Imperial College of Science, Technology and Medicine
Department of Physics

Possibilistic (and other) notions of contextuality

Andrew W. Simmons

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Physics of the University of London and
the Diploma of Imperial College, September 2018

Abstract

Since Bell, Kochen and Specker formulated the first notions of contextuality, a number of frameworks have been created to help make philosophical and mathematical sense of the concept. Two of the most common are the Equivalence-based approach due to Spekkens, and the Sheaf-theoretic framework of Abramsky and Brandenburger. Each of these can support a hierarchy of strengths of contextuality, which can include possibilistic concepts. This thesis will explore the differences between the two approaches, and explore what can be demonstrated by using weaker notions of noncontextuality.

Notions of noncontextuality are important for proving no-go theorems which highlight that many properties that are true of classical mechanics fail to be compatible with quantum predictions. Often, the notions of noncontextuality used are very strong assumptions, and even weaker flavours of noncontextuality can be used to draw the same, or similar conclusions. An important class of such notions are, possibilistic assumptions that only care about whether or not events are possible or impossible, rather than the exact Born rule probabilities that are applied to events. Even these weaker notions are often impossible to resolve with the logical structure and predictions of quantum theory.

Contextuality has been demonstrated to be a resource for many quantum computational operations, and the contextual fraction has been shown to be an important resource. We introduce a metric that can differentiate between different quantum mechanical scenarios with contextual fraction 1, and demonstrate analogues of CHSH and Tsirelson bounds for this quantity. We also investigate the computational complexity of calculating the values of some metrics for nonlocality and contextuality, completing the classification of possibilistic nonlocality scenarios, as well as other metrics of nonlocality analogous to our metric of maximal contextuality.

Acknowledgements

Firstly, I would like to thank my supervisor Terry Rudolph; without his supervision and habit of casually posing me engaging, deep, and ususally difficult questions, this thesis would look very different. It has been a genuine pleasure working together and I feel I have learnt an incredible amount from him.

I would also like to thank EPSRC, the Controlled Quantum Dynamics CDT, and Cambridge Quantum Computing for providing funding during the past few years. I am also greatly indebted to the administrative staff, particularly Miranda Smith and Veena Dhulipala.

I am greatly indebted to all my collaborators for their insights and their interest into my project. Special thanks go to Joel Wallman, Hakop Pashayan and Steve Bartlett, who all put so much work into our shared paper; making it immeasurably better for their efforts.

On a more personal level, I would like to thank Angela Xu for being there for me constantly throughout. I especially thank her for holding my hand through some of the Category theoretic parts of this thesis, and for putting up with endless talk about contextuality over the course of this project.

I would also like to thank my family, especially my parents Alison and David. Without their help and support throughout my life, it goes without saying that I would not be in a position to be able to undertake this kind of project.

Contents

Abstract **3**

Acknowledgements **5**

1 Introduction **11**

1.1 Statement of Originality 11

1.2 Copyright Disclaimer 11

1.3 Motivation and Results 12

2 Flavours of Contextuality **16**

2.1 Historical background 17

2.2 The Ontological Models Formalism 22

2.3 Spekkens, or Equivalence-based, Contextuality 26

2.3.1 Introduction to the equivalence formalism 26

2.3.2 Probabilistic Noncontextuality in the Equivalence Formalism 30

2.3.3 Possibilistic Noncontextuality in the Equivalence Formalism 34

2.4 The Sheaf-theoretic Approach 43

3 Common Ground and Differences Between Approaches **50**

3.1 Unsharp Measurements and Determinism 51

3.2 Kitchen Sink Models 55

3.3 Preparation Noncontextuality 59

3.4 Discussion 66

4 Restrictions on a Noncontextual Reality **69**

4.1 Locality assumptions . . . 70

4.2 Restrictions on noncontextual subtheories . . . 76

4.3 Restrictions on ψ-epistemic models . . . 80

4.3.1 The PBR theorem . . . 80

4.3.2 Restrictions from notions of measurement noncontextuality . . . 83

4.4 Determinism and Counterfactual Independence . . . 90

4.5 Discussion . . . 93

5 Metrics of Contextuality and Nonlocality **98**

5.1 Quantifying Nonlocality and Contextuality . . . 99

5.2 Robustness of Contextuality . . . 100

5.2.1 How maximally contextual is quantum mechanics? . . . 101

5.2.2 Non-Quantum Graphs . . . 111

5.2.3 Applications to quantum verification . . . 116

5.2.4 Notes on the search for small Kochen-Specker sets . . . 120

5.3 Discussion . . . 123

6 The Computational Complexity of Contextuality **126**

6.1 Possibilistic Locality . . . 127

6.1.1 The (2,k,{2,3}) scenario . . . 133

6.2 (2,3)-PossLoc is **NP**-complete . . . 135

6.3 Computing the Robustness of Possibilistic Locality . . . 140

6.4 Discussion . . . 143

7 Conclusion **145**

7.1 Future Work . . . 147

Bibliography **148**

List of Figures

2.1 Punnett square of noncontextuality assumptions 38

2.2 Probability and Possibility tables for a CHSH experiment 44

2.3 The presheaf $\mathcal{D}_R\mathcal{E}$ is defined by composition of the functor $\mathcal{D}_R$ with the event sheaf $\mathcal{E}$. 47

2.4 An example of a possibilistically local table of possibilities 49

3.1 Quantum mechanical probabilities for a proof of preparation contextuality 62

3.2 GPT probabilities for a proof of preparation contextuality 64

3.3 States to which one subsystem can be steered for Theorem 2 65

3.4 A Venn diagram showing the epistemic overlap regions for the different preparations alongside banned areas. 66

4.1 A single point noncontextually assigned a 1 leads to a great circle of points noncontextually assigned a 0. 77

5.1 An illustration of the calcuation of the figure of merit for Peres's 33 ray proof. . 102

5.2 An illustration at how our method of assigning contextual vectors splits Hilbert space into distinct regions. 107

5.3 Contextuality Robustness of different quantum and superquantum constructions 110

5.4 A chart showing the value of the bound on $q(\mathcal{G})$ for quantum-mechanically accessible $\mathcal{G}$ as a function of the Hilbert space dimension d. 110

5.5 The Peres-Mermin magic square, and an illustration of how it can be used for nonlocality games. 118

5.6 Two illustrative cap placements for the 33 ray proof of contextuality due to Peres. 122

6.1 Tables of Probabilities and Possibilities for the CHSH experiment. 128

6.2 A simple signalling situation. 129

6.3 A table of possibilities that does not exhibit nonlocality; it can be decomposed into a sum of deterministic grids that obey no-signalling. 130

6.4 Possibility tables for the Hardy paradox. 130

6.5 Points on an equatorial slice of the Bloch sphere to which Alice's system can be steered. 133

6.6 A generalised Hardy paradox and its minimal possibilistic structure. 134

Chapter 1

Introduction

1.1 Statement of Originality

I declare that the content of this thesis is entirely my own original work, except where appropriately referenced.

1.2 Copyright Disclaimer

Much of this content is adapted from the following papers:

- A. W. Simmons, J. J. Wallman, H. Pashayan, S. D. Bartlett, and T. Rudolph. Contextuality under weak assumptions. *New Journal of Physics*, 19:003030, 2017.
- A. W. Simmons. On the computational complexity of detecting possibilistic locality. *Journal of Logic and Computation*, 28(1):203–217, January 2018.
- A. W. Simmons. How (maximally) contextual is quantum mechanics? arXiv:1712.03766, December 2017.

In each of these cases, the right to use the article in full, or in part of a thesis or dissertation, remained with the author, and material directly taken or adapted from these works is referenced as such. The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

1.3 Motivation and Results

The fact that quantum mechanics was contextual was first discovered by Kochen and Specker [60]; and independently by Bell [15], who discovered that in quantum systems of sufficiently large dimension ($d \geq 3$), the outcome of a measurement of a quantum mechanical observable cannot be considered both deterministic and independent of what it is measured alongside. Classical mechanics does not display this property; the answer to a multiple choice question about the state of a classical system can in no sense depend on the incorrect answers that are placed alongside the correct one. This property serves to confuse the ontology of quantum mechanics; measurements can no longer be thought of as revealing the underlying properties of the system being measured. This form of contextuality is sometimes known as "maximal" contextuality, as the demonstrations show that not even a single ontological state which assigns outcomes to measurements is compatible with the operational statistics of quantum mechanics; any attempt to assign outcomes to observables independent of the context in which they are measured is impossible.

Since then, many different flavours of contextuality have been introduced, as well as new frameworks and perspectives on existing ideas. Two such perspectives that will be explored in this thesis are a possibilistic notion of nonlocality, first discovered for tripartite systems by Greenberger, Horne and Zeilinger [45]; and later in bipartite systems by Hardy [47]. This was later generalised to a possibilistic notion of contextuality [2] by Abramsky and Brandenburger; and a

probabilistic notion of contextuality due to Spekkens [95]. Probabilistic notions of contextuality are often justified by appealing to aesthetic principles such as Leibniz's Identity of Indiscernables; in this thesis we show similar motivations can be applied to possibilistic notions, as well as demonstrating that possibilistic assumptions of noncontextuality allow classical Bayesian consistency results to be applied to quantum mechanical ontology. It is of historical interest that the first notions of nonlocality discovered by Bell [17] were probabilistic in nature; the Bell inequality is sensitive to the exact probabilities of experimental outcomes. Possibilistic nonlocality was not discovered until later, although, as argued by Jevtic and Rudolph [54], Einstein and Schrödinger came close to discovering both flavours of nonlocality in their letters of 1935 [36, 37]. On the other hand, the first notions of contextuality observed were possibilistic in nature, with probabilistic notions coming nearly 40 years later. This happenstance and apparent priority of some notions over others in the two related concepts of nonlocality and contextuality has shaped viewpoints of the two in different directions. In Chapter 2 of this thesis, "Flavours of Contextuality", we follow *Contextuality under weak assumptions* by Simmons, Wallman, Pashayan, Bartlett and Rudolph [93], we will demonstrate a framework which is powerful enough to encompass all of these notions of nonlocality and contextuality, and as alluded to above, use it to provide philosophical motivations for possibilistic noncontextuality that were hitherto unknown. We will also give a brief introduction to the Sheaf-theoretic approach of Abramsky and Brandenburger. In Chapter 3, "Common Ground and Differences Between Approaches" we will use ideas suggested by this approach to compare and contrast these different notions of contextuality. We will also investigate and comment on how concepts natural to each approach can be brought into the other.

Chapter 4, "Restrictions on a Noncontextual Reality" also follows *"Contextuality under weak assumptions"*, and focuses on the use of noncontextuality assumptions as axioms or sets of axioms that are jointly ruled out by no-go theorems. In this chapter, we considerdifferent results originally proved using strong assumptions of noncontextuality and demonstrate that these results, or analogues of these results can be shown to hold using strictly weaker assumptions of contextuality.

Commensurately with the importance of contextuality and nonlocality to the field of quantum foundations, and their importance as a resource to communication and information-processing tasks, a number of metrics have been developed to define sensible metrics for quantification of nonlocality and contextuality. Some examples include the work on nonlocality as a resource theory by de Vicente [32], the use of "mana" to describe amounts of distillable contextuality to provide T-gates [51], and the use of the contextual fraction to describe levels of Kochen-Specker contextuality under the assumption of determinism [5]. In Chapter 5 "Metrics of Contextuality and Nonlocality", following *How (maximally) contextual is quantum mechanics?* [91], we will define a metric that differentiates between maximally contextual scenarios, which are indistinguishable under the contextual fraction metric, being given a contextual fraction of 1. In addition, the author demonstrated an upper bound on the metric of what could be achieved by quantum-mechanically accessible scenarios, and demonstrated that this bound is not respected by other theories in theory-space. This is comparable to a Tisrelson bound for this contextuality robustness metric: there is a classical bound, a higher quantum bound, and an even higher bound for theories that can demonstrate superquantum correlations, which reaches the maximum that is mathematically attainable.

Contextuality is known to be in some sense a resource that enables quantum computation [51]. However, operationally it is not easy to determine if a given resource produces contextual statistics. Avis, Imai, Ito and Sasaki [12], building on work by Pitowsky [79], demonstrated that determining whether or not a table of experimental data could come from a locally realistic theory was a computationally hard problem. In particular, they showed that it was a member of the **NP**-complete class of decision problems, the hardest problems for which positive solutions can be verified in polynomial time. Additionally, it was shown by Abramsky, Gottlob and Kolaitis this problem was **NP**-complete even in the possibilistic nonlocality scenario, in which we only consider whether or not events are possible or impossible, for quantum scenarios featuring three or more nonlocal subsystems, or two nonlocal subsystems with each capable of making measurements with three outcomes [3]. On the other hand, Mansfield and Fritz demonstrated that in the case in which we have two subsystems, capable of only making two-outcome measurements,

the decision problem was in **P** [68], and is therefore computationally tractable. This work left only one case to be classified: the case in which there are two subsystems, one of which has access to measurements with two outcomes only, and the other has access to measurements with three (or more) outcomes. In Chapter 6, "The Computational Complexity of Contextuality" we will follow *On the computational complexity of detecting possibilistic locality* [92], in which it was shown that this case is also **NP**-complete. We will show another way in which the case considered by Mansfield and Fritz is exceptional by demonstrating another generalisation of the problem with two-outcome measurements which is also shown to be **NP**-complete.

Chapter 2

Flavours of Contextuality

The fundamental problem of interpreting quantum mechanics is the inability for the vector subspace structure of quantum measurements and the classical logical structure of their outcomes to be identified in a way that is globally consistent. The irreducible events within the former are vector subspaces within a larger Hilbert space, and in the latter are "atoms", complete sets of measurable properties. In the classical world a system has a complete description of all its measurable properties, and measurement of any of those properties is a deterministic process. Measurement operations that are not fundamental quantities can be described using Boolean operators such as OR and AND, with the underlying properties following the same logic. Quantum mechanics, at least superficially, has vector subspace union in place of OR, and subspace intersection in place of AND; these operations as applied to the subspace structure of quantum events do not directly follow the same logic as the classical concepts they generalise, making it difficult to relate the classical notions we use for interpreting our results to the probabilistic behaviours and potential underlying ontology of the theory. It is difficult to bridge the gap of these two different algebras, but the most common approach is to consider the quantum state as a preprobability: a tool that is usable to calculate probabilities of events occurring such as the probability of seeing different macroscopic outcomes given a measurement procedure. This gives rise to a nondeterministic theory of quantum mechanics. It is this tension between the two lattices that eventually gives rise to much of the behaviour that we consider "quantum", such

as complementarity, the existence of uncertainty relations, nonlocality, and contextuality. Since then, many different approaches have been put forward that give us different views on the unintuitive properties that this notion of context-dependence seems to provide quantum mechanics, and certain quantum subtheories. These approaches and the properties they identify are some of our best attempts to isolate what exactly makes a theory nonclassical, with the additional power for information processing and communication tasks that that implies. Each of these notions of contextuality captures some aspect of the underlying problem that these different event spaces cannot be sewn together in a deterministic fashion. This section will review many of these approaches, and we shall attempt some synthesis between their ideas and frameworks.

2.1 Historical background

The first form of what we would now call contextuality that was noted was nonlocality. Einstein and Schrödinger, amongst others, noted that there was a tension between notions of locality central to classical (and relativistic) thinking, and taking the quantum state as ontologically real. This led them to try and identify a theory that could underpin quantum mechanics while better meeting classical intuitions. This point of view is readily apparent within the famous Einstein-Podolsky-Rosen paper, "Can quantum-mechanical description of physical reality be considered complete?" [40]. The claim of the paper is not that the predictions of quantum mechanical experiments are incompatible with a notion of local realism, which will be explored later. Instead, they claim it is evidence that the wavefunction is not physically real, but rather only a summary of the state of knowledge of an experimenter. Einstein wanted a constructive, causal theory that could be said to underpin quantum mechanics, and did not expect that such a thing would be impossible; the EPR paper states that "no reasonable definition of reality could be expected to permit [nonlocality]". That this theory would be deterministic is somewhat taken for granted by Einstein; indeed, even as late as the 1950s he uses the word "causal" synonymously with "deterministic" [38, 39]. Later, Bell's, and Kochen and Specker's writing demonstrates the same bias [15, 60]; that a hidden variable theory for quantum mechanics would

be deterministic was seen to be almost definitional.

Einstein developed these ideas further, and in the summer of 1935 during the immediate aftermath of the EPR paper, Einstein and Schrödinger engaged in a correspondence in which they attempted to pin down issues of locality and interpretation. Einstein hoped they could hasten the end of the "epistemology-soaked orgy"[1] [36] of the Copenhagen interpretation that had been established 10 years earlier, primarily by Bohr and by Heisenberg. In his letter of June 19 [37] Einstein bemoaned that the central point, as he saw it, of his argument that quantum mechanics was incomplete had been lost somewhat upon Podolsky's writing up of the paper. These letters give us an interesting and unique insight into his own views on the matter and the nuances of how Einstein and Podolsky differed in their approaches to the locality problem. EPR discusses the matter of locality as it applies to measurement of one of two entangled, separated systems with a description of what is effectively steering:

> We see therefore that, as a consequence of two different measurements performed upon the first system, the second system may be left in states with two different wave functions. On the other hand, since at the time of measurement the two systems no longer interact, no real change can take place in the second system in consequence of anything that may be done to the first system. This is, of course, merely a statement of what is meant by the absence of an interaction between the two systems.

The article then moves onto perhaps the more famous argument, wherein it is argued that both position and momentum can be considered elements of physical reality, running contrary to quantum mechanical predictions. However, it is the concept of incompleteness– that a single real state is compatible with multiple quantum states– that was Einstein's concern. In his correspondence with Schrödinger, the concept of steering is defined more clearly, and Einstein he goes on to state that he "doesn't give a sausage[2]" whether or not the states steered to are or are not eigenstates for physically relevant operators. The concept of an "element of physical reality" is to Ein-

[1] "Erkenntnistheorie-getränkte Orgie"

[2] "Ob die ψ_B und $\underline{\psi_B}$ als Eigenfunktionen von Observabeln B, $\underline{B}$ aufgefasst werden können ist mir wurst." Emphasis is Einstein's.

stein immaterial compared to what we would now call the ψ-epistemic property, that the wave-function is not physically instantiated but represents only a state of belief or knowledge about the quantum system by the experimenter. In this block of text, the idea that there might be genuine nonlocality in the theory is dismissed out of hand and it is taken for granted that two objects that are separated cannot have an influence on the real state of the other; Einstein writes and emphasises "The actual state of B cannot be dependent on what sort of measurement I take at A"[3]. This dismissal of the idea of nonlocal effects leads Einstein, Podolsky and Rosen to not make explicit their idea of what exactly *locality* is. Some progress on this conceptual front is made in Einstein's letter of the 19$^{\text{th}}$ of June, 1935. Einstein postulates a "*Trennungsprinzip*", or "disconnection principle"[4], which, while perhaps not concretely defined, is recognisably a movement towards a precise definition of what it means for a theory to be local.

It was not until Bell's famous paper of 1964 [17], nearly 30 years later, that it was noted that in fact quantum mechanics was not just superficially, but rather intrinsically, nonlocal. It was not just the case that the most obvious ontology, with its property of ψ-completeness, seemed to be nonlocal, but rather that the probabilistic predictions could not be met by any local theory. This showed that the assumption of EPR that a local, deterministic theory could exist[5] underpinning quantum mechanics was false. In 1966, Bell published "On the Problem of Hidden Variables in Quantum Mechanics", in which he demonstrated a proof what we would now call contextuality. It is not well known, although it was noted for example by Mermin [72], that this paper was in fact written *before* Bell's famous paper of 1964. However, it was lost in a filing drawer for more than a year which led to its delayed publication. Although his proof has a markedly different characteristic, Bell independently found a proof of contextuality and submitted the paper before Kochen and Specker[6], though Specker had alluded to having a proof

[3] "Der wirkliche Zustand von B kann nun nicht davon abhängen, was für eine Messung ich an A vornehme. Emphasis is again Einstein's.

[4] An alternative translation, used for example by Fine [42], is "separation principle". We will use "disconnection principle", to emphasise that there is no ontological connection between objects which are physically distant; physical separation implies an ontological disconnection.

[5] "[W]e left open the question of whether or not such a description exists. We believe, however, that such a theory is possible."

[6] As Bell notes, the result can be derived as a corollary to Gleason's work of 1957 [44]. Bell's paper, however, seems to be the first time that this corollary of the theorem was specified, and noticed as having fundamental importance.

of the result all the way back in 1960 [94]. In 1967 Kochen and Specker published their famous, and very similarly titled paper "The problem of Hidden Variables in Quantum Mechanics" [60]; their approach to the problem is very different from Bell's, making use of explicit orthogonality graphs and demonstrating for the first time an explicit subtheory of quantum mechanics that does not admit such a canonical hidden variable interpretation.

Each of these approaches had the property that they explicitly looked at deterministic hidden variable models. In the case of Bell's paper of 1964, he illustrates that determinism does not make any difference either way to his result, by demonstrating that a hidden variables theory can always be made deterministic at the expense of some ontological overhead. The claim of Bell that determinism was not restrictive in the locality scenario was expanded upon by Fine [41], who highlighted specifically the equivalence between Bell's condition of "factorisability" and the admission of deterministic ontological models. A generalisation to other scenarios than that considered by Bell and CHSH was later proved by Abramsky *et al* [2]. As we shall see, the assumption, or lack thereof, of determinism for hidden variables shapes the views and the philosophical justifications for different assumptions of noncontextuality outside of the locality scenario.

The deterministic model of quantum mechanics Bell presents, now often called "Bell's first model", served to reignite interest in deterministic hidden variable models which had been suppressed by the work of Von Neumann, who had claimed a disproof of deterministic hidden variables in 1955 [103]. Bell is very critical of this proof, saying that it is "not merely false but foolish" [67]; more generously it could be said to illustrate the pitfalls of trying to apply classical intuitions to quantum mechanical problems. Mermin clearly shared this opinion when in 1993 he wrote [72]

> A few years later Grete Hermann, 1935 [50], pointed out a glaring deficiency in the argument, but she seems to have been entirely ignored. Everybody continued to cite the von Neumann proof. A third of a century passed before John Bell, 1966, rediscovered the fact that von Neumanns no-hidden-variables proof was based on

> an assumption that can only be described as silly— so silly, in fact, that one is led to wonder whether the proof was ever studied by either the students or those who appealed to it to rescue them from speculative adventures.

The mistake made by Von Neumann was to assume that, if A, B are two observables that do not necessarily commute, then for a deterministic hidden variable theory assigning $v(X)$ to an observable X, we must have for $C = A + B$

$$v(C) = v(A) + v(B). \tag{2.1}$$

As Mermin and Bell each point out, this holds in expectation within quantum mechanics, but a simple argument reveals that it cannot hold even for a two-dimensional quantum system: this would imply that a valuation of $\sigma_x + \sigma_y$ could only be in the set $\{-2, 0, 2\}$, and yet this observable has eigenvalues $\{-\sqrt{2}, \sqrt{2}\}$. However, if we restrict Von Neumann's condition to hold only for commensurable operators, we can recover the result of Kochen and Specker; it is really in attempting to apply these relations to observables whose quantum mechanical representations are noncommuting that we encounter an issue. Bell and Mermin clearly felt that the physical content of Von Neumann's proof was limited, however this is not a universal position: see for example a more recent defense of Von Neumann's argument by Bub [23].

Later, questions about contextuality sparked a renewed interest in hidden variable theories of quantum mechanics [48, 49, 89], with a focus on generalised notions of contextuality that did not assume determinism, following the work of Spekkens [95]. This equivalence-based approach makes use of a framework that attempts to capture minimal assumptions for a realist theory of quantum mechanics, called the ontological models formalism. In parallel, the sheaf-theoretic approach, introduced by Abramsky and Brandenburger [2] emerged, which is arguably more faithful to the original concepts of contextuality explored by Kochen and Specker, giving the assumption of determinism a central position, as it has in classical mechanics. This latter approach includes a clear hierarchy of contextuality that includes probabilistic, possibilistic and maximal contextuality which until the paper *Contextuality under weak assumptions* by Simmons

et al was not present within the Spekkens or equivalence-based approach.

2.2 The Ontological Models Formalism

Many modern approaches to contextuality take an approach that builds in a certain amount of operationalism and theory-independence. The starting point for the framework of ontological models is operational, like a principle theory; in a laboratory there are certain preparations one can perform, certain transformations that one can make on a physical system, and certain measurements that can be made at the end. Quantum mechanics can be thought of as a tool that takes in descriptions of experiments in terms of preparations, transformations and measurements, and provides probabilities for the different experimental outcomes. For quantum Bayesians, this is in fact entirely what quantum mechanics is.

The ontological models formalism attempts to capture minimal requirements for realistic theories. Unlike fully relational or subjective views on quantum mechanics, realist theories posit the existence of some definite state of reality, from which probabilistic predictions, or propensities, are derived. We will mostly be considering quantum mechanics, or some subtheory thereof as the operational theory whose predictions our ontological models must match, but this is by no means required by the formalism. We will extend the usual definition of an ontological model and an operational theory to support nonstandard probability theories, and will take the probabilities to be elements of some semiring $\mathcal{R}$, rather than specifically the usual probabilistic semiring; this will make it a little easier to compare to the sheaf-theoretic notion. The notion of a semiring captures the minimal structure necessary for a probability theory; nonintersecting events can have a union, with probability of the union given by $p_1 + p_2$, and independent events can have their probabilities multiplied $p_1 \times p_2$. We do not assume the existence of a notion of subtraction or division. Formally, we will take an operational theory to be denoted $[\mathcal{P}, \mathcal{M}, \mathcal{T}, \mathrm{Pr}]$; a set of preparation procedures $\mathcal{P}$; a set of measurement procedures $\mathcal{M}$; a set of transformation procedures $\mathcal{T}$ closed under composition; and a function Pr that specifies the probabilities of the different possible outcomes once preparation and measurement procedures

have been chosen. We therefore have $\Pr : \mathcal{P} \times \mathcal{M} \times \mathcal{T} \times \mathcal{O} \to \mathcal{R}$, where $\mathcal{O}$ is the set of outcomes, which we can usually take to be $\mathbb{N}$ or a subset thereof. We will denote preparations, transformations, and measurement events (outcomes given a selected measurement procedure) as *elements* of the operational theory.

An operational theory assigns probabilities $\Pr(k|\mathcal{P}, \mathcal{T}, \mathcal{M})$ to outcomes k occurring when procedures for a preparation $\mathcal{P}$, a transformation $\mathcal{T}$, and finally a measurement $\mathcal{M}$ are implemented. Any particular $\mathcal{P}$, $\mathcal{M}$ or $\mathcal{T}$ will be denoted an *element* of the operational theory. Quantum mechanics can be regarded as an operational theory by identifying preparation procedures with density matrices ρ, transformations with CPTP maps $T(|\psi\rangle\langle\psi|) = \sum_{i=1}^{N(T)} K_i|\psi\rangle\langle\psi|K_i^\dagger$, and K-outcome measurements with positive-operator-valued measures (POVM) $\mathcal{M} = \{E_1, \dots E_K\}$. The probability for an outcome k given a preparation described by ρ and a transformation by U is then given by the Born rule,

$$\Pr(k|\rho, U, \mathcal{M}) = \mathrm{Tr}[E_k U \rho U^\dagger]. \tag{2.2}$$

An ontological model for such an operational theory consists of a set Λ, equipped with a σ-algebra Σ, of possible ontic (or "real") states alongside ontological representations of preparations, transformations and measurements, dependent on each other only via states $\lambda \in \Lambda$. Some definitions will also require a topology be defined on Λ, and in this case Σ can be taken to be the Borel σ-algebra. A preparation procedure $\mathcal{P}$ is represented within the ontological model by the preparation of a system in an ontic state λ sampled according to some measure $\mu_\mathcal{P}$ on Λ. In particular, the measure assigned to the full space Λ by any particular $\mu_\mathcal{P}$ must be 1.

A transformation $\mathcal{T}$ in the operational theory is represented in the ontological model by a map that, for a system in the ontic state λ, samples a new $\lambda' \in \Lambda$ according to the conditional probability measure $\Gamma_\mathcal{T}(\lambda'|\lambda)$ on Λ. Finally, a K-outcome measurement $\mathcal{M}$ in the operational theory is represented in the ontological model by a response function $\xi : \Lambda \to R$ such that, for a system in the ontic state λ, the measurement outcome $k \in \{1, \dots, K\} =: \mathbb{N}_K$ is sampled from the conditional probability distribution $\{\xi_\mathcal{M}(k|\lambda)\}$ on $\mathbb{N}_K$. In particular, when $\xi_\mathcal{M}(k|\lambda) \in \{0, 1\}$

for all λ and for all $\mathcal{M}$ we say that the ontological model is is *deterministic*; an ontic state constitutes a deterministic assignment of outcomes to measurements. We will typically denote ontological models by $[\Lambda, \{\Gamma\}, \{\xi\}]$, as we are typically not concerned with state update rules following measurements.

In addition, there are a number of assumptions about the relationships between different elements of the ontological model that are typically (and tacitly) assumed to hold. One important assumption for many proofs is that statistical mixtures of operational procedures are respected ontologically; our ontological models have a convex structure and therefore preparations (or other operational procedures) can be statistically mixed with each other. Without this assumption many of these noncontextuality relations become trivial. An example of a model without such an assumption is quantum mechanics restricted to pure states, unitary transformations, and projective measurements. That this is an explicit postulate of the ontological models formalism was highlighted by Wester [105]. For example, if we have two preparation procedures in the operational theory $\mathcal{P}_1$ and $\mathcal{P}_2$, with ontological representations $\mu_{\mathcal{P}_1}$ and $\mu_{\mathcal{P}_2}$ respectively, then the operational preparation obtained by performing $\mathcal{P}_1$ with probability p and $\mathcal{P}_2$ otherwise has an ontological representation given by

$$\mu_{p\mathcal{P}_1+(1-p)\mathcal{P}_2} = p\mu_{\mathcal{P}_1} + (1-p)\mu_{\mathcal{P}_2}. \tag{2.3}$$

An equivalent assumption is assumed to hold true for both transformations and measurements. For a combination of two measurement effects $m = c_1 m_1 + c_2 m_2$:

$$\xi_m(\lambda) = c_1\xi_{m_1}(\lambda) + c_2\xi_{m_2}(\lambda) \quad \forall\lambda. \tag{2.4}$$

This convexity is completely necessary for defining notions of preparation noncontextuality, and important for defining notions of measurement noncontextuality for unsharp measurements of rank greater than 1; this will be highlighted later on. In addition, an assumption is made about the compositional behaviour of transformation procedures. If $\mathcal{T}_3$ is the transformation

procedure obtained by performing $\mathcal{T}_1$ followed by $\mathcal{T}_2$, then we assume

$$\Gamma_{\mathcal{T}_3}(\lambda''|\lambda) = \Gamma_{\mathcal{T}_2}(\lambda''|\lambda')\Gamma_{\mathcal{T}_1}(\lambda'|\lambda). \tag{2.5}$$

The ontological model reproduces the predictions of the operational theory if

$$\Pr(k|\mathcal{P},\mathcal{T},\mathcal{M}) = \int \mathrm{d}\lambda\, \mathrm{d}\lambda'\, \mu_{\mathcal{P}}(\lambda)\Gamma_{\mathcal{T}}(\lambda'|\lambda)\xi_{\mathcal{M}}(k|\lambda') \tag{2.6}$$

for all $k, \mathcal{P}, \mathcal{T}, \mathcal{M}$.

While all constructions in this paper hold for general operational theories, we will only explicitly consider ontological models of quantum mechanics. For brevity, we will omit transformations and focus solely on preparations and measurements, though all definitions can be directly extended to account for transformations.

The Beltrametti-Bugajski model [18] is perhaps the simplest example of an ontological model of pure-state quantum mechanics. In it, pure quantum states are treated as being physical states, so we have $\Lambda = \mathcal{H}$; $\mu_{|\psi\rangle}(|\lambda\rangle) = \delta_{|\psi\rangle}(|\lambda\rangle)$; and $\xi_M(E||\lambda\rangle) = \mathrm{Tr}(E|\lambda\rangle\langle\lambda|)$. It could be called the ψ-complete ontological model, where the wavefunction is not only real but are the atomic, irreducible states of the ontology.

A less trivial model is that of Kochen and Specker [60], which happens to have the ψ-epistemic property alluded to above. It was presented by Kochen and Specker to demonstrate that the $d \geq 3$ restriction on their famous theorem was necessary, as the model forms a counterexample for $d = 2$. The ontic state space is a sphere, so we have $\Lambda = \mathcal{S}^2$. The preparation distributions are given by

$$\mu_{|\psi\rangle}(\lambda) = \begin{cases} \cos(\theta) & 0 \leq \theta \leq \frac{\pi}{2} \\ 0 & \frac{\pi}{2} < \theta, \end{cases} \tag{2.7}$$

in which θ is the angle subtended at the centre of the sphere by λ and by the Bloch sphere

representation of the state $|\psi\rangle$. The response functions are given by

$$\xi_{|\psi\rangle\langle\psi|} = \begin{cases} 1 & 0 \leq \theta < \frac{\pi}{2} \\ \frac{1}{2} & \theta = \frac{\pi}{2} \\ 0 & \frac{\pi}{2} < \theta. \end{cases} \tag{2.8}$$

Unitary transformations are represented by rotations of the sphere exactly as they would be in the Bloch sphere representation. The model is, as mentioned, ψ-epistemic, and is deterministic[7].

2.3 Spekkens, or Equivalence-based, Contextuality

Contextuality is so named because it is a phenomenon appearing when we have the same element of an operational theory appearing in two different "contexts". These manifestations are treated identically by the operational theory, always leading to the same probabilities. In fact, this is why the same notation is used for the objects in the first place, as a context-independent symbol is all that is needed to calculate probabilities. This feature, which is sometimes called "noncontextuality of probability" [64] appears as a formal axiom needed to prove Gleason's theorem [44][8].

2.3.1 Introduction to the equivalence formalism

There are two ways of recognising that the same element has appeared in two different situations; by defining two elements to be the same if they are operationally equivalent or by fiat, as part of the defined event structure of the theory. Spekkens contextuality takes the former option,

[7] In this presentation, there is a nondeterministic set of measure 0, but this can be avoided and made fully outcome-deterministic without much difficulty.

[8] This could be said to be somewhat tautological. We denote two elements of the operational theory with the same symbol exactly because they lead to the same statistics; good notation tends to strip out unnecessary information and so a symbol represents exactly equivalence classes of operational statistics. To suggest that two projectors could be given different operational probabilistic predictions in different contexts, then, is a strange suggestion, if a logically required one. This property of quantum mechanics is sometimes used to claim that quantum mechanics is not contextual [73]. In the view of the author, this borders on a content-free statement.

which reflects the view that these operational elements are identified because of their operational indistinguishability; and as we will see later, the sheaf approach takes the latter, which reflects the idea that noncontextuality of probability is a given, once we have defined our specific event structure.[9]

There is no formal argument to be made that these elements which are operationally context-independent should also be ontologically context-independent: this must be taken axiomatically. Motivation for such an axiom, or "noncontextuality assumption" cannot be logically deduced or extrapolated from experimental data, but rather must be based on aesthetic principles, such as Leibniz's principle [97]. Leibniz suggested that any two objects that cannot be discerned from one another in any way must be regarded as ontologically identical. This principle is also known as "Leibniz's law", or the "Identity of Indiscernibles". Einstein used the concept to great effect when considering the equivalence principle in general relativity; since it is not possible to differentiate between acceleration and gravitation, he worked on a theory that would combine them into one ontological entity. Another example is the magnetic vector potential in electrodynamics, which is regularly called unphysical because the addition to it of any curl-free vector field causes it to give rise to the same magnetic field and the same operational statistics. Therefore, there are different vector potentials that give rise to the same observable physics: a silent assumption of the identity of indiscernibles is that there can only be ontological (physical) variation between to objects if there is a difference in their physical *behaviour*, and so the existence of specific vector potentials is challenged. The equivalence approach was also championed by Pitowsky [80] in his comments on the usage of de Finetti-style arguments to justify a theory of quantum probability, in which quantum mechanical probabilities are explained as degrees of belief. Pitowsky took the view that when quantum mechanics always assigns the same probabilities to two events existing in different Boolean algebras (compatible contexts), one must regard them as equivalent logical statements.

Much of the language and setting of this section is taken from *Contextuality under weak assumptions* by Simmons, Wallman, Pashayan, Bartlett and Rudolph [93]. The framework developed

[9] In the view of the author, this is closer to the original intent of Gleason.

in that paper allows us to expand the traditional notions of Spekkens contextuality, and to better examine its assumptions. In particular, we will define a generalised form of the standard noncontextuality axioms to be the assumption that a particular relation holding at the level of the ontological theory implies another relation at the level of the ontological theory. These relations will often be similar in their content, as this makes them easier to motivate. This represents a view that noncontextuality assumptions are schemes for making conclusions about ontology based on the operational predictions of a given theory. When a proof of contextuality is found, it then shows that the assumptions about the correspondence between ontological and operational phenomena used to define that scheme were too strong, and therefore incompatible with the operational statistics.

