that would be present in a Boxworld system which has the same number of fiducial measurements and outcomes. To see this in a concrete example, recall the "polygon models" of Chapter 2, in which the local state space takes the form of a regular n-gon, for some integer n. In the case $n = 5$, the set of pairs of fiducial effects $\{e_i, e_{n+i}\}_{i=1}^{5}$ form a set of 5 measurements that is fiducial for the state space. Thus it is possible to characterise a state by assigning outcome probabilities for each of these measurements, rather than specifying its position on the pentagon: in this way the set of outcome distributions is naturally a subset of the set of outcome distributions of a Boxworld systems with 5 measurements and 2 outcomes. This subset is strict, since unlike a Boxworld system it is not possible for a state in a pentagon system to assign a definite outcome to every fiducial measurement (such a vector would have to have inner product 1 with 5 fiducial effect vectors, as well as the unit effect, which it can easily be checked is impossible).

Recall that an individual Boxworld system may be defined by specifying a unit effect $\mathcal{U}$ and a set of fiducial effect vectors $\{X_{a|x}\}_{a=1}^{K_x-1}$, all of which are linearly independent. The remaining fiducial effect vectors are defined via the normalization conditions $X_{K_x|x} = \mathcal{U} - \sum_{a=1}^{K_x-1} X_{a|x}$. Thus a g-bit (binary input/output) system is represented by a 3-dimensional state space, wherein the set of normalised states

lie in the subspace $\{v : \langle \mathcal{U}, v\rangle = 1\}$. Suppose now that we "squash" the g-bit state space, by introducing one extra restriction on the outcome statistics: namely, that the probability of obtaining outcome $a = 1$ is invariant of whether measurement choice $x = 1$ or $x = 2$ was performed. The vector of probabilities representing the outcome distribution must then take the form:

$$(p, 1-p \mid p, 1-p). \tag{5.56}$$

In fact, this is a representation of a classical bit, since only one fiducial measurement ($x = 1$ for example) is sufficient to characterise any such state, and any outcome distribution on measurement 1 is possible. In order to make this "squashing" procedure more interesting we can apply it to a g-trit, i.e. a Boxworld system with two fiducial measurements, each of which has three outcomes. In this case, the vector of outcome probabilities takes the following form:

$$(p,\ q,\ 1-p-q \mid p,\ q',\ 1-p-q'). \tag{5.57}$$

The effect cone which is dual to this set of states may be constructed in the following way: in a real vector space of dimension 4, choose a set of linearly independent vectors $\{\mathcal{U}, X_1, X_{2|1}, X_{2|2}\}$. Unlike Boxworld, the vector X_1 corresponds to the obtaining the first outcome for both measurements $x = 1$ and $x = 2$. The remaining fiducial effect vectors $X_{3|2}$ and $X_{3|3}$ are defined according to the normalisation relations:

$$\mathcal{U} = X_1 + X_{2|1} + X_{3|1} = X_1 + X_{2|2} + X_{3|2}. \tag{5.58}$$

In this Section we will refer to such a system as a *squashed g-trit*.

Proposition 8. *Suppose that a composite system is the max-tensor product of two subsystems, the first subsystem being a squashed g-trit and the second subsystem being a standard g-bit. Then there is an allowed adjoint transformation which switches the composite fiducial effects* $X_1^{(1)} \otimes X_{1|1}^{(2)}$ *and* $X_1^{(1)} \otimes X_{2|1}^{(2)}$*, whilst leaving*

all other composite fiducial effects invariant. Moreover, this transformation is reversible and non-trivial (not a composition of local operations and permutations of subsytems).

Proof. Recall that a g-bit system has fiducial effect vectors $\{X_{1|1}, X_{2|1}, X_{1|2}, X_{2|2}\}$, such that $X_{1|1} + X_{2|1} = X_{1|2} + X_{2|2} = \mathcal{U}$. The given transformation is essentially a kind of *C-NOT* operation on the systems involved: by mapping $X_1^{(1)} \otimes X_{1|1}^{(2)} \leftrightarrow X_1^{(1)} \otimes X_{2|1}^{(2)}$ we are imposing that the outcomes of measurement $x = 1$ on subsystem 2 are switched, *conditional* on outcome 1 occurring on subsystem 1. In order to prove the proposition, we must verify that the given transformation obeys three properties: that it is allowed (i.e. maps allowed states to allowed states in a convex-linear fashion), that it is reversible, and that it is non-trivial.

1. Allowed.

Suppose that vector spaces V_1 and V_2 represent the squashed g-trit and g-bit subsystems respectively. It suffices to show that there exists a linear operator $T^\dagger$ on $V_1 \otimes V_2$ which permutes the set of composite fiducial effect vectors in the required manner. Take bases $\mathcal{B}^{(1)} = \{\mathcal{U}, X_1, X_{2|1}, X_{2|2}\}$ and $\mathcal{B}^{(2)} = \{\mathcal{U}, X_{1|1}, X_{1|2}\}$ for V_1 and V_2 respectively. By expanding tensor products of these basis elements in terms of the composite fiducial effect vectors, and applying the given permutation of fiducial effect vectors (the "*C-NOT*" operation), we may derive the action that $T^\dagger$ must take on the tensor product basis elements. For example, we may begin by demanding that,

$$\begin{aligned} T^\dagger \left(X_1^{(1)} \otimes X_{1|1}^{(2)} \right) &= X_1^{(1)} \otimes X_{2|1}^{(2)} \\ &= X_1^{(1)} \otimes \mathcal{U}^{(2)} - X_1^{(1)} \otimes X_{1|1}^{(2)}. \end{aligned} \tag{5.59}$$

