

Sheaf-Theoretic Methods in Quantum Mechanics and Quantum Information Theory



Carmen Maria Constantin

University College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2015

“All models are wrong, but some are useful.”

George E. P. Box

Abstract

In this thesis we use the language of sheaf theory in an attempt to develop a deeper understanding of some of the fundamental differences - such as entanglement, contextuality and non-locality - which separate quantum from classical physics.

We first present, based on the work of Abramsky and Brandenburger [2], how sheaves, defined over certain posets of physically meaningful contexts, give a natural setting for capturing and analysing important quantum mechanical phenomena, such as quantum non-locality and contextuality. We also describe how this setting naturally leads to a three level hierarchy of quantum contextuality: weak contextuality, logical non-locality and strong contextuality. One of the original contributions of this thesis is to use these insights in order to classify a particular class of multipartite entangled states, which we have named balanced states with functional dependencies. Almost all of these states turn out to be at least logically non-local, and a number of them even turn out to be strongly contextual. We then further extend this result by showing that in fact all n -qubit entangled states, with the exception of tensor products of single-qubit and bipartite maximally-entangled states, are logically non-local. Moreover, our proof is constructive: given any n -qubit state, we present an algorithm which produces $n + 2$ local observables witnessing its logical non-locality.

In the second half of the thesis we use the same basic principle of sheaves defined over physically meaningful contexts, in order to present an elegant mathematical language, known under the name of the Topos Approach [62], in which many quantum mechanical concepts, such as states, observables, and propositions about these, can be expressed. This presentation is followed by another original contribution in which we show that the language of the Topos Approach is at least as expressive, in logical terms, as traditional quantum logic. Finally, starting from a topos-theoretic perspective, we develop the construction of contextual entropy in order to give a unified treatment of classical and quantum notions of information theoretic entropy.

Acknowledgements

The research presented in this thesis has been conducted under the guidance of my supervisors, Prof. Samson Abramsky and Dr. Andreas Döring. Their enthusiasm for research has been contagious and a real source of inspiration for me, throughout my DPhil studies. I am especially grateful for their constant encouragement and support, and for allowing me to enjoy a very large degree of flexibility for the entire duration of this project.

My contact with the other members of the university has both shaped and nurtured my growth as a researcher and as a human being. I would like to thank my colleagues and professors for their friendship and their open attitude towards my ideas and questions, especially Prof. Bob Coecke, Prof. Kobi Kremnitzer, Prof. Boris Zilber, Prof. Prakash Panangaden, Rui Soares Barbosa, Shane Mansfield, Nadish de Silva, Aleks Kissinger, Ray Lal, Jamie Vicary, Chris Heunen, Alex Merry, Brendan Fong, Miriam Backens, Hugo Nava and Ondrej Rypacek.

I am indebted to my husband, Andrei Constantin, far more than I can express in these few lines. He has supported me in every way possible, and has even patiently guided me through some of the more puzzling conceptual aspects of quantum mechanics. He has proofread this thesis, and in fact all my work, and has always helped me organize and express my ideas.

I am also grateful to my teachers and mentors from university and high school which have kindled and nurtured my love for mathematics in general and for sheaf theory in particular: Dr. Antony Maciocia, Enache Pătraşcu and Constantin Tănăsescu.

Special thanks are also due to my formidable grandmother Nona, and to my aunt, Rodica. Without their kind and timely offer to help with child care related matters, this thesis would have taken much longer to complete.

Finally, I would like to acknowledge the financial support I have received from the EPSRC in the form of a generous scholarship including maternity support.

To my late parents, Dana-Maria and Tudor-Alin

Contents

1	Introduction	1
1.1	The Topos Approach	2
1.1.1	A Quantum State Space	3
1.1.2	From Boolean to Intuitionistic Logic and Back	4
1.2	Sheaves and Contextuality	5
1.2.1	Three Levels of contextuality	6
1.3	Outline	7
I	First part	11
2	The Sheaf Theoretic Structure of Contextuality and non-Locality	13
2.1	Entanglement, Non-locality and Contextuality	13
2.1.1	Qubits and qubit states	13
2.1.2	Entanglement	14
2.1.3	Measures of entanglement for bipartite states	15
2.1.4	Measures of entanglement for multipartite states	16
2.1.5	Non-locality, Bell's theorem and hidden variables	17
2.1.6	Measurement contexts	18
2.1.7	Contextuality and the Kochen-Specker theorem	21
2.2	Non-locality and Contextuality in Sheaf-Theoretic Language	21
2.2.1	Measurement covers, the distribution functor and no-signalling	22
2.2.2	Global Sections	26

2.2.3	Existence of Global Sections	27
2.3	A Hierarchy of Contextuality	30
2.3.1	Strong contextuality as a CSP	32
2.3.2	Contextuality for Quantum States	33
2.3.3	The Bell state	34
2.3.4	The possibilistic Hardy model	35
2.3.5	Dicke states	36
2.3.6	Permutationally symmetric states	38
2.3.7	GHZ models	38
3	Classifying a Class of Permutationally Asymmetric Quantum States	41
3.1	Balanced States with Functional Dependency	42
3.2	Contextuality classification for the tripartite case	43
3.2.1	Polynomials of degree zero	43
3.2.2	Degree one polynomials	43
3.2.3	Degree two polynomials	50
3.3	Contextuality classification for the $n + 1$ -partite case, $n > 2$	53
3.3.1	Strongly contextual states	53
3.3.2	Logically contextual states	54
3.3.3	Weakly contextual states	55
3.3.4	Non-contextual states	56
4	Logical Contextuality is Almost Everywhere	57
4.1	Overview of the argument	58
4.2	Auxiliary lemmas	60
4.2.1	Background lemmas	60
4.2.2	The Base Case lemma	61
4.2.3	The Going Up lemmas	62
4.2.4	The Small Difference lemma	65
4.2.5	The 21 lemma	68

4.3	The algorithm	70
II	Second part	75
5	Overview of the Topos Approach	77
5.1	The Basics	77
5.1.1	Categories and functors	77
5.1.2	Limits and colimits	80
5.1.3	Lattices and Heyting algebras	82
5.1.4	Topoi	84
5.2	C*-Algebras and von Neumann Algebras	87
5.2.1	Gelfand duality and Gelfand spectra	90
5.3	The spectral presheaf	91
5.3.1	Daseinisation	94
5.3.2	States as measures on the spectral presheaf	98
6	Quantum Logic vs. the Logic of the Topos Approach	105
6.1	Grouping and splitting	106
6.2	Spiked BSAs	110
6.3	Reconstructing L from $\mathcal{B}(L)$	112
7	Entropy within the Topos Approach	117
7.1	Entropy in Classical and Quantum Mechanics	118
7.1.1	Majorization order	118
7.1.2	Shannon entropy	119
7.1.3	Von Neumann entropy	122
7.1.4	Renyi entropy	124
7.2	Contextual entropy	125
7.2.1	Measures and partial traces	125
7.2.2	The entropy of a measure	127

7.2.3	Properties of the contextual entropy	129
7.3	Reconstructing pure states from global sections	137
7.4	Reconstructing arbitrary quantum states from global sections	137
7.4.1	Two-dimensional Hilbert spaces	141
7.5	Renyi entropy	141
7.5.1	Properties	142
8	Conclusions and Outlook	145
	Appendices	149
A	The Subadditivity Property of Contextual Entropy	151
	Bibliography	155

Chapter 1

Introduction

The emergence of Quantum mechanics at the beginning of the last century has shaken many of the intuitions underlying classical physics. One of its most peculiar aspects comes from the fact that quantum mechanical states induce, in general, only statistical restrictions on the results of measurements, instead of definite outcomes. One could be tempted to draw, as Einstein did, the conclusion that these states are therefore incomplete descriptions of quantum systems. For a while physicists have wondered whether quantum mechanics could be supplemented with a more complete description involving hidden variables. Thus quantum probabilities could be naturally interpreted as epistemic probabilities, of the sort that arise in ordinary statistical mechanics.

The existence of hidden variable theories has been refuted by two powerful theorems. The more famous of these is Bell's theorem which uses correlations between entangled states to show that, assuming the premise of locality (i.e. the idea that spatially separated systems cannot instantaneously influence each other), any hidden variable model can be used to derive an inequality which is violated by the predictions of quantum mechanics and these predictions have been experimentally confirmed.

The second important no-go theorem is that of Kochen-Specker, which shows that it is not possible to assign values to observables under the premise of non-contextuality - which is the assumption that if a quantum system possesses a property (value of an observable), then it does so independently of any measurement context, that is, independently of which other observables are measured alongside the one under consideration.

The more recent rise of Quantum Information Theory is now posing new challenges to our understanding of both computation and physics. Entanglement, non-locality and contextuality, which have proved so problematic to our understanding of quantum theory, have turned out, when seen from a different viewpoint, to offer exciting new possibilities, thus challenging the

usual assumptions of classical computation. Quantum systems have been shown to be able to perform information-theoretic tasks beyond the capabilities of classical systems. Famous examples include secure key commitment [17, 67], quantum teleportation [18] and factoring primes in polynomial time [118].

The work presented in this thesis has been guided and motivated by several recent attempts to use the powerful mathematical tools of sheaf and category theory in order to obtain deeper structural insights into the nature of physical reality. These insights would hopefully allow us to fully harness the resources offered to information processing by quantum physics.

1.1 The Topos Approach

The topos approach to the formulation of physical theories was initiated by Chris Isham and Jeremy Butterfield [28–31, 89], expanded by Andreas Döring and Chris Isham [44–49, 52, 58–63], and further developed by Heunen, Landsman, Spitters and Wolters [84–86, 126], Flori and collaborators [70–72] and others [106, 108, 127]. One of the initial ideas which motivated the programme was that by choosing a suitable universe of mathematical discourse, that is to say a suitable topos, one could express quantum mechanical concepts in a way which would render them structurally similar to their classical mechanical counterparts. It was expected that such a characterisation would bring to light new potentialities of the quantum world.

While recent results, including the approach to non-commutative K -theory developed by Nadish de Silva [40], the new perspective on quantum probability developed by Dewitt and Döring [55, 56] and the non-commutative generalisation of Gelfand duality obtained by Döring [48, 49, 51] give full justice to the claim that the formalism underpinning the topos approach is in itself rich enough to be of mathematical interest, the choice of formulating real world concepts within generalized mathematical universes appears rather counter-intuitive.

Mathematics is the language in which physical theories are formulated and perhaps one of the reasons why the universe of sets and functions and the Boolean logic associated with it have provided the foundation of most of the current mathematical discourse is the fact that mathematics has been itself shaped by our experience of reality which, until relatively recently, has only been directly perceived at the macroscopic level.

It is possible that, if we were living in a quantum world, our primary intuitions would be different and our mathematics would not necessarily be based on set theory. It is at least an interesting coincidence that the change in our basic perception of the natural world brought about by the development of quantum mechanics has taken place in the same century in which

category theory, through the notion of topos, has succeeded in axiomatising set-theory, thus bringing about an entirely new, categorical foundation of mathematics.

With these considerations in mind, one finds it less surprising that the same concepts which have had such an impact on the foundation of mathematics have eventually found their way into the realm of foundations of physics.

However, apart from these general arguments, there is of course no a priori reason why quantum mechanical concepts, for example, should find a formulation in topos theoretic language. The fact that an elegant formulation does exist is in itself rather remarkable and often comes as a result of powerful theorems.

1.1.1 A Quantum State Space

Both classical mechanics and quantum mechanics have been traditionally formulated in the topos of sets. Within this topos, the state space of a classical system is a set, which additionally has the structure of a symplectic manifold. The points of this set are the pure states in which the system can be. Physical observables can be thought of as real-valued functions on the state space and, as such, they form a commutative algebra under pointwise multiplication. This commutative algebra and the state space of the system effectively determine each other. This is a consequence of a mathematical result known as Gelfand duality, which establishes a bijective correspondence between compact Hausdorff spaces and unital commutative C^* -algebras.

This correspondence does not hold for quantum mechanical systems which, in the usual Hilbert space formalism, have non-commutative algebras of observables given by the self-adjoint operators of the Hilbert space corresponding to the system. Non-commuting operators correspond to incompatible physical observables, that is, to observables whose values cannot be measured at the same time such as, for example, position and momentum.

Several approaches [12, 22, 46, 74, 105] to extending Gelfand duality to non-commutative algebras were compared in [90]. One of these has been further developed by Döring in [48, 49, 51]. The underlying idea is to consider all commutative sub-algebras of the non-commutative algebra of bounded operators on the Hilbert space associated with the system. Each of these subalgebras can be interpreted as a classical ‘perspective’ on the physical system, as it only contains pairwise commutative observables. We call it a *classical context*. Since two self-adjoint operators with discrete spectra commute only if they have a joint set of eigenspaces, in finite dimensions a classical context in effect corresponds to a set of pairwise orthogonal projections which add up to identity.

Indeed, such projections generate, via the von Neumann double commutant construction, a commutative von Neumann algebra. We can use Gelfand duality to associate a classical state space to this algebra. The collection of all such classical state spaces forms a presheaf over the base category given by the set of commutative subalgebras, partially ordered by inclusion. The spectral presheaf is hence an object in the topos of presheaves over this base category.

The analogy between the spectral presheaf and a classical state space is further justified by the fact that the set of quantum states is equivalent to the set of measures on the spectral presheaf, just like in classical mechanics mixed states are given by probability measures on state space. Pure states are given by ‘minimal’ measures, but it must be noted that unlike their classical counterparts, these are not concentrated at points. A point of a presheaf, in the category theoretical sense, is given by a global section, but the spectral presheaf has no global sections. The lack of global sections of the spectral presheaf is equivalent [28] to the Kochen-Specker theorem which asserts that it is not possible to simultaneously assign values to all observables on a quantum system. The important insight provided by this elegant reformulation [28] was one of the sources of inspiration behind both the topos programme as well as the later sheaf-theoretic approach to characterising contextuality and non-locality developed in Abramsky and Brandenburger’s paper [2].

1.1.2 From Boolean to Intuitionistic Logic and Back

We have mentioned that topoi include, but are more general than sets. Moreover each topos comes equipped with its own logical calculus. Just like in the universe of sets and functions the principles of classical logic are represented by operations on a certain set, namely the two-element Boolean algebra, each topos carries an analogue of this algebra. This is called a Heyting algebra, and the logical principles which hold in a topos turn out to be precisely those of intuitionistic logic.

The defining characteristic of intuitionistic logic is that the law of excluded middle does not necessarily hold. Hence the main difference between theorems proved using Heyting logic and those using Boolean logic is that proofs by contradiction cannot be used in the former. Döring and Isham argue in [62] that this is not a major restriction and intuitionistic logic is a viable alternative to classical logic for the purpose of building physical theories.

In particular, the internal, multivalued topos logic allows us to assign (generalised) truth values to all propositions about a quantum system. As mentioned above, Kochen and Specker proved [96] that such assignments are impossible if one works with classical two-valued logic.

However, one must carefully distinguish between the internal logic which is adequate to the system in itself, and the meta-logic in which we argue *about* the system. Indeed we, as macroscopic entities, use a meta-logic, typically Boolean, to reason about the world, and to define mathematical structures.

Although interesting work has been done by Heunen, Landsman, Spitters and Wolters [84–86, 126] and Fauser, Raynaud and Vickers [68] using the internal topos logic, it is also possible to reason about a topos in an external fashion. As Döring argues in [50], in order to provide an objective report of some phenomena outside ourselves, we must separate ourselves from the system we are considering, hence in our description and mathematical arguments about a physical system we are free to use the metalogic. In technical terms, this is equivalent to working in an ambient topos which is the familiar topos of sets and functions. This external perspective has been adopted in the present work as well.

1.2 Sheaves and Contextuality

We have mentioned above that one of the basic ingredients of the topos approach is given by the poset of contexts. The contexts capture the physical idea of measurements which can be performed jointly. One can achieve an even higher level of generality by replacing the formalism of operator algebras with abstractly defined families of maximal sets of commuting observables. This is done by Samson Abramsky and Adam Brandenburger in [2] in order to study the key information theoretic resources of contextuality and non-locality. Their approach covers n -partite Bell-type scenarios as well as Kochen-Specker configurations, and many other examples relevant to quantum information theory.

The basic ingredients of the mathematical formalism used by Abramsky and Brandenburger include a set X of labels for observables, a cover $\mathcal{U}$ of X consisting of subsets $U \subseteq X$ (called measurement contexts) which correspond to the different combinations of observables which can be measured together, and a set O which labels the possible outcomes. A joint outcome for any compatible set of measurements U is specified by a function $s : U \rightarrow O$, which is called a section over U . From these ingredients one can build a sheaf of events $\mathcal{E}$ which associates to each U the set O^U of sections over U , together with restriction maps given by function restriction. In a quantum mechanical setting, this sheaf is closely related to Isham and Butterfield’s spectral presheaf.

This sheaf of events is then composed with a certain distribution functor to obtain a presheaf of R -valued distributions, $\mathcal{D}_R\mathcal{E}$, which assigns to each $U \subseteq X$ the set $\mathcal{D}_R(\mathcal{E}(U))$ of

distributions on $\mathcal{E}(U)$. In this thesis the ring R will be taken to be either the positive reals or the Boolean ring.

Measurement covers are defined as maximal sets of compatible measurements. An empirical model over a given measurement cover $\mathcal{M} \subset \mathcal{P}(X)$ is then defined as a compatible family $\{e_C\}_{C \in \mathcal{M}}$ of probability distributions e_C on $\mathcal{E}(C)$.

Using these basic ingredients, a novel three-level characterisation of multipartite quantum states in terms of their degree of contextuality is obtained. This characterisation is linked to the lack of global sections of the presheaf $\mathcal{D}_R\mathcal{E}$, compatible with a given empirical model.

1.2.1 Three Levels of contextuality

The first, and lowest, degree of contextuality described by Abramsky and Brandenburger - weak, or probabilistic contextuality - generalizes the original argument used in Bell's theorem. Bell's argument essentially relies on the probabilistic predictions of quantum mechanics, which are inconsistent with the predictions of any local realistic theory. Abramsky and Brandenburger show that the content of Bell's argument can be summarised by an empirical model over the positive reals, that is by a probability table containing the empirical predictions for all allowed combinations of measurements. The information encoded by this empirical model can be used to derive a proof of Bell's theorem, without recourse to inequalities, but based on the non-existence of a certain joint distribution, or global section.

The second, intermediate degree of contextuality - logical, or possibilistic contextuality - generalizes an argument used by Hardy [82, 83], who showed that an inequality-free proof of Bell's theorem could be given for almost all 2-qubit systems. Hardy's construction works for all bipartite entangled states, except for the maximally entangled states. The content of Hardy's argument can again be summarised by an empirical model, this time over the Boolean ring, as Hardy's argument essentially only relies on the possibility (probability greater than 0) and respectively impossibility (probability 0) of certain measurement outcomes, that is, it only relies on the support of the probability distributions. Abramsky and Brandenburger show that the possibilistic contextuality of the Hardy model is a stronger property than the probabilistic contextuality used in the Bell argument (i.e. the Bell model is not logically contextual, but any probabilistic model whose support coincides with the Hardy model must be both weakly and logically contextual).

The third degree of contextuality - strong contextuality - generalizes an argument used by Greenberger, Horne, Zeilinger and Shimony [78, 79], who used the non-classical properties

of certain quantum states to give a strengthened inequality-free proof of Bell's theorem. Their proof, which predates Hardy's argument, needs – on the other hand – systems of at least three qubits, such as the entangled three-qubit GHZ state. The GHZ argument, like the Hardy one, only relies on the possibility or impossibility of certain measurement outcomes. Yet it is shown by Abramsky and Brandenburger that the three qubit GHZ model satisfies a stronger property than the logical contextuality of the Hardy model, which is logically but not strongly contextual. It is also interesting to note that strong contextuality can be exhibited in the usual n -qubit multipartite Bell-type scenarios only if we have a scenario with three or more parties.

This line of work has been further developed by Abramsky and Hardy in [7] where they introduce a notion of logical Bell inequality, based on logical consistency conditions. Logical Bell inequalities can be used to obtain proofs of Bell's theorem without probabilities, but also to derive testable inequalities with provable violations for a wide variety of situations. It can also be shown that measurement models achieve maximal violations of logical Bell inequalities if and only if they are strongly contextual. Non-maximal violations are achieved by measurement models which are possibilistically contextual – that is, they occupy the middle level of the hierarchy.