We will use this framework to define a "possibilistic" analogue of contextuality, and demonstrate that it too can be motivated by philosophical or aesthetic considerations. This notion of possibilistic contextuality will reflect the assumption that when two elements of an operational theory give rise to the same set of operational possibilities that this is reflected ontologically; they also give rise to the same ontological possibilities. Consideration of operational elements by what possibilities they give rise to has been considered before when considering subjective Bayesian approaches to quantum mechanics; the idea emerges naturally from classical concerns from considerations of what it means for different parties to hold different, yet compatible, beliefs about the state of a quantum system. This problem has been investigated by Brun, Finkelstein and Mermin [22] and separately by Caves, Fuchs and Schack [27]. We shall see that under our assumption of possibilistic noncontextuality, some of these quantum consistency conditions coincide with classical consistency conditions, and therefore after demonstrating that quantum mechanics is possibilistically contextual we see that these classical consistency conditions cannot be thought to directly apply in the quantum setting. We will show that the notion of possibilistic noncontextuality is strictly weaker than that of probabilistic noncontextuality, and in Chapter 4 that analogues of some results originally proved using probabilistic noncontextuality as an axiom in fact hold with this weaker assumption.

During this section we will focus primarily on the contextuality of preparations and measure-

ments. Transformation contextuality is a powerful concept that appears, for example, in some subtheories of single qubits [66], but is not present within the sheaf-theoretic approach.

Definition 1 (Operational Relation)**.** An *Operational Relation* $\sim$ for an operational theory $[\mathcal{P}, \mathcal{M}, \mathcal{T}, \Pr]$ is a symmetric, reflexive relation over elements of the operational theory.

Definition 2 (Ontological Relation)**.** An *Ontological Relation* $\approx$ for an ontological model is a symmetric, reflexive relation over elements of the operational theory.

Note that we do not require that the operational or ontological relations are necessarily transitive by definition. Nevertheless, many noncontextuality assumptions will be built from operational and ontological relations that do obey transitivity and are therefore *equivalence relations*, and the focus of this Chapter is on relations that are equivalence relations. When two elements of an operational theory or ontological model are in the same equivalence class, we will call them *equivalent*.

Definition 3 (Noncontextuality assumption)**.** A noncontextuality assumption is a statement that when two objects in the operational theory are related by a specified *operational relation* $\sim$, their ontological representations must be related by an *ontological relation* $\approx$. For example, applied to preparations these are statements of the form

$$\mathcal{P}_1 \sim \mathcal{P}_2 \Rightarrow \mu_{\mathcal{P}_1} \approx \mu_{\mathcal{P}_2} \tag{2.9}$$

So, the operational relation is a condition that will specify which elements of the operational theory are related to each other. The ontological relation then tells us how the elements of the ontological representations of those operational elements are related. If we have a noncontextuality assumption **X**, we would say that an operational theory exhibits **X**-contextuality if for any ontological model reproducing the probabilistic predictions of the operational theory in its semiring $\mathcal{R}$ contains a contradiction to **X**. That is, there exist some pair of operational elements related by $\overset{\mathbf{x}}{\sim}$ whose ontological representations are not related by $\overset{\mathbf{x}}{\approx}$.

This approach is general enough that it can be made to encapsulate any current paradigm

for contextuality, but it is of particular use when showing how the Spekkens approach can be generalised.

2.3.2 Probabilistic Noncontextuality in the Equivalence Formalism

Let us now consider the most common form of equivalence-based contextuality, that is often called Spekkens contextuality. We will call it "Probabilistic" contextuality, to clearly highlight the sensitivity of the notion to precise experimental probabilities.

Definition 4 (ProbOp). Two preparation procedures $\mathcal{P}_1$ and $\mathcal{P}_2$ in an operational model are *probabilistically equivalent* (denoted $\mathcal{P}_1 \overset{\mathbf{Prob}}{\sim} \mathcal{P}_2$) if for all outcomes k of all measurement procedures $\mathcal{M}$,

$$\Pr(k|\mathcal{P}_1, \mathcal{M}) = \Pr(k|\mathcal{P}_2, \mathcal{M}). \tag{2.10}$$

Two measurement outcomes $k_1|\mathcal{M}_1$ and $k_2|\mathcal{M}_2$ in an operational model are *probabilistically equivalent* (denoted $\mathcal{P}_1 \overset{\mathbf{Prob}}{\sim} \mathcal{P}_2$) if for all preparation procedures $\mathcal{P}$,

$$\Pr(k_1|\mathcal{P}, \mathcal{M}_1) = \Pr(k_2|\mathcal{P}, \mathcal{M}_2). \tag{2.11}$$

In the case of quantum mechanics, all operational predictions for a preparation procedure $\mathcal{P}$ are completely encoded in the associated density matrix $\rho_{\mathcal{P}}$. Therefore we have

$$\mathcal{P}_1 \overset{\mathbf{Prob}}{\sim} \mathcal{P}_2 \Leftrightarrow \rho_{\mathcal{P}_1} = \rho_{\mathcal{P}_1}. \tag{2.12}$$

As a particular example, the maximally mixed state can be prepared in many ways, such as by tracing over half of a maximally-entangled pair, randomly applying a unitary operation or by preparing a system in a basis and forgetting which element was prepared. All of these methods are probabilistically equivalent for a single qubit.

Having specified an operational relation, the next step is to specify an ontological relation.

Definition 5 (**ProbOn**)**.** Two preparation distributions $\mu_{\mathcal{P}_1}$, $\mu_{\mathcal{P}_2}$ in an ontological model are *probabilistically equivalent*, denoted $\mu_{\mathcal{P}_1} \overset{\textbf{Prob}}{\approx} \mu_{\mathcal{P}_2}$, if and only if $\mu_{\mathcal{P}_1}(\lambda) = \mu_{\mathcal{P}_2}(\lambda)\ \forall \lambda \in \Lambda$.

Two measurement effects $\xi_{k_1|\mathcal{M}_1}$, $\xi_{k_2|\mathcal{M}_2}$ in an ontological model are *probabilistically equivalent*, denoted $\xi_{k_1|\mathcal{M}_2} \overset{\textbf{Prob}}{\approx} \xi_{k_2|\mathcal{M}_2}$, if and only if $\xi_{k_1|\mathcal{M}_1}(\lambda) = \xi_{k_2|\mathcal{M}_2}(\lambda)\ \forall \lambda \in \Lambda$.

Definition 6 (**Probabilistic Noncontextuality**)**.** For any two preparations $\mathcal{P}_1$, $\mathcal{P}_2$ related by **ProbOp**, their ontological representations $\mu_{\mathcal{P}_1}, \mu_{\mathcal{P}_2}$ are related by **ProbOn**. That is, the probabilistic noncontextuality assumption states that

$$\mathcal{P}_1 \overset{\textbf{Prob}}{\sim} \mathcal{P}_2 \Rightarrow \mu_{\mathcal{P}_1} \overset{\textbf{Prob}}{\approx} \mu_{\mathcal{P}_2}. \tag{2.13}$$

For any two measurement outcomes $k_1|\mathcal{M}_1$, $k_2|\mathcal{M}_2$ related by **ProbOp**, their ontological representations $\xi_{k_1,\mathcal{M}_1}, \xi_{k_2,\mathcal{M}_2}$ are related by **ProbOn**. That is, the probabilistic noncontextuality assumption states that

$$k_1|\mathcal{M}_1 \overset{\textbf{Prob}}{\sim} k_2|\mathcal{M}_2 \Rightarrow \xi_{k_1|\mathcal{M}_1} \overset{\textbf{Prob}}{\approx} \xi_{k_2|\mathcal{M}_2}. \tag{2.14}$$

These notions of noncontextuality, as applied to preparations and measurements, strongly constrain the space of accessible operational theories, which will be explored further in Chapter 4, Restrictions on a Noncontextual Reality. Some of these restrictions, however, are relevant to the discussion and literature of the interplay between different notions of contextuality, so we will visit them briefly in this chapter. One surprising result is that preparation noncontextuality is already too strong an assumption to apply to the very restricted scenario of a rebit; a qubit restricted to one plane of the Bloch sphere. The proof, attributed to Rudolph, uses the following important properties of ontological models:

Lemma 1. *Let $\mathcal{P}_1$, $\mathcal{P}_2$ be preparation procedures that can be deterministically distinguished, that is to say that there exists some measurement $\mathcal{M}$, with outcomes $\{k_1, k_2, k_e\}$, such that $Pr(k|\mathcal{P}_i, \mathcal{M}) = 1$ for both $i = 1$ and $i = 2$. Then, in any ontological representation, $\mathcal{P}_1$ and $\mathcal{P}_2$ are associated with disjoint regions of the ontic state space; their associated measures $\mu_{\mathcal{P}_1}$, $\mu_{\mathcal{P}_2}$*

have disjoint support, except maybe on a region of measure zero:

$$S(\mu_{\mathcal{P}_1}) \cap S(\mu_{\mathcal{P}_2}) = \emptyset. \tag{2.15}$$

Where the support $S(\mu)$ of a measure μ is the largest set such that every open set with nonempty intersection with $S(\mu)$ has nonzero measure under μ.

Considering $\mathcal{M}$ to be a three-outcome measurement is possible without loss of generality since we have assumed that our operational theory is closed under coarse-graining of measurements. Spekkens uses a slightly more informal definition of support, concluding $\mu_{\mathcal{P}_1}(\lambda)\mu_{\mathcal{P}_2}(\lambda) = 0$ $\forall\lambda$. This is a fairly intuitive property; if there were a region of states in the support of, and therefore compatible with, both distributions, then when the ontic state happened to lie in that region, the measurement $\mathcal{M}$ could not hope to distinguish deterministically between the two preparations. Since $\mathcal{M}$ can do this, there cannot be any such overlaps. The second property used is that statistical mixtures at the operational level correspond to statistical mixtures at the ontological level.

Let us define a sharp measurement to be a measurement outcome for which a preparation exists such that the outcome occurs with probability 1. An example in quantum mechanics would be preparing a state, and performing a PVM containing a projector onto that state. Spekkens demonstrates that under the assumption of preparation noncontextuality, sharp measurements are deterministic; their outcomes are fixed given the ontological state. This is achieved by noting that since the completely mixed state can be made by an even statistical mixture of basis states, for any basis, the union of the supports of any set of basis vectors must be the whole space:

$$\bigcup_{i=0}^{d-1} S\left(\mu_{|i\rangle}\right) = \Lambda. \tag{2.16}$$

We can see this since we can take as a definition that every ontic state must be in the support of the preparation of some pure state, and every pure state can convexly mixed with other states to create the maximally mixed state. If the maximally mixed state has a noncontextual support, then that support must therefore be the entire ontic state space Λ. We note then that every

state is in some basis, and so the response function for the projector onto that state must take the value 1 in the support of the state, and 0 elsewhere. If we prepare a state and then measure with respect to the projector onto that state then we must see the projector deterministically taking a value of 1 and so the response function must take a value of 1; likewise the complement of the support of the preparation of a pure state is the span of the orthogonal basis vectors, and so the response function must take the value 0 on those states. Hence we have recovered determinism for sharp measurements. Wester [105] reiterates these conditions, noting that the state does not have to be maximally mixed in the usual sense, as long as it has a nonzero overlap with every vector.

We see that in the ontological models formalism, the symmetry between preparations and measurements is broken, at least when considering what noncontextuality means for the two concepts. This can be illustrated by the fact that preparation noncontextuality is incompatible with quantum mechanics, but measurement noncontextuality by itself is not, as indeed it isn't with any operational theory. In fact, we should treat the notion of preparation noncontextuality with some suspicion: noncontextuality for measurements is a signature of nonclassicality, since the simplicial structure of the classical ontology allows for every potential measurement to have a definite outcome; on the other hand, it is somewhat nonsensical to even define a notion of preparation contextuality classically, since every mixed state has a unique decomposition into pure states. Preparations cannot have any context dependence, and this is disallowed by the classical structure on arguably a more meaningful level than the comparative statement for measurements. However, this does show that contextuality for preparations, or any epistemic restriction on preparations, should be considered a signature of nonclassicality.

The "probabilistic" title is well deserved; both the operational and ontological relations are probabilistic insofar as they are sensitive to the full set of probabilities associated with their domain. These relations are very restrictive, in that they single out very narrow equivalence classes. As a result, $\overset{\mathbf{Prob}}{\sim}$ is a very selective operational relation; this leads to relatively few operational elements being related. However, the ontological relation is also very restrictive: elements that are related are exactly identical.

2.3.3 Possibilistic Noncontextuality in the Equivalence Formalism

As we have discussed, contextuality is often considered to be a generalisation of nonlocality, especially under the assumption of factorisable probability distributions, or equivalently determinism. Nonlocality is then a specific kind of contextuality, where we see probabilistic predictions that are divergent from the classical, and must conclude that in a nonlocality scenario there is some sort of ontological effect at one location depending on the *context* of the measurement setting at the other site, or give up some other assumptions of realism. In this viewpoint, these probabilistic notions of noncontextuality are equivalent to Bell inequalities; the precise operational probabilities cannot be explained by a noncontextual, or respectively local, model. The approach of Kochen and Specker can be thought of as harnessing the same sort of logic as the GHZ proof of nonlocality, where we demonstrate that no state of affairs, that is, a deterministic assignment of outcomes to measurements, can be globally consistent. In the hierarchy of contextuality as defined by Abramsky and Brandenburger, an intermediate stage exists, which they call Possibilistic contextuality but is sometimes called Logical or Hardy contextuality [2, 63, 65]. This also has a nonlocality equivalent known as logical nonlocality, defined in Reference [2]; an example of a scenario demonstrating this feature is Hardy's paradox. Following Simmons, Wallman, Pashayan, Bartlett and Rudolph [93] we shall give an equivalence-based formalism to capture these possibilistic notions of contextuality, and demonstrate that philosophical justification exists for this notion.

In the quantum case, two preparation procedures $\mathcal{P}_1$ and $\mathcal{P}_2$ are possibilistically equivalent if and only if the density matrices they give rise to have the same support (or equivalently, the same kernel). The kernel, which is spanned by the states with which ρ has no overlap, completely defines the possibilistic structure of measurement.

$$\mathcal{P}_1 \overset{\textbf{Poss}}{\sim} \mathcal{P}_2 \Leftrightarrow \ker \rho_{\mathcal{P}_1} = \ker \rho_{\mathcal{P}_2} \tag{2.17}$$

where $\ker M = \{v : Mv = 0\}$ is the kernel of M.

Definition 7 (**PossOn**)**.** Two preparation distributions $\mu_{\mathcal{P}_1}$, $\mu_{\mathcal{P}_2}$ in an ontological model are

possibilistically equivalent, denoted $\mu_{\mathcal{P}_1} \overset{\textbf{Poss}}{\approx} \mu_{\mathcal{P}_2}$, if $\mathcal{S}(\mu_{\mathcal{P}_1}) = \mathcal{S}(\mu_{\mathcal{P}_2})$, where $\mathcal{S}(\mu)$ denotes the support of the measure μ.

We will use this to define two separate noncontextuality assumptions.

Definition 8 (**Hardy Noncontextuality**)**.** For any two preparations $\mathcal{P}_1$, $\mathcal{P}_2$ related by **ProbOp**, their ontological representations $\mu_{\mathcal{P}_1}, \mu_{\mathcal{P}_2}$ are related by **PossOn**. That is, the Hardy noncontextuality assumption states that

$$\mathcal{P}_1 \overset{\textbf{Prob}}{\sim} \mathcal{P}_2 \Rightarrow \mu_{\mathcal{P}_1} \overset{\textbf{Poss}}{\approx} \mu_{\mathcal{P}_2}. \tag{2.18}$$

For any two measurement outcomes $k_1|\mathcal{M}_1$, $k_2|\mathcal{M}_2$ related by **ProbOp**, their ontological representations $\xi_{k_1,\mathcal{M}_1}, \xi_{k_2,\mathcal{M}_2}$ are related by **ProbOn**. That is, the Hardy noncontextuality assumption states that

$$k_1|\mathcal{M}_1 \overset{\textbf{Prob}}{\sim} k_2|\mathcal{M}_2 \Rightarrow \xi_{k_1|\mathcal{M}_1} \overset{\textbf{Poss}}{\approx} \xi_{k_2|\mathcal{M}_2}. \tag{2.19}$$

Once again using Specker's parable as an example, let us consider first the interpretation of this noncontextuality assumption when applied to preparations. It states that two preparations in the same operational equivalence class, leading to identical operational predictions, must have ontic distributions that agree at the possibilistic level. That is, their associated preparation distributions in the ontological representation must have the same *supports*. While Hardy's notion of nonlocality or noncontextuality is not usually given in this form, this is a fair way to bring the concept into the ontological models formalism. The notion is concerned with operational elements that are represented identically and so are related by $\overset{\textbf{Prob}}{\sim}$; and makes an ontic claim that is possibilistic in nature. This may be surprising, given that in the Sheaf formalism the concept is superficially to do with meeting specific empirical distributions rather than ontological distributions; this will be explored in greater detail below but the two concepts are rendered the same under the assumption of outcome determinism. When split into explicit relationships between operational elements and between ontological elements, this notion seems to contain a mismatch of types: the operational representation is probabilistic in nature, and

the ontological representation is possibilistic in nature. This observation leads to the following definitions, a possibilistic operational relation, and a noncontextuality assumption that leverages it.

Definition 9 (**PossOp**)**.** Two preparation procedures $\mathcal{P}_1$ and $\mathcal{P}_2$ in an operational theory are *possibilistically equivalent*, denoted $\mathcal{P}_1 \overset{\textbf{Poss}}{\sim} \mathcal{P}_2$, if for all outcomes k of all measurement procedures $\mathcal{M}$,

$$\Pr(k|\mathcal{P}_1, \mathcal{M}) = 0 \Leftrightarrow \Pr(k|\mathcal{P}_2, \mathcal{M}) = 0. \tag{2.20}$$

Definition 10 (**Possibilistic Noncontextuality**)**.** For any two preparations $\mathcal{P}_1$, $\mathcal{P}_2$ related by **PossOp**, their ontological representations $\mu_{\mathcal{P}_1}, \mu_{\mathcal{P}_2}$ are related by **PossOn**. That is, the Possibilistic noncontextuality assumption states that

$$\mathcal{P}_1 \overset{\textbf{Poss}}{\sim} \mathcal{P}_2 \Rightarrow \mu_{\mathcal{P}_1} \overset{\textbf{Poss}}{\approx} \mu_{\mathcal{P}_2}. \tag{2.21}$$

For any two measurement outcomes $k_1|\mathcal{M}_1$, $k_2|\mathcal{M}_2$ related by **ProbOp**, their ontological representations $\xi_{k_1,\mathcal{M}_1}, \xi_{k_2,\mathcal{M}_2}$ are related by **PossOn**. That is, the Possibilistic noncontextuality assumption states that

$$k_1|\mathcal{M}_1 \overset{\textbf{Poss}}{\sim} k_2|\mathcal{M}_2 \Rightarrow \xi_{k_1|\mathcal{M}_1} \overset{\textbf{Poss}}{\approx} \xi_{k_2|\mathcal{M}_2}. \tag{2.22}$$

This is a genuinely possibilistic notion of noncontextuality; the operational relation is defined only by operational possibilities, and the ontological relation defines exactly the ontological possibilities. Unlike Hardy contextuality, above, this notion has a Leibnizian flavour. It attempts to "explain" why there exist different preparation procedures, or equivalently measurement effects, that cannot be unambiguously discriminated [53]. This statement perhaps requires some more careful explanation.

If we have two pure states $|\psi\rangle$ and $|\phi\rangle$, where $|\langle\psi|\phi\rangle| \neq 1$, then unless they are orthogonal, $|\langle\psi|\phi\rangle| = 0$ there does not exist a quantum measurement that we can apply to distinguish them with certainty. However, there does exist a measurement that can be used to unambiguously discriminate between them with *some* probability: a POVM with elements given by $E_1 =$

$a|\psi^\perp\rangle\langle\psi^\perp|$, $E_2 = a|\phi^\perp\rangle\langle\phi^\perp|$, $E_3 = \mathbb{1} - E_1 - E_2$ for some $0 < a < 1$, and where $|\psi^\perp\rangle$ indicates some quantum state orthogonal to $|\psi\rangle$. This is a valid POVM, and allows us to sometimes tell which state was prepared: if we see E_1, we know we must have had a preparation of $|\phi\rangle$, and likewise an observation of E_2 tells us we must have had $|\psi\rangle$. Some of the time we will see E_3, in which case we will not know which state we were prepared. Our intuitions are borne out here; the closer to orthogonal $|\psi\rangle$ and $|\phi\rangle$ are, the higher the probability of us successfully distinguishing them.

Quantum mechanics contains preparations and measurements that are, in contrast, impossible to unambiguously discriminate [53]. Possibilistic noncontextuality, for preparations, says that if two preparations are impossible to unambiguously discriminate operationally, giving rise to the same experimental possibilities, then they are not able to be unambiguously discriminated even given ontological access, since their associated distributions have identical supports. The assumption of possibilistic noncontextuality gives an "explanation" for this phenomenon, saying that the reason they are operationally impossible to discriminate is that they are ontologically impossible to discriminate. This allows us to apply classical intuitions to the quantum scenario. The inability to perform unambiguous discrimination is interpretable as a property of the distributions over ontic states themselves, rather than there being a lack of sufficiently fine-grained measurements, either experimentally available or ontologically available. What we have called Hardy noncontextuality, on the other hand, lacks an explanation of this Leibnizian type due to having a mismatch of possibilistic and probabilistic conditions.

We can see that the notion of Hardy noncontextuality applied here is weaker than both the notion we gave earlier of Probabilistic Noncontextuality, and also the notion we gave above of Possibilistic Noncontextuality. We can see that it is weaker than Probabilistic noncontextuality, since we use the exact same relation to identify ontological elements, $\overset{\mathbf{Prob}}{\sim}$, but we demand a less strict relationship between the identified elements.

Formally, we see that Probabilistic noncontextuality implies Hardy noncontextuality because

$$\mathcal{P}_1 \overset{\mathbf{Prob}}{\approx} \mathcal{P}_2 \Rightarrow \mathcal{P}_1 \overset{\mathbf{Poss}}{\approx} \mathcal{P}_2. \tag{2.23}$$

	⟵ Stronger Operational Relation ⟵	
⟶ Stronger Ontological Relation ⟶	Probabilistic (Spekkens) Noncontextuality $\mathcal{P}_1 \overset{\textbf{Prob}}{\sim} \mathcal{P}_2 \Rightarrow \mu_{\mathcal{P}_1} \overset{\textbf{Prob}}{\approx} \mu_{\mathcal{P}_2}$	Contradiction in any nontrivial convex theory
	Hardy Noncontextuality $\mathcal{P}_1 \overset{\textbf{Prob}}{\sim} \mathcal{P}_2 \Rightarrow \mu_{\mathcal{P}_1} \overset{\textbf{Poss}}{\approx} \mu_{\mathcal{P}_2}$	Possibilistic Noncontextuality $\mathcal{P}_1 \overset{\textbf{Poss}}{\sim} \mathcal{P}_2 \Rightarrow \mu_{\mathcal{P}_1} \overset{\textbf{Poss}}{\approx} \mu_{\mathcal{P}_2}$

Figure 2.1: A square showing the different contextuality assumptions, the strongest of which is contradictory in any convex theory. If convexity is not demanded, there are multiple theories that fulfil the stronger form of noncontextuality indicated as a contradiction above, such as quantum mechanics restricted to pure states and unitary operations. This figure is taken from Simmons *et al* [93]. Entries in the table logically imply entries that are to the left of them, and also entries that are directly below them.

The same is true for measurement effects, and also transformations. Similarly, we can see that Possibilistic noncontextuality implies Hardy noncontextuality, since

$$\mathcal{P}_1 \overset{\textbf{Prob}}{\sim} \mathcal{P}_2 \Rightarrow \mathcal{P}_1 \overset{\textbf{Poss}}{\sim} \mathcal{P}_2. \tag{2.24}$$

We might initially assume that Hardy noncontextuality is *strictly* weaker than Possibilistic noncontextuality. They include the same ontological relations, but possibilistic noncontextuality has a operational relation that is coarser grained, and has larger equivalence classes. Any two density matrices with the same support (meant in the linear-algebraic sense, so equivalently with the same kernels) are *possibilistically* equivalent but no two distinct density matrices are *probabilistically equivalent.*

Surprisingly however, when applied to quantum mechanics the two notions defined above of Hardy and of Possibilistic noncontextuality coincide.

Theorem 1. *An ontological model of finite-dimensional quantum mechanics is possibilistically noncontextual if and only if it is Hardy noncontextual.*

Proof. First, we will prove this theorem for preparations, and then for measurements. For transformations, see Simmons *et al* [93]. Note that any two preparation procedures that result in the same density matrix being prepared must have the same ontic supports under an assumption of possibilistic, probabilistic, or Hardy noncontextuality. Hence, we will use the notation $\mathcal{S}(\mu_\rho) \equiv \mathcal{S}(\mu_\mathcal{P})$ for any $\mathcal{P}$ that results in a preparation of ρ. From Equation 2.24, possibilistic preparation noncontextuality implies Hardy preparation noncontextuality: any elements related under the probabilistic operational relation are also related under the possibilistic operational relation. So, we only need to prove the converse, that is, that Hardy noncontextuality implies that density matrices with the same kernel can be represented by distributions with the same support over Λ.

Let ρ_0 and ρ_1 be density matrices with the same kernel and with smallest nonzero and largest eigenvalues $\alpha_{j,\min}$ and $\alpha_{j,\max}$ for $j = 0, 1$ respectively. We define

$$\sigma_0 = \frac{1}{1 - \frac{\alpha_{0,\min}}{\alpha_{1,\max}}} \left(\rho_0 - \frac{\alpha_{0,\min}}{\alpha_{1,\max}} \rho_1 \right). \tag{2.25}$$

We see that σ_0 is positive semi-definite, as the largest eigenvalue of $\frac{\alpha_{0,\min}}{\alpha_{1,\max}}\rho_1$ is $\alpha_{0,\min}$ and ρ_0 and ρ_1 have the same kernel. We note that we must have that $\alpha_{0,\min} \leq \alpha_{1,\max}$, since they have support on subspaces of equal dimensions, say, d, and so both ρ_1 and ρ_0 have d nonzero eigenvalues that sum to 1. Hence, we have $\alpha_{0,\min} = \alpha_{1,\max}$ only in the case that each is the completely mixed state over their support, and $\alpha_{0,\min} < \alpha_{1,\max}$ otherwise. It is easily verified that $\mathrm{Tr}(\sigma_0) = 1$, so σ_0 is a density matrix.

Hence ρ_0 can be prepared as a convex mixture of ρ_1 and σ_0 as

$$\rho_0 = \left(1 - \frac{\alpha_{0,\min}}{\alpha_{1,\max}}\right) \sigma_0 + \frac{\alpha_{0,\min}}{\alpha_{1,\max}} \rho_1. \tag{2.26}$$

Therefore there exists a preparation procedure that prepares ρ_0 such that

$$\mathcal{S}(\mu_{\rho_0}) = \mathcal{S}(\mu_{\sigma_0)} \cup \mathcal{S}(\mu_{\rho_1}) \supseteq \mathcal{S}(\mu_{\rho_1}) \tag{2.27}$$

We can then repeat the argument for ρ_1. Therefore $\mathcal{S}(\mu_{\rho_0}) = \mathcal{S}(\mu_{\rho_1})$ for any ρ_0, ρ_1 with the same kernel, completing the proof for preparations.

For measurements, it is again clear that possibilistic measurement noncontextuality implies Hardy measurement noncontextuality, so we only need to prove the converse.

Quantum mechanically, POVM elements are positive semidefinite matrices, and the assumption of possibilistic measurement noncontextuality is the identification of the ontic supports of any two such preparation procedures whose associated density matrices share the same kernel over $\mathscr{H}_2$. The proof here echoes the proof above strongly: if we have two matrices E_1 and E_2, that share a kernel, and correspond to two measurement effects, we can write $E_1 = aE_2 + bE_3$, for E_3 some matrix with a support strictly contained within that of the other E_i. In a similar fashion to the preparations case, we can construct a POVM element with the same matrix as E_1 as a convex sum of E_2 and E_3, demonstrating that $\mathcal{R}(E_1) = \mathcal{R}(E_2) \cup \mathcal{R}(E_3)$. We also have, by symmetry $\mathcal{R}(E_2) = \mathcal{R}(E_1) \cup \mathcal{R}(E_3)$, and therefore $\mathcal{R}(E_1) = \mathcal{R}(E_2) = \mathcal{R}(\mathbb{1}_{\mathrm{supp}(E_i)})$. This completes the proof. □

We note that in this proof, we are explicitly considering a coarse-graining of two measurement outcomes via classical post-processing to be possibilistically identical to an ontological coarse graining of the relevant POVM elements. This assumption will be revisited in Chapter 4 and we will discuss the validity of the assumption in detail also in that Chapter. We note that we can only apply our noncontextuality assumptions to conclude that this kind of coarse-graining is admissible only during strict prepare-measure scenarios. For scenarios including state update or adaptive measurements that are explicitly time ordered (the observables measured are noncommuting), this will not follow since classical post-processing can be distinguished from higher-rank quantum measurements by successive measurements on the quantum system.

However, in the possibilistic setting, this equivalence is easier to motivate; as long as each of the measurement outcomes that are coarse-grained together via classical post-processing have the same operational support. In this case, it can easily be seen that the two measurement procedures do in fact lead to the same possibilistic statistics on further measurement. It is easy to vary the proof above so that this relationship holds, and ρ_0, ρ_1, σ_0 and σ_1 all have the same support; we need only perturb the coefficient of ρ_1 in the definition of σ_0 by some small $\epsilon < {}^{\alpha_{0,\min}}/_{\alpha_{1,\max}}$; that is:

$$\sigma_0 = \frac{1}{1+\epsilon-\frac{\alpha_{0,\min}}{\alpha_{1,\max}}}\left(\rho_0 - \left(\frac{\alpha_{0,\min}}{\alpha_{1,\max}} - \epsilon\right)\rho_1\right). \tag{2.28}$$

The theorem above was demonstrated for finite-dimensional quantum mechanics; however the argument above can be used in greater generality and applied to theories outside of quantum mechanics.

Corollary 1. *An ontological model of any physical theory whose predictions can be reproduced within finite-dimensional quantum mechanics is possibilistically noncontextual if and only if it is Hardy noncontextual. In particular, this includes all scenarios with a finite number of preparations and measurements.*

As alluded to above, possibilistic versions of operational theories appear in the setting of Bayesian Inference, when considering consistency conditions between the beliefs of rational agents. Brun, Finkelstein and Mermin (BFM) [22]; and Caves, Fuchs and Schack (CFS) [27]; brought these classical concepts into the quantum realm with a set of definitions of what it meant for agents' quantum state assignments to be consistent. Two important classical notions are those of *concordance* and *strong consistency*. If different agents' assignments of probabilities to possible ontic states are *concordant*, then the set of states that are given a nonzero credence are the same. This means that for any measurable quantity, they agree on what set of outcomes are possible. If their probability assignments are *strongly consistent*, then there exists some state, or more accurately subset of states) that they all assign a positive probability to. Being a rational agent in a Bayesian inference scenario when there is a common prior is strong enough

to force all agents to exactly agree on their probability assignments, by Aumann's agreement theorem [11]. However even when priors are allowed to vary, agents assignments can be argued to remain concordant[10]. The BFM consistency condition says that agents' assignments of the states ρ_i to the quantum system are consistent, if the intersection of their supports is nonempty; that is that there exists a projector to which each agent assigns some probability *via* the Born rule. The strongest condition considered by CFS was that each of the ρ_i has support on the same Hilbert subspace, that there is no projector that some agent assigns probability to but some others do not. That these quantum mechanical notions in some sense coincide with their classical counterparts under the assumption of possibilistic measurement noncontextuality can be seen from the fact that it is exactly this assumption that allows us to define an ontological support that is a function of only the density matrix involved and not sensitive to any details about how that density matrix was constructed. For agents to be BFM consistent means that there is a subspace to which each agent assigns some probability; under the assumption of possibilistic measurement contextuality they must each assign probability to those states in the support of mixed states on that subspace. Therefore they are classically strongly consistent. If agents are consistent under the CFS condition, then each agent's state assignment has support on the same Hilbert subspace, and therefore under the assumption of possibilistic preparation noncontextuality they agree on the ontological supports also, so they are classically concordant.

In conclusion, the formalism of equivalence-based contextuality offers a flexible framework to discuss issues of contextuality. Much of the more traditional structure, including outcome determinism, can be inferred from assumptions of "universal noncontextuality", in the case of sharp measurements, as seen above. As we will see in Section 4, it can also be used to test notions of contextuality under different ontological assumptions, such as the ψ-epistemic property. The ability of the framework to consider different preparation settings appears on the surface to be unique, but we shall see that notions of preparation contextuality can be made to work within other formalisms, by direct analogue of the ontological models framework, which was independently noted by Wester [105]. We shall also discuss the possibility of bringing over such a notion by considering preparations as the outcomes of nonlocal measurements, taking a

[10]This is the case only when they vary in probability assignments but not possibility assignments.

point of view more akin to steering.

2.4 The Sheaf-theoretic Approach

The generalisation of Fine's theorem by Abramsky *et al* [2] demonstrates that the existence of a noncontextual probability distribution over deterministic value assignments for a measurement scenario is equivalent to the existence of a factorisable probability distribution. We note that while this assumption is easy to motivate for nonlocality experiments, due to the spacelike separation that subsystems can achieve, it is harder to motivate for general noncontextuality experiments, which will be discussed in more detail later. Nonetheless, it provides a clear unification of nonlocality and contextuality, the former being a special case of the latter. One of the strongest motivators for these concepts of contextuality, that embrace determinism, is that they appear to have a direct operational significance [51].

Similarly to the equivalence-based forms of contextuality, the fundamental goal given any quantum scenario is to create a kind of hidden variable model that matches our theoretical, or experimental, probabilities. These are given to us, as before, in the form of a table of probabilities for outcomes, given the measurement selections. We will for the moment restrict ourselves to the case where we only have one preparation, and we will follow the notation of Abramsky and Brandenburger. We will use examples inspired by nonlocality, and will present them in a manner specifically suited to bipartite scenarios as in Reference [68].

The natural way to view measurements in this framework are as sets of mutually commuting, or in a general theory commensurable, observables. This is in contrast to the equivalence-based formalism, in which the natural view is that of a PVM or POVM, with outcomes that are taken to be mutually exclusive. For this view on contextuality, we will restrict our consideration to a specific measurement scenario which will be taken to have a fixed set X of measurements with outcomes $\mathcal{O}$, which we will typically take to be $\{0, 1\}$ but in general could be any finite set. This mirrors exactly the quantum case, in which any observable with more than two separate eigenspaces could be experimentally replaced with a set of commuting observables that only

	$b_1 = 0$	$b_1 = 1$	$b_2 = 0$	$b_2 = 1$
$a_1 = 0$	$1/2$	0	$3/8$	$1/8$
$a_1 = 1$	0	$1/2$	$1/8$	$3/8$
$a_2 = 0$	$3/8$	$1/8$	$1/8$	$3/8$
$a_2 = 1$	$1/8$	$3/8$	$3/8$	$1/8$

	b_1	b'_1	b_2	b'_2
a_1	1	0	1	1
a'_1	0	1	1	1
a_2	1	1	1	1
a'_2	1	1	1	1

Figure 2.2: Probability and Possibility tables for a CHSH [30] experiment. Alice has two available measurements: measurement A_1 with outcome a_1 and a'_1; and measurement A_2 with outcomes a_2 and a'_2. Bob has measurements denoted similarly. Typically, however, we will not label outcomes as the properties we are consider are invariant under relabelling.

have two eigenspaces.

In general, we will consider a scenario that is described by its set of measurements X, which *measurement contexts*, giving rise to *events*, and then we will wish to match these events to an "empirical model" via *distributions* on those events. We will slowly build up this picture by giving a rigorous definition of each of these concepts within the formalism. This formalism highlights the conceptual difference between the structure, or shape, of a measurement scenario and the probabilities predicted by a specific theory for these events to be observed. As such, it is possible to attribute certain forms of contextuality to the commensurability structure in a way that is completely "theory-independent".

Definition 11 (Section)**.** A *section* over a set of measurements $U \subseteq X$ is a function $s : U \to \mathcal{O}$.

We see that we can consider such a function to be an assignment of outcomes to each measurement and therefore that a section for a set of measurements U is similar to an effective ontic state that predicts measurement outcomes for that particular measurement set; it represents the event in which measurements U were performed, and the outcomes given by s were seen by experimenters. As we shall see later, we can think of such sections as ontic states in a deterministic hidden variable model for the measurements in U. We will now define an important categorial concept that is central to the structure of contextuality and nonlocality in this

setting.

Definition 12 (Presheaf)**.** Let X be a topological space, and $\mathbf{C}$ be a category, typically the category of sets **Set**. A *presheaf* F on $\mathbf{X}$ is a $\mathbf{C}-$valued functor such that each open set U of X has a corresponding $F(U) \in \mathbf{C}$. We have $F : \mathcal{P}(U)^{\mathrm{op}} \to \mathbf{C}$. Additionally, we must have for all $U \subseteq U'$ there exists a restriction map res such that $\mathrm{res}_U^{U'} : F(U') \to F(U)$ in the category $\mathbf{C}$, which obeys the following properties: $\mathrm{res}_U^U = \mathrm{id}_U$, and $\mathrm{res}_U^{U'} \circ \mathrm{res}_{U'}^{U''} = \mathrm{res}_U^{U''}$.