As required, $T^\dagger$ leaves the composite unit effect invariant:

$$\begin{aligned} T^\dagger\left(\mathcal{U}^{(1)}\otimes\mathcal{U}^{(2)}\right) &= T^\dagger\left(\left[X_1^{(1)}+X_{2|1}^{(1)}+X_{3|1}^{(1)}\right]\otimes\left[X_{1|1}^{(2)}+X_{2|1}^{(2)}\right]\right) \\ &= T^\dagger\left(X_1^{(1)}\otimes\left[X_{1|1}^{(2)}+X_{2|1}^{(2)}\right]+\left[X_{2|1}^{(1)}+X_{3|1}^{(1)}\right]\otimes\left[X_{1|1}^{(2)}+X_{2|1}^{(2)}\right]\right) \\ &= X_1^{(1)}\otimes\left[X_{2|1}^{(2)}+X_{1|1}^{(2)}\right]+\left[X_{2|1}^{(1)}+X_{3|1}^{(1)}\right]\otimes\left[X_{1|1}^{(2)}+X_{2|1}^{(2)}\right] \\ &= \mathcal{U}^{(1)}\otimes\mathcal{U}^{(2)}. \end{aligned} \tag{5.60}$$

Note that alternatively decomposing $\mathcal{U}^{(1)}$ as $X_1^{(1)}+X_{2|2}^{(1)}+X_{3|2}^{(1)}$, or decomposing $\mathcal{U}^{(2)}$ as $X_{1|2}^{(2)}+X_{2|2}^{(2)}$ (or both), will not affect this calculation, which reassures us that the action of $T^\dagger$ on the composite unit effect is well-defined. We find that the vector $\mathcal{U}^{(1)}\otimes X_{1|1}^{(2)}$ is transformed in the following manner:

$$\begin{aligned} T^\dagger\left(\mathcal{U}^{(1)}\otimes X_{1|1}^{(2)}\right) &= T^\dagger\left(X_1^{(1)}\otimes X_{1|1}^{(2)}+X_{2|1}^{(1)}\otimes X_{1|1}^{(2)}+X_{3|1}^{(1)}\otimes X_{1|1}^{(2)}\right) \\ &= X_1^{(1)}\otimes X_{2|1}^{(2)}+X_{2|1}^{(1)}\otimes X_{1|1}^{(2)}+X_{3|1}^{(1)}\otimes X_{1|1}^{(2)} \\ &= X_1^{(1)}\otimes\left[\mathcal{U}^{(2)}-X_{1|1}^{(2)}\right]+\left[\mathcal{U}^{(1)}-X_1^{(1)}\right]\otimes X_{1|1}^{(2)} \\ &= X_1^{(1)}\otimes\mathcal{U}^{(2)}+\mathcal{U}^{(1)}\otimes X_{1|1}^{(2)}-2X_1^{(1)}\otimes X_{1|1}^{(2)}. \end{aligned} \tag{5.61}$$

Again, this is well-defined in the sense that we get the same result if the alternative fiducial effect decomposition of $\mathcal{U}^{(1)}$ is used. It turns out that $X_1^{(1)}\otimes X_{1|1}^{(2)}$ and $\mathcal{U}^{(1)}\otimes X_{1|1}^{(2)}$ are the only two members of the tensor product basis not kept invariant by $T^\dagger$, as can be checked by similar calculations. For example,

$$\begin{aligned} T^\dagger\left(X_1^{(1)}\otimes\mathcal{U}^{(2)}\right) &= T^\dagger\left(X_1^{(1)}\otimes\left[X_{1|1}^{(2)}+X_{2|1}^{(2)}\right]\right) \\ &= X_1^{(1)}\otimes\left[X_{2|1}^{(2)}+X_{1|1}^{(2)}\right] \\ &= X_1^{(1)}\otimes\mathcal{U}^{(2)}. \end{aligned} \tag{5.62}$$

At this point we have defined a *linear map* $T^\dagger$, defined on $V_1\otimes V_2$ by its action on the tensor product basis resulting from the bases $\mathcal{B}^{(1)}$ and $\mathcal{B}^{(2)}$. Specifically, $T^\dagger$

maps:

$$X_1^{(1)} \otimes X_{1|1}^{(2)} \to X_1^{(1)} \otimes \mathcal{U}^{(2)} - X_1^{(1)} \otimes X_{1|1}^{(2)}$$
$$\mathcal{U}^{(1)} \otimes X_{1|1}^{(2)} \to X_1^{(1)} \otimes \mathcal{U}^{(2)} + \mathcal{U}^{(1)} \otimes X_{1|1}^{(2)} - 2X_1^{(1)} \otimes X_{1|1}^{(2)}, \tag{5.63}$$

and leaves all other basis vectors invariant. We must now check that $T^\dagger$ correctly permutes the composite fiducial effects in the manner originally specified, i.e. that it switches $X_1^{(1)} \otimes X_{1|1}^{(2)}$ and $X_1^{(1)} \otimes X_{2|1}^{(2)}$, whilst leaving all other composite fiducial effects invariant. This will in turn demonstrate that the specified permutation is allowed, since it necessarily corresponds to a linear map that maps the joint effect cone $\mathcal{E}_+^{max}$ into itself.