1.3 Outline

According to the type of methodology used, the material presented in this thesis has been divided into two parts. Each part begins with a background chapter which lays out the necessary mathematical and physical terminology.

Thus Part I starts with Chapter 2 in which the reader is first given a brief overview of some of the main information theoretic concepts, such as qubits and quantum gates. We also present some of the intriguing properties which distinguish quantum mechanical systems from classical ones, such as entanglement, non-locality and contextuality.

We then proceed to introduce some sheaf theoretic constructions which can be used to describe general experimental scenarios. These are modelled in great generality as compatible families of measurements. We also show how, for such scenarios, compatibility in a sheaf theoretic sense corresponds to a physical condition motivated by Special Relativity, which is known as no-signalling. In particular, the fact that quantum mechanics satisfies no-signalling is taken to indicate a basic consistency between quantum mechanics and relativity. This core mathematical structure is then used to analyse contextuality and non-locality in a unified way. In particular, we show that these phenomena can be characterised precisely in terms of obstructions to the

existence of global sections of a certain distribution presheaf.

The chapter ends by distinguishing the three strengths of degree of contextuality: standard probabilistic non-locality, exhibited by the original example introduced by Bell; possibilistic contextuality, exemplified in our presentation by the Hardy model and by permutationally symmetric states; and strong contextuality, exemplified by GHZ states.

In chapter 3 we use this hierarchy to classify balanced quantum states with functional dependencies. These are non-permutationally symmetric entangled states which are described by Boolean functions. The classification yields a large number of strongly contextual states. Moreover, it turns out that all the states considered, except those which are equivalent to tensor products of pure states and Bell pairs, are at least logically contextual.

Chapter 4 extends the results of the previous chapter's classification by showing that in fact all n -qubit quantum states, with the exception of tensor products of pure states and Bell pairs, are logically contextual for suitable choices of measurements. We moreover show that $n+2$ observables suffice to witness the logical contextuality of any n -qubit state: two observables each for two of the parties, and one each for the remaining $n - 2$ parties. Our proof is constructive, so we are also able to present an algorithm which returns the witnessing local observables for any n -qubit state.

* * *

The second part of this thesis begins with Chapter 5, in which the reader is first guided through the category theoretical notions which are necessary for understanding the definition of an elementary topos.

This is followed by a short presentation of the algebraic-geometric duality (Gelfand's duality) which is used in defining the main construction of the topos approach, the spectral presheaf. Finally the reader is introduced to the way in which several basic physical concepts, such as the state space, real number objects, and the collection of pure and mixed states of a system are formulated within the topos approach. These will be the main building blocks used in the construction of contextual entropy in Chapter 7. We also show how projections, which are used to represent propositions about a system in the Hilbert space formulation of quantum mechanics, can be represented in the topos formalism.

In Chapter 6 we discuss whether the topos-based form of logic for quantum systems is (at least) as rich as the traditional quantum logic derived from the Hilbert space formulation of quantum mechanics. Our discussion revolves around the question whether the orthomodular lattice of projections on a Hilbert space, which is traditionally used in quantum logic to represent

propositions, can be reconstructed from the poset of contexts which underlies the constructions in the topos approach. This question has a positive answer, according to a result obtained by Harding and Navara in [80]. Our contribution in this chapter seeks to answer an open problem stated at the end of [80], which asks for an explicit reconstruction of an orthomodular lattice from its associated poset of distributive sub-lattices. We show that for atomic orthomodular lattices a reasonably direct reconstruction is indeed possible.

In Chapter 7 we take a look at the concept of entropy, one of the cornerstones of information theory, from the perspective of the topos approach. Since our reformulation of entropy does not directly depend on the interpretation of states as density matrices within the usual Hilbert space formalism of quantum mechanics, this endeavour is particularly relevant for the topos programme.

The chapter starts with a brief outline of some of the properties of Shannon entropy, its quantum counterpart, von Neumann entropy and the more general family of Renyi entropies.

Using Döring’s reformulation of quantum states as probability measures [44] on the spectral presheaf, we next show how a classical probability distribution, and hence also its corresponding Shannon entropy, can be naturally associated to each context in the base category over which the spectral presheaf is constructed. The numerical values of these Shannon entropies form a global section of a presheaf of real values over the same base category, the poset of contexts. This global section is the state’s contextual entropy. A powerful result known as the Schur-Horn Lemma can be used to show that this construction unifies the Shannon and von Neumann entropies by encoding the von Neumann entropy of the state in a distinguished way.

A comparison is then drawn between Shannon, von Neumann and contextual entropies. This allows us to observe, for example, that one of the differences between Shannon and von Neumann entropies (the property of being monotone) is precisely due to contextuality, an idea which is not immediately obvious from the definition of these two entropies.

Perhaps the main result of this chapter is the informatic-theoretic characterisation of quantum states provided by contextual entropy which can be shown to be rich enough to allow one to reconstruct the quantum state from which it originated. This also provides us with a new insight into Gleason’s theorem.

At the end of the chapter we show that it is possible to adapt other classical entropies within the formalism of the topos approach, given that they satisfy a certain weak recursivity property. We explicitly describe how the contextual entropy construction can be generalised to include Renyi entropies, and we observe that contextual Renyi entropies also encode sufficient

information to allow for a complete state reconstruction.

This approach has been inspired by one of the key ideas of the topos programme, namely that one can hope to obtain a complete description of a quantum system by looking at it from all possible classical perspectives and keeping track of the resulting information.

The final chapter, namely Chapter 8, contains a summary of the main results presented in this thesis, as well as a number of concluding remarks and directions for further work.

The original material presented in this thesis is drawn from four research papers. The material presented in Chapter 3 is based on:

1. S. Abramsky, C. M. Constantin, *A Classification of Multipartite States by Degree of Non-locality*, EPCS, Proceedings of QPL 2013, arXiv:1412.5213, [5]

Chapter 4 is based on:

2. S. Abramsky, C. M. Constantin, S. Ying, *Hardy is (almost) everywhere: Non-locality without inequalities for almost all entangled multipartite states*, arXiv:1506.01365, [6]

Chapter 6 is based on:

3. C. M. Constantin, A. Döring, *Reconstructing an Atomic Orthomodular Lattice from the Poset of its Boolean Sublattices*, Houston Journal of Mathematics, arXiv:1306.1950, [36]

Finally, Chapter 7 is based on:

4. C. M. Constantin, A. Döring, *Contextual Entropy and Reconstruction of Quantum States*, arXiv:1208.2046, [54]

Part I

First part

Chapter 2

The Sheaf Theoretic Structure of Contextuality and non-Locality

We begin this chapter by reviewing a number of fundamental notions in quantum information theory, such as entanglement, non-locality, hidden variables, Bell inequalities and logic gates. We show that these notions can be re-expressed in an elegant way using sheaf-theoretic language. The presentation is based on the textbook [107] and the original work of Abramsky, Brandenburger and Hardy [2, 7].

2.1 Entanglement, Non-locality and Contextuality

2.1.1 Qubits and qubit states

Classical information theory relies on the concept of *bit*, understood as the basic unit of information, with possible values 0 and 1. Its quantum analogue, the *quantum bit*, or *qubit* for short, is understood as a quantum system with two states $|0\rangle$ and $|1\rangle$. More formally, a qubit is an element of a 2-dimensional Hilbert space¹ (usually over $\mathbb{C}$). While a classical bit can only have one value at a time, 0 or 1, its quantum counterpart is, in general, a superposition of $|0\rangle$ and $|1\rangle$:

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$$

The qubit $|\psi\rangle$ is said to be in a pure state if $|\alpha|^2 + |\beta|^2 = 1$ and in a mixed state if it is a statistical mixture of different pure qubit states.

¹Other frequent notations for a 2-dimensional orthonormal basis, which we will also use in this work, are $\{|-\rangle, |+\rangle\}$ and $\{|\downarrow\rangle, |\uparrow\rangle\}$.

Pure qubit states can be represented as points on the so-called Bloch sphere parametrized by $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$, by setting $\alpha = e^{i\phi} \cos(\theta/2)$ and $\beta = \sin(\theta/2)$. On the Bloch sphere, the North and the South poles correspond to the points accessible to a classical bit, while the interior of the Bloch sphere corresponds to mixed states.

A system of n -qubits represents a finite-dimensional Hilbert space over the complex numbers of dimension 2^n . A state $|\psi\rangle$ in this Hilbert space is represented as a linear combination

$$|\psi\rangle = \sum_{j_1, j_2, \dots, j_n=0,1} c_{j_1, j_2, \dots, j_n} |j_1\rangle \otimes |j_2\rangle \otimes \dots \otimes |j_n\rangle$$

where $|0\rangle$ and $|1\rangle$ are the two states of a qubit. The vectors $|j_1\rangle \otimes \dots \otimes |j_n\rangle$ form an orthonormal basis and are usually labelled by binary strings of length n , $|j_1 j_2 \dots j_n\rangle$.

A quantum computation process on n qubits represents the following sequence of steps. First, we assemble n qubits and prepare them in a standard initial state. Then we apply a unitary transformation U , usually written as a product of *quantum gates*, i.e. unitary transformations that act on a small number of qubits. Finally, we measure all the qubits, by projecting on the $\{|0\rangle, |1\rangle\}$ basis, which represents the (probabilistic) outcome of the computation.

2.1.2 Entanglement

Multiple qubits can exhibit quantum entanglement. *Entanglement* is a striking feature of quantum mechanics, on which many of its information theoretic successes that go beyond classical physics, such as quantum teleportation [18], superdense coding [19, 118] and quantum cryptography [67], rely. As a result of interactions, quantum systems (e.g. qubits), can become *entangled*, giving rise to correlations between the properties of the constituent systems which persist even when these become spatially separated. The possibility of having entangled systems over large distances, leading to non-local correlations, made Albert Einstein, Boris Podolski and Nathan Rosen to conclude in 1935 that quantum mechanics must be an incomplete description of reality [66]. However, nearly 30 years later, John Stewart Bell realised in his groundbreaking paper “On the Einstein Podolsky Rosen Paradox” [15] that any physical theory which would satisfy the principle of locality (stating that an object is directly influenced only by its immediate surroundings) must necessarily satisfy certain inequalities. Bell then showed that the predictions of quantum mechanics violate these inequalities. We will examine this circle of ideas into more detail in Section 2.1.5.

Since quantum entanglement appears as an indispensable resource in quantum information processing, we will give particular attention to understanding the ways in which it can be

quantified. To begin with, let us recall the basic definition of entanglement for bipartite (2-qubit) systems.

Definition 2.1.1. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces and $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$. Then $|\psi\rangle$ is said to be disentangled, or separable or a product state if $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$, for some $|\psi_1\rangle \in \mathcal{H}_1$ and $|\psi_2\rangle \in \mathcal{H}_2$. Otherwise, $|\psi\rangle$ is said to be entangled.

Example 2.1.2. The EPR-state (Einstein-Podolsky-Rosen state) is defined as

$$|\psi\rangle_{EPR} = \frac{1}{\sqrt{2}} (|0\rangle \otimes |1\rangle - |1\rangle \otimes |0\rangle) \quad (2.1.1)$$

One can easily check that the EPR-state cannot be written as a product state.

Example 2.1.3. In the 2-qubit Hilbert space $\mathbb{C}^2 \otimes \mathbb{C}^2$, the Bell states are given by:

$$\begin{aligned} |\Phi^+\rangle &= \frac{1}{\sqrt{2}} (|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle) & |\Phi^-\rangle &= \frac{1}{\sqrt{2}} (|0\rangle \otimes |0\rangle - |1\rangle \otimes |1\rangle) \\ |\Psi^+\rangle &= \frac{1}{\sqrt{2}} (|0\rangle \otimes |1\rangle + |1\rangle \otimes |0\rangle) & |\Psi^-\rangle &= \frac{1}{\sqrt{2}} (|0\rangle \otimes |1\rangle - |1\rangle \otimes |0\rangle) \end{aligned} \quad (2.1.2)$$

where $\{|0\rangle, |1\rangle\}$ represents an arbitrary orthonormal basis of the 1-qubit Hilbert space $\mathbb{C}^2$. The Bell states are also entangled.

2.1.3 Measures of entanglement for bipartite states

The distinction between separable and entangled states is mathematically straightforward. However, in practice, it is generally difficult to discern. As such it is important to find operational criteria to test separability.

The most crude measure of entanglement is the *Schmidt number* (also called the *Schmidt rank*). For a state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ the Schmidt number over $\mathcal{H}_A \otimes \mathcal{H}_B$ is the smallest number $\text{Sch}(|\psi\rangle, \mathcal{H}_A, \mathcal{H}_B)$ such that $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_{j=1}^{\text{Sch}(|\psi\rangle, \mathcal{H}_A, \mathcal{H}_B)} |u_j\rangle \otimes |v_j\rangle$$

where $|u_j\rangle \in \mathcal{H}_A$ and $|v_j\rangle \in \mathcal{H}_B$. Thus a separable state has Schmidt number 1 and an entangled state has Schmidt number greater than 1. The Schmidt number can also be defined using the Schmidt decomposition of $|\psi\rangle$ over $\mathcal{H}_A \otimes \mathcal{H}_B$:

$$|\psi\rangle = \sum_{j=1}^{\min(\dim \mathcal{H}_A, \dim \mathcal{H}_B)} s_j |u_j\rangle \otimes |v_j\rangle$$

where $|u_j\rangle$ form an orthonormal basis of $\mathcal{H}_A$ and $|v_j\rangle$ are an orthonormal basis of $\mathcal{H}_B$ and the coefficients s_i are non-negative. The strictly positive coefficients are called Schmidt coefficients and their number corresponds to the Schmidt number. Alternatively, form the rank 1 matrix $\rho = |\psi\rangle\langle\psi|$ and take the partial trace with respect to either A or B . This will be a diagonal matrix with non-zero elements $|s_i|^2$. Thus the Schmidt number can also be found by counting the non-zero eigenvalues of $\rho_A := \text{tr}_B \rho$ or, equivalently, the non-zero eigenvalues of $\rho_B := \text{tr}_A \rho$.

For bipartite pure states, the above comments lead to another measure of entanglement, the *entropy of entanglement*. Entangled states have Schmidt number greater than 1, therefore a bipartite pure state is entangled if and only if its reduced states are mixed states. Consequently, the von Neumann entropy of either reduced state gives a well defined measure of entanglement:

$$S = -\text{tr}(\rho_A \log \rho_A) = -\text{tr}(\rho_B \log \rho_B)$$

which can be written as the Shannon entropy $S = -\sum_i p_i \log p_i$ where, e.g. ρ_A is written in terms of its eigenvectors $\rho_A = \sum_i p_i |i\rangle\langle i|$.

Many other entanglement measures for bipartite states exist, such as topological entanglement entropy, entropy of formation and dilution, squashed entanglement etc. For a review of these and other related measures, see [109].

2.1.4 Measures of entanglement for multipartite states

For multiple qubit states, one has to distinguish between states that have all subsystems entangled and those in which only certain subsets of qubits are entangled. To this end, one defines the notion of *biseparability* as a property of n -qubit states for which there exists a partition of the qubits into two disjoint subsets A and B , such that $|A| + |B| = n$ and the original state is a product state with respect to the partition $A|B$.

However, the notion of biseparability (or multi-separability) is only a crude measure of entanglement. To see that, consider the following two tripartite states, known as the Greenberger-Horne-Zeilinger (GHZ) state and the W-state:

$$\begin{aligned} |\text{GHZ}_{\pm}\rangle &= \frac{1}{\sqrt{2}} (|000\rangle \pm |111\rangle) \\ |\text{W}\rangle &= \frac{1}{\sqrt{3}} (|001\rangle + |010\rangle + |100\rangle) \end{aligned}$$

Clearly, both states have all three qubits entangled (non-biseparable). The essential difference between these states can be seen if a measurement is performed on one of the three qubits;

after measurement, the state is separated in the case of the GHZ-state (it is either $|000\rangle$ or $|111\rangle$), while the W-state remains entangled. Put differently, if one of the three qubits is lost (traced out), the reduced 2-qubit state is separated for the GHZ-state, while for the W-state it is entangled. One says that the entanglement properties of the W-state are robust with respect to particle loss, and fragile for the GHZ state.

It can be shown [64] that the W- and the GHZ-states cannot be transformed into one another by any protocol involving local quantum operations (LO), i.e. transformations which factor out as tensor products of local operators on each qubit and classical communication systems (CC) between the three parties (LOCC). Moreover, Dür, Vidal and Cirac showed in [64] that any non-biseparable three-qubit state can be transformed into either the W- or the GHZ-state.

2.1.5 Non-locality, Bell's theorem and hidden variables

One could, as Bell did, ask the following question: if qubits were replaced by local classical variables (i.e. classical states with definite values for each qubit at every moment of time), would it then be possible to reproduce the outcome of any quantum computation process by replacing each quantum gate with classical operations which are randomly chosen from a set of possible transformations?

The idea of using classical variables instead of qubits is in accord with the principle of *local realism* (in the sense of Einstein, [65]), according to which any two objects A and B that are far apart in space must exhibit relative independence, such that any external influence on A has no direct (i.e. instantaneous or faster-than-light) influence on B (*Principle of Local Action*) and moreover, material objects have properties independent of any observation, and the results of any possible measurement depend on these properties (*realism*).

The proposal to describe the measurement process by randomly chosen classical operations, instead of quantum gates, is linked to the class of so-called *local hidden variable theories*. Such theories attempt to obtain the non-classical features of Quantum Mechanics as emergent phenomena of a Local Realist theory. This is based on the conjecture that quantum mechanics, as a theory which violates the principle of local realism, cannot be a complete theory. As such one needs to supplement quantum mechanics with additional variables that determine the results of individual measurements, in order to restore causality and locality. The answer to the above question is provided by Bell's theorem:

Theorem 2.1.4. (Bell) *There is no local probabilistic algorithm, or equivalently, no local hidden*

variable theory that can reproduce the conclusions of quantum mechanics.

Bell's theorem expresses the fact that if one considers correlations between ideal measurements on entangled states, the predictions of quantum mechanics violate certain inequalities which are necessary conditions for local realism. Thus Bell's theorem imposes an incompatible alternative between any local hidden variable theory and quantum entanglement as described in quantum mechanics.

Bell's theorem relies on the notion of *non-local correlations*. The existence of non-local correlations makes it impossible to decipher a generic (entangled) multiple-qubit state by dividing the system into parts and studying each part separately. Let us consider the following set-up, in which Alice and Bob have (each of them) access to one qubit of an entangled 2-qubit state. Suppose Alice can perform measurements a and a' and Bob can perform b and b' and assume that the possible outcomes for any of these measurements are 0 and 1. Then, assuming the existence of local hidden variables s , the probability that Alice and Bob obtain the outcomes which we denote by α and β , when measuring the observables a and b , respectively, is given by:

$$P(a, \alpha; b, \beta) = \int \rho(s) P_1(s; a, \alpha) P_2(s; b, \beta) ds, \quad \alpha, \beta \in \{0, 1\} \quad (2.1.3)$$

Here s represents one or several parameters which contain all the relevant information about the past interaction between the two qubits, $\rho(s) \geq 0$ is a probability density normalised to unity and $0 \leq P_1(s; a, \alpha), P_2(s; b, \beta) \leq 1$. The value of $P_1(s; a, \alpha)$ is assumed to be independent of the measurement performed on the second qubit (locality). Following [34, 35], one can show that:

$$-1 \leq P(a, \alpha; b, \beta) - P(a, \alpha; b', \beta') + P(a', \alpha'; b, \beta) + P(a', \alpha'; b', \beta') - P(a', \alpha') - P(b, \beta) \leq 0$$

where $\alpha', \beta' \in \{0, 1\}$ and $P(a, \alpha) = P(a, \alpha; b, 0) + P(a, \alpha; b, 1) = P(a, \alpha; b', 0) + P(a, \alpha; b', 1)$. The above relation is known as the CH74 inequality, obtained by Clauser and Horne in 1974 and we will see in the next section that it is violated by the predictions of quantum mechanics.

