The Mathematical Structure of Non-locality & Contextuality

DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF OXFORD

Shane Mansfield

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2013

To my family.

Acknowledgements

I would like to thank my supervisors Samson Abramsky and Bob Coecke for taking me under their wing, and for their encouragement and guidance throughout this project; my collaborators Samson Abramsky, Tobias Fritz and Rui Soares Barbosa, who have contributed to some of the work contained here; and all of the members of the Quantum Group, past and present, for providing such a stimulating and colourful environment to work in. I gratefully acknowledge financial support from the National University of Ireland Travelling Studentship programme.

I'm fortunate to have many wonderful friends, who have played a huge part in making this such an enjoyable experience and who have been a great support to me. A few people deserve a special mention: Rui Soares Barbosa, who caused me to and kept me from 'banging two empty halves of coconut together' at various times, and who very generously proof-read this dissertation; Daniel 'The Font of all Knowledge' Corbett; Tahir Mansoori, who looked after me when I was laid up with a knee injury; mo chara dhílis, fear mór na Gaeilge, Daithí Ó Ríogáin; Johan Paulsson, my companion since Part III; all of my friends at Wolfson College, you know who you are; Alina, Michael, Jon & Pernilla, Matteo; Bowsh, Matt, Lewis; Antón, Henry, Vincent, Joe, Ruth, the Croatians; Jerry & Joe at the lodge; the Radish, Ray; my flatmates; again, the entire Quantum Group, past and present; the Cambridge crew, Raph, Emma & Marion; Colm, Dave, Goold, Samantha, Steve, Tony and the rest of the UCC gang; the veteran Hardy Bucks, Rossi, Diarmuid & John, and the Abbeyside lads; Brendan, Cormac, and the organisers and participants of the PSM.

I'd also like to thank my teachers, who have had a great influence on me down through the years: especially Mary Walsh, Eddie O'Halloran, Oliver Broderick, Frank Galvin, Michel Vandyck and Niall Ó Murchadha.

A very special thanks is due to Sophie Walon, who was an *étoile* for me during the difficult period of writing-up, and kept me motivated, healthy and sane.

Finally, the biggest thanks of all goes to my family, for their constant love, support and encouragement from the very beginning: to my mum, Joan, and my 'aunt-mother' Eileen, who have always encouraged me to follow my own path; to my father, Jimmy, and uncle Eamonn, who set me up successively in Cambridge and Oxford, and in typical fashion did me the great service of proof-reading the dissertation; to my brothers, Daragh and Ronan, who are an inspiration to me; to my loving grandmother and godmother Nora Cooney; to my aunt Sheila, uncle Jerry, and all the rest of my extended family; to my Nana Mansfield and my aunt Nora, who would have been proud to see this; also to Eilish and to all the family friends who have taken such an interest in my education over the years.

Abstract

Non-locality and contextuality are key features of quantum mechanics that distinguish it from classical physics. We aim to develop a deeper, more structural understanding of these phenomena, underpinned by robust and elegant mathematical theory with a view to providing clarity and new perspectives on conceptual and foundational issues. A general framework for logical non-locality is introduced and used to prove that 'Hardy's paradox' is complete for logical non-locality in all $(2, 2, l)$ and $(2, k, 2)$ Bell scenarios, a consequence of which is that Bell states are the only entangled two-qubit states that are not logically non-local, and that Hardy non-locality can be witnessed with certainty in a tripartite quantum system. A number of developments of the unified sheaf-theoretic approach to non-locality and contextuality are considered, including the first application of cohomology as a tool for studying the phenomena: we find cohomological witnesses corresponding to many of the classic no-go results, and completely characterise contextuality for large families of Kochen-Specker-like models. A connection with the problem of the existence of perfect matchings in k-uniform hypergraphs is explored, leading to new results on the complexity of deciding contextuality. A refinement of the sheaf-theoretic approach is found that captures partial approximations to locality/non-contextuality and can allow Bell models to be constructed from models of more general kinds which are equivalent in terms of non-locality/contextuality. Progress is made on bringing recent results on the nature of the wavefunction within the scope of the logical and sheaf-theoretic methods. Computational tools are developed for quantifying contextuality and finding generalised Bell inequalities for any measurement scenario which complement the research programme. This also leads to a proof that local ontological models with 'negative probabilities' generate the no-signalling polytopes for all Bell scenarios.

Contents

Introduction

At a fundamental level, non-locality and contextuality are key features of quantum mechanics that confound classical intuitions. It was realised early on that the theory displayed certain non-intuitive features: they gave rise to apparent paradoxes such as Schrödinger's cat [101] and quantum 'steering' [97], and led to the Einstein-Podolsky-Rosen argument [48] for the incompleteness of quantum mechanics. The classic no-go theorems of Bell [22], Kochen & Specker [73] et al., however, showed that non-locality and contextuality are necessary features of any theory that agrees with the predictions of quantum mechanics.

While these features are challenging from a conceptual point of view, they have opened the door to radical new possibilities. Bell's insights in particular have been key to developments in quantum information theory, which has grown up around the idea that entanglement and non-locality are a resource that can be exploited. This has led to some remarkable results, including Shor's algorithm [99], which can factor integers in polynomial time, quantum cryptography [26], and the teleportation protocol [25]. More recently, there has also been much work on the experimental realisation of contextuality [21, 71], for which one might hope similar applications can be found.

This dissertation is concerned with understanding the mathematical structure of non-locality and contextuality. Gaining a deeper, structural understanding of these phenomena, underpinned by robust and elegant mathematical theory, is important for a number of reasons. It can provide clarity and new perspectives on conceptual and foundational issues; it exposes connections with diverse fields in which similar structures arise in non-physical contexts, raising interesting possibilities for the transfer of methods and insights in both directions; eventually, one also hopes that it can lead to a more systematic approach to harnessing and utilising both non-locality and contextuality as resources.

Non-locality and contextuality are properties of the correlations or 'empirical models' that arise from the operational predictions of quantum mechanics. Abramsky &

Brandenburger showed that empirical models can be precisely described in sheaf-theoretic terms, and moreover that a very natural unified characterisation of locality and non-contextuality emerges in this setting [4]. This is the language that will be used throughout the dissertation, and is described in detail in chapter 1. Another consequence of the sheaf-theoretic description is the emergence of a hierarchy of non-locality/contextuality:

$$\text{Strong Contextuality} > \text{Logical Contextuality} > \text{Contextuality}.$$

Chapter 2 builds on work published in [79]. It presents a general framework for logical non-locality, which is a precursor to the more general sheaf-theoretic approach and is expressed in similar terms. An advantage to our logical framework is that it comes equipped with a particular representation that provides a powerful means of reasoning about empirical models. This leads to several interesting results. 'Hardy's paradox' [59, 60] is considered to be the simplest non-locality proof for quantum mechanics. We prove a number of completeness theorems which show that it provides a necessary and sufficient condition for logical non-locality in all $(2, 2, l)$ and $(2, k, 2)$ scenarios. It will be seen that these have many consequences and applications. These include a proof that maximally entangled two-qubit states are the only entangled two-qubit states which are not logically non-local. This is surprising since they are perhaps the most studied and utilised of entangled states, even though in this light they appear to be anomalous in terms of non-locality. Much of the literature on Hardy's paradox is concerned with the probability of witnessing a paradox, which has experimental motivations: the highest probability to date is ≈ 0.4 [37]. We also achieve a striking improvement on this, demonstrating that it is possible to witness Hardy non-locality with certainty for a tripartite quantum system.

Chapters 3 and 4 are both concerned with developments of the sheaf-theoretic approach. Non-locality and contextuality are characterised by obstructions to the existence of global sections of empirical models represented on presheaves. Cohomology theories can roughly be thought of as descriptions of obstructions to solving some kind of equation. In chapter 3 we attempt to apply the powerful tools of presheaf cohomology to witness and characterise non-locality and contextuality. The possible use of cohomology to study contextuality in the sense of the Kochen-Specker theorem was first suggested by Isham & Butterfield [68], and this work represents the first progress in this direction. Indeed, we succeed in finding cohomological witnesses of non-locality and contextuality corresponding to many of the classic no-go results.

While the approach is not yet strong enough to characterise contextuality in all models, it can be shown that it yields a complete invariant for contextuality for large families of Kochen-Specker-like models. A connection is also found between contextuality of empirical models and the problem of the existence of perfect matchings in k-uniform hypergraphs, which has been much studied in the mathematics literature, and which leads to results on the complexity of deciding contextuality that are new to the foundations of quantum mechanics.

In chapter 4, the notion of extendability which was shown by Abramsky & Brandenburger to correspond in a unified manner to non-locality and contextuality is refined. This captures partial approximations to locality and non-contextuality and can be useful in characterising the properties of sub-models of an empirical model. The refinement also has another useful application. On practical and foundational levels, the notion of locality in Bell models can more easily be motivated than the corresponding general notion of contextuality. It is shown that a particular, canonical extension, when well-defined, may be used for the construction of Bell models from models of more general kinds in such a way that the constructed model is equivalent in terms of non-locality/contextuality. This construction can be carried out for the Kochen-Specker-like models, which throws up some interesting connections between contextual and non-local models: in particular it relates the simplest possible contextual model, the contextual triangle of Specker's parable [75], with the Popescu-Rohrlich no-signalling correlations [93]. It also suggests a route to proposing Bell tests that correspond to contextuality proofs.

Chapter 5 contains some initial work on attempting to bring recent developments in the foundations of quantum mechanics concerning the nature of the wavefunction within the scope of the logical and structural methods that are set out in the dissertation. As a first step, this involves generalising and reformulating a criterion for the reality of the wavefunction proposed by Harrigan & Spekkens [63], which is central to the PBR theorem [94]. The new criterion has several advantages, including the avoidance of certain technical difficulties. By considering the reality not of the wavefunction but of the observable properties of any ontological physical theory a novel characterisation of non-locality and contextuality is found. A careful analysis of one of the key assumptions of the PBR theorem also leads to some insights on the development of a sheaf-theoretic approach to ontological theories.

Finally, while many of the topics dealt with throughout the dissertation are of quite a theoretical nature, chapter 6 demonstrates that computational exploration can play an important role in the research programme. A number of computational tools have been developed and have been implemented as a *Mathematica* package. These cover the calculation of quantum empirical models, and a computational approach to calculating the degree of contextuality and to finding logical Bell inequalities which is applicable to any measurement scenario (not just to Bell scenarios) using linear programming. This provides a useful setting for formulating and testing conjectures. One particularly interesting result in which computational exploration has played an important role shows that local ontological models with 'negative probabilities' generate the no-signalling polytopes for all Bell scenarios.

Chapter 1

The Sheaf-theoretic Framework

Any physical theory must make predictions for empirical observations. We will refer to any (possibly hypothetical) set of empirical observations, or any set of theoretical predictions for empirical observations, as an *empirical model*, an example of which is the following.

	00	01	10	11
A B	$1/2$	0	0	$1/2$
A B'	$3/8$	$1/8$	$1/8$	$3/8$
A' B	$3/8$	$1/8$	$1/8$	$3/8$
A' B'	$1/8$	$3/8$	$3/8$	$1/8$

This should be read as saying that, if measurements A and B are made jointly, then the probability of A having outcome 0 and B having outcome 0 is $1/2$, etc. This empirical model, which we will return to shortly, arises from the CHSH formulation [40, 24] of Bell's theorem [22]. As the example shows, an empirical model can provide data for joint observations. The data might be probabilistic, as in this case, or deterministic. We will be particularly concerned with empirical models of the kind in which measurements have discrete spectra of outcomes, for the reasons that quantum mechanics gives rise to discrete empirical models, and that the features we are interested in already exhibit themselves at this level.

Non-locality and contextuality are features of correlations in empirical models that contradict the intuitions underlying classical physics. They arise, in particular, in certain quantum mechanical predictions and can be confirmed by empirical observation. A simple example of the non-intuitive nature of these features will be presented in section 2.1.

The first step to a deeper and more structural understanding of non-locality and contextuality is to adopt an appropriate framework and language for dealing with empirical models, and these features in particular. An early approach was the hidden variable framework, which will be encountered in chapter 5. We will introduce a logical framework for non-locality in chapter 2, which is a precursor to the more general unified sheaf-theoretic framework for non-locality and contextuality due to Abramsky & Brandenburger [4]. The unified approach can be shown to subsume the others and is central to the dissertation. This chapter presents an overview of the main ideas of the approach. The approach itself is further developed in chapters 3 and 4.

1.1 Empirical Models

States and Observables

Many of the empirical models that we will be concerned with arise from quantum mechanics. One kind of quantum mechanical empirical model can be obtained by choosing a state and observables and then calculating the expectation values of the various outcomes. For example, we could specify the following two-qubit Bell state

$$\left|\phi^+\right\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B\right)$$

and all pairs of local measurements, where

$$A = B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad A' = B' = \begin{pmatrix} 0 & e^{-i\frac{\pi}{3}} \\ e^{i\frac{\pi}{3}} & 0 \end{pmatrix}$$

are the available measurements on the respective qubits. The model obtained in this case is the Bell-CHSH model from before.

State-independent Models

Another kind of quantum mechanical empirical model is the *state-independent* empirical model, an example of which arises from the Kochen-Specker theorem [73]. We will refer to the simpler, 18-vector proof of the theorem in $\mathbb{R}^4$ [34]. It is shown here that for any state it is always possible to choose 18 vectors (measurements) labelled $A, \ldots, R$ with the following properties:

- Compatible sets of measurements consist of mutually orthogonal subsets of the vectors. These are the columns of the table below. Jointly, each compatible set defines a projective quantum measurement.

A	A	H	H	B	I	P	P	Q
B	E	I	K	E	K	Q	R	R
C	F	C	G	M	N	D	F	M
D	G	J	L	N	O	J	L	O

Joint outcomes assign 1 to the vector onto which the state has been projected, and 0 to all other vectors.

- The probability distribution arising from each compatible set of measurements has the same form. There are non-zero probabilities $\{p_i\}_{i=1}^4$corresponding to the outcomes $\{1000, 0100, 0010, 0001\}$, respectively, such that $\sum_{i=1}^4 p_i = 1$, as in the following example. The precise values of the probabilities need not be known.

	1000	0100	0010	0001
$A\ B\ C\ D$	p_1	p_2	p_3	p_4

State-independent models, therefore, are more general in the obvious sense that they hold for any state. On the other hand, they do not contain precise probabilistic information, effectively only indicating which of the outcomes are possible and which are impossible. Nevertheless, as we will see, non-locality and contextuality can already exhibit themselves at this level.

No-signalling

No-signalling is a property that is satisfied by all correlations that arise from quantum mechanics in either of these ways. This was originally observed in relation to compound systems [53], where it can be seen to be a straightforward consequence of the tensor product structure [70]. It states that if a joint measurement is made then the probabilities of the various outcomes to a measurement on one sub-system should not depend on which measurements are made elsewhere. It is clear that in the case of spatially distributed systems, such behaviour could lead to superluminal signalling; one experimenter could measure her subsystem and immediately affect the probabilities of the outcomes to measurements made by another experimenter on a different subsystem.

However, it is not difficult to show that this is true more generally of any correlations arising from joint measurements of commuting observables in quantum mechanics: this property has been referred to as generalised no-signalling [4] or no-disturbance [95]. One way of stating this is that for any empirical model predicted by quantum mechanics, marginal probability distributions are well-defined. For example, with reference to the Bell-CHSH model, we can speak of the marginal probability distribution

$$p(o_A \mid A) := p(o_A \mid A, B) = p(o_A \mid A, B')$$

where $p(o_A \mid A, B) := \sum_{o_B} p(o_A, o_B \mid A, B)$ 'forgets' the outcome of the second measurement.

Confusion often surrounds this property and its relationship to relativity. First of all, it should be noted that quantum mechanics is a non-relativistic theory. It is true that the property forbids superluminal signalling through the measurement process; but in fact it imposes something even stronger, since it also holds for compatible measurements on a system which is not spatially distributed. It should also be noted that the analogous form of no-signalling holds in classical mechanics. Values of observables in a classical system are represented functions of the system's phase space. Choosing to evaluate an observable at a particular point in phase space does not in any way alter the value of another observable at that point. The non-relativistic feature of classical mechanics is that it allows instantaneous action-at-a-distance: a change of potential instantaneously affects a particle anywhere in classical space. There is a similar action-at-a-distance in non-relativistic quantum mechanics, in terms of potentials, but also (at least in the standard formulation) in terms of collapse of the wavefunction. An attempt at a non-relativistic motivation for the property is contained in [6].

No-signalling does not characterise quantum correlations: there exist no-signalling correlations that cannot be realised by any quantum system: the Popescu-Rohrlich correlations [93], for example. Generally speaking, we will assume no-signalling as a minimum requirement of the empirical models we will be interested in.

1.2 Presheaves & Sheaves

This section contains some basic mathematical background concerning presheaves and sheaves. These are the structures that we will use to describe empirical models. Sheaf theory is pervasive in modern mathematics, allowing the passage from local to global [77]. For the present purposes it suffices to restrict our attention to presheaves and sheaves on a poset. The posets we will be concerned with consist of subsets of some set X ordered by subset inclusion.

Definition 1.2.1. *A* presheaf on a poset $\mathbf{P}$ *is a functor*

$$F : \mathbf{P}^{\mathrm{op}} \to \mathbf{Set}$$

(or, equivalently, a contravariant functor $F : \mathbf{P} \to \mathbf{Set}$*) where* $\mathbf{P}$ *is regarded as a category.*

The objects of the category $\mathbf{P}$ are just the elements of the set $\mathbf{P}$, and there exists a morphism $i_{p,p'} : p \to p'$ whenever $p \leq p'$. We call these *inclusion maps*. Then F assigns a set $F(p)$ to each element $p \in \mathbf{P}$ and a *restriction map* $F(i_{p,p'}) : F(p') \to F(p)$ to each inclusion map $i_{p,p'}$. Functoriality of these assignments implies that

$$F(i_{p,p}) = \mathrm{id}_{F(p)}$$

for all $p \in \mathbf{P}$, and

$$F(i_{p,p''}) = F(i_{p',p''}) \circ F(i_{p,p'})$$

whenever $p \leq p' \leq p''$. Elements of $F(p)$ are called *sections*, and we will use the notation $s|_p := F(i_{p,p'})(s)$ for a restriction of a section $s \in F(p')$. If there exists a top element $\top \in \mathbf{P}$, then a section $s \in F(\top)$ is called a *global section*.

Example 1.2.2. *For any poset* $\mathbf{P}$*, we can define a presheaf* $F : \mathbf{P}^{\mathrm{op}} \to \mathbf{Set}$ *by* $F(p) := \{p' \in \mathbf{P} \mid p' \leq p\}$ *for all* $p \in \mathbf{P}$ *and* $F(p)|_q := \{p' \in F(p) \mid p' \leq q\}$ *for all* $q \in \mathbf{P}$ *such that* $q \leq p$.

A bounded complete poset $\mathbf{P}$ is a poset in which all bounded sets $\{p_j\}_{j\in J}$ have a least upper bound or join $\bigvee_{j\in J} p_j$. For a poset $\mathcal{U} \subseteq \mathcal{P}(X)$ consisting of subsets of some set X ordered by subset inclusion, bounded completeness corresponds to the closure of $\mathcal{U}$ under countable unions.

Definition 1.2.3. *A presheaf on a poset* $\mathbf{P}$ *is a* sheaf *if whenever* $p = \bigvee_{j\in J} p_j$ *and there exists a family of sections* $\{s_j\}_{j\in J}$*, with* $s_j \in F(p_j)$ *for each* $j \in J$*, which satisfies the* compatibility *condition:*

$$\forall\, j,k \in J.\; s_j|_{p_j \wedge p_k} = s_k|_{p_j \wedge p_k},$$

then there exists a section $s \in F(p)$ *such that* $s|_{p_j} = s_j$ *for all* $j \in J$.

A useful intuition is that a presheaf assigns information to a poset in such a way that the assignment for a particular element can be restricted to lower elements in a consistent way. A sheaf has the additional property that if assignments exist and are locally compatible on everything below a particular element then they can be glued or lifted to provide an assignment on that element. The presheaf defined in example 1.2.2 is also a sheaf.

The relevance of these structures to contextuality in the sense of the Kochen-Specker theorem is that it is possible to assign values to certain properties of a quantum system (those measured by the vectors $A, \dots R$) in a way that is consistent over contexts (the sets of compatible measurements) but that cannot be lifted to a global assignment of values to all of these properties at once. Analogous, intuitive examples are the Penrose triangle (figure 1.1) and the Penrose stairs, which are locally but not globally constructible. Indeed, one could present these examples as families of sections on appropriate presheaves which do not arise as restrictions of any global section.

Example 1.2.4. *For the triangle, we could label the edges $\{A, B, C\}$, take as poset subsets of the edges labelled by inclusion, and define a presheaf F that for each subset of edges gives all possible strict total orderings of those edges: e.g.*

$$F(\{A, B\}) = \{A > B, \quad B > A\}.$$

Restrictions arise in the obvious way. If we interpret '$>$' as 'appears closer than' then the Penrose triangle would represent a family of sections

$$\{s_{\{A,B\}} = B > A, \quad s_{\{B,C\}} = C > B, \quad s_{\{C,A\}} = A > C\},$$

which cannot arise from restrictions of any global section $s_{\{A,B,C\}}$, which in this case would be a strict total order on $\{A, B, C\}$.

The Kochen-Specker theorem was first expressed in the language of presheaves by Isham & Butterfield in [68], which instigated the topos approach to physics. While there are some similarities between the topos approach and the sheaf-theoretic approach we are about to set out, we note that there are several key differences. The topos approach deals with contextuality, but is primarily concerned with the spectral presheaf, which is derived from an operator algebra, and thus heavily incorporates much of the mathematical structure of quantum mechanics from the outset. The present approach will avoid this, and assumes a minimum of quantum mechanical baggage. It will therefore provide an elegant language for the discussion of non-local

Figure 1.1: The Penrose Triangle.

and contextual correlations in a more general setting that remains neutral with regard to any underlying physical theory.

1.3 The Framework

With these examples of empirical models in mind, we set out the sheaf-theoretic framework. We assume sets X of measurements and O of outcomes. There is an additional structure on the set of measurements, a cover $\mathcal{M}$ over X, which specifies the sets of compatible measurements: we think of these as sets of measurements that can be performed jointly. In quantum mechanics, for example, this structure would arise as the commutative subalgebras of the algebra of observables.

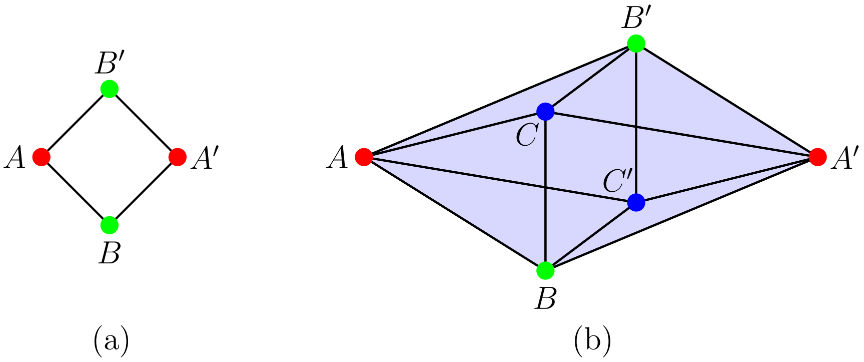
Definition 1.3.1. *We will refer to* $(X, O, \mathcal{M})$ *as a* measurement scenario.

We will mainly be concerned with finite measurement scenarios. Sets in the down-closure $\mathcal{U} := \downarrow \mathcal{M}$ will be referred to as *contexts* and will be denoted by the letters $U, V, \ldots$; elements of the cover $\mathcal{M}$ itself will be usually be referred to as *maximal contexts* and will be denoted by the letters $C, D, \ldots$.

A measurement scenario forms an abstract simplicial complex. For example, figure 1.2 (a) represents the measurement scenario for the Bell-CHSH model, and figure 1.2 (b) represents a similar tripartite measurement scenario in which the shaded faces represent the maximal contexts (this is the compatibility structure of the GHZ-Mermin model [58, 57, 83, 84]).

The *event sheaf* $\mathcal{E} : \mathcal{P}^{\mathrm{op}}(X) \to \mathbf{Set}$ is defined by $\mathcal{E}(U) := O^U$ for each $U \subseteq X$; i.e. $\mathcal{E}(U)$ contains all functional assignments of outcomes to the measurements in U.

Figure 1.2: (a) The compatibility structure of the Bell-CHSH model (b) A similar tripartite measurement scenario.



In order to describe an empirical model we must specify a probability distribution over the assignments $\mathcal{E}(C)$ for each maximal context $C \in \mathcal{M}$. This can be achieved by composing $\mathcal{E}$ with the distribution functor $\mathcal{D}_R : \mathbf{Set} \to \mathbf{Set}$ that takes a set to the set of R-distributions over it, where R is some semiring. Probability distributions are obtained when $R = \mathbb{R}^+$, the non-negative reals. More generally, it can be useful to consider other kinds of distributions: for example 'negative probability' ($R = \mathbb{R}$) or 'possibilistic' ($R = \mathbb{B}$, the Boolean semiring) distributions. The composition of the two functors, $\mathcal{D}_R\mathcal{E}$, is a presheaf in which restriction is given by marginalisation of distributions. Now, an empirical model can be specified by a family of distributions $\{e_C\}_{C\in\mathcal{M}}$, where each $e_C \in \mathcal{D}_R\mathcal{E}(C)$.

To avoid confusion between sections of the event sheaf $\mathcal{E}$ and the presheaf $\mathcal{D}_R\mathcal{E}$, we will refer to sections of the former as *assignments* throughout, since they are understood to assign outcomes to measurements.

We build the property of no signalling into our models by imposing the condition that the marginals of the distributions $\{e_C\}_{C\in\mathcal{M}}$ specifying an empirical model agree wherever contexts overlap; i.e.

$$\forall\, C, D \in \mathcal{M}.\; e_C|_{C\cap D} = e_D|_{C\cap D}.$$

This implies that there are well-defined distributions e_U for all $U \in\downarrow \mathcal{M}$, since we obtain the same distribution no matter which maximal context we marginalise from. This is *compatibility* in the sense of the sheaf condition.

Definition 1.3.2. *An* empirical model e *over a measurement scenario* $(X, O, \mathcal{M})$ *is a compatible family of R-distributions*

$$\{e_C\}_{C\in\mathcal{M}},$$

with $e_C \in \mathcal{D}_R\mathcal{E}(C)$ for each $C \in \mathcal{M}$.

We will use tables as a convenient way of representing empirical models throughout the dissertation. The following example illustrates how such a table is anatomised in the sheaf-theoretic language.

Example 1.3.3 (The Bell-CHSH Model)**.** *The empirical model is again represented in the following probability table.*

	00	01	10	11
$A\ B$	½	0	0	½
$A\ B'$	⅜	⅛	⅛	⅜
$A'\ B$	⅜	⅛	⅛	⅜
$A'\ B'$	⅛	⅜	⅜	⅛

The measurement scenario is described by $X = \{A, A', B, B'\}, O = \{0, 1\}$ and

$$\mathcal{M} = \{\{A, B\}, \quad \{A, B'\}, \quad \{A', B\}, \quad \{A', B'\}\}.$$

The labels for the rows correspond to the maximal contexts, and the cells of each row C (ignoring the entries for now) correspond to the assignments $\mathcal{E}(C)$: for example,

$$\begin{aligned}\mathcal{E}(\{A, B\}) = \{AB \mapsto 00, \quad & AB \mapsto 01,\\ AB \mapsto 10, \quad & AB \mapsto 11\}.\end{aligned}$$

The entries of each row specify the probability distrubution over these assignments (the joint outcomes): for example, the first row of the table

	00	01	10	11
$A\ B$	½	0	0	½

corresponds to the distribution

$$e_{\{A,B\}} \in \mathcal{D}_{\mathbb{R}^+}\mathcal{E}(\{A, B\}).$$

The sheaf-theoretic empirical model e obtained in this way is of course well-defined since it arises from quantum mechanics and is therefore necessarily compatible (no-signalling).

1.4 Locality & Non-contextuality

An important feature of the framework is that it is general enough to provide a unified approach to non-locality and contextuality. The main result of [4] is the following theorem.

Theorem 1.4.1 (Abramsky & Brandenburger)**.** *An empirical model can be realised by a factorisable hidden variable model if and only if the model is extendable to a global section.*

By *factorisability* it is meant that, when conditioned on any particular value of the hidden variable, the probability assigned to a joint outcome should factor as the product of the probabilities assigned to individual outcomes. For Bell scenarios this corresponds exactly to Bell locality [22]. On the other hand, a model is said to be *extendable to a global section* precisely when there exists a $d \in \mathcal{D}_R\mathcal{E}(X)$ such that $d|_C = e_C$ for all $C \in \mathcal{M}$. This corresponds to non-contextuality in the sense of the Kochen-Specker theorem [73].

In the sheaf-theoretic language, then, locality and non-contextuality are characterised in a unified manner by the existence of global sections. Contextuality will therefore sometimes be used as a general term which is assumed to include non-locality. This insight has already led to many interesting results [2, 4, 9, 11, 12, 81, 100]. Non-locality and contextuality are characterised by obstructions to the existence of global sections. In chapter 3 we take this idea further and explore the use of presheaf cohomology as a tool for identifying such obstructions. In chapter 4 we will introduce a refinement of the notion of extendability, which can capture the idea of partial approximations to locality or non-contextuality, and recovers the usual form of extendability in an appropriate limit. We mention also that the set of global assignments $\mathcal{E}(X)$ provides a canonical form of local hidden variable [30, 31]. A simplified proof is given in chapter 5. In this way, the sheaf-theoretic framework can be said to subsume the hidden variable approach.

We note that this greatly generalises earlier work by Fine [50], which showed in certain bipartite Bell-type measurement scenarios[1] that for any local hidden variable model there exists an equivalent deterministic local hidden variable model.

[1] (2,2,2) Bell scenarios, which will be presented in detail in chapter 2.

1.5 Possibilistic Models

Since the fundamental insight of Bell [22, 24], it is known that quantum mechanics gives rise to non-locality. Under some seemingly natural assumptions of locality and realism, it can be shown that any empirical model would have to satisfy certain Bell inequalities, which can be violated quantum-mechanically, from which Bell's conclusion follows.

A more intuitive, logical approach to non-locality proofs was pioneered by Heywood and Redhead [67], Greenberger, Horne, Shimony and Zeilinger [58, 57] (which was formulated in a simplified form by Mermin [83, 84]) and Hardy [59, 60] (also treated by Mermin [86]). This kind of non-locality proof disregards the exact values of the joint outcome probabilities and only records which of them are non-zero and which are zero. In other words, one distinguishes only between possible outcomes and impossible outcomes, and this turns out to be sufficient for demonstrating non-locality in quantum mechanics. Subsequently, several other non-locality proofs of this type have been found (e.g. [29, 36, 54]).

In order to present this kind of 'logical' argument, it suffices to consider what we call *possibilistic empirical models*. One kind of possibilistic empirical model that we have already encountered is the state-independent model, but in fact we can obtain a possibilistic model from any empirical model via the process of *possibilistic collapse*. In a possibilistic empirical model the distributions are Boolean; i.e. the semiring is $R = \mathbb{B} = (\{0,1\}, \vee, 0, \wedge, 1)$ where $\vee$ ('or') is addition modulo 2 and $\wedge$ ('and') is multiplication modulo 2. Boolean '1' is understood to denote 'possible' and '0' to denote 'impossible'.

Possibilistic collapse turns any empirical model into a possibilistic one by conflating all non-zero probabilities to the Boolean '1'. More carefully, its action is described by the natural transformation $\gamma : \mathcal{D}_{\mathbb{R}^+} \to \mathcal{D}_{\mathbb{B}}$ induced by the function

$$h : \mathbb{R}^+ \to \mathbb{B}, \quad p \mapsto \begin{cases} 0 & \text{if } p = 0 \\ 1 & \text{if } p > 0 \end{cases}. \tag{1.1}$$

Example 1.5.1. *The now familiar Bell-CHSH model collapses to the following possibilistic model.*

	00	01	10	11
$A'\ B$	1	0	0	1
$A\ \ B'$	1	1	1	1
$A\ \ B'$	1	1	1	1
$A'\ B'$	1	1	1	1

We introduce a notation that will be extremely useful in dealing with possibilistic models. The *support* of a distribution d over Y is the set

$$\mathsf{supp}(d) := \{y \in Y \mid d(y) \neq 0\} \, .$$

For any $U \subseteq X$ we define

$$S_e(U) := \{s \in \mathcal{E}(U) \mid \forall\, C \in \mathcal{M}.\; s|_{C\cap U} \in \mathsf{supp}(e_C)|_{C\cap U}\} \, .$$

That is to say, the set $S_e(U)$ contains all functional assignments of outcomes to the measurements U that are consistent with the model e. In particular, the set $S_e(X)$ contains all the global assignments that are consistent with the model e, and for each maximal context $C \in \mathcal{M}$ we have $S_e(C) = \mathsf{supp}(e_C)$. It can be shown that $S_e : \mathcal{P}(X)^{\mathrm{op}} \to \mathbf{Set}$ defines a sub-presheaf of the sheaf of events.

The possibilistic content of an empirical model is that which is available at the level of the support of the distributions of which it is made up. That is because a Boolean distribution can be equivalently represented by its support: i.e. there is a bijection

$$\mathsf{supp}(d) \cong \{y \in Y \mid \gamma d(y) = 1\}$$

between the distributions $\mathcal{D}_{\mathbb{B}}(Y)$ and the non-empty subsets of Y, and therefore

$$\{S_e(C)\}_{C\in\mathcal{M}} \cong \{\gamma e_C\}_{C\in\mathcal{M}}.$$

1.6 A Hierarchy of Contextuality

Logical Contextuality

At the possibilistic level, for any empirical model e, we can pose the problem of whether γe is extendable to a global section. As we have seen, a global section $d \in D_{\mathbb{B}}\mathcal{E}(X)$ can be equivalently represented by the set $\mathsf{supp}(d) \subseteq \mathcal{E}(X)$. Then the problem is to find a Boolean distribution over the global assignments $\mathcal{E}(X)$ which restricts to γe_C for each maximal context C. If such a distribution exists we will say that e is *possibilistically extendable* (to a global section).

Using the notation introduced in section 1.5, we are interested in the existence of a Boolean distribution $d \in D_{\mathbb{B}}\mathcal{E}(X)$ for which the following conditions hold.

1. $\mathsf{supp}(d) \subseteq S_e(X)$; i.e. all global assignments in $\mathsf{supp}(d)$ are consistent with the empirical model.

2. $\forall\, C \in \mathcal{M}.\ \forall\, t \in S_e(C).\ \exists\, s \in \mathsf{supp}(d).\ t = s|_C$; i.e. any possible local assignment can be obtained as the restriction of some global assignment in $\mathsf{supp}(d)$.

In short,

$$S_e(C) = \{s|_C \mid s \in \mathsf{supp}(d)\} \tag{1.2}$$

for each $C \in \mathcal{M}$.

There is also an equivalent way to consider possibilistic extendability, which will be especially relevant in chapters 2 and 3.

Proposition 1.6.1. *An empirical model e is possibilistically extendable to a global section if and only if, for all $C' \in \mathcal{M}$, each assignment $s' \in S_e(C')$ belongs to a compatible family of assignments $\{s_C\}_{C \in \mathcal{M}}$ such that $s_{C'} = s'$.*

Proof. If e is possibilistically extendable to a global section d then by (1.2) there exists some global assignment $s \in \mathsf{supp}(d)$ such that $s|_{C'} = s'$. We define the family $\{s_C\}_{C \in \mathcal{M}}$ by $s_C := s|_C$. Then $s_{C'} = s'$ and the family is compatible since it's defined by restriction from a global assignment.