Note that many of the composite fiducial effects are also members of the tensor product basis, but not those with any component belonging to the set

$$\{X_{3|1}^{(1)}, X_{3|2}^{(1)}, X_{2|1}^{(2)}, X_{2|2}^{(2)}\}. \tag{5.64}$$

However, we may expand these non-basis composite fiducial effects in the tensor product basis, and hence check that $T^\dagger$ acts correctly on them. The effects $X_1^{(1)} \otimes X_{1|1}^{(2)}$ and $X_1^{(1)} \otimes X_{2|1}^{(2)}$ are indeed switched:

$$T^\dagger \left(X_1^{(1)} \otimes X_{1|1}^{(2)} \right) = X_1^{(1)} \otimes \mathcal{U}^{(2)} - X_1^{(1)} \otimes X_{1|1}^{(2)} = X_1^{(1)} \otimes X_{2|1}^{(2)} \tag{5.65}$$

$$\begin{aligned} T^\dagger \left(X_1^{(1)} \otimes X_{2|1}^{(2)} \right) &= T^\dagger \left(X_1^{(1)} \otimes \left[\mathcal{U}^{(2)} - X_{1|1}^{(2)} \right] \right) \\ &= X_1^{(1)} \otimes \mathcal{U}^{(2)} - \left(X_1^{(1)} \otimes \mathcal{U}^{(2)} - X_1^{(1)} \otimes X_{1|1}^{(2)} \right) \\ &= X_1^{(1)} \otimes X_{1|1}^{(2)} \end{aligned} \tag{5.66}$$

Moreover, the remaining composite fiducial effect vectors are invariant under $T^\dagger$.

For example,

$$\begin{aligned}
T^\dagger\left(X^{(1)}_{3|1}\otimes X^{(2)}_{2|1}\right) &= T^\dagger\left(\left[\mathcal{U}^{(1)}-X^{(1)}_{1}-X^{(1)}_{2|1}\right]\otimes\left[\mathcal{U}^{(2)}-X^{(2)}_{1|1}\right]\right)\\
&= \left[\mathcal{U}^{(1)}-X^{(1)}_{1}-X^{(1)}_{2|1}\right]\otimes\mathcal{U}^{(2)}+X^{(1)}_{2|1}\otimes X^{(2)}_{1|1}\\
&\quad + T^\dagger\left(X^{(1)}_{1}\otimes X^{(2)}_{1|1}-\mathcal{U}^{(1)}\otimes X^{(2)}_{1|1}\right)\\
&= \left[\mathcal{U}^{(1)}-X^{(1)}_{1}-X^{(1)}_{2|1}\right]\otimes\mathcal{U}^{(2)}+X^{(1)}_{2|1}\otimes X^{(2)}_{1|1}\\
&\quad + X^{(1)}_{1}\otimes X^{(2)}_{1|1}-\mathcal{U}^{(1)}\otimes X^{(2)}_{1|1}\\
&= X^{(1)}_{3|1}\otimes\left[X^{(2)}_{1|1}+X^{(2)}_{2|1}\right]-X^{(1)}_{3|1}\otimes X^{(2)}_{2|1}\\
&= X^{(1)}_{3|1}\otimes X^{(2)}_{1|1}. \qquad (5.67)
\end{aligned}$$

It is worth stressing the order of steps taken in this part of the proof. We were given a permutation of the joint fiducial effects, but with no guarantee that it corresponds to an allowed transformation. We then derived the form of the linear operation which - *if* the given permutation of fiducial effect vectors corresponds to an allowed transformation - must be the *unique* linear operator that performs that permutation. It was then necessary to verify that the linear operator does indeed permute the fiducial effect vectors in the correct manner.

2. Reversible.

Given that $T^\dagger$ as defined above is allowed, and permutes composite fiducial effects as in the statement of the Proposition, it is trivial to see that is is reversible. Composing $T^\dagger$ with itself gives an operator that leaves all composite fiducial effect vectors invariant. Since this set of vectors spans the space $V_1\otimes V_2$, it must be that $T^\dagger$ is its own inverse.

3. Non-trivial.

We now demonstrate that $T^\dagger$ is not a composition of local operations and swapping of subsystems (although, since the subsystems are non-identical, the latter is

impossible anyway). Recall our use of the Hamming distance between pairs of fiducial effect strings in Section 5.1, which is defined as the number of components in which those strings differ. Clearly, local operations and permutations of subsystems must preserve the Hamming distance between pairs of effects. However, consider the composite fiducial effects $e = X_1^{(1)} \otimes X_{1|1}^{(2)}$ and $f = X_{2|1}^{(1)} \otimes X_{1|1}^{(2)}$, which are mapped to the composite fiducial effects $T^\dagger(e) = X_1^{(1)} \otimes X_{2|1}^{(2)}$ and $T^\dagger(f) = f$. The Hamming distance between e and f is 1, whereas the Hamming distance between $T^\dagger(e)$ and $T^\dagger(f)$ is 2. Hence $T^\dagger$ is non-trivial. □

Further insight into why $T^\dagger$ is an allowed transformation can be gained by considering the tabular representation of the max-tensor product of a squashed g-trit and a g-bit system. Suppose that, instead of a squashed g-trit, a similar C-NOT transformation is applied to a max-tensor product of a g-trit and a g-bit system. A slight change of notation is needed, so that the composite fiducial effects $X_{1|1}^{(1)} \otimes X_{1|1}^{(2)}$ and $X_{1|1}^{(1)} \otimes X_{2|1}^{(2)}$ are switched; optionally, the effects $X_{1|2}^{(1)} \otimes X_{1|1}^{(2)}$ and $X_{1|2}^{(1)} \otimes X_{2|1}^{(2)}$ may also be switched. Determining the linear form of such a transformation becomes problematic when considering the tensor product basis vector $\mathcal{U}^{(1)} \otimes X_{1|1}^{(2)}$, due to the fact that a multiform effect is transformed into a non-multiform effect (recall that a bipartite binary effect table in Boxworld is multiform if and only if it contains a sub-row):

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]. \tag{5.68}$$

However, the effect table on the RHS of (5.68) *is* multiform if it represents the max-tensor product of a *squashed* g-trit and g-bit system. It is equivalent to the

following effect table:

$$\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}, \tag{5.69}$$

which can be seen from the fact that they are identical on pure product states (for which the first and fourth lines of the corresponding state table are identical).