2.1.6 Measurement contexts

Suppose a 2-qubit system (or rather, a large collection of identically prepared 2-qubit systems) is prepared in one of the four Bell states, say $|\Phi^+\rangle$. Assume Alice measures in the basis $\{|0\rangle_A, |1\rangle_A\}$ with possible outcomes 0 and 1, and similarly for Bob. Then both Alice and Bob measure 0 and 1 with equal frequency. In fact, they would obtain the same probabilities for measurements

done in any arbitrary orthonormal basis, e.g. of the type:

$$\{|\alpha_0\rangle = \cos \alpha |0\rangle + \sin \alpha |1\rangle, |\alpha_1 = \alpha_0 + \pi/2\rangle = -\sin \alpha |0\rangle + \cos \alpha |1\rangle\}$$

To be more precise, when Alice measures in the basis $\{|\alpha\rangle, |\alpha + \pi/2\rangle\}$, she obtains the outcomes 0 and 1 with equal probabilities independently of Bob's choice of measurement $\{|\beta\rangle, |\beta + \pi/2\rangle\}$. For our set-up, the statement can be easily verified by computing the required probabilities with the formula given by Equation 2.3.1 in Section 2.3.2. In general, the statement that Alice's outcome cannot be influenced by Bob's choice of measurement is known as the *no-signalling condition*.

If Alice and Bob communicate to each other only the frequencies with which they obtain their different outcomes, they have no way of distinguishing between the four Bell states. In order to discern, they have to correlate their measurements, and communicate the frequencies with which the correlated outcomes are obtained, as summarised in the table below:

(0, 0)	(0, 1)	(1, 0)	(1, 1)
1/2	0	0	1/2

With this information, they can decide that the original state can only be $|\Phi^+\rangle$ or $|\Phi^-\rangle$. In order to further distinguish between these, Alice and Bob have to measure other observables. For example, Alice and Bob could perform the following measurements:

Alice / Bob	Eigenbasis	Outcomes
a	$ 0\rangle, 1\rangle$	0, 1
a'	$\cos \frac{\pi}{6} 0\rangle + \sin \frac{\pi}{6} 1\rangle, -\sin \frac{\pi}{6} 0\rangle + \cos \frac{\pi}{6} 1\rangle$	0, 1
b	$ 0\rangle, 1\rangle$	0, 1
b'	$\cos \frac{\pi}{6} 0\rangle - \sin \frac{\pi}{6} 1\rangle, \sin \frac{\pi}{6} 0\rangle + \cos \frac{\pi}{6} 1\rangle$	0, 1

In this scenario, Alice can choose between measurements a and a' and Bob between b and b' . A particular choice will be called, from now on, a *measurement context*. Thus there are four possible measurement contexts, $\{a, b\}, \{a, b'\}, \{a', b\}, \{a', b'\}$.

By elementary operations, one can show that, for $|\alpha\rangle = \cos \alpha |0\rangle + \sin \alpha |1\rangle$ and $|\beta\rangle =$

$\cos \beta |0\rangle + \sin \beta |1\rangle$, the following expressions hold:

$$\begin{aligned} |(\langle \alpha | \otimes \langle \beta |) | \Phi^+ \rangle|^2 &= \frac{1}{2} \cos^2(\alpha - \beta) \\ |(\langle \alpha | \otimes \langle \beta |) | \Phi^- \rangle|^2 &= \frac{1}{2} \cos^2(\alpha + \beta) \end{aligned}$$

Thus, if the entangled state corresponds to Φ^+ , Alice and Bob would report the following set of correlated probabilities:

A	B	(0,0)	(0,1)	(1,0)	(1,1)
a	b	1/2	0	0	1/2
a	b'	3/8	1/8	1/8	3/8
a'	b	3/8	1/8	1/8	3/8
a'	b'	1/8	3/8	3/8	1/8

while if the state was Φ^- , they would obtain:

A	B	(0,0)	(0,1)	(1,0)	(1,1)
a	b	1/2	0	0	1/2
a	b'	3/8	1/8	1/8	3/8
a'	b	3/8	1/8	1/8	3/8
a'	b'	1/2	0	0	1/2

Now let us consider the CH74 inequality for the Φ^+ state:

$$P(a, \alpha; b, \beta) - P(a, \alpha; b', \beta') + P(a', \alpha'; b, \beta) + P(a', \alpha'; b', \beta') \leq P(a', \alpha') + P(b, \beta)$$

In our case, the right hand side is always $1/2 + 1/2$. In order to check the inequality, we need to take one entry from each row of the above table, in total 16 choices. Among these, some violate the CH74 inequality, e.g.:

$$P(a, 0; b, 0) - P(a, 0; b', 1) + P(a', 0; b, 0) + P(a', 0; b', 1) = \frac{1}{2} - \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{9}{8} \not\leq 1$$

2.1.7 Contextuality and the Kochen-Specker theorem

We will refer to a *measurement context* as a set of compatible measurements, i.e. measurements that can be jointly performed in a certain experimental setup. In quantum mechanics, contexts correspond to commutative algebras of self-adjoint operators, which can be simultaneously diagonalised. This is, indeed, the use of contextuality that will be employed in the next chapters on the topos approach to quantum physics. The notion of contextuality discussed in the following section is more general, as it does not rely on the standard operator-algebra formulation of quantum mechanics. As such, the results obtained in this picture will be independent of the particular way in which quantum mechanics has been traditionally formulated and interpreted.

The notion of contextuality arises naturally from the question: is it possible to assign definite values to all (hidden) variables at any given moment of time, independently of the device used to measure them? The Kochen-Specker theorem shows that this is not the case, and for quantum systems with more than two levels, the value of an observable depends on which commuting set of observables is being measured along with it.

The Kochen-Specker theorem asserts that quantum mechanics forbids the simultaneous existence of definite values for observables which cannot be measured together, that is it forbids non-contextual hidden variables.

2.2 Non-locality and Contextuality in Sheaf-Theoretic Language

In a series of papers [2, 3, 7, 8], Abramsky et al. formulated a mathematical description of contextuality phenomena in terms of sheaves over a poset of contexts. Their work was partly inspired by [28], in which Butterfield and Isham realised that the Kochen-Specker theorem can be elegantly reformulated in terms of the non-existence of global sections of a certain presheaf.

While the spectral presheaf used by Butterfield and Isham was based on an operator algebra and, in this sense, it relied on concepts specific to quantum mechanics, the work of Abramsky and Brandenburger was carried out at a much higher level of generality, without using any of the characteristic mathematical structures of quantum mechanics. It moreover introduced many new key structures in order to be able to capture and analyse probability tables such as those used in Section 2.1.6 in the discussion on non-locality and contextuality. We shall describe some of these key structures in the rest of this section.

Let X be a set of measurements and O the set of possible outcomes for each measurement. For the present discussion, it suffices to consider the same O for all measurements, although one

could in general allow for different outcomes for each individual measurement.

Suppose now that a subset $U \subset X$ of measurements are simultaneously performed. The outcomes obtained in this particular context are specified by a *section*: $s : U \rightarrow O$, which associates to a measurement $m \in U$ an outcome $s(m)$. The space of sections over U is denoted by $\mathcal{E}(U)$ and is a subset of O^U , where O^U denotes the set of all functions from U to O . If one considers $\mathcal{E}(U)$, together with the natural restriction maps

$$\begin{aligned} \text{res}_U^{U'} : \mathcal{E}(U') &\longrightarrow \mathcal{E}(U) \quad \text{for any } U \subseteq U' \\ s &\longmapsto s|_U \end{aligned}$$

one immediately has $\text{res}_U^U = \text{id}_U$ and $\text{res}_U^{U'} \circ \text{res}_{U'}^{U''} = \text{res}_U^{U''}$ for any $U \subseteq U' \subseteq U''$, thus $\mathcal{E}$ is a pre-sheaf, namely the pre-sheaf of sections over the poset $\mathcal{P}(X)$. In other words, $\mathcal{E}$ is a functor $\mathcal{E} : \mathcal{P}(X)^{\text{op}} \rightarrow \mathbf{Set}$.

The construction, $\mathcal{E}$ is in fact a sheaf since local sections can be uniquely glued together. That is, given a cover $\{U_i\}_{i \in I}$ of U and a family of sections $\{s_i \in \mathcal{E}(U_i)\}_{i \in I}$ which are compatible on every intersection, i.e. $s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j}$, there is a unique section $s \in \mathcal{E}(U)$ such that its restriction to U_i is s_i for all $i \in I$. The sheaf condition is trivially satisfied, as one can always glue together partial functions on a discrete space which agree on overlaps by taking the union of their graphs. $\mathcal{E}$ will be called the *sheaf of events*.

2.2.1 Measurement covers, the distribution functor and no-signalling

So far we imposed no restrictions on the poset $\mathcal{P}(X)$, which we think of as a set of labels for different basic measurements. However, in quantum mechanics only certain measurements can be performed jointly, thus it makes sense to introduce the notion of *measurement cover* $\mathcal{M} \subset \mathcal{P}(X)$ of X , composed of measurement contexts only, with the property that the union of all contexts contained in $\mathcal{M}$ equals X . Additionally, we will require that $\mathcal{M}$ contains only maximal contexts (maximal sets of compatible measurements), that is if $C, C' \in \mathcal{M}$ and $C \subseteq C'$ then $C = C'$ (anti-chain condition on $\mathcal{M}$).

Example 2.2.1. *Bell-type scenarios.* In the formulation of Bell-type theorems on non-locality, one refers to composed systems, whose parts may be space-like separated, in a fashion similar to our discussion in Section 2.1.6. The class of measurement covers used in Bell-type scenarios and scenarios involving other non-local devices such as PR-boxes, can be described as follows. Let I denote a set of indices labelling the different parts of a composed system and for each $i \in I$ let X_i be the set of basic measurements that can be performed on the part labelled by i . Then X

is constructed as the disjoint union of the family $\{X_i\}_{i \in I}$. The measurement cover $\mathcal{M}$ is defined as the set of contexts containing exactly one measurement from each part, i.e. we consider as compatible any two measurements performed on different parts of the system, but we do not allow for compatible measurements on the same part.

We shall refer to a Bell-type scenario which involves n parts, each of which has k possible choices of measurement, each choice with l possible outcomes, as being of (n, k, l) -type. Note that for a system of (n, k, l) -type, there are k^n measurement contexts, for each of which there are l^n possible assignments of outcomes. Thus there are $(kl)^n$ sections over the contexts. The set of all measurements is of size kn , and there are l^{kn} global assignments.

The last ingredient that we need refers to the probabilistic behaviour of quantum systems in their interaction with classical measuring apparatus. Standard probability distributions are represented by non-negative reals between 0 and 1, for which the usual addition and multiplication rules apply. Put differently, standard probability distributions are valued in the semiring $(\mathbb{R}_{\geq 0}, +, 0, \cdot, 1)$, where multiplication distributes over addition and 0 and 1 denote the units in the commutative monoids $(\mathbb{R}_{\geq 0}, +, 0)$ and $(\mathbb{R}_{\geq 0}, \cdot, 1)$, respectively. If one allows for negative probabilities, the semiring $\mathbb{R}_{\geq 0}$ has to be extended to the reals $\mathbb{R}$. Interestingly, the results of Bell's theorem do not hold if one allows for negative probabilities. A third type of semiring of interest to us is the boolean semiring, $\mathbb{B} = (\{0, 1\}, \vee, 0, \wedge, 1)$.

Fixing a semiring R , we define an R -distribution over a set S as a function $d : S \rightarrow R$ with finite support (i.e. d is non-zero only on a finite subset of S , its support), satisfying the condition

$$\sum_{x \in S} d(x) = 1$$

We denote by $\mathcal{D}_R(S)$ the set of R -distributions over S . In the case of the semiring $\mathbb{R}_{\geq 0}$, this is the set of probability distributions with finite support on S . In the case of the booleans $\mathbb{B}$, it is the set of non-empty finite subsets of S . Furthermore, given a function $f : S \rightarrow T$, we define:

$$\begin{aligned} \mathcal{D}_R(f) : \mathcal{D}_R(S) &\longrightarrow \mathcal{D}_R(T) \\ d &\longmapsto \left(t \mapsto \sum_{f(s)=t} d(s) \right) \end{aligned}$$

It can be easily seen that $\mathcal{D}_R$ is functorial, $\mathcal{D}_R(g \circ f) = \mathcal{D}_R(g) \circ \mathcal{D}_R(f)$ and $\mathcal{D}_R(\text{id}_S) = \text{id}_{\mathcal{D}_R(S)}$. Thus we have defined a functor $\mathcal{D}_R : \mathbf{Set} \rightarrow \mathbf{Set}$.

By composing the distribution functor with the event sheaf $\mathcal{E} : \mathcal{P}(X)^{\text{op}} \rightarrow \mathbf{Set}$, we obtain a presheaf $\mathcal{D}_R \mathcal{E} : \mathcal{P}(X)^{\text{op}} \rightarrow \mathbf{Set}$. This assigns to a set of measurements U the set $\mathcal{D}_R(\mathcal{E}(U))$

of distributions on sections over U . The above arrow function, applied to the restriction maps $\text{res}_U^{U'}$, with $U \subseteq U'$, gives:

$$\begin{aligned} \mathcal{D}_R\mathcal{E}(U') &\longrightarrow \mathcal{D}_R\mathcal{E}(U) \\ d &\longmapsto d|_U \end{aligned}$$

where for each $s \in \mathcal{E}(U)$,

$$d|_U := \sum_{s' \in \mathcal{E}(U'); s'|_U = s} d(s') \quad (2.2.1)$$

Thus $d|_U$ is the *marginal* of the distribution d , in the sense that it associates to each section s in the smaller context U the sum of the weights of all sections s' in the larger context which restrict to s .

Example 2.2.2. *Let us consider a $(2, 2, 2)$ Bell-type scenario, in which Alice can perform measurements a and a' and Bob can perform b and b' . There are two possible outcomes, 0 and 1, for each measurement. The relevant measurement cover $\mathcal{M}$ consists of four maximal contexts:*

$$\mathcal{M} = \{\{a, b\}, \{a', b\}, \{a, b'\}, \{a', b'\}\}$$

Over each maximal context $C \in \mathcal{M}$, there are four sections, e.g. for $C = \{a, b\}$ we have:

$$\mathcal{E}(\{a, b\}) = \{\{a \rightarrow 0, b \rightarrow 0\}, \{a \rightarrow 0, b \rightarrow 1\}, \{a \rightarrow 1, b \rightarrow 0\}, \{a \rightarrow 1, b \rightarrow 1\}\}$$

Further, assume that we can associate a distribution $e_C \in \mathcal{D}_R\mathcal{E}(C)$ to each context $C \in \mathcal{M}$. This gives a number $e_C(s)$ for each section $s \in \mathcal{E}(C)$. The distributions e_C form the rows of the following table:

A	B	$(0, 0)$	$(0, 1)$	$(1, 0)$	$(1, 1)$
a	b	p_1	p_2	p_3	p_4
a	b'	p_5	p_6	p_7	p_8
a'	b	p_9	p_{10}	p_{11}	p_{12}
a'	b'	p_{13}	p_{14}	p_{15}	p_{16}

The data given by $\mathcal{M}$, the sheaf of events E and a family of distributions $\{e_C\}_{C \in \mathcal{M}}$, will be called an *empirical model*. In the standard case of probability tables, the numbers p_i are non-negative integers, and the values along each row sum up to 1; hence the distributions e_C are

probability distributions. One advantage of using empirical models is that they are formulated in an elegant, robust and general mathematical language which does not depend in particular on the Hilbert space formalism of quantum mechanics.

Example 2.2.3. *One could also consider the weights p_i to take values in other semirings, for example the boolean semiring. The following model is a possibilistic version of a non-local Hardy model [83], which can be viewed as specifying the support of a standard probabilistic Hardy model.*

A	B	$(0, 0)$	$(0, 1)$	$(1, 0)$	$(1, 1)$
a	b	1	1	1	1
a	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0

The *no-signalling condition* (the statement that the outcome obtained by Alice cannot be influenced by Bob's choice of a measurement) can be elegantly formulated in terms of the presheaf $\mathcal{D}_R\mathcal{E}$. We define a no-signalling empirical model for a measurement cover $\mathcal{M}$ to be an empirical model for which the family of distributions $\{e_C\}_{C \in \mathcal{M}}$ is compatible in the sense of the sheaf condition: for any $C, C' \in \mathcal{M}$,

$$e_C|_{C \cap C'} = e_{C'}|_{C \cap C'} \quad (2.2.2)$$

where the restrictions $e_C|_{C \cap C'}$ and $e_{C'}|_{C \cap C'}$ are defined as in Eq. 2.2.1.

Example 2.2.4. *For our previous example, involving two parties, Alice and Bob, consider the contexts $C = \{m_a, m_b\}$ and $C' = \{m_a, m'_b\}$. Fixing a section $s_0 \in \mathcal{E}(\{m_a\})$, e.g. $s_0 = \{m_a \rightarrow 0\}$, the no-signalling condition can be expressed as*

$$\sum_{\substack{s \in \mathcal{E}(C) \\ s|_{m_a} = s_0}} e_C(s) = \sum_{\substack{s' \in \mathcal{E}(C') \\ s'|_{m_a} = s_0}} e_{C'}(s')$$

For the above table, with $m_a = a, m_b = b, m'_b = b'$ and $s_0 = \{a \rightarrow 0\}$, we obtain the condition:

$$p_1 + p_2 = p_5 + p_6$$

2.2.2 Global Sections

Let us recapitulate the mathematical structure introduced above. We started with a set of measurements X and we defined the measurement sheaf of events $\mathcal{E}$, which assigns to each $U \subseteq X$ the set of sections $\{s : U \rightarrow O\}$. A particular section over U specifies a particular set of outcomes, one outcome for each measurement in U . (Remember we assumed that all measurements have the same set of possible outcomes O). Further, we defined the presheaf of R -valued distributions, $\mathcal{D}_R\mathcal{E}$, which assigns to each $U \subseteq X$ the set $\mathcal{D}_R(\mathcal{E}(U))$ of distributions on $\mathcal{E}(U)$, which is the set of sections over U . The distribution presheaf satisfied the important property 2.2.1.

Further, we had to deal with the factual requirement that, in general, not all measurements can be performed together. This introduces the notion of contexts and we formed the set $\mathcal{M}$ of maximal contexts (i.e. maximal sets of compatible measurements). Since $\mathcal{M} \subset \mathcal{P}(X)$, passing to the measurement cover does not affect the definitions and properties of the event sheaf $\mathcal{E}$ and the distribution presheaf $\mathcal{D}_R\mathcal{E}$ on X . Each context $C \in \mathcal{M}$ comes with a set of sections O^C . An empirical model is specified by providing, for each $C \in \mathcal{M}$, a probability distribution $e_C \in \mathcal{D}_R\mathcal{E}(C)$ over $\mathcal{E}(C) = O^C$. These probabilities may or may not be compatible, in the sense of Eq. 2.2.1, with a probability distribution over $\mathcal{E}(X) = O^X$. If such a ‘global’ probability distribution exists, the model corresponds to a *deterministic hidden variable* scenario, in the sense that hidden variables determine which section $s \in \mathcal{E}(X)$ is being chosen in any certain situation and context². The sheaf condition implies that the assignment of outcomes is independent of the measurement context, thus the existence of global probability distributions in the above sense is a form of *non-contextuality*.

For no-signalling empirical models, the compatibility condition 2.2.2 satisfied by the family of distributions $\{e_C\}_{C \in \mathcal{M}}$ implies that a global section $d \in \mathcal{D}_R\mathcal{E}(X)$, compatible with $\{e_C\}$, if it exists, will satisfy the sheaf condition with respect to $\mathcal{M}$.