For the converse, suppose that $s' \in S_e(C')$ belongs to a compatible family $\{s_C\}_{C \in \mathcal{M}}$ such that $s_{C'} = s'$. Then we can glue these assignments together to form a global assignment $s : X \to O$ defined by $s(m) := s_C(m)$ for any $C \ni m$. This is well-defined by the compatibility of $\{s_C\}$. Now we can define the Boolean distribution d with support $\mathsf{supp}(d) := \{s \in \mathcal{E}(X) \mid s' \in S_e(C') \text{ for some } C' \in \mathcal{M}\}$. This is a possibilistic global section since conditions 1 and 2 are trivially satisfied. $\square$

Definition 1.6.2. *If an empirical model is not possibilistically extendable to a global section we say that the model is* logically contextual *(or* logically non-local *when appropriate).*

These are the empirical models that admit 'logical' proofs of non-locality.

Some models can be non-local or contextual without exhibiting the properties at the possibilistic level: an example is the Bell-CHSH model. However, it can be shown that a probabilistic model that exhibits logical contextuality at the possibilistic level is necessarily contextual at the probabilistic level, too [4]. Logical contextuality is therefore a strictly stronger form of contextuality. Many familiar empirical models exhibit logical non-locality or contextuality, including the Hardy model [59, 60], which will be considered in detail in chapter 2. A recent result [13] even indicates that for any multipartite qubit state there exists some choice of measurements that will give rise to logical non-locality.

Example 1.6.3 (The Hardy Model)**.** *The support of the Hardy model is represented in the following table.*

	00	01	10	11
$A\ B$	1	1	1	1
$A\ B'$	0	1	1	1
$A'\ B$	0	1	1	1
$A'\ B'$	1	1	1	0

The local assignment $t : AB \mapsto 00$ cannot be obtained as the restriction of any global assignment $s : ABA'B' \mapsto 00o_{A'}o_{B'}$, and therefore condition 2 for possibilistic extendability does not hold.

Strong Contextuality

Recall that $S_e(X)$ consists of those global assignments that are *consistent* with the support of e; i.e. whose restrictions to every context of compatible observables are possible according to e. These are the only global assignments that could be taken to be possible. It has already been observed that if a possibilistic extension $d \in D_{\mathbb{B}}\mathcal{E}(X)$ exists then $\mathsf{supp}(d) \subseteq S_e(X)$, and it is clear that in this case $S_e(X)$ is also a possibilistic extension of e. This follows from condition 2: if any possible local assignment arises as the restriction of an assignment in $\mathsf{supp}(d)$ then, since $\mathsf{supp}(d) \subseteq S_e(X)$, it arises as a restriction of an assignment in $S_e(X)$. For this reason, $S_e(X)$ can be regarded as providing a canonical candidate for a possibilistic extension of the empirical model e.

In general, the set $S_e(X)$ of consistent global assignments can fail to determine an extension of the empirical model e if it isn't large enough to account for all 'local' assignments that are possible in e; that is, if there exists some assignment $s \in S_e(C)$ on some maximal context $C \in \mathcal{M}$ which does not arise as a restriction of a global assignment in $S_e(X)$, as in example 1.6.3. The extreme case happens when $S_e(X)$ is empty (then, $S_e(X)$ does not even determine a distribution over $\mathcal{E}(X)$). This means that there is no global assignment that is consistent with the support of e. In this case, we say that the model e is *strongly contextual* (or strongly non-local, when appropriate).

Note that to have non-empty $S_e(X)$ is a weaker property than possibilistic extendability: it is simply asking for the existence of some global assignment consistent with the support of e. Correspondingly, the negative property is stronger than possibilistic non-extendability (possibilistic contextuality/non-locality). Some of these ideas will be generalised in chapter 4.

The Hardy model of example 1.6.3 is logically non-local but not strongly non-local. Strong contextuality is displayed by many models, however, including the GHZ-Mermin model [83, 84], the 18-vector Kochen-Specker model, the Peres-Mermin 'magic square' [85, 91] and the Popescu-Rohrlich correlations [93].

Example 1.6.4 (The GHZ-Mermin Model)**.** *This model will also be considered in more detail in chapter 2. Its support is represented in the following table.*

	000	001	010	011	100	101	110	111
$A\ B\ C$	1	0	0	1	0	1	1	0
$A\ B\ C'$	1	1	1	1	1	1	1	1
$A\ B'\ C$	1	1	1	1	1	1	1	1
$A\ B'\ C'$	0	1	1	0	1	0	0	1
$A'\ B\ C$	1	1	1	1	1	1	1	1
$A'\ B\ C'$	0	1	1	0	1	0	0	1
$A'\ B'\ C$	0	1	1	0	1	0	0	1
$A'\ B'\ C'$	1	1	1	1	1	1	1	1

Here, no local assignment can be completed to a consistent global assignment.

We thus arrive at a strict hierarchy of contextuality:

$$\text{Strong Contextuality} > \text{Logical Contextuality} > \text{Contextuality}.$$

In terms of familiar representative non-local models of these classes,

$$\text{GHZ} > \text{Hardy} > \text{Bell}.$$

1.7 Towards an Ontological Theory

Many current research programmes are concerned with the problem of reformulating or axiomatising quantum mechanics (e.g. [38, 61, 69]). At the foundational level, a goal of such programs is often to provide a framework for possible theories that might allow one to identify special or defining features of quantum mechanics. Another eventual goal might be to provide a framework that is compatible with quantum mechanics while at the same time being general enough to allow for a possible theory of quantum gravity.

The sheaf-theoretic framework provides and elegant and powerful unified approach to the non-locality and contextuality of correlations in empirical models in a way that is neutral with respect to whatever theory might give rise to the correlations. In this section we outline some steps towards a sheaf-theoretic framework for axiomatising ontological theories, in which this neutrality can be a useful feature, drawing on ideas from [5].

The following is a consequence of Gleason's theorem [56], which provides a motivation to think of empirical models as states (for a more detailed discussion see [45] and [46]).

Proposition 1.7.1. *Let $\mathcal{H}$ be the Hilbert space for a quantum system with observables $\mathcal{A} \subseteq \mathbf{B}(\mathcal{H})$, a von Neumann (sub)algebra of the set of bounded linear operators on $\mathcal{H}$. Let $\mathbf{C}(\mathcal{A})$ be the set of commutative subalgebras of $\mathcal{A}$. There is a one-to-one correspondence between the no-signalling empirical models on the measurement scenario $(\mathcal{A}, \mathbb{R}, \mathbf{C}(\mathcal{A}))$ derived from the compatibility structure of the observables and the set of positive linear functionals on $\mathcal{A}$ (the Gleason states).*

This tells us that if we consider the measurement scenario of all the possible observables on a quantum mechanical system, no-signalling models correspond in a precise way to the quantum states. We will use this as the motivating example for setting up a sheaf-theoretic framework for ontological theories. It should also be noted that if one were to restrict attention to a smaller algebra of observables, this could allow for a larger space of Gleason states, which would no longer coincide with the quantum states [14].

As an aside, the proposition justifies the use of the terms non-local, logically non-local, etc. to describe quantum states for which there exist some sets of compatible observables such that the resulting empirical model has the particular property. This terminology was introduced in [9].

Definition 1.7.2. *A state is said to be* (logically/strongly) contextual *(or* non-local*) if there exist some observables such that the resulting empirical model has that property.*

Since a quantum state can be considered as an empirical model in its own right, the existence of some subset of the observables $\mathcal{A}$ giving rise to a non-local model implies non-locality of the state, because the existence of a global section for the state would imply, by restriction, the existence of a global section for any such sub-model. We note also that if $\mathbb{R}^4$ can be embedded into the Hilbert space of a quantum system, then by the 18-vector Kochen-Specker theorem (which is state-independent) all states of that system are (strongly) contextual in this sense.

For convenience we fix a single outcome set $O = \mathbb{R}$.

- To each system A we associate a system type $(X_A, \mathcal{M}_A)$, and a set of states S_A which are (no-signalling) empirical models over the measurement scenario $(X_A, O, \mathcal{M}_A)$. The *system* A is completely defined by the tuple $(X_A, \mathcal{M}_A, S_A)$.

A morphism of system types is a simplicial map $f : (X_A, \mathcal{M}_A) \to (X_B, \mathcal{M}_B)$; i.e. a map $f : X_A \to X_B$ such that $f(C) \in \downarrow \mathcal{M}_B$ for all $C \in \mathcal{M}_A$. Recall that $\downarrow \mathcal{M}_B$, the down-closure of $\mathcal{M}_B$, which contains all (not necessarily maximal) contexts for the system B, is defined by

$$\downarrow \mathcal{M}_B := \{U \in \mathcal{P}(X_B) \mid \exists\, C' \in \mathcal{M}_B.\, U \subseteq C'\}\,.$$

So every maximal context in the system A maps to a valid context in the system B. This induces a map $f^* : S_B \to S_A$ (note the reversal) on states defined by

$$f^*(e)_C(s) := \sum_{\substack{s' \in \mathcal{E}(f(C)) \\ s' \circ f = s}} e_{f(C)}(s'),$$

for any $e \in S_B$ and $C \in \mathcal{M}_A$. That $f^*(e)$ is a well-defined model follows from the compatibility of e.

- A *morphism of systems* $f : (X_A, \mathcal{M}_A, S_A) \to (X_B, \mathcal{M}_B, S_B)$ is a morphism of system types with the additional property that $f^*(S_B) \subseteq S_A$. This can be interpreted as saying that each state in S_B must be reachable from some state in S_A.

It is clear that identities and compositions are well-defined, so systems and morphisms of systems form a category $\mathcal{C}$, which we call the *category of systems.*

Furthermore, we would like to be able to treat compound systems. For a system of type A and a system of type B there should be a means of composing these to obtain a compound system of type $A \otimes B$ in a coherent way. The appropriate structure is a symmetric monoidal product structure on the category of systems. The idea of using this kind of structure to treat compound systems has been developed extensively in the categorical quantum mechanics programme [7, 8], and we will not labour the point here.

- For systems A given by $(X_A, \mathcal{M}_A, S_A)$ and B given by $(X_B, \mathcal{M}_B, S_B)$ we define the *compound system* $A \otimes B$ by the tuple

  $$(X_{A\otimes B}, \mathcal{M}_{A\otimes B}, S_{A\otimes B})$$

  where $X_{A\otimes B} := X_A \sqcup X_B$, the disjoint union of the measurement sets,

  $$\mathcal{M}_{A\otimes B} := \{C_A \sqcup C_B \mid C_A \in \mathcal{M}_A, C_B \in \mathcal{M}_B\},$$

  and

  $$S_{A\otimes B} := \{e \text{ a state on } (X_{A\otimes B}, O, \mathcal{M}_{A\otimes B}) \mid e|_A \in S_A, e|_B \in S_B\}.$$

  The action on morphisms is the obvious one which lifts from the coproduct (disjoint union) of measurement sets.

$(\mathcal{C}, \otimes)$ forms a symmetric monoidal category.

These three axioms can provide a basic setting in which to consider ontological theories in the sheaf-theoretic language. Of course there may be other restrictions or axioms that we would wish to impose; for example, we might wish to restrict attention to certain types of systems, or certain states on systems, or to impose axioms such as local tomography or the Hardy composition principle [62], etc. A sheaf-theoretic ontological theory, then, would be some symmetric monodical subcategory of $(\mathcal{C}, \otimes)$. In chapter 5 we will suggest some other ways in which this approach might be developed.

1.8 Discussion

The sheaf-theoretic framework provides a precise mathematical language for analysing empirical data or predictions, and can be a powerful, unifying approach to the foundations of quantum mechanics. We have seen that a very natural, unified characterisation of non-locality and contextuality, the key features of quantum correlations, emerges in the general setting. This has already led to a string of interesting results, such as the classification of contextuality of section 1.6. The deeper, more structural approach can raise surprising and interesting connections with other fields. On the one hand, it raises possibilities for the use of new methods and results in the study of non-locality and contextuality: the mathematics of cohomology, which will be considered in chapter 3, or game theory [100], for example. On the other hand, it can also lead to the wider application of foundational research: to relational database theory [2] or linguistics [96], for example. These possibilities have only begun to be investigated.

Chapter 2

Hardy's Paradox as a Logical Condition for Non-locality

In this chapter, which builds on work published in [79], we consider a general framework for logical non-locality proofs, which takes some inspiration from the relational hidden variable framework of Abramsky [3]. Though not as general, it can be considered as a precursor to sheaf-theoretic framework [4], which it predates. More specifically, we study logical Bell inequalities in (n, k, l) Bell scenarios, where n is the number of sites, k is the number of allowed measurements at each site, and l is the number of possible outcomes for each measurement. This is a purely possibilistic version of the sheaf-theoretic framework for such scenarios, which comes with a particular representation for $n = 2$ and $n = 3$ scenarios that can provide a powerful means of reasoning about empirical models.

Hardy's non-locality 'paradox' is a proof without inequalities showing that certain non-local correlations violate local realism [59, 60]. It is considered to be the simplest non-locality proof for quantum mechanics. What we find appears to be a remarkable universality of Hardy's paradox. We prove a number of completeness theorems showing that it is a necessary and sufficient condition for logical non-locality in all $(2, k, 2)$ and $(2, 2, l)$ scenarios, subsuming, for example, ladder paradoxes [29]. We show that we can even interpret the logical versions of the no-signalling condition and the normalisation of probabilities as degenerate cases of the non-occurrence of coarse-grained Hardy paradoxes. However, for the $(2, 3, 3)$ and $(3, 2, 2)$ scenarios we find new logical locality conditions which can be violated without the occurrence of a Hardy paradox.

These completeness results have many interesting consequences. They lead to a constructive argument that the Popescu-Rohrlich box is the only strongly non-local $(2, 2, 2)$ model, and to a proof that Bell states are not logically non-local. Since all other entangled two-qubit states can be shown to witness a Hardy paradox [60] this proves the surprising result that the Bell states are the only such states that are not logically non-local. Together with recent results indicating that all n-partite entangled qubit states for $n > 3$ are logically non-local [13], this shows that the Bell states are anomalous in this respect, in spite of the fact that they are perhaps the most studied and utilised entangled states. It also leads to the discovery of a family of no-signalling empirical models which lie within the Tsirelson bound and can have an arbitrarily small violation of the CHSH inequality though they are not quantum realisable.

Much of the literature on Hardy's paradox is concerned with the probability of witnessing a paradox, which is often considered to be a measure of the quality of Hardy non-locality. This has experimental motivations. The original Hardy paradox can be witnessed with maximum probability $(5\sqrt{5} - 11)/2 \approx 0.09$. It has been shown, however, that it is possible to witness a generalisation of Hardy's paradox with probability 0.125 for a tripartite quantum system [54], and more recently that another generalisation of Hardy's paradox can be witnessed with probability ≈ 0.4 for a high-dimensional bipartite quantum system [37].

Using the present framework, we will achieve a striking improvement on these results, and demonstrate by a much simpler argument that it is possible to witness Hardy non-locality with certainty for a tripartite quantum system. Interestingly, the argument relies on the same state and measurements as the GHZ experiment [57]. We also show that Hardy non-locality can be achieved with certainty for a particular non-quantum, no-signalling $(2, 2, 2)$ empirical model, which turns out to be the Popescu-Rohrlich no-signalling box [93].

2.1 Hardy's Non-locality Paradox

The original Hardy argument concerns the $(2, 2, 2)$ scenario. To give a concrete account of the argument we consider an idealised experiment in which measurements are carried out in Alice's lab and Bob's lab, which share a (possibly entangled) quantum state. Each experimenter can choose to make one of two measurements on their subsystem, which we call polarisation and colour. Each measurement has two possible outcomes: $\{\uparrow, \downarrow\}$ for polarisation, and $\{G, W\}$ for colour. We assume that Alice and Bob perform very many runs of the experiment (each time starting with the same shared state) and then tabulate their results as in table 2.1. A '1' in the table signifies that it was possible to obtain those two outcomes in the same run, and a '0' signifies that this never happened. Such a specification of possibilities is of course an

Table 2.1: An empirical model containing a Hardy paradox. This is a possibilistic table with '1' standing for 'possible' and '0' standing for 'impossible'. The blank entries are not relevant to the argument.

		Bob			
		$\uparrow$	$\downarrow$	G	W
Alice	$\uparrow$	1			0
	$\downarrow$				
	G			0	
	W	0			

empirical model. Recall from chapter 1 that any probabilistic empirical model can be transformed into a possibilistic one in a canonical way via *possibilistic collapse*: the process by which all non-zero probabilities are conflated to '1'.

The partially completed table 2.1 is Hardy's paradox. Notice that the table is of a different form to those of chapter 1. Empirical models will be represented in this way throughout the chapter. We have deliberately chosen this particular representation because, as we will see shortly, it allows one to more easily recognise various features of empirical models and to reason about them. However, it is not used elsewhere in the dissertation since it cannot be straightforwardly generalised beyond $n = 3$. For the present tabular representation of empirical models we will use the terminology that measurements label *rows/columns*, joint measurements label *boxes*, outcomes label *sub-rows/columns*, and joint outcomes label *entries*.

The apparent paradox arises because the table tells us that, when both experimenters measured polarisation, it was possible for them to both get the outcome $\uparrow$; but, when one measured polarisation and the other measured colour, it never happened that they could obtain $\uparrow$ and W together. From these statements it seems that whenever $\uparrow$ was measured in one lab, the colour in the other lab must have had the value G; and since it was possible for both to get the outcome $\uparrow$, then it should have been possible for both to get the outcome G if the experimenters had instead decided to measure colour on those runs. However, the remaining specified entry in the table tells us that it was not possible for both experimenters to measure G. Despite this apparent paradox, such behaviour is actually predicted by quantum mechanics.

Definition 2.1.1. *We say that the joint outcome* $(\uparrow, \uparrow)$ witnesses a Hardy paradox.

Of course, in stating this argument, we have made some tacit assumptions. In particular, we have assumed some form of locality (or, to be more precise, no-signalling) by supposing that, for each run, Bob's choice of measurement did not affect Alice's outcome and vice versa. As discussed in section 1.1, such behaviour could give rise to faster-than-light communication between far distant labs, which is prohibited by special relativity. We have also implicitly assumed some form of realism: that colour and polarisation had definite values even when they were not being measured. Without such an assumption, of course, it would be difficult to give sense to a notion of locality. A further assumption, which concerns the free-choice of experimenters, is that every combined measurement choice has some outcome. This is related to the property of λ-independence, which will be discussed in chapter 5.

Throughout the dissertation we refer to this as Hardy's paradox, though we draw attention to the fact that it is only an apparent paradox. Really, this is a non-locality theorem, which states that models of a certain form cannot satisfy the properties of locality and realism.

We can write the condition for non-occurrence of the Hardy paradox in table 2.1 as a formula in Boolean logic:

$$p(\uparrow,\uparrow) \quad \rightarrow \quad p(\uparrow,W) \vee p(W,\uparrow) \vee p(G,G) \, ,$$

where the $p(i,j) \in \{0,1\}$ are the entries of the table, or the possibility values for Alice to obtain outcome i and Bob to obtain outcome j. This can be thought of as a logical Bell inequality. For the $(2,2,2)$ scenario, there are 64 versions of the Hardy paradox which one obtains from table 2.1 by permuting the order or labelling of measurements and outcomes.

2.2 Properties of Empirical Models

In any discussion of locality, realism, etc. it is important to be careful about which properties are being assumed or inferred. In chapter 1 we mentioned some properties that empirical models might have. We will now present various properties in the context of our tabular representation of $n = 2$ possibilistic empirical models.

Definition 2.2.1. Measurement locality *is the property that at each site the allowed measurements are independent of which measurements are made at the other sites.*

We assume from the outset that all the models we deal with satisfy measurement locality. For $n = 2$, this is equivalent to the property that if the table of a model

Table 2.2: Examples of possibilistic empirical models: (a) a deterministic empirical model; (b) a local model; (c) a signalling model.

	$\uparrow$	$\downarrow$	G	W
$\uparrow$	1	0	1	0
$\downarrow$	0	0	0	0
G	1	0	1	0
W	0	0	0	0

(a)

	$\uparrow$	$\downarrow$	G	W
$\uparrow$	1	0	1	0
$\downarrow$	1	0	0	1
G	1	0	1	0
W	1	0	0	1

(b)

	$\uparrow$	$\downarrow$	G	W
$\uparrow$	1	0	1	0
$\downarrow$	0	0	0	0
G	0	1	1	0
W	0	0	0	0

(c)

has any zero box then that box must belong to a row (or column) of zero boxes, for otherwise the choice of measurement at one site would affect the available measurements at the other. This allows us to omit such rows/columns of zero boxes in the tabular representation and to assume that all tables are totally defined on the domain of measurement choices.

Definition 2.2.2. (Possibilistic) no-signalling (*NS*) *is the property that the choice of measurement at one site does not affect the possible outcomes at another site.*

In terms of the tabular representation, this means that if a sub-row has any '1' then that sub-row must have a '1' in each box, and similarly for sub-columns. For example, table 2.2 (a) and (b) are both no-signalling, while (c) is signalling. In (c), if Alice measures polarisation then the outcome of a polarisation measurement by Bob has to be $\uparrow$, but if Alice measures colour then Bob always gets $\downarrow$. It can be shown that if an empirical model violates possibilistic no-signalling then it also violates probabilistic no-signalling. The converse does not hold in general [3].

Definition 2.2.3. (Strong) determinism *is the property that the outcome at each site is uniquely determined by the measurement at that site.*

In the tabular form, this property says that each box should contain at most one '1', and that the '1's are consistent with no-signalling in that they line up in the same sub-rows/columns where possible. We call such an arrangement of '1's a *deterministic grid* (in the sheaf-theoretic language, these correspond to global assignments). Table 2.2 (a) is an example of a deterministic model. By this definition, determinism implies no-signalling.

In order to define local models, we need a notion of stochastic mixtures in our possibilistic setting. We define the mixture of a model A with entries p_{ij}^A and a model B with entries p_{ij}^B to be the model with entries

$$p_{ij} \equiv p_{ij}^A \vee p_{ij}^B \,.$$

This means that an outcome is possible in the mixture if and only if it is possible in at least one component of the mixture. Note that there is no mixing parameter of the kind that arises when considering stochastic mixtures of probabilistic models.

Definition 2.2.4. *The* local *models are those that can be obtained by taking mixtures of arbitrary sets of deterministic models.*

This corresponds to the existence of a possibilistic global section in the sheaf-theoretic approach. In the tabular representation, a model is local if and only if every '1' in its table belongs to some deterministic grid. An example of a local model is table 2.2 (b). The model in table 2.1 used to explain Hardy's paradox is not local, since the '1' in that table cannot be completed to a deterministic grid. In other words, the assignment does not belong to a compatible family of assignments, one for each context, c.f. proposition 1.6.1.

By theorem 1.4.1, the local models are precisely the models that can be described by (factorisable) local hidden variable models. The decomposition of local models into deterministic models described here can be seen as a canonical form of hidden variable model in which each value of the hidden variable corresponds to a deterministic model. The fact that all local hidden variable models for the $(2, 2, 2)$ scenario can be captured in this way follows from the work of Fine [50], but theorem 1.4.1 holds for all measurement scenarios, even those which are not of the Bell form.

We then obtain the following proposition, which facilitates the application of our results to the usual probabilistic setting:

Proposition 2.2.5. *With these definitions, possibilistic collapse takes probabilistic local models to possibilistic local models. Conversely, every possibilistic local model can be written as the possibilistic collapse of a probabilistic one.*

Proof. The first statement is clear from the fact that a non-trivial convex combination of two probabilities $p^A, p^B \in [0, 1]$ is non-zero precisely when at least one of p^A or p^B is non-zero. For the second statement, we simply write a given possibilistic local model as a mixture of deterministic models and assign an arbitrary non-zero probability to each of these models such that the probabilities sum to 1. This defines a probabilistic local model with the required property. $\square$

Table 2.3: A $(2, 2, l)$ scenario with a $H_{(m_1,m_2)}$ coarse-grained Hardy paradox.

	o'_1	$\cdots$	o'_l	o_1	$\cdots$	o_{m_2}	o_{m_2+1}	$\cdots$	o_l
o'_1	1						0	$\cdots$	0
$\vdots$									
o'_l									
o_1				0	$\cdots$	0			
$\vdots$				$\vdots$	$\ddots$	$\vdots$			
o_{m_1}				0	$\cdots$	0			
o_{m_1+1}	0								
$\vdots$	$\vdots$								
o_l	0								

We interpret this as saying that a non-locality proof without inequalities (or a logical non-locality proof) exists for a given empirical model if and only if it is non-local in the sense of definition 2.2.4.

2.3 Coarse-grained Versions of Hardy's Paradox

For $(2, 2, l)$ scenarios, we consider coarse-grainings of the Hardy paradox. The basic form is the same as in the $(2, 2, 2)$ case (table 2.1), but in the general case (table 2.3) we have $m_1 \times m_2$, $(l-m_1) \times 1$ and $1 \times (l-m_2)$ subtables of '0's, where $0 < m_1, m_2 < l$. Any empirical model whose table is isomorphic (up to permutations of measurements and outcomes) to table 2.3 for some values of m_1 and m_2 is said to have a coarse-grained Hardy paradox. We use the notation $H_{(m_1,m_2)}$ for this property.

Conditions for the non-occurrence of a paradox can still be written as a logical formula. For table 2.3 the corresponding formula is

$$p(o'_1, o'_1) \to \bigvee_{r=m_1+1}^{l} p(o_r, o'_1) \vee \bigvee_{s=m_2+1}^{l} p(o'_1, o_s) \vee \bigvee_{\substack{r\in[1,m_1]\\ s\in[1,m_2]}} p(o_r, o_s)\,.$$

We use the notation $NH_{(m_1,m_2)}$ for the property that all such formulas are satisfied for a particular model.

The coarse-graining includes the degenerate values 0 and l for m_1 and m_2. The cases $m_1 = 0$, $m_2 = l$ and $m_1 = l$, $m_2 = 0$ are especially interesting.

Proposition 2.3.1. *The no-signalling condition can be stated as the logical predicate*

$$NS = NH_{(0,l)} \wedge NH_{(l,0)} \ .$$

Proof. For table 2.3, $NH_{(0,l)}$ and $NH_{(l,0)}$ state that the first sub-column in the lower left box needs to contain some '1', and, respectively, that the first sub-row in the upper right box needs to contain some '1'. These are the possibilistic no-signalling relations. By permutations of measurements and outcomes, these apply to any '1' in the table; so for the no-signalling predicate we get $NS = NH_{(0,l)} \wedge NH_{(l,0)}$. □

The case that $m_1 = m_2 = l$ is also interesting.

Proposition 2.3.2. *The condition that there exists a well-defined Boolean distribution at each context (or the* 'normalisation of possibility'*) can be expressed as* $NH_{(l,l)}$.

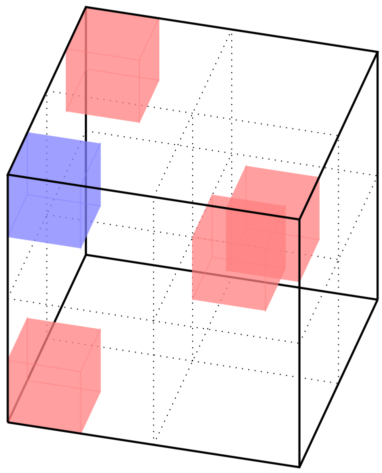
Proof. For table 2.3 $NH_{(l,l)}$ simply expresses that the lower right box in table 2.3 should contain at least some '1'. This is the normalisation of possibility: in order to form a well-defined Boolean distribution at each context, at least one outcome has to be possible for each choice of measurements. So given that at least some '1' occurs somewhere in the table of a no-signalling model, the normalisation of possibility is equivalent to $NH_{(l,l)}$. □

These properties and observations extend to all $(2, k, l)$ Bell scenarios by considering 2×2 subtables. Moreover, when we consider coarse-grainings of the generalised version of Hardy's paradox for $(n, 2, 2)$ Bell scenarios in the next section, it will be clear that these observations extend in an obvious way to all (n, k, l) Bell scenarios.

2.4 An n-partite Hardy Paradox

Wang and Markham have described a generalisation of the Hardy paradox to $(n, 2, 2)$ scenarios which can be used to demonstrate that all symmetric n-partite qubit states for $n > 2$ are logically non-local [104]. This kind of generalisation has been described elsewhere by Ghosh, Kar and Sarkar [54], and is also considered in [36] and [39]. If measurements and outcomes are both labelled by $\{0, 1\}$ at each site, then a generalised Hardy paradox occurs if (up to re-labelling of measurements and outcomes) the following possibilistic conditions are satisfied.

Figure 2.1: The $n = 3$ Hardy paradox. The blue entry corresponds to Boolean '1' or 'possible', and the red entries to '0' or 'impossible'. The blank entries are unspecified.



- $p(\,0,\ldots,0 \mid 0,\ldots,0\,) = 1$
- $p\,(\,\pi(1,0,\ldots,0) \mid \pi(1,0,\ldots,0)\,) = 0$ for all permutations π
- $p(\,0,\ldots,0 \mid 1,\ldots,1\,) = 0$

Then, since all possibilities $p(o_1 \ldots o_n \mid m_1 \ldots m_n)$ are Boolean valued, we can consider these as logical propositions and write the following formula in Boolean logic for the non-occurrence of a generalised Hardy paradox:

$$p(\,0,\ldots,0 \mid 0,\ldots,0\,) \quad \to \bigvee_{\pi \in \text{permutations}} p\,(\,\pi(1,0,\ldots,0) \mid \pi(1,0,\ldots,0)\,) \vee p(\,0,\ldots,0 \mid 1,\ldots,1\,)\,.$$

For the purposes of this chapter it is not necessary to go beyond the $n = 3$ paradox, which can be represented in a three dimensional version of the tabular representation described in section 2.2; see figure 2.1. The advantage of the representation is that it provides a powerful visual means of analysing models.

The axes correspond to different sites, the cubes to joint measurement choices, and individual entries to outcomes, similarly to the $n = 2$ case. The properties of the tabular representation generalise in the obvious way to the third dimension. For example, the blue entry in figure 2.1 cannot be completed to a deterministic grid, and so any $(3, 2, 2)$ model containing this paradox is logically non-local.

2.5 Universality of Hardy's Paradox

In this section we will prove a number of completeness theorems, which show that the occurrence of a (coarse-grained) Hardy paradox is a necessary and sufficient condition for logical non-locality in certain Bell scenarios. In other words, for the scenarios we will describe, logical non-locality is always due to the occurrence of a Hardy paradox.

We write *NH* for the property that no coarse-grained Hardy paradox occurs in a given model.

Theorem 2.5.1. *For the* $(2, 2, 2)$ *scenario, the property of non-occurrence of any coarse-grained paradox is equivalent to possibilistic locality:*

$$NH \quad \leftrightarrow \quad \textit{(Locality)}\,. \tag{2.1}$$

Proof. We have already demonstrated in section 2.1 that an occurrence of the Hardy paradox implies a violation of locality. It only remains to prove that *NH* implies locality. By the observations at the end of the last section, we know in particular that *NH* implies *NS*, so that we can freely use the latter.

From the earlier definition, a model is local if and only if every '1' in its tabular representation belongs to some deterministic grid. We begin by choosing an arbitrary '1' in the table. Without loss of generality (w.l.o.g.) let this be the '1' in table 2.4 (a). Then, by *NS*, the first sub-row must have a '1' in each box, and similarly for the first sub-column. Again w.l.o.g. we let these be the entries in table 2.4 (b). If the starred entry here is a '1', this completes the first entry to a deterministic grid and we're done. Assume that the starred entry is a '0'. Then, by no-signalling, we can fill in the '1's in the lower right box of table 2.4 (c). Now, if either of the starred entries in this table is a '1', this completes the first entry to a deterministic grid. This must be the case, for if it were not then the '0's in these places would form a Hardy paradox together with the first entry and the '0' in the lower right box; but we have assumed the property *NH*. $\square$

This theorem generalises easily to $(2, 2, l)$ scenarios.

Theorem 2.5.2. *For* $(2, 2, l)$ *scenarios, the property of non-occurrence of any coarse-grained Hardy paradox is equivalent to locality; i.e. (2.1) holds for* $(2, 2, l)$ *scenarios.*

Proof. Again, it is enough to show that the left-hand side implies the right-hand side while assuming *NS*. If we take an arbitrary '1' in the table, we can re-label measurements and outcomes such that this '1' appears in the upper-left corner of the

Table 2.4: Stages in the proof of theorem 2.5.1.

	1	

(a)

	1	1
	1	*

(b)

	1	1 *
	1 *	0 1 1

(c)

table, and such that for the upper-right and lower-left boxes, the first sub-row and sub-column, respectively, non-zero entries (of which, by no-signalling, there must be at least one) appear before zero entries, as in table 2.5. Assuming that there is no coarse-grained paradox, at least one of the starred entries must be a '1', and this completes the arbitrarily chosen '1' to a deterministic grid. □

We can also generalise theorem 2.5.1 to $(2, k, 2)$ scenarios.

Theorem 2.5.3. *For $(2, k, 2)$ scenarios, the property of non-occurrence of any Hardy paradox is equivalent to locality; i.e. (2.1) holds for $(2, k, 2)$ scenarios.*

Proof. By theorem 2.5.1, we know that this holds for $k = 2$, and will show by induction that it holds for all k. It is useful to use the tabular representation of models in what follows. In this setting, it must be shown that every '1' in a given table can be completed to a deterministic grid of '1's, assuming that no Hardy paradox occurs. We will show that this property holds for all $k_1 \times k_2$ tables, i.e. for all scenarios with k_1 two-outcome measurements for Alice and k_2 two-outcome measurements for Bob, given that it holds for all $k_1 \times (k_2 - 1)$ tables and all $(k_1 - 1) \times k_2$ tables. As base cases, we know this to be trivially true for all $k_1 \times 1$ and $1 \times k_2$ tables.

First we prove the inductive step in the special case that some sub-row or sub-column in the $k_1 \times k_2$ table consists entirely of '0's. Suppose we have an outcome sub-column of '0's for some measurement setting of Bob. Then if Bob makes this measurement the is just one possible local outcome, which occurs with certainty. We pick any '1' in the table. If this '1' is in the same measurement setting of Bob as the sub-column of '0's, then by no-signalling its sub-row has a '1' in each box of the same setting for Alice. We choose any other of these '1's and complete it to a

Table 2.5: Taking an arbitrary '1' (upper left) in the table of a no-signalling model forces the table to be of this form.

	1	$1 \cdots 1 \quad 0 \cdots 0$
	1 $\vdots$ 1 0 $\vdots$ 0	$\begin{matrix} * & \cdots & * \\ \vdots & \ddots & \vdots \\ * & \cdots & * \end{matrix}$

deterministic grid in the $k_1 \times (k_2 - 1)$ table obtained by ignoring the particular setting of Bob. Then, by no-signalling, this must complete to a $k_1 \times k_2$ deterministic grid. If the initial '1' is in a different measurement setting of Bob to the column of '0's, one can similarly forget the latter setting and apply the induction assumption to the remaining $k_1 \times (k_2 - 1)$ table. Again, the resulting deterministic grid in the sub-table completes uniquely to the whole table by no-signalling. A similar argument holds for sub-rows of '0's.