It is not difficult to give an example in which $T^\dagger$ generates classical correlations, giving a second proof that $T^\dagger$ is non-trivial. The product state

$$(1/2, 1/2, 0 \mid 1/2, 1/2, 0) \otimes (1, 0|1, 0), \tag{5.70}$$

is transformed by $T^\dagger$ in the following manner:

$$\begin{array}{cc|cc} 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \end{array} \rightarrow \begin{array}{cc|cc} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \end{array}. \tag{5.71}$$

The state on the RHS is not a product state, but is rather the following convex combination of pure product states:

$$\frac{1}{2}(1, 0, 0|1, 0, 0) \otimes (0, 1|0, 1) + \frac{1}{2}(0, 1, 0|0, 1, 0) \otimes (1, 0|1, 0). \tag{5.72}$$

On the other hand, it can be shown that $T^\dagger$ does not generate any entanglement in this model. Recall that any reversible transformation must map pure states to pure states: for entanglement to be generated reversibly, at least one pure product state must be transformed to a pure entangled state, otherwise the set of convex

combinations of pure product states will be mapped into itself. We already know that the pure states of a g-bit system are the 4 states that assign deterministic outcomes for both fiducial measurement choices; a similar argument applies to the squashed g-trit, whose pure states are the 5 states that assign deterministic outcomes *and* give equal probabilities for the first outcome of both fiducial measurements. Therefore the state table representations of the pure product states can be enumerated, and it can be verified that $T^\dagger$ permutes the set of pure product states.

Indeed, we can follow this line of reasoning to conclude that *no* reversible transformation of the bipartite system we have described is capable of generating entanglement. Note that the state tables representing pure product states are made up of only 1s and 0s, and conversely that any state table consisting only of 1s and 0s represents a pure product state. In a similar vein to Proposition 7, we may therefore deduce that s is a pure product state if and only if $\langle e, s\rangle \in \{0, 1\}$ for all composite fiducial effects e. The adjoint of any reversible transformation T permutes the set of composite fiducial effects, hence for any pure product state s we have that $\langle e, T(s)\rangle = \langle T^\dagger(e), s\rangle \in \{0, 1\}$ for all composite fiducial effects e. Thus every reversible transformation T maps pure product states to pure product states, and does not generate entanglement.

5.6 Discussion

We have refined and extended the result of [4], and demonstrated that - as long as no subsystem is classical - reversible dynamics in general composite Boxworld systems always take the form of permutations of subsystems and local relabellings of measurement choices and outcomes on individual subsystems. This characterisation, and in particular the method of proof we develop, may be useful as tools for exploring which theories are reversible transitive. Note that we were already able to show in Proposition 7 of Section 2.5.5 that Boxworld is not a reversibly transitive theory, by proving that reversible transformations permute the set of pure product states. It is therefore worth highlighting in what ways it is useful to

provide an explicit characterisation of reversible Boxworld transformations.

One very useful outcome of this work is in developing a broader understanding of reversible dynamics; in particular, developing the study of reversible transformations in general probabilistic theories, and a framework for exploring the application of these techniques to theories other than Boxworld. In Section 5.5 we gave a result which limits the extent to which Theorem 13 may be generalised -- such a result would have been significantly harder to come by without the insights and perspective gained from our construction of the proof of Theorem 13 and the tabular formalism developed along the way. We have consequently shown that the analogous result to Theorem 13 does not generally hold in theories for which the number of local fiducial effects is finite, and for which composite systems combine under the max-tensor product. However, the local squashed g-trit system is not reversibly transitive, as the state which assigns outcome 1 to both fiducial measurements cannot be reversibly transformed to the state which assigns 2 to both fiducial measurements (it can be checked that the adjoint transformation will not be well-defined on the fiducial effect $X_1^{(1)}$, using the notation of Section 5.5).

A natural next step in this direction would be to investigate the set of allowed reversible transformations of the polygonal models given in Section 2.5.5. Such models also admit a finite number of fiducial effects, and in Section 5.5 we saw that an n-gon system may be viewed as a subset of the outcome distributions of a Boxworld system with n fiducial measurements, each of which has two outcomes. The case of even n presents an even greater similarity to Boxworld in that the extreme-ray effects are identical with the fiducial effects, thus improving the applicability of the techniques we have developed. Unfortunately, this equivalence between fiducial and extreme-ray effects is not true for odd n, where only half of the fiducial effects are extreme-ray effects.

At the beginning of the present chapter we discussed the centrality of reversible transitivity in derivations of quantum theory; an interesting open conjecture is that this feature alone is sufficient as a "physical axiom" from which to derive quantum theory. In order to prove such a conjecture it would be necessary to demonstrate

that any general probabilistic theory which is not in a well-defined sense "embedded in" quantum theory, is not reversibly transitive either. We may say that a theory is "embedded in" quantum theory if the systems of that theory can be mapped to quantum systems, and the effects and states mapped to positive operators and density operators respectively on those systems, in such a manner that composite systems are mapped to the corresponding compositions of quantum systems, and that the predictions (both local and composite) of the theory may equivalently be calculated by taking the trace of the product of the relevant quantum operators. For example, Boxworld is not contained within quantum theory, since it allows for a strictly greater violation of the CHSH inequality than quantum theory does.