The existence of a global section $d \in \mathcal{D}_R\mathcal{E}(X)$ which restricts to yield the probabilities specified by the empirical model on each context $C \in \mathcal{M}$, i.e. $d|_C = e_C$, has very important

²The unfamiliar reader might find helpful the following remarks: we have in mind the preparation of a state in the same conditions over and over again. If one had access to the complete set of variables that characterise this state, he should obtain the same set of outcomes each time he performs a ‘complete’ measurement, as the state was prepared in the same conditions. The fact that, in practice (i.e. in quantum mechanics), this does not happen, is explained in theories invoking hidden variables, by the existence of unaccessible variables which can alter the values of the accessible variables.

implications. Thus given a section $t \in \mathcal{E}(C)$,

$$e_C(t) = d|_C(t) = \sum_{\substack{s \in \mathcal{E}(X) \\ s|_C = t}} d(s) = \sum_{s \in \mathcal{E}(X)} \delta_s|_C(t) \cdot d(s) \quad (2.2.3)$$

where, for a global section $s \in \mathcal{E}(X)$, $\delta_s \in \mathcal{D}_R \mathcal{E}(X)$ is the distribution defined globally by $\delta_s(s) = 1$ and $\delta_s(s') = 0$ for $s \neq s'$. Hence, the presence of a global section $d \in \mathcal{D}_R \mathcal{E}(X)$, such that $d|_C = e_C$ turns out to be equivalent with the statement that the empirically observed probabilities $e_C(t)$ can be obtained by *averaging* over the hidden variables with respect to the distribution d . Further, from the definition of the globally induced distributions δ_s , it follows that

$$\delta_s|_C(t) = \prod_{x \in C} \delta_{s|_{\{x\}}}(t|_{\{x\}}) \quad (2.2.4)$$

which means that the probability distribution induced by s *factorizes* as a product of probabilities assigned to individual measurements, independent of the context in which they appear. Note that this is precisely the notion of *locality* employed in Bell-type scenarios, e.g. in Equation 2.1.3 we assumed that the probability $P(s; a, \alpha; b, \beta)$ of the joint outcome $\{\alpha \mapsto a, \beta \mapsto b\}$ determined by the hidden variable s is the product of the probabilities that s determines for the outcomes $\{\alpha \mapsto a\}$ and $\{\beta \mapsto b\}$, i.e. $P(s; a, \alpha; b, \beta) = P_1(s; a, \alpha) P_2(s; b, \beta)$ leading to:

$$P(a, \alpha; b, \beta) = \int \rho(s) P_1(s; a, \alpha) P_2(s; b, \beta) ds$$

in agreement with Equation 2.2.3.

We summarise the above discussion through the following:

Proposition 2.2.5. *The existence of a global section for an empirical model implies the existence of a local (non-contextual) deterministic hidden-variable model which realises it.*

It is interesting to note that the converse of the above proposition also holds, as Abramsky and Brandenburger have shown in Theorem 8.1 [2].

2.2.3 Existence of Global Sections

The above discussion leads to the following problem: *given an empirical model, determine it if has a global section.* Let us refine the expression of this problem. An empirical model consists of a measurement cover $\mathcal{M}$ (which covers a set of measurements X), together with the choice of a (local) section of the distribution presheaf $e_C \in \mathcal{D}_R \mathcal{E}(C)$, one over each context $C \in M$ – more precisely, one distribution e_C over each set of local sections (events) $\mathcal{E}(C)$. In addition, for no-

signalling empirical models the local distributions e_C satisfy the compatibility condition 2.2.2. What we mean by a global section for an empirical model $(\mathcal{M}, \{e_C\})$ is a global section of the presheaf $\mathcal{D}_R\mathcal{E}$, whose restriction to each context $C \in \mathcal{M}$ agrees with the local sections e_C . Of course, global sections for the presheaf $\mathcal{D}_R\mathcal{E}$ always exist, since $\mathcal{D}_R\mathcal{E}(X)$ represents the set of distributions on the set $\mathcal{E}(X) = O^X$.

We shall give a general linear-algebraic method of finding global sections for a given empirical model $(\mathcal{M}, \{e_C\})$. The main step involves constructing a *incidence matrix* for the empirical model. This matrix is defined using only $\mathcal{M}$ and the event sheaf $\mathcal{E}$. It can be applied to any empirical model $(\mathcal{M}, \{e_C\})$ with respect to any distribution functor $\mathcal{D}_R$.

To define the incidence matrix, we first form the disjoint union $\coprod_{C \in \mathcal{M}} \mathcal{E}(C)$ of all the sections over the contexts in $\mathcal{M}$, and specify an enumeration $t_1, \dots, t_p$ of this set. We also specify an enumeration $s_1, \dots, s_q$ of all the global sections of the sheaf $\mathcal{E}$. We then form the $(p \times q)$ -matrix $\mathbf{M}$, with entries defined as follows:

$$\mathbf{M}[i, j] = \begin{cases} 1, & s_j|_C = t_i, \text{ where } t_i \in \mathcal{E}(C) \\ 0, & \text{otherwise} \end{cases}$$

It acts by matrix multiplication on distributions in $\mathcal{D}_R\mathcal{E}(X)$, seen as row vectors: $d \mapsto (d|_C)_{C \in \mathcal{M}}$. The result of all multiplications of this form will be the set of families $\{e_C\}_{C \in \mathcal{M}}$ which arise from global sections.

A given empirical model assigns an element in the semiring R to each section $t_i \in \coprod_{C \in \mathcal{M}} \mathcal{E}(C)$. Thus it can be specified by a vector $\mathbf{V}$ of length p , where $\mathbf{V}[i] = e_C(t_i)$. We can also introduce a vector $\mathbf{X}$ of length q of ‘unknowns’, one for each global section $s_j \in \mathcal{E}(X)$. Now a solution for the linear system $\mathbf{M}\mathbf{X} = \mathbf{V}$ will be a vector of values in R , one for each s_j .

In the case of Bell-type scenarios of (n, k, l) -type, such solutions are automatically distributions - due to the regular structure of the incidence matrix for these cases - and they are in bijective correspondence with global sections of the empirical model. For more general scenarios, one can simply augment $\mathbf{M}$ with an extra row, every entry in which is 1, and similarly to augment $\mathbf{V}$ with an extra element, also equal to 1. A solution for this augmented system enforces the constraint

$$\mathbf{X}[1] + \dots + \mathbf{X}[q] = 1$$

and hence ensures that the first q entries of $\mathbf{X}$ define a distribution on $\mathcal{E}(X)$.

Example 2.2.6. Consider a general $(2, 2, 2)$ Bell-type scenario:

A	B	$(0, 0)$	$(0, 1)$	$(1, 0)$	$(1, 1)$
a	b	p_1	p_2	p_3	p_4
a	b'	p_5	p_6	p_7	p_8
a'	b	p_9	p_{10}	p_{11}	p_{12}
a'	b'	p_{13}	p_{14}	p_{15}	p_{16}

Here $X = a, b, a', b'$ and the measurement cover is $\mathcal{M} = \{\{a, b\}, \{a', b\}, \{a, b'\}, \{a', b'\}\}$. Assume $O = \{0, 1\}$ is the set of possible outcomes for all observables. Assume also that $R = \mathbb{R}_{\geq 0}$, i.e. the numbers $p_1, \dots, p_{16}$ are probabilities in the usual sense. Note that any measurement performed at part A is compatible with any measurement performed by part B . The event sheaf $\mathcal{E}$ has 16 global sections corresponding to:

$$\begin{aligned}
(a, b, a', b') \in & \{(0, 0, 0, 0), (0, 0, 0, 1), (0, 0, 1, 0), (0, 0, 1, 1) \\
& (0, 1, 0, 0), (0, 1, 0, 1), (0, 1, 1, 0), (0, 1, 1, 1) \\
& (1, 0, 0, 0), (1, 0, 0, 1), (1, 0, 1, 0), (1, 0, 1, 1) \\
& (1, 1, 0, 0), (1, 1, 0, 1), (1, 1, 1, 0), (1, 1, 1, 1)\}
\end{aligned} \tag{2.2.5}$$

Suppose now there is a probability distribution d over $\mathcal{E}(X)$ which associates probabilities $q_1, \dots, q_{16}$ to the above global sections in the given order. Then one must have:

$$\begin{aligned}
p_1 &= q_1 + q_2 + q_3 + q_4 & p_2 &= q_5 + q_6 + q_7 + q_8 \\
p_3 &= q_9 + q_{10} + q_{11} + q_{12} & p_4 &= q_{13} + q_{14} + q_{15} + q_{16} \\
p_5 &= q_1 + q_3 + q_5 + q_7 & p_6 &= q_2 + q_4 + q_6 + q_8 \\
p_7 &= q_9 + q_{11} + q_{13} + q_{15} & p_8 &= q_{10} + q_{12} + q_{14} + q_{16} \\
p_9 &= q_1 + q_2 + q_9 + q_{10} & p_{10} &= q_3 + q_4 + q_{11} + q_{12} \\
p_{11} &= q_5 + q_6 + q_{13} + q_{14} & p_{12} &= q_7 + q_8 + q_{15} + q_{16} \\
p_{13} &= q_1 + q_5 + q_9 + q_{13} & p_{14} &= q_2 + q_6 + q_{10} + q_{14} \\
p_{15} &= q_3 + q_7 + q_{11} + q_{15} & p_{16} &= q_4 + q_8 + q_{12} + q_{16}
\end{aligned} \tag{2.2.6}$$

and the (16×16) incidence matrix is

$$\mathbf{M} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.2.7)$$

One can solve the linear system $\mathbf{M}\mathbf{X} = \mathbf{V}$ and obtain $\mathbf{X} = [q_1, \dots, q_{16}]$ in terms of the given probabilities $\mathbf{V} = [p_1, \dots, p_{16}]$ and require that $\mathbf{X} \geq \mathbf{0}$. If this system has a non-negative solution, then the model admits a realisation in terms of local deterministic hidden variables and is non-contextual, otherwise, the model is said to be weakly contextual.

So far we have assumed that the distributions in $\mathcal{D}_R(\mathcal{E}(C))$ are real valued. We made this assumption explicitly or implicitly, by talking about probability distributions. It is both interesting and important to discuss models involving $\mathbb{B}$ -valued distributions (where $\mathbb{B}$ is the boolean semiring). In fact, any *probabilistic model* (i.e. a model involving $\mathbb{R}_{\geq 0}$ -distributions) can be turned into a *possibilistic model* (i.e. a model involving $\mathbb{B}$ -distributions) by replacing all non-zero probabilities with “1” (True) and all null probabilities with “0” (False). Many quantum information-theoretic results, such as Bell’s theorem, generalise to possibilistic models (see [7]). These considerations motivate the hierarchy of contextuality discussed below.

2.3 A Hierarchy of Contextuality

A probabilistic model $(X, \mathcal{M}, \{e_C\}, \mathbb{R}_{\geq 0})$ is *non-contextual* if there exists a global probability distribution $d \in \mathcal{D}_{\mathbb{R}_{\geq 0}}\mathcal{E}(X)$ compatible with the local probability distributions $\{e_C\}$. The

compatibility condition can be expressed by the equivalence between $(X, \mathcal{M}, \{e_C\}, \mathbb{R}_{\geq 0})$ and $(X, \mathcal{M}, \{d|_C\}, \mathbb{R}_{\geq 0})$. If no such global distributions exist, the model is *weakly contextual*.

The probabilistic model $(X, \mathcal{M}, \{e_C\}, \mathbb{R}_{\geq 0})$ induces a possibilistic model $(X, \mathcal{M}, \{e_C^{\mathbb{B}}\}, \mathbb{B})$ with boolean-valued distributions. The boolean-valued distributions $\{e_C^{\mathbb{B}}\}$ assign the boolean value “1” to those sections which are in the support of the corresponding probability distributions $\{e_C\}$. If the model $(X, \mathcal{M}, \{e_C^{\mathbb{B}}\}, \mathbb{B})$ admits a global boolean distribution $d \in \mathcal{D}_{\mathbb{B}}\mathcal{E}(X)$ compatible with $\{e_C^{\mathbb{B}}\}$, then $(X, \mathcal{M}, \{e_C\}, \mathbb{R}_{\geq 0})$ is either non-contextual or weakly contextual. Otherwise, it is said to be *logically contextual*.

Remark 2.3.1. *A logically contextual model is automatically weakly contextual, since a compatible global probability distribution $d \in \mathcal{D}_{\mathbb{R}_{\geq 0}}\mathcal{E}(X)$ induces a compatible global boolean distribution $d \in \mathcal{D}_{\mathbb{B}}\mathcal{E}(X)$. The converse, however, is not true: there are weakly contextual models which are not logically contextual. In general, we say that an empirical model is **probabilistically non-extendable** if it has no global section over $\mathcal{D}_{\mathbb{R}_{\geq 0}}$, and **possibilistically non-extendable** if it has no global section over $\mathcal{D}_{\mathbb{B}}$.*

Working over the boolean ring comes with a number of simplifications. For instance, the set of distributions $\mathcal{D}_{\mathbb{B}}\mathcal{E}(C)$ can be identified with $\mathcal{P}(\mathcal{E}(C))$ for any subset $C \subseteq X$. Indeed, a boolean distribution $d \in \mathcal{D}_{\mathbb{B}}\mathcal{E}(X)$ is a way of distinguishing between ‘possible’ sets of outcomes and ‘impossible’ sets of outcomes. Moreover, the boolean distribution d induces a possibilistic model $(X, \mathcal{M}, \{d|_C\}, \mathbb{B})$. If this model reproduces $(X, \mathcal{M}, \{e_C^{\mathbb{B}}\}, \mathbb{B})$, or, equivalently, the support of the model $(X, \mathcal{M}, \{e_C\}, \mathbb{R}_{\geq 0})$, the model is not logically contextual.

There is a further refinement to this hierarchy of contextuality. Let $\{s_1, \dots, s_n\} = \mathcal{E}(X)$ be the set of global sections of the events sheaf. The global sections of the presheaf $\mathcal{D}_{\mathbb{B}}\mathcal{E}$ are boolean distributions over $\mathcal{E}(X)$. Let $d_1, \dots, d_n$ be the global distributions with minimal support $\{s_1\}, \dots, \{s_n\}$, respectively. These distributions induce possibilistic models $(X, \mathcal{M}, \{d_1|_C\}, \mathbb{B})$, $\dots$, $(X, \mathcal{M}, \{d_n|_C\}, \mathbb{B})$. If all these models are such that their support falls outside of the support of the original model $(X, \mathcal{M}, \{e_C\}, \mathbb{B})$, then any other distribution with a larger support $\{s_{i_1}, s_{i_2}, \dots\}$ will produce a possibilistic model whose support will also fall outside of the support of the original model $(X, \mathcal{M}, \{e_C\}, \mathbb{B})$. In this case, the model $(X, \mathcal{M}, \{e_C\}, \mathbb{B})$ is said to be *strongly contextual*. In particular, the model is also logically contextual. However, there are logically contextual models which are not strongly contextual.

In the case of dichotomic measurements (i.e. measurements with only two outcomes) we can make a connection to logic by interpreting one outcome as *true* and the other outcome as *false*. This allows us to think of the set of measurements X as a set of boolean variables. If C is

a finite context, any subset of $\mathbf{2}^C$ can be defined by a propositional formula. For example, each joint outcome $s : C \rightarrow \mathbf{2}$ determines a propositional formula

$$\varphi_s = \bigwedge_{x \in C, s(m)=True} m \wedge \bigwedge_{x \in C, s(m)=False} \neg m$$

The only satisfying assignment of φ_s in $\mathbf{2}^C$ is s .

The propositional formula whose set of satisfying assignments is the support of C is

$$\varphi_C := \bigvee_{s \in S(C)} \varphi_s.$$

An empirical model is *logically contextual* if there exists some $C \in \mathcal{C}$ and some $s \in S(C)$ such that the formula

$$\Phi = \varphi_s \wedge \bigwedge_{V \in \mathcal{C} \setminus C} \varphi_V$$

is not satisfiable. This says that there is a possible joint outcome s which cannot be accounted for by any valuation on all the variables in X which is consistent with the support of the model. This immediately implies that there is no joint distribution on all the observables which marginalizes to yield the empirically observable probabilities as in Remark 2.3.1. As originally shown in the bipartite case by Fine [69], and in a very general form in [2], the non-existence of a joint distribution is equivalent to the usual definition of non-locality as given by Bell [15].

2.3.1 Strong contextuality as a CSP

A constraint satisfaction problem (CSP) [97, 104] is specified by a triple $(V, K, \mathcal{R})$ where V is a finite set of variables, K is a finite set of values, and $\mathcal{R}$ is a finite set of constraints. A constraint is a pair (C, S) where $C \subseteq V$ and $S \subseteq K^C$. This formulation is equivalent to the more common one where a constraint is specified as a list of k variables together with a set of k -tuples of values. An assignment $s : V \rightarrow K$ satisfies a constraint (C, S) if $s|_C \in S$. A solution of the CSP $(V, K, \mathcal{R})$ is an assignment $s : V \rightarrow K$ which satisfies every constraint in $\mathcal{R}$.

To any empirical model e defined over a cover $\mathcal{M}$, with outcome set O , we can associate the CSP $(X, O, \{\text{supp}(e_C)\}_{C \in \mathcal{M}})$ with e . An empirical model e is maximally contextual if and only if the corresponding CSP has no solution.

In the case of dichotomic measurements (i.e. measurements with only two possible outcomes), the CSP reduces to a boolean satisfiability problem. In this case global sections corre-

spond precisely to satisfying assignments for the formula

$$\varphi_e = \bigwedge_{C \in \mathcal{M}} \varphi_C$$

Hence an empirical model e is strongly contextual if and only if the above formula is unsatisfiable.

2.3.2 Contextuality for Quantum States

The hierarchy of contextuality which has so far been defined for abstract empirical models can be naturally lifted to apply to quantum states. If we fix observables for each party, a n -qubit quantum state gives rise to an empirical model. We shall mainly be concerned with $(n, 2, 2)$ scenarios where each of the n parties has access to one qubit of a quantum state and can choose to perform one out of two available measurements on their qubit, each with two possible outcomes. The two available measurements can be represented by 2×2 self-adjoint unitaries, A and B . The two possible outcomes of a given measurement correspond to the two eigenvalues, $+1$ and -1 of the corresponding unitary. We usually associate the Boolean value *True* to the $+$ eigenvalue and the Boolean value *False* to the $-$ eigenvalue.

A compatible set of measurements is given by a choice of either A or B at each of the n measurement sites. A given quantum state $|\Psi\rangle$ determines a probability distribution on the set of possible outcomes for each compatible set of measurements. Thus if $X_1 X_2 \dots X_n$ is a compatible set of measurements (i.e. $X_i \in \{A, B\}$ for all i) the probability of obtaining the outcome $\sigma_1 \sigma_2 \dots \sigma_n$ where $\sigma_i \in \{+, -\}$ is given by the squared norm of the inner product

$$P(\Psi, \sigma_1 \sigma_2 \dots \sigma_n) = |\langle e_1 | \otimes \langle e_2 | \otimes \dots \otimes \langle e_n | \Psi \rangle|^2 \quad (2.3.1)$$

where e_i is the eigenvector corresponding to the σ_i eigenvalue of the unitary X_i representing the measurement performed by the i^{th} party.

Thus for fixed observables A and B as above, a quantum state gives rise to an empirical model with rows indexed by n -tuples $X_1 X_2 \dots X_n$, $X_i \in \{A, B\}$ and columns indexed by n -tuples $\sigma_1 \sigma_2 \dots \sigma_n$. The section on a given row and column is given by formula (2.3.1) above. The section is in the support of the model if and only if the inner product (2.3.1) is non-zero.

The observables most frequently used in this thesis are the three Pauli matrices:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.3.2)$$

2.3.3 The Bell state

We look again at the $(2, 2, 2)$ -type scenario which can be represented in quantum mechanics by measuring the $|\Phi^+\rangle$ Bell state, as discussed in Section 2.1.6. This scenario is specified by an empirical model summarized in the following table:

A	B	$(0, 0)$	$(0, 1)$	$(1, 0)$	$(1, 1)$
a	b	$1/2$	0	0	$1/2$
a	b'	$3/8$	$1/8$	$1/8$	$3/8$
a'	b	$3/8$	$1/8$	$1/8$	$3/8$
a'	b'	$1/8$	$3/8$	$3/8$	$1/8$

We are interested in finding a probability distribution on the global assignments $\mathcal{E}(X)$. This amounts to solving $\mathbf{M}\mathbf{X} = \mathbf{V}$ over the reals, subject to the constraint $\mathbf{X} \geq \mathbf{0}$, which is in fact a linear programming problem.