Now we need to prove the inductive step in the case that there are no sub-rows or sub-columns of '0's. By no-signalling, this is equivalent to no individual box having a sub-row/column of '0's. Hence we can assume that every box has a diagonal or anti-diagonal of '1's. We choose an arbitrary '1' in the table, which w.l.o.g. we can write in the upper left corner. By the inductive hypothesis, this can be completed to a $k_1 \times (k_2 - 1)$ deterministic grid, which w.l.o.g. we write in the upper left corners of all boxes up to Bob's $(k_2 - 1)$th setting (see table 2.6).

Assume that this deterministic grid does not complete to Bob's k_2th setting. Then there must be a '0' in the upper right corner of some box(es) of Bob's k_2th setting, and a '0' in the upper left corner of some box(es) in the same setting. In table 2.6, we have illustrated a representative situation, including the diagonals or anti-diagonals that these boxes must have. In order to avoid a Hardy paradox triggered by the '1's

Table 2.6: Table for the proof of theorem 2.5.3.

	1	$\cdots$	1	1 0 1
	1 *	$\cdots$	1 *	0 1 1
	1	$\cdots$	1	
		$\ddots$		

in the top sub-row, we must have '1's in the starred places, corresponding to all those sub-rows where the '0' in the k_2th setting of Bob occurs on the upper left. But now we can find a deterministic grid including the initial '1' for table 2.6 by choosing the second outcome for Alice in the case of a starred row and the first outcome otherwise, while choosing the first outcome for Bob in all measurements. $\square$

2.6 Applications

We now present some results that follow from the completeness theorems of the previous section.

Complexity

The theorems can tell us something about the computational complexity of recognising logical non-locality, which in the relevant scenarios is equivalent to deciding whether a Hardy paradox occurs.

Proposition 2.6.1. *Polynomial algorithms can be given for deciding non-locality in* $(2, 2, l)$ *and* $(2, k, 2)$ *models.*

Proof. For $(2, k, 2)$ scenarios, deciding whether a model is local or non-local simply amounts to checking all 2×2 sub-tables for such a Hardy paradox, which gives an algorithm that is polynomial in the size of the input: we check for the 64 possible Hardy configurations in each of $\binom{k}{2}^2$ sub-tables, which is clearly $O(k^4)$. For $(2, 2, l)$

Table 2.7: A ladder paradox. The $(2, 2, 2)$ ladder paradox is just the standard Hardy paradox.

	1	0		
	0	$*$	$\ddots$	
		$\ddots$		0
			0	0

scenarios, one has to check whether each '1' in the table can be completed to a deterministic grid. Following the illustration in table 2.5, it must be checked whether there is some '1' among the starred entries, which is equivalent to the non-occurrence of the coarse-grained Hardy paradox. There are $4l^2$ entries in the table, and each check is clearly $O(l^2)$. Again, we have an algorithm that is polynomial in the size of the input. $\square$

Ladder Paradoxes & Other Generalisations

The ladder paradox [29] has been proposed as a generalisation of the original Hardy paradox and was used for experimental tests of quantum non-locality [17]. Up to symmetries, there is one ladder paradox for any number of settings k. It can be presented neatly in tabular form (table 2.7). We will not explain here how the ladder paradox is in contradiction with locality, as our theorem 2.5.3 makes it clear that the ladder paradox has to be subsumed by the original Hardy paradox in terms of its strength for proving non-locality.

Proposition 2.6.2. *For $(2, k, 2)$ scenarios, the occurrence of a ladder paradox implies the occurrence of a Hardy paradox.*

Proof. This follows as a corollary of theorem 2.5.3, but one can also prove the proposition more directly. If the starred entry in table 2.7 is a '0', then the Hardy paradox

Table 2.8: The Chen et al. paradox occurs when at least one of the starred entries is non-zero. The relevant entries for each joint measurement are either those above or those below the diagonal.

$$\begin{array}{c|c|c|}
 & & \\ \hline
 & \begin{matrix} * & \cdots & * \\ & \ddots & \vdots \\ & & * \end{matrix} & \begin{matrix} 0 & \cdots & 0 \\ & \ddots & \vdots \\ & & 0 \end{matrix} \\ \hline
 & \begin{matrix} 0 & \cdots & 0 \\ & \ddots & \vdots \\ & & 0 \end{matrix} & \begin{matrix} 0 & & \\ \vdots & \ddots & \\ 0 & \cdots & 0 \end{matrix} \\ \hline
\end{array}$$

occurs; if it is a '1', then the ladder paradox for $k-1$ settings is triggered by this '1'. Applying the argument recursively, we find that either the Hardy paradox occurs somewhere in the table, or the ladder paradox for two settings occurs. Since the latter is again just a Hardy paradox, we find that Hardy's paradox occurs in any case. Hence the occurrence of a ladder paradox always implies the occurrence of a Hardy paradox. □

We also comment on a very recent paper by Chen et al. [37] which claims to provide a generalisation of Hardy's paradox for high-dimensional (qudit) systems. In the present terminology, their argument applies to $(2, 2, l)$ Bell scenarios. This will be relevant to the discussion in section 2.7. It is presented in tabular form in table 2.8. For this, theorem 2.5.2 (first published in [79]) implies that there must exist a coarse-graining of the measurements considered for which the model contains an ordinary Hardy paradox.

Proposition 2.6.3. *The occurrence of a Chen et al. paradox implies the occurrence of a Hardy paradox.*

Proof. Again, this follows directly from theorem 2.5.2, but one can also prove the proposition more directly. Suppose one of the starred entries corresponding to outcomes (o'_i, o_j) of table 2.8 is non-zero. We write $p(i, j) > 0$ for short. Then we can see from the table that for the joint measurement represented by the upper-right box, we must have $p(r, j) = 0$ for all $r > (l - j)$. Similarly, for the measurement represented by the lower-left box, $p(i, s) = 0$ for all $s > (l - i)$. In the lower-right box, we have $p(r, s) = 0$ when $r \leq (l - j)$ and $s \leq (l - i)$. This is a $(2, 2, l)$ Hardy paradox, or more precisely the $H_{(l-j,l-i)}$ paradox. □

The PR Box

Theorem 2.5.1 can be used to provide the first constructive proof of a result originally due to Lal [4, 74] that the only strongly non-local $(2, 2, 2)$ models are the Popescu-Rohrlich no-signalling boxes [93].

Proposition 2.6.4. *The only strongly non-local no-signalling* $(2, 2, 2)$ *models are the PR boxes.*

Proof. Recall from chapter 1 that strong non-locality is the property that no assignment of outcomes that is possible in the model can belong to a global assignment. In terms of the tabular representation this is simply the property that no '1' can be completed to a deterministic grid; and by the proof of theorem 2.5.1, strong non-locality is equivalent to the property that every '1' witnesses a Hardy paradox. Simply by using this characterisation of strong contextuality and the requirement that the model must be no-signalling we can prove the required result.

For any choice of measurements there must be some possible outcome (this is the 'normalisation of possibilities': the requirement that possibilities at each context form a well-defined Boolean distribution). This possible assignment is represented by a '1' in the table, and it must witness a Hardy paradox. After re-labelling as necessary, we can represent the model as in table 2.1. For this to be a no-signalling model, it is necessary to fill in '1's as in table 2.9 (a). Using the fact that the '1's in the lower-right box must also witness Hardy paradoxes, we must fill in '0's as in table 2.9 (b). By no-signalling, the remaining unspecified entry in the upper-left box must be a '1', and by the fact that it must witness a Hardy paradox, the remaining entry in the lower-right box must be a '0' .We thus arrive at table 2.9 (c), which is the PR box. □

Bell States are Anomalous

Projective measurements can be prescribed for almost all entangled two-qubit states such that the resulting empirical model will contain a Hardy paradox [60]. The prescription breaks down for the maximally entangled states, or the familiar Bell states.

Table 2.9: Stages in the proof of proposition 2.6.4.

	1		1	0
				1
	1		0	1
	0	1	1	

(a)

	1	0	1	0
	0		0	1
	1	0	0	1
	0	1	1	

(b)

	1	0	1	0
	0	1	0	1
	1	0	0	1
	0	1	1	0

(c)

This naturally raises the question of whether there are any projective measurements that can be chosen for the maximally entangled states such that the resulting empirical model contains a Hardy paradox. The question gains even more importance in light of the completeness theorems of section 2.5 which show that it is equivalent to asking whether the maximally entangled states are logically non-local.

Somewhat surprisingly, we answer this question in the negative, and show that no projective measurements can be chosen that lead to a Hardy paradox (and thus logical non-locality) for a maximally entangled state. To the author's knowledge, this is the first full proof of the fact. A related result showing that if the same two measurements are available at each site then it is impossible to realise a Hardy paradox was proved independently by Abramsky & Constantin [9]. The proof we are about to present holds for any number of measurements per qubit, and without the restriction that the same set of measurements should be available for each qubit.

This is remarkable since it shows that the Bell states are the only entangled two-qubit states not to be logically non-local. In fact, recent results indicate that all n-qubit entangled states are logically non-local for $n > 2$. It appears, therefore, that despite being perhaps the most studied and utilised states in the field of quantum information, the Bell states are actually anomalous.

Theorem 2.6.5. *Bell states are not logically non-local.*

Proof. We prove the statement for the Bell state

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right).$$

Since all other maximally entangled states are equivalent to this one up to local unitaries, which can easily be incorporated into the local measurements, the proof will extend to all maximally entangled states.

Any quantum mechanical empirical model obtained by making local projective measurements on $|\phi^+\rangle$ will necessarily give rise to a $(2, k, 2)$ model. By theorem 2.5.3 we know that Hardy's paradox completely characterises logical non-locality for such scenarios, and that logical non-locality implies the occurrence of a Hardy paradox in some $(2, 2, 2)$ sub-model. It therefore suffices to show that for any observables $\{A_1, A_2\}$ for the first qubit and $\{B_3, B_4\}$ for the second qubit the resulting model does not contain a Hardy paradox.

The $+1$ and -1 eigenvectors for these measurements will be given by

$$\begin{aligned}|0_i\rangle &= \cos\frac{\theta_i}{2}|0\rangle + e^{i\phi_i}\sin\frac{\theta_i}{2} \\ |1_i\rangle &= \sin\frac{\theta_i}{2}|0\rangle + e^{-i\phi_i}\cos\frac{\theta_i}{2}\end{aligned}$$

where $\{(\theta_i, \phi_i)\}_{i\in\{1,2,3,4\}}$ label the coordinates of the $+1$ eigenvector of the respective measurements on the Bloch sphere. The amplitudes of the outcomes of the various joint measurements are calculated to be:

$$\begin{aligned}\langle 0_j 0_k|\psi\rangle &= \frac{1}{\sqrt{2}}\left(\cos\frac{\theta_j}{2}\cos\frac{\theta_k}{2} + e^{-i(\phi_j+\phi_k)}\sin\frac{\theta_j}{2}\sin\frac{\theta_k}{2}\right) \\ \langle 0_j 1_k|\psi\rangle &= \frac{1}{\sqrt{2}}\left(\cos\frac{\theta_j}{2}\sin\frac{\theta_k}{2} + e^{-i(\phi_j-\phi_k)}\sin\frac{\theta_j}{2}\cos\frac{\theta_k}{2}\right) \\ \langle 1_j 0_k|\psi\rangle &= \frac{1}{\sqrt{2}}\left(\sin\frac{\theta_j}{2}\cos\frac{\theta_k}{2} + e^{i(\phi_j-\phi_k)}\sin\frac{\theta_j}{2}\cos\frac{\theta_k}{2}\right) \\ \langle 1_j 1_k|\psi\rangle &= \frac{1}{\sqrt{2}}\left(\sin\frac{\theta_j}{2}\sin\frac{\theta_k}{2} + e^{i(\phi_j+\phi_k)}\cos\frac{\theta_j}{2}\cos\frac{\theta_k}{2}\right)\end{aligned}$$

where $j \in \{1, 2\}$ and $k \in \{3, 4\}$. We see that $\langle 0_j 0_k|\psi\rangle = e^{-i(\phi_j+\phi_k)}\langle 1_j 1_k|\psi\rangle$ and $\langle 0_j 1_k|\psi\rangle = \langle 1_j 0_k|\psi\rangle$ for each choice of measurements. Thus the symmetry of the underlying state manifests itself as a symmetry in the probabilities of the joint outcomes for each choice of measurements:

$$p(01 \mid AB) = p(10 \mid AB) \tag{2.2}$$

$$p(00 \mid AB) = p(11 \mid AB). \tag{2.3}$$

We know from proposition 2.6.4 that the only strongly contextual $(2, 2, 2)$ models are the PR boxes, which are not quantum realisable [93]. So even though the PR box satisfies these symmetries, it cannot be realised by measurements on $|\phi^+\rangle$. We show that there is a unique $(2, 2, 2)$ model (up to re-labelling) that satisfies the symmetries (2.2) and (2.3) and is logically but not strongly non-local.

Table 2.10: Stages in the proof of theorem 2.6.5.

	1		1	
	1		1	

(a)

	1		1	
		1		1
	1		1	
		1		1

(b)

	1		1	0
	0	1		1
	1		1	
	1	1	0	1

(c)

	B_3		B_4	
A_1	1	0	1	0
	0	1	0	1
A_2	1	1	1	0
	1	1	0	1

(d)

If a model is not strongly non-local then there exists at least one global assignment compatible with the model, or in tabular form at least one deterministic grid. Up to re-labelling this is represented in table 2.10 (a). By the symmetry (2.3) there must exist a second global assignment, as in table 2.10 (b). It is clear from the configuration of the table that none of the entries that have already been specified can witness a Hardy paradox. If the model is logically non-local, therefore, at least one of the unspecified entries in table 2.10 (b) must witness a Hardy paradox. Up to re-labelling, this can be represented as in table 2.10 (c). By the symmetry (2.2) the table must be completed to table 2.10 (d). This (up to re-labelling) is the only possibilistic empirical model that respects the symmetries and is logically non-local without being strongly non-local. The question now is whether it can be realised by measurements on $|\phi^+\rangle$.

Consider the measurement statistics for the joint measurement A_1B_3 required by table 2.10. If these are to arise from quantum observables A_1 and B_3, then $\langle\phi^+|0_1 0_3\rangle = \langle\phi^+|1_1 1_3\rangle = \frac{1}{\sqrt{2}}$ and $\langle\phi^+|0_1 1_3\rangle = \langle\phi^+|1_1 0_3\rangle = 0$. So, either $|0_1\rangle = |0_3\rangle = |0\rangle$ and $|1_1\rangle = |1_3\rangle = |1\rangle$ up to an overall sign or vice versa. The eigenvectors of both observables are $\{|0\rangle, |1\rangle\}$, so they must simply be Pauli X operators (up to a common sign, which would allow for re-labelling the outcomes):

$$A_1 = B_3 = \pm X. \tag{2.4}$$

A similar argument applies for the joint measurements A_1B_4 and A_2B_4, showing that

$$A_1 = B_4 = \pm X, \tag{2.5}$$

$$A_2 = B_4 = \pm X. \tag{2.6}$$

Equations (2.4–2.6) imply that

$$A_1 = A_2 = B_3 = B_4 = \pm X;$$

but therefore the measurement statistics for A_2B_3 must be the same as for each of the other joint measurements, and it is not possible to realise table 2.10 (d). This completes the proof that no quantum mechanical logically non-local empirical model can be obtained by considering (any number of) local projective measurements on the Bell state. □

Symmetry is important here: the symmetry of the underlying state manifests itself as a symmetry of the probabilities of outcomes for each joint measurement. By theorem 2.5.1, logical non-locality implies a particular relationship between certain probabilities in each of these distributions (a Hardy paradox). However, quantum mechanically there cannot exist local projective measurements that realise these correlations and respect the symmetries at the same time. On the other hand, there exists a whole family of no-signalling empirical models which are logically non-local and respect the symmetries. These are the no-signalling models with support as in table 2.10 (d), and the PR box.

Fritz [51] has considered quantum analogues of Hardy's paradox. These are not realisable quantum mechanically, but can arise in more general no-signalling empirical models. An interesting point is that table 2.10 (d) contains two such paradoxes, and so the fact that any model with this support is not quantum realisable also follows from the results of [51].

We have mentioned already that this result singles out the Bell states as being unique among entangled qubit states. In section 2.8 we will see that the completeness of Hardy's paradox for logical non-locality breaks down outside of the scenarios that we have considered so far. Beyond qubit states, since there are more ways of being logically non-local, it appears less likely that such a situation might arise.

Tsirelson's Bound

We consider in more detail the family of logically non-local, no-signalling empirical models with support given by table 2.10 (d) that appeared in the proof of theorem 2.6.5. These models have some interesting properties.

Tsirelson [102] proved the existence of an upper bound on the degree to which any quantum mechanical $(2, 2, 2)$ empirical model can violate a CHSH inequality [40]. Several attempts have been made to find physical principles that account for this bound, such as the absence of third-order interference [42], information causality [89], and non-trivial communication complexity [33]. These last two are somewhat complicated by the fact that the properties must be proved on a case by case basis by finding appropriate protocols.

The models that we are interested in are all non-local, and we will show that many lie within the Tsirelson bound. In fact one can find models in this family that will violate the CHSH inequality by an arbitrarily small amount, and in this sense come arbitrarily close to the polytope of local models. What is surprising is that all models in the family are provably not quantum realisable.

This is important since it shows directly that the Tsirelson bound, which in any case only applies to the $(2, 2, 2)$ Bell scenario, does not completely characterise quantum correlations even here, and only provides a necessary condition for quantum realisability. Recently there has been some progress on completely characterising the set of $(2, 2, 2)$ quantum correlations by means of a convergent hierarchy of semi-definite programs [88]. At any rate, the fact that the Tsirelson bound does not provide a necessary and sufficient condition for quantum realisability to some extent weakens the argument that physical principles that account for the bound should necessarily be of fundamental importance to quantum mechanics, and single it out in the space of all no-signalling theories.

Similar families of models to this one have been discussed in [15], where it is shown that information causality can be used to provide an improvement over the Tsirelson bound in characterising quantum correlations. We note that this family can also be seen to violate information causality by means of the protocol from [89].

The Bell version [24] of the CHSH correlation function is

$$S := |E(A, B) + E(A', B') - E(A', B) + E(A', B')| , \tag{2.7}$$

where $E(A, B)$ is the probability that the outcomes to the joint measurement AB are correlated minus the probability that they are anti-correlated. Permuting the measurement labels at each site, or equivalently the signs of the terms in this expression

Table 2.11: A family of non-quantum, non-local empirical models, which lie within the Tsirelson bound for $0 < q < \sqrt{2} - 1$ and can be arbitrarily close to the local polytope.

	½	0	½	0
	0	½	0	½
	(1-q)⁄2	q⁄2	½	0
	q⁄2	(1-q)⁄2	0	½

will give other CHSH correlation functions (we have chosen the one that will achieve the maximum for the models we are interested in). It can easily be shown that for a local model,

$$S_{\max} \leq 2.$$

This is the CHSH inequality. For a deterministic model, $S = 2$ for each correlation function. Using the triangle inequality, any model that can be expressed as a stochastic mixture of deterministic models must therefore have $S \leq 2$ for each inequality. Tsirelson showed that the maximum achievable for a quantum empirical model is $S_{\max} = 2\sqrt{2}$ [102]. However, this is less than the algebraic maximum of $S = 4$, which Popescu and Rohrlich showed to be attainable by a no-signalling empirical model (the PR box).

Proposition 2.6.6. *The probabilistic empirical models defined by table 2.11 such that* $0 < q \leq 1$ *are logically non-local, not quantum realisable, and violate the CHSH inequality by* $2q$. *For* $0 < q < \sqrt{2} - 1$, *the models lie within the Tsirelson bound.*

Proof. This is the family of models that arose in the proof of theorem 2.6.5 together with the PR box ($q = 1$). We first show that any model in this family violates the CHSH inequality by $2q$. By inspection of table 2.11 it is clear that the only correlation function that can violate the CHSH bound is that of equation (2.7). This function has value $S = 2 + 2q$, whereas for the other functions, $S = 2q$. So $S_{\max} = 2 + 2q$ and the model violates the CHSH inequality by $2q$. The Tsirelson bound is achieved for $q = \sqrt{2} - 1$, and for $0 < q < 2\sqrt{2} - 1$ the models are non-local but lie within the Tsirelson bound.

To show that no model in this family is quantum realisable, it would suffice to notice that every model in the family contains several of the Fritz quantum analogues of Hardy's paradox [51]. However, we can also prove this more directly. We rely on a result due to Masanes [82] showing that any quantum mechanical $(2, 2, 2)$ empirical model is realisable by projective measurements on a two-qubit state. Therefore, if a model in this family is quantum realisable, it must be realised by some observables on a two-qubit state, say $\{A_1, A_2\}$ on the first and $\{B_3, B_4\}$ on the second. Each of these observables defines a basis $\{|0_i\rangle, |1_i\rangle\}$ for $\mathbb{C}^2$. If we consider the measurements A_1 and B_3, for example, then we can define a basis

$$\{|0_1\rangle \otimes |0_3\rangle, |0_1\rangle \otimes |1_3\rangle, |1_1\rangle \otimes |0_3\rangle, |1_1\rangle \otimes |1_3\rangle\}$$

of $\mathbb{C}^2 \otimes \mathbb{C}^2$ for the joint system. The underlying state can be decomposed in this basis; but then according to the upper-left box in table 2.11 we must have

$$|\psi\rangle = \pm\frac{1}{\sqrt{2}}\left(|0_1\rangle \otimes |0_3\rangle + |1_1\rangle \otimes |1_3\rangle\right) = \left|\phi^+\right\rangle.$$

Referring back to the proof of theorem 2.6.5, we have already shown that no model in this family is quantum mechanically realisable by measurements on the $|\phi^+\rangle$ state. We note that the local model for which $q = 0$, which this family approaches, is realisable with $A_1 = A_2 = B_3 = B_4 = \pm X$. □

2.7 Hardy Non-locality with Certainty

While Hardy's paradox is considered to be an 'almost probability free' non-locality proof, much of the literature on Hardy's paradox is concerned with the value of the *paradoxical probability*; i.e. the probability of obtaining the particular outcome assignment that witnesses a Hardy paradox (e.g. [29, 37]). This is especially relevant for experimental tests. In this section, we will show how Hardy non-locality can be demonstrated in such a way that even this probability becomes irrelevant.

As previously mentioned, Hardy [60] prescribed measurements for all entangled two-qubit states (excluding the maximally entangled ones) such that the resulting empirical model contains a Hardy paradox. For this family of models the maximum paradoxical probability is

$$p_{\max} = \frac{5\sqrt{5} - 11}{2} \approx 0.09\,. \tag{2.8}$$

We might think of this as providing a candidate Tsirelson-like bound for the paradoxical probability in the $(2, 2, 2)$ scenario, which by theorem 2.5.3 would extend to any

$(2, k, 2)$ scenario. We note, however, that this is only relevant to logical non-locality. For example, it is possible to achieve the Tsirelson violation of the CHSH inequality with the state $|\phi^+\rangle$ by choosing equatorial measurements at $\phi = \pi/8, 5\pi/8$ on the Bloch sphere for each qubit, though the resulting model will not exhibit logical non-locality (see chapter 6). So there exist non-local quantum mechanical empirical models for which the value of the paradoxical probability is always '0' .On the other hand, we will see shortly that it fares better than the Tsirelson bound in managing to exclude the family of non-quantum models from proposition 2.6.6.

A model has also been found for which the tripartite Hardy paradox can be witnessed with probability 0.125 [54], and in [39] it is demonstrated that for a generalised no-signalling theory it is possible to witness a $(2, 2, 2)$ Hardy paradox with probability 0.5. For the $(2, 2, l)$ scenario, Chen et al. have recently argued that it is possible to witness logical non-locality with probability ≈ 0.4 in the large d limit for two qu$d$it systems with the paradox presented in table 2.8. From our proposition 2.6.3, it becomes clear that they are essentially summing the probabilities of witnessing $(l-1)^2/2$ (coarse-grained) Hardy paradoxes.

In this section, we use our framework to gain a new perspective on this problem and achieve a striking improvement on these results. In particular, we will demonstrate by much simpler arguments how Hardy non-locality can be witnessed with certainty for a tripartite quantum system, and for a particular non-quantum $(2, 2, 2)$ empirical model. Interestingly, the models required for these arguments turn out to be the familiar GHZ-Mermin model, and the PR box.

The PR Box

We begin with a simple example to illustrate the idea.

Proposition 2.7.1. *The PR box witnesses a Hardy paradox with certainty.*

Proof. The probabilistic version of the PR box is given in table 2.12. We have already observed in the proof of proposition 2.6.4 that every assignment of outcomes that has non-zero probability witnesses a Hardy paradox. Each entry in the table therefore represents a paradoxical probability of 0.5; but for any joint measurement the probability of obtaining an outcome that witnesses a Hardy paradox is 1. □

The PR box achieves the maximum paradoxical probability of 0.5 for a no-signalling model found by Choudhary et al. in [39], but by a much simpler argument and using a familiar and well-studied model. Moreover, we see that the more relevant

Table 2.12: The PR box.

	½	0	½	0
	0	½	0	½
	½	0	0	½
	0	½	½	0

parameter, the *probability of witnessing some Hardy paradox* is actually 1 for any choice of measurements.

So the value of the paradoxical probability and the probability of witnessing a Hardy paradox need not be the same. This is what lies behind the fact that the Chen et al. paradox appears to violate the Hardy bound (2.8). However, we can prove that for any quantum realisable $(2, 2, 2)$ empirical model these necessarily coincide.

Proposition 2.7.2. *For any quantum realisable* $(2, k, 2)$ *empirical model, the paradoxical probability and the probability of witnessing a Hardy paradox coincide.*

Proof. First, we note that by theorem 2.5.3 it suffices to prove the proposition for $(2, 2, 2)$ models. If the probabilities do not coincide, then it must be the case that, for some joint measurement, more than one Hardy paradox may be witnessed. Working within our formalism, it is clear that any such empirical model must be of the form of table 2.11 up to re-labelling of measurements and outcomes. In this family, for a particular joint measurement, the probability of witnessing a Hardy paradox is always '1'. However, we have proved in proposition 2.6.6 that no model in this family is quantum realisable, and the result follows. □

This shows that even taking into account that the probability of witnessing a Hardy paradox may in general be higher than the paradoxical probability, the Hardy bound for $(2, k, 2)$ models still holds. In fact, it excludes the family of non-quantum models from proposition 2.6.6 which were seen to lie within the Tsirelson bound.

GHZ

We now consider the $(3, 2, 2)$ empirical model used in the Mermin version [83] of the GHZ non-locality argument [57], which we note is not of the tripartite Hardy form. The model was encountered already in chapter 1 as an example of a strongly contextual model. Here, we need only consider a subset of the measurement contexts.

	000	001	010	011	100	101	110	111
$X\ X\ X$	1	0	0	1	0	1	1	0
$X\ Y\ Y$	0	1	1	0	1	0	0	1
$Y\ X\ X$	0	1	1	0	1	0	0	1
$Y\ Y\ X$	0	1	1	0	1	0	0	1

The suppressed rows of the table $\{XXY, XYX, YXX, YYY\}$ have full support. Figure 2.2 (a) depicts the model in the three dimensional representation.

Proposition 2.7.3. *The GHZ model witnesses a Hardy paradox with certainty.*

Proof. The three dimensional representation makes it easy to identify an n-partite Hardy paradox, which is shown in figure 2.2 (b). It can be expressed algebraically as follows.

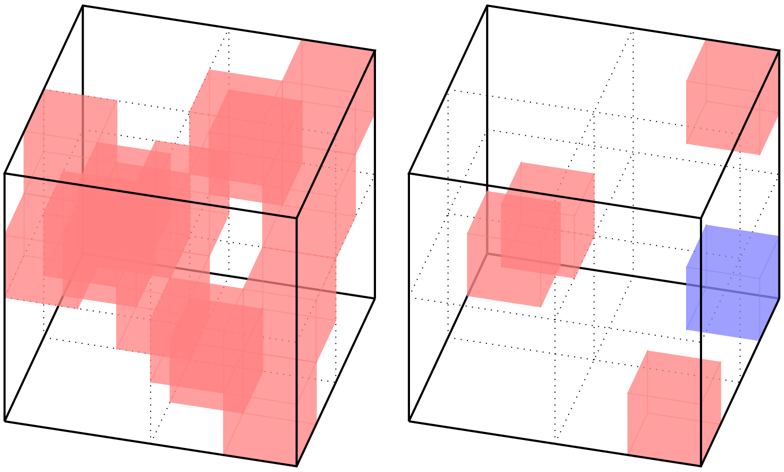
- $p(\,1,1,1 \mid Y,Y,Y\,) > 0$
- $p(\,1,1,0 \mid Y,Y,X\,) = p(\,1,0,1 \mid Y,X,Y\,) = p(\,0,1,1 \mid X,Y,Y\,) = 0$
- $p(\,0,0,0 \mid X,X,X\,) = 0$

Up to re-labelling, this is the form of the n-partite Hardy paradox defined in section 2.4. Moreover, it can similarly be demonstrated that any joint outcome for the measurement context YYY witnesses a Hardy paradox (a more careful treatment will be given in the proof of proposition 2.7.4).

The paradoxical probability is $p(1,1,1 \mid Y,Y,Y) = 0.125$. However, since every outcome to the measurement YYY witnesses some Hardy paradox, then it is clear that the probability of witnessing a Hardy paradox is actually 1. $\square$

This provides a much simpler tripartite Hardy argument than that of Ghosh, Kar and Sarkar [54]. We obtain their maximum of 0.125 for the paradoxical probability, which was also obtained on the GHZ state but with different measurements. We do better, however, since with certainty we must witness some Hardy paradox for the joint measurement YYY. The model here is exactly the GHZ-Mermin model, since the observables available at each subsystem are simply the X and Y operators, so in fact what we have shown is that the GHZ experiment [57] witnesses Hardy non-locality with certainty.

Figure 2.2: (a) The GHZ model. We represent only the red, impossible outcomes; all other entries are possible. (b) Hardy's paradox within the GHZ model; the blue outcome is possible.



(a) (b)

GHZ(n)

Mermin gave logical non-locality proofs for n-partite generalisations of the GHZ state [87] for all $n > 2$. Again, his arguments were not of the Hardy form, but we will now show how to generalise proposition 2.7.3 to some of the GHZ(n) models.

The GHZ(n) states are:

$$|\mathrm{GHZ}(n)\rangle := \frac{1}{\sqrt{2}} \left(|0 \cdots 0\rangle + |1 \cdots 1\rangle \right), \tag{2.9}$$

where n is the number of qubits. Note that for $n = 2$ the state obtained is the $|\phi^+\rangle$ Bell state. For $n > 2$, Mermin considered models in which each each party can make Pauli X or Y measurements. With a little calculation, it is possible to concisely describe the resulting empirical models in a logical form.

The eigenvectors of the X operator are

$$|0_x\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + e^{i0} |1\rangle \right), \qquad |1_x\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + e^{i\pi} |1\rangle \right). \tag{2.10}$$

The vector $|0_x\rangle$ has eigenvalue $+1$ and the vector $|1_x\rangle$ has eigenvalue -1. These are more usually denoted $|+\rangle$ and $|-\rangle$, respectively, but we use an alternative notation to agree with our usual $\{0, 1\}$ labelling for outcomes. Similarly, the $+1$ and -1 eigenvectors of the Y operator are

$$|0_y\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + e^{i\pi/2} |1\rangle \right) \qquad |1_y\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + e^{-i\pi/2} |1\rangle \right). \tag{2.11}$$

The phases have been made explicit since they will play the crucial role in the following calculations.

The various probabilities for the empirical model predicted by quantum mechanics can be calculated as

$$|\langle \mathrm{GHZ}(n)|v_1 \dots v_n\rangle|^2$$

where the v_i are the appropriate eigenvectors. This evaluates to

$$\left|\frac{1+e^{i\phi}}{\sqrt{2^{n+1}}}\right|^2 = \frac{1}{2^n}\left(1+\cos\phi\right), \tag{2.12}$$

where ϕ is the sum of the phases of the v_i. From the phases of the possible eigenvectors, (2.10) and (2.11), it is clear that we must have $\phi = k\,\pi/2$ for some $k \in \mathbb{Z}_4$, the four element cyclic group. For $k = 0 \pmod 4$, the probability will be $\frac{1}{\sqrt{2^{n-1}}}$; for $k = 1$ or $3 \pmod 4$ the probability will be $\frac{1}{\sqrt{2^n}}$; and for $k = 2 \pmod 4$ the probability will be '0'.

We can now reduce the calculation of probabilities for any such model into a simple counting argument. If k_{0_x} is the number of $|0_x\rangle$ eigenvectors, k_{1_x} is the number of $|1_x\rangle$ eigenvectors, and so on, then

$$\begin{aligned} k &= k_{0_y} + 2\cdot k_{1_x} + 3\cdot k_{1_y} \pmod 4 \\ &= \left(k_{0_y} + k_{1_y}\right) + 2\cdot\left(k_{1_x} + k_{1_y}\right) \pmod 4. \end{aligned}$$

- For contexts containing an odd number of Y's, every outcome is possible with equal probability $\frac{1}{\sqrt{2^n}}$, since $k = 1$ or $3 \pmod 4$.

- For contexts containing 0 mod 4 Y's, outcomes are possible if and only if they contain an even number of 1's. For these outcomes, $k = 0 \pmod 4$ and the probabilities are $\frac{1}{\sqrt{2^{n-1}}}$. If there were an odd number of 0's in the outcome then $k = 2 \pmod 4$ and the probability would be 0.

- Similarly, for contexts that contain 2 mod 4 Y's, outcomes are possible if and only if they contain an odd number of 1's. Again, the non-zero probabilities are $\frac{1}{\sqrt{2^{n-1}}}$.

Though the probabilities are seen to be easily be calculated in this way, we need only concern ourselves with the possibilistic information.

Proposition 2.7.4. *All GHZ(n) models for $n = 3 \bmod 4$ witness an n-partite Hardy paradox with certainty.*

Proof. Proposition 2.7.3 showed that this holds for GHZ(3). Let $\overline{o} = (o_1, \ldots, o_n)$ be any binary string of length n, let ϕ_i be the function that changes the ith entry of a binary string, and let $\overline{o}^{-1}$ denote the binary string of length n which differs in every entry from $\overline{o}$. We show that every outcome $\overline{o}$ to the measurements $(Y, \ldots, Y)$ witnesses a Hardy paradox. We deal with the cases that $\overline{o}$ has an even or odd number of 1's separately.

Suppose $\overline{o}$ has an even number of 1's.

- $p\,(\overline{o} \mid Y, \ldots, Y) > 0$, since there are an odd number of Y measurements;
- $p\,(\,\overline{o} \mid \pi(X, Y, \ldots, Y)\,) = 0$, for all permutations π, since there are 2 mod 4 Y's and $\overline{o}$ has an even number of 1's;
- $p\,(\overline{o}^{-1} \mid X, \ldots, X) = 0$, since there are 0 mod 4 Y's and $\overline{o}^{-1}$ has an odd number of 1's.

Suppose $\overline{o}$ has an odd number of 1's.