From this perspective, the squashed g-trit example given in Section 5.5 may be seen as providing a counter-example to the strongest of three possible versions of this conjecture. In order of strongest to weakest, these versions of the conjecture may be explicitly stated as:

Strong Any theory with non-trivial reversible joint dynamics is embedded in quantum theory.

Medium Any theory which is reversibly transitive on local systems and has non-trivial, reversible, joint dynamics is embedded in quantum theory.

Weak Any theory which is reversibly transitive on local *and* composite systems is embedded in quantum theory.

Note that the bipartite system consisting of a max-tensor product of a squashed g-trit and a g-bit is not embedded in quantum theory, since it allows for maximal CHSH violations in much the same way that a max-tensor product of two g-bits does, via fiducial measurements on the following allowed state:

0	0	0	0
1/2	0	1/2	0
0	1/2	0	1/2
0	0	0	0
1/2	0	0	1/2
0	1/2	1/2	0

(5.73)

Thus the squashed g-trit provides a counter-example to the strong version of the conjecture, but not to the medium or weak versions. As already mentioned, it would be interesting to look at whether polygon models (which are locally reversibly transitive) give any insight into whether the medium conjecture holds in general. One result by Masanes *et al* [2] is in a similar vein to the Medium conjecture, and states that any general probabilistic theory whose local systems are qubit systems, and which admits at least one non-trivial, continuous, reversible interaction between systems, must also be identical to quantum theory on its composite systems. This suggests that structure of the local state space, in conjunction with reversible transitivity, may have a lot to tell us about the nature of quantum theory.

If it is true that all reversibly transitive theories are embedded in quantum theory, then quantum theory is the maximal reversibly transitive theory that could possibly describe Nature; this would go some way towards settling the matter of understanding why quantum theory is the most accurate description of Nature. On the other hand, if a "foil" theory is discovered, which is reversibly transitive but which makes predictions that quantum theory cannot, then this is of great interest in itself. Perhaps such a counterexample will eventually supersede quantum theory as a theory of Nature; at the very least it will give us some indication of what further axioms are necessary (rather than simply sufficient) to distinguish quantum theory from the full set of general probabilistic theories. Whichever outcome turns out to be true, the question of reversible transitivity is undoubtedly worthy of further investigation.

Chapter 6

Conclusion and Outlook

The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. We should be grateful for it and hope that it will remain valid in future research and that it will extend, for better or for worse, to our pleasure, even though perhaps also to our bafflement, to wide branches of learning.

"The Unreasonable Effectiveness of Mathematics"
Eugene Wigner

In this thesis we have derived various results which attempt to provide a more intuitive and comprehensive understanding of some of the stranger features of quantum theory. In Chapter 3 we investigated the remarkable fact that if one relaxes either one of two assumptions in quantum theory's Hilbert space formalism - that states be represented by positive operators, or that observables be represented by positive operators - then the predictions of quantum theory extend to the full class of non-signaling distributions, and hence encompass the full class of general probabilistic theories. We have also shown that the same feat is possible by introducing quasiprobabilies into local, classical outcome distributions, even in the case where almost no correlation exists between the involved systems. These results are useful in that they can be applied to the study of general probabilistic theories (for example, in deriving quantum theory from physical principles [2]) and that they provide a neat categorisation of quantum-achievable correlations as those which are generated so long as we *do* enforce the positivity of all operators.

Chapter 4 explores the uniqueness of quantum theory from the perspective of information theory. We prove the conjectured tightness of a bound on how well entanglement allows us to perform random access codes, and introduce a quadratic bias bound $\sum_{\mathbf{y}} E_{\mathbf{y}}^2 \leq 1$ which seems to capture a great deal of information about the set of quantum-achievable correlations. We also argue that the existence of a sensible measure of entropy precludes many general probabilistic theories whose non-locality is stronger than that of quantum theory, and have discussed how this relates to recent research on entropies in general probabilistic theories. Both the quadratic bias bound and the suggested link between entropy and non-locality throw up many intriguing open questions which may in future lead to fruitful lines of research.

Chapter 5 explores the role that reversible transitivity plays as a characteristic and fundamental feature of quantum theory. It is demonstrated that the triviality of reversible Boxworld transformations, previously only known in the case that all subsystems are identical [4], extends to the most general case, so long as none of the subsystems are classical. A key insight in this proof was to consider a special

class of effects known as sub-unit effects, and to show that the set of sub-unit effects is mapped to itself by the adjoint of any reversible transformation. As a visual aid to motivate the central idea of the proof, and to provide a simpler proof for the bipartite case, we introduced and further developed a tabular formalism of Boxworld states and effects first introduced in [5]. The techniques developed in this chapter may provide useful tools in exploring the reversible dynamics of other general probabilistic theories. We give one example of a model which is more non-local than quantum theory, but *does* allow for non-trivial reversible dynamics, although this model is not reversibly transitive. There exists a fascinating interplay between how non-local a theory is and the richness of its reversible dynamics. Further research might illuminate how the dynamical properties of the universe inform the strength of correlations allowed between distant parties.

The reader may make the justified observation that the content of each chapter is somewhat disparate, in that there is little interplay between their respective results, and in that they do not appear to constitute a linear progression towards some unified goal. Nevertheless, in the interests of tying together several years spent thinking about the same physical theory, it is worth analysing what similarities can be found amongst these chapters. There is one quite conspicuous feature that they share in common: each chapter constitutes an investigation of quantum theory "from the outside"; that is to say, as a special member of the class of general probabilistic theories. We may go one step further and say that the main results of each chapter are an attempt to mark out quantum theory from all other theories; to discern what makes quantum theory special in the first place.