However, it is easier in this case to give a direct argument, as in [2], showing that there is no such solution. This implies that the above model has no hidden-variable realization, which proves Bell's theorem [15]. The argument starts by considering 4 out of the 16 equations, corresponding to rows 1, 6, 11 and 13 of the incidence matrix $\mathbf{M}$ given in Equation 2.2.7. For ease, we write X_i instead of $\mathbf{X}[i]$.

$$\begin{aligned}
X_1 + X_2 + X_3 + X_4 &= \mathbf{V}[1] = 1/2 \\
X_2 + X_4 + X_6 + X_8 &= \mathbf{V}[6] = 1/8 \\
X_3 + X_4 + X_{11} + X_{12} &= \mathbf{V}[11] = 1/8 \\
X_1 + X_5 + X_9 + X_{13} &= \mathbf{V}[13] = 1/8
\end{aligned}$$

Adding the last three equations yields

$$X_2 + 2X_4 + X_6 + X_8 + X_3 + X_{11} + X_{12} + X_1 + X_5 + X_9 + X_{13} = 3/8$$

Since all these terms must be non-negative, the left hand side of this equation must be greater than or equal to the left-hand side of the first equation, yielding the required contradiction.

In terms of our hierarchy, it is easy to see that, although this model is weakly contextual,

it is not logically contextual. If we exclude from the set $\mathcal{E}(X)$ those global sections which map a and b to different outcomes, we are left with one example of a set of global sections (or equivalently, with one example of a boolean distribution over the set of global sections) whose restriction to each context is equal to the model's support.

2.3.4 The possibilistic Hardy model

The original purpose of the Hardy model, introduced by Lucien Hardy in [82, 83], was to give a 'logical' proof of Bell's theorem in the bipartite case. Its support is specified by the following table:

A	B	$(0, 0)$	$(0, 1)$	$(1, 0)$	$(1, 1)$
a	b	1	1	1	1
a	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0

It has been shown in [2] that this model is logically contextual, as it admits no solutions over the Boolean semiring. As for the previous example, this can be done by a direct argument. This time the equations specified by the incidence matrix are equations over $\mathbb{B}$ which can also be interpreted as logical clauses. For example, the equation specified by the fifth row of the incidence matrix is

$$X_1 \vee X_3 \vee X_5 \vee X_7 = \mathbf{V}[5] = 0$$

This yields the equivalent logical formula

$$\neg X_1 \wedge \neg X_3 \wedge \neg X_5 \wedge \neg X_7$$

We focus on the formulas corresponding to rows 1, 5, 9 and 16 of the incidence matrix:

$$\begin{aligned} & X_1 \vee X_2 \vee X_3 \vee X_4 \\ & \neg X_1 \wedge \neg X_3 \wedge \neg X_5 \wedge \neg X_7 \\ & \neg X_1 \wedge \neg X_2 \wedge \neg X_9 \wedge \neg X_{10} \\ & \neg X_4 \wedge \neg X_8 \wedge \neg X_{12} \wedge \neg X_{16} \end{aligned}$$

Since every disjunct in the first formula appears as a negated conjunct in one of the other three

formulas, there is no satisfying argument.

According to Remark 2.3.1, the Hardy model satisfies a stronger non-locality property than the Bell model. So far we have proved this directly, but it also follows from the general results in [102], which show that models which rely on Hardy's paradox are complete for the $(2, 2, 2)$ -type cases, and in particular that there must be at least three null sections in the support in order for non-locality to hold, while the Bell model has only two zero entries.

The Hardy model on the other hand is not strongly contextual, as the global assignment $\{a \mapsto 1, b \mapsto 1, a' \mapsto 0, b' \mapsto 0\}$ is one example of a global section whose restriction to each context is compatible with the model's support.

2.3.5 Dicke states

A permutation-symmetric n -qubit state is one which is invariant under the action of the full symmetry group S_n . A natural basis for the permutation-symmetric states is provided by the *Dicke states* [42], which are also physically significant. For each $n \geq 2$, $0 < k < n$ we define:

$$S(n, k) := K \sum_{\text{perm}} |0^k 1^{n-k}\rangle.$$

Here $K = \binom{n}{k}^{-1/2}$ is a normalization constant, and we sum over all products of k 0-kets and $n - k$ 1-kets.

Note that the W state is the $S(3, 2)$ Dicke state in the above notation. For each $n > 2$, and $0 < k < n$, the Dicke state $S(n, k)$ is logically contextual.

We have excluded the cases $k = 0$ and $k = n$, since in these cases $S(n, k) = |0^n\rangle$ or $|1^n\rangle$, and these are obviously product states. We have also excluded the bipartite case, for which $S(2, 1)$ is the Bell state $\frac{|01\rangle + |10\rangle}{\sqrt{2}}$.

If each party is allowed choose to measure one of the Pauli observables X and Z , a Dicke state $S(n, k)$ gives rise to an $(n, 2, 2)$ empirical model. The model is specified by a table with 2^n rows, corresponding to the possible choices of an observable at each site. We shall focus firstly on the $\frac{n(n-1)}{2}$ rows R_{ij} , where X observables are selected at sites i and j , and Z observables at the remaining sites. Let T_{ij} be the support of the model at row R_{ij} .

Now consider any joint outcome t for this row in which there are k outcomes corresponding to the $+$ eigenvalue and $(n - k)$ outcomes corresponding to the $-$ eigenvalue, and the outcome for X^i is different to the outcome for X^j . We claim that t is not in T_{ij} . If we compute the inner product whose squared norm gives the probability for t , we see that there are two terms, of the

form $+1/c$ and $-1/c$ respectively. Thus the probability of t is 0, and it is not in the support. We can express this in logical terms by saying that T_{ij} satisfies the formula

$$\bigwedge_{k \neq i, j, t(k)=+} Z^k \wedge \bigwedge_{k \neq i, j, t(k)=-} \neg Z^k \Rightarrow (X^i \leftrightarrow X^j). \quad (2.3.3)$$

We now consider the row where Z measurements are selected by every party. The support of this row is described by the formula

$$\bigvee_{\pi \in S_n} \left[\bigwedge_{i=1}^k Z^{\pi(i)} \wedge \bigwedge_{j=k+1}^n \neg Z^{\pi(j)} \right]. \quad (2.3.4)$$

This is the logical counterpart of the description of $S(n, k)$ in the Z -basis.

From each disjunct D of (2.3.4) together with the relevant instances of (2.3.3), we can prove that $X^i \leftrightarrow X^j$ for all i, j such that Z^i and Z^j appear with opposite polarity in D . Note that, by the conditions on k and n , both polarities do appear in D . By the transitivity of logical equivalence, it follows that $X^i \leftrightarrow X^j$ can be derived for all i, j . Thus D , together with the formulas (2.3.3), implies the formula

$$\bigwedge_{i,j} X^i \leftrightarrow X^j. \quad (2.3.5)$$

Thus (2.3.4) together with the conjunction of all instances of (2.3.3) implies (2.3.5).

It follows that any global section which satisfies these formulas must restrict to just two joint outcomes in the row where X measurements are selected by every party, namely those with the same outcome at every part.

To complete the argument, it suffices to show that these two outcomes form a proper subset of the support at that row. If we calculate the probability for each of these events, we obtain

$$\left(\frac{\binom{n}{k}}{(\sqrt{2})^n \sqrt{\binom{n}{k}}} \right)^2 = \frac{\binom{n}{k}}{2^n}.$$

Thus we must show that

$$\frac{\binom{n}{k}}{2^n} < \frac{1}{2},$$

or equivalently

$$\binom{n}{k} < 2^{n-1} = \sum_{l=0}^{n-1} \binom{n-1}{l}$$

which follows from Pascal's rule:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Note however that to obtain a strict inequality, we need the assumption that $n > 2$; the argument for the Bell state $S(2, 1)$ fails at exactly this point.

We also note that logical contextuality is preserved by the action of local unitaries $U_1 \otimes \cdots \otimes U_n$. If a state $|\psi\rangle$ is logically contextual with respect to measurement bases

$$\eta_1^+, \eta_1^-, \dots, \eta_n^+, \eta_n^-,$$

then $U_1 \otimes \cdots \otimes U_n |\psi\rangle$ is logically contextual with respect to the measurement bases

$$U_1 \eta_1^+, U_1 \eta_1^-, \dots, U_n \eta_n^+, U_n \eta_n^-.$$

This follows since inner products and hence probabilities are preserved:

$$\begin{aligned} \langle U_1 \eta_1^\pm \otimes \cdots \otimes U_n \eta_n^\pm | (U_1 \otimes \cdots \otimes U_n) |\psi\rangle &= \langle (U_1 \otimes \cdots \otimes U_n) \eta_1^\pm \otimes \cdots \otimes \eta_n^\pm | (U_1 \otimes \cdots \otimes U_n) |\psi\rangle \\ &= \langle (U_1 \otimes \cdots \otimes U_n)^\dagger (U_1 \otimes \cdots \otimes U_n) \eta_1^\pm \otimes \cdots \otimes \eta_n^\pm | \psi \rangle \\ &= \langle \eta_1^\pm \otimes \cdots \otimes \eta_n^\pm | \psi \rangle. \end{aligned}$$

Thus the orbits of the Dicke states under the actions of local unitaries are all logically contextual.

2.3.6 Permutationally symmetric states

In [125] it is shown that all permutation-symmetric states *except* the unitary orbit of the Dicke states admit a Hardy argument, making use of the Majorana representation of permutation-symmetric states. This is easily converted into a proof of logical contextuality. By combining this result with the one presented above, we can conclude that all permutation-symmetric n -partite entangled states, for $n > 2$, are logically contextual.

2.3.7 GHZ models

A GHZ model of type $(n, 2, 2)$ can be specified abstractly as follows. Each part can choose between two measurements labeled as X^i and Y^i , respectively, with $1 \leq i \leq n$. The outcomes are labeled as 0 and 1. For each context C , every section t in the support of the model must satisfy two conditions. First, if the number of Y measurements in C is a multiple of 4, the

number of 1's in the outcomes allowed by t is even. Second, if the number of Y measurements in C is $4k + 2$, the number of 1s in the outcomes allowed under the assignment is odd.

In quantum mechanics, such models can be obtained by measuring the X and Y Pauli observables for the n -partite GHZ state

$$|GHZ\rangle = \frac{|0 \dots 0\rangle + |1 \dots 1\rangle}{\sqrt{2}}$$

To show that these are strongly contextual, we proceed by contradiction.

We first consider the case when $n = 4k, k \geq 1$. Assume that s is a global section compatible with the support of the GHZ model. If Y measurements are taken at every part, the number of 1 outcomes under the assignment is even.

If we replace any two Y 's by X 's, we must have the opposite parity for the number of 1 outcomes under the assignment. Thus for any Y^i, Y^j which have been assigned the same outcome, if we substitute X 's in those positions, they must receive different values. Similarly, for any Y^i, Y^j assigned different values, the corresponding X^i, X^j must receive the same value.

Suppose firstly that not all Y^i are assigned the same value by s . Then there exist some i, j, k such that Y^i is assigned the same value as Y^j , and Y^j is assigned a different value to Y^k . Then X^i is assigned the same value as X^k , and X^j is assigned the same value as X^k . By transitivity, X^i is assigned the same value as X^j , yielding a contradiction.

The remaining cases are those where all Y 's receive the same value. Then any pair of X 's must receive different values. But by taking any 3 X 's, this yields a contradiction, since there are only two values, so some pair must receive the same value.

The case when $n = 4k + 2, k \geq 1$, is proved in the same fashion, interchanging the parities. When $n \geq 5$ is odd, we start with a context containing one X , and again proceed similarly.

The most familiar case, for $n = 3$ does not admit this argument, which relies on having at least 4 Y 's in the initial configuration. However, this case is also strongly contextual, [2]. The proof uses a case analysis to show that there are 8 possible global sections satisfying the parity constraint on the 3 measurement combinations with two Y s and one X , and all of these violate the constraint for the XXX measurement.

Chapter 3

Classifying a Class of Permutationally Asymmetric Quantum States

In Chapter 2 we have seen that there is a hierarchy of forms of non-locality or contextuality which empirical models may satisfy. This hierarchy has three levels:

- A model is **strongly contextual** if its support has no global section; that is, there is no simultaneous assignment of outcomes to all the measurements whose restriction to each compatible set of measurements is in the support.
- A model is **logically contextual** if there are events in the support of some compatible family of measurements F_j which are not consistent with the supports of the other measurement contexts.
- Finally, a model is **weakly contextual** if it is contextual, but neither logically nor strongly contextual.

These notions form a proper hierarchy. Strong contextuality implies logical contextuality, which implies contextuality in the usual sense. On the other hand, there are weakly contextual models which are not logically contextual, and logically contextual models which are not strongly contextual.

The contextual characterisation of empirical models can also be used to characterise n -partite quantum states shared by n observers, as long as we allow each observer to choose one out of a finite set of measurements, each with a fixed set of possible outcomes. For example,

when the set of allowed measurements contains two elements, each with two possible outcomes, the resulting empirical model corresponds to a $(n, 2, 2)$ Bell scenario.

Note that the bipartite case seems to be anomalous within the landscape of multipartite entangled states. For example, the only strongly contextual bipartite models are given by super-quantum devices known as PR-boxes [2, 110]. These are however not quantum realizable. By contrast, for all $n > 2$, the n -partite GHZ states are strongly contextual [2]. Moreover, it is known that in the bipartite case, all entangled states *except* the maximally entangled ones admit Hardy arguments, and hence are logically contextual [83]

We have seen at the end of the previous chapter that (almost) all permutationally symmetric entangled states are at least logically contextual. We will now look at a class of highly non-permutation-symmetric entangled states, the *balanced states with functional dependencies*. These states are described by Boolean functions, and have a rich structure, allowing a detailed analysis. We provide logical contextuality witnesses (i.e. measurement choices) for almost all of these states and we also reveal a large collection of strongly contextual multipartite entangled states.

3.1 Balanced States with Functional Dependency

For each $n \geq 2$, a n -ary Boolean function is a function $F : \{0, 1\}^n \rightarrow \{0, 1\}$. Each n -ary Boolean function can be expressed as a multivariate polynomial over $GL(2)$:

$$F(x_1, \dots, x_n) = a_0 + \sum_i a_1^i x_i + \sum_{i,j} a_2^{i,j} x_i x_j + \dots + a_n^{1,2,\dots,n} x_1 x_2 \dots x_n$$

There are $2^n = 1 + n + \binom{n}{2} + \dots + \binom{n}{n}$ summands in the expression of the above polynomial, each of which containing a binary coefficient $a_t^{i_1, \dots, i_t}$. Hence there are 2^{2^n} distinct n -variate polynomials over $GF(2)$. Alternatively, each n -ary Boolean function can be expressed as a propositional formula in the Boolean variables $x_1, \dots, x_n$ [38].

We define a balanced $n + 1$ -qubit quantum state with a functional dependency given by a n -variate polynomial F as above to be a state which has the form

$$\Psi_F(n + 1) = \frac{1}{\sqrt{2^n}} \sum_{q_1 q_2 \dots q_n = 00 \dots 0}^{11 \dots 1} |q_1 q_2 \dots q_n F(q_1, q_2, \dots, q_n)\rangle$$

when expressed in the Z -basis.

We start by classifying the tripartite functionally dependent balanced states in terms of

their contextuality properties. A classification of the $n + 1$ -qubit states for $n > 2$ can then be obtained using the results from the tripartite scenarios.

3.2 Contextuality classification for the tripartite case

3.2.1 Polynomials of degree zero

There are $2^{2^2} = 16$ tripartite balanced states with a functional dependency. Two of these, namely

$$\frac{1}{2}|000\rangle + |010\rangle + |100\rangle + |110\rangle = \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)^{\otimes 2} \otimes |0\rangle$$

and

$$\frac{1}{2}|001\rangle + |011\rangle + |101\rangle + |111\rangle = \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)^{\otimes 2} \otimes |1\rangle$$

are obviously product states, and hence non-contextual. They correspond to the constant polynomials $F_0(q_1, q_2) = 0$ and $F_1(q_1, q_2) = 1$ respectively.

3.2.2 Degree one polynomials

There are six states whose corresponding polynomials have degree one. Two of these are given by the functional dependencies which correspond to the two-variable propositional formulas XOR and $NXOR$. Another four states are given by dictatorships, i.e. the value of the last qubit is dictated either by the value of the first qubit or by the value of the second qubit. We shall look at these two classes of states below.

XOR and NXOR The polynomials corresponding to the XOR and $NXOR$ states have the form

$$F_{XOR}^a(q_1, q_2) = a + q_1 + q_2$$

with $a = 0$ for XOR and $a = 1$ for $NXOR$.

Theorem 3.2.1. *The XOR state is strongly contextual if each party chooses between Y and Z measurements.*

Proof: The support of the probability table for the XOR state is

	+++	++-	+ - +	+ - -	- ++	- + -	- - +	- - -
YYY	1	1	1	1	1	1	1	1
YYZ	0	1	1	0	1	0	0	1
YZY	0	1	1	0	1	0	0	1
ZYY	0	1	1	0	1	0	0	1
YZZ	1	1	1	1	1	1	1	1
ZYZ	1	1	1	1	1	1	1	1
ZZY	1	1	1	1	1	1	1	1
ZZZ	1	0	0	1	0	1	1	0

One can simply inspect the table above and check that none of the sections in the support of the ZZZ row can be extended to global sections (i.e. each possible global assignment consistent with the support of the ZZZ row will restrict to a section outside the support on at least one of the three rows YYZ , YZY and ZYY). Thus there cannot be any global assignment of outcomes whose restriction to each set of compatible measurements is in the support of the model.

It is worth at this point to give a more formal expression to this argument in order to gain a better understanding of what is actually going on. For this recall that the $+$ and $-$ eigenstates of the Z observable are $|0\rangle$ and $|1\rangle$ respectively while for the Y observable they are (modulo some normalization constant which does not play any role in our argument) $|Y^+\rangle := |0\rangle + i|1\rangle$ and $|Y^-\rangle := |0\rangle - i|1\rangle$ respectively

We start our argument by assuming that a global section does exist. Assume next that this global section makes the assignment $Z^3 = +$. The probability of obtaining the outcome $Z^1 Z^2 +$ with $Z^i \in \{+, -\}$ is given by the squared norm of the inner product

$$\langle e_{Z^1} e_{Z^2} 0 | XOR \rangle = \langle e_{Z^1} e_{Z^2} 0 | \frac{|000\rangle + |011\rangle + |101\rangle + |110\rangle}{2}$$

where $e_+ = 0$ and $e_- = 1$. If we regard each e_{Z^i} as an element of $GF(2)$ then the inner product above is non-zero only if

$$F_{XOR}^0(e_{Z^1}, e_{Z^2}) = e_{Z^1} + e_{Z^2} = 0$$

So the sections in the support of the ZZZ for which $Z^3 = +$ must have $Z^1 = Z^2$, as the table confirms.

Next consider the YYZ set of compatible measurements. The probability (modulo normalization constants) of obtaining the outcome $Y^1 Y^2 +$ with $Y^i \in \{+, -\}$ for this set of mea-

surements is given by the squared norm of the inner product

$$\langle Y^{Y^1} Y^{Y^2} 0 | XOR \rangle = \langle Y^{Y^1} Y^{Y^2} 0 | \frac{|000\rangle + |011\rangle + |101\rangle + |110\rangle}{2} \rangle \quad (3.2.1)$$

We have

$$\begin{aligned} \langle Y^+ Y^+ | &= \langle 00 | + i\langle 01 | + i\langle 10 | - \langle 11 | \\ \langle Y^+ Y^- | &= \langle 00 | - i\langle 01 | + i\langle 10 | + \langle 11 | \\ \langle Y^- Y^+ | &= \langle 00 | + i\langle 01 | - i\langle 10 | + \langle 11 | \\ \langle Y^- Y^- | &= \langle 00 | - i\langle 01 | - i\langle 10 | - \langle 11 | \end{aligned}$$

and since $F_{XOR}^0(0, 1) = F_{XOR}^0(1, 0) \neq 0$ the imaginary part of the tensor products above will not bring any contribution towards the value of the inner product (3.2.1). The only contribution will come from the real part of the tensor products above, and it is easy to see that the inner product (3.2.1) will vanish when $Y^1 = Y^2$. So we must have $Y^1 \neq Y^2$ in any global assignment which sends Z^3 to + in order to stay within the support of the YYZ row.