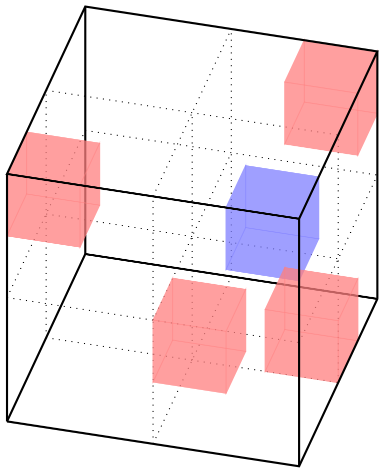
- $p\,(\overline{o} \mid Y, \ldots, Y) > 0$, since there are an odd number of Y measurements;
- $p\,(\,\phi_i(\overline{o}) \mid \phi_i(Y, Y, \ldots, Y)\,) = 0$, for all permutations $i = 1, \ldots, n$, since there are 2 mod 4 Y's and an even number of 1's in $\phi_i(\overline{o})$;
- $p\,(\overline{o} \mid X, \ldots, X) = 0$, since there are 0 mod 4 Y's and an odd number of 1's in $\overline{o}$.

□

It should be pointed out that even though we can say with certainty that some Hardy paradox will be witnessed in these models, the paradoxical probabilities are $\frac{1}{2^n}$, and so the maximal paradoxical probability is obtained for the tripartite GHZ model.

This kind of result does not hold for GHZ(n) models for which $n \neq 3$ mod 4, as it can be shown that these models do not contain n-partite Hardy paradoxes. This is because any $(n, 2, 2)$ Hardy paradox must take the form of one of the paradoxes in the proof of proposition 2.7.4, but it can easily be verified that the counting arguments only allow these for $n = 3$ mod 4.

Figure 2.3: A non-Hardy $(3, 2, 2)$ paradox. We assume that all unspecified entries are possible. This model does not contain a Hardy paradox, but is logically non-local since the blue entry cannot be completed to a deterministic grid.



2.8 Non-universality of Hardy's Paradox

The completeness results of section 2.5 might raise the conjecture that the Hardy paradox could be universal in the same sense for any $(2, k, l)$ scenario. However, we have found that the equivalence of locality to the absence of Hardy-type non-locality does not hold for $(2, k, l)$ scenarios in general: consider the probabilistic empirical model displayed in table 2.13 (b), for example. This concerns a Bell scenario with three two-outcome measurements for Alice, and one two-outcome and one three-outcome measurement for Bob. (This can easily be expanded to a probabilistic empirical model in the $(2, 3, 3)$ scenario, but we find the example easier to understand in the form of table 2.13.) By direct inspection, we find that no coarse-grained Hardy paradox occurs for this empirical model. Nevertheless, it displays logical (and hence probabilistic) non-locality: the '1' in the upper left corner of table 2.13 (a) cannot be completed to a deterministic grid.

We have already seen at the end of section 2.7 that the Hardy paradox does not occur in GHZ(n) models for $n \neq 3 \bmod 4$, though it is known that these are logically non-local. So for $(4, 2, 2)$ scenarios we know that completeness must break down. However, already for the $(3, 2, 2)$ scenario we have been able to find a logically non-local model (figure 2.3) for which no Hardy paradox occurs.

In conclusion, the Hardy paradox and its coarse-grainings cannot account for all non-local behaviour in scenarios with at least three parties, or with at least three settings and at least three outcomes. In general, the non-occurrence of a Hardy paradox is necessary but not sufficient for possibilistic locality.

Table 2.13: (a) A non-locality proof without inequalities; (b) a probabilistic no-signalling model to which it applies although it displays no (coarse-grained) Hardy paradox.

1		0		
0				
			0	
0				
				0

(a)

1/16	3/16	0	1/8	1/8
3/16	9/16	1/2	1/8	1/8
0	1/2	1/8	1/4	1/8
1/4	1/4	3/8	0	1/8
0	1/2	1/8	1/8	1/4
1/4	1/4	3/8	1/8	0

(b)

2.9 Discussion

To begin with, we have investigated the scope of Hardy's non-locality paradox in terms of non-local behaviour. We have proved a number of completeness theorems showing that it is a necessary and sufficient condition for logical non-locality in $(2, 2, l)$ and $(2, k, 2)$ Bell scenarios. In this sense, it is the only non-locality proof without inequalities for these Bell scenarios. We can even interpret the possibilistic versions of the no-signalling condition and the normalisation of probabilities as degenerate cases of the non-occurrence of a coarse-grained Hardy paradox.

However, we have found that this universality does not extend to the $(2, 3, 3)$ Bell scenario, nor does it extend to n-partite scenarios for $n > 2$. This raises the question of finding other logical non-locality conditions that do not belong to the class of Hardy paradoxes for $n, k, l > 2$. The GHZ(n) models of section 2.7, for example, are logically non-local but do not contain $(n, 2, 2)$ Hardy paradoxes for $n \neq 3 \bmod 4$.

The completeness theorems of section 2.5 have led to a number of interesting applications. We have seen that for $(2, 2, l)$ and $(2, k, 2)$ scenarios, polynomial algorithms can be given for deciding non-locality. It was conjectured in [79] that the general decidability problem for possibilistic local models with k as the free input is NP-hard when $n > 2, l \geq 2$ or $n \geq 2, l > 2$; as is the case for probabilistic models [92]. It was shown that the problem is NP by Abramsky in [3], and the it has since been proved to be NP-complete by Abramsky, Gottlob and Kolaitis [10]. This gives some reason to suspect that it is not possible to obtain a classification of conditions that are necessary and sufficient for logical non-locality in full generality.

Another direct consequence is that the Hardy paradox must subsume all other non-locality arguments for $(2, 2, l)$ and $(2, k, 2)$ scenarios, and we have demonstrated this for the ladder paradoxes and the Chen et al. paradox. Furthermore, the theorems have been used to provide the first constructive proof that the PR boxes are the only strongly contextual $(2, 2, 2)$ models, as well as the first full proof that the Bell states, despite being maximally entangled, are the only entangled two-qubit states that are not logically non-local.

Together with recent work by Ying [13] which shows that all entangled n-partite qubit states are logically non-local, this singles out the Bell states as being anomalous in terms of non-locality. This is quite surprising in light of the fact that they are perhaps the most studied and utilised of entangled states. We mention, however, that it remains to be seen whether the result still holds when we allow for POVM's. The proof of theorem 2.6.5 is quite interesting in itself, and led to the discovery of a family of non-quantum empirical models which lie within the Tsirelson bound and can have an arbitrarily small violation of the CHSH inequality. Interestingly, the models violate information causality, which has been proposed a candidate property for characterising quantum correlations, and also violate the Hardy bound on paradoxical probabilities.

In section 2.7, we have taken advantage of the perspective gained within our framework to demonstrate a striking improvement on the probability of witnessing a Hardy paradox, which is often used in the literature as a measure of the quality of Hardy non-locality. With much simpler arguments, it has been demonstrated that a tripartite quantum system can in fact witness Hardy non-locality with certainty. Interestingly, the empirical model used for this proof was exactly that of the GHZ-Mermin non-locality proof. Though it is not quantum realisable, we have also shown that the PR box has this property.

A further interesting point is that Abramsky [2] has uncovered a correspondence between possibilistic empirical models and relational database theory. It remains to be explored whether the completeness theorems of this chapter might find applications in database theory, or indeed whether similar results already exist in the field that might lead to further insights.

Chapter 3

The Cohomology of Non-locality & Contextuality

We have seen in chapter 1 that the mathematics of sheaf theory can provide a natural framework in which to analyse the structure of non-locality and contextuality. Empirical models form compatible (no-signalling) families of sections on a presheaf of distributions that is defined on a cover corresponding to the sets of compatible measurements. Locality and non-contextuality are characterised in a unified manner by the existence of global sections. Therefore, the phenomena of non-locality and contextuality can be characterised in terms of obstructions to the existence of global sections.

Roughly speaking, cohomology theories can be thought of as descriptions of obstructions to solving some kind of equation (see e.g. chapter 3 of [64] for some intuitive examples). The aim in this chapter, which is largely based on [12], is to build on these results, using the powerful tools of presheaf cohomology to study the structure of non-locality and contextuality and provide a positive characterisation of obstructions to global sections, and by extension of non-locality and contextuality. The possible application of cohomology to the study of contextuality in the sense of the Kochen-Specker theorem was first suggested by Isham & Butterfield [68]. The results here provide the first steps in this direction.

We succeed in finding cohomological witnesses of non-locality and contextuality which correspond to many of the classic no-go results. The approach is not yet strong enough, however, to completely characterise these phenomena for all models. We will discuss certain situations in which cohomology can fail to identify contextuality, which merit further investigation.

More precisely, we use the Čech cohomology on an abelian presheaf derived from the support of the model in order to define a cohomological obstruction for the family as a certain cohomology class. This class vanishes if the family has a global section.

Thus, in general, the non-vanishing of the obstruction provides a sufficient (but not necessary) condition for the model to be contextual. It can be demonstrated that for a number of salient examples, including PR boxes [93], the GHZ-Mermin model [83], and the 18-vector configuration giving a proof of the Kochen-Specker theorem in $\mathbb{R}^4$ [34], the obstruction does not vanish, thus yielding cohomological witnesses for contextuality. Moreover, we prove that for large classes of models generalising the state-independent models of these Kochen-Specker proofs, the cohomological characterisation provides a complete invariant for contextuality. These general results also raise an interesting connection between contextuality of Kochen-Specker models and the existence of perfect matchings in hypergraphs, leading to a number of useful insights.

3.1 Čech Cohomology of a Presheaf

Let X be a topological space, $\mathcal{U}$ be an open cover of X, and $\mathcal{F}$ be a presheaf of abelian groups on X. So $\mathcal{F}(U)$ is an abelian group for each open set $U \in \mathcal{U}$, and when $U \subseteq V \in \mathcal{U}$, there exists a group homomorphism $\rho^V_U : \mathcal{F}(V) \to \mathcal{F}(U)$. These assignments are functorial:

$$\rho^U_U = \mathsf{id}_U,$$

and if $U \subseteq U' \subseteq U''$ then

$$\rho^{U'}_U \circ \rho^{U''}_{U'} = \rho^{U''}_U .$$

The *nerve* $\mathsf{N}(\mathcal{U})$ of the cover $\mathcal{U}$ is defined to be the abstract simplicial complex comprising those finite subsets of $\mathcal{U}$ with non-empty intersection. Concretely, we take a q-simplex to be a list $\sigma = (U_0, \ldots, U_q)$ of elements of $\mathcal{U}$, with $|\sigma| := \cap^q_{i=0} U_i \neq \emptyset$. Thus a 0-simplex (U) is a single element of the cover $\mathcal{U}$. We write $\mathsf{N}(\mathcal{U})^q$ for the set of q-simplices.

Given a $q+1$-simplex $\sigma = (U_0, \ldots, U_{q+1})$, we can obtain q-simplices

$$\partial_j(\sigma) := (U_0, \ldots, \widehat{U_j}, \ldots, U_{q+1}), \qquad 0 \leq j \leq q$$

by omitting any one of the elements of the $q+1$-simplex. Note that:

$$|\sigma| \;\subseteq\; |\partial_j(\sigma)|.$$

We now define the *Čech cochain complex*. For each $q \geq 0$, we define the abelian group $C^q(\mathcal{U}, \mathcal{F})$:

$$C^q(\mathcal{U}, \mathcal{F}) \;:=\; \prod_{\sigma \in \mathsf{N}(\mathcal{U})^q} \mathcal{F}(|\sigma|).$$

We also define the *coboundary maps*

$$\delta^q : C^q(\mathfrak{U}, \mathcal{F}) \longrightarrow C^{q+1}(\mathfrak{U}, \mathcal{F}).$$

For $\omega = (\omega(\tau))_{\tau \in \mathsf{N}(\mathfrak{U})^q} \in C^q(\mathfrak{U}, \mathcal{F})$ and $\sigma \in \mathsf{N}(\mathfrak{U})^{q+1}$,

$$\delta^q(\omega)(\sigma) \ := \ \sum_{j=0}^{q+1} (-1)^j \, \rho_{|\sigma|}^{|\partial_j(\sigma)|} \, \omega(\partial_j \sigma).$$

For each q, δ^q is a group homomorphism.

Proposition 3.1.1. *For each q, $\delta^{q+1} \circ \delta^q = 0$.*

By this proposition, $C^\bullet(\mathfrak{U}, \mathcal{F})$ is a *cochain complex*. We will also consider the *augmented complex*

$$\mathbf{0} \to C^0(\mathfrak{U}, \mathcal{F}) \to \cdots .$$

We define $Z^q(\mathfrak{U}, \mathcal{F})$, the *$q$-cocycles*, to be the kernel of δ^q, and $B^q(\mathfrak{U}, \mathcal{F})$, the *$q$-coboundaries*, to be the image of δ^{q-1}. These are subgroups of $C^q(\mathfrak{U}, \mathcal{F})$ and, by proposition 3.1.1, we have $B^q(\mathfrak{U}, \mathcal{F}) \subseteq Z^q(\mathfrak{U}, \mathcal{F})$. We define the *$q$-th Čech cohomology group* $\check{H}^q(\mathfrak{U}, \mathcal{F})$ to be the quotient group $Z^q(\mathfrak{U}, \mathcal{F})/B^q(\mathfrak{U}, \mathcal{F})$. Note that $B^0(\mathfrak{U}, \mathcal{F}) = \mathbf{0}$, so $\check{H}^0(\mathfrak{U}, \mathcal{F}) \cong Z^0(\mathfrak{U}, \mathcal{F})$.

Given a cocycle $z \in Z^q(\mathfrak{U}, \mathcal{F})$, the *cohomology class* $[z]$ is the image of z under the canonical map

$$Z^q(\mathfrak{U}, \mathcal{F}) \longrightarrow \check{H}^q(\mathfrak{U}, \mathcal{F}).$$

A *compatible family* with respect to a cover $\mathfrak{U}$ is a family $\{r_i \in \mathcal{F}(U_i)\}$ for $U_i \in \mathfrak{U}$, such that, for all i, j:

$$r_i|_{U_i \cap U_j} = r_j|_{U_i \cap U_j}.$$

Proposition 3.1.2. *There is a bijection between compatible families and elements of the zeroth cohomology group $\check{H}^0(\mathfrak{U}, \mathcal{F})$.*

Proof. Cochains $c = (r_i)_{U_i \in \mathfrak{U}}$ in $C^0(\mathfrak{U}, \mathcal{F})$ correspond to families $\{r_i \in \mathcal{F}(U_i)\}$. For each 1-simplex $\sigma = (C_i, C_j)$,

$$\delta^0(c)(\sigma) \ = \ r_i|_{C_i \cap C_j} \ - \ r_j|_{C_i \cap C_j}.$$

Hence $\delta^0(c) = 0$ if and only if the corresponding family is compatible. □

We will also use the *relative cohomology* of $\mathcal{F}$ with respect to an open subset $U \subseteq X$. We define two auxiliary presheaves related to $\mathcal{F}$. First, $\mathcal{F}|_U$ is defined by

$$\mathcal{F}|_U(V) := \mathcal{F}(U \cap V).$$

There is an evident presheaf morphism

$$p : \mathcal{F} \longrightarrow \mathcal{F}|_U, \quad p_V :: r \mapsto r|_{U \cap V}.$$

Then $\mathcal{F}_{\bar{U}}$ is defined by $\mathcal{F}_{\bar{U}}(V) := \ker(p_V)$. Thus we have an exact sequence of presheaves

$$\mathbf{0} \longrightarrow \mathcal{F}_{\bar{U}} \longrightarrow \mathcal{F} \xrightarrow{p} \mathcal{F}|_U.$$

The relative cohomology of $\mathcal{F}$ with respect to U is defined to be the cohomology of the presheaf $\mathcal{F}_{\bar{U}}$.

We have the following refined version of proposition 3.1.2.

Proposition 3.1.3. *For any $U_i \in \mathfrak{U}$, the elements of the relative cohomology group $\check{H}^0(\mathfrak{U}, \mathcal{F}_{\bar{U}_i})$ correspond bijectively to compatible families $\{r_j\}$ on $\mathcal{F}$ such that $r_i = 0$.*

Proof. By proposition 3.1.2, compatible families correspond to cocycles $r = (r_j)$ in $C^0(\mathfrak{U}, \mathcal{F})$. By compatibility, $r_i|_{C_i \cap C_j} = r_j|_{C_i \cap C_j}$ for all j. Hence r is in $C^0(\mathfrak{U}, \mathcal{F}_{\bar{U}_i})$ if and only if $r_i = p_{U_i}(r_i) = 0$. □

3.2 Cohomological Obstructions

Recall that the *support* $\mathsf{supp}(\phi)$ of a function $\phi : X \to R$, where R is any commutative ring, is the set of all $x \in X$ such that $\phi(x) \neq 0$. We define a functor $F_R : \mathbf{Set} \longrightarrow \mathbf{Set}$ such that $F_R(X)$ is the set of functions $\phi : X \to R$ with finite support. There is an embedding $x \mapsto 1 \cdot x$ of X in $F_R(X)$, which we will use implicitly throughout. Given $f : X \to Y$, we define:

$$F_R f : F_R X \longrightarrow F_R Y :: \phi \mapsto [y \mapsto \sum_{f(x)=y} \phi(x)].$$

This assignment is easily seen to be functorial.

In fact, $F_R(X)$ is the *free R-module generated by X*. It is an abelian group, and $F_R(f)$ is a group homomorphism for any function f. In particular, taking $R = \mathbb{Z}$, $F_{\mathbb{Z}}(X)$ is the *free abelian group generated by X*. Thus, given any presheaf of sets P on X, we obtain a presheaf of abelian groups $F_{\mathbb{Z}}P$ by composition: $F_{\mathbb{Z}}P(U) := F_{\mathbb{Z}}(P(U))$.

Given an empirical model e defined on the the measurement scenario $(X, O, \mathcal{M})$, we will be interested in the relative Čech cohomology groups $\check{H}^q(\mathcal{M}, \mathcal{F}_{\bar{C}})$ for the abelian presheaf $\mathcal{F} := F_{\mathbb{Z}}S_e$ and the open cover $\mathcal{M}$ of maximal contexts of X. Note that $\mathcal{F}(C)$ is the set of *formal $\mathbb{Z}$-linear combinations of assignments* in the support of e_C for any set of measurements $C \in \mathcal{M}$.

To each $s \in S_e(C)$ we associate an element $\gamma(s)$ of a cohomology group, which can be regarded as an obstruction to s having an extension within the support of e to a global section. In particular, the existence of such an extension implies that the obstruction vanishes. In good cases, these two conditions are equivalent, yielding *cohomological characterisations* of contextuality and strong contextuality.

For notational convenience, we fix an element $s = s_1 \in S_e(C_1)$. Due to the compatibility (no-signalling) of the empirical model $\{e_C\}$, there must exist some family $\{s_i \in S_e(C_i)\}$ with $s_1|C_1 \cap C_i = s_i|C_1 \cap C_i$ for $i = 2, \ldots, n$. We define the cochain $c := (s_1, \ldots, s_n) \in C^0(\mathcal{M}, \mathcal{F})$. The coboundary of this cochain is $z := \delta^0(c)$.

Proposition 3.2.1. *The coboundary z of c vanishes under restriction to C_1, and hence is a cocycle in the relative cohomology with respect to C_1.*

Proof. We write $C_{i,j} := C_i \cap C_j$. For all i, j, we define $z_{i,j} := z(C_{i,j}) = s_i|_{C_{i,j}} - s_j|_{C_{i,j}}$. Due to the compatibility of the family $\{s_i\}$, for all i, j,

$$s_i|_{C_1 \cap C_{i,j}} = (s_1|_{C_1 \cap C_i})|_{C_j} = s_1|_{C_1 \cap C_{i,j}}.$$

Similarly, $s_j|_{C_1 \cap C_{i,j}} = s_1|_{C_1 \cap C_{i,j}}$. Hence $z_{i,j}|_{C_1} = 0$, and so $z_{i,j} \in \mathcal{F}_{\bar{C}_1}(C_i \cap C_j)$. Thus $z = (z_{i,j})_{i,j} \in C^1(\mathcal{M}, \mathcal{F}_{\bar{C}_1})$. Note that $\delta^1 : C^1(\mathcal{M}, \mathcal{F}_{\bar{C}_1}) \to C^2(\mathcal{M}, \mathcal{F}_{\bar{C}_1})$ is the restriction of the coboundary map on $C^1(\mathcal{M}, \mathcal{F})$. Hence $z = \delta^0(c)$ is a cocycle. $\square$

Definition 3.2.2. *We define $\gamma(s_1)$ to be the cohomology class $[z] \in \check{H}^1(\mathcal{M}, \mathcal{F}_{\bar{C}_1})$.*

Note that, although $z = \delta^0(c)$, it is not necessarily a coboundary in $C^1(\mathcal{M}, \mathcal{F}_{\bar{C}_1})$, since c is not a cochain in $C^0(\mathcal{M}, \mathcal{F}_{\bar{C}_1})$, as $p_{C_i}(s_i) = s_i|C_1 \cap C_i \neq 0$. Thus, in general, we need not have $[z] = 0$.

Remark There is a more conceptual way of defining this obstruction, using the connecting homomorphism from the long exact sequence of cohomology; see [55]. We have given a more concrete formulation, which may be easier to grasp, and is also convenient for computation.

Proposition 3.2.3. *The following are equivalent:*

1. *The cohomological obstruction vanishes:* $\gamma(s_1) = 0$.

2. *There is a family* $\{r_i \in \mathcal{F}(C_i)\}$ *with* $s_1 = r_1$*, and for all* i, j*,*

$$r_i|C_i \cap C_j = r_j|C_i \cap C_j.$$

Proof. The obstruction vanishes if and only if there is a cochain $c' = (c'_1, \ldots, c'_n) \in C^0(\mathcal{M}, \mathcal{F}_{\bar{C}_1})$ with $\delta^0(c') = \delta^0(c)$, or equivalently $\delta^0(c - c') = 0$ (i.e. such that $c - c'$ is a cocycle). By proposition 3.1.2, this is equivalent to $\{r_i := s_i - c'_i\}$ forming a compatible family. Moreover, $c' \in C^0(\mathcal{M}, \mathcal{F}_{\bar{C}_1})$ implies that $c'_1 = p_{C_1}(c'_1) = 0$, so we have $r_1 = s_1$.

For the converse, suppose we have a family $\{r_i \in \mathcal{F}(C_i)\}$ as in the second statement. We define $c' := (c'_1, \ldots, c'_n)$, where $c'_i := s_i - r_i$. Since $r_1 = s_1$, we find that $p_{C_i}(c'_i) = s_1|C_{1,i} - r_1|C_{1,i} = 0$ for all i, and $c' \in C^0(\mathcal{M}, \mathcal{F}_{\bar{C}_1})$. We must show that $\delta^0(c') = z$ (i.e. that $z_{i,j} = c'_i|C_{i,j} - c'_j|C_{i,j}$); but this holds since $r_i|C_{i,j} = r_j|C_{i,j}$. □

As an immediate application to contextuality, we have the following.

Proposition 3.2.4. *If a model* e *is possibilistically extendable, then the obstruction vanishes for every assignment in the support of the model. If* e *is not strongly contextual, then the obstruction vanishes for some assignment in the support.*

Proof. If e is possibilistically extendable, then for every $s \in S_e(C_i)$, there is a compatible family $\{s_j \in S_e(C_j)\}$ with $s = s_i$. Applying the embedding of $S_e(C_j)$ into $\mathcal{F}(C_j)$, by proposition 3.2.3 we conclude that $\gamma(s) = 0$. The same argument can be applied to a single assignment witnessing the failure of strong contextuality. □

Thus the non-vanishing of the obstruction is a sufficient condition for contextuality. The non-necessity of the condition arises from the possibility of 'false positives': families $\{r_i \in \mathcal{F}(C_i)\}$ that do not determine *bona fide* global assignments in $\mathcal{E}(X)$.

The Hardy Model

This first example shows that false positives do indeed arise. It is the Hardy model [60], which was examined in detail in chapter 2. In the more usual representation, the support is described as follows.

	00	01	10	11
$A\ B$	1	1	1	1
$A\ B'$	0	1	1	1
$A'\ B$	0	1	1	1
$A'\ B'$	1	1	1	0

For convenience, we enumerate the assignments.

	00	01	10	11
$A\ B$	s_1	s_2	s_3	s_4
$A\ B'$	s_5	s_6	s_7	s_8
$A'\ B$	s_9	s_{10}	s_{11}	s_{12}
$A'\ B'$	s_{13}	s_{14}	s_{15}	s_{16}

As discussed in chapter 2, the assignment s_1 provides a witness for the non-locality of the Hardy model. It is not a member of any compatible family of assignments in the support. However, we do have the following family of $\mathbb{Z}$-linear combinations of assignments:

$$r_1 = s_1, \quad r_2 = s_6 + s_7 - s_8, \quad r_3 = s_{11}, \quad r_4 = s_{15}.$$

One can check that

$$\begin{array}{rclcl} r_2|_A & = & 1 \cdot (A \mapsto 0) + 1 \cdot (A \mapsto 1) - 1 \cdot (A \mapsto 1) & = & r_1|_A, \\ r_2|_{B'} & = & 1 \cdot (B' \mapsto 1) + 1 \cdot (B' \mapsto 0) - 1 \cdot (B' \mapsto 1) & = & r_4|_{B'}. \end{array}$$

Thus the family $\{r_i\}$ meets the conditions of proposition 3.2.3, and the obstruction $\gamma(s_1)$ vanishes.

3.3 Non-locality Results by Example

The PR Box

There is better news when we look at the PR box [93]. The support of this model is described again in the following table.

	00	01	10	11
$A\ B$	1	0	0	1
$A\ B'$	1	0	0	1
$A'\ B$	1	0	0	1
$A'\ B'$	0	1	1	0

This is a strongly contextual model (see proposition 2.6.4), so no assignment in the support is a member of a compatible family. The coefficients for a candidate family $\{r_i\}$ can be labelled as follows.

	00	01	10	11
$A\ B$	a	0	0	b
$A\ B'$	c	0	0	d
$A'\ B$	e	0	0	f
$A'\ B'$	0	g	h	0

The constraints arising from the requirements that $r_i|_{C_{i,j}} = r_j|_{C_{i,j}}$ are:

$$a = c, \quad b = d, \quad a = e, \quad b = f, \quad c = h, \quad d = g, \quad e = g, \quad f = h,$$

implying that all the coefficients must be equal.

Checking that an assignment in the support is a member of such a family requires assigning 1 to the coefficient labelling that assignment and 0 to all the other assignments in that row. Clearly such an assignment is incompatible with the above constraints, since it implies $1 = 0$. Hence there can be no such family, and the obstruction does not vanish for any assignment in the support, witnessing the non-locality of the PR box.

GHZ

We now consider the GHZ-Mermin model [83], which is also strongly contextual. This model, however, is realisable in quantum mechanics, whereas the previous example is not. The support for the relevant part of the model is described in the following table.

	000	001	010	011	100	101	110	111
$A\ B\ C$	1	0	0	1	0	1	1	0
$A\ B'\ C'$	0	1	1	0	1	0	0	1
$A'\ B\ C'$	0	1	1	0	1	0	0	1
$A'\ B'\ C$	0	1	1	0	1	0	0	1

The other contexts $\{ABC', AB'C, A'BC, A'B'C'\}$ have full support. Coefficients for a candidate family are labelled as follows.

	000	001	010	011	100	101	110	111
$A\ B\ C$	a	0	0	b	0	c	d	0
$A\ B'\ C'$	0	e	f	0	g	0	0	h
$A'\ B\ C'$	0	i	j	0	k	0	0	l
$A'\ B'\ C$	0	m	n	0	o	0	0	p

The constraints arising from the requirements that $r_i|_{C_{i,j}} = r_j|_{C_{i,j}}$ are:

$$
\begin{array}{rclrcl}
a+b & = & e+f & c+d & = & g+h \\
a+c & = & i+k & b+d & = & j+l \\
a+d & = & n+o & b+c & = & m+p \\
f+g & = & j+k & e+h & = & i+l \\
e+g & = & m+o & f+h & = & n+p \\
i+j & = & m+n & k+l & = & o+p
\end{array}
$$

Again, to check whether an assignment in the support is a member of such a family requires setting the coefficient for that assignment to 1, and the coefficients for all other assignments in that row to 0.

It suffices to show that these constraints cannot be satisfied over the integers modulo 2. This implies that they cannot be satisfied over $\mathbb{Z}$, since otherwise such a solution would descend via the homomorphism $\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$. Of course, this will also show that the cohomological obstruction does not vanish even if we use $\mathbb{Z}/2\mathbb{Z}$ as the coefficient group.

All cases for GHZ have been machine-checked in mod 2 arithmetic, and it has been confirmed that the cohomological obstruction witnesses the impossibility of extending any assignment in the support to all measurements. Thus cohomology witnesses the non-locality of the GHZ model.

3.4 Contextuality Results by Example

The Peres-Mermin Square

The Peres-Mermin 'magic square' [85, 90] is an important example of a contextual model which can be realised in quantum mechanics using two-qubit observables. The model consists of nine measurements $\{A, \ldots, I\}$. The compatible families of measurements are the rows and columns of the following table.

A	B	C
D	E	F
G	H	I

For 'row contexts' the support of the model contains only those assignments with an odd number of 1's, while for 'column contexts' it contains only the assignments with an even number of 1's. Thus the model has the following support table.

	000	001	010	011	100	101	110	111
$A\ B\ C$	0	1	1	0	1	0	0	1
$D\ E\ F$	0	1	1	0	1	0	0	1
$G\ H\ I$	0	1	1	0	1	0	0	1
$A\ D\ G$	1	0	0	1	0	1	1	0
$B\ E\ H$	1	0	0	1	0	1	1	0
$C\ F\ I$	1	0	0	1	0	1	1	0

We label the coefficients for a candidate compatible family of $\mathbb{Z}$-linear combinations of assignments at each context.

	000	001	010	011	100	101	110	111
$A\ B\ C$	0	c_1	b_1	0	a_1	0	0	t_1
$D\ E\ F$	0	c_2	b_2	0	a_2	0	0	t_2
$G\ H\ I$	0	c_3	b_3	0	a_3	0	0	t_3
$A\ D\ G$	$\bar{t}_4$	0	0	$\bar{a}_4$	0	$\bar{b}_4$	$\bar{c}_4$	0
$B\ E\ H$	$\bar{t}_5$	0	0	$\bar{a}_5$	0	$\bar{b}_5$	$\bar{c}_5$	0
$C\ F\ I$	$\bar{t}_6$	0	0	$\bar{a}_6$	0	$\bar{b}_6$	$\bar{c}_6$	0

Compatibility gives rise to the following constraints.

$$\begin{array}{rclcrcl}
a_1 + t_1 & = & \overline{b}_4 + \overline{c}_4 & & \overline{a}_4 + \overline{t}_4 & = & b_1 + c_1 \\
b_1 + t_1 & = & \overline{b}_5 + \overline{c}_5 & & \overline{a}_5 + \overline{t}_5 & = & a_1 + c_1 \\
c_1 + t_1 & = & \overline{b}_6 + \overline{c}_6 & & \overline{a}_6 + \overline{t}_6 & = & a_1 + b_1 \\
\\
a_2 + t_2 & = & \overline{a}_4 + \overline{c}_4 & & \overline{b}_4 + \overline{t}_4 & = & b_2 + c_2 \\
b_2 + t_2 & = & \overline{a}_5 + \overline{c}_5 & & \overline{b}_5 + \overline{t}_5 & = & a_2 + c_2 \\
c_2 + t_2 & = & \overline{a}_6 + \overline{c}_6 & & \overline{b}_6 + \overline{t}_6 & = & a_2 + b_2 \\
\\
a_3 + t_3 & = & \overline{a}_4 + \overline{b}_4 & & \overline{c}_4 + \overline{t}_4 & = & b_3 + c_3 \\
b_3 + t_3 & = & \overline{a}_5 + \overline{b}_5 & & \overline{c}_5 + \overline{t}_5 & = & a_3 + c_3 \\
c_3 + t_3 & = & \overline{a}_6 + \overline{b}_6 & & \overline{c}_6 + \overline{t}_6 & = & a_3 + b_3
\end{array}$$

To check whether some assignment in the support belongs to such a family, we set its coefficient to 1 and the coefficients of the other assignments in the same context to 0. It has been machine-checked in mod 2 arithmetic that there is no solution to the system for any choice of starting assignment. So cohomology also witnesses the contextuality of the Peres-Mermin model.

The Contextual Triangle

We will introduce a general notion of a Kochen-Specker-type model in section 3.5, the simplest example of which is the contextual triangle. This is the model that arises from Specker's parable [98, 75]. It has also appeared in a somewhat different context in [4] and [81], and is related to the Penrose triangle of figure 1.1.

The model is defined on the following measurement cover:

$$\mathcal{M} \quad = \quad \{\{A, B\}, \quad \{B, C\}, \quad \{A, C\}\}.$$

The cover is not realisable by projective measurements in quantum mechanics since the pairwise compatibility would imply that $\{A, B, C\}$ should also be an allowed measurement context. It is, however, realisable by POVM's [65], though it is not known if the actual model we are interested in on this cover is realisable. It is nevertheless a useful example to set the scene.

We are interested in the *Kochen-Specker support*, which contains those assignments with exactly one 1 among the outcomes. Thus we have the following table:

	00	01	10	11
A B	0	1	1	0
B C	0	1	1	0
C A	0	1	1	0

The coefficients for a candidate family are labelled as follows.

	00	01	10	11
A B	0	a	b	0
B C	0	c	d	0
C A	0	e	f	0

The constraints on the coefficients for a compatible family are:

$$a = f, \quad b = e, \quad a = d, \quad b = c, \quad d = e, \quad c = f,$$

implying that all the coefficients must be equal.

As before, checking that an assignment in the support has a non-vanishing obstruction requires setting the coefficient labelling that assignment to 1, and the other coefficients in its row to 0. Clearly there is no such solution, since it would imply that $1 = 0$.

The 18-Vector Kochen-Specker Configuration

The 18-vector construction in $\mathbb{R}^4$ from [34], gives rise to a model that is state-independent at the level of the support. This is the model with Kochen-Specker support on the measurement cover given by the columns of the following table.

A	A	H	H	B	I	P	P	Q
B	E	I	K	E	K	Q	R	R
C	F	C	G	M	N	D	F	M
D	G	J	L	N	O	J	L	O

We label the coefficients for a candidate family as below.

	1000	0100	0010	0001
$A\ B\ C\ D$	a	b	c	d
$A\ E\ F\ G$	a	e	f	g
$H\ I\ C\ J$	h	i	c	j
$H\ K\ G\ L$	h	k	g	l
$B\ E\ M\ N$	b	e	m	n
$I\ K\ N\ O$	i	k	n	o
$P\ Q\ D\ J$	p	q	d	j
$P\ R\ F\ L$	p	r	f	l
$Q\ R\ M\ O$	q	r	m	o

Note that some of the constraints on the coefficients take the form of simple equations between coefficients (see proof of proposition 3.5.5) allowing us to reduce from 36 to 18 coefficients; we have used this reduction in the table.

The remaining constraints are expressed by the following equations.

$$
\begin{array}{rclrcl}
b+c+d & = & e+f+g & a+b+d & = & h+i+j \\
a+c+d & = & e+m+n & a+b+c & = & p+q+j \\
a+f+g & = & b+m+n & a+e+f & = & h+k+l \\
a+e+g & = & p+r+l & i+c+j & = & k+g+l \\
h+c+j & = & k+n+o & h+i+c & = & p+q+d \\
h+g+l & = & i+n+o & h+k+g & = & p+r+f \\
b+e+n & = & q+r+o & b+e+m & = & i+k+o \\
i+k+n & = & q+r+m & q+d+j & = & r+f+l \\
p+d+j & = & r+m+o & p+f+l & = & q+m+o
\end{array}
$$

It has been machine-checked in mod 2 arithmetic that no cohomological obstruction vanishes, confirming that we have a cohomological witness for the Kochen-Specker theorem.