Chapters 4 and 5 are each concerned with a fundamental property of quantum theory, the imposition of which renders one or more alternative theories as unnatural, or at least inconvenient. One moral that can be drawn equally well from both chapters is thus: if the world were not as described by quantum theory, then one or another fundamental physical aspect of the universe would be violated. These results strive towards a characterisation of quantum theory as the only reasonable theory that could possibly describe nature. Of course, we should always keep at

the back of our minds the possibility that quantum theory is not the most accurate description of nature, or indeed that the attempt to model nature in a way which is somehow agreeable to our common sense is futile, since nature might simply be too weird for us to grasp in a way that we find intuitive. Despite this reservation, we can take the optimistic view from these chapters (and other recent results in quantum foundations) that there is still more ground to be gained in obtaining an intuitive and reasonable explanation of quantum phenomena.

Chapter 3 stands apart from Chapters 4 and 5 in that it makes no real *judgement* about non-quantum theories. Rather, the results in this chapter explore the relation between the outcome distributions achievable in classical theory, quantum theory and Boxworld, by means of extending the normal rules of probability. The overarching moral of the chapter takes on a more negative tone concerning the uniqueness of quantum theory: it is not likely a special feature of quantum theory that all non-signaling distributions admit quasiprobabilistic quantum representations, and negative probabilities should probably be regarded as a useful calculational tool more than as a profound insight into the workings of the universe. However, it is very possible that future work on quasiprobabilistic representations will, for example, draw a stronger connection between non-locality and negative probabilities, in such a way as to provide a reasonable principle by which strongly non-local outcome distributions are ruled out. Even if this does not turn out to be the case, it is still useful to have a unified, quantum-like framework for non-signaling correlations, and to have a variety of local quasiprobabilistic models for these correlations.

Quantum non-locality has played almost as prominent a role in our discussions as have general probabilistic theories. Each chapter draws out interactions between non-locality and some other object of physical or mathematical interest, be it negative probability, information-based games, entropy, or reversible dynamics. An interesting take on the local quasiprobabilistic distributions of Chapter 3 is the idea that one is replacing the weirdness of non-locality with the weirdness of negative probabilities. Assigning positive probabilities to all possible outcomes

is clearly an assumption of Bell's Theorem; we have shown that Bell's Theorem does not hold if one just slightly relaxes to the extent that negative probabilities are assigned to outcomes which never actually occur. Despite its treatment by various notable physicists including Dirac [66] and Feynman [91], the idea of negative probabilities has never gained much traction. Is it simply a lesser degree of discomfort that seems to incline us towards accepting non-locality over negative probabilities? Perhaps our comfort *with* locality is the result of cultural and historical familiarity; if Newtonian physics had been written in a different conceptual language, would the language of quantum theory - being heavily influenced by Newtonian mechanics - have developed differently too, perhaps even having a general probabilistic theory slant from the beginning? These questions are difficult to answer, but as long as the presiding theories of nature admit distinct interpretations and distinct mathematical representations, it is in our own nature to differentiate these interpretations and representations according to our sense of reason, usefulness and beauty.

Bibliography

[1] A. Acín, R. Augusiak, D. Cavalcanti, C. Hadley, J. K. Korbicz, M. Lewenstein, Ll. Masanes, and M. Piani. Unified framework for correlations in terms of local quantum observables. *Phys. Rev. Lett.*, 104:140404, 2010.

[2] G. de la Torre, L. Masanes, A. J. Short, and M. P. Müller. Deriving quantum theory from its local structure and reversibility. *Phys. Rev. Lett.*, 109:090403, 2012.

[3] M. Pawlowski and M. Zukowski. Entanglement-assisted random access codes. *Phys. Rev. A*, 81:042326, 2010.

[4] D. Gross, M. P. Mueller, R. Colbeck, and O. C. O. Dahlsten. All reversible dynamics in maximally non-local theories are trivial. *Phys. Rev. Lett.*, 104:080402, 2010.

[5] A.J. Short and J. Barrett. Strong nonlocality: a trade-off between states and measurements. *New J. Phys.*, 12:033034, 2010.

[6] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802, 1982.

[7] S. Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. *Indiana Univ. Math. J.*, 17:59, 1967.

[8] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195, 1964.

[9] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777, 1935.

[10] D. Bohm and Y. Aharonov. Discussion of experimental proof for the paradox of Einstein, Rosen, and Podolsky. *Phys. Rev.*, 108:1070, 1957.

[11] 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.*, 49:91, 1982.

[12] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger. Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett.*, 81:5039, 1998.

[13] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland. Experimental violation of a Bell's inequality with efficient detection. *Nature*, 409:791, 2001.

[14] D. N. Matsukevich, P. Maunz, D. L. Moehring, S. Olmschenk, and C. Monroe. Bell inequality violation with two remote atomic qubits. *Phys. Rev. Lett.*, 100:150404, 2008.

[15] M. Giustina, A. Mech, S. Ramelow, B. Wittman, J. Kofler, J. Beyer, A. Lita, B. Calkins, T. Gerrits, S. W. Nam, R. Ursin, and A. Zeilinger. Bell violation using entangled photons without the fair-sampling assumtption. *Nature*, 497:227, 2013.

[16] S. Groblacher, T. Paterek, R. Kaltenbaek, C. Brukner, M. Zukowski, M. Aspelmeyer, and A. Zeilinger. An experimental test of non-local realism. *Nature*, 446:871, 2007.

[17] N. Gisin and H. Zbinden. Bell inequality and the locality loophole: Active versus passive switches. *Phys. Lett. A*, 264(2–3):103, 1999.