On the other hand, the probabilities of obtaining the outcomes $Y^1 z Y^3$ and $z Y^2 Y^3$, where $z = Z^1 = Z^2$, for the YZY and ZYY sets of compatible measurements are given by the inner products

$$\begin{aligned} \langle Y^{Y^1} e_z Y^{Y^3} | XOR \rangle &= (\langle 0e_z 0 | + iY^3 \langle 0e_z 1 | + iY^1 \langle 1e_z 0 | - (Y^1 Y^3) \langle 1e_z 1 |) | XOR \rangle \\ \langle e_z Y^{Y^2} Y^{Y^3} | XOR \rangle &= (\langle e_z 00 | + iY^3 \langle e_z 01 | + iY^2 \langle e_z 10 | - (Y^2 Y^3) \langle e_z 11 |) | XOR \rangle \end{aligned} \quad (3.2.2)$$

If $e_z = 0$ the imaginary part of the two expressions in (3.2.2) will be equal to zero for all values of Y^i . If $e_z = 1$ the real part of the two expressions in (3.2.2) will vanish for all values of Y^i . In the first case the expressions are non-zero only if $Y^1 = Y^2 = -Y^3$ and in the second case they are non-zero only if $Y^1 = Y^2 = Y^3$. But both these assignments violate the previous requirement that $Y^1 \neq Y^2$.

So far we have established the fact that no global section can assign the outcome + to Z^3 . If on the other hand the outcome - is assigned to Z^3 , we can construct a similar argument which yields a contradiction. This time the sections in the support of ZZZ for which $Z^3 = -$ must have $Z^1 = -Z^2$. The sections in the support of YYZ must have $Y^1 = Y^2$, while those in the support of YZY and ZYY must either have $Y^1 = -Y^3 = -Y^2$ for $e_{Z^2} = 0, e_{Z^3} = 1$ or $Y^1 = Y^3 = -Y^2$ for $e_{Z^2} = 1$ and $e_{Z^3} = 0$. $\square$

Theorem 3.2.2. *The $NXOR$ state is also strongly contextual if each party chooses between Y and Z measurements.*

Proof: The support of the probability table for the $NXOR$ state is

	+++	++-	+ - +	+ - -	- ++	- + -	- - +	- - -
YYY	1	1	1	1	1	1	1	1
YYZ	1	0	0	1	0	1	1	0
YZY	1	0	0	1	0	1	1	0
ZYY	1	0	0	1	0	1	1	0
YZZ	1	1	1	1	1	1	1	1
ZYZ	1	1	1	1	1	1	1	1
ZZY	1	1	1	1	1	1	1	1
ZZZ	0	1	1	0	1	0	0	1

The argument for strong contextuality follows the same pattern as for the XOR state. We assume by contradiction that a global section exists, and that it makes the assignment $Z^3 = +$. Then from the ZZZ row we obtain the requirement that $Z^1 \neq Z^2$. From the YYZ row we obtain that $Y^1 = Y^2$ and from the YZY and ZYY rows we obtain that $Y^1 \neq Y^2$, which is a contradiction.

Similarly, if $Z^3 = -$ we must have $Z^1 = Z^2$ and $Y^1 \neq Y^2$ from the ZZZ and YYZ rows. This means we must also have $Y^1 = Y^2$ from the YZY and ZYY rows, which again is a contradiction.

Note at this point that the similarity between these two arguments for strong contextuality is due to the similar structure of the tables for the XOR and $NXOR$ states. Namely, the second table can be obtained from the first by interchanging the $+$ and $-$ signs which label the table columns. Thus the second argument is the same as the first, only with the $+$ and $-$ signs interchanged. $\square$

Dictatorships

The four degree one polynomials of the form $F_1^a(q_1, q_2) = a + q_1$ and $F_1^a(q_1, q_2) = a + q_2$ where $a \in \{0, 1\}$ correspond to the dictatorship states, where the value of the last qubit is dictated by the value of either the first or of the second qubit. In the Z basis these states are

$$\Delta_2^+ := \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

or

$$\Delta_2^- := \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|01\rangle + |10\rangle}{\sqrt{2}}$$

if the dictatorship is given by the second qubit. Similarly, if the dictatorship is given by the first

qubit, we have two possible states

$$\Delta_1^+ := \frac{|0_2\rangle + |1_2\rangle}{\sqrt{2}} \otimes \frac{|0_1 0_3\rangle + |1_1 1_3\rangle}{\sqrt{2}}$$

and

$$\Delta_1^- := \frac{|0_2\rangle + |1_2\rangle}{\sqrt{2}} \otimes \frac{|0_a 1_3\rangle + |1_1 0_3\rangle}{\sqrt{2}}$$

where the subscripts 1, 2 and 3 indicate whether the qubit belongs to the first, second or third party respectively.

Proposition 3.2.3. *The four dictatorship states are weakly contextual for suitable dichotomic choices of measurements.*

Proof:

Consider the general form of an observable, given in terms of angles θ and ϕ on the Bloch sphere

$$U(\theta, \phi) := \begin{pmatrix} \cos \theta & e^{-i\phi} \sin \theta \\ e^{i\phi} \sin \theta & -\cos \theta \end{pmatrix}$$

We will use the fact that the bell basis states $\Phi^+ = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$ and $\Phi^- = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$ are weakly contextual with respect to suitable choices of measurements.

It can be machine checked that the state Φ^+ is weakly contextual if we allow each party to choose between the measurements $A := U\left(\frac{\pi}{2}, \frac{\pi}{8}\right)$ and $B := U\left(\frac{\pi}{2}, \frac{5\pi}{8}\right)$, while the state Φ^- is weakly contextual if we allow each party to choose between the measurements $C := U\left(\frac{\pi}{8}, \frac{\pi}{2}\right)$ and $D := U\left(\frac{5\pi}{8}, \frac{\pi}{2}\right)$

In fact, it can also be machine checked that this choice of measurements gives a maximal violation of Bell inequalities for both states.

The probability models of the dictatorship states can be obtained from the probability models of the states Φ^+ and Φ^- in a straightforward way. Let $|+_A\rangle$ and $|-_A\rangle$ stand for the eigenstates of A and $|+_B\rangle$ and $|-_B\rangle$ stand for the eigenstates of B .

Define the two constants

$$a_+ := \frac{1}{\sqrt{2}}(\langle +_A | 0 \rangle + \langle +_A | 1 \rangle)$$

$$a_- := \frac{1}{\sqrt{2}}(\langle -_A | 0 \rangle + \langle -_A | 1 \rangle)$$

Note that $a_+ + a_- = 1$. Similarly, define the two constants

$$b_+ := \frac{1}{\sqrt{2}}(\langle +_B|0\rangle + \langle +_B|1\rangle)$$

$$b_- := \frac{1}{\sqrt{2}}(\langle -_B|0\rangle + \langle -_B|1\rangle)$$

Up to two decimal points precision, the probability table of the Φ^+ state for the observables A and B is

	++	+-	-+	--
AA	0.43	0.07	0.07	0.43
AB	0.07	0.43	0.43	0.07
BA	0.07	0.43	0.43	0.07
BB	0.07	0.43	0.43	0.07

The inner product formula (2.3.1) for computing probabilities implies that the probability table of the dictatorship state Δ_2^+ can be expressed in terms of the constants a_+ , a_- , b_+ and b_- and the probability table of Φ^+ :

	+++	++-	+ - +	+ - -	- + +	- + -	- - +	- - -
AAA	$0.43a_+$	$0.07a_+$	$0.07a_+$	$0.43a_+$	$0.43a_-$	$0.07a_-$	$0.07a_-$	$0.43a_-$
AAB	$0.07a_+$	$0.43a_+$	$0.43a_+$	$0.07a_+$	$0.07a_-$	$0.43a_-$	$0.43a_-$	$0.07a_-$
ABA	$0.07a_+$	$0.43a_+$	$0.43a_+$	$0.07a_+$	$0.07a_-$	$0.43a_-$	$0.43a_-$	$0.07a_-$
ABB	$0.07a_+$	$0.43a_+$	$0.43a_+$	$0.07a_+$	$0.07a_-$	$0.43a_-$	$0.43a_-$	$0.07a_-$
BAA	$0.43b_+$	$0.07b_+$	$0.07b_+$	$0.43b_+$	$0.43b_-$	$0.07b_-$	$0.07b_-$	$0.43b_-$
BAB	$0.07b_+$	$0.43b_+$	$0.43b_+$	$0.07b_+$	$0.07b_-$	$0.43b_-$	$0.43b_-$	$0.07b_-$
BBA	$0.07b_+$	$0.43b_+$	$0.43b_+$	$0.07b_+$	$0.07b_-$	$0.43b_-$	$0.43b_-$	$0.07b_-$
BBB	$0.07b_+$	$0.43b_+$	$0.43b_+$	$0.07b_+$	$0.07b_-$	$0.43b_-$	$0.43b_-$	$0.07b_-$

Note also that the table of the dictatorship state Δ_1^+ will have the same values as the one above, but the rows will be indexed in the order AAA , AAB , BAA , BAB , ABA , ABB , BBA , BBB , since the coefficients $a_{+/-}$ and $b_{+/-}$ come from the second qubit's contribution to the inner product.

It is now straightforward to deduce that the states Δ_1^+ and Δ_2^+ are indeed weakly contextual for the same choice of measurements for which the Φ^+ state is weakly contextual, since any probability distribution on the set of global sections of one of these two dictatorship states would restrict to a probability distribution on the set of global sections of the Φ^+ state.

Next note that up to two decimal points precision, the probability table of the Φ^- state for the observables C and D is

	++	+-	-+	--
AA	0.43	0.07	0.07	0.43
AB	0.07	0.43	0.43	0.07
BA	0.07	0.43	0.43	0.07
BB	0.07	0.43	0.43	0.07

and the probability tables of the Δ_1^- and Δ_2^- dictatorship states can be expressed in terms of the table above and four suitably defined constants $c_{+/-}$ and $d_{+/-}$, so by analogy with the Δ_1^+ and Δ_2^+ case, these states will also be weakly contextual. $\square$

Theorem 3.2.4. *None of the four dictatorship states is logically contextual, for **any** dichotomic choice of measurements.*

Proof: The relationship between probability tables discussed in Proposition 3.2.3 allows us to reduce the problem to the bipartite scenario. Thus we seek to prove that neither of the two Bell basis states is logically contextual for any given choice of measurements.

Let $A := U(\theta_1, \phi_1)$ and $B := U(\theta_2, \phi_2)$. Let c , s and f stand for $\cos \frac{\theta_1}{2}$, $\sin \frac{\theta_1}{2}$ and $e^{i\phi_1}$ respectively. Similarly, let k , z and v stand for $\cos \frac{\theta_2}{2}$, $\sin \frac{\theta_2}{2}$ and $e^{i\phi_2}$ respectively. Then the general form of the probability model of the Φ^+ state is

	++	+-	-+	--
AA	$ c^2 + f^2 \cdot s^2 ^2$	$ cs - f^2 \cdot cs ^2$	$ cs - f^2 \cdot cs ^2$	$ s^2 + f^2 \cdot c^2 ^2$
AB	$ ck + fv \cdot sz ^2$	$ cz - fv \cdot sk ^2$	$ sk - fv \cdot cz ^2$	$ sz + fv \cdot ck ^2$
BA	$ ck + fv \cdot sz ^2$	$ sk - fv \cdot cz ^2$	$ cz - fv \cdot sk ^2$	$ sz + fv \cdot ck ^2$
BB	$ k^2 + v^2 \cdot z^2 ^2$	$ kz - v^2 \cdot kz ^2$	$ kz - v^2 \cdot kz ^2$	$ z^2 + v^2 \cdot k^2 ^2$

In most cases, all of the sections in the model of Φ^+ will be in the support, in which case the state is clearly not logically contextual. However, for certain values of c , f , v and k (which may be chosen independently of each other) the entries of the table above may vanish, which will exclude certain sections from the support. It suffices therefore to check that the resulting possibilistic models are not logically contextual for any choices of c , f , v and k (and implicitly also of s and z) which would allow one or more of the above table entries to vanish. We therefore need to consider each element in the powerset of the following set of conditions on c , s , f , v , z and k :

$$\left\{ c \vee k \in \{0, \pm 1\}, f \vee v \in \{\pm 1, \pm i\}, f = \pm \frac{1}{v}, c = \pm s, k = \pm z, ck = \pm sz, cz = \pm sk \right\}$$

A computer can easily verify that no subset of the above set of conditions leads to a logically contextual probability model.

Finally, using the relation between probability tables from Proposition 3.2.3, we note that any global section of the model above can be easily extended to a global section of the corresponding dictatorship state model by adding the assignment $+$ to the third party's outcome for the A measurement, if $a_+ \neq 0$ and $-$ otherwise, and similarly for the third party's outcome corresponding to the B measurement. We can therefore conclude that for all possible choices of measurements, the dictatorship states corresponding to Φ^+ can not be logically contextual.

For the Φ^- state note that the observables $C := U(\phi_1, \theta_1)$ and $D := U(\phi_2, \theta_2)$ will give the probability model

	$++$	$+-$	$-+$	$--$
AA	$ cs - f^2 \cdot cs ^2$	$ c^2 + f^2 \cdot s^2 ^2$	$ s^2 + f^2 \cdot c^2 ^2$	$ cs - f^2 \cdot cs ^2$
AB	$ cz - fv \cdot sk ^2$	$ ck + fv \cdot sz ^2$	$ sz + fv \cdot ck ^2$	$ sk - fv \cdot cz ^2$
BA	$ sk - fv \cdot cz ^2$	$ ck + fv \cdot sz ^2$	$ sz + fv \cdot ck ^2$	$ cz - fv \cdot sk ^2$
BB	$ kz - v^2 \cdot kz ^2$	$ k^2 + v^2 \cdot z^2 ^2$	$ z^2 + v^2 \cdot k^2 ^2$	$ kz - v^2 \cdot kz ^2$

where c, k, s, z now take $\phi_i/2$ as arguments while f and v take θ_i as arguments.

We can show that this model is also not logically contextual, using an argument completely analogous to the one used for the Φ^+ state. Hence the dictatorship states corresponding to the Φ^- state are also not logically contextual. $\square$

3.2.3 Degree two polynomials

There are eight balanced functionally dependent states whose corresponding polynomials have degree two. Four of these correspond to the two-variable propositional formulas AND , $NAND$, OR and NOR . Their respective polynomials have the form

$$F_{AND}^a(q_1, q_2) = a + q_1 q_2$$

and

$$F_{OR}^a = a + q_1 + q_2 + q_1 q_2$$

with $a = 0$ for *AND* and *OR* and $a = 1$ for *NAND* and *NOR*.

The other four states correspond to logical implication and its negation. We use L_1 , L_2 , NL_1 and NL_2 to denote the propositional formulas $q_1 \Rightarrow q_2$, $q_2 \Rightarrow q_1$ and $\overline{q_1 \Rightarrow q_2}$, $\overline{q_2 \Rightarrow q_1}$ respectively. The polynomials corresponding to these propositional formulas are of the form

$$F_{NL_i}^a = a + q_i + q_1 q_2$$

with $i \in \{1, 2\}$, $a = 0$ for NL_i and $a = 1$ for L_i .

All the eight states described above turn out to be logically contextual if we choose Y and Z measurements in each part.

Theorem 3.2.5. *The AND state is logically contextual.*

Proof: The support of the probability table for the *AND* state is

	+++	++-	+ - +	+ - -	- ++	- + -	- - +	- - -
YYY	1	1	1	1	1	1	1	1
YYZ	1	1	1	1	1	1	1	1
YZY	1	1	0	1	1	1	1	0
ZYY	1	1	1	1	0	1	1	0
YZZ	1	0	1	1	1	0	1	1
ZYZ	1	0	1	0	1	1	1	1
ZZY	1	1	1	1	1	1	1	1
ZZZ	1	0	1	0	1	0	0	1

The global assignment $Z^1 Z^2 Z^3 Y^1 Y^2 Y^3 = + + + + ++$ is clearly consistent with the support of the *AND* table, so this state is not strongly contextual for Y and Z measurements. However, not all sections in the support can be extended to global sections. Consider for example the section $Y^1 Y^2 Z^3 = + - -$ which is in the support. The only section on the *ZZZ* row consistent with it is $Z^1 Z^2 Z^3 = - - -$. But it is now impossible to assign an outcome to Y^3 which will make the resulting global section restrict to sections in the support of both of the rows *YZY* and *ZYY*. In fact, there are only two sections in the support of the *YYZ* row which cannot be extended to global ones. These are the sections where the two Y measurements are assigned different outcomes, while the Z measurement is assigned the outcome $-$. $\square$

Theorem 3.2.6. *The NAND state is logically contextual.*

Proof: The support of the probability table for the *NAND* state is

	+++	++-	+ - +	+ - -	- ++	- + -	- - +	- - -
YYY	1	1	1	1	1	1	1	1
YYZ	1	1	1	1	1	1	1	1
YZY	1	1	1	0	1	1	0	1
ZYY	1	1	1	1	1	0	0	1
YZZ	0	1	1	1	0	1	1	1
ZYZ	0	1	0	1	1	1	1	1
ZZY	1	1	1	1	1	1	1	1
ZZZ	0	1	0	1	0	1	1	0

Note that this table can be obtained from the *AND* table by simply relabeling the columns. The relabeling sends the first + to +, the second + to + and the third + to -, and it sends the first two -s to - and the third one to +.

The same argument used in the proof of Theorem 3.2.5 can therefore be used to prove the logical contextuality of the *NAND* state, with the provision that the new labeling replaces the one used within the old argument's statements. $\square$

Remark 3.2.7. *The notation $+++ \mapsto ++-$ unambiguously describes the relabeling used in the proof of Theorem 3.2.6, and we shall use this shorthand notation in further proofs.*

Theorem 3.2.8. *The *OR*, *NOR*, L_1 , NL_1 , L_2 and NL_2 states are all logically contextual.*

Proof: The support of the probability tables for these states are also obtained from the *AND* table by column relabelings, so the argument used in the proof of Theorem 3.2.5 can again be used to prove the logical contextuality of these states. The necessary relabelings are

- 1) $+++ \mapsto ---$ for the *OR* state
- 2) $+++ \mapsto --+$ for *NOR*
- 3) $+++ \mapsto +- -$ for L_1
- 4) $+++ \mapsto + - +$ for NL_1
- 5) $+++ \mapsto - + -$ for L_2
- 6) $+++ \mapsto - ++$ for NL_2

$\square$

Remark 3.2.9. *The relabelings above can also be used for the probability tables themselves, not only for their supports, but only for Y, Z measurements. For general choices of measurements there is no simple relation between the probability tables of the balanced states with functional dependency given by degree two polynomials, nor between their supports.*

3.3 Contextuality classification for the $n + 1$ -partite case, $n > 2$

We can use the results of the previous section to classify the $n + 1$ -partite balanced states which have a functional dependency. In the rest of this section, let F_n denote a polynomial in n variables.

3.3.1 Strongly contextual states

Theorem 3.3.1. *Given a $n + 1$ -partite balanced quantum state whose functional dependency is given by the polynomial $F_n(q_1, \dots, q_n)$, the state is strongly contextual if the polynomial F_n is of the form*

$$F_n(q_1, \dots, q_n) = q_i + q_j + F_{n-2}(q_1, \dots, \hat{q}_i, \dots, \hat{q}_j, \dots, q_n)$$

for some variables q_i and q_j and some polynomial F_{n-2} .

Proof: If Y and Z measurements are chosen by each party, then we can show that none of the sections in the support of the $ZZZ \dots Z$ row can be extended to a global section.

Consider any fixed assignment of outcomes to the Z measurements performed by the first n parties, except the i^{th} and the j^{th} party. Let $\sigma_k \in \{+, -\}$, $k \neq i, j$ denote the outcome corresponding to the measurement performed by the k^{th} party. Next evaluate the polynomial F_{n-2} at the values of $q_1, \dots, \hat{q}_i, \dots, \hat{q}_j, \dots, q_n$ corresponding to the fixed assignment of outcomes, using the convention that 0 corresponds to the $+$ outcome and 1 corresponds to the $-$ outcome. Use a to denote the result of the evaluation.