3.5 General Results I

The previous examples, while providing cohomological non-locality and contextuality proofs, needed to be analysed on a case by case basis. Therefore, one might view these as providing a proof of concept. In this section, however, we are concerned with finding general results which prove the effectiveness of the cohomological approach for whole classes of models. In particular, we will be interested in models that can be used for contextuality arguments similar to that of Kochen & Specker [73].

We begin by introducing a general notion of a Kochen-Specker-type model for any measurement cover. These models will be encountered again in chapter 4. We assume an outcome set $\{0, 1\}$. For any maximal context $C \in \mathcal{M}$ and measurement $m \in C$, we define $s_{C,m} \in \mathcal{E}(C)$ to be the section that assigns 1 to m and 0 to all other measurements in C. Possible outcomes for each context are precisely those that assign 1 to a single measurement. This kind of condition arises when one considers projective measurements in quantum mechanics.

Definition 3.5.1. *Let $O = \{0, 1\}$. The* Kochen-Specker support *for the cover $\mathcal{M}$ is the presheaf given by*

$$S_{\mathsf{KS}}(C) = \{s_{C,m} \mid m \in C\}.$$

The Kochen-Specker model *on the measurement scenario $(X, O, \mathcal{M})$ is the possibilistic model $\{e_C\}_{C\in\mathcal{M}}$ defined by*

$$\mathsf{supp}\,(e_C) = S_e(C) = S_{\mathsf{KS}}(C),$$

for all $C \in \mathcal{M}$.

A necessary condition for Kochen-Specker models to have a consistent global assignment (i.e. $S_e(X) \neq \emptyset$) is given in [4]. The negation of this condition, therefore, provides a sufficient condition for a model to be strongly contextual.

Proposition 3.5.2 (Abramsky & Brandenburger [4])**.** *The existence of a consistent global assignment implies that*

$$\mathsf{gcd}\{d_m \mid m \in X\} \quad | \quad |\mathcal{M}|, \tag{3.1}$$

where gcd is the greatest common divisor and $d_m := |\{C \in \mathcal{M} \mid m \in C\}|$.

We refer to (3.1) as the GCD condition, and to each d_m as the *degree* of the measurement m. All models that do not satisfy the GCD condition are therefore strongly contextual.

Definition 3.5.3. $\neg\mathsf{GCD}$ *is the class of Kochen-Specker models that do not satisfy the GCD condition.*

For the results that follow, it is necessary to assume connectedness of measurement scenarios in the following sense. Given any measurement scenario $(X, O, \mathcal{M})$ one can define a hypergraph $(X, \mathcal{M})$ with the measurements X as vertices and with hyperedges given by the maximal contexts $\mathcal{M}$. This differs from the abstract simplicial complex defined by the measurement scenario in that it only takes into account the maximal contexts. As we will see later, it can also be useful to consider the dual hypergraph $(\mathcal{M}, X)$, which has a vertex for each maximal context and a hyperedge $e_m = \{C \in \mathcal{M} \mid x \in C\}$ for each measurement $m \in X$.

Definition 3.5.4. *A measurement scenario* $(X, O, \mathcal{M})$ *is said to be* connected *if its hypergraph* $(X, \mathcal{M})$ *is connected.*

That is to say, a measurement scenario is connected if, for any maximal contexts $C, C' \in \mathcal{M}$, one can find a finite sequence of maximal contexts

$$C = C_0, C_1, C_2, \ldots, C_n, C_{n+1} = C'$$

such that

$$\forall\, i \in \{0, \ldots, n\}.\ \ C_i \cap C_{i+1} \neq \emptyset.$$

We will now show that cohomology captures strong contextuality for a class of connected Kochen-Specker models using an argument related to proposition 3.5.2. Of course, we note that cohomology witnesses strong contextuality in some connected models outside of this class (e.g. the PR box, the GHZ and Peres-Mermin models) so it captures the property more finely than this.

Proposition 3.5.5. *If the cohomological obstruction vanishes for some assignment in a Kochen-Specker model, then the* GCD *condition holds for that model.*

Proof. Assume that $\gamma(s_1) = 0$ for some assignment $s_1 \in S_e(C_1)$ in the support. If we enumerate the maximal contexts $\mathcal{M} = \{C_i\}_{i \in I}$ then this implies by proposition 3.2.3 that there exists a compatible family $\{r_i \in \mathcal{F}(C_i)\}_{i \in I}$ of $\mathbb{Z}$-linear combinations of assignments of $S_e$such that $r_1 = s_1$. Recall that the support of each maximal context is $S_e(C) = \{s_{C,m} \mid m \in C\}$. Let $c_{i,m}$ denote the coefficient of the assignment $s_{C_i,m}$ in the linear combination r_i.

If for some i, j there exists $m \in C_i \cap C_j$, then compatibility gives the following constraints:

$$c_{i,m} = c_{j,m}, \qquad\qquad \sum_{\substack{m' \in C_i \\ m' \neq m}} c_{i,m'} = \sum_{\substack{m'' \in C_j \\ m'' \neq m}} c_{j,m''}.$$

Using equations of the first kind we can identify the coefficients $c_{i,m}$ for all i, and unambiguously denote these coefficients by c_m alone, regardless of the context. Summing the two equations above gives

$$\sum_{m \in C_i} c_m = \sum_{m' \in C_j} c_{m'};$$

i.e. the sums of the coefficients of r_i and r_j are the same. By connectedness, and since the sum is equal to 1 for the context C_1, the coefficients of r_k sum to 1 for each maximal context C_k.

Hence, we have

$$|\mathcal{M}| = \sum_{C \in \mathcal{M}} 1 = \sum_{C \in \mathcal{M}} \sum_{m \in C} c_m = \sum_{m \in X} d_m c_m = g \sum_{m \in X} \frac{d_m}{g} c_m$$

where $d_m := |\{C \in \mathcal{M} \mid m \in C\}|$ as before and $g := \mathsf{gcd}\{d_m \mid m \in X\}$. Since g divides d_m for all m, we can conclude that g divides $|\mathcal{M}|$. $\square$

Then for any model in the class $\neg\mathsf{GCD}$, no cohomological obstruction vanishes, and we have the following corollary.

Corollary 3.5.6. *Cohomology witnesses contextuality for all* $\neg\mathsf{GCD}$ *Kochen-Specker models.*

We have already considered two familiar models from this class: the contextual triangle and the state-independent model arising from the 18-vector proof of the Kochen-Specker theorem, from which we provided a cohomological proof of the theorem.

3.6 General Results II

The fact that cohomology can be shown to witness contextuality for this class of strongly contextual models, as well as the success of the cohomological approach in all of the strongly contextual and strongly non-local examples that have been considered (recall that the only example where a false positive has been seen to arise was the Hardy model, which is not strongly non-local) might lead us to suspect that the cohomological characterisation is complete for strong contextuality. However, it has been possible to construct a strongly contextual model for which a false positive does arise. This is the Kochen-Specker model for the cover

$$\mathcal{M} \quad = \quad \{\{A, B, C\}, \{B, D, E\}, \{C, D, E\}, \{A, D, F\}, \{A, E, G\}\}.$$

In contrast with all of the earlier examples, this model does not satisfy any reasonable criterion for symmetry, nor does it satisfy any strong form of connectedness. In fact, the existence of measurements with degree 1 is crucial in this example (F and G each belong to a single maximal context). It means that it is always possible to choose coefficients for $s_{\{A,D,F\},F}$ and $s_{\{A,E,G\},G}$ that will make the coefficients of the respective contexts sum to 1 without imposing any constraints on the other contexts. This leads to the following conjecture.

Conjecture 3.6.1. *Under suitable assumptions of symmetry and connectedness, the cohomological obstruction is a complete invariant for contextuality.*

We will now present some results which support this conjecture, including a proofs of the conjecture for other classes of Kochen-Specker models. Note that this a strengthening of a conjecture made by the author and his collaborators in [12], which only proposed that the cohomological obstruction might be a complete invariant for strong contextuality. This is because, for models with appropriate properties, we can prove (proposition 3.6.4) that contextuality and strong contextuality are equivalent.

The first step is to introduce the appropriate notion of symmetry. Again, we will define this in relation to the hypergraph derived from a measurement scenario. A hypergraph $(X, \mathcal{M})$ is *vertex-symmetric* if its automorphism group is transitive. That is to say that for every pair of vertices $m, m' \in X$ there exists a hypergraph automorphism $\alpha : X \to X$ such that $\alpha(m) = m'$. Every vertex-symmetric hypergraph is necessarily *k-regular*, which is to say that there exists some $k \in \mathbb{N}$ such that the degree $d_m = k$ for all $m \in X$.

Definition 3.6.2. *A measurement scenario* $(X, O, \mathcal{M})$ *is said to be* symmetric *if the hypergraph* $(X, \mathcal{M})$ *is vertex-symmetric.*

It can be useful to phrase the problem of contextuality of Kochen-Specker models in terms of hypergraphs. This naturally leads to some interesting connections with ideas from (hyper)graph theory. A *transversal* of a hypergraph $(X, \mathcal{M})$ is a subset $Y \subseteq X$ such that $Y \cap C \neq \emptyset$ for all $C \in \mathcal{M}$. A *stable transversal* of $(X, \mathcal{M})$ is a transversal Y such that no two elements $m, m' \in Y$ are adjacent. In these terms, it is possible to characterise contextuality in Kochen-Specker models as follows (this was first pointed out in [4]).

Proposition 3.6.3. *The Kochen-Specker model on* $(X, \mathcal{M})$ *is non-contextual if and only if each* $m \in X$ *belongs to a stable transversal of the hypergraph* $(X, \mathcal{M})$.

Proof. Recall from proposition 1.6.1 that a model is non-contextual if and only if every 'local' assignment $s \in \mathcal{E}(C)$ belongs to a compatible family of assignments

$$\{s_i \in \mathcal{E}(C_i)\}_{C_i \in \mathcal{M}}$$

in the model. Since every possible assignment in a Kochen-Specker model is of the form $s = s_{C,m}$, it follows that in such a model non-contextuality is equivalent to every section belonging to a compatible family $\{s_{C_i,m_i}\}_{C_i \in \mathcal{M}}$ for some $\{m_i \in C_i\}_{C_i \in \mathcal{M}}$. A family of assignments, therefore, is defined by a family of measurements $\{m_i\}_{C_i \in \mathcal{M}}$, and compatibility of the assignments translates to the property that if $m_i \in C_i \cap C_j$ then $m_i = m_j$, or, equivalently, if $m_i \neq m_j$ then $m_i \notin C_j$ and $m_j \notin C_i$. This is precisely a stable transversal of $(X, \mathcal{M})$. □

Proposition 3.6.4. *A symmetric Kochen-Specker model is contextual if and only if it is strongly contextual.*

Proof. Consider the Kochen-Specker model on $(X, \mathcal{M})$. Suppose there exists some $m \in X$ such that m belongs to a stable transversal Y of the hypergraph $(X, \mathcal{M})$. Since the model is symmetric, for any $m' \in X$ there exists a hypergraph automorphism $\alpha : X \to X$ such that $\alpha(m) = m'$. Since α is a hypergraph automorphism, it must be that $\alpha(Y)$ is also a stable transversal of $(X, \mathcal{M})$. It is therefore the case that every $m \in X$ belongs to a stable transversal. □

The fact that symmetry implies an equivalence between contextuality and strong contextuality is an interesting point: all of the familiar strongly contextual models are symmetric. On the other hand, the Hardy model, which is contextual but not strongly contextual, is inherently asymmetric. This was a crucial consideration in the proof of theorem 2.6.5, in which it was shown that in order to realise a Hardy model, the underlying entangled state must have some asymmetry.

Some interesting connections arise when we consider the problem of contextuality in Kochen-Specker models in terms of dual hypergraphs $(\mathcal{M}, X)$. A *matching* M of the hypergraph $(\mathcal{M}, X)$ is a set of pairwise non-adjacent edges $M \subseteq X$. A *perfect matching* is a matching that matches all vertices. This is the dual notion to a stable transversal. Now we arrive at the following characterisation of contextuality for Kochen-Specker models.

Corollary 3.6.5. *The Kochen-Specker model on $(X, \mathcal{M})$ is non-contextual if and only if every edge belongs to a perfect matching of the dual hypergraph $(\mathcal{M}, X)$.*

Proof. This is the dual statement of proposition 3.6.3. □

In this dual picture, *edge-transitivity* is the relevant notion of symmetry: for any two edges $e_m, e_{m'} \in X$ there exists a hypergraph automorphism $\alpha : X \to X$ such that $\alpha(e_m) = e_{m'}$. An edge-symmetric hypergraph is necessarily *k-uniform*, which is to say that $|e_m| = k$ for all $e_m \in X$. Of course, in our case, $|e_m|$ is just the degree d_m of the measurement m.

Such problems have been studied in the mathematics literature, and there are several results that can find interesting applications to the problem of contextuality in Kochen-Specker models. The first results we mention relate to the decidability problem for strong contextuality in Kochen-Specker models with constant degree $d = k$, which is equivalent to checking for the existence of a perfect matching in a k-uniform hypergraph.

For $d = 2$, the 'blossom algorithm' [47] provides an efficient method[1] of finding a maximum matching (i.e. a matching M such that $|M|$ is maximised) and hence for deciding strong contextuality.

Corollary 3.6.6. *Decidability of strong contextuality for Kochen-Specker models with constant degree $d = 2$ is polynomial with respect to the number of maximal contexts.*

For $d \geq 3$, however, this is known to be an NP-complete problem [52].

Corollary 3.6.7. *Decidability of strong contextuality for Kochen-Specker models with constant degree $d \geq 3$ is NP-complete with respect to the number of maximal contexts.*

These results, which as far as the author is aware are new to the foundations of quantum mechanics literature, complement the work of Pitowsky [92] and Abramsky, Gottlob & Kolaitis [10] which prove similar NP-completeness results for probabilistic and, respectively, possibilistic Bell scenarios, as well as our proposition 2.6.1, which gave efficient algorithms for deciding logical non-locality in certain Bell scenarios.

In the $d = 2$ case, Tutte's theorem [76] provides a necessary and sufficient condition for a 2-uniform hypergraph (i.e. a graph or multigraph) to have a perfect matching.

Corollary 3.6.8. *A Kochen-Specker model with constant degree $d = 2$ has a global assignment if and only if for each $S \subseteq \mathcal{M}$ the subgraph of $(\mathcal{M}, X)$ induced by $(\mathcal{M} - S)$ has at most $|S|$ connected components with an odd number of vertices.*

[1] $O(|\mathcal{M}|^4)$.

This has the same flavour as Vorob'ev's theorem [103], which was originally proved in relation to game theory, but which in our setting characterises measurement covers for which empirical models are necessarily local.

In the general case, despite the fact that the decidability problem for perfect matching of a k-uniform hypergraph is NP-complete, there do exist a number of sufficient conditions. One such condition, due to Daykin & Häggkvist [44], is sufficient for any k-uniform hypergraph satisfying the GCD condition to have a perfect matching. The condition is

$$m \geq \left(1 - \frac{1}{d}\right) \binom{|\mathcal{M}| - 1}{d - 1}, \tag{3.2}$$

where d is the degree and $m := \min_{C \in \mathcal{M}} |C|$. We can use this to prove the following result.

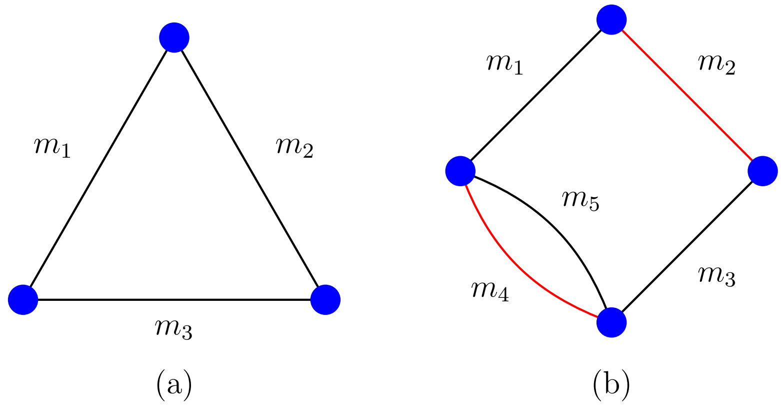
Proposition 3.6.9. *Cohomology provides a complete characterisation of contextuality for the class of symmetric Kochen-Specker models which satisfy the Daykin-Häggkvist condition (3.2).*

Proof. A symmetric model has constant degree $d = k$ for some k, and its dual hypergraph is k-uniform. If the GCD condition and condition (3.2) are both satisfied satisfied, the Daykin-Häggkvist theorem guarantees the existence of a perfect matching, and hence that that the model is not strongly contextual. In fact, we know by proposition 3.6.4 that strong contextuality and contextuality are equivalent for symmetric models, so assuming the GCD condition to hold, Daykin-Häggkvist theorem actually guarantees non-contextuality. Together with proposition 3.5.2, this proves that the GCD condition is necessary and sufficient for non-contextuality in this class of models. Therefore, by proposition 3.5.5, cohomology witnesses contextuality for all models in the class. □

This proves a restricted version of conjecture 3.6.1. If it could be shown that the GCD condition is necessary and sufficient for an edge-symmetric hypergraph to have a perfect matching, then by a similar argument we could prove the conjecture for all Kochen-Specker models. However, it is possible to find a counter-example to this, and there is some reason to suspect that classifying the edge-symmetric hypergraphs that have a perfect matching is not an easy problem [35].

We also mention another class of symmetric models for which we can prove that the conjecture holds by virtue of the fact that the GCD condition is necessary and sufficient for non-contextuality. These are the Kochen-Specker models whose dual hypergraphs are simply graphs consisting of a closed chain of vertices, or multigraphs

Figure 3.1: Dual (multi)graphs of chain Kochen-Specker models: (a) the contextual triangle; (b) an example with multiple edges which has a perfect matching (red).



whose underlying graph is of this form. A couple of examples are given in figure 3.1. We will refer to such models as *chain* Kochen-Specker models. This class contains the model for the Klyachko proof of the Kochen-Specker theorem [72] and its generalisations [75].

Proposition 3.6.10. *Cohomology provides a complete characterisation of contextuality for chain Kochen-Specker models.*

Proof. For the dual hypergraph of such a model, we claim that there exists a perfect matching if and only if there exists a Hamiltonian cycle (a closed cycle which passes through each vertex exactly once) of even length. Since we have assumed that the dual hypergraph is a multigraph, each measurement must have degree $d = 2$. A matching pairs vertices, so the existence of a perfect matching implies that there must be an even number of vertices. Since the underlying graph consists simply of a chain of vertices, there necessarily exists a Hamiltonian cycle, which, since there are an even number of vertices, must be of even length. Conversely, if there exists a Hamiltonian cycle of even length, then by selecting alternating edges in the cycle one obtains a perfect matching.

This shows that a model in this class is contextual if and only if it is strongly contextual, since if some edge belongs to a perfect matching there exists a Hamiltonian cycle of even length, but then all edges can be shown to belong to a perfect matching. Furthermore, it is clear that the GCD condition is necessary and sufficient for non-contextuality since it holds precisely when $|\mathcal{M}|$ is even. By proposition 3.5.5, cohomology completely characterises contextuality for this class of models. $\square$

3.7 Discussion

We have succeeded, in sections 3.3 and 3.4, in finding cohomological non-locality and contextuality proofs that are counterparts to many of the well-known theorems (GHZ, Kochen-Specker, etc.). There are some immediate limitations to the results described in these sections, however. One point is that the obstructions are simply computed by brute force enumeration, so the results we have obtained can only be considered a proof of concept, and are not as conceptually illuminating as one might hope. Ideally, we would like to use the machinery of homological algebra and exact sequences to obtain more conceptual and general results. A second point is that, in general, the cohomological condition for contextuality is sufficient, but not necessary. This is a consequence of the fact that the presheaf we use $\mathcal{F} = F_{\mathbb{Z}}S_e$ is only an approximation to the presheaf S_e that we are really interested in. Overcoming these limitations is an objective for future work.

The results of sections 3.5 and 3.6 represent some progress on the issue of generality. We have seen that for large classes of Kochen-Specker models cohomology provides a complete characterisation of contextuality. Moreover, the investigations in these sections have led to several insights which are illuminating in their own right. In particular, we have found a connection between Kochen-Specker-type contextuality proofs and the problem of the existence of perfect matchings in hypergraphs. This has been quite fruitful in that known results about hypergraphs have allowed us to show that decidability of strong contextuality for Kochen-Specker models with constant degree $d = 2$ is a polynomial problem, while for $d \geq 3$ it is NP-complete, and Tutte's theorem was seen to provide a necessary and sufficient condition for non-contextuality in $d = 2$ Kochen-Specker models.

The conjecture made in [12] that under suitable assumptions for symmetry and connectedness it might be shown that cohomology provides a complete characterisation for strong contextuality has also been strengthened in light of proposition 3.6.4, which shows that for symmetric Kochen-Specker models contextuality and strong contextuality are equivalent. Propositions 3.6.9 and 3.6.10 prove the conjecture in restricted cases.

Another idea is that 'good' cases (in which cohomology succeeds in witnessing contextuality) may somehow be related to the notion of Vorob'ev regularity of measurement covers. In [103], Vorob'ev characterised the covers, or more precisely the simplicial complexes these generate on which any model is extendable; i.e. non-contextual.

These are exactly the complexes which can be reduced to an empty complex by removing certain extremal maximal contexts. From the proof of the theorem, one can see that the non-extendability of a model would be already noticed in its reduced form, which allows us to focus on witnessing non-contextuality for irreducible (Vorob'ev regular) covers. A necessary condition for a context to be extremal is that it contains measurements that do not belong to any other maximal context. Even though the strongly contextual model that gave a false positive in section 3.6 has no extremal contexts, and is therefore irreducible, it does have this weaker property.

Chapter 4

Bell Models from Kochen-Specker Models

Again we recall that in chapter 1 extendability of empirical models was seen to correspond in a unified manner to both locality and non-contextuality, an insight that has initiated diverse lines of research (e.g. [2, 4, 9, 11, 12, 100]). In this chapter, which is based on [78], we introduce a refinement of the notion of extendability that captures the idea of partial approximations to locality/non-contextuality. This can be useful in characterising the properties of sub-models.

The refinement has also found more practical applications. Certain empirical models, such as those considered in chapter 2, have measurements that can be partitioned into sites, and can be considered to abstract spatially distributed systems: these are the Bell-type models. We are especially interested in a particular, canonical extension, which, when well-defined, may be used for the construction of equivalent Bell models from models of the more general kind. On both foundational and practical levels, an advantage of having an equivalent Bell form of a contextual model is that it is much easier to motivate a notion of locality in a Bell scenario than the corresponding notion of non-contextuality in a more general measurement scenario, making non-local behaviour all the more striking.

In chapters 2 and 5, for example, we present locality as the conjunction of determinism and parameter independence (this is closely related to no-signalling, which we discussed in some detail in chapter 1). Of course, in a spatially separated system, one might appeal to compatibility with relativity to motivate parameter independence as a reasonable physical assumption: the choice of measurement in one system should not instantaneously affect the outcomes of measurements on other systems. On the other hand, the justification for this kind of argument is less clear for sets of compatible measurements that are made on a single system: one could simply 'coarse-grain' the measurement set so that the sets of compatible measurements are treated as the basic measurements, and the problem is somewhat mitigated. This was realised by Bell, who had observed a similar result [23][1] to that of Kochen & Specker [73] before going on to prove his more well-known non-locality theorem [22] (this is also discussed in [85]). A further advantage is that non-locality can be exploited as an information theoretic resource [20], whereas contextuality has yet to be developed for such purposes.

We can find equivalent Bell models for many familiar examples of contextual models: the entire family of symmetric Kochen-Specker models from chapter 3, which includes the contextual triangle and the 18-vector Kochen-Specker model, for example. One connection that arises is that the equivalent Bell model for the contextual triangle is essentially a folding of several Popescu-Rohrlich boxes [93]. The Peres-Mermin square [85] is also treated. This represents a step in the direction of proposing equivalent Bell tests for contexuality results, though an important issue that remains to be addressed is that of quantum realisability.

4.1 Bell Scenarios

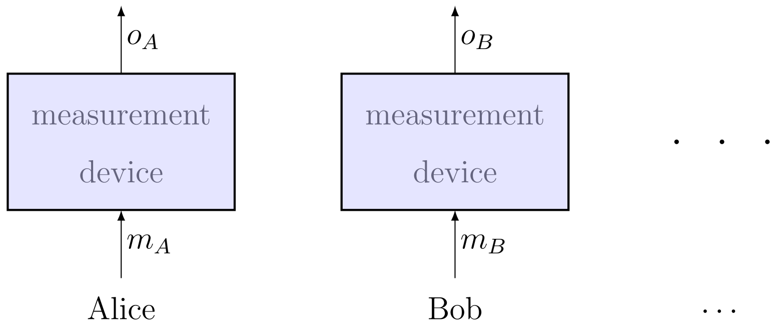
Bell scenarios are measurement scenarios that can be thought of as abstracting of spatially distributed systems (see figure 4.1). We encountered this kind of measurement configuration in chapter 2, in which we introduced a logical framework for (n, k, l) Bell scenarios. These were to be thought of as n-partite models, in which each party could choose to perform one of k different measurements, each of which could have l possible outcomes. For example, the model arising from the Bell-CHSH theorem from chapter 1 or the original Hardy model [60] from chapter 2 are both $(2, 2, 2)$ models. An example of a $(3, 2, 2)$ model is that which arises from the GHZ-Mermin non-locality argument [83]. Recall that in such scenarios extendability corresponds to the usual notion of Bell locality.

More carefully speaking, these are measurement scenarios $(X, O, \mathcal{M})$ for which the set of measurements can be written as a disjoint union $X = \coprod_{i=1}^{n} X_i$ such that the maximal contexts are given by the cartesian product $\mathcal{M} = \prod_{i=1}^{n} X_i$. We define $l := |O|$ and $k := \max_{1 \leq i \leq n} |X_i|$.

As a technical remark, there is a slight abuse of notation here. Elements of $X = \coprod_{i=1}^{n} X_i$ are of the form $\langle x, i \rangle$ where $i \in \{1, \dots, n\}$ identifies the site and $x \in X_i$. An element of the cover $\prod_{i=1}^{n} X_i$ is an n-tuple $\langle x_1, \dots, x_n \rangle$ with each $x_i \in X_i$, which can be seen as a subset of $\coprod_{i=1}^{n} X_i$ if we interpret it as $\{\langle x_1, 1 \rangle, \dots, \langle x_n, n \rangle\}$. We will

[1] This paper was written earlier but finally published later than [22].

Figure 4.1: A Bell scenario.



denote it as a tuple, however, to simplify notation. Therefore, the maximal contexts of a Bell scenario are the sets of measurements that contain one measurement from each site.

4.2 Kochen-Specker Models

Of course, not all measurement scenarios are of the Bell type. We have already encountered some examples: the state-independent model for the 18-vector proof of the Kochen-Specker theorem [34] and the Peres-Mermin square [85]. Both of these models make use of measurement scenarios that are of a more general form and cannot be partitioned into sites. Many 'non-Bell' models fall into the general class of *Kochen-Specker models* from definition 3.5.1.

Recall that these are the possibilistic models on any connected measurement scenario for which, at each maximal context, an assignment is possible if and only if it maps a single measurement to the outcome 1. As we saw in chapter 3, the cohomological characterisation of strong contextuality is complete for certain classes of Kochen-Specker models.

	00	01	10	11
A B	0	1	1	0
B C	0	1	1	0
C A	0	1	1	0

The contextual triangle, above, which we encountered in chapter 3, is the simplest example. We will return to this model in section 4.6.

Note that in this chapter we use an equivalent definition of the Kochen-Specker support, which is more convenient for our present purposes:

$$S_{\mathsf{KS}}(C) := \{s \in \mathcal{E}(C) \mid o(s) = 1\},$$

where $o(s) := |\{x \in C \mid s(x) = 1\}|$ for any assignment $s \in \mathcal{E}(C)$.

4.3 No-signalling Extensions of Models

We consider the problem of extending an empirical model to a cover that allows increased compatibility of measurements. For notational convenience, in this section, we fix sets X of measurements and O of outcomes, so that a measurement scenario can be identified by its cover $\mathcal{M}$ of maximal contexts alone. Also, when we refer to models in this section it will be assumed that we refer to possibilistic models, as introduced in chapter 1.

Definition 4.3.1. *Let $\mathcal{M}$ and $\mathcal{M}'$ be two measurement covers on X. We write $\mathcal{M} \preceq \mathcal{M}'$ when $\downarrow\mathcal{M} \subseteq \downarrow\mathcal{M}'$; i.e.*

$$\forall\, D \in \mathcal{M}. \quad \exists\, C \in \mathcal{M}'. \quad D \subseteq C.$$

Definition 4.3.2. *Let $\mathcal{M} \preceq \mathcal{M}'$, and let e be a model defined on $\mathcal{M}$. A model f on $\mathcal{M}'$ is said to* extend *e if*

$$\forall\, D \in \mathcal{M}. \qquad f_D = e_D.$$

When such an f exists, we say that e is extendable *to $\mathcal{M}'$.*

Note that the cover $\mathcal{M}_\top := \{X\}$, in which any subset of measurements is jointly compatible, is larger than all other covers; i.e. it is the top element in the poset of measurement covers. Asking for extendability to the top cover amounts to asking for extendability in the usual sense: in other words, locality or non-contextuality. The notion of extendability to any cover $\mathcal{M}' \prec \mathcal{M}_\top$ therefore captures partial approximations to the usual notion. One cover that will be of particular interest in section 4.4 is the following.

Definition 4.3.3. *For any cover $\mathcal{M}$ we define $n(\mathcal{M}) := \max_{C \in \mathcal{M}} |C|$ to be the maximum size of contexts in $\mathcal{M}$ (where no confusion arises, this will be simply denoted n). Then we can define another cover*

$$\mathcal{P}_n X := \{Y \subseteq X \mid |Y| = n\}$$

over X. It necessarily holds that $\mathcal{M} \preceq \mathcal{P}_n X$.

We now consider a construction that provides a candidate for a canonical extension of a model to any larger cover (much like $S_e(X)$ for the usual notion of extendability). We note, however, that this will not necessarily yield a well-defined model. The idea is to allow every assignment except those that are directly forbidden by the compatibility/no-signalling condition; i.e. to allow every assignment in e' that is *consistent* with the possible assignments in e.

Definition 4.3.4. *Let $\mathcal{M} \preceq \mathcal{M}'$ and e be a model on $\mathcal{M}$. For each $C \in \mathcal{M}'$ and $s \in \mathcal{E}(C)$, we define:*

$$e'_C(s) := \bigwedge_{W \subseteq C, W \in \downarrow\mathcal{M}} e_W(s|_W).$$

If $\{e'_C\}_{C \in \mathcal{M}'}$ is a well-defined model extending e, we call it the canonical extension *of e to $\mathcal{M}'$, and say that e is* canonically extendable to $\mathcal{M}'$.

According to the definition, the support of the extended model e' at each maximal context $C \in \mathcal{M}'$ is

$$\mathsf{supp}(e'_C) = \{s \in \mathcal{E}(C) \mid \forall\, W \in\downarrow\mathcal{M}, W \subseteq C.\ s|_W \in S_e(W)\} = S_e(C);$$

i.e. the support contains those assignments on C that are consistent with the model e. We can equivalently express this in a way that mentions only maximal contexts:

$$\mathsf{supp}(e'_C) = \{s \in \mathcal{E}(C) \mid \forall\, D \in \mathcal{M}.\ \exists\, t \in \mathcal{E}(D).\ t \in S_e(D) \wedge t|_{C \cap D} = s|_{C \cap D}\}.$$

Clearly, for $\mathcal{M}' = \mathcal{M}_\top$,

$$\mathsf{supp}(e'_X) = S_e(X);$$

i.e. the assignments consistent with e' are precisely the global assignments consistent with e.

We saw in chapter 1 that $S_e(X)$ provides a canonical local hidden variable space for the model e. Although the present construction does not satisfy properties that are quite as strong, the next two propositions show why this construction, when it yields a well-defined extension, can be regarded as canonical in some sense, especially with regard to strong contextuality.

Proposition 4.3.5. *Let $\mathcal{M} \preceq \mathcal{M}'$, e be a model on $\mathcal{M}$, and f be a model on $\mathcal{M}'$ that extends e. Then, for all $C \in \mathcal{M}'$,*

$$\mathsf{supp}(f_C) \subseteq \mathsf{supp}(e'_C)$$

(i.e. $f_C(s) = 1$ implies $e'_C(s) = 1$ for any $s \in \mathcal{E}(C)$).

Proof. Consider any maximal context $C \in \mathcal{M}'$ and any assignment $s \in \mathsf{supp}(f_C)$. Then, by virtue of f being an extension of e, all restrictions of s to contexts in $\mathcal{M}$ must be consistent with e. That is, for all $W \subseteq C$ with $W \in\downarrow \mathcal{M}$, we have $s|_W \in S_e(W)$. Then, by the definition of e', it must be that $s \in \mathsf{supp}(e'_C)$. $\square$

This tells us that any extension has less possible assignments than the canonical one. This is not surprising, since the canonical construction picks out all the assignments that are consistent with the model e. It is clear that, in the extreme, e'_C might fail to be a distribution for some $C \in \mathcal{M}'$; i.e.

$$\mathsf{supp}(e'_C) = S_e(C) = \emptyset.$$

Obviously, in that case there can be no extension of e to $\mathcal{M}'$ whatsoever. We say that e is *strongly non-extendable to* $\mathcal{M}'$, in analogy with the notion of strong contextuality.

The following proposition will be relevant for the construction of Bell models in section 4.4. It can be read as saying that e' is the most conservative extension that can be made in terms of not introducing any extra (global) strong contextuality.

Proposition 4.3.6. *Let $\mathcal{M} \preceq \mathcal{M}'$, let e be a model on $\mathcal{M}$, and suppose that e is extendable to $\mathcal{M}'$. Then e' is strongly contextual if and only if e is strongly contextual.*

Proof. It is enough to show that $S_e(X) = S_{e'}(X)$; i.e. that the sets of global assignments consistent with each model coincide. For a global assignment $s \in \mathcal{E}(X)$,

$$
\begin{aligned}
& s \in S_e(X) \\
\Leftrightarrow & \\
& \forall\, U \in\downarrow\mathcal{M}.\ s|_U \in S_e(U) \\
\Leftrightarrow & \quad \{\ \text{“}\Leftarrow\text{”: all } U\text{s above satisfy } U \in\downarrow\mathcal{M} \subseteq\downarrow\mathcal{M}'\text{, hence are covered by some } C \in \mathcal{M}'\ \} \\
& \forall\, C \in \mathcal{M}'.\ \forall\, W \subseteq C, W \in\downarrow\mathcal{M}.\ s|_W \in S_e(W) \\
\Leftrightarrow & \\
& \forall\, C \in \mathcal{M}'.\ s_C \in S_{e'}(C) \\
\Leftrightarrow & \\
& s \in S_{e'}(X).
\end{aligned}
$$

$\square$

The situation here is more complicated than in the usual case of extensions to the top cover. The key issue is whether compatibility (no-signalling) holds for the extended model. This is by no means guaranteed. It might happen that the canonical construction e' has too many possible assignments, causing it to be signalling. The reason is that e' picks out all the assignments that are 'locally' consistent with e, but when overlaps of contexts arise that were not contained in the original cover it is possible that assignments are not compatible. We give an example to illustrate how such behaviour might arise.