We will define $\mathcal{E} : U \mapsto \mathcal{O}^U$ which assigns to each measurement set its set of sections. This maps a subset of available measurements to every combination of outcomes of those measurements; this should put us in mind of a classical hidden variable. We can note that there is a natural concept of a restriction map, if we have $U \subseteq U'$, there is a restriction map given by

$$\mathrm{res}_U^{U'} : \mathcal{E}(U') \to \mathcal{E}(U) :: s \mapsto s|U. \tag{2.29}$$

This structure builds in an important notion of noncontextuality for our ontic states. Since this map has the properties that $\mathrm{res}_U^{U'} = \mathrm{id}_u$, and $\mathrm{res}_U^{U'} \circ \mathrm{res}_{U'}^{U''} = \mathrm{res}_U^{U''}$, and because of this we can consider $\mathcal{E}$ to be a functor $\mathcal{P}(X)^{\mathrm{op}} \to \mathbf{Set}$. That is, $\mathcal{P}(X)^{\mathrm{op}}$ is the category of open sets of X, and to say that $\mathcal{E}$ is a functor is to say that $\mathcal{E}$ maps $U \in \mathcal{P}(X)^{\mathrm{op}}$ to some $\mathcal{E}(U) \in \mathbf{Set}$, and also maps arrows $U \subseteq U'$ to $\mathcal{E}(U) \to \mathcal{E}(U')$, an arrow in the category of sets. We also note that $\mathcal{E}$ has the structure of a *presheaf*. However, we can demonstrate that $\mathcal{E}$ also has additional structure: the structure of a *sheaf*.

Definition 13 (Sheaf)**.** We will define a *sheaf* to be a presheaf F taking values in **Set** that also satisfies the following properties.

- A locality condition: if $\{U_i\}$ is an open covering of an open set U and two elements $s, t \in F(U)$ have $s|U_i = t|U_i$ for all U_i, then $s = t$.
- A gluing condition: if $\{U_i\}$ is an open covering of an open set U, and for each i we have a section s_i such that for all i, j the restrictions of s_i and s_j to the overlap of U_i and U_j agree, $s_i|U_i \cap U_j = s_j|U_i \cap U_j$, then there exists a section $s \in F(U)$ such that $s|U_i = s_i$.

That is, given a cover of U, when the set of sections s_i on the U_i agree locally on the overlaps between the U_i, then they can be seen as restrictions of a unique global section $s \in F(U)$ such that $s|U_i = s_i$. The existence of the global section comes from the gluing condition, and its uniqueness from the locality condition. The name "locality" here is a category theory term, it is important perhaps to note that this does not have any physical significance. We can interpret these conditions as saying that unique global data assignment can be derived from local data which agrees on overlaps. It is not difficult to prove that $\mathcal{E}$ also obeys these two conditions, so we call $\mathcal{E}$ the "sheaf of events". These two conditions complete the notion of noncontextuality for the individual ontic states in the formalism. These properties cement the interpretation of global sections of the event sheaf as deterministic hidden variables.

As with the ontological formalism above, we require the probabilities associated with events to live within a commutative semiring R. As before, three semirings with immediate relevance to quantum mechanics are $\mathbb{R}_+$, which captures classical notions of probability; $\mathbb{R}$, which recovers the concept of quasiprobability distributions and Wigner functions; and the Boolean semiring, which as before allows us to capture notions of possibilistic contextuality. Given a semiring, R, an R-distribution is a function $d : X \to R$ with finite support and $\sum_{x \in X} d(x) = 1$. The set of R-distributions on X is called $\mathcal{D}_R(X)$. We can also extend $\mathcal{D}_R$ to a functor by defining its action on arrows in a way that captures and generalises the notion of coarse-graining outcomes (for example by marginalising over outcomes at a different subsystem). Given a function $f : X \to Y$, defining

$$\mathcal{D}_R(f) : \mathcal{D}_R(X) \to \mathcal{D}_R(Y) :: d \mapsto \left(y \mapsto \sum_{x:\, f(x)=y} d(x) \right). \tag{2.30}$$

We can think of this functor as offering a kind of coarse-graining; moving from a set X to a coarser Y, and identifying outcomes x which share an outcome after application of f. This can be composed with the event sheaf $\mathcal{E} : \mathcal{P}(X)^{\mathrm{op}} \to \mathbf{Set}$, forming a functor $\mathcal{D}_R\mathcal{E} : \mathcal{P}(X)^{\mathrm{op}} \to \mathbf{Set}$. We can see that $\mathcal{D}_R\mathcal{E}$ is a presheaf by considering its functorial action. When $U \subseteq U'$, we define

$$\mathcal{D}_R\mathcal{E}(U') \to \mathcal{D}_R\mathcal{E}(U) :: d \mapsto d|U \quad \text{s.t. } d|U(s) := \sum_{s' \in \mathcal{E}(U'), s'|u = s} d(s'). \tag{2.31}$$

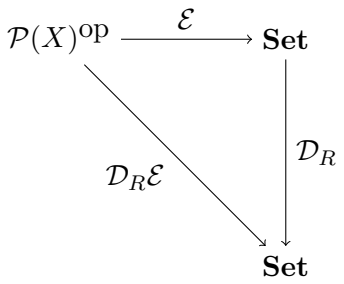


Figure 2.3: We define the presheaf $\mathcal{D}_R\mathcal{E}$ by composition of the functor $\mathcal{D}_R$ with the event sheaf $\mathcal{E}$. This assigns to sections the set of distributions over U-sections, capturing the full set of operational theories compatible with our deterministic, noncontextual ontology.

This object assigns to each set of measurements U the set $\mathcal{D}_R(\mathcal{E}(U))$, which is the set of all R-distributions over sections of U. This is the full set of possible distributions over the sections of U and therefore captures all probability distributions over the outcomes of U that are noncontextual.

A crucial part of our formalism, which will cover nonclassical scenarios, is that there are sets of measurements that are not commensurable. When considering measurement subsets U, we have until now not considered whether or not each individual U actually contains a set of measurements that can be simultaneously measured. In order to describe this structure, we introduce the notion of the measurement cover.

Definition 14 (Measurement cover)**.** A *measurement cover* $\mathcal{M}$ on the set X is a family of subsets of X such that $\bigcup \mathcal{M} = X$, and $\mathcal{M}$ is an anti-chain: $\mathcal{C} \subseteq \mathcal{C}' \in \mathcal{M} \Rightarrow \mathcal{C} = \mathcal{C}'$.

The anti-chain condition is not strictly neccesary, but reflects the definition of a context to be a *maximal context*; there are no more commensurable information-granting measurements that could be added. It also means that given the entirety of quantum mechanics, every context is infinite, containing all classical post-processings of data within the context itself. As an example, if a particular observable exists within a context then every other observable sharing the same eigenspaces will also be in the context. This formalises the idea that observables within each context $\mathcal{C} \in \mathcal{M}$ are simultaneously measurable, and that each *context* $\mathcal{C}$ is a compatible set of measurements. The necessary condition for contextuality, identical objects appearing in

different contexts, is specified by the compatibilities of measurements within the scenario under consideration, independently of any specific theory. We can note that for classical scenarios, $\mathcal{M} = X$, since every measurement is simultaneously commensurable.

We now introduce the concept of an *no-signalling empirical model* as a compatible family, *i.e.* locally consistent family in the sense of the gluing condition for sheaves, for the cover $\mathcal{M}$ with respect to the presheaf $\mathcal{D}_{\mathcal{R}}\mathcal{E}$, which assigns to each set of measurements U the set of distributions over U-sections. For each context $C \in \mathcal{M}$, there is a distribution $e_C \in \mathcal{D}_R\mathcal{E}(C)$ a distribution over the noncontextual ontic states for the context C. This family of distributions must be compatible in the sense of the gluing condition for sheaves:

$$e_C|C \cap C' = e_{C'}|C \cap C' \quad \forall C, C' \in \mathcal{M}. \tag{2.32}$$

This sheaf gluing condition, or presheaf condition, can be thought of as capturing what is often referred to as the no-signalling restriction. However, as mentioned before, it is exactly such a restriction that empowers us to represent quantum measurements with context dependent symbols, and is necessary to endow any physical significance to our choice to identify elements of X between different measurement contexts C. We will only consider empirical models that have this property.

We are concerned with whether for a specific $\{e_C\}_{C\in\mathcal{M}}$ there exists a *global section* $d \in \mathcal{D}_R\mathcal{E}(X)$. We recall that a global section for $\mathcal{E}$ is a deterministic value assignment for all the measurements in our scenario, X; a global section for $\mathcal{D}_R\mathcal{E}(X)$ is then an R-distribution over such global sections for $\mathcal{E}$. Such global sections for $\mathcal{D}_R\mathcal{E}(X)$ can then be thought of as distributions over the admissible hidden variables for the whole scenario. This distribution must restrict to yield the probabilities for measurement outcomes in all contexts predicted by the empirical model: for all $C \in \mathcal{M}$ $d|C = e_C$. The existence of such a global section is equivalent to noncontextuality for a scenario, regardless of whether $\mathcal{D}_R\mathcal{E}$ is a sheaf.

The theory independence of the framework, combined with the fact that the structure lends itself most easily to the view of measurements as a set of mutually commuting observables

	$b_1 = 0$	$b_1 = 1$	$b_2 = 0$	$b_2 = 1$
$a_1 = 0$	$1/3$	0	0	$1/3$
$a_1 = 1$	0	$2/3$	0	$2/3$
$a_2 = 0$	0	0	0	0
$a_2 = 1$	$1/3$	$2/3$	0	1

$=$

$$\frac{1}{3}\left(\begin{array}{c|cc|cc} & & & & \\ \hline & 1 & 0 & 0 & 1 \\ & 0 & 0 & 0 & 0 \\ \hline & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 1 \end{array}\right) + \frac{2}{3}\left(\begin{array}{c|cc|cc} & & & & \\ \hline & 0 & 0 & 0 & 0 \\ & 0 & 1 & 0 & 1 \\ \hline & 0 & 0 & 0 & 0 \\ & 0 & 1 & 0 & 1 \end{array}\right)$$

Figure 2.4: A table of possibilities that does not exhibit noncontextuality/nonlocality; it can be decomposed into a sum of deterministic grids that obey no-signalling.

means that in some situations, there can be some ambiguity as to whether impossible events ought not to have been considered part of the structure of the scenario. One example is that there may be events that are quantum mechanically impossible that might not be impossible in another physical theory, or for a different realisation of the same measurement compatibility structure. This is not in and of itself a problem for the framework, although it does mean that maximal contextuality often does not arise out of the compatibility structure itself, but rather from the events that are impossible given a particular empirical model.

Chapter 3

Common Ground and Differences Between Approaches

The equivalence approach and the sheaf-theoretic approach have much in common, especially with regards to their physical content rather than the mathematical machinery used to express those ideas. While the two approaches differ greatly in terms of their philosophical approach, the technical differences of which scenarios are considered by each to be contextual is determined by their position on outcome-determinism. In this chapter, we will explore this assumption of outcome-determinism, especially as applied to unsharp measurements; the ability to define contextuality as applied to such POVM measurements is a feature of the Equivalence formalism, but we shall argue is not always well-motivated philosophically. Another feature of the Equivalence formalism is being able to talk about preparations being contextual, rather than just measurements; we will explore the challenge of bringing this notion into the Sheaf-theoretic approach, both by considering deterministic ontological models and also by considering scenarios motivated by quantum steering.

3.1 Unsharp Measurements and Determinism

There is an inherent tension between unsharp measurements and the axiom of outcome determinism, as discussed by Spekkens [96]. In quantum mechanics, proper POVMs are irreducibly unsharp when considering their action on single systems. For instance, we can note that POVM elements with eigenvalues all less than or equal to $1/2$ can appear in POVMs alongside another copy of themselves, which is inherently contradictory under outcome determinism since the outcomes of two POVM elements are strictly exclusive within a single context. If the POVM element has a noncontextual valuation of 1 or 0, then if such a POVM element had a 1-valuation, then both copies of it within the context would have a 1-valuation. Therefore, such POVM elements can only take 0-valuations, which is manifestly contradictory to the probabilistic predictions of quantum mechanics. One partial resolution is to assume outcome determinism only for sharp measurements, or for sharp measurements and POVMs with an eigenvalue strictly greater than $1/2$ [73]. In addition, a set of POVMs cannot always be thought of as being in bijection with a set of PVMs on a larger Hilbert space, in which contexts are preserved by bijection in the sense that a POVM resolution of the identity is mapped to a PVM resolution of the identity. One trivial counterexample for this is exactly that used above, the same POVM element appearing twice in one context. However the impossibility can be seen to be structural, since PVMs obey Specker's principle [25] and POVMs do not. That is, if a set of projectors are pairwise commensurable then they are jointly commensurable, but a set POVM elements that are pairwise commensurable, summing to less than the identity, can together sum to an operator with an eigenvalue of greater than 1. That such a mapping between POVMs and PVMs is impossible strongly suggests a possible ontological distinction between measurement outcomes represented by the same POVM element, even though these share the same operational statistics. We must ask whether or not applying the concept of noncontextuality to POVMs in such circumstances is of itself naïve.

Spekkens points out that a single POVM has multiple Naimark dilations, and each of these Naimark dilations would be considered its own measurement, distinguishable with different preparations on the ancilla system [96]. He argues that this is problematic for the claim that

POVM elements should be able to be assigned a single deterministic outcome, regardless of context. The issue arises when trying to combine two innocuous seeming statements about POVMs and noncontextuality in quantum mechanics: firstly, two measurement effects in the same operational equivalence class in quantum mechanics are represented by the same POVM element on the system in question; secondly, every POVM element can be represented by a PVM on a larger Hilbert space comprising of the system and a suitably-prepared ancillary system. Since we have ceded that it is not unreasonable to demand determinism of projector valued measures, it seems then to not be unreasonable to demand this of POVMs that are not projective. However, since the Naimark dilations are not unique, we cannot use an argument from noncontextuality to demand that the two dilations be treated the same. We either take the two unsharp POVM elements, and accept that those must be treated identically by noncontextuality, or we allow the dilation but lose the ability to apply the noncontextuality assumption to the dilated measurements.

This argument certainly holds true within quantum mechanics taken as a whole. However, this objection does not necessarily apply in full to quantum subtheories or specific measurement scenarios. In this setting, the idea of using out-of-scope measurements to differentiate between different POVM dilations on the ancilla is questionable. Let us assume we have a set of measurable POVM elements $\{E_i\}$, which can be made to form various complete POVM contexts. We suggest that it is only unwarranted to take the POVM elements $\{E_i\}$ as deterministic if it is *impossible* to dilate each of the POVM elements to a unique projector, in a manner that is independent of the context in which they appear. We note that this will still be, in some sense, a violation of noncontextuality for unsharp measurements; there are other ways we could have dilated these elements, and we are not claiming that our representation is independent of which dilation we use. The exact opposite, in fact: we are identifying these POVM elements exactly because we can dilate them consistently to the same PVM element. We are, in effect, claiming a noncontextuality assumption for PVMs only, and accepting that unsharp measurements need not obey the notion of noncontextuality. However, to try and associate the outcomes of POVMs in situations where their dilations are necessarily contextual is not well-founded. The form of the

necessary and sufficient conditions for this to be possible is an open question, however for sets of POVM elements which are closed in the sense that for all $E \in \{E_i\}$, we have $\mathbb{1} - E \in \{E_i\}$, certainly the condition that each POVM element must contain at least one eigenvalue that is greater than $1/2$ is necessary to avoid the case where the same POVM element appears twice in the same POVM. As mentioned above, this immediately rules out the possibility that each POVM element can have a noncontextual dilation.

An illustration of the differences between the two approaches can be illustrated by appeal to the works of Specker [25]. In Specker's parable, an "overprotective seer" has a daughter, with many suitors. The seer has three boxes, each of which may or may not contain a gem. A prospective suitor must guess which boxes contain gems, and which do not; in any such assignment there must be two boxes that either are claimed to both contain gems or to both not, and the seer invites the suitor to open these. Each suitor is surprised to find exactly one gemstone within the two boxes, and never the zero or two he predicted.

This scenario certainly seems nonclassical, and to exhibit contextuality. The subtext of the story puts us in mind of a classical assignment of values to the different boxes, perhaps because instinctively we note that the opening of different boxes should give rise to a factorisable probability distribution. The general structure of the scenario is as follows: we have some odd number of measurements events, say 3, which have a cyclic compatibility structure. Each pair of adjacent measurements is commensurable, and we find that when a pair is measured, each measurement event occurs exactly $1/2$ of the time. We cannot measure all three simultaneously; already this should itself be enough to make us suspicious, as epistemic restrictions are a hallmark of nonclassicality.

If we take the sheaf-theoretic view, and try to find deterministic hidden variables to mimic this behaviour, we find, just as in the original story, that this is a contextual situation. However, this is not the position taken by the equivalence approach. In the equivalence formalism, each of the three measurement effects is identified, since they all have the same statistics. Since the respoinse function $\xi(\lambda)$ must be able to compose exactly with itself, and two copies of it forms a complete event set, we have the restriction that $\xi(\lambda) + \xi(\lambda) = 1 \ \forall \lambda \in \Lambda$ from the ontological

models axioms, so $\xi(\lambda) = 1/2\ \forall\lambda \in \Lambda$ follows immediately. This is manifestly not contextual, since there is no explicit or implicit context dependence, only a constant function, and we meet notions of preparation noncontextuality trivially, since there is only one preparation in our theory. Therefore this scenario is not considered to be contextual under the equivalence formalism, and we have a tension between the two approaches. Unsurprisingly, POVMs also pose an issue in the Cabello-Severini-Winter framework for contextuality, which also definitionally restricts itself to projective measurements since orthogonality graph approaches have Specker's principle as a central part of their structure. Trying to extend it naïvely to POVMs, as pointed out by Kunjwa l [61], is problematic since trivial POVMS which are merely resolutions of the identity operator can be made to maximally violate the framework's supposed nonclassicality inequalities. Both approaches to this problem have arguments in their favour. Let us consider that we are presented with a box, with three measurement settings corresponding to opening two of the boxes, and lights which indicate which outcome occurred for our chosen context. When we choose a context, we see that one and only one of the two measurement events in the context occurs, each with probability 1/2. It would be strange to conclude in this setting that there were anything unusual going on, and definitely not something particularly nonclassical. To amend a quotation from Pusey, if it turns out that inside the box is just a man flipping a coin, we had better not have concluded that what was happening inside was outside the realm of classical physics.

In this instance, when we do not have sharp measurements to motivate the notion of outcome determinism, it seems difficult to resolve the idea of contextuality with this behaviour if we want to consider contextuality to be a signal of nonclassicality. The counter to this argument is that there *must* be something contextual happening; if the situation were that the box really *does* contain a little man who flips a coin to determine experimental outcomes, he cannot be doing this in a way that is completely noncontextual. If in the context of measurement settings $\{1, 2\}$ a heads denotes outcome 1 and a tails denotes outcome 2; and in the setting $\{2, 3\}$ a heads denotes outcome 3 and a tails denotes outcome 2; in the context $\{1, 3\}$ we must have one outcome attached to a tails, demonstrating a contextual process. This problem can be

easily seen to persist even with more complicated state spaces for the same problem, since the contextuality of the scenario in the sheaf setting shows that any classical state space, with associated factorisable probability distributions cannot reproduce the statistics of the scenario. We must rely on actual, inherent randomness in order to have a noncontextual explanation, and this is in and of itself a hallmark of nonclassicality. We have to decide, however, whether we want this apparently trivial form of nonclassicality to be part of the fold of our theory of contextuality, or to be noted, but not to be considered fundamentally contextual.

3.2 Kitchen Sink Models

As discussed, one of the key differences between these models of contextuality is the treatment of determinism, and through determinism the status of intrinsic randomness; in particular we are concerned with whether or not intrinsic randomness is considered nonclassical, especially when considering POVMs. The *kitchen sink model* is a folkloric ontological model that is deterministic, and can be ψ-epistemic or ψ-ontic depending on the specific measurements and preparations available. As an obvious way to construct a deterministic ontological model, it crops up throughout the literature [49], originating with Bell [15], although the concept was not initially fully fleshed-out.

The kitchen sink model for an operational theory $[\mathcal{P}, \mathcal{M}, \mathcal{T}, \mathrm{Pr}]$ is a particular ontological model defined on the state space of functions $\lambda : \mathcal{M} \to \mathcal{O}_{\mathcal{M}}$, deterministic assignments of outcomes to measurements.

The intuition behind the model is that when a system is prepared, an outcome is selected for each of the measurements that could be made on the system, at random, according to the statistics of that measurement. These outcomes are chosen independently, and as such are highly contextual; knowing the outcome of one measurement yields no information about other measurements. The response functions associated with the outcomes of the measurement just look at the recorded outcome for their measurement, and are 1 if their outcome is the one in the ontic state and 0 otherwise. As such, the model is outcome-deterministic.

For example, let us consider the case in which we have one preparation and two measurements $\mathcal{M}_1$ and $\mathcal{M}_2$, each of which has two outcomes. The statespace has four elements, since there are four different value assignments to $\mathcal{M}_1$ and $\mathcal{M}_2$ jointly. The preparation distribution is then defined to be:

$$\mu(\lambda) = \Pr(\lambda(\mathrm{M}_1)|\mathcal{P}, \mathrm{M}_1) \times \Pr(\lambda(\mathrm{M}_2)|\mathcal{P}, \mathrm{M}_2), \tag{3.1}$$

The probability of being in state λ is the probability of each of the measurements having the otucome attributed to them by λ. So, were we to have $\lambda(\mathcal{M}_1) = 0$, $\lambda(\mathcal{M}_2) = 1$, and the probability for these outcomes were $1/2$ and $1/3$ respectively, then λ would be given a probability of $1/6$. The response functions are then defined to be $\xi_{\mathcal{M}}(k|\lambda) = \delta(k, \lambda(\mathcal{M}))$.

There is clearly no obstacle to defining such a model for any finite set of measurements, however we must when we are dealing with an infinite set of measurements, even in finite dimensional quantum theory. We need to be careful when defining our preparation distributions to check that they do indeed form a measure. Following a construction of Barrett [13] for the specific case of quantum theory, let us define $S_i^{\mathcal{M}} \subset \Lambda$ as the support of the measurement effect $i|\mathcal{M}$; that is the set of λ such that $\lambda(\mathcal{M}) = i$. Then, we let $\mathcal{A}$ be the algebra of subsets of Λ generated by the $S_i^{\mathcal{M}}$ under finite intersections, unions, and complements. We now define a premeasure ρ_P on $\mathcal{A}$. Since any element $a \in \mathcal{A}$ can be written as a finite union of finite intersections of $S_i^{\mathcal{M}}$, it is enough to specify the premeasure on such an a. So, writing

$$a = \bigcup_{i=1}^{N_1} \bigcap_{j=1}^{N_2} S_{k_{i,j}}^{\mathcal{M}_{i,j}} \tag{3.2}$$

we set

$$\rho_P(a) = \sum_{i=1}^{N_1} \prod_{j=1}^{N_2} \Pr\left(\mathrm{k_{i,j}}|\mathcal{P}, \mathcal{M}_{\mathrm{i,j}}\right) \tag{3.3}$$

We note while this decomposition of a may not be unique, $\rho_P(a)$ is unique and well defined since given any two such decompositions, we can consider a common fine-graining of the two. Then, each can be recovered by varying the order of summation so each of the two decompositions

must have the same measure. Now, we need only let Σ be the sigma algebra generated by $\mathcal{A}$ and define for any $E \in \Sigma$:

$$\mu_P(E) = \inf \left(\sum_{j \geq 0} \rho_P(a_j) : a_j \in \mathcal{A}, E \subset \bigcup_{j \geq 0} a_j \right). \tag{3.4}$$

This is a valid probability measure which can be shown to be the unique measure on Σ whose restriction to the algebra $\mathcal{A}$ is ρ. It is easy to see that this model reproduces the operational predictions of the theory $[\mathcal{P}, \mathcal{M}, \mathcal{T}, \Pr]$.

We then define our response functions as before by $\xi_{\mathcal{M}}(k|\lambda) = \delta(k, \lambda(\mathcal{M}))$, following the interpretation of the response functions as revealing the measurement outcome determined by the state λ.

Having defined this implementation of the Kitchen Sink model for such an operational theory, we will define a generalisation of the concept as follows:

Definition 15 (Generalised Kitchen Sink Model)**.** An ontological model is a *Generalised Kitchen Sink Model* if its state space and endowed σ-algebra are the same as that of a Kitchen Sink Model.

This is something of a broad definition; we can note that any outcome-deterministic ontological model can be considered a fine-graining of a generalised Kitchen Sink Model. However this name does capture structural similarities with a well-known ontological model.

Noncontextuality emerges operationally *via* the identification of outcomes or coarse-grainings of outcomes that have equal probabilities regardless of the specific preparation used.

The Kitchen Sink Model is, perhaps, a not a good candidate for the ontology of the universe for multiple reasons, both concrete and aesthetic. The state space Λ is exceptionally large, compared both to the naïve ψ-complete ontology of Beltrametti and Bugajski [18]; and also to the deterministic extension of Bell's first model [16]. Additionally, as was first pointed out to the author by Wallman, when the allowed preparations and measurement are such that the

model is ψ-epistemic, there are no possible measurement update rules for the ontic states that are in agreement with quantum theory; this will be covered in more detail in Chapter 4.

It is apparent when considering the basic objects for the Kitchen Sink model and the sheaf-theoretic framework that they have sprung from very different views about what it means to be contextual. The definition of contextuality in the ontological models formalism takes an operational theory first, drawing conclusions about which outcomes, or coarse-grainings of outcomes, can be identified from probabilistic cues; the sheaf-theoretic definition takes first the measurement scenario, which by fiat declares that certain outcomes, or coarse-grainings of outcomes, should be identified. The operational theory used in the equivalence formalism can be thought of as defining both a measurement scenario and an empirical model; a unique measurement scenario can be recovered from an operational theory $[\mathcal{P}, \mathcal{M}, \mathcal{T}, \mathrm{Pr}]$, but many are many operational theories are consistent with any given measurement scenario. This property makes it easier to clearly define theory-independent, or model-independent, definitions of contextuality in the sheaf formalism [2].

We would like to recover the theory-independence of the sheaf-theoretic mode within the ontological models formalism. Since any noncontextual measurement scenario has a representation as a noncontextual generalised kitchen sink model, then this can be achieved. By quantification over all operational theories that give rise to the same space of noncontextual generalised Kitchen Sink models we will be able to recover the theory-independence of the sheaf approach within the ontological models setting. This concept could in theory then be applied to models outside the (deterministic) scope of the sheaf-theoretic framework.

We must define, then, mappings between the two settings that allow us to translate over machinery. First, we note that the full sheaf of events $\mathcal{E}$ can be interpreted as Λ, the ontic statespace for a generalised kitchen sink model.

We then identify which elements of our ontological model we will consider to be the same, but appearing in different contexts, by fiat from the measurement scenario. We see, then, that this allows us to define an analogue of preparation noncontextuality for the formalism by

enforcing a noncontextuality relation *via* the introduction of a functional composition principle, analogously to the functional composition principle for measurements that yields the Kochen-Specker theorem. However, it should be noted that this notion of preparation noncontextuality in the sheaf formalism coincides with that of the ontological models formalism only for the restricted class of noncontextual generalised kitchen sink models detailed here.

What, then, about maximal contextuality? Our generalised kitchen sink constructions all explicitly carry the semiring $\mathcal{R}$ with them, and so there does not seem to be an obvious way in which we can bring this notion over into the ontological models setting. While it may be unsatisfactory, we can see that it corresponds to this statement: a measurement scenario is "maximally contextual" if any generalised kitchen sink model that has the same measurement scenario contextual. If there were a single ontic state compatible with the requirements of a noncontextual generalised kitchen sink model, the operational theory of that one ontic state would be a valid one. Since the existence of a single global section implies that there is such an ontic state, we see that this is the form that maximal contextuality takes in this framework. Again, a degree of theory-independence has been recovered by quantification.

3.3 Preparation Noncontextuality

As noted above, the global sections of the sheaf formalism are effectively determinstic, noncontextual ontic states. By thinking of them as such, it is possible to bring the notion of preparation noncontextuality into the sheaf formalism *via* this parallel. Each preparation is considered as giving rise to their own empirical models, and if this model is noncontextual, distributions over global sections. With a prepare and measure scenario, we can recover a notion of *universal contextuality*, the combination of measurement and preparation noncontextuality in this way.

We need to cover the case of multiple preparations, not usually contained within the sheaf approach. We can do this by considering the case where we have multiple different empirical models, families $E_i = \{e_C^{(i)}\}_{C \in \mathcal{M}}$ compatible with a common set of contexts $\mathcal{M}$. As part of our scenario, we take a set of empirical models of interest E, which is a no-signalling subset

of the full set of empirical models $\prod_{C\in\mathcal{M}} \mathcal{D}_R\mathcal{E}(C)$. We can associate each of these compatible families with some operational preparation. As mentioned in Chapter 2, the existence of a noncontextual deterministic hidden variable model for E_i depends on whether there is a global section $d \in \mathcal{D}_R\mathcal{E}(X)$ compatible with it. We can consider then a map $\mathcal{G} : E \to \mathcal{D}_R\mathcal{E}(X)$, such that $\mathcal{G}(E_i)|C = e_C^{(i)}$ for all $C \in \mathcal{M}$, which is simply to say that each of the empirical models are mapped to a global section of $\mathcal{D}_R\mathcal{E}(X)$ that recovers the empirical model in question when restricted to each context. The existence of such a map is equivalent to a statement that for each of the empirical models E_i has a noncontextual representation.

We can now borrow a notion of equivalence and define a notion of preparation noncontextuality by demanding a convex structure on $\mathcal{G}$. We will define some new notation for R-mixing empirical models E_i. For a countably supported $R-$distribution $r : E \to R$, writing $r_i := r(E_i)$ with $\sum_i d(E_i) = 1$, we define:

$$\sum_i r_i E_i = \left\{ \sum_i r_i e_C^{(i)} \right\}_{C\in\mathcal{M}} \tag{3.5}$$

In which we define $\sum_i r_i e_C^{(i)}$ as the weighted sum over the functions $e_C^{(i)}$: $\sum_i r_i e_C^{(i)}(s) := \sum_i r_i e_C^{(i)}(s)\ \forall s \in \mathcal{E}(C)$. Importantly, we will take E to be closed under such mixtures; this directly parallels our operational preparations being closed under statistical mixtures. Our preparation noncontextuality condition is then that all such convex mixtures are assigned unique global sections of $\mathcal{D}_R\mathcal{E}(X)$, that respect their probabilities. This approach explicitly mirrors notions of preparation noncontextuality; if different R-distributions mix to the same empirical model, then these are assigned by $\mathcal{G}$ an identical distribution over the global assignments $\mathcal{E}(X)$. We see this is exactly the notion from the equivalence approach; this is of course to be expected as we are identifying preparations based on equivalence classes of their empirical models, or operational predictions. As mentioned briefly above, this introduces a concept like universal contextuality; we are unable to separate the notions of preparation and measurement noncontextuality completely.

One other potentially appealing way of introducing the idea, however, is inspired by quantum steering, and by using measurements to perform preparations. The sheaf theoretic approach

mostly deals with single preparations with multiple possible measurements. This can be encapsulated by a steering scenario, since any set of states to which a system can be steered by measurement choice on an separate, entangled subsystem must mix to the same local reduced state by no-signalling.

Using steering, we can attempt to rebuild the proof of contextuality for preparations by Rudolph in Spekkens *Contextuality for preparations, transformations and unsharp measurements* [95]. Consider a nonlocality scenario in which the nonlocal resource is the symmetric state $|\phi_+\rangle = 1/2(|00\rangle + |11\rangle)$. Following Spekkens, let us define the following states:

$$
\begin{aligned}
|\psi_a\rangle &= |0\rangle \\
|\psi_A\rangle &= |1\rangle \\
|\psi_b\rangle &= \frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle \\
|\psi_B\rangle &= \frac{\sqrt{3}}{2}|0\rangle - \frac{1}{2}|1\rangle \\
|\psi_c\rangle &= \frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle \\
|\psi_C\rangle &= \frac{\sqrt{3}}{2}|0\rangle + \frac{1}{2}|1\rangle
\end{aligned}
\tag{3.6}
$$

At site A, we have access to two Trine POVM measurements and three projective measurements. In particular, we have access to the projective contexts $\{|\psi_a\rangle\langle\psi_a|, |\psi_A\rangle\langle\psi_A|\}$, $\{|\psi_b\rangle\langle\psi_b|, |\psi_B\rangle\langle\psi_B|\}$, $\{|\psi_c\rangle\langle\psi_c|, |\psi_C\rangle\langle\psi_C|\}$ and the trines $\{\sqrt{2/3}|a\rangle\langle a|, \sqrt{2/3}|b\rangle\langle b|, \sqrt{2/3}|c\rangle\langle c|\}$ and (the inverted image) $\{\sqrt{2/3}|A\rangle\langle A|, \sqrt{2/3}|B\rangle\langle B|, \sqrt{2/3}|C\rangle\langle C|\}$. At site B, we have access only to the projective contexts. This gives rise to the empirical model or probabilistic predictions shown in Figure 3.1.

This can be found to be contextual by linear methods as described by Abramsky [3]. We might, at first thought, believe that we have demonstrated a measure of preparation contextuality *via* showing the existence of nonlocality; there does not exist a local distribution of ontic states describing the state of the system, independently of which mixture of pure states we are steered to. However, further inspection reveals that the proof of contextuality relies on the exact overlaps between the quantum states, whereas the proof in the equivalence formalism does not

	a	A	b	B	c	C
a	$1/2$	0	$1/8$	$3/8$	$1/8$	$3/8$
A	0	$1/2$	$3/8$	$1/8$	$3/8$	$1/8$
b	$1/8$	$3/8$	$1/2$	0	$1/8$	$3/8$
B	$3/8$	$1/8$	0	$1/2$	$3/8$	$1/8$
c	$1/8$	$3/8$	$1/8$	$3/8$	$1/2$	0
C	$3/8$	$1/8$	$3/8$	$1/8$	0	$1/2$
a	$1/3$	0	$1/12$	$1/4$	$1/12$	$1/4$
b	$1/12$	$1/4$	$1/3$	0	$1/12$	$1/4$
c	$1/12$	$1/4$	$1/12$	$1/4$	$1/3$	0
A	0	$1/3$	$1/4$	$1/12$	$1/4$	$1/12$
B	$1/4$	$1/12$	0	$1/3$	$1/4$	$1/12$
C	$1/4$	$1/12$	$1/4$	$1/12$	0	$1/3$

Figure 3.1: Quantum mechanical predictions for the two-party steering scenario reproducing the proof of preparation contextuality from Reference [95].

make use of these facts at all. Only the orthogonality structure as well as the coefficients of the convex mixtures are used. In fact, the altered table shown in Figure 3.2 despite having the exact same structure of orthogonality and mixing properties, is a local table. This can be seen by noting that it is a uniform convex mixture of the twelve permissible ontic states.

Deterministic ontological states are defined here entirely by their assignments of outcomes to rows, since outcomes of the top three rows define the outcomes for the top three columns. The twelve ontic states can therefore be denoted as λ_X, for

$$X \in \{abCaC, abCbC, aBcaB, aBccB, AbcbA, AbcCa, \\ aBCaB, aBCaC, AbCbA, AbCbC, ABccA, ABccB\}, \quad (3.7)$$

where the subscript denotes the assignments to the rows in order. For example, λ_{abCaC} yields an outcome of a when measuring with respect to a and A, the first row or a, b, c, the fourth row; b when measuring with respect to the second row; and C when measuring with respect to rows 2 or 5. The table is local, so there exists a distribution over these sections, in this case the uniform distribution, that is compatible with all of the possible steering options. In addition, there are sharp measurements which fully enforce the compatibility structure, and each ontic state being bound to specific measurement subrows allow us to associate a conditional distribution

over ontic states with a specific outcome. Consideration of these outcomes will reveal that the normal notion of preparation contextuality is not being enforced by this structure. This can be seen immediately by noting that states like λ_{abCbC} are permissible, which are part of the support of $|a\rangle$ within a mixture of $\{|a\rangle, |A\rangle\}$, but not within a mixture of $\{|a\rangle, |b\rangle, |c\rangle\}$.

$$\mu_{a|\{a,A\}} = \frac{1}{6}\mathbb{1}_{\{abCaC,abCbC,aBcaB,aBccB,aBCaB,aBCaC\}} \tag{3.8}$$

$$\mu_{a|\{a,b,c\}} = \frac{1}{4}\mathbb{1}_{\{abCaC,aBcaB,aBCaB,aBCaC\}} \tag{3.9}$$

In order to recover the original proof, we must put in by hand the restriction that such states are inadmissible; however this is not part of the standard formalism at all, as we must in effect make the claim that whenever we would get the outcome $|a\rangle$ on a measurement of $\{|a\rangle\langle a|, |A\rangle\langle A|\}$, we must also for the POVM $\{\sqrt{2/3}|a\rangle\langle a|, \sqrt{2/3}|b\rangle\langle b|, \sqrt{2/3}|c\rangle\langle c|\}$. The operators $|A\rangle\langle A|$ and $\sqrt{2/3}|A\rangle\langle A|$ are not identified in any contextuality paradigm regardless of context; they give rise to different probabilistic predictions, and have different compatibility structures. However, if we do put such an argument in by hand, then the original argument follows; we can show that $1/2\mu_{aX} = 1/3\mu_{aX}$ for all X by no-signalling. Another way to produce the contradiction is to note that there are no ontic states compatible with those conditions for all the column measurements. This gives a slightly different perspective on the original argument, and also perhaps serves to highlight why even weakened forms of preparation noncontextuality are incompatible with quantum mechanics; the modifications to measurement noncontextuality to reconstruct even this weakened version[1] of preparation noncontextuality are large, and obviously in contention with much of the motivations of measurement noncontextuality.