Depending on the value of a we can use the argument made for the strong contextuality of either the XOR or the $NXOR$ state in order to show that there is no consistent assignment of outcomes which will restrict to sections in the support for all four of the following rows:

$$Z \dots ZZ_i Z \dots ZZ_j Z \dots ZZ$$

$$Z \dots ZY_i Z \dots ZY_j Z \dots ZZ$$

$$Z \dots ZZ_i Z \dots ZY_j Z \dots ZY$$

$$Z \dots ZY_i Z \dots ZZ_j Z \dots ZY$$

Since this can be done for all possible assignments of outcomes to the Z measurements performed by the first n parties, except the i^{th} and the j^{th} party, the quantum state we are considering must be strongly contextual. $\square$

3.3.2 Logically contextual states

Theorem 3.3.2. *Any $n+1$ -partite balanced quantum state whose functional dependency is given by a polynomial $F_n(q_1, \dots, q_n)$ of degree at least two which is not of the form*

$$F_n(q_1, \dots, q_n) = q_i + q_j + F_{n-2}(q_1, \dots, \hat{q}_i, \dots, \hat{q}_j, \dots, q_n)$$

for any choice of variables q_i and q_j and polynomial F_{n-2} is logically contextual.

Proof: Consider any two variables q_i and q_j which appear in at least one of the terms with degree at least two of the polynomial F_n . The polynomial F_n can be rewritten as

$$F_n(q_1, \dots, q_n) = F_{n-2}^1 + q_i F_{n-2}^2 + q_j F_{n-2}^3 + q_i q_j F_{n-2}^4$$

where F_{n-2}^i are $n-2$ variable polynomials in $q_1, \dots, \hat{q}_i, \dots, \hat{q}_j, \dots, q_n$.

Next choose any assignment of outcomes to the Z measurements performed by the first n parties, except the i^{th} and the j^{th} party, such that the polynomial F^4 evaluates to 1 at the values of $q_1, \dots, \hat{q}_i, \dots, \hat{q}_j, \dots, q_n$ corresponding to this assignment. Using this assignment, we have obtained a degree two polynomial in two variables, q_i and q_j .

We can now use one of the arguments in Section 3.2.3 in order to identify at least two sections in the support of the

$$Z \dots ZY_i Z \dots ZY_j Z \dots ZZ$$

row which cannot be extended to a global section consistent with the support of the rows

$$\begin{aligned}
& Z \dots ZZ_i Z \dots ZY_j Z \dots ZY \\
& Z \dots ZY_i Z \dots ZZ_j Z \dots ZY \\
& \text{and} \\
& Z \dots ZZ_i Z \dots ZZ_j Z \dots ZZ \quad \square
\end{aligned}$$

Note however that showing that at least one global section *does* exist for the class of states considered in the Theorem above is not as simple as in the tripartite case, so strong contextuality cannot be immediately ruled out for these states even in the special case when one considers only Y and Z measurements.

3.3.3 Weakly contextual states

Theorem 3.3.3. *Any $n+1$ -partite balanced quantum state whose functional dependency is given by a polynomial $F_n(q_1, \dots, q_n)$ of degree one which is not of the form*

$$F_n(q_1, \dots, q_n) = q_i + q_j + F_{n-2}(q_1, \dots, \hat{q}_i, \dots, \hat{q}_j, \dots, q_n)$$

for any choice of variables q_i and q_j and polynomial F_{n-2} is weakly contextual.

Proof: Any degree one polynomial which is not of the above form must contain precisely one term. Thus the state we are dealing with is a dictatorship state, i.e. the value of the last qubit is dictated by the value of its i^{th} qubit, and the state is either of the form

$$\Delta_i^+ := \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \right)^{\otimes n} \otimes \frac{|0_i 0_{n+1}\rangle + |1_i 1_{n+1}\rangle}{\sqrt{2}}$$

or

$$\Delta_i^- := \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \right)^{\otimes n} \otimes \frac{|0_i 1_{n+1}\rangle + |1_i 0_{n+1}\rangle}{\sqrt{2}}$$

and its probability table can be expressed in terms of a suitable choice of $n-2$ constants and the probability table of either the Φ^+ or of the Φ^- state.

A straightforward inductive argument based on the argument used in Proposition 3.2.3 will show that the $n+1$ -partite dictatorship states are also weakly contextual for the measurements $U\left(\frac{\pi}{2}, \frac{\pi}{8}\right)$, $U\left(\frac{\pi}{2}, \frac{5\pi}{8}\right)$ and $U\left(\frac{\pi}{8}, \frac{\pi}{2}\right)$, $U\left(\frac{5\pi}{8}, \frac{\pi}{2}\right)$ respectively.

Moreover, the generalization of the argument used in Theorem 3.2.4 shows that the $n + 1$ -partite dictatorship states are not logically contextual for any possible dichotomic choice of measurements. $\square$

3.3.4 Non-contextual states

Any $n + 1$ -partite balanced quantum state whose functional dependency is given by a constant polynomial is clearly a product state and hence non-contextual.

Chapter 4

Logical Contextuality is Almost Everywhere

We have seen in previous chapters that quantum states together with local observables (i.e. one-qubit measurements which can be made by each of the parties on their part of the state) give rise to probability tables which are, respectively, weakly, logically or strongly contextual.

This leads to the natural question of finding the maximum level of contextuality which a quantum state can exhibit for some choice of local observables.

We have seen in the previous chapter that maximally entangled two-qubit states are weakly contextual, but they cannot be logically contextual, regardless of the local observables chosen. This is consistent with Hardy's results [82, 83], who showed that an inequality-free proof of Bell's theorem (or equivalently, in the language of Abramsky and Brandenburger [2], a logically contextual empirical model) can be given for all entangled two-qubit quantum states which are not maximally entangled.

We have also seen that among the multipartite states considered so far, Dyck states, GHZ states, and balanced states with functional dependencies, almost all of them have turned out to be at least logically contextual, for a suitable choice of observables. The only exceptions were given by those states which were equivalent to tensor products of maximally entangled two-qubit states and single states. In this chapter we shall see that this is not simply a coincidence, and in fact the following statement holds.

Theorem 4. *Let $|\psi\rangle$ be an n -qubit pure state. Then exactly one of the following two cases must hold:*

1. $|\psi\rangle$ can be written (up to permutation of tensor factors) as a product

$$|\psi\rangle = |\psi_1\rangle \otimes \cdots \otimes |\psi_k\rangle \quad (4.1)$$

where each $|\psi_i\rangle$ is either a 1-qubit state, or a 2-qubit maximally entangled state.

2. There are local observables such that the probability table arising from $|\psi\rangle$ and these local observables is logically contextual, that is, it admits a “Hardy paradox”, i.e. an inequality-free, probability-free proof of non-locality.

If (2) holds, then only $n + 2$ local observables are needed; two each for two of the parties, and one each for the other $n - 2$ parties. Moreover, there is an algorithm to decide which of the above cases holds, and which in case (2) explicitly computes the witnessing local observables. The complexity of this algorithm is $O(d \log^3 d)$ in the dimension $d = 2^n$.

The remainder of this chapter is devoted to the detailed proof of this theorem. In particular, while the algorithm is quite simple, the arguments justifying its correctness are non-trivial.

4.1 Overview of the argument

In this section, we shall state a number of main lemmas, and show how Theorem 4 is proved from these lemmas. The proofs of the lemmas will be given in the following section.

Notation. We shall write $\mathcal{P}_n$ for the set of n -qubit pure states of the form (4.1) given in Theorem 4. For succinctness, we shall write the tensor product of a ket $|\psi\rangle$ with a 1-qubit ket $|i\rangle$, $i = 0, 1$, as $|\psi\rangle |i\rangle$ rather than $|\psi\rangle \otimes |i\rangle$.

We shall prove Theorem 4 by induction on n . The case $n = 1$ is trivial. The $n = 2$ case is given by the following lemma.

Lemma 4.1.1 (The Base Case Lemma). *Every 2-qubit state $|\psi\rangle$ is either in $\mathcal{P}_2$ (i.e. it is either a product state, or a maximally entangled bipartite state), or there are two local observables for each party, which can be computed directly from the Schmidt decomposition of $|\psi\rangle$, and which witness the logical contextuality of $|\psi\rangle$.*

The following lemmas and corollary will be used in the induction step.

Lemma 4.1.2 (Going Up Lemma I). *If an n -qubit state $|\psi\rangle$ is logically contextual with some choice of local observables, then for any n -qubit state $|\theta\rangle$, and $\alpha, \beta \in \mathbb{C}$ with $\alpha \neq 0$ and*

$|\alpha|^2 + |\beta|^2 = 1$, the states

$$\alpha |\psi\rangle |0\rangle + \beta |\theta\rangle |1\rangle, \quad \beta |\theta\rangle |0\rangle + \alpha |\psi\rangle |1\rangle$$

are also logically contextual with the same choice of observables, augmented with a single additional observable for the $n + 1^{th}$ party.

Lemma 4.1.3 (Going Up Lemma II). *If $|\theta\rangle = \alpha|\psi\rangle + \beta|\phi\rangle$ is a logically contextual n -qubit state under some choice of local observables, then for any non-zero $x, y \in \mathbb{C}$ such that $|x\alpha|^2 + |y\beta|^2 = 1$, the state $|\omega\rangle = x\alpha|\psi\rangle|0\rangle + y\beta|\phi\rangle|1\rangle$ is also logically contextual under the same choice of observables, augmented with a single additional observable for the $n + 1^{th}$ party.*

As an easy consequence of these lemmas, we have:

Corollary 4.1.4. *If $|\theta\rangle$ is logically contextual under some choice of local observables, so are $|\theta\rangle \otimes |\eta\rangle$ and $|\eta\rangle \otimes |\theta\rangle$ for any state $|\eta\rangle$, with the same choice of local observables for each party in $|\theta\rangle$, and a single observable for each party in $|\eta\rangle$.*

We now consider the induction step where we have an $n + 1$ -qubit state $|\omega\rangle$, $n > 1$. We can write

$$|\omega\rangle = \alpha |\psi\rangle |0\rangle + \beta |\phi\rangle |1\rangle. \quad (4.1.1)$$

By the Going Up Lemma I, if either $|\psi\rangle$ or $|\phi\rangle$ are logically contextual, so is $|\omega\rangle$, and we are done.

Suppose now that $|\psi\rangle$ and $|\phi\rangle$ are both in $\mathcal{P}_n$. We consider the parameterised family of states

$$\tau(a) = a |\psi\rangle + \sqrt{1 - a^2} |\phi\rangle, \quad a \in [0, 1]. \quad (4.1.2)$$

If for some a , $\tau(a)$ is logically contextual, so is $|\omega\rangle$, by the Going Up Lemma II. For the remaining case, we have the following rather remarkable result.

Lemma 4.1.5 (The Small Difference Lemma). *Let $|\psi\rangle$ and $|\phi\rangle$ be states in $\mathcal{P}_n$, and suppose that for all $a \in [0, 1]$, $\tau(a)$ is in $\mathcal{P}_n$, where $\tau(a)$ is defined by (4.1.2). Then $|\psi\rangle$ and $|\phi\rangle$ differ in at most one qubit.*

Applying this result to our decomposition (4.1.1) of $|\omega\rangle$, we have the following possibilities:

- $|\psi\rangle = |\phi\rangle$, in which case, from (4.1.1) and the bilinearity of the tensor product:

$$|\omega\rangle = \alpha |\psi\rangle |0\rangle + \beta |\psi\rangle |1\rangle = |\psi\rangle \otimes (\alpha |0\rangle + \beta |1\rangle).$$

Since by assumption $|\psi\rangle$ is in $\mathcal{P}_n$, $|\omega\rangle$ is in $\mathcal{P}_{n+1}$.

- $|\psi\rangle \neq |\phi\rangle$, in which case (up to permutation) we can write

$$\begin{aligned} |\psi\rangle &= |\Psi\rangle \otimes |\eta\rangle \\ |\phi\rangle &= |\Psi\rangle \otimes |\theta\rangle \end{aligned}$$

where $|\Psi\rangle$ is in $\mathcal{P}_{n-1}$ and $|\eta\rangle$ and $|\theta\rangle$ are 1-qubit states. From this and (4.1.1), using the bilinearity of tensor product again we have

$$|\omega\rangle = |\Psi\rangle \otimes |\xi\rangle$$

where $|\xi\rangle$ is a 2-qubit state. We can apply the Base Case Lemma to $|\xi\rangle$ to conclude that $|\xi\rangle$ is either in $\mathcal{P}_2$, in which case $|\omega\rangle$ is in $\mathcal{P}_{n+1}$, or $|\xi\rangle$ is logically contextual, in which case $|\omega\rangle$ is logically contextual by the corollary to the Going Up Lemma.

At this point, we have established (1) and (2) of Theorem 4, but it seems that we require an infinite search to determine if there exists some $a \in [0, 1]$ for which $\tau(a)$ is logically contextual.

However, the following lemma shows that we only need to test a fixed, finite number of values for a to determine this.

Lemma 4.1.6 (The 21 Lemma). *With the same notation as in the Small Difference Lemma, suppose that $\tau(a)$ is in $\mathcal{P}_n$ for 21 distinct values of a in $[0, 1]$. Then $\tau(a)$ is in $\mathcal{P}_n$ for all $a \in [0, 1]$.*

Thus this lemma allows us to determine which case applies on the basis of a finite number of tests.

In the next section, we shall give proofs of these lemmas. We shall then give an explicit algorithm in Section 5 to complete the proof of Theorem 4.

4.2 Auxiliary lemmas

Firstly, we collect a few useful basic properties.

4.2.1 Background lemmas

We consider relations $\sim = \{\sim_n\}_{n \in \mathbb{N}}$, where $\sim_n$ is an equivalence relation on n -qubit states. We say that $\sim$ is *LC invariant* if for all n -qubit states $|\psi\rangle, |\phi\rangle$:

- If $|\psi\rangle \sim_n |\phi\rangle$, then $|\psi\rangle \in \mathcal{P}_n$ iff $|\phi\rangle \in \mathcal{P}_n$.

- If $|\psi\rangle \sim_n |\phi\rangle$, then $|\psi\rangle$ is logically contextual iff $|\phi\rangle$ is logically contextual.

Lemma 4.2.1. *The relation induced by permutation of tensor factors is LC invariant.*

Lemma 4.2.2. *The relation of LU equivalence is LC invariant. Here LU equivalence refers to the relation induced by the action of local (1-qubit) unitaries.*

The following result will be used in the proof of the Small Difference lemma. It refers to the “partial inner product” operation described e.g. in [116, p. 129].¹

Lemma 4.2.3. *Let $|\phi\rangle$ be a state in $\mathcal{H} \otimes \mathcal{K}$. For any states $|\eta\rangle$ in $\mathcal{H}$ and $|\theta\rangle$ in $\mathcal{K}$, if for all $|\eta^\perp\rangle$ orthogonal to $|\eta\rangle$, $\langle \eta^\perp | \phi \rangle = \mathbf{0}$, and for all $|\theta^\perp\rangle$ orthogonal to $|\theta\rangle$, $\langle \theta^\perp | \phi \rangle = \mathbf{0}$, then (up to global phase) $|\phi\rangle = |\eta\rangle \otimes |\theta\rangle$.*

Proof: We extend $|\eta\rangle$ into an orthonormal basis $|\eta_1\rangle, \dots, |\eta_n\rangle$ with $|\eta\rangle = |\eta_1\rangle$, and similarly extend $|\theta\rangle$ into $|\theta_1\rangle, \dots, |\theta_m\rangle$ with $|\theta\rangle = |\theta_1\rangle$. Then $B = \{|\eta_i\rangle \otimes |\theta_j\rangle\}_{i,j}$ forms an orthonormal basis of $\mathcal{H} \otimes \mathcal{K}$. Note that

$$\mathcal{H} \otimes \mathcal{K} = (|\eta\rangle \otimes |\theta\rangle)^{\perp\perp} \oplus S^{\perp\perp}$$

where $S = \{|\eta_i\rangle \otimes |\theta_j\rangle \mid (i, j) \neq (1, 1)\}$. Hence $S^{\perp\perp} = (|\eta\rangle \otimes |\theta\rangle)^\perp$. By our assumption and the defining property of the partial inner product [116, Equation (6.47)], $S \subseteq |\phi\rangle^\perp$. Hence

$$|\phi\rangle^{\perp\perp} \subseteq S^\perp = S^{\perp\perp\perp} = (|\eta\rangle \otimes |\theta\rangle)^{\perp\perp}.$$

Since these are one-dimensional subspaces, this implies that, up to global phase, $|\phi\rangle = |\eta\rangle \otimes |\theta\rangle$.

We now turn to detailed proofs of the main lemmas. For convenience, we shall repeat the statements of the lemmas.

4.2.2 The Base Case lemma

Lemma 4.1.1 (The Base Case Lemma). *Every 2-qubit state $|\psi\rangle$ is either in $\mathcal{P}_2$ (i.e. it is either a product state, or a maximally entangled bipartite state), or there are two local observables for each party, which can be computed directly from the Schmidt decomposition of $|\psi\rangle$, and which witness the logical contextuality of $|\psi\rangle$.*

Proof: This is essentially Hardy’s construction in [83]. Using the Schmidt decomposition, every two-particle entangled state can be written in the form

$$|\psi\rangle = \alpha|+\rangle_1|+\rangle_2 + \beta|-\rangle_1|-\rangle_2$$

¹This is actually the application of a linear map to a vector under Map-State duality [4].

for an appropriate choice of basis states $|\pm\rangle_i$ for each particle i , and normalized non-zero real constants α and β .

The logical contextuality of $|\psi\rangle$ is witnessed by a set of four dichotomic observables, two for each of the two parties, namely U_i and D_i , $i = 1, 2$. These observables can be defined as $U_i = |u_i\rangle\langle u_i|$ and $D_i = |d_i\rangle\langle d_i|$ where

$$|u_i\rangle = \frac{1}{\sqrt{|\alpha| + |\beta|}}(\beta^{\frac{1}{2}}|+\rangle_i + \alpha^{\frac{1}{2}}|-\rangle_i) \quad (4.2.1)$$

$$|d_i\rangle = \frac{1}{\sqrt{|\alpha|^3 + |\beta|^3}}(\beta^{\frac{3}{2}}|+\rangle_i - \alpha^{\frac{3}{2}}|-\rangle_i) \quad (4.2.2)$$

Hardy's paper also explains why it is not possible to run this particular non-locality argument when either α or β are equal to zero (product states), or when $|\alpha| = |\beta|$ (maximally entangled states), that is, when the state $|\psi\rangle$ belongs to $\mathcal{P}_2$.

There is one subtle remaining point. To show that the dichotomy asserted in the lemma is strictly disjoint, we must show that in the maximally entangled case, there is *no* choice of local observables which can give rise to logical contextuality. This is shown for the case where each party has the same two local observables as in Theorem 3.2.4 in the previous chapter, and more generally for any finite sets of local observables as Theorem 2.6.5 in [101].

4.2.3 The Going Up lemmas

The proofs of the two Going Up lemmas are quite similar. We shall prove the second, which is somewhat harder.

Lemma 4.1.3 (Going Up Lemma II). *If $|\theta\rangle = \alpha|\psi\rangle + \beta|\phi\rangle$ is a logically contextual n -qubit state under some choice of local observables, then for any non-zero $x, y \in \mathbb{C}$ such that $|x\alpha|^2 + |y\beta|^2 = 1$, the state $|\omega\rangle = x\alpha|\psi\rangle|0\rangle + y\beta|\phi\rangle|1\rangle$ is also logically contextual under the same choice of observables, augmented with a single additional observable for the $n + 1^{th}$ party.*

Proof: Since $|\theta\rangle$ is logically contextual there must be some context U' and some $s' \in S(U')$ such that the formula $\Psi = \varphi_{s'} \wedge \bigwedge_{U \in \mathcal{U} \setminus U'} \varphi_U$ is not satisfiable. We will show that it is possible to construct a similar unsatisfiable formula in order to prove the logical non-locality of $|\omega\rangle$.