Example 4.3.7. *The model e on the cover $\mathcal{M} = \{AB, BC, CD, DA\}$ is given by the following table.*

	00	01	10	11
$A\ B$	1	1	1	1
$B\ C$	1	0	0	1
$C\ D$	1	0	0	1
$D\ A$	1	1	1	1

We consider the canonical extension e' to the cover $\mathcal{M}' = \{ABC, BCD\} \succeq \mathcal{M}$ (see figure 4.2).

	000	001	010	011	100	101	110	111
$A\ B\ D$	1	1	1	1	1	1	1	1
$B\ C\ D$	1	0	0	0	0	0	0	1

The extension e' is clearly not compatible. For example,

$$1 = e'_{ABD}|_{BD}(01) \neq e'_{BCD}|_{BD}(01) = 0.$$

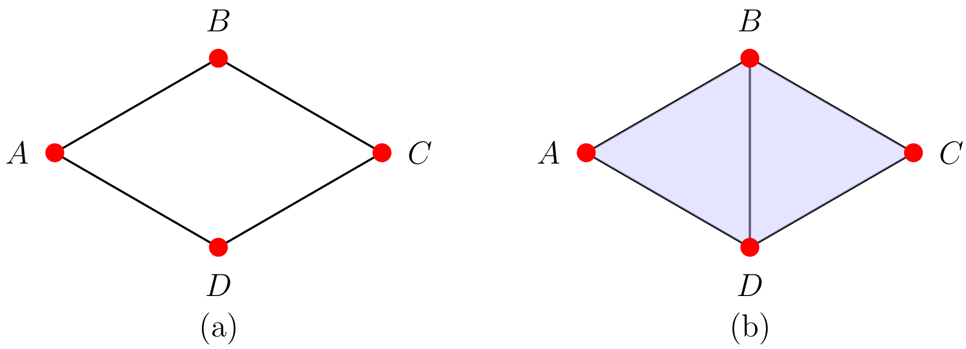
Sub-models

No-signalling extensions can also be related to the contextuality of sub-models of an empirical model.

Definition 4.3.8. *Let e be a model on $\mathcal{M}$. For any $U \subseteq X$, the* induced sub-model of e on U *is $\{e_{U\cap C}\}_{C\in\mathcal{M}}$.*

By compatibility of the original model it is clear that any induced sub-model will be a well-defined empirical model. We note that this definition holds for any empirical model, not just the possibilistic ones we are concerned with in this section.

Figure 4.2: The measurement covers of example 4.3.7: (a) $\mathcal{M} = \{AB, BC, CD, DA\}$ (b) $\mathcal{M}' = \{ABD, BCD\}$.



Proposition 4.3.9. *Let e be a model on $\mathcal{M}$. If f extends e to $\mathcal{M}'$ then all of the sub-models induced by $\mathcal{M}'$ are non-contextual.*

Proof. We show that if $\{f_C\}_{C \in \mathcal{M}'}$ extends e to $\mathcal{M}'$ then each f'_C is a global section of the induced sub-model of e on C. The induced sub-model on C is a model defined on the measurement cover $(C, O, \mathcal{M}|_C)$ where $\mathcal{M}|_C = \{D \cap C\}_{D \in \mathcal{M}}$. Since f is an extension, $f_C|_{D \cap C} = e_{D \cap C}$ for all $D \in \mathcal{M}$, and f_C must be a global section. □

The converse is not necessarily true, however. It is possible that all sub-models that are induced in this way by elements of a cover $\mathcal{M}' \succeq \mathcal{M}$ have a global section but that one cannot find an extension $\{f_C\}_{C \in \mathcal{M}'}$ (canonical or otherwise) that is no-signalling. This is the situation for example 4.3.7, for which the induced sub-models on ABD and BCD are non-contextual, but, as we have seen, the model cannot be extended to the cover $\mathcal{M}' = \{ABD, BCD\}$.

Nevertheless, more can be said about the relationship between extendability and induced sub-models when we talk of the strong properties.

Corollary 4.3.10. *Let $\mathcal{M} \preceq \mathcal{M}'$ and let e be a model on $\mathcal{M}$. Then e is strongly non-extendable to $\mathcal{M}'$ if and only if there exists some $C \in \mathcal{M}'$ such that the induced sub-model of e on C is strongly contextual. In particular, e is strongly non-extendable to $\mathcal{M}' = \mathcal{P}_{n(\mathcal{M})}X$ if and only if it has a strongly contextual induced sub-model of size $n(\mathcal{M})$.*

Proof. This follows from proposition 4.3.5. □

4.4 Construction of Bell Models

We turn now to a second construction, which builds from empirical models on certain kinds of measurement covers Bell models that are equivalent in terms of contextuality. Note that this construction is not restricted to possibilistic models, and will work for any kind of empirical model. The idea is to start with a model on some measurement scenario $(X, O, \mathcal{M})$, and to transform this into a model on the Bell scenario $(\coprod_{i=1}^n X, O, \prod_{i=1}^n X)$. Measurements here are of the form $\langle x, i \rangle$ with $x \in X$ and $i \in \{1, \dots, n\}$ identifying the site (copy of X).

Definition 4.4.1. *Let $U \subseteq \coprod_{i=1}^n X$. An assignment $s \in \mathcal{E}(U)$ is said to be* codiagonal *if it satisfies*

$$\forall\, x \in X \,\text{.}\; \forall\, i, j \in \{1, \dots, n\} \,\text{.}\; \langle x, i \rangle, \langle x, j \rangle \in U \implies s(\langle x, i \rangle) = s(\langle x, j \rangle);$$

i.e. copies of the same measurement at different sites are assigned the same outcome. Equivalently, in categorical terms, an assignment $s : U \to O$ is codiagonal when it factors as

$$U \hookrightarrow \coprod_{i=1}^n X \xrightarrow{\nabla_n} X \longrightarrow O \,.$$

We denote the set of such assignments by $\mathcal{E}^{\nabla}(U)$.

With each set $U \subseteq \coprod_{i=1}^n X$ of measurements on the new scenario, we associate a set $\underline{U} \subseteq X$ of measurements on the original, which is obtained by forgetting the site information; i.e. $\underline{U} := \{x \in X \mid \exists\, i \,\text{.}\, \langle x, i \rangle \in U\}$. It is clear that there is a bijection $\mathcal{E}^{\nabla}(U) \cong \mathcal{E}(\underline{U})$, which commutes with restrictions to smaller contexts, and we write $\underline{s}$ for the image of an assignment $s \in \mathcal{E}^{\nabla}(U)$ under this map. Recall also that $\mathcal{P}_n X := \{Y \subseteq X \mid |Y| = n\}$.

Definition 4.4.2. *With any empirical model f defined on a measurement scenario $(X, O, \mathcal{P}_n X)$ we associate an n-partite empirical model f^{Bell} on the Bell scenario $(\coprod_{i=1}^n X, O, \prod_{i=1}^n X)$ defined by*

$$f^{\mathsf{Bell}}_C(s) = \begin{cases} f_{\underline{C}}(\underline{s}) & \text{if } s \in \mathcal{E}^{\nabla}(C) \\ 0 & \text{if } s \notin \mathcal{E}^{\nabla}(C) \end{cases} .$$

Proposition 4.4.3. *f^{Bell} is an empirical model.*

Proof. It is clear that in the Boolean case, f^{Bell}_C is a distribution over the assignments $\mathcal{E}^{\nabla}(C)$, and hence also over $\mathcal{E}(C)$. Moreover, in the general case, f^{Bell}_C is equivalent to $f_{\underline{C}}$, and since $\underline{C}$ is a context (not necessarily maximal) of the measurement scenario for f it is therefore a well-defined distribution. As for compatibility, let C_1 and C_2 be two maximal contexts, and let $C_{1,2} := C_1 \cap C_2$. For $t \in \mathcal{E}^{\triangledown}(C_{1,2})$, we have:

$$
\begin{aligned}
& f^{\mathsf{Bell}}_{C_1}\big|_{C_{1,2}}(t) \\
= \quad & \{ \text{ definition of marginalisation } \} \\
& \sum_{\substack{s \in \mathcal{E}(C_1) \\ s|_{C_{1,2}} = t}} f^{\mathsf{Bell}}_{C_1}(s) \\
= \quad & \{ \text{ since } f^{\mathsf{Bell}}_{C_1}(s) \neq 0 \text{ only if } s \text{ is codiagonal } \} \\
& \sum_{\substack{s \in \mathcal{E}^{\triangledown}(C_1) \\ s|_{C_{1,2}} = t}} f_{\underline{C_1}}(\underline{s}) \\
= \quad & \{ \text{ using the bijection } s \mapsto \underline{s} \ \} \\
& \sum_{\substack{\underline{s} \in \mathcal{E}(\underline{C_1}) \\ \underline{s}|_{\underline{C_{1,2}}} = \underline{t}}} f_{\underline{C_1}}(\underline{s}) \\
= \quad & \{ \text{ compatibility condition for the model } f \ \} \\
& \sum_{\substack{\underline{r} \in \mathcal{E}(\underline{C_2}) \\ \underline{r}|_{\underline{C_{1,2}}} = \underline{t}}} f_{\underline{C_2}}(\underline{r}) \\
= \quad & \{ \text{ same steps in reverse order for } C_2 \ \} \\
& f^{\mathsf{Bell}}_{C_2}\big|_{C_{1,2}}(t).
\end{aligned}
$$

For $t \in \mathcal{E}(U) \setminus \mathcal{E}^{\triangledown}(U)$, any assignment $s \in \mathcal{E}(C_i)$ that restricts to t is not codiagonal. Therefore $f^{\mathsf{Bell}}_{C_i}(s) = 0$ and compatibility holds trivially. We conclude that

$$f^{\mathsf{Bell}}_{C_1}\big|_{C_{1,2}} = f^{\mathsf{Bell}}_{C_2}\big|_{C_{1,2}}.$$

□

Proposition 4.4.4. *There is a bijection between the global sections of f and of f^{Bell}. In particular,*

- f^{Bell} *is non-local if and only if f is contextual,*
- f^{Bell} *is logically non-local if and only if f is logically contextual,*
- f^{Bell} *is strongly non-local if and only if f is strongly contextual.*

Proof. By the definition of f^{Bell}, it is clear that any assignments that are deemed possible by a global section $d \in \mathcal{D}_R\mathcal{E}(\coprod_{i=1}^n X)$ of the model must belong to $\mathcal{E}^\nabla(\coprod_{i=1}^n X)$. Recall that there exists a bijection $\mathcal{E}^\nabla(U) \cong \mathcal{E}(\underline{U}), s \mapsto \underline{s}$ for each $U \subseteq \coprod_{i=1}^n X$, and that these bijections commute with restrictions. The correspondence lifts to provide bijections between $\mathcal{D}_R\mathcal{E}^\nabla(U)$ and $\mathcal{D}_R\mathcal{E}(\underline{U})$ that commute with marginalisation. Therefore, if $U = \coprod_{i=1}^n X$ (and $\underline{U} = X$), the resulting bijection gives the desired correspondence between global sections. □

On a related note, it is worth pointing out that any compatible family of $\mathbb{Z}$-linear combinations of assignments in f can be lifted to a compatible family of this kind on f^{Bell} by taking its pre-image under diagonalisation, leading to the following proposition. It is unclear whether the converse holds.

Proposition 4.4.5. *The existence of a non-vanishing cohomological obstruction in f^{Bell} implies the existence of a non-vanishing obstruction in f.*

In good cases, we can use the canonical extension of the previous section to extend a model e on $(X, O, \mathcal{M})$ to a model e' on $(X, O, \mathcal{P}_n X)$ and then apply the present construction to obtain a Bell model which, in general, is equivalent to the original in terms of strong contextuality.

Corollary 4.4.6. *Let e be an empirical model on $(X, O, \mathcal{M})$ and suppose that e is canonically extendable to $(X, O, \mathcal{P}_n X)$. Then e'^{Bell} is strongly non-local if and only if e is strongly contextual.*

4.5 Bell Models from Kochen-Specker Models

The construction of an equivalent Bell model can be carried out for all Kochen-Specker models in which maximal contexts are all of the same size.

Proposition 4.5.1. *The Kochen-Specker model for any scenario $(X, O, \mathcal{M})$ in which all maximal contexts are of the same size (i.e. $|C| = n$ for all $C \in \mathcal{M}$) is canonically extendable to $(X, O, \mathcal{P}_n X)$.*

Proof. Let e be the Kochen-Specker model for the scenario $(X, O, \mathcal{M})$. For a context $C \in \mathcal{P}_n X$ and an assignment $s \in \mathcal{E}(C)$, we have

$$s \in S_{e'}(C) \;\;\leftrightarrow\;\; \forall\, W \subseteq C, W \in \downarrow\!\mathcal{M}.\; o(s|_W) \leq 1 \;\wedge\; \forall\, D \subseteq C, D \in \mathcal{M}.\; o(s|_D) = 1$$

Since $C \in \mathcal{P}_n X$, it cannot have any proper sub-context that is a maximal context in $\mathcal{M}$. Thus, we can write:

$$e'_C(s) = \begin{cases} 1 & \text{if } C \in \mathcal{M} \wedge o(s) = 1 \\ 1 & \text{if } C \notin \mathcal{M} \wedge \forall\, W \subseteq C, W \in \downarrow\mathcal{M}.\ o(s|_W) \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

This is clearly a distribution for every $C \in \mathcal{P}_n X$, as there is always at least one possible assignment.

Now, let $U \subsetneq C$ and consider the marginalisation $e'_C|_U$. First, we look at the case that $C \notin \mathcal{M}$. Then it is easy to see that, for any $t \in \mathcal{E}(U)$, $t \in S_{e'}(C)|_U$ implies that $\forall\, V \subseteq U, V \in \downarrow\mathcal{M}.\ o(t|_V) \leq 1$. Conversely, if the latter holds, one can extend t to C by assigning 0 to all other measurements, giving an assignment $s \in \mathcal{E}(C)$ that satisfies $\forall\, W \subseteq C, W \in \downarrow\mathcal{M}.\ o(s|_W) \leq 1$, since $o(s|_W) = o(s|_{W\cap U}) = o(t|_{W\cap U}) \leq 1$ for all such W. We then have $s \in S_{e'}(C)$, and hence $s|_U = t \in S_{e'}(C)|_U$. So,

$$t \in S_{e'}(C)|_U \quad \leftrightarrow \quad \forall\, V \subseteq U, V \in \downarrow\mathcal{M}.\ o(t|_V) \leq 1. \tag{4.1}$$

For the case that $C \in \mathcal{M}$, a section $t \in \mathcal{E}(U)$ is in $S_{e'}(C)|_U$ if and only if $o(t) \leq 1$. So equation 4.1 holds in this case, too, since U itself is one of the V's in the formula.

This shows that the marginalisation to any U is independent of the maximal context from which one starts, proving compatibility as required. □

Combining this with corollary 4.4.6 gives the following.

Corollary 4.5.2. *For any Kochen-Specker model e, the Bell model e'^{Bell} is well-defined, and is strongly non-local if and only if e is strongly contextual.*

Proposition 3.6.4 showed that symmetric Kochen-Specker models are contextual if and only if they are strongly contextual. This means that for the whole class of symmetric Kochen-Specker models (including the contextual triangle and the 18-vector model), we can construct Bell models that are equivalent in terms of contextuality.

Corollary 4.5.3. *For any symmetric Kochen-Specker model e, the Bell model e'^{Bell} is well-defined, and is non-local if and only if e is contextual.*

4.6 Examples

The Contextual Triangle

Carrying out the construction on the triangle yields the following $(2, 3, 2)$ model.

	00	01	10	11
$A\ A'$	1	0	0	1
$A\ B'$	0	1	1	0
$A\ C'$	0	1	1	0
$B\ A'$	0	1	1	0
$B\ B'$	1	0	0	1
$B\ C'$	0	1	1	0
$C\ A'$	0	1	1	0
$C\ B'$	0	1	1	0
$C\ C'$	1	0	0	1

We include dashes to make it clear that different measurements in the same context are now considered to belong to different sites. It is especially interesting that the model can be seen to contain many different PR boxes [93] as sub-models. These are:

	00	01	10	11
$A\ A'$	1	0	0	1
$A\ C'$	0	1	1	0
$B\ A'$	0	1	1	0
$B\ C'$	0	1	1	0

	00	01	10	11
$B\ A'$	0	1	1	0
$B\ B'$	1	0	0	1
$C\ A'$	0	1	1	0
$C\ B'$	0	1	1	0

	00	01	10	11
$B\ A'$	0	1	1	0
$B\ C'$	0	1	1	0
$C\ A'$	0	1	1	0
$C\ C'$	1	0	0	1

and those obtained by reversing the order of the measurements.

We note that neither the triangle nor the PR box is realisable in quantum mechanics. The triangle provides the simplest possible example of a contextual model, and the PR box is the only strongly contextual $(2, 2, 2)$ model (proposition 2.6.4).

The Peres-Mermin Square

The Peres-Mermin square [85] is another example of a strongly contextual model, though it does not fall into the class of Kochen-Specker models. It, too, has the desirable property that it can be extended to $\mathcal{P}_n X$. We can therefore construct a tripartite Bell model from it which is equivalent in terms of strong contextuality. The constructed model contains 36 different $(3, 5, 2)$ non-local sub-models, which are essentially 'padded-out' versions of the square. The following table represents the non-local sub-model on the measurement cover

$$\mathcal{M} = \{A, B, C, D, G\} \times \{B, D, E, F, H\} \times \{C, F, G, H, I\},$$

though we only explicitly write those rows that do not have full support.

	000	001	010	011	100	101	110	111
$A\ \ B'\ C''$	0	1	1	0	1	0	0	1
$D\ \ E'\ F''$	0	1	1	0	1	0	0	1
$G\ \ H'\ I''$	0	1	1	0	1	0	0	1
$A\ \ D'\ G''$	1	0	0	1	0	1	1	0
$B\ \ E'\ H''$	1	0	0	1	0	1	1	0
$C\ \ F'\ I''$	1	0	0	1	0	1	1	0

Ignoring the additional rows, and the dashes, which are there as a reminder that the measurements belong to different sites, this looks just like the table for the Peres-Mermin model itself; though it is a genuinely new strongly non-local Bell model.

An interesting point is that since the Peres-Mermin contextuality proof is based on a parity argument, just like the GHZ proof, one might expect that the Bell model we have constructed should contain a GHZ sub-model. However, it can easily be shown by comparison with the table for the GHZ-Mermin model (chapters 1 & 2) that this is not the case. A further point is that it contains no tripartite Hardy paradox (chapter 2).

4.7 Discussion

We have dealt with two related ideas. The refinement of extendability introduced here is a development of the sheaf-theoretic framework, which captures partial approximations to locality and non-contextuality. This allows us to characterise contextuality and strong contextuality in sub-models of an empirical model, as we have seen in section 4.3.

The second idea is to introduce a method of constructing Bell models from models of a more general kind in such a way that these are equivalent in terms of non-locality/contextuality. This can even work at the level of probabilities. Equivalent Bell forms of models are desirable since, both practically and theoretically, it is easier to motivate a notion of locality in such scenarios than the equivalent notion of non-contextuality in a more general scenario, as one can always appeal to relativity as a justification for certain assumptions. We have also mentioned that non-locality is better understood as an information theoretic resource.

These two ideas are related by the fact that, for any model, the existence of the canonical extension to $\mathcal{P}_n X$ will guarantee the ability to construct a Bell model that is equivalent in terms of strong contextuality. We have proved that for the entire class of Kochen-Specker models with maximal contexts of constant size we can carry out this construction, and that for the symmetric models the equivalence even holds for contextuality. Even in the more general form, which applies to strong contextuality only, this is a very useful result since so many of the familiar examples of non-local/contextual models are strongly contextual: we have mentioned the GHZ-Mermin model, the 18-vector Kochen-Specker model, the Peres-Mermin model, and the Popescu-Rohrlich correlations.

There are several open questions arising from this work. We would like to know whether there is an analogue of Vorob'ev's theorem [103] for this partial notion of extendibility; that is, given any measurement cover $\mathcal{M}'$, can there be a complete characterisation of the measurement covers $\mathcal{M} \preceq \mathcal{M}'$ such that any empirical model defined on $\mathcal{M}$ is extendable to $\mathcal{M}'$. This could potentially lead to applications to macroscopic realism similar to [100]. We would also like to know whether there are other general classes of 'good' models for which we can guarantee the ability to extend to $\mathcal{P}_n X$ and thereby construct Bell models that are equivalent either in terms of contextuality or strong contextuality: a class of Peres-Mermin-like models for example.

It is especially interesting that when we constructed the equivalent Bell model for the contextual triangle, we ended up with what is essentially a folding of PR boxes. This appears to point to a deeper relationship between the models, which merits further investigation. The PR box has been much studied and has been considered, for example, as a possible unit of non-locality [19]. Since the triangle is the simplest possible example of a contextual model, it is conceivable that it could be a unit of contextuality. One might hope for some sort of analogous result to Kuratowski's theorem for graphs, which states that a graph is planar if and only if it does not contain K_5 or $K_{3,3}$ as subgraphs. For example, it could be that, for some notion of reducibility, contextual models must reduce to the triangle or to elements of some set of irreducible models containing the triangle.

Another important issue that has not been dealt with so far is that of quantum realisability: given that a model is quantum realisable, we would like to understand when its extensions are and vice versa. The issue of quantum realisability was touched on in a different context in chapter 2. Here, it is especially relevant to the example of the Peres-Mermin model. An aim of this work is to propose Bell tests that correspond to contextuality proofs such as that of Peres & Mermin by giving a means of quantum mechanically reproducing its equivalent Bell model. First, however, it will be necessary to understand how quantum realisability relates to our constructions.

Chapter 5

On the Reality of Observable Properties

The issue of the reality of the wavefunction has received a lot of attention recently (see especially [94, 41]). In this chapter, we will show that insights may also be gained by considering the reality of objects and properties in physical theories more generally, and in particular that such an approach can provide a new perspective on non-locality and contextuality. The first step will be to introduce a suitably general criterion for reality inspired by the Harrigan-Spekkens criterion for the reality of the wavefunction [63], which was the subject of the Pusey-Barrett-Rudolph theorem [94].

The aim is to formulate such ideas in a manner that can allow for a deeper, structural understanding of what is at play, and to attempt to bring considerations of this kind within the scope of the methods of the sheaf-theoretic approach. Indeed, the new criterion has several advantages over the original. It avoids technical difficulties, and due to its generality it can be applied within any ontological physical theory: generalised probabilistic theories [18], or classical mechanics, for example.

The initial research contained in this chapter also demonstrates that such considerations can provide an alternative perspective on foundational questions in general. We find a novel characterisation, for any predictive theory, of both local and non-contextual correlations as those that can arise from observations of properties that are 'real'. This ties together the notions of locality and reality, bringing to light a link between the Bell and PBR theorems, which deal, respectively, with these properties.

Much of the foundations of quantum mechanics literature, including the recent developments on the reality of the wavefunction that we have mentioned, deals with hidden variable models (or ontological models). We will therefore begin with a brief overview of this framework. It has already been pointed out in chapter 1 that local hidden variable models can be subsumed by the sheaf-theoretic framework; we will see in more detail why this is so in section 5.3.

5.1 Ontological Models

We are concerned with theories that give operational predictions for outcomes to measurements; in other words, empirical models. Quantum mechanics is one such theory. We have seen how it gives rise to empirical models in chapter 1. To give this a more operational treatment, we associate a density matrix ρ^p with each preparation p, a POVM $\{E_o^m\}_{o \in O}$ with each measurement m, and prescribe the probability of the outcome o given preparation p and measurement m by

$$p(o \mid m, p) = \mathrm{tr}(\rho^p \, E_o^m).$$

We wish, more generally, to consider theories with this kind of operational structure. For each system we assume spaces P of preparations, X of measurements, and O of outcomes. Again, there may be some compatibility structure on the space of measurements, say $\mathcal{M} \subseteq \mathcal{P}(X)$, specifying which sets of measurements can be made jointly (in quantum mechanics, this is specified by the commutative sub-algebras of the algebra of observables). We additionally assume a space Λ of *ontic states*, over which each preparation induces a probability distribution.

In an effort to simplify notation, we will use an overline to denote a tuple of joint measurements $\overline{m} \in \mathcal{M}$ or joint outcomes $\overline{o} \in \mathcal{E}(\overline{m})$, whereas $m \in X$ and $o \in O$ denote individual measurements and outcomes, respectively. We will treat preparations and ontic states similarly in section 5.4. Recall from chapter 1 that $\mathcal{E} : X \to O^X$ is the event sheaf, and that $\mathcal{E}(\overline{m})$ denotes the set of functions $\overline{o} : \overline{m} \to O$.

Definition 5.1.1. *An* ontological *or* hidden variable model h *over* Λ *specifies:*

1. *A distribution $h(\lambda \mid p)$ over Λ for each preparation $p \in P$;*
2. *For each $\lambda \in \Lambda$ and set of compatible measurements $\overline{m} \in \mathcal{M}$, a distribution*

   $$h(\overline{o} \mid \overline{m}, \lambda)$$

   over functional assignments $\mathcal{E}(\overline{m})$ of outcomes to these measurements.

The operational probabilities *are then prescribed by*

$$h(\overline{o} \mid \overline{m}, p) = \int_\Lambda d\lambda \; h(\overline{o} \mid \overline{m}, \lambda) \; h(\lambda \mid p). \tag{5.1}$$

The terms ontological model and hidden variable model are both used in the literature, but recently the term ontological model has gained some popularity. It may be a more suitable term in the sense that the 'hidden' variable need not necessarily be hidden at all: it could be directly observable. In Bohmian mechanics [27, 28], for example, position and momentum play the role of the hidden variable. It also carries the connotation that such a model is an attempt to describe some underlying ontological reality.

Definition 5.1.2. *A theory which determines the operational probabilities will be referred to as an* ontological theory *over* Λ*.*

We are especially interested in ontological models and theories that can reproduce quantum mechanical predictions. Of course the simplest such theory is quantum mechanics itself, regarded as an ontological theory.

Example 5.1.3 (ψ-complete Quantum Mechanics)**.** *The ontic state is identified with the quantum state. A preparation produces a density matrix, which is viewed as a distribution over the projective Hilbert space associated with the system. By construction, the operational probabilities are those given by the Born rule.*

Of course, quantum mechanics, treated as an ontological theory in itself in this way, has certain non-intuitive features. Einstein, Podolsky & Rosen provided one early discussion of the fact [48]; but later results such as Bell's theorem [22] and the Kochen-Specker theorem [73] provided more clarity on the fact that non-locality and contextuality are necessary features of any theory that can account for quantum mechanical predictions. In order to address these issues, we now consider some relevant properties in the setting of ontological models. These are similar to the properties of models in the possibilistic framework for Bell models from section 2.2.

Definition 5.1.4. *An ontological model is* λ-independent *if and only if the distributions over* Λ *induced by each preparation* $p \in P$ *do not depend on the measurements* $\overline{m} \in \mathcal{M}$*.*

We have already implicitly assumed this in definition 5.1.1, but it is worth making this clear since it is crucial for all of the familiar no-go theorems. A λ-*dependent* model, on the other hand, would have $h(\lambda \mid p, \overline{m})$ rather than $h(\lambda \mid p)$ in equation (5.1).

Definition 5.1.5. *An ontological model is* deterministic *if and only if for each* $\lambda \in \Lambda$ *and set of compatible measurements* $\overline{m} \in \mathcal{M}$ *there exists some joint outcome* $\overline{o} \in \mathcal{E}(\overline{m})$ *such that* $h(\overline{o} \mid \overline{m}, \lambda) = 1$*.*

Such a model is deterministic with respect to the ontic states; the outcomes to all measurements are determined with certainty by the ontic state.

Definition 5.1.6. *An ontological model is* parameter-independent *if and only if the marginal probabilities* $h(o \mid m, \lambda)$ *are well-defined for all* $o \in O$, $\overline{m} \in \mathcal{M}$ *and* $\lambda \in \Lambda$.

For any $m \in \overline{m}$ and $\lambda \in \Lambda$ we can find a distribution $h(o \mid m, \lambda)$ over O by marginalising from $h(\overline{o} \mid \overline{m}, \lambda)$. Parameter independence requires that the same distribution be obtained regardless of which set of measurements we marginalise from; and thus asserts that the probabilities of outcomes to a particular measurement do not depend on the other measurements being performed. It essentially amounts to imposing no-signalling with respect to the ontic states.

Definition 5.1.7. *An ontological model is* local/non-contextual *if and only if it is both deterministic and parameter-independent; empirical correlations are* local/non-contextual *if and only if they can be realised by a local/non-contextual model.*

This says that for each ontic state there is a certain outcome to any measurement that can be performed, and that this does not depend on which other measurements are made. The term local is generally only used when the system being modelled is spatially distributed; where such an arrangement is not assumed, the model is said to be non-contextual.

As we pointed out in chapter 1, another definition of locality that is common in the literature concerns the factorisability of the distributions $h(\overline{o} \mid \overline{m}, \lambda)$. These were shown to be equivalent in the sense that they generate the same sets of empirical models in [4], which built on work by Fine [50] that was specific to $(2, 2, 2)$ Bell scenarios.

5.2 A Criterion for Reality

In this section we will use the terminology of Harrigan & Spekkens [63], as that which has been established in the literature. We will first present the Harrigan-Spekkens criterion for the reality, or *onticity*, of the wavefunction, which will then be reformulated and generalised. For this, we need only postulate, for each system, a space Λ of ontic states. These can be considered to correspond to real, physical states of the system. The idea will be that objects or properties that are determined with certainty by the ontic state are themselves ontic. Indeed, the term ontic was chosen as meaning that which relates to real as opposed to phenomenal existence.

On the other hand, objects or properties that are not determined with certainty are said to be *epistemic*. The dictionary meaning of this word is that which relates to knowledge or to its degree of validation. Here, the term reflects the fact that objects and properties of this kind are necessarily probabilistic and could thus be assumed to represent a degree of knowledge about some underlying ontic object or property. We also note, however, that regardless of any physical significance attached to these definitions, the results that follow will hold on the purely mathematical level.

As well as the existence of an ontic state space, Harrigan & Spekkens also assume that the preparation of any quantum state $|\psi\rangle$ induces a distribution $\mu_{|\psi\rangle}$ over the ontic state space Λ for that system, which represents the probability of the system being in each ontic state given that is was prepared in this way.

Definition 5.2.1 (Harrigan & Spekkens)**.** *If for each system, and for all wavefunctions $|\psi\rangle \neq |\phi\rangle$ the distributions $\mu_{|\psi\rangle}$ and $\mu_{|\phi\rangle}$ have non-overlapping supports, then the wavefunction is said to be* ontic*. If not, then there exist some $|\psi\rangle \neq |\phi\rangle$ such that $\mu_{|\psi\rangle}(\lambda) > 0$ and $\mu_{|\phi\rangle}(\lambda) > 0$ for some $\lambda \in \Lambda$, and the wavefunction is said to be* epistemic*.*

We now present our more general reformulation of the definition which can be applied to any object or property. Though the wavefunction would more usually be considered as an object than a property of a system, for simplicity we only use the term property from now on. It may not immediately be clear how these relate, but this will be addressed by proposition 5.2.3.

Definition 5.2.2. *A $\mathcal{V}$-*valued property *over Λ is a function $f : \Lambda \to \mathcal{D}(\mathcal{V})$, where $\mathcal{D}(\mathcal{V})$ is the set of probability distributions over $\mathcal{V}$. The property is said to be* ontic *in the special case that, for all $\lambda \in \Lambda$, the distribution $f(\lambda)$ over $\mathcal{V}$ is a delta function. Otherwise, it is said to be* epistemic*.*

Ontic properties, therefore, are generated by functions $\widehat{f} : \Lambda \to \mathcal{V}$, which map each ontic state to a unique value. For epistemic properties, however, there is at least one ontic state that is compatible with two or more distinct values in $\mathcal{V}$.

We now show how this definition relates to that of Harrigan & Spekkens. Any $\mathcal{V}$-valued property f specifies probability distributions over $\mathcal{V}$, conditioned on Λ. Bayesian inversion can be used to obtain probability distributions over Λ, conditioned on $\mathcal{V}$, which we (suggestively) label $\{\mu_v\}_{v \in \mathcal{V}}$. Explicitly,

$$\mu_v(\lambda) := \frac{f(\lambda)(v) \cdot p(\lambda)}{\int_\Lambda f(\lambda')(v) \cdot p(\lambda')\, d\lambda'}, \tag{5.2}$$

assuming a uniform distribution $p(\lambda)$ on Λ. Note that this is only well-defined for finite Λ, and that a more careful measure theoretic treatment, which will not be provided here, is required for the infinite case.

Proposition 5.2.3. *A $\mathcal{V}$-valued property over finite Λ is ontic (according to definition 5.2.2) if and only if the distributions $\{\mu_v\}_{v\in\mathcal{V}}$ have non-overlapping supports.*

Proof. Let $\lambda \in \Lambda$ and let $v, v' \in \mathcal{V}$ such that $v \neq v'$. Suppose the property f is ontic in the sense of definition 5.2.2. We assume for a contradiction that $\mu_v(\lambda) > 0$ and $\mu_{v'}(\lambda) > 0$. Then, by (5.2), $f(\lambda)(v) > 0$ and $f(\lambda)(v') > 0$; but since f is ontic,

$$v_\lambda = v \neq v' = v_\lambda,$$

where $v_\lambda := \widehat{f}(\lambda)$.

Conversely, suppose that the distributions $\{\mu_v\}_{v\in\mathcal{V}}$ have non-overlapping supports and assume for a contradiction that $f(\lambda)(v) > 0$ and $f(\lambda)(v') > 0$. Then, by (5.2), $\mu_v(\lambda) > 0$ and $\mu_{v'}(\lambda) > 0$. □

One way of thinking about this correspondence is as a special case of the dual equivalence between the category of von Neumann algebras and $*$-homomorphisms and the category of measure spaces and measurable functions [66].

To illustrate, some simple examples of ontic and epistemic properties are the following.

Example 5.2.4 (Classical Mechanics)**.** *The phase space of a system is taken to be the ontic state space. Classical mechanical observables (energy, momentum, etc.) are represented by real-valued functions on phase space, and are therefore always* ontic.

Example 5.2.5. *Consider an experiment in which a bag is prepared containing two coins, which can be green or white, with equal probability, but are otherwise identical. We claim that the process of removing one and checking its colour measures an* epistemic *property. If the ontic states are $\Lambda = \{GG, GW, WG, WW\}$, the property cannot be represented by a $\{G, W\}$-valued function on Λ. Given the ontic state GW, for example, both G and W are compatible, and can arise with equal probability.*

In example 5.2.5, according to our definition, the information gained by making the measurement described is epistemic. In other words, the property that is actually being measured is epistemic with respect to the state of the bag. It might also be said the example describes a fuzzy measurement on the state of the bag.

The ontic criterion for reality set out in definition 5.2.2 has several advantages.

- It is fully general and can be applied to any object or property in any ontological theory.
- It avoids measure theoretic problems relating to sets of measure zero that are inherent to that of Harrigan & Spekkens.
- It is mathematically straightforward and conceptually transparent.