This association however, can in certain cases in which the assumption of determinism can be made without loss of generality, be made to work in the opposite direction; as noted in Reference [90], with details given in Reference [93], The structure of Hardy's paradox can be

[1]The notion is weakened because a steering scenario can only reflect the structure of a preparation contextuality proof where decompositions of a single reduced state are considered. General proofs may involve consideration of multiple reduced states, and these proofs cannot be reflected in this formalism.

	a	A	b	B	c	C
a	$1/2$	0	$1/6$	$1/3$	$1/6$	$1/3$
A	0	$1/2$	$1/3$	$1/6$	$1/3$	$1/6$
b	$1/6$	$1/3$	$1/2$	0	$1/6$	$1/3$
B	$1/3$	$1/6$	0	$1/2$	$1/3$	$1/6$
c	$1/6$	$1/3$	$1/6$	$1/3$	$1/2$	0
C	$1/3$	$1/6$	$1/3$	$1/6$	0	$1/2$
a	$1/3$	0	$1/12$	$1/4$	$1/12$	$1/4$
b	$1/12$	$1/4$	$1/3$	0	$1/12$	$1/4$
c	$1/12$	$1/4$	$1/12$	$1/4$	$1/3$	0
A	0	$1/3$	$1/4$	$1/12$	$1/4$	$1/12$
B	$1/4$	$1/12$	0	$1/3$	$1/4$	$1/12$
C	$1/4$	$1/12$	$1/4$	$1/12$	0	$1/3$

Figure 3.2: GPT predictions for the two-party steering scenario reproducing the orthogonality and mixing properties of the proof of preparation contextuality from Reference [95], but obeying locality constraints. This difference is achieved by modifying only the overlaps, or distinguishability, of the states a, b and c. This change, however, makes the data table not quantum mechanically admissible.

seen as a proof of possibilistic preparation contextuality, again *via* considering the action of steering measurements as a method of state preparation on a subsystem

Theorem 2. *Any ontological model of quantum mechanics is possibilistically contextual for preparations.*

Proof. Let

$$|\phi\rangle = \cos\left(\frac{\phi}{2}\right)|0\rangle + \sin\left(\frac{\phi}{2}\right)|1\rangle \tag{3.10}$$

for some $0 < \phi < \pi/2$. We note that all mixed states are related under the possibilistic operational relation. In particular then, a mixed state made up of $|\phi\rangle\langle\phi|$ and $|-\pi/2\rangle\langle-\pi/2|$, and one made up of $|-\phi\rangle\langle-\phi|$ and $|\pi/2\rangle\langle\pi/2|$ are related. Therefore, such matrices must have the same support.

$$\mathcal{S}(|-\phi\rangle\langle-\phi|) \cup \mathcal{S}\left(\left|\frac{\pi}{2}\right\rangle\left\langle\frac{\pi}{2}\right|\right) = \mathcal{S}(|\phi\rangle\langle\phi|) \cup \mathcal{S}\left(\left|-\frac{\pi}{2}\right\rangle\left\langle-\frac{\pi}{2}\right|\right). \tag{3.11}$$

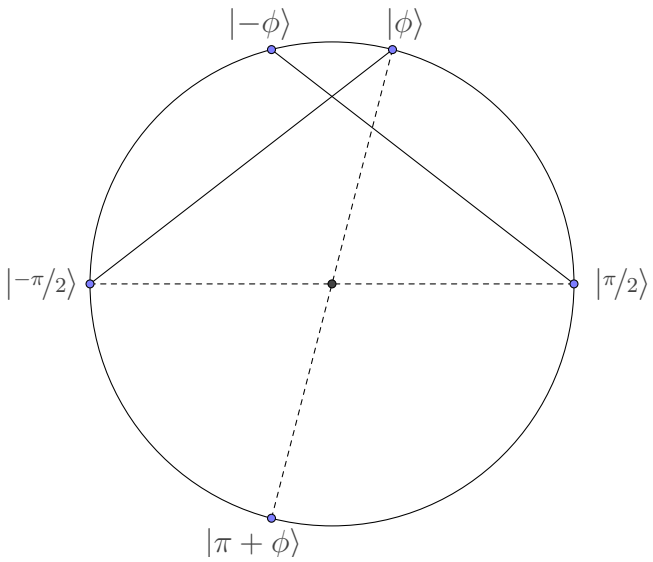


Figure 3.3: An example of a set of states to which one subsystem can be steered, demonstrating Possibilistic preparation noncontextuality as in Theorem 2.

Taking the intersection of both sides with $\mathcal{S}(\left|\frac{\pi}{2}\right\rangle\left\langle\frac{\pi}{2}\right|)$ gives

$$\begin{aligned}
\mathcal{S}\left(\left|\frac{\pi}{2}\right\rangle\left\langle\frac{\pi}{2}\right|\right) &= \left[\mathcal{S}(|\phi\rangle\langle\phi|) \cap \mathcal{S}\left(\left|\frac{\pi}{2}\right\rangle\left\langle\frac{\pi}{2}\right|\right)\right] \cup \left[\mathcal{S}\left(\left|-\frac{\pi}{2}\right\rangle\left\langle-\frac{\pi}{2}\right|\right) \cap \mathcal{S}\left(\left|\frac{\pi}{2}\right\rangle\left\langle\frac{\pi}{2}\right|\right)\right] && (3.12)\\
&= \left[\mathcal{S}(|\phi\rangle\langle\phi|) \cap \mathcal{S}\left(\left|\frac{\pi}{2}\right\rangle\left\langle\frac{\pi}{2}\right|\right)\right] && (3.13)\\
&\subseteq \mathcal{S}(|\phi\rangle\langle\phi|) && (3.14)
\end{aligned}$$

where the second line follows since $\left|\frac{\pi}{2}\right\rangle$ and $\left|-\frac{\pi}{2}\right\rangle$ are completely operationally distinguishable, and hence must be fully ontologically distinct. This shows that the set of states consistent with $|\pi/2\rangle\langle\pi/2|$ is a subset of those consistent with $|\phi\rangle\langle\phi|$, and so any operational predictions compatible with a preparation of $|\pi/2\rangle\langle\pi/2|$must also be consistent with a preparation of $|\phi\rangle\langle\phi|$. However, in order to be consistent with quantum mechanics, preparing $\left|\frac{\pi}{2}\right\rangle$ and measuring in the basis $\{|\phi\rangle, |\pi+\phi\rangle\}$ must give the outcome $|\pi+\phi\rangle$ with nonzero probability. □

The above proof is similar to Hardy's proof of Bell's theorem based only upon possibilistic arguments [47]. It leverages the same structure, but comes to a different, but related conclusion. This kind of proof highlights some of the inherent tension between preparation and measurement

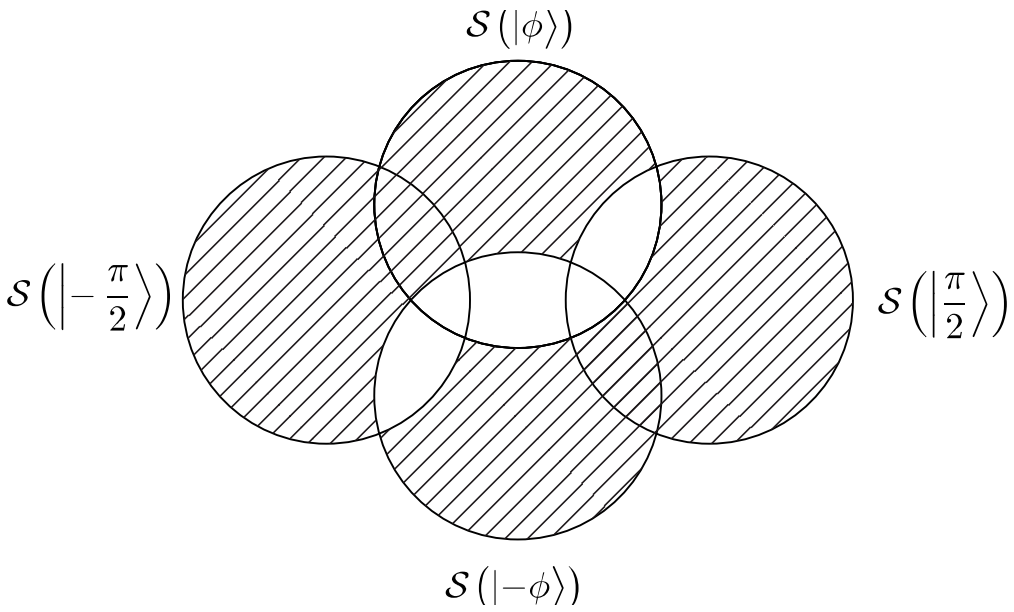


Figure 3.4: A Venn diagram showing the epistemic overlap regions for the different preparations. Regions that must be empty by the assumption of possibilistic noncontextuality for preparations are shown hatched. We see that the set of ontic states compatible with the states $|\pm\pi/2\rangle$ lie entirely within the supports of $|\mp\phi\rangle$ respectively, in contradiction with the possibilistic predictions of quantum mechanics.

noncontextuality; measurements can be used to perform preparations of mixed states, and in this scenario noncontextuality assumptions directly capture notions of locality.

3.4 Discussion

The different approaches to quantum contextuality all stem from the same concerns, and can be seen to be coincident with one another in certain restricted scenarios. The main point of contention, as discussed, is the way in which the different approaches consider the assumption of outcome determinism. Except in specific scenarios, such as those considered by Wester [105], this leads to an interesting difference in whether or not a situation can be considered contextual. Intrinsic randomness in the form of ontological coin flipping may not be the most interesting phenomenon to declare nonclassical, but the inherent randomness of contextual scenarios has been leveraged in proposals for quantum technology in fields such as certified secure randomness

generation [20]. As such, there is good philosophical and operational ground to consider such behaviour deeply nonclassical. In such devices, the violation of a Bell inequality is used as proof that a device's output is ontologically random and unknown to any adversary; since no deterministic model of hidden variables can explain the correlations, this rules out theories in which some knowledge possessed by an adversary (thought of as a hidden variable) can inform the outcomes in a meaningful way. The alternative is that the outputs are inherently, and ontologically random, providing a security guarantee that is uniquely quantum in that it is outside of the operational capacity of classical physics. Epistemic restrictions and complementarity, which are ultimately necessary for contextuality, are strong hallmarks of nonclassical behaviour. However, the strength of not mandating determinism in the equivalence framework is the additional flexibility to deal with proper POVMs, although one must still exercise caution as to whether such assumptions are well-motivated, as discussed above. However, this additional freedom can be useful, for example in proving no-go theorems. Perhaps, however, the key message of the sheaf approach in terms of the treatment of POVMs in the equivalence formalism should be that we should be much more careful in justifying notions of noncontextuality for POVMs in cases where they cannot be dilated uniquely. Assuming noncontextuality for POVMs is a powerful axiom; for example it leads to Gleason's theorem applying in $d = 2$ [28]. rather than only in $d \geq 3$. Assuming that POVMs have an ontological representation that is insensitive to context when those contexts cannot share a common dilation of the element is a very strong assumption and we should keep this in mind whenever something is demonstrated using this axiom.

One advantage of the sheaf-theoretic framework, which we were able to show above is capturable in a sense by the equivalence approach, is that many of the notions of contextuality can be defined in a way that is independent of a ny specific theory. On the other hand, we would note that the assumption of a measurement scenario is itself theory-laden, as it requires an experimental ontology in which measurements *can* be identified between contexts, side-stepping any finite-precision loopholes such as those discussed in detail by Barrett and Kent [14]. In a nonlocality scenario, when we assume that an outcome of a measurement at subsystem A is independent of the measurement choice at subsystem B, we do this because of an inherent

belief that the operational outcomes at subsystem A are the same. The difference in the two approaches is that in the equivalence formalism, this fact is taken axiomatically; the equivalence of the two is the reason that we identify their behaviour ontologically. In the sheaf-theoretic approach, however, this is deduced from a locality axiom and so the chain of reasoning exists in the opposite direction. The assumption made in the equivalence framework, that relationships of operational probabilities reveal some ontological relationships, is of course a very strong claim and more theory laden.

An interesting relationship between the two approaches can be found in the case above, where a proof of preparation contextuality in the equivalence formalism constituted a proof of what we might call measurement contextuality in the sheaf formalism. The reason for such coincidences, can be thought of as being a consequence, once again, of the differing treatment of determinism within the two approaches. The assumption of determinism forms a direct mapping between a preparation distribution to the ontic states and the distribution over the operational outcomes, by completely removing the degrees of freedom added to the system by the response functions $\{\xi_M\}$. This correspondence is possible in this case because the measurements made at the subsystem in question each have a preparation for which they are a sharp measurement with probability 1 or 0. This means that the assumption of determinism can be made in general, when this holds, proofs of nonlocality in the sheaf formalism can be turned into proofs of a corresponding flavour of preparation contextuality in the equivalence formalism.

Chapter 4

Restrictions on a Noncontextual Reality

Many of the most successful results in quantum foundations are *no-go* theorems, which can be thought as ruling out certain classes of physical theories. Many no-go theorems have as their underlying assumptions notions of noncontextuality; such theorems lead to restrictions on a noncontextual reality. The first of these to be demonstrated were proofs of quantum nonlocality such as Bell's theorem and variants such as that of GHZ. The locality assumptions of such theorems are easy to motivate; nonlocality is an explicit rejection of an ontological restriction that is intuitively true given our classical experiences. Indeed, the results of Bell's theorem are shocking enough that the flow of papers attempting to disprove it, or side-step it in some way have been unceasing since the paper's publication in 1964. In the following sections, we will attempt to demonstrate the difficulty in pinning down a precise notion of locality in the ontological models formalism, alongside some comments on attempts to create local theories which violate one or more of these assumptions. The awkwardness in some cases of defining whether or not a given model is nonlocal is of course in contrast to the question of whether a table of data from an experiment of two separated subsystems demonstrates nonlocality, which is easy to define, although hard to compute.

After, we will consider restrictions on ψ-epistemic models; that is, ones in which the quantum state is not ontologically instantiated but represents a state of knowledge. The most famous no-go theorem in this area is the Pusey-Barrett-Rudolph (PBR) theorem [83]; we will discuss the assumptions of this theorem as well as provide analogous results that use different or weaker assumptions. Some of the material in this section is from *"Contextuality under weak assumptions"* by Simmons, Wallman, Pashayan, Bartlett and Rudolph; in particular Theorem 3 from that paper will be discussed, the proof of which is primarily due to Wallman.

4.1 Locality assumptions

Building on the historical background given in Chapter 2, let us start off by considering Einstein's first attempt at a locality axiom, taken from his letters to Schroëdinger in the aftermath of the Einstein-Podolsky-Rosen [40] (EPR) paper. In this section we will bring this principle, which Einstein referred to as the "*Trennungsprinzip*", into the framework of ontological models [48]. To motivate this principle of locality, Einstein begins by considering an example of superposition applied to a macroscopic system of two boxes and a ball. The system is prepared such that the ball has an equal chance to be found in either of the two boxes if one or the other is opened. He notes that one way to force a "classical" interpretation to the experiment is to imagine the boxes separated and to assume "the content of the second box is independent of what happens to the first box"[1], noting that this locality assumption excludes the ψ-complete (Schrödinger-esque) interpretation, leaving only the more classical ψ-epistemic (Born-esque) interpretation remaining.

Later, with a more concrete example, he characterises the principle as saying "the actual state of B cannot be dependent on what sort of measurement I take at A[2]". We see that unlike more operationally motivated definitions of locality, Einstein's disconnection principle has an immediately ontological flavour; this is very much in keeping with Einstein's philosophical style of wanting to posit complete theories rather than principles or restrictions on agents. He then

[1] "[D]ie zweite Schachtel nebst allem, was ihren Inhalt passiert."
[2] "Der wirkliche Zustand von B kann nun nicht davon abhängen, was für eine Messung ich an A vornehme."

asks whether that theory is a better explanation of reality than ones lacking the properties in question. However, we argue that these remarks, amongst others in the letter, reflect a certain *naïvété* on Einstein's part. It does not seem as though he took the idea that quantum theory might actually preclude his *Trennungsprinzip* at all seriously; he takes the concept of locality, as Podolsky does in EPR, as a somewhat obvious fact that need not necessarily be rigorously defined. Demonstrating the inherent tension between ψ-completeness and locality, he is sure that it is ψ-completeness that will falter. By 1950, he was less sure, observing that physicists supporting interpretations with an ontologically-instantiated wavefunction were playing a "risky game with reality[3] [42]".

In particular, it is arguable that Einstein's wording of the *Trennungsprinzip* betrays one bias in particular; he makes the implicit assumption that separated objects *have* separate descriptions at all. This will form the first of the four elements of our definition of locality [90]. In this section we will focus on bipartite systems for notational clarity and to illustrate the guiding principles behind the ideas; however each of these definitions and theorems can be shown to hold true for multipartite cases.

Definition 16 (KL). An ontological model satisfies the *kinematic localtity* or *separability* condition if the ontic state of a collection of disconnected systems is equal to the Cartesian product of the states of each subsystem.

$$\mathbf{\Lambda} = \Lambda_A \times \Lambda_B \tag{4.1}$$

$$\mu_{P_{AB}}(\lambda) = \mu_{P_{AB}}(\lambda_A, \lambda_B), \tag{4.2}$$

for all preparations $\mu_{P_{AB}}$.

Now that we have defined some notion of kinematic locality, we can begin to define assumptions that capture Einstein's idea that actions taken on one subsystem cannot ontologically affect an other; we wish to define local dynamics in the case that we have local kinematics. Given the representational split between transformations and measurements in both the ontological

[3] "... ein gewagtes Spiel mit der Wirklichkeit treiben."

models formalism, it is worth splitting out the locality constraints on caused by each of these into different definitions.

Definition 17 (**DL**)**.** An ontological model satisfies the *dynamical locality* condition just when the temporal evolution of an ontic state of a subsystem A, $\lambda \in \Lambda_A$, is independent of an operation or measurement performed in any separate subsystem.

$$\Gamma_{T_A}(\lambda_A \times \lambda_B) = \Gamma'_{T_A}(\lambda_A) \tag{4.3}$$

Definition 18 (**RL**)**.** An ontological model for quantum mechanics obeys the *response locality* condition if the response function for a local measurement is a function only of the local ontic state.

$$\xi_{k_a|M_A,k_b|M_B}(\lambda_A \times \lambda_B) = \xi'_{k_a|M_A}(\lambda_A)\xi'_{k_b|M_B}(\lambda_B) \tag{4.4}$$

These two conditions are enough to define locality for simple prepare-and-measure scenarios, but not for scenarios involving transformations on states that have already undergone a measurement procedure. Such scenarios, and therefore a complete ontological theory for quantum mechanics must possess a state update rule.

Definition 19 (**UL**)**.** An ontological model for quantum mechanics obeys the *measurement update locality* condition if the update rule for a measurement affects, and depends on, only the local ontic state.

Definition 20 (**EL**)**.** An ontological model is *Einstein local* if it obeys **KL**, **DL**, **RL** and **UL**.

This seems a fair definition of Einstein's view on what counted as a local theory, brought into the formalism of ontological models. He tacitly assumes **KL**, as he talks about quantum subsystems having defined "real states"; his first statement about the disconnection principle, that the content of a second system should be independent of what happens to the first system is a description of **DL**; his notations for measurements being taken make it very clear that he is assuming **RL**; he highlights and underlines a statement that the actual state of a system cannot be dependent on which sort of measurement is taken at A, which if we consider also the outcome

of the measurement is a statement of **UL**. Many of these can be thought of as reminiscent of noncontextuality assumptions; this is not surprising, because nonlocality is a special case of contextuality where our compatibility graphs for measurements are complete bipartite.

These definitions are worthy of some questioning. One notable feature is our differentiation of **DL** and **RL**; it may seem as though the concept of dynamical locality already gives us the concept of response locality, as we are denying any causal influence between two separated systems. On the other hand, given the split between unitary transformations, governed by the Schrödinger equation, and nonunitary collapses associated with measurements that appear in many interpretations of quantum mechanics, including the standard interpretation for ontological models, differentiating between these assumptions may be seen as helpful. One argument for splitting these locality constraints apart is that this allows a more precise view on how exactly certain models for quantum mechanics fail to be local theories in the Einstein sense. An example of a model for quantum mechanics that obeys **DL** but not **RL** is Pienaar's [78] promotion of the Deutsch-Hayden approach [33] to a model for quantum mechanics [90]. Pienaar's use of the phrase "dynamically local" in the title and body of the paper refers, as it does here, to locality of transformation processes other than measurements.

Only the combination of all of these separate assumptions has power to restrict the set of available physical theories. An extreme case is that without kinematic locality, we cannot even begin to define dynamical and response locality. This separability condition is essential for any ontological model to be considered truly local. This is a contentious issue for the Many-Worlds interpretation; while the manifestly nonlocal wavefunction collapse is removed, and the dynamics are local, the theory still relies on a highly-nonlocal kinematic space to enforce the correct quantum correlations. This is not an unresolvable issue; we can define a model for quantum mechanics where at every spacetime point, the local state encodes the image of the point's past lightcone including all wavefunction information. This is a much larger memory burden on the model than just having a nonlocal wavefunction, however, and such memory burdens are characteristic of attempts to create fully local Many Worlds interpretations such as that of Rubin [88], and of Brassard and Raymond-Robichaud [21]. Here, as always in

the ontological models formalism, we will be working under the assumption of a single-world interpretation of quantum mechanics.

Theorem 3. *Any operational theory has a compatible ontological model* $[\Lambda, \{\mu\}, \{\Gamma\}, \{\xi\}]$ *satisfying* ***KL****,* ***DL****, and* ***UL****.*

Proof. The construction used to demonstrate this result is a "history-ontic" model, where there is an ontological record of the complete history of the systems rather than their current state. Such models are unappealing as real candidates for the ontology of the universe, but have been leveraged successfully to prove technical results about nonlocality, such as the aforementioned work on local versions of many worlds by Rubin [88].

For clarity, we will assume that time is discretised. Given an operational theory $[\mathcal{P}, \mathcal{M}, \mathcal{T}, \Pr]$ with n nonlocal sites, our ontological model has a statespace given by a Cartesian product of n functions $\mathbb{N} \to \mathcal{P} \cup (\mathcal{O} \times \mathcal{M}) \cup \mathcal{T}$: functions from the time index to the operational event happening at that time at that site. We can assume that the starting state of the system maps all $\mathbb{N}$ to the null transformation, and the (local) dynamics merely keep track of what events are happening at each of the nonlocal sites over time. When a transformation involves an interaction between these otherwise nonlocal sites, this is recorded as a transformation T_{n_1,n_2} at each of the two sites involved.

Clearly, the state of this history-ontic system is enough to determine probabilities of measurement outcomes; this is tautological, since it is axiomatic for operational theories that the history is enough to determine these probabilities. Hence, the nonlocal response functions are able to reproduce any probability function Pr. There is no state update in this model other than recording an outcome $o|M$ in a subsystem where M is measured and outcome o observed. □

An example of a model of this type that has local dynamics but not local representation of local measurements is Pienaar's Deutsch–Hayden-inspired model [78]. In this Heisenberg picture model, each subsystems track in some sense their own states but also capture information about states they have interacted with in the past, up to the point of interaction. While measurements

on a single system have a local representation, separate, commuting measurements on two or more systems have a representation that is nonlocal.

Theorem 4. *Any operational theory has a compatible ontological model* $[\Lambda, \{\mu\}, \{\Gamma\}, \{\xi\}]$ *satisfying* ***KL****,* ***RL****, and* ***UL****.*

Proof. We consider a model with the following state space:

$$\Lambda' = \Lambda \times \Lambda \tag{4.5}$$

which should be interpreted as each separated subsystem possessing a complete copy of the universal state. Such a model is clearly dynamically nonlocal, as any action taken on any subsystem is propagated to the states of every other system, regardless of spatial separation. However, since each subsystem possesses full information about the rest of the universe, any set of measurement probabilities are replicable. If state update is local, the nonlocal dynamics can nonetheless carry this information nonlocally. We see that **UL** is entirely trivial in the absence of **DL**. □

Theorem 5. *Any operational theory has a compatible ontological model* $[\Lambda, \{\mu\}, \{\Gamma\}, \{\xi\}]$ *satisfying* ***KL****,* ***DL****, and* ***RL****.*

Proof. The construction for this case is reminiscent of our "history-ontic" model from before, but with some slight differences. We now must have **RL** and so response functions must be local; however we do not have **UL**, and so our state update rule can act nonlocally.

We will use the same statespace as before: if our operational theory has n nonlocal sites, our statespace is a Cartesian product of n functions $\mathbb{N} \to \mathcal{P} \cup (\mathcal{O} \times \mathcal{M}) \cup \mathcal{T}$. Our local dynamics are the same, updating each local system so that the state of a system is a map between time steps and the operational element that occurred at that time step. Consider the case in which we make a measurement M, on some system, and obtain outcome o. The operational procedure up until that point is a valid preparation procedure for each system, and is represented in particular within $\mathcal{P}$. The state update rule then assigns to that time step, at each of the other

sites, the relevant preparation from $\mathcal{P}$. For notational clarity, we assume that there are no measurements made at the exact same timestep on each system, although this restriction is not particularly meaningful. The state of each of the subsystems at each point contains enough information to predict the probabilities of local measurements by the definition of operationally local measurements, so such a model can realise any probability function Pr. □

An example of an ontological model without update locality can be found in a traditional view of state collapse and quantum steering; a measurement is performed on an entangled state, collapsing both its state and the state of the other entangled subsystems.

4.2 Restrictions on noncontextual subtheories

The Bell-Kochen-Specker theorem itself strongly constrains the set of measurements that a noncontextual subtheory of quantum mechanics can contain. We will present a novel proof of the Bell-Kochen-Specker theorem that illustrates just how overconstrained a model for quantum mechanics is that attempts to be both noncontextual and deterministic.

Theorem 6 (Bell-Kochen-Specker)**.** *There is no ontological model* $[\Lambda, \{\mu_i(\lambda)\}, \{\xi_i(\lambda)\}]$ *of quantum mechanics in a Hilbert space of dimension* $d \geq 3$ *with a set of noncontextual response functions* $\{\xi_{P_i} i(\lambda)\}$ *with the property that for all* $\{P_i\}$*, we have* $\xi_{P_i}(\lambda) \in \{0, 1\}$*.*

There exist many proofs of this theorem, of which the majority are by demonstrating a specific set of operators which admit no noncontextual assignment of 0 and 1 variables, the first of which, the original proof by Kochen and Specker, involved a set of 117 projection operators that could not simultaneously be assigned values consistent with these axioms. Here, I present a proof more similar to that of Bell, displaying that the geometric relationships demanded by such an assignment are not possible, given the assumption of a complete state assignment for the real qutrit.

Proof. Let us assume that such an assignment of values to projectors exists. We will then

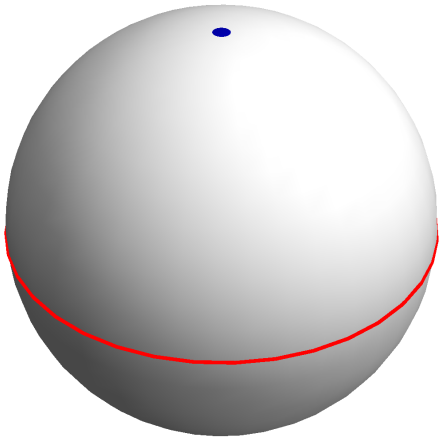


Figure 4.1: A single point noncontextually assigned a 1, in blue, leads to a great circle of points noncontextually assigned a 0, in red.

consider the subspace of projectors projecting onto states with real coefficients, that is to say, states of the form $|\psi\rangle = a|0\rangle + b|1\rangle + c|2\rangle$, with $a, b, c \in \mathbb{R}$. We can see that this statespace is isomorphic to a sphere, embedded into $\mathbb{R}^3$ in the obvious way, and we will identify points on the sphere with these states, and equivalently the possible projectors. Since bases with respect to which we can make a projective measurement are sets of mutually orthogonal states, these correspond naturally to sets of mutually orthogonal vectors in the sphere. Our problem is now to colour a sphere such that whenever we colour a point blue (equivalent to the projector related to it being assigned the value 1), we colour the antipodal point blue (as antipodal points are identified) and colour the great circle equidistant from those two points red (assigning those projectors a value of 0), as shown in figure 4.1. We see that the existence of a blue point necessitates the existence of a red great circle associated to them, and by contrast the existence of two perpendicular red points necessitates that they lie on a red great circle, with the associated antipodal points blue. This is because for any basis with respect to which we decide to measure, and equivalently three mutually orthogonal vectors we choose, there must be a single outcome which is represented in our model by a single red point. Since if an operator is determined to have a positive outcome, say $|0\rangle\langle 0|$, then it must be the outcome yielded whether we measure it with respect to $|1\rangle\langle 1|$ and $|2\rangle\langle 2|$; or $|+_{1,2}\rangle\langle +_{1,2}|$ and $|-_{1,2}\rangle\langle -_{1,2}|$. These correspond to rotations of the sphere about the vector identified with $|0\rangle\langle 0|$.

The proof structure is roughly as follows. First, we will prove a lemma stating that there are no closed curves on the sphere that enclose an area that is entirely coloured blue, by considering the colouration of great circles directly adjacent to that boundary. Then, we will demonstrate that the smaller area between any two red great circles that do not meet at a right angle must be entirely coloured red. Since any blue point must, by the lemma, be between two such red great circles, then we reach a contradiction. In fact, this second fact could be used to prove the theorem by itself; the only possible configuration of blue points it admits, a pair of antipodal spherical cap of blue points, covers less than $1/3$ of the sphere, whereas a randomly chosen measurement must have one outcome, dictating that the measure of blue points must be $1/3$.

Lemma 2. *No closed, non-self-intersecting curve on the sphere can enclose an area entirely coloured blue.*

Proof of Lemma. Let us assume that such a curve exists, and then consider a maximal such curve. We can choose some arbitrary point strictly inside it, A, and then note that since the curve is non-self-intersecting there is an infimal distance d under the geodesic distance metric on the surface of the sphere (or indeed, the Euclidean distance metric in the 3-d space in which the sphere is embedded) between any red point and the point A, and so there also exists a family of sets, $\mathcal{P}_\epsilon = \{X \in \mathcal{S} | d(X, A) \leq d + \epsilon\}$ indexed by $\epsilon \in \mathbb{R}$, of red points whose distance from A differ from d by at most ϵ. The case in which this infimal distance is realised is easy; choose a point B which realises this infimal distance, and consider the set of great circles passing through B. All but one contain points that are closer to A than B is, and so none of these great circles can be entirely red, as the distance between A and B was minimal. Hence, the great circle associated to the point B is almost all blue, since a red point on this great circle means that there would be a red great circle through that point to B. But then we have a circle, less two points, of blue points; clearly we have two orthogonal blue points and therefore a contradiction.

The case in which this infimum is not minimised is similar but trickier. For some ϵ we choose arbitrarily some member of $\mathcal{P}_\epsilon$, B, and consider the set of great circles passing through B, as a family of curves parametrised by θ, their angle from the element of the set passing through A.

Note, that for $\theta \neq \pi/2$, these great circles contain points strictly closer to A than B is, so they contain points which, if coloured red, are members of $\mathcal{P}_{\epsilon'}$, with $\epsilon' < \epsilon$. Consider the great circles corresponding to $\theta = \pm\pi/4$. It is a fairly simple exercise in coordinate geometry to determine that for sufficiently small ϵ, these great circles contain points with $\epsilon < 0$. Hence, by the prior logic these great circles cannot be entirely red, since they intersect the blue cap.

However, if we consider the points perpendicular to B on these two great circles, they must be coloured blue, since neither great circle is entirely red, as they would have to be by noncontextuality if there were two perpendicular red points. Then, we have two blue points perpendicular to one another– a contradiction. Hence, there cannot be any closed, non-self-intersecting curve on the sphere enclosing an entirely blue area. □

As an immediate corollary to our lemma, we see that red points are dense on the sphere, all of which lie on red great circles. We will now consider two red great circles that do not meet at a right angle. Choose one of the two antipodal points where these circles meet, and we will define the point $\vec{x}(\theta)$ to be the point along the first great circle that subtends an angle of θ with the chosen point. We then define $\vec{y}(\theta)$ in a similar way.

It can easily be seen that $\forall\theta\ \exists\phi_1$, ϕ_2 such that $\vec{x}(\theta)$ subtends an angle of $\pi/2$ with $\vec{y}(\phi_1)$ or $\vec{y}(\phi_2)$. Additionally, we can find functions $\phi_{1,2}(\theta)$ that are continuous. Therefore, we have a solid red area.

To show this, we first take without loss of generality our first great circle to be the one defined by the following parametric equation:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cos(\theta) + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \sin(\theta). \tag{4.6}$$

Again without loss of generality, the equation for our second circle is given by

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cos(\phi) + \begin{pmatrix} 0 \\ a \\ \sqrt{1-a^2} \end{pmatrix} \sin(\phi). \tag{4.7}$$

For any given θ, the points which are orthogonal to $\vec{x}(\theta)$ are those on the circle given by:

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} \sin(\theta) \\ -\cos(\theta) \\ 0 \end{pmatrix} \cos(\Theta) + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \sin(\Theta). \tag{4.8}$$

The latter two great circles must intersect, since all great circles intersect. Therefore there exists a solution $\forall \theta, a$. In fact, by solving the above equations we see that we get $\phi = -\tan^{-1}(1/a \cot\theta)$, and $\Theta = -\tan^{-1}(\sqrt{1-a^2}/a \cos\theta)$. So the solutions are continuous functions of θ, yielding our result. Hence, the smaller area between any two red great circles that do not meet at a right angle is entirely red. However, as a consequence of our above lemma, it can be seen that *all* points on the sphere are in such a sector for some two red great circles. Hence, our initial assumption of such a colouring is found to be contradictory and this concludes the proof. □

4.3 Restrictions on ψ-epistemic models

4.3.1 The PBR theorem

The Pusey-Barrett-Rudolph theorem [83] is a powerful no-go result that strongly restricts ψ-epistemic theories, under a condition motivated by locality concerns. The set up is that there are some number of sources of states, that may be spacelike separated from one another. These sources produce one of two quantum states that, in some putative model, are expected to have some overlap in the set of ontic states that they produce.

Let us quickly examine the case in which the two states that can be produced are $|0\rangle$ and

$|+\rangle = {}^1\!/\!\sqrt{2}\,(|0\rangle + |1\rangle)$, and there are two spacelike separated sources. We might assume that the distribution associated with these preparations has the following form:

$$\mu_{|\psi_A\rangle|\psi_B\rangle} = \mu_{|\psi_A\rangle}\mu_{|\psi_B\rangle}. \tag{4.9}$$

This property is known as *Preparation Independence*, and reflects the locality assumption that the choice of preparation on subsystem A and the choice of preparation on subsystem B make completely independent contributions to the overall distribution, with no holistic effects or interaction between the two. This can be defined even in the absence of Kinematic locality, although it is easier to justify when the action on each subsystem is ontologically local. This can be the case within a model that is not overall kinematically local, such as one in which there is a nonlocal component to entangled states enabling superclassical correlations, but separable states have local kinematics. This also can be thought to be a kind of noncontextuality assumption; the context of what choice is made at A does not affect the contribution of the choice made at B to the overall state. Weaker versions of this assumption are known to be strong enough to prove analogues of the PBR result, for example allowing classical shared randomness during preparation procedures [69].

In the thought experiment, there are two spacelike separated quantum sources, each producing $|0\rangle$ or $|1\rangle$. Therefore, the joint states that can be produced by the sources together are $|00\rangle$, $|0+\rangle$, $|+0\rangle$ and $|++\rangle$. In a ψ-epistemic model in which $|0\rangle$ and $|1\rangle$ have ontological overlap when prepared on either of the systems A, B, then these four states must share some joint ontological overlap. The authors show that there exists a measurement with four outcomes, such that for each of the four possible states that can be prepared by the separated systems one of the measurement outcomes is impossible. The measurement is a projection onto the states

$$\begin{aligned} |\xi_1\rangle &= \frac{1}{\sqrt{2}}\,(|01\rangle + |10\rangle) & |\xi_2\rangle &= \frac{1}{\sqrt{2}}\,(|0-\rangle + |1+\rangle) \\ |\xi_3\rangle &= \frac{1}{\sqrt{2}}\,(|+1\rangle + |-0\rangle) & |\xi_4\rangle &= \frac{1}{\sqrt{2}}\,(|+-\rangle + |-+\rangle)\,. \end{aligned} \tag{4.10}$$

A state that is in the support of $|00\rangle$, say, must assign zero probability to the projector $|\xi_1\rangle\langle\xi_1|$,

since we have that $|\langle 00|\xi_1\rangle|^2 = 0$. However, each of the states that can be produced is orthogonal to one of these, so a state that is in the support of all four states must, in order for these possibilities to be reproduced, assign no probability to any of the four outcomes. However, this is in violation of the axioms of ontological models, since for any ontic state and any measurement, there must be some outcome and so the probabilities assigned to each outcome must sum to 1. In the language of Caves, Fuchs and Schack when discussing compatibility of quantum state assignments, the set $\{|00\rangle, |0+\rangle, |+0\rangle, |++\rangle\}$ has been found to be *Post-Peierls incompatible*.

This result is then extended to demonstrate that for any two states $|\psi_1\rangle$, $|\psi_2\rangle$, there is some number of such sources N such that $\{\bigotimes_i^N |\psi_i\rangle\}$ is Post-Peierls incompatible, by demonstrating an existence proof of a measurement, each element of which is orthogonal to one of the states that can be prepared.