[18] D. Salart, A. Baas, J. A. W. van Houwelingen, N. Gisin, and H. Zbinden. Spacelike separation in a Bell test assuming gravitationally induced collapses. *Phys. Rev. Lett.*, 100:220404, 2008.

[19] T. Scheidl, R. Ursin, J. Kofler, S. Ramelow, X.-S. Ma, T. Herbst, L. Ratschbacher, A. Fedrizzi, N. K. Langford, T. Jennewein, and A. Zeilinger. Violation of local realism with freedom of choice. *Proc. Natl. Acad. Sci. U.S.A.*, 107(46):19708, 2010.

[20] G. Blaylock. The E. P. R. paradox, Bell's inequality, and the question of locality. *Am. J. Phys.*, 78:111, 2009.

[21] T. Maudlin. What Bell proved: A reply to Blaylock. *Am. J. Phys.*, 78:121, 2010.

[22] P. H. Eberhard. Bell's theorem without hidden variables. *Nuovo Cimento B*, 38:75, 1977.

[23] T. Norsen. Against 'realism'. *Found. Phys.*, 37(3):311--340, 2007.

[24] T. Maudlin. Space-time in the quantum world. In *Bohmian mechanics and quantum theory: an appraisal*. Kluwer Academic Publishers, 1996.

[25] P. M. Pearle. Hidden-variable example based upon data rejection. *Phys. Rev. D*, 2:1418, 1970.

[26] A. Garg and N. D. Mermin. Detector inefficiencies in the Einstein-Podolsky-Rosen experiment. *Phys. Rev. D*, 35:3831, 1987.

[27] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67:661, 1991.

[28] L. Hardy. Quantum theory from five reasonable axioms, 2001. *arXiv:quant-ph/0101012*.

[29] J. Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75:032304, 2007.

[30] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A*, 71:022101, 2005.

[31] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Cloning and broadcasting in generic probabilistic theories, 2006. *arXiv:quant-ph/0611295*.

[32] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Probabilistic theories with purification. *Phys. Rev. A*, 81:062348, 2010.

[33] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84:012311, 2011.

[34] L. Masanes and M. P. Müller. A derivation of quantum theory from physical requirements. *New J. Phys.*, 13:063001, 2011.

[35] H. Barnum, M. P. Mueller, and C. Ududec. Higher-order interference and single-system postulates characterizing quantum theory, 2014. *arXiv:1403.4147 [quant-ph]*.

[36] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner. Limits on nonlocal correlations from the structure of the local state space. *New J. Phys.*, 13:063024, 2011.

[37] C. Pfister. One simple postulate implies that every polytopic state space is classical, 2012. (Master's Thesis, Institute for Theoretical Physics, ETH Zurich) *arXiv:1203.5622 [quant-ph]*.

[38] A. J Short and S. Wehner. Entropy in general physical theories. *New J. Phys.*, 12:033023, 2010.

[39] G. Kimura, K. Nuida, and H. Imai. Distinguishability measures and entropies for general probabilistic theories. *Rep. M. Phys.*, 66:175, 2010.

[40] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke. Entropy and information causality in general probabilistic theories. *New J. Phys.*, 12:033024, 2010.

[41] G. M. D'Ariano. No-signaling, dynamical independence, and the local observability principle. *J. Phys. A*, 40:8137, 2007.

[42] A. J. Short. No purification for two copies of a noisy entangled state, 2008. *arXiv:0809.2622 [quant-ph]*.

[43] S. W. Al-Safi and A. J. Short. Simulating all nonsignaling correlations via classical or quantum theory with negative probabilities. *Phys. Rev. Lett.*, 111:170403, 2013.

[44] S. W. Al-Safi and A. J. Short. Information causality from an entropic and a probabilistic perspective. *Phys. Rev. A*, 84:042323, 2011.

[45] S. W. Al-Safi and A. J. Short. Reversible dynamics in strongly non-local boxworld systems. *J. Phys. A: Math. Theor.*, 47:325303, 2014.

[46] J. Clauser, M. Horne, A. Shimony, and R. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880, 1969.

[47] B. Tsirelson. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.*, 4:93, 1980.

[48] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Found. Phys.*, 24:379, 1994.

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

[50] T. H. Yang, M. Navascués, L. Sheridan, and V. Scarani. Quantum bell inequalities from macroscopic locality. *Phys. Rev. A*, 83:022105, 2011.

[51] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski. Information causality as a physical principle. *Nature*, 461:1101, 2009.

[52] N. Linden, S. Popescu, A. J. Short, and A. Winter. Quantum nonlocality and beyond: Limits from nonlocal computation. *Phys. Rev. Lett.*, 99:180502, 2007.

[53] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

[54] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Teleportation in general probabilistic theories. In *Mathematical Foundations of Information Flow (Proceedings of the Clifford Lectures 2008)*, page 25. American Mathematical Society, 2012.

[55] M. P. Müller and C. Ududec. The structure of reversible computation determines the self-duality of quantum theory. *Phys. Rev. Lett.*, 108:130401, 2012.

[56] H. Barnum, C. P. Gaebler, and A. Wilce. Ensemble steering, weak self-duality, and the structure of probabilistic theories, 2009. *arXiv:0912.5532 [quant-ph]*.

[57] A. Wilce. Symmetry, self-duality and the jordan structure of quantum mechanics, 2011. *arXiv:1110.6607 [quant-ph]*.

[58] C. Piron. Axiomatique quantique. *Helv. Phys. Acta*, 37:439, 1964.

[59] D.J. Foulis and C.H. Randall. Empirical logic and quantum mechanics. *Synthese*, 29(1-4):81, 1974.