5.3 Observable Properties

If we assume that the outcomes of measurements provide the values of properties of a system, then for each measurement $m \in X$ there should exist an O-valued property $f_m : \Lambda \to \mathcal{D}(O)$ such that $f_m(\lambda)(o) = h(o \mid m, \lambda)$ for all $\lambda \in \Lambda$ and $o \in O$.

Definition 5.3.1. *The* observable properties *of an ontological model h over Λ are the O-valued properties $f_m : \Lambda \to \mathcal{D}(O)$ given by*

$$f_m(\lambda)(o) := h(o \mid m, \lambda) \tag{5.3}$$

for each $m \in X$ such that the marginal $h(o \mid m, \lambda)$ is well-defined.

Theorem 5.3.2. *An ontological model is local/non-contextual if and only if all measurements are of ontic observable properties.*

Proof. First, we claim that a model is deterministic if and only if its observable properties are ontic. This holds since, by (5.3),

$$h(o \mid m, \lambda) = 1 \qquad \Leftrightarrow \qquad f_m(\lambda)(o) = 1.$$

Next, we claim that a model is parameter independent if and only if all measurements are of observable properties. This holds since, by definition 5.3.1, all measurements are of observable properties if and only if all marginals $h(o \mid m, \lambda)$ are well-defined. The result follows. $\square$

This is a new characterisation of locality, which falls out easily from the definitions. It is similar to the Kochen-Specker [73] or topos approach [68] treatment of non-contextuality. This can provide an alternative and sometimes simpler approach to many results. The first result we mention shows that local ontological models have a canonical form. In fact, it shows that local ontological or hidden variable models can equivalently be expressed as distributions over the set of global assignments. In this sense it shows how local ontological models are subsumed by the sheaf-theoretic approach. It has recently been proved in measure theoretic generality [31], and can also be seen to generalise earlier work by Fine [50]. An interesting related point that will be proved in chapter 6 is that, by allowing for negative probabilities, canonical models can also generate all no-signalling correlations.

Theorem 5.3.3. *Local models can be expressed in a* canonical form*, with an ontic state space* $\Omega := \mathcal{E}(X)$*, and probabilities*

$$h(\overline{o} \mid \overline{m}, \omega) = \prod_{m \in \overline{m}} \delta\left(\omega(m), \overline{o}(m)\right)$$

for all $\overline{m} \in \mathcal{M}$*,* $\overline{o} \in \mathcal{E}(\overline{m})$*, and* $\omega \in \Omega$*.*

Proof. By theorem 5.3.2, a local model h over Λ has a set $\{\widehat{f_m} : \Lambda \to O\}_{m \in X}$ of ontic observable properties. For each $\lambda \in \Lambda$, we define a function $\omega_\lambda \in \mathcal{E}(X)$ by $\omega_\lambda(m) := \widehat{f_m}(\lambda)$. Then the function $c : \Lambda \to \mathcal{E}(X)$ defined by $c(\lambda) := \omega_\lambda$ takes the original to the canonical ontic state space.

We first prove the claim that if $\lambda, \lambda' \in c^{-1}(\omega)$ for some $\omega \in \mathcal{E}(X)$ then

$$h(\overline{o} \mid \overline{m}, \lambda) = h(\overline{o} \mid \overline{m}, \lambda')$$

for all $\overline{m} \in \mathcal{M}$ and $\overline{o} \in \mathcal{E}(\overline{m})$. Since $\lambda, \lambda' \in c^{-1}(\omega)$, then $\omega_\lambda = \omega_{\lambda'}$, and therefore $\widehat{f_m}(\lambda) = \widehat{f_m}(\lambda')$ for all $m \in X$. It follows that

$$\begin{aligned}
h(\overline{o} \mid \overline{m}, \lambda) &= \prod_{m \in \overline{m}} h(\overline{o}(m) \mid m, \lambda) \\
&= \prod_{m \in \overline{m}} f_m(\lambda)\left(\overline{o}(m)\right) \\
&= \prod_{m \in \overline{m}} \delta\left(\widehat{f_m}(\lambda), \overline{o}(m)\right) \\
&= \prod_{m \in \overline{m}} \delta\left(\widehat{f_m}(\lambda'), \overline{o}(m)\right) \\
&= \cdots \\
&= h(\overline{o} \mid \overline{m}, \lambda'),
\end{aligned}$$

where the first equality can be shown to hold by locality.

The canonical model h over Ω is defined by

$$h(\overline{o} \mid \overline{m}, \omega) := h(\overline{o} \mid \overline{m}, \lambda_\omega)$$

and

$$h(\omega \mid p) := \sum_{\lambda \in c^{-1}(\omega)} h(\lambda \mid p)$$

for all $\overline{m} \in \mathcal{M}$, $\overline{o} \in \mathcal{E}(X)$, $\omega \in \Omega$, $\lambda \in \Lambda$, and any $\lambda_\omega \in c^{-1}(\omega)$. The canonical model realises the same operational probabilities as the original, since

$$\begin{aligned}
\sum_{\omega \in \Omega} h(\overline{o} \mid \overline{m}, \omega)\, h(\omega \mid p) &= \sum_{\omega \in \Omega} \left(h(\overline{o} \mid \overline{m}, \lambda_\omega) \sum_{\lambda \in c^{-1}(\omega)} h(\lambda \mid p) \right) \\
&= \sum_{\omega \in \Omega} \sum_{\lambda \in c^{-1}(\omega)} h(\overline{o} \mid \overline{m}, \lambda)\, h(\lambda \mid p) \\
&= \sum_{\lambda \in \Lambda} h(\overline{o} \mid \overline{m}, \lambda)\, h(\lambda \mid p),
\end{aligned}$$

where the second equality holds by the previous claim. Moreover, the operational probabilities can be simplified as follows.

$$\begin{aligned}
h(\overline{o} \mid \overline{m}, \omega) &= h(\overline{o} \mid \overline{m}, \lambda_\omega) \\
&= \prod_{m \in \overline{m}} \delta \left(\widehat{f_m}(\lambda_\omega), \overline{o}(m) \right) \\
&= \prod_{m \in \overline{m}} \delta \left(\omega(m), \overline{o}(m) \right)
\end{aligned}$$

□

The next proposition will not be surprising in light of the EPR argument [48]. It shows that if one were to take the view that quantum mechanics is ψ-complete then all non-trivial observables are epistemic or inherently probabilistic. Indeed, we can obtain a re-statement of the EPR result as a corollary.

Proposition 5.3.4. *Any non-trivial quantum mechanical observable is epistemic with respect to ψ-complete quantum mechanics.*

Proof. Any observable $\hat{A} \neq \mathbf{I}$ has eigenvectors, say $|v_1\rangle$ and $|v_2\rangle$, corresponding to distinct eigenvalues, say o_1 and o_2. Consider any state $|\psi\rangle$ such that $\langle v_1|\psi\rangle > 0$ and $\langle v_2|\psi\rangle > 0$. In a ψ-complete model, the wavefunction is the ontic state, so $\lambda = |\psi\rangle$. Then

$$f_{\hat{A}}(\lambda)(o_1) = h(o_1 \mid \hat{A}, \lambda) = |\langle v_1|\psi\rangle|^2 > 0,$$

and similarly $f_{\hat{A}}(\lambda)(o_2) > 0$. Therefore $f_{\hat{A}}$ is epistemic. □

Corollary 5.3.5 (EPR)**.** *Under the assumption of locality, quantum mechanics cannot be ψ-complete.*

Proof. By proposition 5.3.4, any non-trivial quantum observable is epistemic with respect to ψ-complete quantum mechanics. Therefore, by theorem 5.3.2, ψ-complete quantum mechanics is not local. □

This is the same result that was argued for by EPR, though the proof has more in common with an earlier argument by Einstein at the 1927 Solvay conference [16], and also with a more recent, general treatment found in [32] and mentioned again in [3].

5.4 The PBR Theorem

In this section we briefly make some observations relating to the PBR theorem, which deals with the reality of the wavefunction. One of the assumptions for this result is *preparation independence* [94]:

> systems that are prepared independently have independent physical states.

The other assumptions are implicit in the present framework.

Theorem 5.4.1 (PBR)**.** *For any preparation independent theory that reproduces (a certain set of) quantum correlations, the wavefunction is ontic.*

The preparation independence assumption is concerned with the composition of systems and has not appeared in other no-go results. We will attempt to give this a more careful treatment. First of all, the PBR theorem describes a *preparation scenario*. Generalising, this can be thought of as a kind of dual to a measurement scenario, in which the preparations P play the role of measurements and the ontic states Λ play the role of outcomes. Just as we had a compatibility structure $\mathcal{M}$ for measurements, which in Bell scenarios allowed us to chose one measurement from each site, we should in general have a compatibility structure $\mathcal{P}$ for preparations, which in the case of the PBR result allows us to chose one preparation from each site. We should allow for joint ontic states $\overline{\lambda} : \overline{p} \to \Lambda$, just as we allowed for joint outcomes. It is possible to modify the definitions of an ontological model and the properties from section 5.1 in an obvious way to account for this additional structure.

Definition 5.4.2. *An ontological theory h over Λ is* preparation independent *if and only if we can factor*

$$h(\overline{\lambda} \mid \overline{p}) = \prod_{p \in \overline{p}} h(\lambda_p \mid \overline{p}) \tag{5.4}$$

for all $\overline{p} \in \mathcal{P}$, where $\lambda_p := \overline{\lambda}|_p$.

Presented in this way, this is clearly seen to be analogous to non-contextuality or Bell locality of an empirical model. An intriguing question is what happens if this is relaxed to an assumption analogous to no-signalling, in which we only assume that the marginal distributions $h(\lambda_p \mid \overline{p})$ are well-defined: a sort of 'no-preparation-signalling' assumption. In this case, it is easy to see that the PBR argument no longer holds. The argument even makes tacit assumptions that each sub-system has a definite hidden state and that these are not correlated, analogous to those pointed out in the naïve introduction to the Hardy paradox in chapter 2. It is true that the relaxed assumption would allow for global or non-local correlations in the joint ontic state $\overline{\lambda}$, but perhaps, in light of the Bell and Kochen-Specker theorems, this should not be so surprising. An important question that remains to be answered, therefore, is whether by another argument a result similar to (or indeed contrasting with) that of PBR can be proved.

Another interesting observation, which is also pointed out in [63], is that onticity of the wavefunction is actually inconsistent with locality. This can be demonstrated as a consequence of what Schrödinger called *steering* [97]. If a local measurement in the basis $\{|0\rangle, |1\rangle\}$ is made on the first qubit of the state

$$\left|\phi^+\right\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right)$$

then this can be considered as a remote preparation of the second qubit in one of the states $|0\rangle$ or $|1\rangle$, and similarly for a measurement in the basis $\{|+\rangle, |-\rangle\}$. If the second sub-system has an ontic state λ that is independent of measurements made elsewhere, then λ must be consistent with one state from each of the sets $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$, but this contradicts the onticity of the wavefunction.

We therefore arrive at the following theorem, which we propose to think of as a weak Bell theorem, since it draws the same conclusion as Bell's theorem [22] but with the extra assumption of preparation independence.

Theorem 5.4.3. *Quantum mechanics is not realisable by any preparation independent, local ontological theory.*

Proof. This follows from the PBR theorem and the occurrence of steering in quantum mechanics. □

5.5 Discussion

We have presented a generalised reformulation of the Harrigan-Spekkens criterion for the reality or onticity of the wavefunction. The reformulation aspect of the present definition can be thought of as a special case of the dual equivalence between the category of von Neumann algebras and $*$-homomorphisms and the category of measure spaces and measurable functions. It has been seen to have several advantages: it avoids measure theoretic technicalities and is mathematically and conceptually straightforward. Of course, it is also general enough to apply to any object or property in any ontological theory.

The first obvious application of the criterion to an object or property other than the wavefunction is to the observable properties of a system. This led to a new characterisation of locality and non-contextuality in terms of the nature of the observed properties. This can provide a useful tool for looking at foundational results: we have used it to obtain a new proof that local ontological models have a canonical form (which allows them to be subsumed by the sheaf-theoretic approach), and also to gain another perspective on the EPR argument. The characterisation is similar to the Kochen-Specker [73] or topos approach [68] treatment of non-contextuality.

It is interesting that the characterisation draws a connection between locality and onticity: these are the properties that are dealt with by the Bell and PBR theorems, respectively. A further connection was found in theorem 5.4.3, which showed that a weakened version of Bell's result can be obtained by an argument that combines the PBR result with the incompatibility of steering and the onticity of the wavefunction.

In relation to the PBR result itself, we have attempted to give a more careful treatment of the assumption of preparation independence, and made a concrete analogy between this property and locality/non-contextuality. It is possible to relax the assumption to something analogous to no-signalling, in which case we have pointed out that the PBR argument no longer holds. This amounts to introducing global or non-local correlations in the joint ontic state, which at least seems consistent with the Bell and Kochen-Specker theorems. An open question is whether by another argument the result can be shown to hold under the relaxed, 'no-preparation-signalling' assumption.

Taken further, the analogy between measurement and preparation scenarios of section 5.4 suggests that a sheaf-theoretic approach can also be taken for preparation scenarios. An important question, then, is how to give a unified treatment of both kinds of scenario. It seems possible that the approach to ontological theories of section 1.7 could be extended to allow for such a treatment, with the ontic states defined to be the Gleason states of corollary 1.7.1.

Chapter 6

Computational Tools

Though many of the topics discussed and presented throughout this dissertation are of quite a theoretical nature, computational exploration can play an important role the research programme. In this chapter, we present a computational approach to calculating the degree of contextuality of any empirical model and to finding logical Bell inequalities [11] using linear programming methods. This has been implemented as a *Mathematica* package, which allows one to calculate quantum empirical models given a (pure or mixed) state and sets of compatible observables, and to calculate the degree of contextuality of any empirical model. We stress that this is fully general and applies to any measurement scenario, including of course all Bell scenarios.

This kind of tool can be useful, for example, in attempting to classify non-local states [9], which will be a goal of future research. As a demonstration of how the package works, we use it to explore the non-locality of empirical models arising from the ϕ^+ and GHZ(n) states. In this way, new sets of measurements on the ϕ^+ Bell state which give rise to empirical models that achieve the maximum violation of the CHSH inequality are found, as well as new sets of measurements on the GHZ(n) states that lead to similar logical proofs of non-locality.

A particularly interesting result in which computational exploration has already been of importance shows that an empirical model is no-signalling if and only if it can be realised by a local ontological or hidden variable model with negative probabilities. This is proved in [4] for any measurement cover; we present a precursor of that result, which shows that the equivalence holds for all Bell scenarios, and in which the role of computational exploration as a guiding tool will be apparent.

6.1 Linear Algebra & Contextuality

Recall from chapter 1 that a model e on a measurement scenario $(X, O, \mathcal{M})$ is local/non-contextual if and only if it has a global section; i.e. a distribution $d \in \mathcal{D}_R\mathcal{E}(X)$ such that $d|_C = e_C$ for all $C \in \mathcal{M}$. This means that the empirical model can equivalently be expressed as a mixture of global assignments $\mathcal{E}(X)$. Similarly, it was shown in chapter 2 that local models on Bell scenarios are those which can be obtained as stochastic mixtures of local deterministic models.

The tabular representation of Bell models from chapter 2 is convenient for illustrating this point with a simple example. The following is the local model obtained by local X and Y measurements on each qubit of the $|\phi^+\rangle$ state.

	½	0	¼	¼
	0	½	¼	¼
	¼	¼	0	½
	¼	¼	½	0

It is clear that this model can be decomposed into a uniform distribution over each of the four compatible deterministic grids.

With this example in mind, it is possible to show that, for any model, the problem of finding a global section is equivalent to finding a solution to a particular system of linear equations. This was pointed out in [11] and in early versions of [4]. We can represent empirical models as vectors: for example, the previous model would be

$$\mathbf{v} = \{1/2, 0, 0, 1/2, \quad 1/4, 1/4, 1/4, 1/4, \quad 1/4, 1/4, 1/4, 1/4, \quad 0, 1/2, 1/2, 0\}.$$

For any measurement scenario $(X, O, \mathcal{M})$, each global assignment $t \in \mathcal{E}(X)$ can also be written as a vector. We can form a matrix, $\mathbf{M}$, with these as columns, which we call the *incidence matrix*. In a Bell scenario, this corresponds to writing the deterministic grids as vectors and forming a matrix with these as columns. More carefully, we specify an enumeration $\{s_i\}_{i=1}^p$ of the set of all 'local' assignments $\coprod_{C \in \mathcal{M}} \mathcal{E}(C)$, an enumeration $\{t_j\}_{j=1}^q$ of the set of global assignments $\mathcal{E}(X)$, and define

$$\mathbf{M}[i, j] := \begin{cases} 1 & \text{if } s_i \in \mathcal{E}(C) \text{ and } t_j|_C = s_i \\ 0 & \text{otherwise} \end{cases}.$$

Proposition 6.1.1. *A probabilistic empirical model e is non-contextual if and only if there exists a solution to the system of linear equations*

$$\mathbf{M}.\mathbf{x} = \mathbf{v} \tag{6.1}$$

subject to the constraint

$$\mathbf{x} \geq \mathbf{0} \tag{6.2}$$

(i.e. such that $\mathbf{x}[j] \geq 0$ for each component $\mathbf{x}[j]$ of $\mathbf{x}$).

Proof. For a solution $\mathbf{x}$ to this system, we have $\sum_j \mathbf{x}[j]\, \mathbf{t}_j = \mathbf{v}$, where each $\mathbf{t}_j$ is the global assignment t_j represented as a vector. By restricting to the rows corresponding to any maximal context $C \in \mathcal{M}$, we have $\sum_j \mathbf{x}[j] = 1$. It is clear, then, that solutions to this system of equations correspond bijectively to global sections for the model e. □

The constraint ensures that solutions correspond to distributions over $\mathbb{R}^+$, the non-negative reals, and hence that we have a well-defined probability distribution. For possibilistic models we obtain a similar problem over the integers modulo 2, where we substitute $\mathbf{v}$ with the vector $\mathbf{v}_B$ in which all non-zero components are replaced by 1; i.e. if $h : \mathbb{R}^+ \to \mathbb{B}$ is the semiring homomorphism (1.1) then

$$\mathbf{v}_B := h(\mathbf{v}),$$

where h acts component-wise on $\mathbf{v}$.

Proposition 6.1.2. *If the system of linear equations $\mathbf{M}\,\mathbf{x} = \mathbf{v}$ has a solution over $\mathbb{R}^+$, then the system $\mathbf{M}\,\mathbf{x} = \mathbf{v}_B$ has a solution over $\mathbb{B}$.*

Proof. Since $h : \mathbb{R}^+ \to \mathbb{B}$ is a semiring homomorphism, by applying it component-wise to any solution for the system $\mathbf{M}\,\mathbf{x} = \mathbf{v}$ over $\mathbb{R}^+$ one obtains a solution for the system $\mathbf{M}\,\mathbf{x} = \mathbf{v}_B$ over $\mathbb{B}$. □

Corollary 6.1.3. *If a probabilistic empirical model is logically contextual then it is contextual.*

6.2 Quantifying Contextuality

Suppose that an empirical model is contextual; we would like to quantify how contextual it is. This is similar to, but more general than, asking by how much a non-local $(2, 2, 2)$ model violates the CHSH inequality. One reason why this is of interest is that it can help to provide a more robust basis for experimental verification of contextuality by providing some tolerance for inaccuracy of measurements or state preparation. The ideas in this section are based on [1].

Linear programming is an optimisation technique (see e.g. [43]) that can allow us to do precisely this. Linear programs can be expressed in the following canonical form:

$$\begin{array}{ll} \text{maximise} & \mathbf{c}^T \mathbf{x} \\ \text{subject to} & \mathbf{M}\,\mathbf{x} \leq \mathbf{b} \\ \text{and} & \mathbf{x} \geq \mathbf{0} \end{array}$$

where $\mathbf{b}$ and $\mathbf{c}$ are vectors and $\mathbf{M}$ is a matrix with known coefficients; $\mathbf{c}^T \mathbf{x}$ is referred to as the *objective function.* We can use this method to find how close a model comes to being contextual in the sense of finding the maximum $\mathbf{1} \cdot \mathbf{x} = \sum_j \mathbf{x}[j]$ can obtain such that $\mathbf{M}\,\mathbf{x} \leq \mathbf{v}$; i.e.

$$\begin{array}{ll} \text{maximise} & \mathbf{1} \cdot \mathbf{x} \\ \text{subject to} & \mathbf{M}\,\mathbf{x} \leq \mathbf{v} \\ \text{and} & \mathbf{x} \geq \mathbf{0} \end{array} \tag{6.3}$$

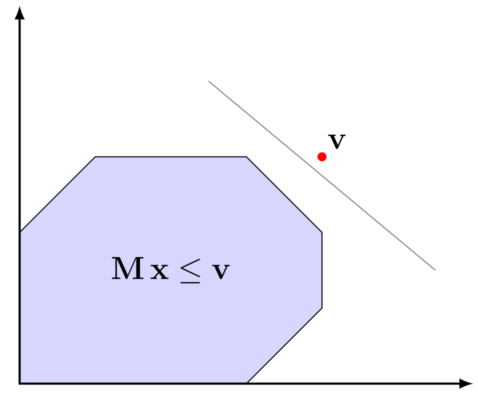
We note that $\mathbf{x} = \mathbf{0}$ always satisfies the constraints, and so the linear program is always feasible. A two-dimensional analogy is shown in figure 6.1.

Proposition 6.2.1. *If $\mathbf{x}^*$ is an optimal solution to the linear program (6.3), then the following statements hold.*

1. $\mathbf{1} \cdot \mathbf{x}^* \in [0, 1]$.
2. $\mathbf{1} \cdot \mathbf{x}^* = 1$ *if and only if e is non-contextual.*
3. $\mathbf{1} \cdot \mathbf{x}^* = 0$ *if and only if e is strongly contextual.*

Proof. Since $\mathbf{x} = \mathbf{0}$ always satisfies the constraints, we must have $\mathbf{1} \cdot \mathbf{x}^* \geq 0$. Let $C \in \mathcal{M}$ be any maximal context, and suppose $\mathbf{M}\,\mathbf{x}^* \leq \mathbf{v}$. Summing over the rows indexed by C, we find that $\mathbf{1} \cdot \mathbf{x}^* \leq \mathbf{1} \cdot \mathbf{v}|_C$; but since $\mathbf{v}|_C$ is a distribution, it follows that $\mathbf{1} \cdot \mathbf{x}^* \leq 1$. This proves the first statement.

Figure 6.1: A two-dimensional analogue of the linear program for contextuality, with the image of a level set of the objective function in grey.



For the second statement, we show that

$$\mathbf{M}\,\mathbf{x}^* = \mathbf{v} \quad \Leftrightarrow \quad \mathbf{1} \cdot \mathbf{x}^* = 1.$$

Suppose $\mathbf{M}\,\mathbf{x}^* = \mathbf{v}$. Again, by considering only the rows corresponding to any maximal context, it can be seen that $\mathbf{1} \cdot \mathbf{x}^* = \mathbf{1} \cdot \mathbf{v}|_C = 1$, similarly to before. For the converse, suppose that $\mathbf{M}\,\mathbf{x}^* < \mathbf{v}$. Then there exists some maximal context C and some assignment $s_i \in \mathcal{E}(C)$ such that $\mathbf{M}\,\mathbf{x}^*[i] < \mathbf{v}[i]$; but then $\mathbf{1} \cdot \mathbf{x}^* < \mathbf{1} \cdot \mathbf{v}|_C = 1$.

For the final statement: if $\mathbf{1} \cdot \mathbf{x}^* > 0$ then $\mathbf{x}^*[j] > 0$ for some j; but then $t_j \in S_e(X)$, so the model is not strongly contextual. For the converse, suppose the model is not strongly contextual. Then there exists some $t_j \in S_e(X)$ and we can define $\epsilon := \min_{C \in \mathcal{M}} \mathbf{v}[t_j|_C]$, which must be positive. The vector $\mathbf{x}$ defined by

$$\mathbf{x}[k] = \begin{cases} \epsilon & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases}$$

satisfies the constraints, and $\mathbf{1} \cdot \mathbf{x} = \epsilon > 0$. Therefore $\mathbf{1} \cdot \mathbf{x}^* > 0$. $\square$

So the linear programming method provides a fully general measure of the contextuality of any empirical model. Since linear programming has polynomial time complexity, it even seems that this might provide an efficient algorithm for deciding contextuality. However, this is not the case since the incidence matrix has $|O^X|$ columns and grows exponentially with respect to the number of measurements.

We mention briefly how these methods can be related to a form of Bell inequality. Using the duality principle of linear programming [43], we can re-cast (6.3) as the following dual program.

$$\begin{array}{ll} \text{minimise} & \mathbf{y} \cdot \mathbf{v} \\ \text{subject to} & \mathbf{M}^T \mathbf{y} \geq \mathbf{1} \\ \text{and} & \mathbf{y} \geq \mathbf{0} \end{array} \tag{6.4}$$

The weak duality theorem tells us that any feasible solution to the dual problem places a bound on the objective function of the original: if $\mathbf{y}$ is a feasible solution to (6.4) and $\mathbf{x}$ is a feasible solution to (6.3) then

$$\mathbf{1} \cdot \mathbf{x} \leq \mathbf{y} \cdot \mathbf{v}.$$

The strong duality theorem tells us that if $\mathbf{x}^*$ is an optimal solution to (6.3) and $\mathbf{y}^*$ is an optimal solution to (6.4) then

$$\mathbf{1} \cdot \mathbf{x}^* = \mathbf{y}^* \cdot \mathbf{v}.$$

If a model is contextual, then we have a vector $\mathbf{y}^*$ such that $\mathbf{M}^T \mathbf{y}^* \geq \mathbf{1}$ and $\mathbf{y}^* \cdot \mathbf{v} = \mathbf{1} \cdot \mathbf{x}^* < \mathbf{1}$. If we set $\mathbf{z}^* := \mathbf{y}^* - \mathbf{1}$, then

$$\mathbf{M}^T \mathbf{z}^* \geq \mathbf{0}, \qquad \mathbf{z}^* \cdot \mathbf{v} < \mathbf{0},$$

and $\mathbf{z}^*$ defines a separating hyperplane which witnesses the fact that $\mathbf{v}$ is not in the cone generated by the non-contextual polytope, since it makes an angle less than $\pi/2$ with each local deterministic model and an angle greater than $\pi/2$ with the model $\mathbf{v}$ (the existence of this hyperplane is guaranteed by the Farkas Lemma). Moreover, since $\mathbf{y}^*$ is an optimal solution, it provides a tight bound. This can be thought of as a generalised form of Bell inequality.

6.3 *Mathematica* Package

Computational tools in the form of a *Mathematica* package [80] (see figure 6.2) have been developed for:

1. calculating quantum empirical models from any state and any sets of compatible measurements;

2. calculating the incidence matrix for any measurement scenario;

3. quantifying the degree of contextuality of any empirical model using the linear programming method of section 6.2.

We stress that these tools are completely general: they can be applied to any pure or mixed quantum state in any Hilbert space and to any sets of compatible observables in that space, including Bell scenarios as a special case.

Figure 6.2: A screenshot of the *Mathematica* package.

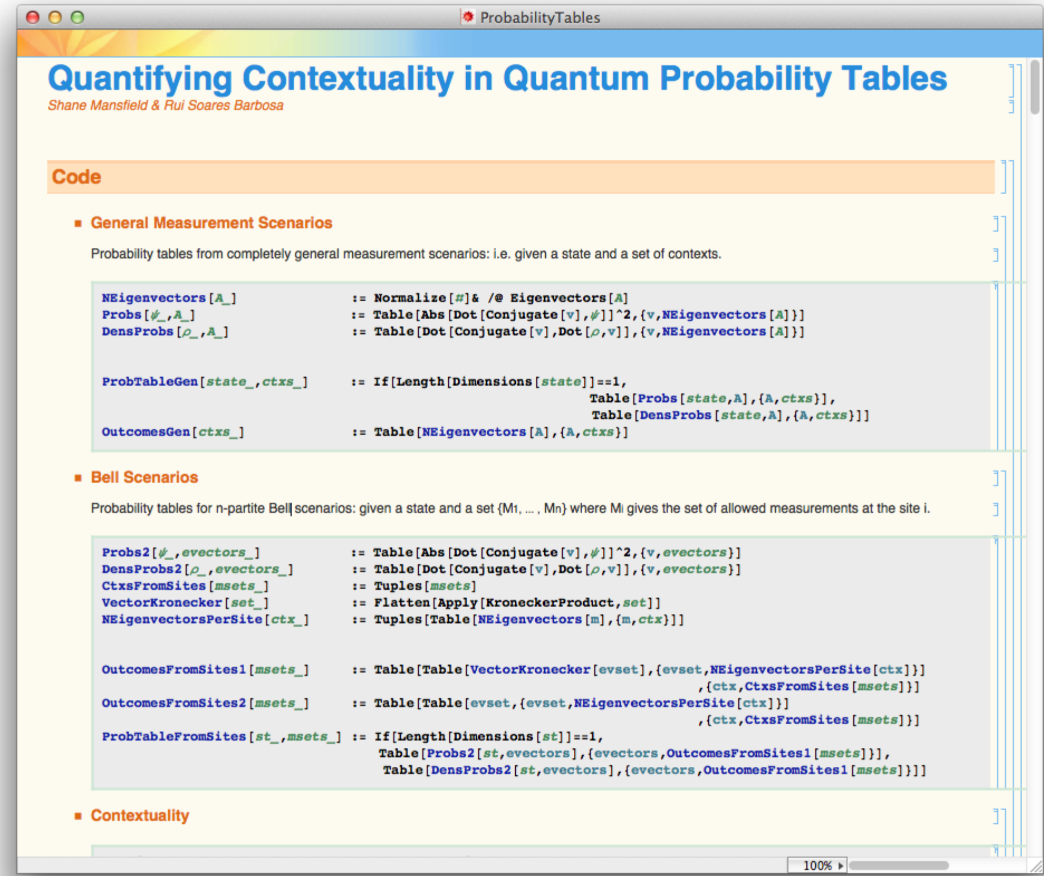
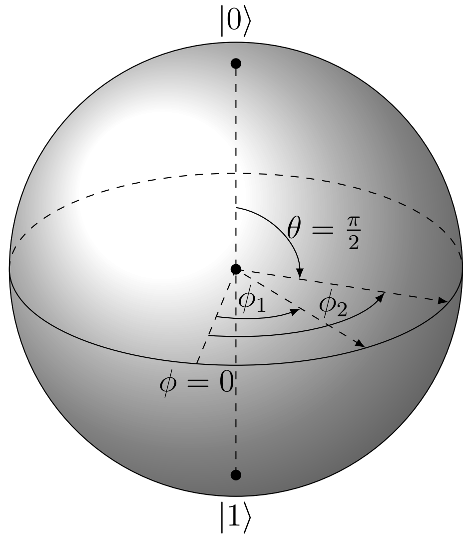


Figure 6.3: Equatorial measurements at ϕ_1 and ϕ_2 on the Bloch sphere.



Equatorial Measurements on $|\phi^+\rangle$

As an example of how the package can be used, we consider a family of empirical models that can be obtained by considering local measurements on the two-qubit state

$$|\phi^+\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle) . \tag{6.5}$$

Recall that projective measurements on a qubit can equivalently be represented by a point on the Bloch sphere. Suppose that we allow the same two local measurements on each qubit, and that these are equatorial on the Bloch sphere (figure 6.3). One such model is the Bell-CHSH model from chapter 1, which is obtained when

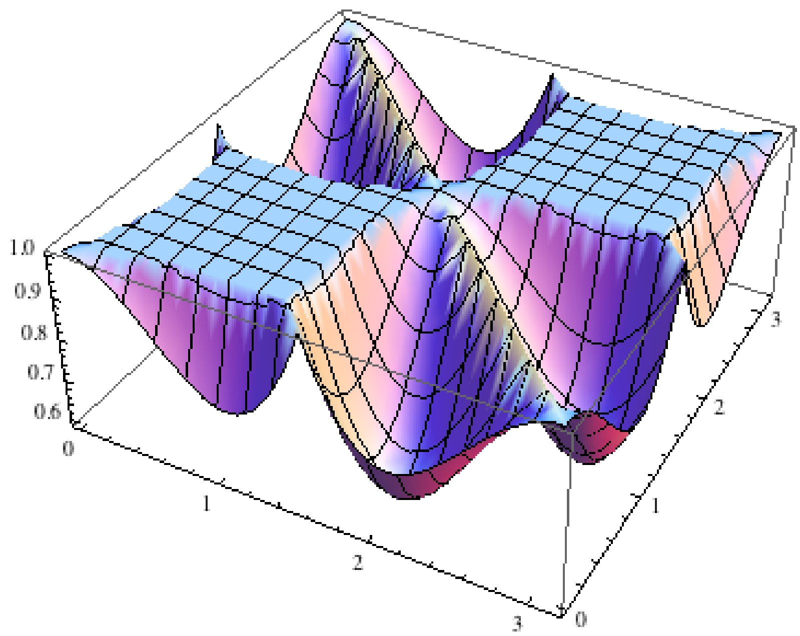
$$(\phi_1, \phi_2) = (0, \pi/3).$$

We can use the package to plot the degree of contextuality of the resulting models versus ϕ_1 and ϕ_2 (figure 6.4). It is interesting to note that the Bell-CHSH model does not achieve the maximum degree of contextuality. The minima of the plot (which correspond to maximum contextuality) occur when

$$(\phi_1, \phi_2) \in \left\{ \left(\frac{\pi}{8}, \frac{5\pi}{8} \right), \left(\frac{7\pi}{8}, \frac{3\pi}{8} \right) \right\}$$

and vice versa. All of the corresponding empirical models take the form of the following table

Figure 6.4: Contextuality of empirical models obtained with equatorial measurements at ϕ_1 and ϕ_2 on each qubit of $|\phi^+\rangle$.



	00	01	10	11
A B	p	$(1/2-p)$	$(1/2-p)$	p
A B'	$(1/2-p)$	p	p	$(1/2-p)$
A' B	$(1/2-p)$	p	p	$(1/2-p)$
A' B'	$(1/2-p)$	p	p	$(1/2-p)$

where

$$p = \frac{\sqrt{2}+2}{8}.$$

These can easily be shown to achieve the Tsirelson violation of the CHSH inequality. Note that none of these models are strongly contextual: this is consistent with theorem 2.6.5, and provided one motivation for attempting to find a general proof of that result.

It may seem surprising at first that the empirical models are not constant with respect to the relative angle $(\phi_2 - \phi_1)$ between measurements; a fact which is apparent from figure 6.4. For example, the empirical model obtained when $(\phi_1, \phi_2) = (0, \pi/4)$ is local, but if these values are shifted by $\pi/8$ the resulting model achieves the maximum violation of the CHSH inequality. Nevertheless, this must be the case since a rotation by ϕ around the Z-axis for each of the qubits is described by

$$\begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix} \otimes \begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix} = \begin{pmatrix} e^{-i\phi} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\phi} \end{pmatrix} \tag{6.6}$$

and thus introduces a relative phase of 2ϕ between the terms in $|\phi^+\rangle$ (6.5).