Our conclusion must be that if we want to have this notion of preparation independence, then the size of our Hilbert space severely constrains which states can be ψ-epistemic. This result can be interpreted as something of a complement to Bell's theorem: the PBR result most obviously rules out kinematically local, ψ-epistemic states with local preparations and an entangled measurement; Bell's theorem rules out kinematically local states (which would have to be ψ-epistemic, since entangled states enable steering) using an entangled preparation and local measurements. While the conclusion of the PBR theorem is less forceful, it is in some senses more surprising. In the case of Bell's theorem, the two systems have been in contact in the past, and have interacted in a way that the normal quantum formalism represents as being describable only holistically. It is perhaps somewhat palatable that in these entangled cases, there is some "communication channel" forged between the two subsystems that enable locality to be sidestepped in one of the manners described in Section 4.1. In any case, once particles have been brought into contact and interacted in a fundamental way, we are perhaps more tolerant of the idea that those particles really do no longer have simple, local descriptions. However, the PBR theorem places restrictions on particles that have not interacted at all! If one wishes to find a model for quantum mechanics that supports a ψ-epistemic interpretation, one either has to accept a large amount of nonlocality, or place other arguably artificial restrictions on the

model to sidestep the PBR argument[4].

4.3.2 Restrictions from notions of measurement noncontextuality

Another known restriction on ontological models under the assumption of probabilistic measurement noncontextuality is that such models cannot be ψ-epistemic, a result first shown by Morris [75] and by Chen and Montina [29]. Each of these proofs has something of a different flavour, and can be interpreted slightly differently. We will present a streamlined version of the proof by Morris, demonstrating that it can be thought of as similar to Rudolph's proof of preparation contextuality in Reference [95]. This streamlining maintains the assumptions and logical structure of the original proof, but simplifies the algebra from the original formulation.

Theorem 7. *A ψ-epistemic ontological model for quantum mechanics in dimension $d \geq 2$ must be measurement contextual.*

Proof. Let us define the following states

$$\begin{aligned} |\psi_a\rangle &= \cos(\theta)|0\rangle + \sin(\theta)|1\rangle \\ |\psi_b\rangle &= \cos(\theta)|0\rangle - \sin(\theta)|1\rangle \\ |\psi_c\rangle &= |1\rangle. \end{aligned} \tag{4.11}$$

We will assume without loss of generality that we have $0 < \theta < \pi/4$. We will denote the orthogonal state to each of these with a capital letter, such as $|\psi_A\rangle = |\psi_a^\perp\rangle$. We assume that there are some ontic states $\lambda_{ab} \in \Lambda_{ab}$ that are in the epistemic overlap of $\mu_{|\psi_a\rangle}$ and $\mu_{|\psi_b\rangle}$.

We will consider a single measurement, obtained by convex mixture of the three pairs of PVM measurements uniformly, generating the following two POVM elements. We note that this is not a *proper* POVM such as the trine, but rather something that can be performed on a single

[4] A somewhat trivial example of such an artificial restriction is to build an ontological model in which there is exactly one spacetime point at which a preparation can be ψ-epistemic, and otherwise the model is ψ-ontic. Since the argument relies on being able to compose multiple copies of ψ-epistemic systems, this approach can reproduce the quantum mechanical predictions for some restricted quantum subtheories.

system by randomly selecting a PVM. Our two measurement effects are

$$E_0 = \alpha\left(|\psi_a\rangle\langle\psi_a| + |\psi_b\rangle\langle\psi_b|\right) + \beta|\psi_c\rangle\langle\psi_c| = \frac{\mathbb{1}}{2} \tag{4.12}$$
$$E_1 = \alpha\left(|\psi_A\rangle\langle\psi_A| + |\psi_B\rangle\langle\psi_B|\right) + \beta|\psi_C\rangle\langle\psi_C| = \frac{\mathbb{1}}{2}$$

It is easy to calculate that $\alpha = 1/4\cos^2(\theta)$, and $\beta = 1-\tan^2(\theta)/2$. In particular, note that we have $\beta < 1/2$ for all $0 < \theta < \pi/4$. Since these measurement effects simulate a coin flip operationally, the assumption of measurement noncontextuality says that they must also do so ontologically, giving:

$$\alpha\left(\xi_{|\psi_A\rangle\langle\psi_A|}(\lambda) + \xi_{|\psi_B\rangle\langle\psi_B|}(\lambda)\right) + \beta\xi_{|\psi_C\rangle\langle\psi_C|}(\lambda) = \frac{1}{2} \quad \forall\lambda \in \Lambda \tag{4.13}$$

Now, consider those states λ_{ab}, in the epistemic overlap of $|\psi_a\rangle$ and $|\psi_b\rangle$. We have

$$\alpha\left(\xi_{|\psi_A\rangle\langle\psi_A|}(\lambda_{ab}) + \xi_{|\psi_B\rangle\langle\psi_B|}(\lambda_{ab})\right) + \beta\xi_{|\psi_C\rangle\langle\psi_C|}(\lambda_{ab}) = \frac{1}{2} \tag{4.14}$$
$$\beta\xi_{|\psi_C\rangle\langle\psi_C|}(\lambda_{ab}) = \frac{1}{2}. \tag{4.15}$$

However, this is a contradiction since $\beta < 1/2$ and $\xi_{|\psi_C\rangle\langle\psi_C|}(\lambda_{ab}) \leq 1$. □

As discussed before, the two measurement effects $\mathbb{1}/2$ and E_1 in general lead to different state updates ontologically, and so even in a ψ-epistemic theory it is arguable as to whether or not it is fully justified to claim that they should be ontologically identified.

However, a similar result can also be demonstratedby considering the POVM measurement whose elements are proportional to the projectors onto the $|\psi_a\rangle$, $|\psi_b\rangle$ and $|\psi_c\rangle$ [90] . In this case, we obtain

$$1 = \frac{1}{2\cos^2(\theta)}\left(\xi_{|\psi_a\rangle\langle\psi_a|}(\lambda) + \xi_{|\psi_b\rangle\langle\psi_b|}(\lambda)\right) + (1-\tan^2(\theta))\xi_{|\psi_c\rangle\langle\psi_c|}(\lambda) \tag{4.16}$$
$$> \frac{1}{2\cos^2(\theta)}\left(\xi_{|\psi_a\rangle\langle\psi_a|}(\lambda) + \xi_{|\psi_b\rangle\langle\psi_b|}(\lambda)\right) \tag{4.17}$$
$$> \frac{1}{\cos^2(\theta)} > 1. \tag{4.18}$$

We note that this is a proper POVM measurement, requiring another system to be interacted with the first in order for the measurement to be made possible.

By contrast, the proof of Chen and Montina holds in $d \geq 3$, and relies on an application of Gleason's theorem. Since Gleason's result shows that any noncontextual value assignment to the quantum measurement operators is reproducible *via* the Born rule applied to some density matrix, it is immediate that there cannot be two distinct measurement operators that are both given a valuation of 1. While the result of Chen and Montina does not require consideration of unsharp measurement operators, and therefore any requirement of convexity, it does require a larger quantum dimension. It also naïvely requires the full set of measurement operators. However, we can see this result as a corollary to the "logical indeterminacy" principle that was demonstrated by Pitowsky [81] in 1998. Pitowsky showed that that given any two quantum states that there was some finite measurement scenario, such that for any noncontextual probability assignment to that scenario, the two states were never given valuations in $\{0, 1\}$ unless they were both given 0. Since the ψ-epistemic property implies that it must be possible to give each state a 1-valuation, the result immediately follows. Hence, the result of Chen and Montina is possible to prove even within a finite scenario. The size of the measurement scenario needed to demonstrate that two given quantum states cannot share epistemic overlap is dependent on their overlap, just as in the PBR case, although the exact scaling is an open problem. Pitowsky shows that for any ϵ, there is some finite scenario such that the Born-rule probabilities are met to within deviances of ϵ, and the size of the necessary scenario increases as ϵ decreases. To demonstrate that two states cannot simultaneously be given a 1-valuation requires a value of ϵ that scales with the squared inner product of the two vectors in question.

In this tradition of restricting ψ-epistemic theories using notions of measurement noncontextuality, we will demonstrate an analogous result, in which the notion of noncontextuality applied is strictly weaker than probabilistic measurement noncontextuality. This proof follows that of Simmons, Wallman, Pashayan, Bartlett and Rudolph [93], and was primarily due to Wallman.

We define two sets associated with a response function for outcome k of a measurement $\mathcal{M}$ in

an ontological model:

$$\mathcal{C}[\xi_M(k|\lambda)] := \{\lambda : \xi_{\mathcal{M}}(k|\lambda) = 1\} \tag{4.19}$$

$$\mathcal{T}[\xi_M(k|\lambda)] := \{\lambda : \xi_{\mathcal{M}}(k|\lambda) > 0\}. \tag{4.20}$$

These we will refer to as the *certain* set, $\mathcal{C}$; and the *tracking* set $\mathcal{T}$. We will here consider the important special case in which these sets are noncontextual insofar as they depend only upon the POVM element E_k associated with outcome k and not the context of the POVM in which it appears. In these case, we will be able to denote the sets by $\mathcal{C}(E_k)$ and $\mathcal{T}(E_k)$ respectively. We will see that when this is the case, restrictions are imposed on which states can have ontic overlap. Models in which the sets $\mathcal{C}$ and $\mathcal{T}$ are noncontextual in this sense will be called *faithful*, since this is a stronger form of the faithfulness condition introduced in Reference [89]. We will briefly note that under certain natural assumptions, models obeying possibilistic measurement noncontextuality are faithful; this shall be discussed in more detail below.

Theorem 8. *Let $|\psi_1\rangle$ and $|\psi_2\rangle$ be two states in a Hilbert space of dimension $d \geq 2^N$, with $|\langle\psi_1|\psi_2\rangle|^2 \leq \cos^{2N}(2\tan^{-1}(2^{1/2N} - 1))$. Then in any faithful ontological model they have disjoint supports:*

$$\mathcal{S}(\mu_{|\psi_1\rangle}) \cap \mathcal{S}(\mu_{|\psi_2\rangle}) = \emptyset. \tag{4.21}$$

Proof. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be fixed preparation procedures for $|\psi_1\rangle$ and $|\psi_2\rangle$ and let $|\psi_1\rangle = |a\rangle^{\otimes N}$ and $|\psi_2\rangle = |b\rangle^{\otimes N}$ where $|a\rangle = |\phi\rangle$ and $|b\rangle = |-\phi\rangle$, where we have

$$|\phi\rangle = \cos\left(\frac{\phi}{2}\right)|0\rangle + \sin\left(\frac{\phi}{2}\right)|1\rangle. \tag{4.22}$$

This form for our states can be made without loss of generality, since for any pair of states $|\varphi_1\rangle$, $|\varphi_2\rangle$, there exists a unitary rotation mapping them into a pair of states of our chosen form, and unitary actions must preserve overlaps of support. If $|\langle\varphi_1|\varphi_2\rangle|^2 = |\langle\psi_1|\psi_2\rangle|^2$, then there exists a unitary U such that $U|\varphi_1\rangle = |\psi_1\rangle$ and $U|\varphi_2\rangle = |\psi_2\rangle$, and so any state that is in the support of $|\varphi_1\rangle$ and $|\varphi_2\rangle$ is mapped by such a transformation into a state that is in the support of $|\psi_1\rangle$

and $|\psi_2\rangle$ since this is a legitimate preparation procedure for those states. An alternative (but equivalent) viewpoint would be that we can define our tensor product structure such that $|\varphi_1\rangle$, $|\varphi_2\rangle$ have this form.

By considering each of the measurements formed by independently measuring $\{|a\rangle, |\bar{a}\rangle\}$ or $\{|b\rangle, |\bar{b}\rangle\}$ on each of the N qubits (where $|\bar{a}\rangle$ denotes the orthogonal state to $|a\rangle$, and similarly for $|b\rangle$), we have

$$\mathcal{S}(\mu_{|\psi_1\rangle}) \cap \mathcal{S}(\mu_{|\psi_2\rangle}) \subseteq \cap_{\vec{\nu}\in\{a,b\}^N} \mathcal{C}(|\vec{\nu}\rangle\langle\vec{\nu}|), \tag{4.23}$$

except perhaps on an unphysical set of measure 0, as either $|a\rangle^{\otimes N}$ or $|b\rangle^{\otimes N}$ is orthogonal to all but one outcome in each measurement. To see this, consider the case in which we measure with respect to $\{|a\rangle, |\bar{a}\rangle\}$ on the first qubit, and $\{|b\rangle, |\bar{b}\rangle\}$ on the second. Here, $|a\rangle|a\rangle$ is incompatible with $|\bar{a}\rangle|b\rangle$ and $|\bar{a}\rangle|\bar{b}\rangle$, with a similar condition holding for $|b\rangle|b\rangle$. The only outcome that is compatible with both $|a\rangle|a\rangle$ and $|b\rangle|b\rangle$ is $|a\rangle|b\rangle$. We note that by assumption, we have $|\langle a|b\rangle| \leq \cos^2(2\tan^{-1}(2^{1/2N}-1))$ and so we can apply the construction by PBR in Reference [83]; a rank 1 PVM such that for every one of the states $|\vec{\nu}\rangle$, there is a unique element of the PVM orthogonal to it. That is, this demonstrates that there exists a rank-1 projector-valued measure (PVM) $M = \{\Pi_{\vec{\nu}} : \vec{\nu} \in \{a,b\}^N\}$ such that $\Pi_{\vec{\nu}}|\vec{\nu}\rangle\langle\vec{\nu}| = 0$ for all $\vec{\nu}$ [83]. For all $\vec{\nu}$, $\Pi_{\vec{\nu}}|\vec{\nu}\rangle\langle\vec{\nu}| = 0$ and so there exist PVMs containing $\Pi_{\vec{\nu}}$ and $|\vec{\nu}\rangle\langle\vec{\nu}|$, for example $\{|\vec{\nu}\rangle\langle\vec{\nu}|, \Pi_{\vec{\nu}}, \mathbb{1} - |\vec{\nu}\rangle\langle\vec{\nu}| - \Pi_{\vec{\nu}}\}$. Hence

$$\mathcal{C}(|\vec{\nu}\rangle\langle\vec{\nu}|) \cap \mathcal{T}(\Pi_{\vec{\nu}}) = \emptyset \tag{4.24}$$

for all $\vec{\nu}$, since a lambda in the intersection of $\mathcal{C}(|\vec{\nu}\rangle\langle\vec{\nu}|$ and $\mathcal{T}(\Pi_{\vec{\nu}})$ would, in a measurement containing both such as the one given above, have to yield the $|\vec{\nu}\rangle\langle\vec{\nu}|$ outcome with probability 1, and the $\Pi_{\vec{\nu}}$ outcome with nonzero probability, a clear contradiction. Therefore

$$\begin{aligned}
\mathcal{S}(\mu_{|\psi_1\rangle}) \cap \mathcal{S}(\mu_{|\psi_2\rangle}) &\subseteq \left[\cap_{\vec{\nu}\in\{a,b\}^n} \mathcal{C}(|\vec{\nu}\rangle\langle\vec{\nu}|)\right] \cap \Lambda && (4.25)\\
&\subseteq \left[\cap_{\vec{\nu}\in\{a,b\}^n} \mathcal{C}(|\vec{\nu}\rangle\langle\vec{\nu}|)\right] \cap \left[\cup_{\vec{\nu}\in\{a,b\}^n} \mathcal{T}(\Pi_{\vec{\nu}})\right] && (4.26)\\
&\subseteq \emptyset, && (4.27)
\end{aligned}$$

giving the desired result, where the second line follows as some outcome must occur when the

measurement M is performed. □

Since $\lim_{N\to\infty} \cos^{2N}(2\tan^{-1}(2^{1/2N}-1)) = 1$, which can be verified by computing the expression's Laurent series expansion at infinity, which is given by $1-{}^{\ln^2(2)}/_{x}+O({}^{1}/_{x^2})$, this demonstrates that a pair of states with any overlap can be shown to be ontologically distinct, given a large enough Hilbert space in which to make measurements. As touched on before, this proof does differ from those of Morris [75] and of Chen and Montina [29] in the sense that the dimensional bound is dependent on the overlap between the states; this can be seen as a trade off for the weaker assumptions involved with the proof. In particular, we see that any ontological models for infinite dimensional quantum mechanics that obey the faithfulness condition must be ψ-ontic, since we can consider any two as living within an arbitrary dimensional subspace and perform an (infinite rank) PVM that differentiates states in this subspace. In most interpretations of quantum mechanics, consideration of a quantum system as closed is an arbitrary distinction and in general, any quantum system can be thought of as a small part of a larger quantum system. Results such as this, therefore, demonstrate an inherent tension between possibilistic measurement noncontextuality and ψ-epistemic ontological models or can be thought of as demonstrating that the assumption of preparation independence used by PBR can be replaced by that of faithfulness: an assumption about measurements rather than preparations. While both the necessary conditions for PBR and Theorem 8 are possibilistic in nature, they are distinct and constrain preparations and measurements respectively.

Earlier, we mentioned that the faithfulness property required for this proof could, under some assumptions, be considered to be a consequence of Possibilistic measurement noncontextuality. We recall from Chapter 2 that this means that when the same quantum measurement appears in two different contexts, the possibility of that outcome being selected is the same in both for all ontic states λ.

We can note first that the two conditions used in defining the certain and tracking sets, $\xi_M(k|\lambda) > 0$ and $\xi_M(k|\lambda) = 1$, can be recognised even after performing a possibilistic collapse. The former is easily identifiable, since it is the definition of a possible event and so is

strictly preserved by possibilistic collapse. The second is a little trickier; it can be identified within a particular context by one outcome being possible and all others being impossible. However, the notion of possibilistic noncontextuality only states explicitly that this (certain) event must be possible in other contexts, not that it must be certain. In order to bridge this gap we will need to make explicit some common assumptions about the structure of ontological models, or to use a more powerful notion of noncontextuality.

We note that if a particular measurement outcome, represented in quantum theory by $|\psi\rangle\langle\psi|$ is certain in some context, we can perform classical post-processing and coarse grain the other outcomes in its context, and this coarse-grained outcome will be impossible. However, in any other context in which $|\psi\rangle\langle\psi|$, we can also perform the same coarse graining, which in a single prepare-and-measure scenario has equivalent statistics, and therefore must also be impossible. This assumption is common, occurring for example in Reference [95]. One argument for this approach, for example made by Peres is to point out that quantum mechanically, one can perform coarse grained measurements and then further refine them; this is statistically indistinguishable from measuring the fine-grained measurement immediately. However, ontologically speaking this too requires some notion of state update. As discussed above, different coarse-grained measurement operators constructed from classical post-processing carry different state update rules ontologically, and so this assumption is at least arguable.

The other alternative is to adopt a more restrictive version of measurement noncontextuality. A seemingly minimal candidate for such a role would be a *trichotomistic* notion, in which quantum probabilities are collapsed into the three classes of 0, $1/2$ and 1. In such a notion, events are either impossible, indeterminate, or certain. Interestingly, such a concept of noncontextuality is not possible to replicate within the sheaf-theoretic framework, because it does not form a semiring; the union of two indeterminate events can either be indeterminate or certain. Clearly, such a trichotomistic notion is strictly weaker than probabilistic noncontextuality, just as possibilistic noncontextuality is.

4.4 Determinism and Counterfactual Independence

In Chapter 2, it was briefly mentioned that ψ-epistemic instances of the standard Kitchen Sink model in quantum dimensions $d \geq 3$ are incompatible with the Luders rule for state update. In this section, we will prove this result, reproducing a result of Wallman, as well as discuss the physical relevance of the result.

Theorem 9. *There is no possible state update rule for ψ-epistemic Kitchen Sink models in quantum dimensions $d \geq 3$.*

Proof. Let $|\psi_1\rangle = a\left(|0\rangle + |1\rangle\right) + b|2\rangle$, $a, b \in \mathbb{R}$, and $|\psi_2\rangle = a(|0\rangle - |1\rangle) + b|2\rangle$. Let $M = \{|0\rangle\langle 0| + |1\rangle\langle 1|, |2\rangle\langle 2|\}$. We see that quantum mechanically, for both $|\psi_1\rangle$ and $|\psi_2\rangle$, there is a $2a^2$ chance of getting the outcome associated with $|0\rangle\langle 0| + |1\rangle\langle 1|$, and a $b^2 = 1 - 2a$ chance of getting the outcome associated with $|2\rangle\langle 2|$.

The post update state for $|\psi_1\rangle$, given the outcome $|0\rangle\langle 0| + |1\rangle\langle 1|$ is $1/\sqrt{2}\left(|0\rangle + |1\rangle\right)$; for $|\psi_2\rangle$, it is $1/\sqrt{2}\left(|0\rangle - |1\rangle\right)$. These are orthogonal, whereas $|\psi_1\rangle$ and $|\psi_2\rangle$ were not. The sketch of the proof is as follows: $|\psi_1\rangle$ and $|\psi_2\rangle$ have epistemic overlap, and in particular there are some states λ that are associated with each preparation. When the measurement occurs, those lambda must be associated with an outcome of $|2\rangle\langle 2|$; if the outcome were $|0\rangle\langle 0| + |1\rangle\langle 1|$, then those states would have to update to states compatible with both $1/\sqrt{2}\left(|0\rangle + |1\rangle\right)$ and $1/\sqrt{2}\left(|0\rangle - |1\rangle\right)$; if the state λ is compatible with $|\psi_1\rangle$, then it has to be compatible with that state after state update. However, there are no states compatible with both the post-update states and we have a contradiction. The last remaining element is to show that the Kitchen Sink model does have some states that are compatible with both $|\psi_1\rangle$ and $|\psi_2\rangle$ and are bound to the $|0\rangle\langle 0| + |1\rangle\langle 1|$ outcome.

Firstly, we must show that there is some a for which there is epistemic overlap; this is simple, since by axiom our model is ψ-epistemic, and we have the relative size of a and b as a tuneable parameter. Since our model is ψ-epistemic, there are two vectors $|\phi_1\rangle$, $|\phi_2\rangle$ that share ontic states; we let $b = \sqrt{|\langle\phi_1|\phi_2\rangle|}$ and choose our basis appropriately.

Secondly, we must demonstrate that the structure of the Kitchen Sink model has states compatible with both $|\psi_1\rangle$ and $|\psi_2\rangle$ and are bound to the $|0\rangle\langle 0| + |1\rangle\langle 1|$ outcome. This is true by the independent structure of the Kitchen Sink model; conditioning on the outcomes of the $\{|\psi_1\rangle\langle\psi_1|, |\psi_1^\perp\rangle\langle\psi_1^\perp|\}$ and $\{|\psi_2\rangle\langle\psi_2|, |\psi_2^\perp\rangle\langle\psi_2^\perp|\}$ yields no information about the outcome of $\{|0\rangle\langle 0|+|1\rangle\langle 1|, |2\rangle\langle 2|\}$. Because of this, $2a^2$ of the epistemic states will be bound to $|0\rangle\langle 0|+|1\rangle\langle 1|$. This completes the proof. □

This result is perhaps less narrow than it appears at first blush, because it does put a constraint on deterministic ontological models that obey a notion of *counterfactual independence*, which we can define as follows.

Definition 21 (Counterfactual Independence)**.** An outcome-deterministic ontological model $[\Lambda, \{\mu\}, \{\Gamma\}, \{\xi\}]$ demonstrates *Counterfactual Independence* if for all measurements M, M', outcomes k, k', and pure preparations P we have

$$\int_\Lambda \xi_M(k|\lambda)\mu_P(\lambda)\,\mathrm{d}\lambda \int_\Lambda \xi_{M'}(k'|\lambda)\mu_P(\lambda)\,\mathrm{d}\lambda = \int_\Lambda \xi_M(k|\lambda)\xi_{M'}(k'|\lambda)\mu_P(\lambda)\,\mathrm{d}\lambda. \tag{4.28}$$

This property can be understood, as the name suggests, in terms of counterfactual reasoning. If one performs a known preparation procedure, and then performs a measurement and get some outcome, Counterfactual Independence encapsulates the idea that this outcome does not give any information about what the outcome would have been had one used a different measurement procedure; different measurement outcomes are independent of each other, conditional on the preparation distribution used.

It is clear that the Kitchen sink model does possess this independence property; if we prepare a known state ψ and measure with respect to some M and we get a certain outcome, then that tells us nothing about what would have happened counterfactually if we had instead measured M'. Some ψ-complete models for quantum mechanics, such as the Beltrametti-Bugowski model, or most Everettian viewpoints which perhaps are more physically reasonable examples of the property than the Kitchen Sink model. However, it is not true that all ψ-complete models have

this property: in both Bell's first model and also Bohmian mechanics, one can counterfactually reason about changes in the likelihoods outcomes of measurements that were not performed, based on an observed measurement. The Kochen-Specker model for a qubit can also be seen to lack this independence property.

Corollary 2. *There are no state update rules compatible with quantum mechanics for ontological models that are deterministic and obey Counterfactual Independence in quantum dimensions $d \geq 3$; consequently all such models in situations where an update rule is mandated, are ψ-ontic.*

Proof. The proof of Theorem 9 goes through with any such model obeying Counterfactual independence. □

Corollary 2 tells us that this counterfactual independence property is incompatible with deterministic models for quantum mechanics that are ψ-epistemic. This is of particular interest because classical ontologies, such as the noncontextual deterministic states of the sheaf-theoretic approach, meet these definitions in many finite subtheories, in the case that multiple preparations are considered. Counterfactual independence seems a reasonably intuitive physical property; the existence of contextuality within quantum mechanics teaches us that seeing a particular projector P occur as the outcome of some PVM should not necessarily be taken as any kind of evidence that P would have been the outcome in a different context. While strict statistical independence is of course a much stronger statement than not being exactly correlated, it does seem like a reasonable property to consider. In Chapter 5, similar questions will be asked, such as investigating how many different projectors from a quantum scenario can be considered to possess noncontextual valuations, and therefore to have a direct correlation between their valuation in different contexts. We note that in the terminology used in Chapter 3, the only generalised Kitchen Sink model that possesses Counterfactual Independence for pure states is the original Kitchen Sink model. We note that it is not true that it holds for proper mixtures of pure states; if we prepare a convex mixture where we have a $1/2$ chance of preparing $|0\rangle$ and a $1/2$ chance of preparing $|1\rangle$, then in a measurement $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$, seeing an outcome of $|0\rangle\langle 0|$

tells us that our preparation prepared $|0\rangle$, and our counterfactual probabilities must be updated to reflect this.

4.5 Discussion

Notions of contextuality, especially when paired with assumptions that seem intuitive or natural, can provide compelling restrictions on what the underlying ontology of reality can be, as well as restricting the probabilistic predictions that can exist in the theory. For example, we can interpret the fact that maximal contextuality exists within quantum theory as providing a strong restriction on the measurement properties of ontic states given measurement noncontextuality and determinism.

The status of locality within quantum theory has been one of constant debate within the foundations of physics community, fuelled in part by a lack of clear, consistent definitions. This chapter contained a specific view on the interpretation of local realism: that we must first have *realism* in the sense that we posit the existence of an actual ontic state of reality, independent of measurement choices or the presence of observers; and then that such states must be *local* in the sense that separated objects have separate descriptions in terms of states that are local to those systems, and that these states undergo (unitary and nonunitary) dynamics that are themselves local to those systems. Many do not share this interpretation. For example, in *Quantum Computation and Quantum Information* by Nielsen and Chuang [76], they give the following description of local realism into "local" and "realistic" parts, both halves of which they call "questionable":

1. The assumption that [physical properties] have definite values, which exist independent of observation. This is sometimes known as the assumption of *realism*.

2. The assumption that Alice performing her measurement does not influence the result of Bob's measurement. This is sometimes known as the assumption of

locality.

This could be considered somewhat of a classical view. Bell phrases his result in similar terms [17], applying what is in essence the EPR argument to conclude that the "result of [a measurement at a single site] must actually be predetermined". Bell formally assumes determinism in his proof of the theorem, justifying it by stating that there is in general no issues with assuming determinism of a single system. He might, however, have been served well by elucidating this point, because, as we shall see by returning to our previous example, it has caused the implications of the theorem to be misunderstood by many.

Let us look a little more carefully at the notion of realism defined by Nielsen and Chuang, as compared to the one used in this chapter. We can see that our definition is considerably weaker; we take a position of philosophical realism and state only that there is some definite state of reality that can be defined, distancing ourselves from positions such as quantum bayesianism which do not posit the existence of a state of reality that all observers should in principle agree on, whereas the definition of Nielsen and Chuang goes further, positing that that state defines definite, deterministic properties of reality that can be revealed by measurement. This goes further, and can be thought of as an assumption of noncontextuality. In particular, this assumption is not even true for single systems, as Bell knew, having already demonstrated that quantum mechanics in $d > 3$ was contextual prior to his celebrated theorem on the issue of locality. I would posit that it was not the intention of Bell to define realism in this way.

Faced with dropping either "realism" in the form of noncontextuality, or the principle of locality, is it any surprise that they conclude that "[m]ost physicists take the point of view that it is the assumption of realism which needs to be dropped from our worldview in quantum mechanics"? If dropping the already-falsified assumption of noncontextuality, and therefore determinism, were really all that it took to fully recover locality for quantum mechanics, that would be very uninteresting indeed. Another perspective on this comes from Abramsky and Brandenburger's extension of Fine's theorem [2], as they demonstrate that the introduction of *local* irreducible randomness does not yield any additional explanatory power in nonlocality experiments.

The search for ψ-epistemic models for quantum mechanics was arguably started by Einstein and Schrödinger, who noted in their correspondence [37, 54] that locality of the type looked for in EPR [40] indicated that the ontic states of quantum systems must be compatible with multiple different quantum states. From this point of view, Bell's theorem provides a significant restriction to such ψ-epistemic models, and tells us something about the nonlocality we must accept in the kinematics or dynamics of such a model. Perhaps the first such restrictions were set by Gleason's theorem, however, which provides a very severe constraint on any models for subtheories of quantum mechanics that obey noncontextuality for measurements. There are parallels between the Gleason-type approaches to ruling out ψ-epistemic models for quantum mechanics and the approaches we have seen here, for example those of PBR. Given two states, they demonstrate a finite set of measurements that are sufficient to rule out ψ-epistemic interpretations for subtheories containing those states and those measurements, and demand resources (either quantum dimensions or additional contexts) that depend on the squared overlap of the two states under consideration.

In our discussion of Theorem 8, we introduced a trichotomistic notion of contextuality that we may adopt if we do not wish to accept a notion of possibilistic measurement noncontextuality that includes coarse-grainings. This notion is somewhat reminiscent of the three-valued logic approach taken by Reichenbach [86], but ultimately reflect the structure of the underlying quantum logic rather than nonstandard classical logics created to reason about quantum structures. Reichenbach's approach to resolving the differences between the Boolean algebra of classical events and the subspace algebra of quantum events was to introduce a system of classical logic that was more like the quantum logic, rather than the Einstein approach of seeking models for quantum mechanics that were more classical. To do this, he introduced an "intermediate" logical value, that applied to statements that were not "unknown", but neither true nor false. This approach was later analysed and endorsed by Putnam [84]. The approach of Reichenbach is somewhat different, however, to the approach needed to motivate that particular notion of contextuality. Reichenbach's "true" and "false" statements apply only to statements made about measurements that have actually occurred; knowledge that a quantum state is in the the

relevant support of a projector is not enough for Reichenbach to assign a value of truth to the proposition about the system corresponding to that projector. Any statements about unperformed measurements are considered indeterminate, echoing the maxim of Peres "unperformed experiments have no results". This avoids the issue wherein the disjunction of two intermediate propositions can be either indeterminate or true (or *certain*, in our noncontextuality framework), which is necessary for the algebraic structure. However, it seems perhaps unsatisfactory that quantum systems have no properties except at the moment of measurement; in particular, it suggests a conceptual difference between preparing a quantum state by measurement or by other constructions, while this distinction is meaningless in most interpretations On the other hand, while these features are in contrast to this three-valued logic approach, the notion of noncontextuality discussed maintains common definitions of truth and falsehood in quantum logic [19, 52], in which a statement is true if it holds in a measurement with probability 1, and false if it holds with probability 0. These quantum logics, however, are harder to reconcile with the Boolean classical logic and are arguably less useful for interpretation.

It is worth discussing the apparent impact on Corollary 2 for the ψ-epistemic status of the interpretation of quantum mechanics called QBism, or Quantum Bayesianism. This is an interpretation of quantum mechanics that is agent focused and can be thought of as a quantum generalisation of the subjective Bayesianism interpretation of classical probabilities, in which probability is defined only as quantifying a personal belief. Restrictions on which state assignments are considered rational are provided by arguments from decision theory, or Dutch book arguments. In quantum Bayesianism, the Born rule is a normative statement containing information about how one *ought* to ascribe probabilities to events in a globally rational manner, rather than a statement about objective probabilities. It could be argued that the Kitchen Sink model provides something akin to an ontology for QBism; agents believe that if a measurement is made, there is a definite outcome to that measurement, and the interpretation supports the notion of counterfactual independence as defined above. We see that an agent's beliefs about the outcomes of experiments are faithfully expressed by an epistemic distribution over Kitchen-Sink ontic states; since we also have counterfactual independence we might conclude

that we are left with exactly the Kitchen Sink model. This model, however, is ψ-ontic, which is seemingly in contradiction to the ψ-epistemic nature of the theory, in which quantum states really do just represent epistemic beliefs of rational agents. Interestingly, however, this logic does not follow because of how the Kitchen Sink model and a QBist agent differ in their state assignments when dealing with mixed quantum states. In the Kitchen Sink model, since the quantum state is ontologically instantiated, different mixtures of pure states that yield the same density matrix are ontologically distinct. A Bayesian agent, however, can perform calculations directly using the mixed state that represents his degrees of belief. The Kitchen Sink model reproduces the convex mixing of preparations at the ontological level, which is axiomatic to the ontological models formalism. However, QBist agents do not need to use these notions of convexity when expressing their credences; care must be taken when applying to QBism notions that were defined for the ontological models formalism. This logic holds only when dealing with a pure quantum state, which is not in contravention with QBist principles since it could be argued that in that case it has a physical representation within the brain of the agent. It is also possible to argue that QBist agents do not have to ascribe to counterfactual independence; since the Born rule is viewed as a normative rule about placement of bets, and one cannot bet on counterfactual events, one can argue that QBists can assign any probabilities to counterfactual events without being inconsistent.

Chapter 5

Metrics of Contextuality and Nonlocality

In John Bell's seminal *On the Einstein Podolsky Rosen Paradox* [17], he demonstrates that a realist interpretation of quantum mechanics cannot, in some sense, be local. Bell, however, is not visibly concerned with quantitative questions about quantum nonlocality, and is content to demonstrate that quantum mechanics is somehow qualitatively nonlocal. Even Clauser, Horne, Shimony, and Holt [30], who proposed a direct experimental test of Bell's proposition, and characterised what is now known as the CHSH inequality, were content to consider only simple possibilities of quantum measurements on two entangled qubits, and were interested in maximum violations only to give the experiment as large a chance of succeeding as possible despite noise.

In fact, the question of how nonlocal quantum mechanics was, compared to other possible physical theories, was not raised until 1978, when Boris Tsirelson was giving a seminar about Bell inqualities at the Leningrad department of the Steklov Institute of mathematics at the USSR academy of sciences [58, 100, 102]. The chairman, Anatoly Vershik, asked him that, given that the CHSH inequality is bounded within classical theory, is it bounded within quantum theory, compared to other theories in theoryspace? Tsirelson was reluctant to grant the idea merit;

the point was that quantum observations ruled out classical, local, theories. He thought it only useful in the case that in the future some superquantum observations would be found, and similarly dismiss all quantum theories. Nonetheless, two years later in 1980, Tsirelson published a paper containing the result now known as Tsirelson's bound; quantum mechanically, the maximum value for the CHSH inequality is $2\sqrt{2}$. This marked the first time that the question of the extent of the nonlocality within quantum mechanics, was explored, albeit compared to the very specific yardstick if the CHSH inequality. Since then, a lot of progress has been made applying this kind of quantitative approach to nonlocality, and more generally to contextuality, to quantum mechanics as an entire theory and to specific quantum subtheories. The contextual fraction also relates directly to the violation of some Bell inequalities. This is fairly suggestive of a general class of measures of nonclassicality, based on the differentiability of a set of probabilistic predictions from the classical polytope.

5.1 Quantifying Nonlocality and Contextuality

The nonclassical properties of quantum mechanics are arguably most interesting when they allow some operational tasks to be conducted with speed, efficiency, or security that nature does not afford to classical mechanics. As such, a measure of nonlocality or contextuality in terms of what can be achieved over what is allowed within classical mechanics, equipped with that resource is an appealing metric to use for that quantity; it allows us to think of intangible concepts such as entanglement in a way more analogous to how we think of petrol in litres. Arguably, we already take this view for more familiar concepts such as bandwidth and information storage. Similarly, we can investigate if superquantum resources can offer advantages that are not available within the confines of quantum mechanics.

Resource theories have been successfully established for nonlocality scenarios, particularly by considering entanglement as a resource. By considering the operational power of entangled resources, entanglement is able to be quantified in a way that is meaningful; we consider the entangled resources as a fuel that is consumed by our information processing. Recently, the *con-*

textual fraction has been used as a measure of contextuality [5], motivated by clear operational advantage in information processing tasks and games. The contextual fraction can be calculated by linear methods, and is calculated by decomposing an empirical model into a noncontextual and contextual part. The contextual fraction is the smallest fraction of this decomposition that can be assigned to the contextual part.

The work of Anders and Browne [8] showed that maximal contextuality, of the sort found in the Greenberger-Horne-Zeilinger state, could also be seen as a resource enabling nonlocal classical computation in a measurement-based-computation scenario. To perform measurement-based quantum computation in the usual setting requires adaptive measurements to be made on quantum systems. However, the computer tasked with keeping track of the necessary adaptations does not need to be endowed with the ability to perform universal classical computation; it is enough for it to keep track of parities which can be done with only access to XOR or CNOT gates. Such a computer can calculate the parity of bit strings, but it can also perform simulations of quantum circuits with gates confined to those in the Clifford group [1], using the Gottesman-Knill theorem. This is suggestive, since despite demonstrating many of the properties often given credit for improved computational power, such as large-scale superposition; or entanglement including maximal entanglement; and maximal contextuality for example the Peres-Mermin magic square; Clifford circuits are simulable classically and do not offer increased computational power.