[60] B. Coecke S. Abramsky. A categorical semantics of quantum protocols. In *Proceedings of the 19th IEEE conference on Logic in Computer Science*, page 415. IEEE Computer Science Press, 2004.

[61] R. Webster. *Convexity*. Oxford University Press, Oxford, 1994.

[62] C. Carathéodory. Über den variabilitätsbereich der fourier'schen konstanten von positiven harmonischen funktionen. *Rend. Circ. Mat. Palermo*, 32(1):193, 1911.

[63] B. Grünbaum. *Convex Polytopes*. Interscience Publishers, 1967.

[64] N. S. Jones and L. Masanes. Interconversion of nonlocal correlations. *Phys. Rev. A*, 72:052312, 2005.

[65] S. Pironio, J.-D. Bancal, and V. Scarani. Extremal correlations of the tripartite no-signaling polytope. *J. Phys. A: Math Theor.*, 44:065303, 2011.

[66] P. A. M. Dirac. Bakerian lecture. the physical interpretation of quantum mechanics. *Proc. R. Soc. A*, 180(980):1, 1942.

[67] E. Wigner. On the quantum correction for thermodynamic equilibrium. *Phys. Rev.*, 40:749, 1932.

[68] C. Ferrie. Quasi-probability representations of quantum theory with applications to quantum information science. *Rep. Prog. Phys.*, 74(11):116001, 2011.

[69] T. J. Barnea, J.-D. Bancal, Y.-C. Liang, and N. Gisin. Tripartite quantum state violating the hidden-influence constraints. *Phys. Rev. A*, 88:022123, 2013.

[70] H. Buhrman, M. Christandl, F. Unger, S. Wehner, and A. Winter. Implications of superstrong nonlocality for cryptography. *Proc. Roy. Soc. A*, 462:1919, 2006.

[71] A. Ambainis, D. Leung, L. Mancinska, and M. Ozols. Quantum random access codes with shared randomness, 2008. (Master's thesis, University of Waterloo) *arXiv:0810.2937 [quant-ph]*.

[72] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger Limit on nonlocality in any world in which communication complexity is not trivial. *Phys. Rev. Lett.*, 96:250401, 2006.

[73] J. Allcock, N. Brunner, M. Pawlowski, and V. Scarani. Recovering part of the boundary between quantum and nonquantum correlations from information causality. *Phys. Rev. A*, 80:040103, 2009.

[74] M. Navascués, S. Pironio, and A. Acín. Bounding the set of quantum correlations. *Phys. Rev. Lett.*, 98:010401, 2007.

[75] M. Navascues, S. Pironio, and A. Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New J. Phys.*, 10:073013, 2008.

[76] R. Franco and V. Penna. Discrete wigner distribution for two qubits: a characterization of entanglement properties. *J. Phys. A*, 39(20):5907, 2006.

[77] J. Degorre, M. Kaplan, S. Laplante, and J. Roland. The communication complexity of non-signaling distributions. *Quantum Info. Comput.*, 11:649, 2011.

[78] E. R. Loubenets. Local quasi hidden variable modelling and violations of bell-type inequalities by a multipartite quantum state. *J. Math. Phys.*, 53:022201, 2012.

[79] E. R. Loubenets. Nonsignaling as the consistency condition for local quasi classical probability modelling of a general multipartite correlation scenario. *J. Phys. A: Math. Theor.*, 45:185306, 2012.

[80] R. Cleve, P. Høyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. In *19th Annual IEEE Conference on Computational Complexity*, page 236. IEEE, 2004.

[81] P. K. Aravind. A simple demonstration of bell's theorem involving two observers and no probabilities or inequalities, 2002. *arXiv:quant-ph/0206070v2*.

[82] J. Briet, H. Buhrman, T. Lee, and T. Vidick. Multiplayer xor games and quantum communication complexity with clique-wise entanglement, 2009. *arXiv:0911.4007 [quant-ph]*.

[83] M. L. Almeida, J.-D. Bancal, N. Brunner, A. Acín, N. Gisin, and S. Pironio. Guess your neighbor's input: A multipartite nonlocal game with no quantum advantage. *Phys. Rev. Lett.*, 104:230404, 2010.

[84] W. van Dam. Implausible consequences of superstrong nonlocality. *Natural Computing*, 12:9, 2013.

[85] L. Masanes, A. Acin, and N. Gisin. General properties of nonsignaling theories. *Phys. Rev. A*, 73:012112, 2006.

[86] B. Tsirelson. Quantum analogues of Bell inequalities: The case of two spatially separated domains. *J. Sov. Math.*, 36:557, 1987.

[87] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Quantum theory, namely the pure and reversible theory of information. *Entropy*, 14:1877, 2012.

[88] P. Skrzypczyk, A. J. Short, and S. Popescu. Work extraction and thermodynamics for individual quantum systems. *Nat. Commun.*, 5:4185, 2014.

[89] N. Brunner, M. Kaplan, A. Leverrier, and P. Skrzypczyk. Dimension of physical systems, information processing, and thermodynamics, 2014. *arXiv:1401.4488 [quant-ph]*.

[90] M. Huber, M. Perarnau-Llobet, K. V. Hovhannisyan, P. Skrzypczyk, C. Klöckl, N. Brunner, and A. Acín. Thermodynamic cost of creating correlations, 2014. *arXiv:1404.2169 [quant-ph]*.

[91] R. P. Feynman. Negative probability. In F. David Peat and Basil Healy, editors, *Essays in Honour of David Bohm*. Routledge & Kegan Paul Ltd, London, 1987.