Equatorial Measurements on GHZ(n) States

We can consider similar families of models for the GHZ(n) states (2.9), where again we allow the same two local measurements on each qubit and assume that these are equatorial on the Bloch sphere. For GHZ(3) and GHZ(4) we obtain the plots shown in figure 6.5. The minima of the plot for GHZ(3) reach 0, indicating strong contextuality, and occur when

$$(\phi_1, \phi_2) \in \left\{ \left(\frac{\pi}{2}, 0\right), \left(\frac{2\pi}{3}, \frac{\pi}{6}\right), \left(\frac{5\pi}{6}, \frac{\pi}{3}\right) \right\} \tag{6.7}$$

and vice versa. Of course, $(\phi_1, \phi_2) = (\pi/2, 0)$ corresponds to the GHZ(3) model described in section 2.7. The empirical models corresponding to other minima are identical up to re-labelling, so these provide alternative sets of measurements that can be made on the GHZ state that still lead to the familiar parity argument for non-locality [87]. The situation is similar for the GHZ(4) state, in which minima of 0 are seen to occur at

$$(\phi_1, \phi_2) \in \left\{ \left(\frac{\pi}{2}, 0\right), \left(\frac{5\pi}{8}, \frac{\pi}{8}\right), \left(\frac{3\pi}{4}, \frac{\pi}{4}\right), \left(\frac{7\pi}{8}, \frac{3\pi}{8}\right) \right\}. \tag{6.8}$$

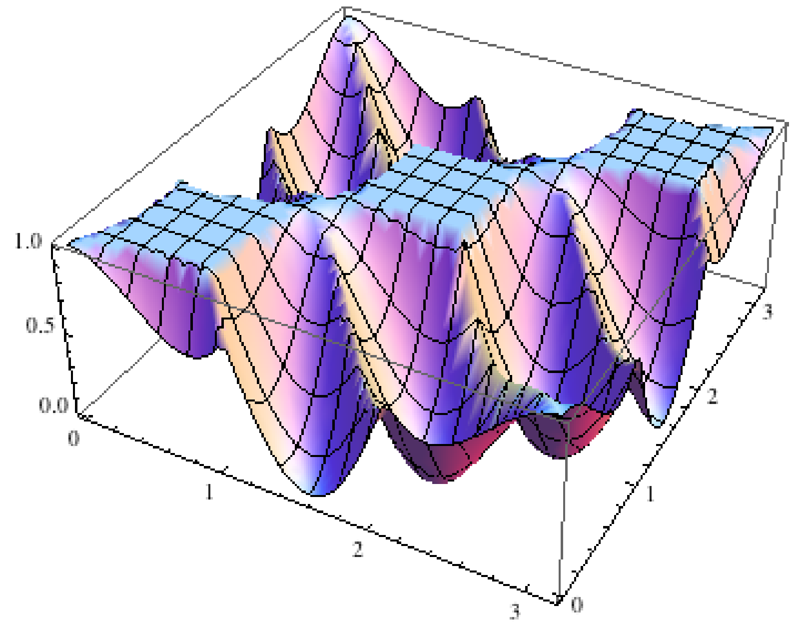
We can see a pattern beginning to emerge in (6.7) and (6.8), which leads to the following proposition.

Proposition 6.3.1. *Equatorial measurements at*

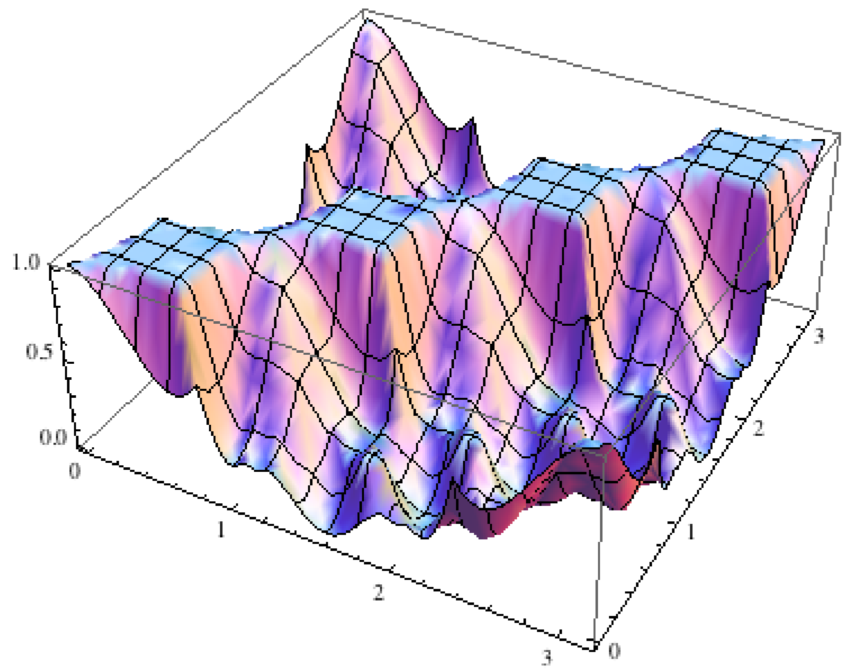
$$(\phi_1, \phi_2) \in \left\{ \left(\frac{(n+k)\,\pi}{2n}, \frac{k\,\pi}{2n}\right) \,\middle|\, 0 \leq k < n \right\}$$

on each qubit of a GHZ(n) state give rise to the GHZ(n) model of section 2.7, and in particular are strongly contextual.

Figure 6.5: Contextuality of empirical models obtained with equatorial measurements at ϕ_1 and ϕ_2 on each qubit of: (a) the GHZ state; (b) the GHZ(4) state.



(a)



(b)

Proof. First, we know that this holds for $k = 0$, since in that case we simply have Pauli X and Y measurements, which were the measurements prescribed for obtaining the GHZ(n) model in section 2.7. For $0 < k < n$, we can rotate each qubit by the phase $\overline{\phi} = k\,\pi/n$, so that we continue to deal with X and Y measurements. It is necessary, however, to take account of the relative phase introduced by this operation on the overall state. By generalising (6.6) it is clear that the state after rotations will be

$$\left|\mathrm{GHZ}(n, \overline{\phi})\right\rangle = \frac{1}{\sqrt{n}} \Big(|0 \cdots 0\rangle + e^{i2n\overline{\phi}} \, |1 \cdots 1\rangle \Big) \, .$$

Notice that for the relevant values of $\overline{\phi}$ the relative phase vanishes and we're left with the state $|\mathrm{GHZ}(n)\rangle$ from (2.9). Then the probabilities of the various outcomes can simply be calculated using equation (2.12), as before, and it is clear that we must obtain the strongly contextual GHZ(n) models described in section 2.7. □

6.4 Negative Probabilities & No-Signalling

In this final section, we prove that an empirical model on any Bell scenario has a local ontological realisation with negative probabilities if and only if it is no-signalling. This built on a result from an earlier version[1] of [4], which proved the equivalence for $(n, 2, 2)$ Bell scenarios. It was later generalised to arbitrary measurement scenarios without the restriction that they be of the Bell form in [4]. It is a remarkable result in that, while probability distributions on local ontological models allow us to generate the local polytope of empirical models, it shows that simply by allowing for negative probabilities we can generate the entire no-signalling polytope. The earlier results in particular were guided by computational exploration of the structure and ranks of incidence matrices.

Negative probability realisations correspond to solutions of the system of equations $\mathbf{M}\,\mathbf{x} = \mathbf{v}$ over $\mathbb{R}$, without constraints. We have seen in proposition 6.1.2 that there exists a semiring homomorphism $h : \mathbb{R}^+ \to \mathbb{B}$ by which any solution over $\mathbb{R}^+$ can be transformed into a solution over $\mathbb{B}$. However, there can be no such homomorphism $h : \mathbb{R} \to \mathbb{B}$ from the reals: if this were the case, we would have

$$0 = h(0) = h(1 - 1) = h(1) \vee h(-1) = 1 \vee h(-1) = 1.$$

The result will be proved inductively, and it is useful to define an inductive enumeration of the local assignments of an (n, k, l) Bell model. We may assume any

[1] Available online at http://arxiv.org/abs/1102.0264v5

enumeration of measurements and outcomes, and let s_{ij} denote the assignment of the outcome j to the measurement i. For a $(1,k,l)$ model we use the enumeration

$$S_1 = (s_{11}, \dots, s_{1l}, \qquad \dots \qquad , s_{k1}, \dots, s_{kl}).$$

and inductively define

$$S_n = (s_{11} \cdot S_{n-1}, \dots, s_{1l} \cdot S_{n-1}, \qquad \dots \qquad , s_{k1} \cdot S_{n-1}, \dots, s_{kl} \cdot S_{n-1})$$

for any (n,k,l) model such that $n > 1$. A vector $\mathbf{v}$ written in this enumeration can be decomposed into blocks,

$$\begin{aligned} \mathbf{v} &= (\mathbf{v}_1, \quad \dots \quad , \mathbf{v}_k) \\ &= (\mathbf{v}_{11}, \dots, \mathbf{v}_{1l}, \quad \dots \quad , \mathbf{v}_{k1}, \dots, \mathbf{v}_{kl}), \end{aligned}$$

where $\mathbf{v}_{ij} := s_{ij} \cdot S_{n-1}$.

Proposition 6.4.1. *For any no-signalling (n,k,l) model, the sums of the probabilities in each block $\mathbf{v}_i$ are constant.*

Proof. Let $S_{n-1} = (t_1, \dots, t_m)$, let $i, i' \in \{1, \dots, k\}$, and let σ_i and $\sigma_{i'}$ denote the sums of the the probabilities over the blocks $\mathbf{v}_i$ and $\mathbf{v}_{i'}$, respectively. By no-signalling, it follows that

$$\sum_{j=1}^{l} s_{ij} \cdot t_p = \sum_{j=1}^{l} s_{i'j} \cdot t_p \tag{6.9}$$

for all $1 \leq p \leq m$, since the choice of measurement i or i' at site n should not alter the probability of the assignment t_p at the other sites. Then, by (6.9),

$$\begin{aligned} \sigma_i &= \sum_{p=1}^{m} \sum_{j=1}^{l} s_{ij} \cdot t_p \\ &= \sum_{p=1}^{m} \sum_{j=1}^{l} s_{i'j} \cdot t_p = \sigma_{i'}. \end{aligned}$$

In the case that $n = 1$, we have $\sigma_i = \sigma_{i'} = 1$ since then each block is simply a probability distribution. □

Next, we consider the form of the incidence matrices with respect to the inductive enumeration. These can be defined inductively on k and n. Let $\mathbf{M}_{(n,k,l)}$ denote the (n,k,l) incidence matrix. For an arbitrary enumeration of global sections, each $\mathbf{M}_{(1,1,l)}$ is simply the $l \times l$ identity matrix after some permutation of columns. We can choose our enumeration of global sections such that $\mathbf{M}_{(1,1,l)} = \mathbf{I}_l$, the $l \times l$ identity matrix, and such that

$$\mathbf{M}_{(1,k,l)} = \begin{bmatrix} & \mathbf{I}_l \otimes \mathbf{1}_l^T & \\ \mathbf{M}_{(1,k-1,l)} & \cdots & \mathbf{M}_{(1,k-1,l)} \end{bmatrix}$$

for $k > 1$, where $\mathbf{1}_l^T$ is the row matrix whose l entries are all 1's. For example, the $(1,3,2)$ incidence matrix is

$$\mathbf{M}_{(1,3,2)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Now, due to the inductive enumeration of local assignments, for $n > 1$ we must have

$$\mathbf{M}_{(n,k,l)} = \mathbf{M}_{(1,k,l)} \otimes \mathbf{M}_{(n-1,k,l)}. \tag{6.10}$$

Proposition 6.4.2. *The rank of any (n,k,l) incidence matrix is given by*

$$\mathsf{rank}\left(\mathbf{M}_{(n,k,l)}\right) = \left(k\,(l-1)+1\right)^n. \tag{6.11}$$

Proof. A $(1,k,l)$ incidence matrix can be divided into k blocks of l rows. Notice that the rows in each block are linearly independent, and that the sum of the rows in each block is $\mathbf{1}_{l^2}^T = (1,\ldots,1)$. This means that given any two blocks we can write any of the rows as a linear combination of all the others. The rank of a matrix is equal to the number of linearly independent rows. Therefore, we have

$$\mathsf{rank}\left(\mathbf{M}_{(1,1,l)}\right) = \mathsf{rank}\left(\mathbf{I}_l\right) = l$$

for all $l > 0$, and

$$\begin{aligned} \mathsf{rank}\left(\mathbf{M}_{(1,k,l)}\right) &= \mathsf{rank}\left(\mathbf{M}_{(1,k-1,l)}\right) + (l-1) \\ &= k\,(l-1)+1 \end{aligned}$$

for all $k > 1$, since each increment of k introduces $(l-1)$ new linearly independent rows. Finally, from (6.10) and by the fact that

$$\mathsf{rank}(A \otimes B) = \mathsf{rank}(A)\,\mathsf{rank}(B)$$

for any matrices A and B, it follows that the rank of any (n,k,l) incidence matrix is given by (6.11). $\square$

Theorem 6.4.3. *An empirical model on an (n, k, l) Bell scenario can be realised by a local ontological model with negative probabilities if and only if it satisfies no-signalling.*

Proof. Realisability by a local ontological model with negative probabilities corresponds precisely to the existence of a solution to the system of linear equations $\mathbf{M}\,\mathbf{x} = \mathbf{v}$ over $\mathbb{R}$. It is a standard result of linear algebra that such a system has a solution if and only if

$$\mathsf{rank}\,(\mathbf{M}) = \mathsf{rank}\,([\mathbf{M} \mid \mathbf{v}])\,,$$

where $[\mathbf{M} \mid \mathbf{v}]$ is the augmented matrix; but this follows from propositions 6.4.1 and 6.4.2 since the rows of $\mathbf{v}$ have the same linear dependencies as $\mathbf{M}$. □

6.5 Discussion

We have presented a number of computational tools which have been implemented as a *Mathematica* package and which form a useful complement to the sheaf-theoretic approach in general. Indeed, the package was used to calculate many of the probability tables found throughout this dissertation, and as a means of testing results when they were at the conjectural stage. Examples include theorems 2.6.5 and 6.4.3, as well as some of the results of [9]. It is hoped that the tools and methods described in this chapter can continue to play an important role in guiding future results and developments within the sheaf-theoretic research programme. For example, the tools can be especially useful in attempting to classify the non-locality of states, which will be a goal of future work.

An important feature of the tools is that they are applicable to empirical models on any measurement scenario; not just to Bell scenarios. The linear programming approach to finding the degree of contextuality of a model, for example, works in full generality and even provides a means of finding a general analogue of a Bell inequality which witnesses contextuality given any contextual or non-local model. This is especially relevant for experimental verification of contextuality, where it can be used to ensure robustness of contextuality with respect to inaccuracies in state preparation and measurements in scenarios where the CHSH or other inequalities are not applicable. It is also an interesting development in itself, which is worthy of further investigation. For example, one might consider how this relates to the semi-definite programming approach of Navascues, Pironio & Acín [88] to characterising the set of quantum correlations in the $(2, 2, 2)$ scenario.

Theorem 6.4.3 also deserves a further mention. The idea of negative probabilities in quantum mechanics has a long history, which is briefly outlined in [4]. This result gives a perspective on their role. Feynman once said [49, p. 480],

> The only difference between a probabilistic classical world and the equations of the quantum world is that somehow or other it appears as if the probabilities would have to go negative ...

In fact, theorem 6.4.3 and its subsequent generalisation to arbitrary measurement covers in [4] show that, in a certain sense, allowing probabilities to 'go negative' is the difference between a probabilistic classical world and the no-signalling world.

Conclusion

We have aimed to develop a deeper, more structural understanding of non-locality and contextuality, and to this end have presented a framework for logical non-locality, which sits neatly within the unified sheaf-theoretic approach to non-locality and contextuality, and have also presented several developments of the more general framework. These are underpinned by robust mathematical theory and offer clarity and new perspectives on a variety of issues. Indeed, this programme of research has already been seen to lead to many interesting results. We have also seen the value of the approach in its ability to highlight connections with diverse fields in which similar structures arise, which allows for the cross-fertilisation of results and ideas. Of course, there remain many intriguing open questions and possibilities for further developments and applications.

The completeness theorems of chapter 2 prove that Hardy non-locality completely characterises logical non-locality in all $(2, 2, l)$ and $(2, k, 2)$ scenarios, and have led to numerous applications. Polynomial algorithms can be given for deciding non-locality in these scenarios, even though it has been shown that in general the problem is NP-hard [10]. A constructive proof that the PR boxes are the only strongly contextual $(2, 2, 2)$ models was found. The first full proof that Bell states, despite being maximally entangled, are the only entangled two-qubit states that are not logically non-local was also given. This is surprising in that it singles out the Bell states as being anomalous in terms of non-locality. The proof of the result is interesting in that it led to the discovery of a new family of non-quantum empirical models lying within the Tsirelson bound but which can have an arbitrarily small violation of the CHSH inequality. Another remarkable result that emerges within the logical framework is that it can be proved that the GHZ experiment [57] should witness Hardy non-locality with certainty. It is often the case in the literature that the probability of witnessing a Hardy paradox is used as a measure of the quality of non-locality, and this represents a striking improvement on the previous best probability of ≈ 0.4 [37]. The possibility of further applications remains to be explored. It will also be interesting to see whether there can be any transfer of ideas between this framework and relational database theory via the correspondence established in [2].

We have found the first application of cohomology as a tool for studying non-locality and contextuality, finding cohomological witnesses corresponding to many of the classic no-go results and completely characterising contextuality for large families of Kochen-Specker models. However, there is room for improvement, especially since it is not yet possible to completely characterise non-locality and contextuality in all models. The examples considered often relied on brute force enumeration, and one obvious possibility is to try to use the machinery of homological algebra and exact sequences to obtain results that are more general, but that are also more conceptually illuminating. Other refinements such as considering higher order cohomology groups might also be used to achieve a finer invariant.

A novel connection between contextuality of Kochen-Specker models and the existence of a perfect matching in the dual hypergraph of the measurement scenario also appeared in relation to this material, and was crucial to the proofs of the general results about completeness of the cohomological characterisation for certain classes of models. This is a connection that has only begun to be explored, but which has already been seen to lead to new insights. In fact, one of these general results follows directly from a theorem proved in the setting of k-uniform hypergraphs. It was also possible to use results proved in that setting to find an efficient algorithm for deciding strong contextuality for Kochen-Specker models with constant degree $d = 2$, and to show that for $d \geq 3$ the problem is NP-complete. As far as we are aware, these results are new to the foundations of quantum mechanics. They complement the results of Pitowsky [92], Abramsky, Gottlob & Kolaitis [10] and those of chapter 2 for Bell models. Furthermore, it was seen that Tutte's theorem can provide a necessary and sufficient condition for contextuality in $d = 2$ Kochen-Specker models.

The refinement of extendability introduced in chapter 4 was seen to be useful for characterising contextuality in sub-models of an empirical model. Some open questions remain: we would like to know if there is some analogue of Vorob'ev's theorem [103] in this case, which could potentially lead to applications to macroscopic realism similar to [100], or whether there exist other classes of 'good' models that can be guaranteed to extend to $\mathcal{P}_n X$ and thereby to have equivalent Bell models. With regard to the construction of Bell models, an interesting connection was found between the contextual triangle and the PR box, which merits further investigation. One thought arising from this, given that the PR box has been considered as a candidate unit of non-locality [19], is whether it might be shown that there exists some set of 'irreducible' contextual models including the triangle that all other contextual models must reduce to. The construction also suggests the possibility of proposing Bell tests that correspond to contextuality proofs such as that of Peres & Mermin [85]. However, the issue of quantum realisability of constructed models will need to be considered first.

Some progress has been made on bringing recent developments on the reality of the wavefunction within the scope of the logical and sheaf-theoretic methods. One result of this work is a generalised reformulation of the Harrigan-Spekkens criterion for the reality of the wavefunction [63], which among other things has the advantage of avoiding measure theoretic technicalities. Applying this to the observable properties of a system led to a new characterisation of non-locality and contextuality in these terms. A more careful treatment of preparation independence, which was a key assumption of the PBR theorem [94], also leads to some interesting questions. It was shown that preparation independence is analogous to Bell non-locality, and if it is weakened to 'no-preparation-signalling', an assumption analogous to no-signalling, then the PBR argument no longer holds. This amounts to introducing non-local correlations on the joint ontic state, which at least is consistent with the Bell and Kochen-Specker theorems. An important question then is whether under this relaxed assumption the result can hold. These considerations also suggest the need to introduce a notion of 'preparation scenario' analogous to a measurement scenarios, in which preparations play the role of measurements and ontic states play the role of outcomes. It seems the sheaf-theoretic approach to ontological theories described in chapter 1 might be adapted to allow for a unified treatment of both kinds of scenario, where ontic states are identified with the Gleason states of corollary 1.7.1.

A number of computational tools have been implemented as a *Mathematica* package forming a useful complement to the sheaf-theoretic approach in general. This played an important guiding role for the results of [9] and in proving that realisability by a local ontological model with negative probabilities is equivalent to no-signalling. This is an interesting result, which shows that, in a certain sense, the only difference between a classical probabilistic world and the no-signalling world is that probabilities are allowed to be negative. It is hoped that the linear programming approach to quantifying contextuality and finding logical Bell inequalities which applies to any measurement scenario can find many applications due to its generality, and that these tools can play an important role in leading to future developments and applications within the research programme. In particular, these can be useful tool in attempting to classify the non-locality of states.

References

[1] S. Abramsky. Quantifying contextuality. Private communication, 2012.

[2] S. Abramsky. Relational databases and Bell's theorem. *arXiv:1208.6416*, 2012. ArXiv preprint.

[3] S. Abramsky. Relational hidden variables and non-locality. *Studia Logica*, 101(2):411–452, 2013.

[4] S. Abramsky and A. Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13(11):113036, 2011.

[5] S. Abramsky and A. Brandenburger. Axiomatizing physical theories in a sheaf-theoretic framework. Private communication, 2013.

[6] S. Abramsky, A. Brandenburger, and A. Savochkin. No-signalling is equivalent to free choice of measurements. In *Proceedings of Quantum Physics & Logic*, 2013.

[7] S. Abramsky and B. Coecke. A categorical semantics of quantum protocols. In *Logic in Computer Science, 2004. Proceedings of the 19th Annual IEEE Symposium on*, pages 415–425. IEEE, 2004.

[8] S. Abramsky and B. Coecke. Categorical quantum mechanics. *Handbook of quantum logic and quantum structures: quantum logic*, pages 261–324, 2009.

[9] S. Abramsky and C. Constantin. A classification of multipartite states by degree of non-locality. In *Proceedings of Quantum Physics & Logic*, 2013.

[10] S. Abramsky, G. Gottlob, and P. Kolaitis. Robust constraint satisfaction and local hidden variables in quantum mechanics. In *Proceedings of the International Joint Conference in Artificial Intelligence (IJCAI)*, 2013.

[11] S. Abramsky and L. Hardy. Logical Bell inequalities. *Physical Review A*, 85(6):062114, 2012.

[12] S. Abramsky, S. Mansfield, and R. Soares Barbosa. The cohomology of non-locality and contextuality. In *Electronic Proceedings in Theoretical Computer Science*, volume 95, pages 1–14, 2012.

[13] S. Abramsky and S. Ying. Private communication, 2013.

[14] A. Acín, R. Augusiak, D. Cavalcanti, C. Hadley, J. Korbicz, M. Lewenstein, L. Masanes, and M. Piani. Unified framework for correlations in terms of local quantum observables. *Physical Review Letters*, 104(14):140404, 2010.

[15] J. Allcock, N. Brunner, M. Pawlowski, and V. Scarani. Recovering part of the quantum boundary from information causality. *arXiv:0906.3464*, 2009. ArXiv preprint.

[16] G. Bacciagaluppi, A. Valentini, and M. Jähnert. Quantum theory at the crossroads: Reconsidering the 1927 Solvay Conference. *Physics Today*, 63(10):53, 2010.

[17] M. Barbieri, C. Cinelli, F. De Martini, and P. Mataloni. Test of quantum nonlocality by full collection of polarization entangled photon pairs. *The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics*, 32(2):261–267, 2005.

[18] J. Barrett. Information processing in generalized probabilistic theories. *Physical Review A*, 75(3):032304, 2007.

[19] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Nonlocal correlations as an information-theoretic resource. *Physical Review A*, 71(2):22101, 2005.

[20] J. Barrett and S. Pironio. Popescu-Rohrlich correlations as a unit of nonlocality. *Physical Review Letters*, 95(14):140401, 2005.

[21] H. Bartosik, J. Klepp, C. Schmitzer, S. Sponar, A. Cabello, H. Rauch, and Y. Hasegawa. Experimental test of quantum contextuality in neutron interferometry. *Physical Review Letters*, 103(4):40403, 2009.

[22] J. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(3):195–200, 1964.

[23] J. S. Bell. On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics*, 38(3):447–452, 1966.

[24] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics: Collected papers on quantum philosophy.* Cambridge University Press, 1987.

[25] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70(13):1895, 1993.

[26] C. H. Bennett, G. Brassard, et al. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, volume 175(0). New York, 1984.

[27] D. Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables. I. *Physical Review*, 85(2):166, 1952.

[28] D. Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables. II. *Physical Review*, 85(2):180, 1952.

[29] D. Boschi, S. Branca, F. De Martini, and L. Hardy. Ladder proof of nonlocality without inequalities: Theoretical and experimental results. *Physical Review Letters*, 79(15):2755–2758, 1997.

[30] A. Brandenburger and H. J. Keisler. What does a hidden variable look like, 2011.

[31] A. Brandenburger and H. J. Keisler. Use of a canonical hidden-variable space in quantum mechanics. In *Computation, Logic, Games, and Quantum Foundations. The Many Facets of Samson Abramsky*, pages 1–6. Springer, 2013.

[32] A. Brandenburger and N. Yanofsky. A classification of hidden-variable properties. *Journal of Physics A: Mathematical and Theoretical*, 41:425302, 2008.

[33] N. Brunner and P. Skrzypczyk. Nonlocality distillation and postquantum theories with trivial communication complexity. *Physical Review Letters*, 102(16):160403, 2009.

[34] A. Cabello, J. Estebaranz, and G. García-Alcaine. Bell-Kochen-Specker theorem: A proof with 18 vectors. *Physics Letters A*, 212(4):183–187, 1996.

[35] P. Cameron and A. Treglown. Private communication, 2013.

[36] J. L. Cereceda. Hardy's nonlocality for generalized n-particle GHZ states. *Physics Letters A*, 327(5-6):433 – 437, 2004.

[37] J.-L. Chen, A. Cabello, Z.-P. Xu, H.-Y. Su, C. Wu, and L. Kwek. Hardy's paradox for high-dimensional systems: Beyond Hardy's limit. *arXiv:1308.4468v1*, 2013. ArXiv preprint.

[38] G. Chiribella, G. M. DAriano, and P. Perinotti. Probabilistic theories with purification. *Physical Review A*, 81(6):062348, 2010.

[39] S. K. Choudhary, S. Ghosh, G. Kar, S. Kunkri, R. Rahaman, and A. Roy. Hardy's non-locality and generalized non-local theory. *arXiv:0807.4414*, 2008. ArXiv preprint.

[40] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, 1969.

[41] R. Colbeck and R. Renner. Is a system's wave function in one-to-one correspondence with its elements of reality? *Physical Review Letters*, 108(15):150402, 2012.

[42] D. Craig, F. Dowker, J. Henson, S. Major, D. Rideout, and R. D. Sorkin. A Bell inequality analog in quantum measure theory. *Journal of Physics A: Mathematical and Theoretical*, 40(3):501, 2007.

[43] G. B. Dantzig and M. N. Thapa. *Linear programming: 1: Introduction*, volume 1. Springer, 1997.

[44] D. E. Daykin and R. Häggkvist. Degrees giving independent edges in a hypergraph. *Bulletin of the Australian Mathematical Society*, 23(1):103–109, 1981.

[45] N. de Silva. Quantum states are consistent probability distributions. Forthcoming, 2013.

[46] A. Döring. Generalised Gelfand spectra of nonabelian unital C*-algebras ii: Flows and time evolution of quantum systems. *arXiv:1212.4882*, 2012. ArXiv preprint.

[47] J. Edmonds. Paths, trees, and flowers. *Canadian Journal of Mathematics*, 17(3):449–467, 1965.

[48] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47(10):777, 1935.

[49] R. P. Feynman. Simulating physics with computers. *International journal of theoretical physics*, 21(6):467–488, 1982.

[50] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48(5):291–295, 1982.

[51] T. Fritz. Quantum analogues of Hardy's nonlocality paradox. *Foundations of Physics*, 41(9):1493–1501, 2011.

[52] M. R. Garey and D. S. Johnson. Computers and intractability: A guide to the theory of NP-completeness, 1979.

[53] G.-C. Ghirardi, A. Rimini, and T. Weber. A general argument against superluminal transmission through the quantum mechanical measurement process. *Lettere Al Nuovo Cimento (1971–1985)*, 27(10):293–298, 1980.

[54] S. Ghosh, G. Kar, and D. Sarkar. Hardy's nonlocality for entangled states of three spin-½ particles. *Physics Letters A*, 243(5):249–255, 1998.

[55] R. Ghrist and Y. Hiraoka. Applications of sheaf cohomology and exact sequences to network coding, 2011. Preprint.

[56] A. M. Gleason. Measures on the closed subspaces of a Hilbert space. *Journal of Mathematics and Mechanics*, 6(6):885–893, 1957.

[57] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger. Bell's theorem without inequalities. *American Journal of Physics*, 58:1131, 1990.

[58] D. M. Greenberger, M. A. Horne, and A. Zeilinger. Going Beyond Bell's Theorem. In M. Kafatos, editor, *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, pages 69–72. Kluwer, 1989.

[59] L. Hardy. Quantum mechanics, local realistic theories, and Lorentz-invariant realistic theories. *Physical Review Letters*, 68(20):2981–2984, 1992.

[60] L. Hardy. Nonlocality for two particles without inequalities for almost all entangled states. *Physical Review Letters*, 71(11):1665–1668, 1993.

[61] L. Hardy. Reformulating and reconstructing quantum theory. *arXiv:1104.2066*, 2011. ArXiv preprint.

[62] L. Hardy. On the theory of composition in physics. In *Computation, Logic, Games, and Quantum Foundations. The Many Facets of Samson Abramsky*, pages 83–106. Springer, 2013.

[63] N. Harrigan and R. W. Spekkens. Einstein, incompleteness, and the epistemic view of quantum states. *Foundations of Physics*, 40(2):125–157, 2010.

[64] A. Hatcher. *Algebraic Topology.* Cambridge University Press, Cambridge, 2002.

[65] T. Heinosaari, D. Reitzner, and P. Stano. Notes on joint measurability of quantum observables. *Foundations of Physics*, 38(12):1133–1147, 2008.

[66] C. Heunen. Private communication, 2013.

[67] P. Heywood and M. Redhead. Nonlocality and the Kochen-Specker paradox. *Foundations of Physics*, 13:481–499, 1983. 10.1007/BF00729511.

[68] C. Isham and J. Butterfield. Topos perspective on the Kochen-Specker theorem: I. quantum states as generalized valuations. *International Journal of Theoretical Physics*, 37(11):2669–2733, 1998.

[69] A. Kapustin. Is there life beyond quantum mechanics? *arXiv:1303.6917*, 2013. ArXiv preprint.

[70] J. Kennedy. On the empirical foundations of the quantum no-signalling proofs. *Philosophy of Science*, pages 543–560, 1995.

[71] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. Roos. State-independent experimental test of quantum contextuality. *Nature*, 460(7254):494–497, 2009.

[72] A. Klyachko. Coherent states, entanglement, and geometric invariant theory. *arXiv:0206.012*, 2002. ArXiv preprint.

[73] S. Kochen and E. P. Specker. Logical structures arising in quantum theory. In *The Logico-Algebraic Approach to Quantum Mechanics*, pages 263–276. Springer, 1975.

[74] R. Lal. A sheaf-theoretic approach to the contextuality of cluster states. Forthcoming, 2013.

[75] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman. Specker's parable of the overprotective seer: A road to contextuality, nonlocality and complementarity. *Physics Reports*, 506(1-2):1 – 39, 2011.

[76] L. Lovász and M. D. Plummer. *Matching theory.* American Mathematical Society, 1986.

[77] S. Mac Lane and I. Moerdijk. *Sheaves in geometry and logic: A first introduction to topos theory.* Springer, 1992.

[78] S. Mansfield. On the reality of observable properties. *arXiv:1306.3216*, 2013. ArXiv preprint.

[79] S. Mansfield and T. Fritz. Hardy's non-locality paradox and possibilistic conditions for non-locality. *Foundations of Physics*, 42:709–719, 2012. 10.1007/s10701-012-9640-1.

[80] S. Mansfield and R. Soares Barbosa. Quantifying contextuality in quantum probability tables, 2012. *Mathematica* package.

[81] S. Mansfield and R. Soares Barbosa. Construction of Bell models from Kochen-Specker models. In *Proceedings of Quantum Physics & Logic*, 2013.

[82] L. Masanes. Extremal quantum correlations for N parties with two dichotomic observables per site. *arXiv:0512.100*, 2005. ArXiv preprint.

[83] N. Mermin. Quantum mysteries revisited. *American Journal of Physics*, 58(8):731–734, 1990.

[84] N. Mermin. Simple unified form for the major no-hidden-variables theorems. *Physical Review Letters*, 65(27):3373–3376, 1990.

[85] N. Mermin. Hidden variables and the two theorems of John Bell. *Reviews of Modern Physics*, 65(3):803–815, 1993.

[86] N. Mermin. Quantum mysteries refined. *American Journal of Physics*, 62:880, 1994.

[87] N. D. Mermin. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Physical Review Letters*, 65(15):1838–1840, 1990.

[88] M. Navascués, S. Pironio, and A. Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

[89] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, 2009.

[90] A. Peres. Incompatible results of quantum measurements. *Physics Letters A*, 151(3-4):107–108, 1990.

[91] A. Peres. Two simple proofs of the Kochen-Specker theorem. *Journal of Physics A: Mathematical and General*, 24:175, 1991.

[92] I. Pitowsky. Correlation polytopes: their geometry and complexity. *Mathematical Programming*, 50(1):395–414, 1991.

[93] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[94] M. F. Pusey, J. Barrett, and T. Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):476–479, 2012.

[95] R. Ramanathan, A. Soeda, P. Kurzyński, and D. Kaszlikowski. Generalized monogamy of contextual inequalities from the no-disturbance principle. *Physical Review Letters*, 109(5):050404, 2012.

[96] M. Sadrzadeh. Private communication, 2013.

[97] E. Schrödinger. Probability relations between separated systems. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 32, pages 446–452. Cambridge University Press, 1936.

[98] M. Seevinck. E. Specker: "The logic of non-simultaneously decidable propositions"(1960). *arXiv:1103.4537*, 2011. ArXiv preprint.

[99] P. W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings of 35th Annual Symposium on Foundations of Computer Science*, pages 124–134. IEEE, 1994.

[100] R. Soares Barbosa. A structural reason for monogamy and local average macroscopic behaviour. Forthcoming, 2013.

[101] J. D. Trimmer. The present situation in quantum mechanics: A translation of Schrödinger's "cat paradox" paper. In *Proceedings of the American Philosophical Society*, volume 124(5), pages 323–338. JSTOR, 1980.

[102] B. S. Tsirelson (Cirel'son). Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980.

[103] N. N. Vorob'ev. Consistent families of measures and their extensions. *Theory of Probability and its Applications*, VII(2):147–163, 1962. (translated by N. Greenleaf, Russian original published in *Teoriya Veroyatnostei i ee Primeneniya*).

[104] Z. Wang and D. Markham. Nonlocality of symmetric states. *Physical Review Letters*, 108(21):210407, 2012.

Index