5.2 Robustness of Contextuality

Almost as an exact parallel to the case of nonlocality, as described above the proofs of contextuality due to Bell, and to Kochen and Specker, had a qualitative flavour. They defined a binary notion of contextuality, one that a theory either obeyed or did not; there is no quantitative element, or way of comparing different theories that obey this condition. As mentioned in the previous section, the notion of *maximal contextuality* corresponds to a contextual fraction of 1, the highest possible. In this section, following Reference [91], *"How (Maximally) Contextual*

is Quantum Mechanics?", we will define metrics of contextuality that discriminate between different physical theories or subtheories possessing a contextual (or nonlocal) fraction of 1.

5.2.1 How maximally contextual is quantum mechanics?

Put simply, the Bell-Kochen-Specker theorem dictates that there exist subsets of the quantum mechanically available projectors $\{|\psi_i\rangle\langle\psi_i|\}$ that do not support a noncontextual value assignment, that is, a function $f : \{|\psi_i\rangle\langle\psi_i|\} \to \{0, 1\}$, such that

$$\forall C \in \mathcal{C}, \quad |f^{-1}(1) \cap C| = 1, \tag{5.1}$$

where $\mathcal{C}$ is the set of all contexts and $f^{-1}(1)$ denotes the subset of those $|\psi\rangle\langle\psi| \in \{|\psi_i\rangle\langle\psi_i|\}$ such that $f(|\psi\rangle\langle\psi|) = 1$. That is to say, f has the property that in every context C, there is exactly one projector that is assigned a 1 valuation by f. This allows us to interpret f as a set of noncontextual predictions for a measurement, and interpret a 1-valuation for a projector by saying that no matter the context in which we measure that projector, we will see it as the outcome of the measurement. The projector can be thought of as representing a property of the system. This is analogous to the classical case in which such a property might be the object being within a certain area or not; if the object is in the area A, it does not matter how we divide up the rest of the universe in our measurement, we will always find the object in A. We can define a quantitative version of this notion by allowing a contextual valuation; that is to say a function $g : \{|\phi_i\rangle\langle\phi_i|\} \times \mathcal{C} \to \{0, 1\}$.

From this g, we can define a contextual version of $f : \{|\psi_i\rangle\langle\psi_i|\} \to \{0, 1, c\}$ as follows:

$$f(|\psi\rangle\langle\psi|) = \begin{cases} 0 & \text{if } g(|\psi\rangle\langle\psi|, C) = 0 \quad \forall\ C \in \mathcal{C} \\ 1 & \text{if } g(|\psi\rangle\langle\psi|, C) = 1 \quad \forall\ C \in \mathcal{C} \\ c & \text{o/w.} \end{cases} \tag{5.2}$$

That is, if g assigns a particular value to $|\psi\rangle\langle\psi|$ in a noncontextual way, then the f assigns

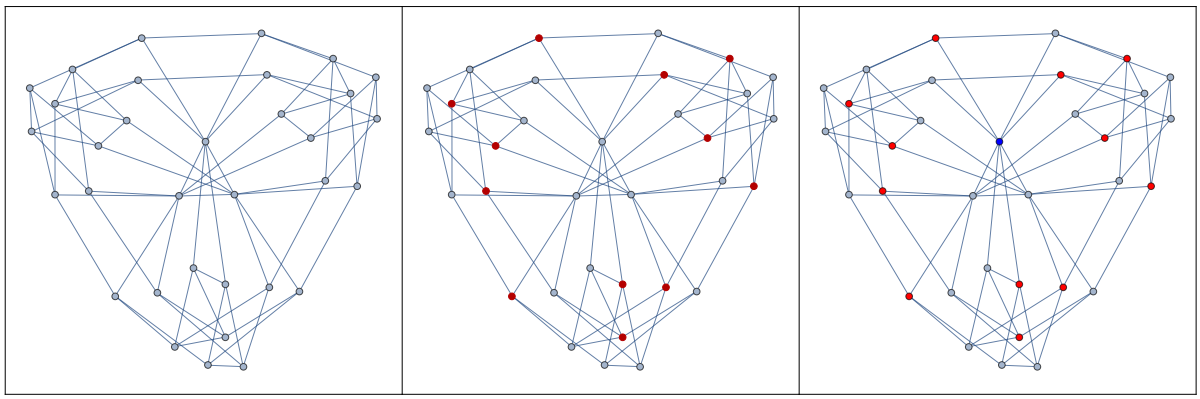


Figure 5.1: Three images illustrating the calculation of the figure of merit for the 33-ray proof of the Kochen-Specker theorem due to Peres [77]. The first is the orthogonality graph with no special highlighting. In the second, a maximal independent set is highlighted in red; this represents an assignment of noncontextual 1-valuations that cannot be extended without having a context containing two or more 1-valued projectors. We see that only the single central context lacks a 1-valued projector. In the third, a projector whose value can be taken to be contextual is highlighted in blue; this provides a contextual hidden variable model for the scenario if it takes a 1-valuation when measured in the central context and a 0-valuation otherwise. This contextuality proof has the figure of merit $q_s(\mathcal{G}) = 0.\dot{0}\dot{3}$.

$|\psi\rangle\langle\psi|$ the same valuation. If the valuation given by g to $|\psi\rangle\langle\psi|$ is in fact context dependent, then f will assign a label of c, for "contextual", to $|\psi\rangle\langle\psi|$. Considering only this f function, then, we want it to have the following properties in order to be able to interpret it as giving a prediction for an experiment:

$$\forall C \in \mathcal{C}, \quad |C \cap f^{-1}(0)| < |C|, \tag{5.3}$$

that is, there is at least $|\psi\rangle\langle\psi|$ in any C that has a nonzero assignment, representing the fact that no matter which context is measured there must be an outcome; and

$$\forall C \in \mathcal{C}, \quad |C \cap f^{-1}(1)| \leq 1, \tag{5.4}$$

that is, there is at most one element from each C that is given a 1-assignment, representing the fact that for any context, there cannot be two possible outcomes that are both certain to occur.

Past proofs that quantum mechanics is maximally contextual have focussed only on showing

that there must be one c-valued projector amongst the "Kochen-Specker set" used. We note that in the past, graph theoretic methods have been shown to be very effective at capturing the structures that lead to maximal contextuality [10, 24]. We will see that graph-theoretic methods will be useful in the case of this figure-of-merit, also. In particular, we will demonstrate that while there is a strict quantum mechanical bound on the robustness of maximal contextuality, there is no inherent limit imposed by graph theory on what is attainable in a larger theoryspace.

As before, contexts will appear in our representation as *maximum cliques*, connected subgraphs of maximum size which is denoted $\omega(\mathcal{G})$. In a quantum system of dimension d, there can be at most d mutually independent vectors, and this forms a measurable basis, so as long as the set of projectors we are considering contains at least one measurable basis, we will have $\omega(\mathcal{G}) = d$, noting that in general we have $\omega(\mathcal{G}) \leq d$. Motivated by our concern that in our contextual model for a scenario, there must be at least one projector in each basis that can take a noncontextual or a contextual 1-valuation, we are motivated to define the following, taken from reference [91]:

Definition 22 (Maximum-clique hitting-set)**.** For a graph $\mathcal{G} = \{V, E\}$, a vertex subset $T \subset V$ is a *maximum-clique hitting-set* or *maximum-clique transversal* if for all cliques C, $|C| = \omega(\mathcal{G})$, $T \cap C \neq \emptyset$.

We are here considering partial transversals; as noted by Abramsky and Brandenburger [2], the Kochen-Specker theorem can be represented in a similar way using stable transversals. Since maximum cliques are contexts, the set of projectors that are given either a 1 or c valuation must form a maximum-clique hitting set. This allows us to begin to define our figure-of-merit graph theoretically:

$$q_s(\mathcal{G}) = \min_{T} \min_{A \subset T} |T| - |A|, \tag{5.5}$$

where the minimisations are performed over all maximum-clique hitting-sets T, and all their independent subsets A. This requires some explanation: consider the set of vectors which are noncontextually assigned a 1-valuation. No two of them can appear in a context together, or there would be some context that had two projectors that were certain to appear as the measurement outcome. More stringently, two such vectors can never be orthogonal to each

other; this is equivalent if our set of projectors has the property that any two orthogonal projectors can be extended to a basis,which many historically relevant KS-sets possess. In the literature, this is known as a weak KS-set [87], and can be motivated either by the idea that weak-KS sets can be extended to have the aforementioned property by adding only $O(n^2 d)$ extra projectors, where $n = |\{|\psi\rangle\langle\psi|\}|$, the number of projectors. Alternatively, incomplete bases can be considered as completed by (possibly coarse-grained projectors) outside the scope of the measurement scenario, whose valuations are allowed to vary contextually.

This quantity, then, has a direct interpretation in terms of contexts; we seek a set, T, of projectors that are not noncontextually assigned 0 valuations, and we subtract from that set the number of those projectors that can support a noncontextual 1 valuation. We are left with the number of vectors that must be assigned a c valuation. However, this particular figure-of-merit suffers from an undesirable scaling property that means it is trivial to achieve an arbitrarily high q_s. This is done by taking an equal number of noninteracting copies of some contextuality scenario, such as Peres's or Penrose's 31-vector solutions. In graph theoretic terms, we note that this figure-of-merit is additive under disjoint graph union: $q_s(\mathcal{G} \oplus \mathcal{H}) = q_s(\mathcal{G}) + q_s(\mathcal{H})$. Quantum mechanically we note that if we supplement our set of projectors $\{|\psi_i\rangle\langle\psi_i|\}$ by taking the union with $\{U|\psi_i\rangle\langle\psi_i|U^\dagger\}$, for some specially chosen U such that for no pair of $|\psi_j\rangle\langle\psi_j|$, $|\psi_k\rangle\langle\psi_k|$ in $\{|\psi_i\rangle\langle\psi_i|\}$, we have $\mathrm{Tr}(|\psi_j\rangle\langle\psi_j|U|\psi_k\rangle\langle\psi_k|U^\dagger) = 0$, then we double our valuation of q_s. Slightly more generally, q_s is additive under taking a union of two contextuality scenarios, as long as this orthogonality condition is respected. We can think of this as a generalisation of a distance measure as discussed above, where our distance is defined as the number of deviations from a classical framework.

In order to avoid this scaling behaviour, we will define our main quantity of interest to be the *graph density* of q_s:

$$q(\mathcal{G}) = \frac{q_s(\mathcal{G})}{|V(\mathcal{G})|}. \tag{5.6}$$

We are of course interested in the range of values of this quantity that are attainable in different physical theories: classically; quantum mechanically; and in other nonclassical theories. We are also interested in the calculation of this quality, especially with respect to its computational

complexity classification.

We note that the following lower bound is easy to derive:

$$q_s(\mathcal{G} = \min_T \min_{A \subset T} |T| - |A| \geq \min_T |T| - \min_A(-|A|), \tag{5.7}$$

where this inequality holds since by relaxing the requirement that our independent set A be a subset of the maximum-clique hitting-set A we potentially allow further minimisation;

$$q_s(\mathcal{G}) \geq \min_T |T| + \max_A(|A|) = \min_T |T| - \alpha(\mathcal{G}). \tag{5.8}$$

This bound, then, is the difference between the sizes of the smallest maximum-clique hitting-set and the largest clique of the graph $\mathcal{G}$. We note that it is **NP**-complete to calculate the value of $\min |T|$, even if we know the independence number $\omega(\mathcal{G})$. This can be seen since for triangle-free graphs, maximum-cliques are edges; this problem is therefore equivalent to finding a minimal vertex cover: this was shown by Poljak to be **NP**-complete [82]. In addition, the independence number itself, $\alpha(\mathcal{G})$ is well-known to be **NP**-complete, being one of Karp's original 21 **NP**-complete problems, in the guise of the Clique problem [43, 56].

We will now demonstrate an upper bound on the quantity $q(\mathcal{G})$ for any quantum-mechanically accessible $\mathcal{G}$ in terms of the quantum dimension of the vectors that give rise to it. This proof is reproduced with few changes from reference [91], "*How (maximally) contextual is quantum mechanics?*".

Theorem 10. *Any orthogonality graph $\mathcal{G}$ induced by a set of vectors $|\psi_i\rangle$ in a d-dimensional Hilbert space has*

$$q(\mathcal{G}) \leq \left(1 - \frac{1}{d}\right)^{d-1} - \frac{1}{2^{d-1}}. \tag{5.9}$$

Proof. The proof proceeds by demonstrating an explicit procedure for finding a T and an $A \subset T$, for any $\{|\psi_i\rangle\}$, with upper-bounded worst-case performance.

First, we wish to find some region of the Hilbert space such that for any basis, at least one

element of the basis lies in the region. This region will be of the form of a closed complex hyperspherical cap centred at some arbitrary state $|\psi\rangle$; that is the set of vectors $|\phi\rangle \in \mathcal{H}_d$ such that $|\langle\psi|\phi\rangle|^2 \geq t$ for some $t \in \mathbb{R}$. We want to choose some amount of squared overlap t so that given our arbitrary quantum state $|\psi\rangle$, its overlap with one of the elements of our basis is at least t. The minimal such t can be seen to be $1/d$, where d is the quantum dimension of our space; when we perform a measurement on $|\psi\rangle$ with respect to any basis, there must be at least one of the d outcomes with a probability greater than or equal to $1/d$.

This puts an upper bound on the size of the largest angle that can separate an arbitrary vector from its closest basis element in any given basis. A closed complex hyperspherical cap subtending this angle, $\left\{|\psi\rangle \middle| |\langle\psi|\phi\rangle|^2 \geq 1/d\right\}$, will then necessarily capture at least one element from the basis, and so any set of vectors within such a hyperspherical cap C constitutes a maximum-clique hitting-set, T_C. We wish, now, to calculate the proportion of the Hilbert space that lies within such a spherical cap under the Haar measure.

Since columns of a Haar-random unitary matrix over $\mathbb{C}^d$ are uniformly random unit vectors, the volume V of this hyperspherical cap as a proportion of that of the sphere is given by $1 - F(t)$, where $F(t)$ is the cumulative distribution function of the squared modulus of an entry in a Haar-random unitary. This quantity is well known in the field of random matrix theory and it can be found, for example, as a special case of a result by Życzkowski and Sommers [107], and will be revisited below.

$$\frac{V}{V_S} = (1-t)^{d-1}\,. \tag{5.10}$$

Clearly, the value of $q(\mathcal{G})$ for a quantum graph is upper bounded by the size of any maximum-clique hitting set, so if we can bound the minimum number of vectors we must intersect with such a hyperspherical cap, this also constitutes a bound on the value $q(\mathcal{G})$. However, we would not expect this bound to be tight, since we have ignored the independent set subtrahend entirely. We would like to demonstrate an independent set within our maximum-clique hitting-set T_C, and therefore give a better upper bound on the quantity $q_s(\mathcal{G}) = \min_T \min_{A \subset T} |T| - |A|$, giving

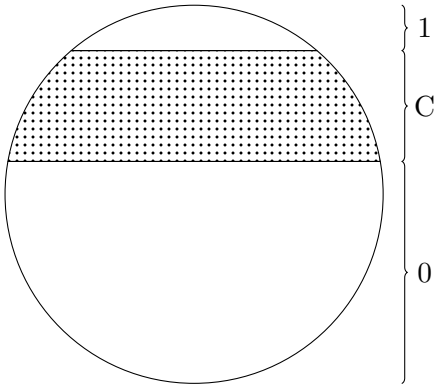


Figure 5.2: An illustration of how our construction to assign potentially contextual values to projectors splits up the Hilbert space into different regions, that can be associated with projectors given noncontextual 1-assignments, contextual projectors, and projectors given noncontextual 0-assignments.

an explicit construction procedure for a T and $A \subset T$.

We note that a hyperspherical cap containing the vectors $|\phi\rangle$ such that $|\langle\psi|\phi\rangle|^2 > 1/2$ cannot contain two orthogonal vectors as a simple application of the triangle inequality. We can place this smaller hyperspherical cap anywhere within the larger hyperspherical cap, and this will form an independent set A that is a subset of T_C. This might put us in mind of the colouring proof of the Bell-Kochen-Specker theorem from Chapter 4; the freedom to have contextually labelled projectors rather than 1-labelled outside of this spherical cap allows us to sidestep the argument of Lemma 2 and build a consistent labelling. For the rest of this proof, we will assume that the two hyperspherical caps are centred on the same point; this will make no difference to the worst case scenario which forms the bound. The region of Hilbert space, then, in which vectors count towards the figure of merit, is a member of a two-parameter family of *annuli* on the complex hypersphere, illustrated in Figure 5.2 and defined by

$$t_1 \leq |\langle\psi|\phi\rangle|^2 \leq t_2. \tag{5.11}$$

Such an annulus takes up a proportion

$$p(t_1, t_2) = (1 - t_1)^{d-1} - (1 - t_2)^{d-1} \tag{5.12}$$

of the Hilbert space. Our construction has $t_1 = 1/d$, and $t_2 = 1/2$, giving us a hyperspherical annulus defined by $1/d \leq |\langle\psi|\phi\rangle|^2 \leq 1/2$. It now remains to be proved that no matter the positions of the set of vectors $\{\psi\}$, there must be some place in which we can place this annulus such that we only intersect strictly less than $p(t_1, t_2)$ of them.

Consider the canonical action of $SU(d)$ on the unit hypersphere in $\mathbb{C}^d$, which we shall denote $S^{d-1}_{\mathbb{C}}$. We will integrate over the space $SU(d) \times S^{d-1}_{\mathbb{C}}$, where $SU(d)$ is equipped with the Haar measure, and $S^{d-1}_{\mathbb{C}}$ is equipped with the uniform measure. Equivalently, we could have taken H_d rather than $S^{d-1}_{\mathbb{C}}$, in which case it would be equipped with the Fubini-Study metric.

At each $|\psi_i\rangle$, we place a bump function $f^{\epsilon}_{|\psi_i\rangle}$ with unit integral and support only within a disc of radius ϵ around $|\psi_i\rangle$. Let A_{t_1,t_2} denote the hyperspherical annulus $\{|\phi\rangle \mid t_1 \leq |\langle 0|\phi\rangle|^2 \leq t_2\}$. By Fubini's theorem, we have:

$$\int_{SU(d)} \int_{S^{d-1}_{\mathbb{C}}} \mathbb{1}_{g.z \in A_{t_1,t_2}} \sum_i f^{\epsilon}_{|\psi_i\rangle}(z)\, \mathrm{d}z\, \mathrm{d}g = \int_{S^{d-1}_{\mathbb{C}}} \int_{SU(d)} \mathbb{1}_{g.z \in A_{t_1,t_2}} \sum_i f^{\epsilon}_{|\psi_i\rangle}(z)\, \mathrm{d}g\, \mathrm{d}z \tag{5.13}$$

We note that if we extend our annulus's extent by ϵ, so that we have $t_1 \to t_1 - \epsilon$ and $t_2 \to t_2 + \epsilon$, we capture the entirety of the measure of the bump functions centred inside the original annulus. Hence, for all $g \in SU(d)$ we have $\left|\left\{|\psi_i\rangle \middle| |\psi_i\rangle \in g^{-1} A_{t_1,t_2}\right\}\right| \leq \int_{S^{d-1}_{\mathbb{C}}} \mathbb{1}_{g.z \in A_{t_1-\epsilon,t_2+\epsilon}} \sum_i f^{\epsilon}_{|\psi_i\rangle}(z)\, \mathrm{d}z$, and so we have

$$\int_{SU(d)} \left|\left\{|\psi_i\rangle \middle| |\psi_i\rangle \in g^{-1} A_{t_1,t_2}\right\}\right| \mathrm{d}g \leq \int_{S^{d-1}_{\mathbb{C}}} \int_{SU(d)} \mathbb{1}_{g.z \in A_{t_1-\epsilon,t_2+\epsilon}} \sum_i f^{\epsilon}_{|\psi_i\rangle}(z)\, \mathrm{d}g\, \mathrm{d}z, \tag{5.14}$$

Rearranging:

$$\int_{SU(d)} \left|\left\{|\psi_i\rangle \middle| |\psi_i\rangle \in g^{-1} A_{t_1,t_2}\right\}\right| \mathrm{d}g \leq \sum_i \int_{S^{d-1}_{\mathbb{C}}} f^{\epsilon}_{|\psi_i\rangle}(z) \int_{SU(d)} \mathbb{1}_{g.z \in A_{t_1-\epsilon,t_2+\epsilon}}\, \mathrm{d}g\, \mathrm{d}z \tag{5.15}$$

Taking the average of the indicator function $\mathbb{1}_{g.z \in A_{t_1-\epsilon,t_2+\epsilon}}$ with respect to the Haar-random g

over $SU(n)$ reduces it to a constant function with value $p(t_1 - \epsilon, t_2 + \epsilon)$, the proportion of the Hilbert space taken up by the annulus.

$$\int_{SU(d)} \left|\left\{|\psi_i\rangle \middle| |\psi_i\rangle \in g^{-1} A_{t_1,t_2}\right\}\right| \mathrm{d}g \leq p(t_1 - \epsilon, t_2 + \epsilon) \sum_i \int_{S^{d-1}_{\mathbb{C}}} f^{\epsilon}_{|\psi_i\rangle}(z)\, \mathrm{d}z \tag{5.16}$$

$$\leq p(t_1 - \epsilon, t_2 + \epsilon)\, ||\psi_i\rangle| \tag{5.17}$$

Since the left-hand side of Equation 5.16 represents an average over the rotation group, there must be some $g \in SU(d)$ such that $\left|\left\{|\psi_i\rangle \middle| |\psi_i\rangle \in g^{-1} A_{t_1,t_2}\right\}\right| \leq p(t_1 - \epsilon, t_2 + \epsilon)\, |\{|\psi\rangle\langle\psi|\}|$, and this is true for all ϵ. Therefore:

$$q(\mathcal{G}) \leq \lim_{\epsilon \to 0} p(t_1 - \epsilon, t_2 + \epsilon) \tag{5.18}$$

$$\leq \left(1 - \frac{1}{d}\right)^{d-1} - \frac{1}{2^{d-1}}. \tag{5.19}$$

□

Corollary 3. *Any quantum mechanically accessible $\mathcal{G}$ has $q(\mathcal{G}) \leq {}^{4251920575}/_{11019960576} \sim 0.385838$.*

Proof. Calculating the derivative of $\left(1 - \frac{1}{d}\right)^{d-1} - \frac{1}{2^{d-1}}$ with respect to d shows that it takes a maximum between $d = 9$ and $d = 10$. Evaluation of the quantity at these points reveal the maximum to be at $d = 9$, when we get $(1 - {}^1/_9)^8 - {}^1/_{2^8} = {}^{4251920575}/_{11019960576}$. This, then, forms a hard limit of $q(\mathcal{G})$ for any quantum mechanically achievable $\mathcal{G}$. □

We note that as the quantum dimension approaches infinity, we achieve a limiting value of ${}^1/_e$, so quantum systems with very high dimension are viable candidates for providing robust contextuality scenarios. In Figure 5.4, we see the bound on $q(\mathcal{G})$ plotted as a function of d.

Name	Quantum Dimension	Transversal Size per copy	Independence Number per copy	Figure of Merit
ABK Triangle-Free Graphs [7]	✗	$1 - c\sqrt{\frac{\log n}{n}}$	$c\sqrt{\frac{\log n}{n}}$	$1 - O\left(\frac{\log n}{n}\right)$
Quantum Upper Bound	d	$\leq \left(1 - \frac{1}{d}\right)^{d-1}$	$\geq \frac{1}{2^{d-1}}$	≤ 0.385838
E_8 root system	8	$\leq \frac{11}{60}$	$\frac{1}{15}$	$\leq 0.11\dot{6}$
Two-Qubit Stabiliser Quantum Mechanics	4	$\frac{3}{10}$	$\frac{1}{5}$	0.1
Peres-Mermin Magic Square [77]	4	$\frac{7}{24}$	$\frac{5}{24}$	$0.08\dot{3}$
Cabello's 18-ray proof [26]	4	$\frac{5}{18}$	$\frac{4}{18}$	$0.0\dot{5}$
Peres's 33-ray proof [77]	3	$\frac{9}{33}$	$\frac{12}{33}$	$0.\dot{0}\dot{3}$

Figure 5.3: A table comparing a number of contextual constructs; each concrete example represents the graph union of multiple noninteracting copies of the KS proof. Two-qubit stabiliser quantum mechanics forms a KS proof using 60 rays and 105 contexts; the E_8 root system forms a KS proof using 120 rays and 2025 contexts. The quantities related to each were calculated using a proof by exhaustion, or are the best found during a non-exhaustive search, such as is the case for the E_8 system. It is perhaps of note that the lower bound of the size of a traversal less the independence number forms a trivial bound only for Peres's 33-ray proof, and in fact is optimal for Cabello's 18-ray proof and the Peres-Mermin magic square.

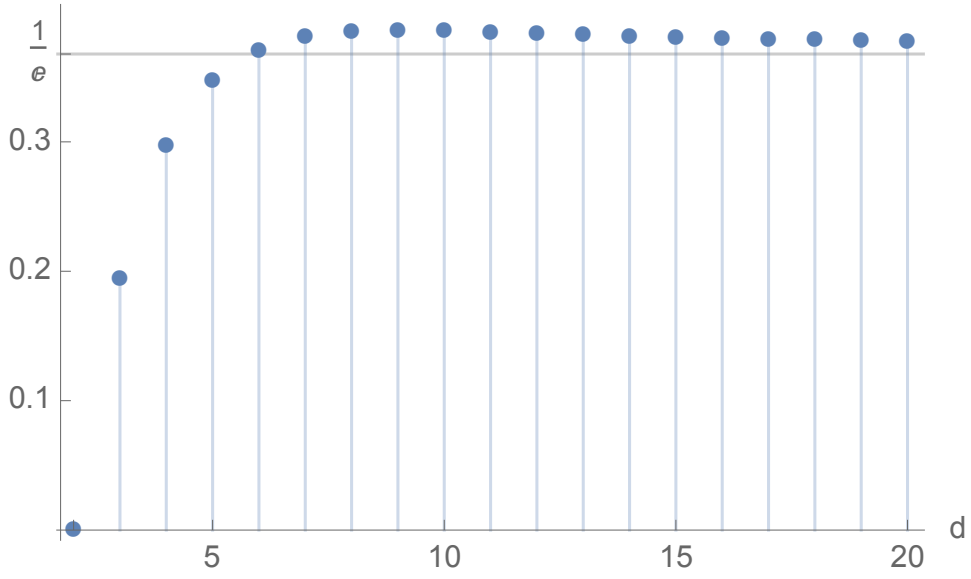


Figure 5.4: A chart showing the value of the bound on $q(\mathcal{G})$ for quantum-mechanically accessible $\mathcal{G}$ as a function of the Hilbert space dimension d.

5.2.2 Non-Quantum Graphs

Analogously to Tsirelson's bound, we wish to ask whether theories other than quantum theory could demonstrate more contextuality, by our metric. In order to do this, we will demonstrate that there is no graph-theoretic obstruction to super-quantum values of this figure of merit. In order to do so, we will consider triangle free graphs, for whom the figure of merit reduces to one previously studied in the literature of graph theory.

In triangle-free graphs, the maximum cliques are merely edges, and so each context consists of two measurement objects. Quantum mechanically, the only such graphs are uninteresting: the disjoint graph union of K_2 graphs. These do not have enough structure to form a proof of the Bell-Kochen-Specker theorem; this reflects the fact that each projector appears in only a single context. Many such graphs, then, are not quantum-mechanically accessible. However, they do allow us to prove a concrete bound on what values of $q(\mathcal{G})$ are achievable within the framework of generalised probabilistic theories. In such graphs, the concept of a maximum-clique hitting-set reduces to that of a covering of edges by vertices, known as a *vertex cover*. The size of the minimal vertex cover of $\mathcal{G}$ is denoted $\tau(\mathcal{G})$. This reduction of our figure-of-merit allows us to make use of a bound derived for Ramsey type problems.

In the triangle-free case, we have a powerful relation between maximum-clique hitting-sets and independent sets; namely that they are graph complements of each other. A maximum-clique hitting-set, or vertex cover, is a set of vertices T such that for every edge, at least one of that edge's vertices is in T. We can see, then, that the set $V - T$ must be independent, since if there were two vertices in $V - T$ connected by an edge, then neither of those vertices would be in T, a contradiction. Conversely, if we have an independent set A, then $V - A$ must be a vertex cover; if there were an edge with neither vertex in $V - A$, then both vertices are in A, a contradiction. This means that we can bound $q_s(\mathcal{G})$ as

$$q_s(\mathcal{G}) \geq \min_{\mathcal{G}} |T| - \alpha(\mathcal{G}) = |V(\mathcal{G})| - 2\alpha(\mathcal{G}). \tag{5.20}$$

In other words, to be able to lower bound $q_s(\mathcal{G})$, we need only consider the independence number,

$\alpha(\mathcal{G})$. We seek, therefore, triangle-free graphs with low independence number. We can invoke here a theorem due to Alon, Ben-Shimon and Krivelevich [7].

Theorem 11. *There exists a constant c such that for all $n \in \mathbb{N}$ there exists a regular triangle-free graph $\mathcal{G}_n$, with $V(\mathcal{G}_n) = n$, and $\alpha(\mathcal{G}_n) \leq c\sqrt{n \log n}$.*

Hence, we have

$$\lim_{n\to\infty} q(\mathcal{G}_n) \geq \lim_{n\to\infty} \frac{n - 2c\sqrt{n \log n}}{n} \tag{5.21}$$

$$\geq 1 - 2c \lim_{n\to\infty} \sqrt{\frac{\log n}{n}} = 1. \tag{5.22}$$

Therefore, $q(\mathcal{G}_n) = 1$ is approachable in the limit of $n \to \infty$. We see then that quantum mechanics does not display maximal contextuality as robustly, given this definition of robustness, as other possible GPTs. This could potentially have an impact on attempts to recreate quantum mechanics as a principle theory.

We now extend the result to apply to projective valued measures including projectors of rank greater than 1, although we are still considering the case in which contexts are made up of a set of projectors which sum to the identity, rather than the more general case of a set of commuting projectors. We will allow a rank-k projector P to "inherit" a labelling as a function of the assignments to some $\left\{\left|\psi_P^{(i)}\right\rangle\right\}$, which form a specially chosen decomposition of P as $P = \sum_i \left|\psi_P^{(i)}\right\rangle\left\langle\psi_P^{(i)}\right|$.

Given a variable assignment to states $f : \mathcal{H}_d \to \{0, C, 1\}$, we can define a variable assignment to all projection operators on $\mathcal{H}_d$, $g : \mathcal{P}(\mathcal{H}_d) \to \{0, C, 1\}$ by the following:

$$g(P) = \begin{cases} 0 & \text{if } \exists \left\{\psi_P^{(i)}\right\},\ P = \sum_i \left|\psi_P^{(i)}\right\rangle\left\langle\psi_P^{(i)}\right|,\ f\left(\left|\psi_P^{(i)}\right\rangle\right) = 0\ \forall i, \\ 1 & \text{elif } \exists \left\{\psi_P^{(i)}\right\},\ P = \sum_i \left|\psi_P^{(i)}\right\rangle\left\langle\psi_P^{(i)}\right|,\ f\left(\left|\psi_P^{(i)}\right\rangle\right) \in \{0,1\}\ \forall i, \\ C & \text{o/w.} \end{cases} \tag{5.23}$$

The motivation for this notion of value inheritance is as follows: in any context containing

P, we can consider replacing it with one of its maximally fine-grained decompositions, which then inherits value assignments from f derived *via* the annulus method. Since the assignment of values to this rank-1 decomposition of P is necessarily consistent with the restrictions of the scenario, we can consider a post-processing that consists of a coarse-graining of those fine-grained results, "forgetting" which one of them occurred. This process must also be consistent. Hence, as long as there exists some decomposition of P into rank-1 projectors, each of which is assigned a noncontextual valuation, then P can inherit a noncontextual valuation. However if there must be a contextual vector in any such decomposition, we must treat P as having a contextual valuation.

The challenge, then, is to characterise which projectors P must be given a contextual valuation under this schema, and then prove a bound analogous to that of Theorem 10.

Theorem 12. *In a scenario in which the available measurements are rank-r projectors* $\{P_i\}$*, peformed on a quantum system of dimension d, then at most* $(I_{1/2}(r, d-r) - I_{r/d}(r, d-r))|\{P_i\}|$ *are given a contextual valuation, where* $I_x(\alpha, \beta)$ *is the* regularised incomplete beta function.

Proof. The proof proceeds similarly to the proof of Theorem 10. We will identify a generalised annulus within the space of rank-r projectors on $\mathcal{H}_d$ which can be associated with contextual valuations, and then use its volume in proportion to that of the overall Hilbert space to form a bound on how many must be given contextual valuations.

Lemma 3. *Under the extension of a valuation of rank-1 projectors given by the annulus method from Theorem 10 centred on* $|\phi\rangle$*, a rank-r projector* P_r *is contextual only if* $r/d \leq \mathrm{Tr}(P_r|\phi\rangle\langle\phi|) \leq 1/2$.

Proof of Lemma. First, consider the case that $\mathrm{Tr}(P_r|\phi\rangle\langle\phi|) > 1/2$. We could equivalently write this as $\langle\phi|P_r|\phi\rangle = \langle\phi|P_rP_r|\phi\rangle > 1/2$. We note that $P_r|\phi\rangle$ is a vector in the eigenvalue-1 subspace of P_r. As such, we can find a basis for P_r that includes $P_r|\phi\rangle$, up to a normalisation constant. By design, each of these other basis elements $|i\rangle$ will have $\langle i|\phi\rangle = \langle i|P_r|\phi\rangle = 0$. Hence, this

corresponds to a decomposition of P_r in which every vector would be given either a 1 or a 0 valuation, and so P_r inherits a 1 valuation.

Next, we consider the case that $\mathrm{Tr}(P_r|\phi\rangle\langle\phi|) = r/d - \epsilon < r/d$. We wish to decompose P_r in such a way that each of the elements $|i\rangle$ of the decomposition have $|\langle i|\phi\rangle| < 1/d$. Again, we consider the vector $P_r|\phi\rangle$ and choose a basis for the 1-eigenspace of P_r so that we have

$$P_r|\phi\rangle = \sum_{i=0}^{r-1} \left(\sqrt{\frac{1}{d} - \frac{\epsilon}{r}}\right) |i\rangle. \tag{5.24}$$

Each of these elements in the decomposition would recieve a noncontextual 0 valuation, so P_r inherits a 0 valuation. This concludes the proof. □

Using the result of the lemma, we can apply a proof method identical to that in Theorem 10 if we can calculate the proportion of the Hilbert space taken up by the set $\{P_r | r/d \leq \mathrm{Tr}(P_r|\phi\rangle\langle\phi|) < 1/2\}$.

As before, entries of a uniformly random unit vector in $\mathbb{C}^d$ have the same distribution as the entries of a column in a Haar-random $d \times d$ unitary matrix. Applying Equation 7 from Życzkowski and Sommers [107], if U is a Haar-random $d \times d$ unitary, then defining $Y = \sqrt{\sum_{k=1}^{r} |U_{k,1}|^2}$, we have

$$P_{d,r}^{Y}(y) = c_{d,r} y^{2r-1} (1-y^2)^{d-r-1}, \tag{5.25}$$

where $P_{d,r}^{Y}(y)$ is the probability density function for the random variable Y, and

$$c_{d,r} = \frac{2}{B\left(r, d-r\right)} = \frac{2\Gamma(d)}{\Gamma(r)\Gamma(d-r)}, \tag{5.26}$$

in which B is the Euler beta function. and Γ is the gamma function. Performing a change of variables, then, in which $T = Y^2$, we get:

$$P_{d,r}^{T}(t) = \left|\frac{\mathrm{d}}{\mathrm{d}t}\left(\sqrt{t}\right)\right| P_{d,r}^{Y}\left(\sqrt{t}\right) \tag{5.27}$$

$$= \frac{t^{r-1}\left(1-t\right)^{d-r-1}}{B(r, d-r)} \tag{5.28}$$

This is exactly a Beta distribution with shape parameters $\alpha = r$ and $\beta = d - r$; we have derived that $T \sim \text{Beta}(r, d-r)$. We note taking $r = 1$, and integrating, we recover our earlier result used in Theorem 10. Since the CDF for a Beta distributed random variable is given by the regularised incomplete beta function, $I_x(\alpha, \beta)$, the proportion of the Hilbert space taken up by the set $\{P_r|^r/_d \leq \text{Tr}(P_r|\phi\rangle\langle\phi|) \leq {}^1/_2\}$ is given by

$$I_{1/2}(r, d-r) - I_{r/d}(r, d-r). \tag{5.29}$$

To complete the proof we once again apply the argument from Fubini's theorem used in the proof of Theorem 10. □

Corollary 4. *For any choice of d, r, the proportion of projectors requiring a contextual valuation is always less than ${}^1/_2$.*

Proof. We have for any d, r, that the proportion of projectors requiring a contextual valuation is bounded above by $I_{1/2}(r, d-r) - I_{r/d}(r, d-r)$. We wish, then, to bound this quantity above by ${}^1/_2$.

Since $I_x(\alpha, \beta)$ is the cumulative distribution function for a random variable with distribution $\text{Beta}(\alpha, \beta)$, a trivial upper bound for this quantity is given by

$$I_{1/2}(r, d-r) - I_{r/d}(r, d-r) < 1 - I_{r/d}(r, d-r). \tag{5.30}$$

In fact for large d, this is a reasonable approximation, since we can apply Chebyshev's inequality to show that $I_{1/2}(r, d-r) \geq 1 - O(d^{-2})$. We wish, then, to lower bound the value of $I_{r/d}(r, d-r)$. The mean-median-mode inequality [57] for the Beta distribution with $\alpha \leq \beta$, which is met here since r must be a nontrivial factor of d, is given by

$$\frac{\alpha - 1}{\alpha + \beta - 2} \leq m(\alpha, \beta) \leq \frac{\alpha}{\alpha + \beta}, \tag{5.31}$$

where $m(\alpha, \beta)$ represents the median of a random variable with a $\text{Beta}(\alpha, \beta)$ distribution. This

is an equality only when $\alpha = \beta$; although this corresponds to a trivial contextuality scenario. Importantly, we have

$$m(r, d-r) \leq \frac{r}{d}. \tag{5.32}$$

Hence by the monotonicity of $I_x(\alpha, \beta)$, we have $I_{r/d}(r, d-r) \geq I_{m(r,d-r)}(r, d-r) = 1/2$, and the result follows.

□

5.2.3 Applications to quantum verification

Maximal contextuality has been harnessed to verify the adherence of delegated quantum computation systems to a specified protocol, for example by the verifier-on-a-leash proposal of Coladangelo, Grilo, Jeffery and Vidick [31]. In the scenario considered, two agents who are, for the sake of argument, spacelike separated, claim to have quantum computational capabilities, and to share entangled qubits with each other. One or both of the two is being tasked with performing a quantum computation for a third (principal) agent, who does not immediately trust that the two devices in question are performing the protocol faithfully. The principal agent only has access to classical computation, and in particular shares no entanglement with either prover. In this section, we will consider a stripped down version of the scenario covered by Coladangelo *et al*, and look only at the simple case of entanglement games; as with their result, we will work with noiseless systems. We have one principal agent, who sends measurement instructions to two separated devices, which claim to have quantum computational ability and control entangled bits. The aim of the principal agent is to find out if the provers are behaving dishonestly while expending a minimal amount of resource.

We will consider first the case in which the two quantum provers are asked to measure observables from the Peres-Mermin magic square, which was considered as an example of a locality violation by Aravind [9]. Each row and column of the Peres-Mermin magic square is a set of three commensurable observables and forms a context. The product of observables in each row (well-defined, since they commute), is the identity, and that of each column is the negative

identity. This means the product of the three results of the observables must be 1 or -1, respectively. Since the product of all the squares entries must be 1, found by multiplying the 1 values for each row, but it must also be -1, found by multiplying the -1 values for each column. This demonstrates that a classical "magic square" cannot exist, and so a classical strategy cannot be always successful.

The promotion of the Peres-Mermin magic square from a proof of maximal BKS contextuality to a proof of maximal nonlocality comes when each party performs the two-qubit measurements on the state $|\phi_+\phi_+\rangle = 1/2\,(|0_A0_B\rangle + |1_A1_B\rangle) \otimes (|0_A0_B\rangle + |1_A1_B\rangle)$, distributed as two EPR pairs rather than separable over the two systems. This state, symmetric over the two parties, has the nice property that for a measurement observable O in the magic square, measured at both sites A and B, we have $O_AO_B|\phi_+\phi_+\rangle = |\phi_+\phi_+\rangle$. Hence, when the same observable is chosen at each site, the measurements must yield the same eigenvalue. Working in the PVM picture we can easily demonstrate a similar result; if the device at site A measures with respect to the basis $\{|\psi_i\rangle\}$, and the device at site B measures with respect to $\{|\psi_i^*\rangle\}$, where the complex conjugate is taken of the coefficients, then if the outcome at A is $|\psi_i\rangle$ then the outcome at B will be $|\psi_i^*\rangle$. The states of course coincide when the coefficients are real, and this can be thought of just as a difference in phase convention between the two sites.

As shown in Figure 5.5, a dishonest pair of provers can do quite well with their classical strategy. It is not hard to demonstrate that the Nash equilibrium for the game is that the prover will ask site A about a uniformly random row, and site B about a uniformly random column. Their deceit is only discovered in the case that the row and column intersect at the one observable marked with a "?", which must have a contextual value. Hence, the principal agent is fooled $8/9$ of the time. Can the agent do better, by picking a different set of observables to be measured? If so, how much better is possible? We ask whether it is possible the result of Theorem 10 in order to demonstrate upper bounds of the performance of one of these agents attempting to catch out classical deceivers.

There are a number of resources that could be considered when trying to perform such a verification. One obvious quantum resource that is often thought to be difficult to generate

$Z \otimes \mathbb{1}$	$\mathbb{1} \otimes Z$	$Z \otimes Z$	$\mathbb{1}$
$\mathbb{1} \otimes X$	$X \otimes \mathbb{1}$	$X \otimes X$	$\mathbb{1}$
$Z \otimes X$	$X \otimes Z$	$Y \otimes Y$	$\mathbb{1}$
$-\mathbb{1}$	$-\mathbb{1}$	$-\mathbb{1}$	

1	-1	-1
-1	-1	1
1	-1	?

Figure 5.5: On the left hand side, the nine observables making up the Peres-Mermin magic square are shown. On the right, a maximally-filled out square is shown. The observable in the bottom right corner must be considered contextual: it takes the value of 1 if measured in a column or -1 in a row.

is the entangled pairs of qubits used by the two provers. If this is the resource we wish to minimise then we would want to minimise the number of rounds of the protocol, since each round will consume more entangled pairs, and also the quantum dimension of the system used, since higher quantum dimensional BKS proofs require more Bell states in order to coordinate their results over the two parties. It might be that it is expensive to communicate between the principal agent and the two provers; in this case we will again want to minimise the number of protocol rounds, but we will also want to have small absolute sizes of the quantum scenario, so that a smaller number of bits suffices to tell each prover which context they must measure with respect to. For example, specifying one of the 2025 contexts of the E_8 quantum set requires a comparatively large 11 bits. We note that one potential restriction from our previous setting is that these strategies need not necessarily invoke every single context that could be measured in the set in question; for the Peres-Mermin magic square, for example, there are additional contexts that correspond to adaptive measurements of the observables. Since we are dealing with spacelike separation of provers, our measurements are nonadaptive, which is what enables us to take the PVM approach in full generality; the entire measurement is specified at once.

However, none of these metrics intersect completely with $q(\mathcal{G})$. If we have a set of projective

measurements with labels in $\{0, 1, c\}$, with a minimal number assigned c, we may find, for example, that one vector labelled c is in a number of contexts, say ten. If in nine of those ten contexts the vector must contextually take the value of 0, and in only one of the ten take the value of 1, then despite being contextually labelled, and being counted as such by $q(\mathcal{G})$, it may not make much of a contribution to the sensitivity of the test in question, since it can be fairly well simulated classically by assigning it a value of 0.

One approach might be to attempt to convert a bound on $q(\mathcal{G})$ into a bound on resource costs as discussed above. The other quantities that are required to be bounded in this approach are the number of c-valued and 1-valued projectors that are orthogonal to an average c-valued projector. This can be seen by considering the manner in which a classical deception is detected. First, a context is chosen to be measured at site A. Since at least one of the two contexts must contain a contextual valued projector for a deception to be discovered, we will assume without loss of generality that the context on subsystem A does and denote that projector $|c\rangle\langle c|$. We can note also that in the limit of high quantum dimension, in the construction in Section 5.2, the vast majority of contexts will contain a contextually-valued projector by construction. Now, the context measured at site B must contain information that is contradictory with the contextual valuation of $|c\rangle\langle c|$ in the first context; that means that $|c\rangle\langle c|$ must have a 1 valuation and the new context must contain either a 1-valued projector or a c-valued projector, orthogonal to $|c\rangle\langle c|$, that has received a contextual 1-value.

We can however find trivial bounds, by bounding only the probability that our first-chosen context contains a contextually-valued projector. We will show now that this tends to 1 as quantum dimension increases if we are only able to apply the general bounds proved above. One naive strategy to assigning these 1-values, would be to, in any context, if the context does not have a 1-valued projector, assign the 1 to the value closest to the top of the cap. This is not guaranteed to be the best strategy for the deceitful provers, but can provide an intuition of what such a bound would look like. Potentially, such an approach could be useful in scenarios in which the claimed provers have restricted computing ability.

A basis can be represented as a unitary matrix, where the columns (or equivalently, rows) of

the matrix are the elements of the basis. For a random basis, defined from a Haar-random unitary matrix, the coefficients of each basis element for any given basis vector themselves form a random unit vector drawn from the Fubini-Study metric. If we choose a basis vector $|0\rangle$, the distribution of the squared overlap of the closest vector from a random basis with $|0\rangle$ is, then, the distribution of the largest element drawn from a Fubini-Study random vector of the same dimension. This distribution was calculated by Lakshminarayan, Tomsovic, Bohigas and Majumdar [62]. Defining $I_k = [1/k+1, 1/k]$, and $F(t, d)$ as the CDF of the largest absolute value of a Fubini-Study random vector in $\mathbb{C}^d$, they show

$$F(t \in I_k, d) = \sum_{m=0}^{k} \binom{d}{m} (-1)^m (1 - mt)^{d-1}. \tag{5.33}$$

However, since we are looking only at the special case of $t = 1/2$, the events are independent and either simply by subtracting from 1 the probability that any of the d entries could be over $1/2$, or by application of Equation 5.33, we find

$$F\left(\frac{1}{2}, d\right) = 1 - \frac{d}{2^{d-1}}. \tag{5.34}$$

This gives us an estimate of the number of contexts whose 1-valued projector is contextual; although this reflects a naïve strategy on the part of the provers, it can help bound the effectiveness of these strategies. We see that this does tend to 1 as the quantum dimension increases.

5.2.4 Notes on the search for small Kochen-Specker sets

Since the inception of the Bell-Kochen-Specker theorem, one research programme of interest has been that of searching for Kochen-Specker sets of minimal size in different dimensions [104]. In 2011, Arends, Ouaknine and Wampler [10] showed that there is no graph-theoretic obstruction to Kochen-Specker sets in $d = 3$ smaller than the best-known example of Conway and Kochen's 31 vector set. They also show that the only candidate graph with 17 vertices is not quantum-

mechanically producible, and therefore an example admissible by quantum mechanics must have 18 or more vectors. In 2016, by considering additional restrictions that a minimal Kochen-Specker graph would have to meet, to reduce their search space, Uijlen and Westerbaan showed that a quantum-mechanically accessible Kochen-Specker set must have at least 22 vectors [101].

In the previous section, we determined bounds on the robustness of contextuality of quantum-mechanically accessible systems. Unfortunately, these bounds are not sufficient to demonstrate sensible lower bounds on the size of Kochen-Specker sets in any dimension. However, given the intuition that the spherical cap techniques for finding noncontextual sets gives us, we might hope to find numerical evidence of quantum-mechanical or GPT constraints on the size of systems demonstrating contextual behaviour.

In this section, we consider what happens when dropping certain geometrical constraints that apply to quantum systems; we will look for arrangements of vectors in d (real) quantum dimensions that cannot be assigned noncontextual valuations by the cap method. However, we will not require that they meet the orthogonality relations that quantum mechanical systems must meet. It is certainly true that by loosening this restriction, we expect the number of vectors required for an apparent contradiction with noncontextual assumptions will be nonincreasing. The aim, then, is to place the contextual annulus in a region of the space that misses all of the selected points. Instead of asking for a set of points, forming a measurement scenario that must be contextual, we ask for the intuitively easier question of how many points it takes to find an arrangement that is always found to be contextual by our cap construction, in the sense that any placement of the cap must intersect at least one point.

From the proof of Theorem 10, we can intuit that a successful strategy for avoiding cap placement will involve points that are in some sense evenly distributed throughout Hilbert space. Any clumping results in more points being able to be captured by our 1-assignment cap, or avoidance entirely by our 0-assignment cap. However, except for specific vector counts, there is no entirely natural definition of what constitutes "even spacing", even in low-dimensional cases such as $d = 3$. We are able to borrow techniques from the field of spherical coding, with additional restrictions on our vector set due to the phase (or in the real case, sign) invariance

of the vectors in the scope of our interest. Two figures of merit were used in order to derive the results in each section, and each of these metrics is in some sense arbitrary.

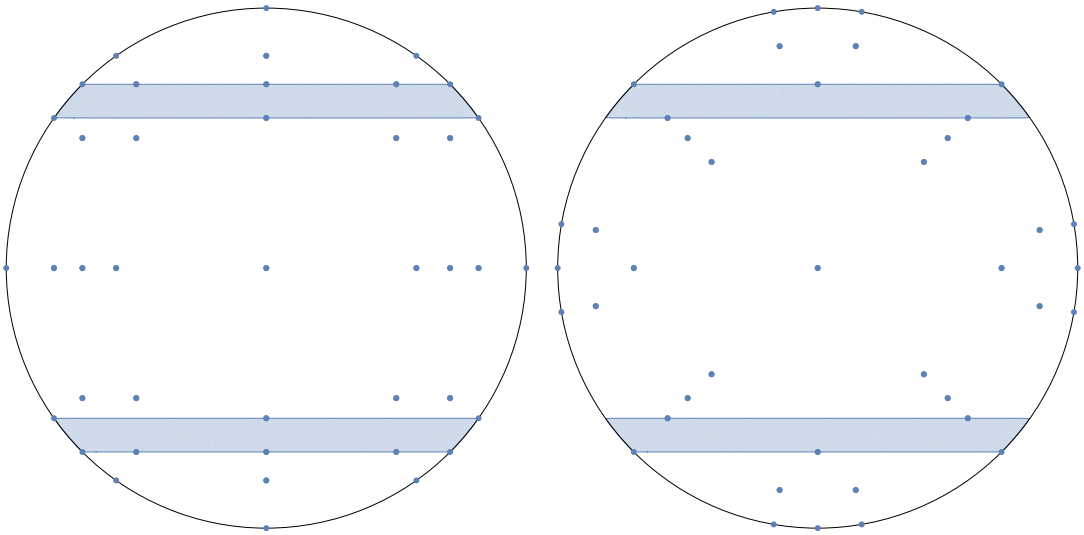


Figure 5.6: Two illustrative cap placements, shown highlighted in light blue, for the 33 ray proof of contextuality due to Peres [77], with one dimension projected out for clarity. We see that multiple vectors, in dark blue, sit exactly on the edges of the annular sections, seemingly restricting their placement.

The first figure of merit applied is that asked by the *Thomson problem* [99]; which was first asked in relation to the "plum-pudding" model of the atom. Thomson's problem asks what the minimal-energy distribution is of charged particles confined to the surface of a sphere. The connections with our normed vector space is obvious, although we require the points to maintain symmetry under inversion through the centre of the sphere. This restriction is unusual, and so detailed numerical results are unknown. Approximations to this problem were found by placing vectors randomly on the sphere, alongside their reflection through the centre, and then simulating their repulsion electrically until a steady state was found. This was repeated a number of times to avoid global minima. Other than initially setting up points to have the necessary symmetry condition, no other changes need to be made, since the symmetry is preserved by the electrical effects.

The second figure of merit is to mimimise the largest overlap between any pair of the vectors, an optimisation problem known as the Tammes problem [98]. To find approximations to optimal

solutions to this problem, electrical-energy minimising solutions were perturbed iteratively to improve the Tammes figure of merit.

However, we find that the "evenly spaced" vectors found using both of these methods do not perform as well as might be expected; there exist positions between the points around which the difference between the two spherical caps can be slotted to create a "noncontextual" model. This suggests that the intuition gained from the above was not completely correct, and the contextuality of quantum systems cannot be explained by the set's exploration of the Hilbert space, or equivalently by the "largeness" of Hilbert space which is sometimes blamed for nonclassicality.

These results strongly suggest that in order for a set of vectors to minimally exhibit contextuality, it must possess some structure that goes above and beyond merely spanning a large region of Hilbert space. Indeed, looking to the literature for examples we see that certain minimal-known vector sets seem almost to be designed *explicitly* rather than implicitly, to capture spherical cap differences like the ones we have considered. An example of this is the proof of the Kochen-Specker theorem due to Peres. Since this is a proof in three quantum dimensions and makes use only of the real subspace, we can once again illustrate the set using a 2-sphere, which is convenient for visualisation purposes, as in Figure 5.6. We see that many of the rays involved in the proof lie exactly on the edges of the pictured cap placements, restricting the possible variable assignments in a structured manner.

5.3 Discussion

Many metrics of nonlocality or of contextuality have a direct operational significance. While our notion of the robustness of contextuality does not have this feature, we have shown that the ideas involved in the work do have potential applications to quantum verification as well as the search for small Kochen-Specker sets. It is perhaps surprising, when we are familiar with quantum subtheories being "maximally contextual" that there is a sense in which quantum mechanics is restricted in how contextual it is compared to other physical theories. However, this directly parallels the case of nonlocality, where we have proofs such as GHZ that demonstrate maximal

nonlocality as well as constructions such as the PR box that are more nonlocal with respect to certain metrics. In the case of quantum verification, certain further work that could be performed includes numerically testing the other examples of maximal contextuality we have analysed here to see if any yield a better chance of discovery than the Peres-Mermin magic square; another bound could also be calculated using the result of Theorem 10 by bounding the probability that the same $c-$valued projector exists in both contexts sent to the provers.

There is a big gap in between the analytical upper bound on our figure of merit for robustness of contextuality and our best concrete example that serves as a lower bound for what is possible within quantum mechanics. One way in which the bound could potentially be improved is by finding a region of the hypersphere, free from orthogonal points, that covers an area greater than that of the spherical cap used in the proof. This relates to an open problem in mathematics known as Witsenhausen's problem, after the author of the paper that formulated the problem and proved the trivial bound of $1/d$ of the area of the sphere in dimension d [106]. Spherical caps cover $1/\sqrt{2}^d \sim 0.707^d$ of the surface of the sphere, whereas the best known upper bound for such an area is $(\sqrt{3}/2)^{\sqrt{2}d} \sim 0.815^d$ [85]. Depending on the shape of a possible refinement to the area used for this purpose, we might also have to vary the shape of the area designed to capture at least one element from each basis, so that this latter area is a strict superset of the former. This is one avenue to improvement of the upper bound, however it is conjectured that the double-cap construction is optimal, even in the case in which our vector space is complex rather than real; this complex extension was first conjectured by Montina [74]. It is the opinion of the author that any major improvement to this bound will have to leverage a different proof technique.

Recent work by Karanjai, Wallman and Bartlett [55] demonstrated that contextuality can be thought of as bounding the efficiency of classical simulations of quantum mechanics, by mandating a larger statespace. The internal memory state of a simulator is analogous to an ontic state, and if we wish to have fewer memory states than quantum states, our effective ontological model must be ψ-ontic. Karanjai *et al* demonstrate that any relatively small set of quantum states within the qubit stabiliser formalism displays a proof of state-independent contextuality

akin to the Peres-Mermin magic square. This result seems to be morally in tension with the result of Theorem 10, in that the take home message seems to be that contextuality is omnipresent within the stabiliser formalism rather than linearly bounded. The main difference in the two approaches is that Karanjai *et al* have state update by Lüders rule as a key part of their formalism, whereas in the approach of this chapter measurements are fixed PVMs and are not adaptive. We have seen in Chapter 4 that state update rules do in some cases provide strong bounds on ψ-epistemic systems, and here we have an example of these bounds being applied *via* the injection of additional contextuality into the system. The literature on contextuality for ontological models has been primarily concerned with contextuality as applied to preparations and measurements, these forming a simple picture of quantum experiments. There is significant further work to be done in exploring and understanding the role of contextuality in quantum transformations, and in quantum state update.

Chapter 6

The Computational Complexity of Contextuality

The computational complexity of tasks involving correlations was first investigated by Pitowsky in 1991 [79], when considered the decision problem of whether a set of correlations of a specific form were valid in the sense of being statistically possible. He found that for a set of measurements of which any two were commensurable and no sets of three were commensurable, determining membership of the correlation polytope was **NP**-complete. Such measurement situations do not occur in quantum mechanics, and this extension of our objects of interest to Generalised Probabilistic Theories outside the scope of quantum mechanics has been fairly universal in the literature for such problems. As such, the computational complexity of many of these problems when restricted to quantum objects is open. He also demonstrated that finding all the facets of the polytope was an even harder problem, and bravely conjectured a structure for the general facet inequalities that, if true, would imply **NP**=**co-NP**. Avis, Imai, Ito and Sasaki determined that for a bipartite nonlocality scenario, with a complete bipartite compatibility graph for measurements, the decision problem was also **NP**-complete. Again, this was proved within a GPT framework rather than strictly within quantum mechanics [12].

One question of obvious foundational relevance that we might ask about a table of data is

whether or not it contains evidence of *signalling* behaviour. Since quantum mechanics and classical mechanics are no-signalling theories, this can be a clean way of demonstrating that a certain set of correlations are outside of both the classical and quantum sets. However, in real experiments it is likely that due to noise there will be a small amount of apparent signalling that will grow weaker with larger sample sizes; this problem was explored by Dzafharov, Kujala and Cervantes [35].

Given a table of probabilities or possibilities, for example in Figure 6.1, there are two questions of particular foundational relevance that we might consider: whether the table provides an example of *signalling*, and if not, whether it still provides an example of nonlocality. Such questions of locality can be used to explore the quantum-classical boundary, as violation of classical locality constraints is an explicit demonstration of quantum-mechanical behaviour. Detecting signalling within distributions can easily be shown to be computationally tractable, leading us to consider the question of locality.

6.1 Possibilistic Locality

In this section, we will focus on a weaker notion of locality, or equivalently a stronger notion of nonlocality: one that is possibilistic in nature. The form will be that of a classical Bell experiment, with two parties at spacelike separated locations performing measurements on a pair of possibly quantum systems that may have interacted at some point in the past. At each measurement site, there exists a range of measurements that may be performed, from potentially different sets. Alice at the first site chooses some measurement $a \in A$, and Bob at the second site chooses some $b \in B$. These measurements may possess a different number of outcomes: Alice's measurements have an outcome set $\{0, 1, 2, \ldots, l-1\}$, and Bob's have $\{0, 1, 2, \ldots, m-1\}$. The most studied case is that in which Alice and Bob have the same number of measurement choices, and the same number of outcomes per measurement. If there are n parties, each with k measurement options possessing l outcomes, this is known as a (n, k, l) scenario [3, 68].

Quantum mechanics can be thought of as a calculational tool that, given a particular context

	$b_1 = 0$	$b_1 = 1$	$b_2 = 0$	$b_2 = 1$
$a_1 = 0$	$1/2$	0	$3/8$	$1/8$
$a_1 = 1$	0	$1/2$	$1/8$	$3/8$
$a_2 = 0$	$3/8$	$1/8$	$1/8$	$3/8$
$a_2 = 1$	$1/8$	$3/8$	$3/8$	$1/8$

	b_1	b'_1	b_2	b'_2
a_1	1	0	1	1
a'_1	0	1	1	1
a_2	1	1	1	1
a'_2	1	1	1	1

Figure 6.1: Tables of Probabilities and Possibilities tables for a CHSH [30] experiment. Alice has two available measurements: measurement A_1 with outcome a_1 and a'_1; and measurement A_2 with outcomes a_2 and a'_2. Bob has measurements denoted similarly. Typically, however, we will not label outcomes as the properties we are consider are invariant under relabelling.

of measurement settings, yields probabilities for the different outcome combinations that can occur. Since we are looking primarily at bipartite nonlocality scenarios, we can adopt the table notation of Mansfield and Fritz [68]. Each "row" and "column", separated by lines, consists of multiple "subrows" and "subcolumns". Each row and column represents a measurement setting adopted by a party, and a subrow or subcolumn represents a specific outcome of that measurement. At the intersection of a subrow and subcolumn is the probability that those two events are observed simultaneously, given the context of the measurement settings containing them. For example, the probabilities for a classic CHSH experiment are shown in in the upper half of Figure 6.1. If Alice were to choose setting A_2, and Bob B_1, then the probability of Alice's outcome being $a_2 = 1$ and Bob's $b_2 = 1$ is given by $1/8$.

A table of probabilities can be converted into a table of possibilities *via* the mapping

$$f(p) = \begin{cases} 1 & p \neq 0 \\ 0 & p = 0, \end{cases} \tag{6.1}$$

which acts as a homomorphism between the normal probability semiring $\mathbb{R}_+$, and the Boolean semiring on $\{0, 1\}$. This has been referred to as a *possibilistic collapse*, for example by Abramsky [4], and by Mansfield and Barbosa [71].

	$b_1 = 0$	$b_1 = 1$	$b_2 = 0$	$b_2 = 1$
$a_1 = 0$	1	1	1	1
$a_1 = 1$	0	0	1	1

Figure 6.2: A simple signalling situation; if Alice sees the outcome $a_1 = 1$, she knows Bob must have chosen measurement setting B_2.

The lower half of Figure 6.1 displays the outcome of such a collapse applied to the CHSH experiment probabilities. As before, when considering ontological models we can represent such a possibility table as a function $T : A \times B \times \{0, \ldots, l-1\} \times \{0, \ldots, m-1\} \to \{0, 1\}$, a mapping from chosen measurement settings alongside a specified outcome event, and returning whether or not that event is possible or impossible. It is natural to consider this object as a conditional possibility $P(L, M|A, B)$, which tells us whether our given quantities (here, measurement settings) are compatible or incompatible with another quantity, such as in this case an observed measurement event.

For the decision problems in this section, we will focus on tables of data that do not offer the possibility of superluminal signalling, however this is just for notational convenience; allowing signalling inputs does not meaningfully change the computational results. For possibilistic tables, the no-signalling property can be summarised as saying that if a subrow contains a 1 entry in some column, then it must contain a 1 entry in every column, and the equivalent statement for subcolumns. This just encapsulates the fact that if an outcome is possible in one context, it must be possible in every context.

It is worth taking a moment to explore what locality looks like within a possibilistic context. Just as in the probabilistic case, a distribution is local if it can be generated by a theory of local hidden variables, ones in which each of the systems has a separate state and response. By a generalisation of Fine's theorem due to Abramsky and Brandenburger [2], such distributions are exactly characterised by being a convex mixture of states that are deterministic and non-signalling. If the distribution in question is outside of this convex hull, it cannot be generated in this way and is *nonlocal*. Convex mixing has a simple form within the Boolean semiring, since the only two choices of coefficients for a function are 0 and 1. Because of this the mixture

	1	0	0	1			1	0	0	1			0	0	0	0
	0	1	0	1	=		0	0	0	0	+		0	1	0	1
	0	0	0	0			0	0	0	0			0	0	0	0
	1	1	0	1			1	0	0	1			0	1	0	1

Figure 6.3: A table of possibilities that does not exhibit nonlocality; it can be decomposed into a sum of deterministic grids that obey no-signalling.

a)

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

b)

1	0
0	
	0

Figure 6.4: Possibility tables for the Hardy paradox. First, the quantum mechanically achievable table of possibilities; second, the structure that directly causes the paradox, blank spaces can be replaced with either a 0 or a 1 entry while maintaining the nonlocality.

of two tables (respresented, as above, by functions T_1, T_2) is given by $T_1 + T_2$. An outcome is possible in the statistical mixture of two distributions if and only if it is possible in at least one of the distributions. Figure 6.3 is a simple example of a local table, taken from Reference [92].

The Hardy paradox, a possibilistic structure given in Figure 6.4 is quantum mechanically realisable. As demonstrated by Hardy [47], there exists a quantum state and projective measurements performable on the two separate subsystems that give rise to this structure of possibilities. It is not an example of maximal contextuality; there exist multiple deterministic grids representing canonical hidden variables that are compatible with the empirical model, but the "1" entry in the upper left corner cannot be part of any such grid. We can see this quickly by attempting to construct such a grid and we can quickly see that it is impossible.

Given the presence of possibilistic nonlocality as a stronger form of nonlocality, we are motivated to investigate the associated decision problem. Interestingly, none of the results from the probabilistic case can be brought over immediately into the possibilistic domain; we are seeking a stronger form of nonlocality, and yet only are giving ourselves possibilistic data to work with. Unlike in Chapter 2, in which Possibilistic notions of noncontextuality were simply weaker forms of contextuality, we have a restricted information here that potentially increases the difficulty of

the decision problem. On the other hand, our quantity of interest is more extreme, which would imply it could not be harder than the probabilistic case, and might indeed be easier. Let us, then, following Reference [92], define the Possibilistic Locality, or PossLoc decision problem.

Definition 23 ((l, m)-PossLoc)**.**
Instance: a possibilistic table of data represented by a function $T : A \times B \times \{1, \ldots, l\} \times \{1, \ldots, m\} \to \{0, 1\}$.

Question: does there exist a set of functions T_i, such that the T_i are no-signalling, deterministic, and $T = \sum_i T_i$?

One thing to note is that the number of outcomes per measurement available to each party is part of the definition of the decision problem; we consider a two-party nonlocality scenario in which Alice's measurements have m outcomes, and Bob's have l. As we will see, varying this parameter will make a critical difference to the complexity of the problem. When these parameters are fixed, the scaling parameter for the class of decision problems is given by the number of measurements available to each party. The instance of the problem is a given table of possibilities, and we are asked to decide whether or not it displays nonlocality. In the case $m = l$, we will denote the decision problem by m-PossLoc.

Such decision problems have been studied before in the literature, for example by Mansfield and Fritz [68, 70], as well as by Abramsky, Gottlob and Kolaitis [3]. Mansfield and Fritz studied the case of 2-PossLoc, and showed by an analysis of the structure of such instances, that the decision problem was in **P**.

Theorem 13. *[70] 2-*PossLoc *is in* ***P****.*

Proof. This proof proceeds by showing that the only kind of possibilistic nonlocality that can occur in the two-outcome case is the one originally identified by Hardy [47], as depicted in Figure 6.4 (b); a table is nonlocal if and only if it contains an example of the Hardy paradox.

Since the appearance of this structure is equivalent to possibilistic nonlocality in this case, an algorithm to determine whether or not a given data table is possibilistically local or nonlocal

reduces to checking each possible set of four contexts for the structure. There are $n^2(n-1)^2/4$ contexts, so since checking for the appearance of the Hardy paradox is possible in constant time, our algorithm runs in $O(n^4)$. □

However, this argument is specific to the case in which we have measurements with at most two outcomes on each side. If we extend one of the parties measurements to have three outcomes, then we can create possibilistic nonlocality scenarios that are not a fine graining of Hardy's paradox. As an illustration that such structures exist, and are even quantumly-accessible, a novel possibilistic nonlocality scenario in the (2,3) case will be given below.

Abramsky, Gottlob and Kolaitis [3] considered the 3-PossLoc decision problem, as well as the 2-PossLoc decision problem with three parties rather than two, and demonstrated that **NP**-hard problems could be embedded in possibilistic structures for each of them. Since membership within **NP** is not hard to prove for these problems, this demonstrated that m-PossLoc is **NP**-complete when m is 3 or greater, and that 2-PossLoc variants were also **NP**-complete for three or more parties.

Theorem 14. *[3] 3-*PossLoc *is **NP**-complete.*

Proof. The proof proceeds by embedding a robust graph colouring problem into the grid. This technique of defining and using robust constraint satisfaction problems to prove Possibilistic Locality results was instrumental in proving the below results. □

The combination of these two results left only the (2,k,{2,3}) scenario with unknown complexity classification. As we shall see, the (2,3)-PossLoc case is also **NP**-complete, meaning that only 2-PossLoc is a tractable problem as we scale the number of measurements available to each party.

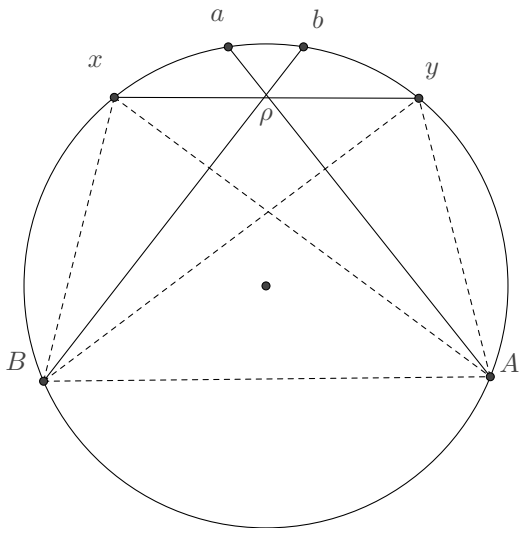


Figure 6.5: Points are shown on the Bloch sphere representing states to which Alice's subsystem can be steered; the possibilistic structure of the experiment demonstrates an example of Possibilistic nonlocality that cannot be reduced to Hardy's paradox. Solid lines denote pairs of states that can be jointly steered to, and dotted triangles between triplets highlight the existence of a valid POVM.

6.1.1 The (2,k,{2,3}) scenario

For this construction, we will confine our steering scenario to the real subspace of a qubit, known sometimes as a *rebit*. This state space corresponds to an equatorial slice through the Bloch sphere, and allows us to consider our vectors as existing in $\mathbb{R}^2$ instead of $\mathbb{C}^2$. This mapping is a case of the Hopf fibration.

As in Chapter 3 we can more easily visualise the structure of the scenario by picturing a steering diagram. The two parties share some entangled state, which can be any that is not maximally entangled[6]. Bob can perform one of a number of two-outcome measurements on his side of the system; any one that he chooses corresponds to steering Alice to one of a pair of states that have her reduced state lying along the line that joins them. With a pure entangled state as the resource for the steering scenario, the accessible states constitute the whole state space; Alice can steer Bob to any pair of points as long as the line between them contains Bob's reduced state. Bob's measurements are similarly unconstrained, and can create a POVM with elements

	b	$b^\perp$	$A^\perp$	$B^\perp$	$x^\perp$	$A^\perp$	$B^\perp$	$y^\perp$
a	1	1	1	1	1	1	1	1
A	1	1	0	1	1	0	1	1
b	1	0	1	1	1	1	1	1
B	1	1	1	0	1	1	0	1
x	1	1	1	1	0	1	1	1
y	1	1	1	1	1	1	1	0

	1	0	0
	0		
		0	0
		0	
			0

Figure 6.6: On the left hand side, a generalised Hardy paradox available for a two qubit measurement, with a two-outcome projective measurement at one site and a three-outcome POVM measurement at the other. On the right hand side, the minimal possibilistic structure.

proportional to any set of points on the sphere, so long as the sphere's centre is strictly contained within their convex hull, or in the case of a pair of points, along the line joining them, so they must be antipodal. This can be seen to be a necessary condition, since the local state must have a decomposition in terms of the states to which it can be steered. For a pure state resource, it can also easily be seen to be a sufficient condition, even when the steering measurement is a POVM rather than a PVM, although this is not necessary for this demonstration. The antipodal points that make up valid contexts for Bob to measure are not shown in Figure 6.5. We can see that an impossible event will occur whenever Bob can be steered to a state $|\psi\rangle$, and has access to a measurement containing a measurement effect proportional to $|\psi^\perp\rangle\langle\psi^\perp|$. If we construct the possibility table for the steering scenario described by Figure 6.5, we find the structure of possibilities shown in Figure 6.6. The measurement rows denote pairs of states to which Alice can steer Bob; the measurement columns represent measurements that Bob is able to perform. We also present this paradox minimally, and it can be seen that it is neither a Hardy paradox, nor a fine-graining of a Hardy paradox. Therefore, the following statement holds:

Proposition 1. *The Hardy paradox is not universal even for quantumly-accessible possibilistic nonlocality as long as one party has access to three-outcome measurements.*

As we will explore later, this possibilistic structure can be thought of as a logical generalisation of Hardy's paradox. The logical structure of the two rightmost measurement columns acts very similarly to the rightmost column of the traditional Hardy paradox, with the third row enforcing

this by mandating that it cannot be the case that both the rightmost measurement rows have 1 entries in their rightmost columns. Therefore, one of them must fulfil the role of the right hand column of the Hardy paradox. This is a truly quantum mechanically viable possibility table, and does not fall victim to the criticisms of assuming the existence of deterministic assignments to proper POVMs that were mentioned in Chapter 2. This is because we can dilate each of these POVM elements in a way that is noncontextual; in fact each POVM element appears in only a single context and so this statement is somewhat trivial. To clarify the notation used in the tables such as Figure 6.6, the column $A^{\perp}$ represents a POVM element proportional to $\left|A^{\perp}\right\rangle\left\langle A^{\perp}\right|$, and the coefficient of these projectors varies between the two contexts; they are genuinely different POVMs, are dilated differently to PVMs which can be considered deterministic, and in any case our structure does not in any way imply that they must be given the same deterministic assignment.

6.2 (2,3)-PossLoc is NP-complete

As with the case of 2- and 3-PossLoc, $(2,3)$-PossLoc is in **NP**. There are at $O(n^2)$ 1 entries in the instance, and for each of these, a deterministic grid (or deterministic hidden variable) must be specified, and each of these is also $O(n^2)$ sized. Checking that the sum of these deterministic grids is equal to the original instance is also achievable in polynomial time, as we have at most $O(n^4)$ operations; $O(n^2)$ semiring additions per table, and $O(n^2)$ tables. The rest of this section will be given to proving that $(2,3)$-PossLoc is **NP**-hard. To do this we will first introduce the concept of a robust satisfiability problem, following the logic of Abramsky, Gottlob and Kolaitis [3]. Then, we will use this as an intermediary step to effectively embed arbitrary 3-SAT instances into $(2,3)$-PossLoc instances, demonstrating that $(2,3)$-PossLoc is at least as hard as 3-SAT, and therefore is **NP**-hard. This proof is reproduced from *On the Computational Complexity of detecting Possibilistic Locality* [92].

Definition 24 (r-Robust satisfiability problems)**.** For a satisfiability decision problem S, the problem r-Robust S to be the decision problem that, given an instance of S, asks whether or

not every assignment of r variables in the problem that do not directly cause one of the clauses to be false, can be extended to a satisfying assignment of the problem.

Theorem 15. *The* (2,3)-PossLoc *decision problem is* ***NP****-complete.*

Proof. We will now introduce the notion of 0- and 1-valid 3-SAT. This restriction is not strictly necessary for this proof, but we will see later that it has a bearing on the computational hardness of deciding locality for quantum-mechanically accessible possibility tables.

Definition 25. 0-valid 3-SAT is the set of 3-SAT decision problems that are satisfied by assigning a value of 0 to all variables.

We will define 1-valid 3-SAT analogously. Now, we will demonstrate that 2-Robust 0-valid 1-valid 3-SAT, the set of 2-Robust 3-SAT instances which are both 0-valid and 1-valid, is **NP**-complete. We will then demonstrate how such a problem can be embedded into $(2,3)$-PossLoc.

Lemma 4. 2-Robust 0-valid 1-valid 3-SAT *is* ***NP****-complete.*

Proof of Lemma. Consider a general 3-SAT instance $C = \bigwedge_i c_i$, where c_i are clauses consisting of the disjunction of three literals. We will introduce two new variables x and y, and apply the mapping f to each clause c_i defined by:

$$f(l_1 \vee l_2 \vee \circ l_3) = (l_1 \vee l_2 \vee \circ l_3 \vee \neg x), \tag{6.2}$$

$$f(\neg l_1 \vee \neg l_2 \vee \circ l_3) = (\neg l_1 \vee \neg l_2 \vee \circ l_3 \vee y), \tag{6.3}$$

in which the $\circ$ symbol is being used to denote the presence or absence of a $\neg$ symbol, and should be taken as constant across instances of the same variable within a specific equation. By the pigeonhole principle, each clause must contain either at least two positive literals, or at least two negated literals, and so f is total. We note that now, every clause of $C' = \bigwedge_i f(c_i)$ contains at least one positive literal and at least one negative literal, and so C' is both 0-valid and 1-valid.

We note that it is possible now to convert this 4-SAT instance back to a 3-SAT instance by the following procedure, mapping each clause to a pair of equisatisfiable clauses (which we will refer to as an effective clause):

$$(l_1 \vee l_2 \vee \circ l_3 \vee \neg x) \mapsto (l_1 \vee \circ l_3 \vee \neg z) \wedge (z \vee l_2 \vee \neg x) \tag{6.4}$$

$$(\neg l_1 \vee \neg l_2 \vee \circ l_3 \vee y) \mapsto (\neg l_1 \vee \circ l_3 \vee z) \wedge (\neg z \vee \neg l_2 \vee y) \tag{6.5}$$

We note also that this adjustment maintains 0-validity and 1-validity. Since each effective clause contains either $\neg x$ or y, setting x to 0 and y to 1 leads to a satisfying assignment for any assignment choices of the other variables.

We can now verify that this transformed instance is 2-RobustTly satisfiable if and only if our original 3-SAT instance was satisfiable. We will consider the possibilities on a case-by-case basis, depending on whether or not the variables x or y have been assigned valuations of 1 and 0 respectively:

1. If neither x is fixed to 1 nor y is fixed to 0, then we set x to 0 and y to 1. As mentioned above, this yields a satisfying assignment. If one or more z has been assigned a value, we may need to assign some l_i to make the affected clauses valid. Since this affects at most two clauses it can be seen that it is always possible to satisfy such a modified pair of clauses.

2. If exactly one of x is fixed to 1 or y is fixed to 0, then, removing all instances of the $\neg x$ or y literals respectively still leaves the instance either 0-valid or 1-valid, and so the instance is satisfiable by definition.

3. If we fix both x to 1 and y to 0, then removing the $\neg x$ and y literals from C' transform it back into C, and therefore this restriction has a satisfying assignment if and only if C did originally.

Therefore, since 3-SAT is **NP**-hard, so is 2-Robust 0-valid 1-valid 3-SAT. It is in **NP**

since there are only $O(n^2)$ different pairs of variables to set, and so the witness to demonstrate a set of assignments that display robustness is still only polynomially sized. Hence, 2-Robust 0-valid 1-valid 3-SAT is **NP**-complete.

□

We note that a more general results in particular imply this, such as that of Ham [46].

Corollary 5. 2-Robust 3-SAT *is* ***NP****-complete.*

We note now that we will only need the fact that 2-Robust 3-SAT is **NP**-complete for the rest of this proof; the motivation for having proven that the 0-valid 1-valid is also **NP**-complete is helpful when determining which of these possibility tables are realisable within quantum mechanics. For a more complete exposition of this problem, see Reference [92], from which most of the material of this chapter is drawn.

Lemma 5. *There is a polynomial-time embedding of* 2-Robust 3-SAT *into* (2,3)-PossLoc.

Proof of Lemma. The reduction algorithm is as follows:

- For each variable in the 2-Robust 3-SAT instance, we add a measurement to the party with two-outcome measurements available to them, to which we will assign the measurement rows. We pick any ordering of the variables to do this.
- For each clause in the 2-Robust 3-SAT instance, we add a measurement to the party with three-outcome measurements available to them; these are our measurement columns. This is a departure from the strategy of Abramsky, Gottlob, and Kolaitis, whose constructions have a direct symmetry between the rows and columns.
- For each intersection of a variable row and clause column, such that the variable is represented positively in the clause, we have the measurement possibilities given by:

$$\left[\begin{array}{ccc} 0 & 1 & 1 \\ 1 & 1 & 1 \end{array}\right], \left[\begin{array}{ccc} 1 & 0 & 1 \\ 1 & 1 & 1 \end{array}\right], \text{or} \left[\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 1 & 1 \end{array}\right],$$

depending on whether the variable is the first, second, or third variable in the clause with respect to our variable ordering.

- For each intersection of a variable row and clause column, such that the variable is represented negatively in the clause, we have the measurement possibilities given by:

$$\begin{array}{|ccc|}\hline 1 & 1 & 1 \\ 0 & 1 & 1 \\ \hline\end{array}, \begin{array}{|ccc|}\hline 1 & 1 & 1 \\ 1 & 0 & 1 \\ \hline\end{array}, \text{or} \begin{array}{|ccc|}\hline 1 & 1 & 1 \\ 1 & 1 & 0 \\ \hline\end{array},$$

depending on whether the variable is the first, second, or third variable in the clause with respect to our variable ordering.

We can see that the choice of subrow in a measurement row correspond to an assignment of the associated variable, and that the clause structure effectively bans the assignment disallowed by that specific clause, for example the clause $(x_1 \vee x_2 \vee x_3)$ corresponds to the possibility table

$x_1 = 0$	0	1	1
$x_1 = 1$	1	1	1
$x_2 = 0$	1	0	1
$x_2 = 1$	1	1	1
$x_3 = 0$	1	1	0
$x_3 = 1$	1	1	1

We see we cannot simultaneously assign a 1 to each of the top subrows as part of a deterministic grid. It now only remains to be shown that choosing a specific "1" instance, which needs to be extended to a deterministic grid, is the equivalent of fixing at most two variable assignments in the embedded 3-SAT instance. We note that choosing any "1" entry fixes the valuation of the variable associated with its measurement row; the "1" is in some measurement subrow and corresponds to a value assignment for that variable. Additionally, each measurement subcolumn has exactly one "0" entry by construction: if this "0" instance is not in the same measurement row as the chosen "1" entry, this also fixes a variable assignment for the measurement row

containing the "0" entry. In the example directly above, if the bottom-left "1" is chosen, then we see from the table that this is equivalent to asking that $x_3 = 1$, $x_1 = 1$. Therefore this construction constitutes a polynomial-time reduction from 2-Robust 3-SAT to $(2,3)$-PossLoc.

□

Therefore, since 2-Robust 3-SAT is **NP**-hard; any instance of 2-Robust 3-SAT can be reduced to an instance of (2,3)-PossLoc in polynomial time; and (2,3)-PossLoc is in **NP**, (2,3)-PossLoc is **NP**-complete. □

6.3 Computing the Robustness of Possibilistic Locality

We saw that for a contextuality situation, a measure of the strength of a maximal contextuality scenario was the number of projectors whose values needed to depend on context. For a nonlocality scenario, which is of course inherently state-dependent, we have a choices in how we interpret this condition. We could count the minimum number of maximally-fine grained projectors whose outcomes depend on context, so if Alice measures in the basis $\{|a_0\rangle, |a_1\rangle \dots |a_n\rangle\}$, and Bob measures in the basis $\{|b_0\rangle, |b_1\rangle \dots |b_n\rangle\}$, we are counting projectors of the form $|a_i\rangle|b_j\rangle$ which have context-dependent outcomes. However, in many nonlocality scenarios, each of these fine-grained projectors may only occur in a single context, and so while this is the most obvious analogue to the maximal contextuality scenario described above, this metric does not capture the structure of the nonlocality problem.

Perhaps a more fruitful approach is to consider how many of the projective measurements available to a single party must be found to have outcomes that depend on the context of which measurement choice is chosen by the other partner. A different generalisation is that we could count the minimum number of contexts that can deviate in their outcomes from a specified set of global valuations. Both of these are valid measures of possibilistic nonlocality, but have very differing properties with respect to how much nonlocality can be considered achievable by

quantum mechanics, as well as the computational complexity of calculating each value. In this section, we will consider one such generalisation, restricting ourselves for clarity to the case in which there are two nonlocal sites and each has access to only two-outcome measurements.

Definition 26 (k-PL)**.** We define the decision problem k-PL as follows; the input is a table of data as in [68] and an integer k, and we ask the following question: is there an explanation of the table in which the number of measurements available to the two parties whose outcomes depend on context is less than or equal to k.

Theorem 16. 2-PL *is* ***NP****-complete when each party has access to measurements with two outcomes.*

Proof. The proof will be by reduction from Vertex Cover, whose definition, as given by Garey and Johnson [43], is as follows:

Definition 27 (Vertex Cover)**.**
Instance: a graph $\mathcal{G} = (V, E)$ and a positive integer $k \leq V$.
Question: is there a $V' \subset V$ such that $|V'| \leq k$ and for all $e \in E$, $V \cap e \neq \emptyset$?

We will now construct a nonlocality scenario from a given instance of Vertex Cover such that there exists a vertex cover for $\mathcal{G}$ of size k if and only if there exists an explanation of the nonlocality scenario in which the number of projectors with context-independent valuation is k.

Each of the four contexts that makes up the Hardy paradox will have criteria that must be met by a set of contextual rows and columns to be able to explain the paradox. For the context containing the 1 entry that cannot be extended to a deterministic grid, the sufficient conditions are always satisfied by one of the following: "the measurement row must be contextual", "the measurement column must be contextual", or "the measurement row and column must be contextual". The remaining contexts can always be satisfied by *either* one of the following:"the measurement row must be contextual", or "the measurement column must be contextual". We never require both the measurement column and row to vary, as this would be a violation of no-signalling.

Given a graph $\mathcal{G} = (V, E)$, we construct a nonlocality scenario in which there are $2|E|$ measurement rows and $|V|$ measurement columns, noting that since there are only $O(|V|^2)$ possible edges, the number of measurements required for the instance is polynomial in $|V|$. Let us label the edges $e_i = \{v_1^i, v_2^i\}$, and for each edge, we place a Hardy paradox in the contexts at the intersections of the measurement rows labelled $2i$ and $2i + 1$, and the measurement columns v_1^i, and v_2^i. Every other context should contain all 1-entries. The Hardy paradoxes we have placed may not be the only Hardy paradoxes in the scenario, due to "interactions" between the paradoxes that share measurement rows.

By the result of Mansfield and Fritz [68], an appearance of the Hardy paradox is the only way in which possibilistic nonlocality can manifest itself for the two-outcome scenario. Consider, then, an unintentionally appearing Hardy paradox, appearing at the intersection of rows r_1, r_2 and columns v_1, v_2. By the form of the Hardy paradox, there must be one row, say r_1, whose intersections with both c_1 and c_2 contain a nonzero entry. This can only occur if the vertices associated with v_1 and v_2 are joined by an edge, and so these paradoxes are resolved by a either of v_1 or v_2 being allowed to contextually vary. Because of this, we do not need to worry about unintentionally introduced paradoxes as they are resolved by the same contextuality assignments as the deliberately seeded paradoxes.

Assume there is a vertex cover for $\mathcal{G}$ of size k, namely $K = \{v_1^K, v_2^K \dots v_k^K\}$, and then consider an explanation for the nonlocality scenario in which measurement columns indexed by K are allowed to have outcomes that vary contextually. We will now show that this is a sufficient explanation for the nonlocality scenario.

We note that for such a scenario, a k minimal for the graph is also a k minimal for the nonlocality scenario: allowing any measurement row to vary contextually affects exactly one of the Hardy paradoxes introduced deliberately by the graph structure, while allowing a measurement column to vary contextually affects affects strictly more than one. Hence, if in a given explanation of the nonlocality scenario, there is a measurement row with a context-dependent outcome, we can replace it with one of the columns involved in the Hardy paradox that involves contexts on that row. By iterating this, we are left with an explanation in which only measurement columns, not

rows, are required to vary contextually, which, by construction, is exactly a vertex cover for $\mathcal{G}$. Hence, the size of a minimal vertex cover for $\mathcal{G}$ and the minimal explanation for the nonlocality scenario are equal.

We have exhibited a polynomial-time reduction from Vertex Cover to k-PL. Since Vertex Cover is **NP**-hard, so too, then, is k-PL. In addition, k-PL is in **NP** since there are only polynomially many 1 entries in the grid and therefore a witness that is an explicit set of data tables that combine to the target table is of polynomial size. □

By this reduction from Vertex Cover, we find that unless **P**=**NP**, k-PL is not approximable in polynomial time to a factor of better than 1.3606 [34], or better than 2 if the Unique Games conjecture holds [59].

The cases in which one party has a number of measurements greater than two can be seen to be **NP**-hard since they are already **NP**-hard for $k = 0$ [2, 92]. This demonstrates another sense in which the case treated by Mansfield and Fritz [68, 70] is unique.

6.4 Discussion

The classification of different variants on the possibilistic locality problem is now largely complete; the hole left in the classification started by Mansfield and Fritz [68], and by Abramsky, Gottlob and Kolaitis [3] has been filled, demonstrating that the $(2, 3)$-PossLoc problem is **NP**-complete. In addition, we have demonstrated an additional way in that the two-outcome version studied by Mansfield and Fritz is unique in its tractability, by extending the decision problem in a way inspired by the approach of Chapter 5 and demonstrating that this too is **NP**-hard.

The main open questions unfortunately capture much of the interest of the problem; what is the complexity classification when applied only to quantum-mechanically accessible distributions? As we have seen in earlier chapters, quantum scenarios carry with them strict geometrical rules that tightly constrain the algebra of events. In *On the computational complexity of detecting*

possibilistic locality [92], the author demonstrates an embedding of a class of 3-SAT instances, of open computational complexity. However, the author conjectures this class to be computationally tractable. This embedding uses POVMs on a two-dimensional quantum system; as discussed earlier there are some conceptual issues with such an approach, and even if the full set of 3-SAT instances were embedded it would be questionable whether this constituted a fully quantum embedding. This embedding makes use of the fact that under possibilistic collapse, POVM elements that differ by a constant factor are identified[1]; even making use of this additional slack in how we can build instances, the problem of embedding 3-SAT instances is not easy. This difficulty with embeddings suggests the following conjecture:

Conjecture 1. PossLoc *restricted to quantum mechanically available instances over two or more sites is in* ***P****; it is computationally tractable no matter the number of outcomes available for measurements at each site.*

[1]However, as seen in Chapter 3, identifying POVM elements that differ by a constant factor can be possible under even probabilistic notions of preparation noncontextuality.

Chapter 7

Conclusion

In this thesis, we have aimed to gain a deeper understanding of notions of contextuality by promoting understanding of differing frameworks for understanding contextuality; by introducing our own framework for equivalence notions of contextuality; by demonstrating that results that were demonstrated using stronger notions also hold when making weaker assumptions; and by investigating the structure and computational complexity of different metrics of nonlocality and contextuality.

We reviewed two major approaches to contextuality in Chapter 2, *Flavours of Contextuality*: the Spekkens, or Equivalence approach; and the Sheaf-theoretic approach of Abramsky and Brandenburger. Original contributions in this chapter included the introduction of Possibilistic contextuality to the Spekkens formalism, as first explored in *Contextuality under weak assumptions* by Simmons, Wallman, Pashayan, Bartlett and Rudolph. Additionally, philosophical justification was given for the possibilistic notions of contextuality by a Leibnizian principle.

In Chapter 3, *Common ground and differences between approaches*, we demonstrated novel equivalences between the sheaf formalism and the equivalence formalism. This included a method of obtaining a notion of preparation noncontextuality within the sheaf approach, by considering mappings from empirical models to global sections of the sheaf $\mathcal{D}_R\mathcal{E}(X)$, and asking that convexity at the operational level be respected at the ontological level, as a direct parallel

to the assumptions of the ontological models formalism and the representation of preparation noncontextuality in that framework. We also draw comparisons between a well-known class of ontological models which capture much of the physical content of the sheaf framework. We also highlight why steering cannot be used to bring a notion of preparation noncontextuality into the Sheaf theoretic approach, and use this example to highlight how strong an assumption noncontextuality for preparations is when compared to noncontextuality for measurements. This intuitively makes sense, given that quantum mechanics cannot support preparation noncontextuality, but measurement noncontextuality without the assumption of determinism can be applied to any operational theory.

Nonlocality is perhaps one of the most misunderstood concepts in quantum foundations, even by physicists. In Chapter 4, *Restrictions on a noncontextual reality* we looked at issues of defining ontological notions of locality, demonstrating that within the ontological models framework, if any ontological elements are allowed to be nonlocal then any operational probabilities can be matched by the theory. Only when the theory has local kinematics, dynamics, state update and measurement responses are the operational probabilities constrained to the local polytope. We also demonstrated a novel proof of the Kochen-Specker theorem that demonstrates how overconstrained a potential solution would be. In this Chapter, we also reviewed restrictions on ψ-epistemic ontological models, demonstrating a proof simpler than current constructive proofs that measurement noncontextuality for POVM measurements implies that an ontological model is ψ-ontic, and demonstrated that possibilistic notions of contextuality provide strong dimension-dependent constraints on which quantum states can have ontological overlap. We also explored restrictions relating to counterfactual independence for deterministic models, in which knowledge of a measurement outcome yields no information about counterfactual statements when conditioning on the preparation of a system.

In Chapter 5, *Metrics of Contextuality and Nonlocality*, we introduced a new metric for determining how contextual a quantum scenario is, that differentiates between scenarios with contextual fraction 1. We proved an analytic upper bound on what is achievable for this figure of merit within quantum mechanics, as well as providing a number of example-based lower

bounds. We also provided an example of an event structure which achieved a maximal limiting value of the figure of merit as the number of projectors increased. We also discussed implications of the result to the search for small Kochen-Specker sets and the problem of quantum verification of untrusted computational devices.

We demonstrated two original results in Chapter 6, *The computational complexity of contextuality*. The first is the classification of the (2,3)-PossLoc decision problem, which filled in the only gap in the classification of the possibilistic locality decision problem that was started by Mansfield and Fritz [68] and continued by Abramsky, Gottlob and Kolaitis [3]. In addition, we demonstrated that a generalisation of the problem inspired by our figure of merit for Chapter 5 was also **NP**-complete, which together with the result of Abramsky, Gottlob and Kolaitis forms a complete classification of the complexity of that generalised decision problem.

7.1 Future Work

The connections between the sheaf-theoretic and equivalence framework seem to be mostly untangled, although one assumes that debates about the status of determinism within these approaches will continue, as will the use of the Identity of Indiscernibles rather than operational principles such as no-signalling to identify ontological elements in the purported physical theory. There are however a number of open questions about the existence of certain ψ-epistemic models in the ontological models formalism. One that is of particular interest concerns the existence of ψ-epistemic ontological models for a single qubit that support POVMs in their operational theory. It was noted by Morris [75] that it is impossible to define representations of POVMs for the Kochen-Specker model for the qubit, even contextual ones[1]. The proof leverages the specific state space and preparation distributions for the Kochen-Specker model. It would be desirable to have a proof of this fact that only made use of the fact that the Kochen-Specker qubit is maximally ψ-epistemic: all the operational indistinguishability is caused by ontological indistinguishability. An important open question then is how ψ-epistemic a model can be while

[1]We note that the fact that even contextual realisations of POVMs are impossible to define for the Kochen-Specker model demonstrates that there is more to this problem than proving analogues of Gleason's theorem.

still supporting POVM measurements. This is linked to a conjecture of the author made in Reference [90], which is that the Kochen-Specker qubit is essentially the only ontological model for a qubit with this property. It is noteworthy that there are no natural ψ-epistemic models for quantum subtheories that are not effectively the Kochen Specker qubit model: the eight-state model is a coarse-graining of the Kochen-Specker model into orthants, and the Spekkens Toy Model is the same, but post selected so that the state lies necessarily on one of four of those orthants.

As mentioned in the discussion section of Chapter 5, there is a significant gap between the example-based lower bound and the analytic upper bound of which values of our "robustness of contextuality" figure of merit are achievable in quantum mechanics. Refining the upper bound will likely take a different proof method and new insight, but it seems unlikely that the current bound is tight, and so refinements are likely to be possible. On the other side, there are a number of good candidates for searching for good example-based lower bounds, such as qubit stabiliser quantum mechanics with more qubits, and the $E7$ sporadic group. Another open question is the existence of an event structure, realising in the limit of large numbers of projectors the maximal figure of merit of 1, that is not triangle-free. Such an event structure would have an easier interpretation as a GPT and finding one would yield more insight into the graph theoretic properties of such examples. In addition, as mentioned in the discussion section of Chapter 5, there are multiple open problems relating to applying techniques used in the proof of Theorem 10 to the problem of verifying untrusted quantum computational devices, including calculating bounds on detection efficiently for different quantum scenarios by bounding the probability that the same contextually-valued projector lies in the contexts measured by both provers; in extending the analysis to the case in which there may be three or more alleged quantum computational devices; and in generalising the figure of merit q from that chapter to better reflect operational needs when performing such informational tasks and finding bounds for these generalised figures of merit.

Bibliography

[1] S. Aaronson and D. Gottesman. Improved simulation of stablizer circuits. *Phys. Rev. A*, 70:052328, 2004.

[2] S. Abramsky and A. Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13:113036, 2011.

[3] S. Abramsky, G. Gottlob, and P. Kolaitis. Robust constraint satisfaction and local hidden variables in quantum mechanics. In *Artificial Intelligence (IJCAI '13), 2013, 23rd International Joint Conference on*, pages 440–446. AAAI Press, 2013.

[4] S. Abramsky. Relational hidden variables and non-locality. *Studia Logica*, 101(2):411–452, 2013.

[5] S. Abramsky, R. S. Barbosa, and S. Mansfield. Contextual fraction as a measure of contextuality. *Phys. Rev. Lett.* 119:050504, 2017.

[6] S. Abramsky, C. M. Constantin, and S. Ying Hardy is (almost) everywhere: nonlocality without inequalities for almost all entangled multipartite states *Information and Computation* vol. 250, pages 3–14, 2016.

[7] N. Alon, S. Ben-Shimon, and M. Krivelevich. A note on regular Ramsey graphs. *Journal of Graph Theory*, 64(3):244–249, 2010.

[8] J. Anders and D. E. Browne. Computational power of correlations. *Phys. Rev. Lett.*, 102:050502, 2009.

[9] P. K. Aravind. Bell's theorem without inequalities and only two distant observers. *Foundations of Physics Letters*, 14(4):397–405, 2002.

[10] F. Arends, J. Ouaknine, and C. W. Wampler. On searching for small kochen-specker vector systems. In P. Kolman and J. Kratochvíl, editors, *Graph-Theoretic Concepts in Computer Science*, pages 23–34. Springer Berlin Heidelberg, 2011.

[11] R. J. Aumann Agreeing to Disagreee *The Annals of Statistics* 4(6): 1236–1239, 1976.

[12] D. Avis, H. Imai, T. Ito, and Y. Sasaki. Two-party Bell inequalities derived from combinatorics via triangular elimination. *Journal of Physics A: Mathematical and General*, 38(50):10971–10987, 2005.

[13] J. Barrett. The Kitchen Sink model. Personal Communication, 2010.

[14] J. Barrett and A. Kent. Noncontextuality, finite precision measurement and the Kochen-Specker theorem. *Stud. Hist. Phil. Sci.*, 35:151–176, 2004.

[15] J. S. Bell. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38:447–452, 1966.

[16] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press, 2002.

[17] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1(3):195–200, 1964.

[18] E. G. Beltrametti and S. Bugajski. A classical extension of quantum mechanics. *J. Phys. A: Math Gen*, 28:3329–3343, 1995.

[19] G. Birkhoff and J. Von Neumann. The logic of quantum mechanics. *Annals of Mathematics*, 37(4):823–843, 1936.

[20] F. G. S. L. Brandão, R. Ramanathan, A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, T. Szarek, and H. Wojewódka. Realistic noise-tolerant randomness amplification using finite number of devices. *Nature Communications*, 7:11345, 2016.

[21] G. Brassard, and P. Raymond-Robichaud Parallel lives: a local-realistic interpretation of "nonlocal boxes. *Entropy* 21(1):87, 2019.

[22] T. A. Brun, J. Finkelstein, and N. D. Mermin. How much state assignments can differ. *Phys. Rev. A*, 65:032315, 2002.

[23] J. Bub. Von Neumann's 'no hidden variables' proof: a re-appraisal. *Foundations of Physics*, 40(9):1333–1340, 2010.

[24] A. Cabello, S. Severini, and A. Winter. Graph-theoretic approach to quantum correlations. *Phys. Rev. Lett.*, 112:040401, 2014.

[25] A. Cabello. Speckers fundamental principle of quantum mechanics. arXiv:1212.1756, 2012.

[26] A. Cabello, J. M. Estebaranz, and G. García-Alcaine. Bell-kochen-spcker theorem: A proof with 18 vectors. *Phys. Lett. A*, 212(4):183–187, 1996.

[27] C. Caves, C. Fuchs, and R. Schack. Conditions for compatibility of quantum-state assignments. *Phys. Rev. A*, 66:062111, 2002.

[28] C. Caves, C. Fuchs, K. Manne, and J. Renes Gleason-Type Derivations of the Quantum Probability Rule for Generalized Measurements *Foundations of Physics*, 34(2):193-209, 2004.

[29] Z. Chen and A. Montina. Measurement contextuality is implied by macroscopic realism. *Phys. Rev. A*, 83:042110, 2011.

[30] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23(15):880–884, 1969.

[31] A. Coladangelo, A. Grilo, S. Jeffery, and T. Vidick. Verifier-on-a-leash: new schemes for verifiable delegated quantum computation, with quasilinear resources. arXiv:1708.07359, 2017.

[32] J. I. de Vicente. On nonlocality as a resource theory and nonlocality measures. *J. Phys. A: Math Theor*, 47(424017), 2014.

[33] D. Deutsch and P. Hayden. Information flow in entangled quantum sytems. *Proc. Roy. Soc. A: Math. Phy.*, 456:1759–1774, 2000.

[34] I. Dinur and S. Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, 162(1):439–485, 2005.

[35] E. N. Dzhafarov, J. V. Kujala, and V. H. Cervantes. Contextuality-by-default: A brief overview of ideas, concepts, and terminology. In H. Atmanspacher, T. Filk, and E. Pothos, editors, *Quantum Interaction*, pages 12–23. Springer International Publishing, 2016.

[36] A. Einstein. Letter to E. Schrödinger, June 17. Einstein Archives Call Number EA 71-836, 1935.

[37] A. Einstein. Letter to E. Schrödinger, June 19. Einstein Archives Call Number EA 22-47, 1935.

[38] A. Einstein. Letter to J. J. Fehr, March 25. Einstein Archives Call Number EA 26-77, 1952.

[39] A. Einstein. Letter to A. Lamouche, March 20. Einstein Archives Call Number EA 15-43, 1955.

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

[41] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.*, 48(5):291–295, 1982.

[42] A. Fine. *The Shaky Game (2nd ed.)*. University of Chicago Press, 1996.

[43] M. R. Garey and D. S. Johnson. *Computers and Intractability; a guide to the theory of NP-completeness*. W.H. Freeman and Company, 1979.

[44] A. M. Gleason. Measures on the closed subspaces of a Hilbert space. *Indiana University Mathematics Journal*, 6:885–893, 1957.

[45] D. M. Greenberger, M. A. Horne, A. Zeilinger, A. Shimony Bell's Theorem without Inequalities. *Am. J. Phys.* (58):1131–43 1990.

[46] L. Ham. Gap theorems for robust satisfiability: Boolean CSPs and beyond. *Theor. Comput. Sci.*, 676:69–91, 2017.

[47] L. Hardy. Nonlocality for two particles without inequalities for almost all entangled states. *Phys. Rev. Lett.*, 71:1665–1668, 1993.

[48] N. Harrigan and T. Rudolph. Ontological models and the interpretation of contextuality. arXiv:0709.4266, 2007.

[49] N. Harrigan, T. Rudolph, and S. Aaronson. Representing probabilistic data via ontological models. arXiv:0709.1149, 2007.

[50] G. Hermann. Die naturphilosophischen Grundlagen der Quantenmechanik. *Naturwissenschaften*, 23(42):718–721, 1935.

[51] M. Howard, J. J. Wallman, V. Veitch, and J. Emerson. Contextuality supplies the 'magic' for quantum computation. *Nature*, 510:351–355, 2014.

[52] C. J. Isham. *Lectures on quantum theory.* Imperial College Press, 1995.

[53] I. D. Ivanovic. How to differentiate between non-orthogonal states. *Phys. Lett. A*, 123(6):257–259, 1987.

[54] S. Jevtic and T. Rudolph. How Einstein and/or Schrödinger should have discovered Bell's theorem. *Journal of the Optical Society of America B*, 32(4):50–55, 2015.

[55] A. Karanjai, J. J. Wallman, and S. Bartlett. Contextuality bounds the efficiency of classical simulation of quantum processes. arXiv:1802.07744.

[56] R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum New York, 1972.

[57] J. Kerman. A closed-form approximation for the median of the beta distribution. arXiv:1111.0433, 2011.

[58] L. A. Khalfin and B. S. Tsirelson. Quantum/classical correspondence in the light of Bell's inequalities. *Foundations of Physics*, 22(7):879–948, 1992.

[59] S. Khot and O. Regev. Vertex cover might be hard to approximate to within 2-ϵ. In *In proceedings of the 18th annual IEEE conference on computational complexity*, pages 379–386. IEEE, 2003.

[60] S. Kochen and E. Specker. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17:59–87, 1967.

[61] R. Kunjwal. Beyond the Cabello-Severini-Winter framework: making sense of contextuality without sharpness of measurements. arXiv:1790.01098, 2017.

[62] A. Lakshminarayan, S. Tomsovic, O. Bohigas, and S. N. Majumdar. Extreme statistics of complex random and quanutm chaotic states. *Phys. Rev. Lett.*, 100:044103, 2008.

[63] R. Lal. Personal Communication, 2014.

[64] M. Leifer. "It from bit and the quantum probability rule. In *It from bit or bit from it?*, pages 5–23, Springer, 2015.

[65] 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.

[66] P. Lillystone, J. J. Wallman, and J. Emerson. Contextuality and the single-qubit stabilizer formalism. arXiv:1802.06121, 2018.

[67] C. Mann and R. Crease. Interview: John bell, 1988.

[68] S. Mansfield and T. Fritz. Hardy's non-locality paradox and possibilistic conditions for non-locality. *Foundations of Physics*, 42(5), 2011.

[69] S. Mansfield. Reality of the quantum state: Towards a stronger ψ-ontology theorem *Phys. Rev. A.*, 94, 042124, 2016.

[70] S. Mansfield. Consequences and applications of the completeness of Hardy's nonlocality. *Phys. Rev. A*, 95:022122, 2017.

[71] S. Mansfield and R. S. Barbosa. Extendability in the sheaf-theoretic approach: Construction of Bell models from Kochen-Specker models. *Quantum Physics and Logic X - 2013 (ICFO, Barcelona)*, pages 1–18, 2013.

[72] N. D. Mermin. Hidden variables and the two theorems of John Bell. *Rev. Mod. Phys.*, 65:803–815, 1993.

[73] A. A. Méthot. Minimal Bell-Kochen-Specker proofs with POVMs on qubits. *International Journal of Quantum Information*, 5(3):353–357, 2007.

[74] A. Montina. Communication cost of classically simulating a quantum channel with subsequent rank-1 projective measurement. *Phys. Rev. A*, 84:06303, 2011.

[75] R. Morris. Topics in quantum foundations: Ontological models, and distinguishability as a resource. (Master's thesis), 2009.

[76] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[77] A. Peres. Two simple proofs of the Kochen-Specker theorem. *J. Phys. A: Math Gen*, 24:175–178, 1991.

[78] J. Pienaar. A seperable, dynamically local ontological model of quantum mechanics. *Foundations of Physics*, 46(1):1–16, 2015.

[79] I. Pitowsky. Correlation polytopes: Their geometry and complexity. *Mathematical Programming*, A50:395–414, 1991.

[80] I. Pitowsky. Betting on the outcomes of measurements: A bayesian theory of quantum probability. *Stud. Hist. Phil. Sci.*, 34:395–414, 2003.

[81] I. Pitowsky. Infinite and finite Gleason's theorems and the logic of indeterminacy. *J. Math. Phys.*, 39(1):218–228, 1998.

[82] S. Poljak. A note on stable sets and colorings of graphs. *Commentationes Mathematica Universitatis Carolinae*, 15(2):307–309, 1974.

[83] M. F. Pusey, J. Barrett, and T. Rudolph. On the reality of the quantum state. *Nature Physics*, 8:475–478, 2012.

[84] H. Putnam. *The Logico-Algebraic Approach to Quantum Mechanics*, volume 1: Historical Evolution, chapter Three-Valued Logic, pages 99–108. D. Reidel Publishing Company, 1975.

[85] A. M. Raigorodskii. On a bound in Borsuk's problem. *Communications of the Moscow Mathematical Society*, 54(2):453–454, 1999.

[86] H. Reichenbach. *The Logico-Algebraic Approach to Quantum Mechanics*, volume 1: Historical Evolution, chapter Three-Valued Logic and the Interpretation of Quantum Mechanics, pages 53–98. D. Reidel Publishing Company, 1975.

[87] R. Renner and S. Wolf. Quantum pseudo-telepathy and the Kochen-Specker theorem. In *International Symposium on Information Theory, 2004, Proceedings of the*, pages 322–322, 2004.

[88] M. A. Rubin. Locality in the Everett Interpretation of Heisenberg-Picture Quantum Mechanics. *Foundations of Physics Letters*, 14:301–322, 2001.

[89] T. Rudolph. Ontological models for quantum mechanics and the Kochen-Specker theorem. arXiv:quant-ph/0608120, 2006.

[90] A. W. Simmons. Contextuality, nonlocality, preparation independence and the power of quantum computers. (Master's thesis), 2015.

[91] A. W. Simmons. How (maximally) contextual is quantum mechanics? arXiv:1712.03766, 2017.

[92] A. W. Simmons. On the computational complexity of detecting possibilistic locality. *Journal of Logic and Computation*, 28(1):203–217, 2018.

[93] A. W. Simmons, J. J. Wallman, H. Pashayan, S. D. Bartlett, and T. Rudolph. Contextuality under weak assumptions. *New Journal of Physics*, 19:003030, 2017.

[94] E. Specker. Die Logik nicht gleichzeitig entscheidbarer Aussagen. *Dialectica*, 14:239–246, 1960.

[95] R. W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Phys. Rev. A*, 71:052108, 2005.

[96] R. W. Spekkens. The status of determinism in proofs of the impossibility of a noncontextual model of quantum theory. *Foundations of Physics*, 44(11):1125–1155, 2014.

[97] R. W. Spekkens. Leibniz's principle of the identity of indiscernables as a foundational principle for quantum theory. Talk at Information-theoretic interpretations of quantum mechanics conference, 2016.

[98] P. M. L. Tammes. On the origin of number and arrangement of the places of exit on pollen grains. *Diss. Groningen*, 1930.

[99] J. J. Thomson. On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure. *Philosophical Magazine*, 7(39):237–265, 1904.

[100] B. S. Tsirelson. Personal Communication, 2018.

[101] S. Uijlen and B. Westerbaan. A Kochen-Specker system has at least 22 vectors. *New Generation Computing*, 34(1-2):3–23, 2016.

[102] A. M. Vershik. Personal Communication, 2018.

[103] J. Von Neumann. *Mathematical foundations of quantum mechanics*. Princeton University Press, 1955.

[104] M. Waegell and P. K. Aravind. The minimum complexity of Kochen-Specker sets does not scale with dimension. *Phys. Rev. A*, 95:050101, 2017.

[105] L. Wester. Almost equivalent paradigms of contextuality. In B. Coecke and A. Kissinger, editors, *Proceedings 14th International conference on quantum physics and logic*, pages 1–22, 2018.

[106] H. S. Witsenhausen. Spherical sets without orthogonal point pairs. *American Mathematical Monthly*, pages 1101–1102, 1974.

[107] K. Życzkowski and H.-J. Sommers. Truncations of random unitary matrices. *Journal of Physics A: Mathematical and General*, 33(10):2045, 2000.