The $n + 1^{th}$ party is assigned a single observable $B = B(x, y)$, whose eigenvectors are $|b_+\rangle = \bar{y}|0\rangle + \bar{x}|1\rangle$ and $|b_-\rangle = x|0\rangle - y|1\rangle$. The observable B is given by the self-adjoint matrix

$$B(x, y) = \begin{pmatrix} -|x|^2 + |y|^2 & 2x\bar{y} \\ 2\bar{x}y & |x|^2 - |y|^2 \end{pmatrix} \quad (4.2.3)$$

For any n -qubit state $|\mu\rangle$ we have

$$\begin{aligned}
\langle \mu | \langle b_+ | \cdot | \omega \rangle &= x\alpha \langle \mu | \psi \rangle \langle b_+ | 0 \rangle + y\beta \langle \mu | \phi \rangle \langle b_+ | 1 \rangle \\
&= xy\alpha \langle \mu | \psi \rangle + xy\beta \langle \mu | \phi \rangle \\
&= xy(\alpha \langle \mu | \psi \rangle + \beta \langle \mu | \phi \rangle) \\
&= xy \langle \mu | \theta \rangle
\end{aligned}$$

which implies

$$|\langle \mu | \theta \rangle|^2 = 0 \Leftrightarrow |\langle \mu | \langle b_+ | \cdot | \omega \rangle|^2 = 0 \quad (4.2.4)$$

The augmented set of allowed measurements $\tilde{X} = X \cup (B, n+1)$ is then covered by the family of compatible subsets

$$\tilde{\mathcal{U}} := \{\tilde{U} = U \cup (B, n+1) \mid U \in \mathcal{U}\}.$$

Let $T(\tilde{U})$ denote the support of $\tilde{U} \in \tilde{\mathcal{U}}$. Equation (4.2.4) implies that $S(U) = \{s \mid s+ \in T(\tilde{U})\}$ where $s+ : \tilde{U} \rightarrow \mathbf{2}$ extends s by mapping $(B, n+1)$ to $+$. As a side remark, note that the presence of an analogously defined section $\sigma-$ in $T(\tilde{U})$ does not necessarily imply the presence of σ in $S(U)$.

Now let $t' := s'+$. We have $t' \in T(\tilde{U}')$ and we can define the analogue of the proposition $\varphi_{s'}$ as

$$\begin{aligned}
\varphi_{t'} &= \bigwedge_{x \in \tilde{U}', t'(x)=+} x \wedge \bigwedge_{x \in \tilde{U}', t'(x)=-} \neg x \\
&= \bigwedge_{x \in U', s'(x)=+} x \wedge \bigwedge_{x \in U', s'(x)=-} \neg x \wedge z_{n+1} \\
&= \varphi_{s'} \wedge z_{n+1}
\end{aligned}$$

where z_{n+1} denotes the boolean variable corresponding to the outcome on the $n+1^{th}$ qubit. Recall that $+$ stands for true and $-$ stands for false.

It also holds that

$$\begin{aligned}
\varphi_{\tilde{U}} &= \bigvee_{t \in T(\tilde{U})} \varphi_t = \bigvee_{t \in T(\tilde{U}), t=s+} \varphi_t \vee \bigvee_{t \in T(\tilde{U}), t=s-} \varphi_t \\
&= \left(\bigvee_{s \in S(U)} (\varphi_s \wedge z_{n+1}) \right) \vee \left(\bigvee_{t \in T(\tilde{U}), t=\sigma-} \varphi_\sigma \wedge \neg z_{n+1} \right) \\
&= \left(\left(\bigvee_{s \in S(U)} \varphi_s \right) \wedge z_{n+1} \right) \vee \left(\underbrace{\left(\bigvee_{t=\sigma- \in T(\tilde{U})} \varphi_\sigma \right)}_{\gamma_U} \wedge \neg z_{n+1} \right) \\
&= (\varphi_U \wedge z_{n+1}) \vee (\gamma_U \wedge \neg z_{n+1})
\end{aligned}$$

We can now define the formula Ω which specifies the joint outcome $t' = s' + \in \tilde{U}'$ as well as the joint outcomes within the supports of all compatible sets of measurements $\tilde{U} \neq \tilde{U}'$. This formula is just the conjunction of Ψ and z_{n+1} . In order to show that $|\omega\rangle$ is logically contextual, it suffices to show that Ω has no satisfiable assignment. This is indeed the case, as the fact that Ψ is not satisfiable implies that Ω is also not satisfiable, thus completing our proof. Indeed, we have

$$\begin{aligned}
\Omega &= \varphi_{t'} \wedge \bigwedge_{\tilde{U} \in \tilde{\mathcal{U}} \setminus \tilde{U}'} \varphi_{\tilde{U}} \\
&= (\varphi_{s'} \wedge z_{n+1}) \wedge \bigwedge_{\tilde{U} \in \tilde{\mathcal{U}} \setminus \tilde{U}'} [(\varphi_{\tilde{U}} \wedge z_{n+1}) \vee (\gamma_U \wedge \neg z_{n+1})] \\
&= \bigwedge_{\tilde{U} \in \tilde{\mathcal{U}} \setminus \tilde{U}'} [(\varphi_{s'} \wedge z_{n+1}) \wedge ((\varphi_{\tilde{U}} \wedge z_{n+1}) \vee (\gamma_U \wedge z_{n+1}))] \\
&= \bigwedge_{\tilde{U} \in \tilde{\mathcal{U}} \setminus \tilde{U}'} [(\varphi_{s'} \wedge z_{n+1}) \wedge (\varphi_{\tilde{U}} \wedge z_{n+1})] \\
&= \varphi_{s'} \wedge \left(\bigwedge_{\tilde{U} \in \tilde{\mathcal{U}} \setminus \tilde{U}'} \varphi_{\tilde{U}} \right) \wedge z_{n+1} = \Psi \wedge z_{n+1}
\end{aligned}$$

The proof of the Going Up Lemma I follows by a similar argument, where the observables are augmented by the Z measurement for the $n+1^{th}$ party.

The Going Up lemmas have the following useful corollary.

Corollary 4.1.4. *If $|\theta\rangle$ is logically contextual under some choice of local observables, so are $|\theta\rangle \otimes |\eta\rangle$ and $|\eta\rangle \otimes |\theta\rangle$ for any state $|\eta\rangle$, with the same choice of local observables for each party*

in $|\theta\rangle$, and a single observable for each party in $|\eta\rangle$.

Proof: We argue by induction on the number of qubits in $|\eta\rangle$. If $|\eta\rangle = \alpha|0\rangle + \beta|1\rangle$, then by bilinearity of the tensor product,

$$|\theta\rangle \otimes |\eta\rangle = \alpha|\theta\rangle|0\rangle + \beta|\theta\rangle|1\rangle$$

and we can apply the Going Up Lemma I.

For the inductive case, we can write

$$|\eta\rangle = \alpha|\eta_0\rangle|0\rangle + \beta|\eta_1\rangle|1\rangle$$

and by bilinearity

$$|\theta\rangle \otimes |\eta\rangle = \alpha|\theta\rangle|\eta_0\rangle|0\rangle + \beta|\theta\rangle|\eta_1\rangle|1\rangle.$$

By induction hypothesis, $|\theta\rangle|\eta_0\rangle$ and $|\theta\rangle|\eta_1\rangle$ are logically contextual, and we can apply the Going Up Lemma I again to conclude.

4.2.4 The Small Difference lemma

Lemma 4.1.5 (The Small Difference Lemma). *Let $|\psi\rangle$ and $|\phi\rangle$ be states in $\mathcal{P}_n$, and suppose that for all $a \in [0, 1]$, $\tau(a)$ is in $\mathcal{P}_n$, where $\tau(a)$ is defined by (4.1.2). Then $|\psi\rangle$ and $|\phi\rangle$ differ in at most one qubit.*

Proof: Firstly, we note that each state $|\psi\rangle$ in $\mathcal{P}_n$ has an *entanglement type*, which can be described by a graph on n vertices with an edge from i to j when the corresponding qubits of $|\psi\rangle$ are maximally entangled. There are finitely many such graphs, and we can partition $\mathcal{P}_n$ into $P_1, \dots, P_M$ according to the entanglement type. All the states in each P_l are LU equivalent.

As before, we can write $|\psi\rangle$, up to permutation of tensor factors, as

$$|\psi\rangle = |\psi_1\rangle \otimes \dots \otimes |\psi_k\rangle \tag{4.2.5}$$

where each $|\psi_i\rangle$ is either a 1-qubit state, or a 2-qubit maximally entangled state.

We recall the definition of the parameterised family of states $\tau(a)$, $a \in [0, 1]$:

$$\tau(a) = a|\psi\rangle + g(a)|\phi\rangle$$

where $g(a) = \sqrt{1 - a^2}$.

Now for each $\delta \in [0, 1]$, we define a set

$$R_\delta = \{\tau(a) \mid 1 - \delta < a \leq 1\}$$

Under our assumption on $\tau(a)$, each set R_δ , which is infinite, is partitioned among the sets $P_1, \dots, P_M$. Also, $\delta < \delta'$ implies $R_\delta \supset R_{\delta'}$. Hence, by an application of König's infinity lemma [98], we can conclude that there is l with $1 \leq l \leq M$, and an infinite increasing sequence $\{a_i\}$ with supremum 1, such that $\tau(a_i)$ is in P_l for all i .

Since all the states $\tau(a_i)$ are LU-equivalent, we can express them in terms of a representative state $|\Theta\rangle \in P_l$ as

$$\tau(a_i) = U_{a_i}^1 \otimes U_{a_i}^2 \otimes \dots \otimes U_{a_i}^n |\Theta\rangle \quad (4.2.6)$$

Since $\mathbf{U}(2)^n$ is compact, there is a convergent subsequence $\{U_{b_i}^1 \otimes \dots \otimes U_{b_i}^n\}_{b_i}$, whose limit as $i \rightarrow \infty$ is $W^1 \otimes \dots \otimes W^n$. The limit of the corresponding subsequence $\{a_{b_i}\}$ is still 1. Hence $|\psi\rangle$ is also a member of P_l , as

$$|\psi\rangle = \lim_{i \rightarrow \infty} \tau(a_i) = \lim_{i \rightarrow \infty} \tau(a_{b_i}) = W^1 \otimes \dots \otimes W^n |\Theta\rangle.$$

This, together with Equation (4.2.6), implies that we can express each $\tau(a_i)$ as

$$\tau(a_i) = [(U_{a_i}^1 W^{1\dagger}) \otimes \dots \otimes (U_{a_i}^n W^{n\dagger})] |\psi\rangle.$$

Equivalently, using Equation 4.2.5 and the definition of $\tau(a_i)$, we can obtain a family of equations, one for each a_i :

$$\begin{aligned} a_i |\psi\rangle + g(a_i) |\phi\rangle &= [(U_{a_i}^1 W^{1\dagger}) \otimes \dots \otimes (U_{a_i}^n W^{n\dagger})] |\psi\rangle \\ &= |\psi_1^i\rangle \otimes \dots \otimes |\psi_k^i\rangle \end{aligned} \quad (4.2.7)$$

where $|\psi_j^i\rangle$ is the state obtained after the LU transformation of $|\psi_j\rangle$, $1 \leq j \leq k$.

We shall now make use of the “partial inner product” operation described e.g. in [116, p. 129]. We will use this operation to probe the components of (4.2.7).

By Lemma 4.2.3, if for all j , and for any state $|\psi_j^\perp\rangle$ orthogonal to $|\psi_j\rangle$, the application of $|\psi_j^\perp\rangle$ to $|\phi\rangle$ results in a null vector, we must have $|\phi\rangle = |\psi\rangle$, and the lemma is proved.

Otherwise, assume there is some j , and some $|\psi_j^\perp\rangle$, such that $\langle \psi_j^\perp | \phi \rangle$ is a non-zero vector. For ease of notation, we take $j = k$.

Applying $\langle \psi_k^\perp |$ on both sides of Equation (4.2.7), we obtain

$$g(a_i) \langle \psi_k^\perp | \phi \rangle = |\psi_1^i\rangle \otimes \dots \otimes |\psi_{k-1}^i\rangle \underbrace{\langle \psi_k^\perp | \psi_k^i \rangle}_{\neq 0} \quad (4.2.8)$$

$$\underbrace{\langle \psi_k^\perp | \phi \rangle}_{\text{constant vector}} = \epsilon_i |\psi_1^i\rangle \otimes \dots \otimes |\psi_{k-1}^i\rangle \quad (4.2.9)$$

The fact that the LHS of (4.2.9) is constant implies that for any i and j we must have $\epsilon_i = \epsilon_j$. We write ϵ for this common value. We must also have $|\psi_t^i\rangle = |\psi_t^j\rangle$ for all $t < k$, and we write $|\psi_t'\rangle$ for the common value. In fact we have

$$|\psi_t\rangle = \lim_{i \rightarrow \infty} |\psi_t^i\rangle = |\psi_t'\rangle, \quad 1 \leq t < k \quad (4.2.10)$$

This means that we can rewrite Equation (4.2.9) as

$$\langle \psi_k^\perp | \phi \rangle = \epsilon |\psi_1\rangle \otimes \dots \otimes |\psi_{k-1}\rangle.$$

A similar analysis will apply to any state $|\eta\rangle$ orthogonal to $|\psi_k\rangle$ for which $\langle \eta | \phi \rangle$ is non-zero. Using Equation (6.48) from [116], and bilinearity of the tensor product, we obtain

$$|\phi\rangle = x |\phi_2\rangle \otimes |\psi_k\rangle + \epsilon |\psi_1\rangle \otimes \dots \otimes |\psi_{k-1}\rangle \otimes |\xi\rangle \quad (4.2.11)$$

for some $|\phi_2\rangle$ and $|\xi\rangle$. We can use Equation (4.2.10) to rewrite Equation (4.2.7) as

$$a_i |\psi\rangle + g(a_i) |\phi\rangle = |\psi_1\rangle \otimes \dots \otimes |\psi_{k-1}\rangle \otimes |\psi_k^i\rangle.$$

For any $j < k$, if we apply any state $|\psi_j^\perp\rangle$, orthogonal to $|\psi_j\rangle$, to the above equation we obtain $\langle \psi_j^\perp | \phi \rangle = 0$. Together with Equation (4.2.11), this implies that $|\phi_2\rangle = |\psi_1\rangle \otimes \dots \otimes |\psi_{k-1}\rangle$. So $|\phi\rangle$ and $|\psi\rangle$ can differ by at most one component.

These components cannot be two-qubit maximally entangled states, since by assumption all linear combinations $\tau(a)$ of $|\psi\rangle$ and $|\phi\rangle$, with $a \in [0, 1]$, belong to $\mathcal{P}_n$, and hence, using bilinearity again, the corresponding linear combinations of these components would also have to be maximally entangled, yielding a contradiction.

Thus we conclude that $|\psi\rangle$ and $|\phi\rangle$ can differ at most in a one-qubit component.

4.2.5 The 21 lemma

We shall need some elementary facts about partial traces (see e.g. [107]):

- A pure state in $\mathcal{H} \otimes \mathcal{K}$ is a product state if and only if tracing out over $\mathcal{H}$ results in a pure state.
- Tracing out over one party of a maximally entangled bipartite state yields a maximally mixed state.
- A mixed state ρ is pure if and only if $\text{Tr} \rho^2 = 1$.

Lemma 4.1.6 (The 21 Lemma). *With the same notation as in the Small Difference Lemma, suppose that $\tau(a)$ is in $\mathcal{P}_n$ for 21 distinct values of a in $[0, 1]$. Then $\tau(a)$ is in $\mathcal{P}_n$ for all $a \in [0, 1]$.*

Proof:

If a state belongs to $\mathcal{P}_n$ then all partial traces over $n - 1$ parties result either in a pure state or in the maximally mixed state $\frac{1}{2}I_2$.

We can express $|\psi\rangle$ and $|\phi\rangle$ as

$$\begin{aligned} |\psi\rangle &= \sum_{\sigma_i} a_{\sigma_i}^0 |\sigma_i\rangle |0\rangle + \sum_{\sigma_i} a_{\sigma_i}^1 |\sigma_i\rangle |1\rangle \\ |\phi\rangle &= \sum_{\sigma_i} b_{\sigma_i}^0 |\sigma_i\rangle |0\rangle + \sum_{\sigma_i} b_{\sigma_i}^1 |\sigma_i\rangle |1\rangle \end{aligned}$$

where the σ_i index the elements of the computational basis on $n - 1$ qubits.

This means we can write the density matrix corresponding to $|\tau(a)\rangle = a|\phi\rangle + b|\psi\rangle$, with $b = g(a)$, as

$$\begin{aligned} |\tau(a)\rangle\langle\tau(a)| &= \sum_{\sigma_i, \sigma_j} (aa_{\sigma_i}^0 + bb_{\sigma_i}^0)(\overline{aa_{\sigma_j}^0} + \overline{bb_{\sigma_j}^0}) |\sigma_i\rangle\langle\sigma_j| \otimes |0\rangle\langle 0| \\ &+ \sum_{\sigma_i, \sigma_j} (aa_{\sigma_i}^0 + bb_{\sigma_i}^0)(\overline{aa_{\sigma_j}^1} + \overline{bb_{\sigma_j}^1}) |\sigma_i\rangle\langle\sigma_j| \otimes |0\rangle\langle 1| \\ &+ \sum_{\sigma_i, \sigma_j} (aa_{\sigma_i}^1 + bb_{\sigma_i}^1)(\overline{aa_{\sigma_j}^0} + \overline{bb_{\sigma_j}^0}) |\sigma_i\rangle\langle\sigma_j| \otimes |1\rangle\langle 0| \\ &+ \sum_{\sigma_i, \sigma_j} (aa_{\sigma_i}^1 + bb_{\sigma_i}^1)(\overline{aa_{\sigma_j}^1} + \overline{bb_{\sigma_j}^1}) |\sigma_i\rangle\langle\sigma_j| \otimes |1\rangle\langle 1| \end{aligned}$$

The partial trace over the first $n - 1$ qubits of $|\tau(a)\rangle$ is given by

$$\begin{aligned}
\rho_n &= \text{Tr}_{n-1} |\tau(a)\rangle \langle \tau(a)| = \sum_{\sigma_i} \langle \sigma_i | \tau(a) \rangle \langle \tau(a) | \sigma_i \rangle \\
&= \sum_{\sigma_i} (aa_{\sigma_i}^0 + bb_{\sigma_i}^0)(\overline{aa_{\sigma_i}^0} + \overline{bb_{\sigma_i}^0}) |0\rangle \langle 0| + \sum_{\sigma_i} (aa_{\sigma_i}^0 + bb_{\sigma_i}^0)(\overline{aa_{\sigma_i}^1} + \overline{bb_{\sigma_i}^1}) |0\rangle \langle 1| \\
&\quad + \sum_{\sigma_i} (aa_{\sigma_i}^1 + bb_{\sigma_i}^1)(\overline{aa_{\sigma_i}^0} + \overline{bb_{\sigma_i}^0}) |1\rangle \langle 0| + \sum_{\sigma_i} (aa_{\sigma_i}^1 + bb_{\sigma_i}^1)(\overline{aa_{\sigma_i}^1} + \overline{bb_{\sigma_i}^1}) |1\rangle \langle 1| \\
&= |0\rangle \langle 0| \sum_{\sigma_i} (a^2 a_{\sigma_i}^0 \overline{a_{\sigma_i}^0} + (1 - a^2) b_{\sigma_i}^0 \overline{b_{\sigma_i}^0} + a \sqrt{1 - a^2} (a_{\sigma_i}^0 \overline{b_{\sigma_i}^0} + b_{\sigma_i}^0 \overline{a_{\sigma_i}^0})) \\
&\quad + |0\rangle \langle 1| \sum_{\sigma_i} (a^2 a_{\sigma_i}^0 \overline{a_{\sigma_i}^1} + (1 - a^2) b_{\sigma_i}^0 \overline{b_{\sigma_i}^1} + a \sqrt{1 - a^2} (a_{\sigma_i}^0 \overline{b_{\sigma_i}^1} + b_{\sigma_i}^0 \overline{a_{\sigma_i}^1})) \\
&\quad + |1\rangle \langle 0| \sum_{\sigma_i} (a^2 a_{\sigma_i}^1 \overline{a_{\sigma_i}^0} + (1 - a^2) b_{\sigma_i}^1 \overline{b_{\sigma_i}^0} + a \sqrt{1 - a^2} (a_{\sigma_i}^1 \overline{b_{\sigma_i}^0} + b_{\sigma_i}^1 \overline{a_{\sigma_i}^0})) \\
&\quad + |1\rangle \langle 1| \sum_{\sigma_i} (a^2 a_{\sigma_i}^1 \overline{a_{\sigma_i}^1} + (1 - a^2) b_{\sigma_i}^1 \overline{b_{\sigma_i}^1} + a \sqrt{1 - a^2} (a_{\sigma_i}^1 \overline{b_{\sigma_i}^1} + b_{\sigma_i}^1 \overline{a_{\sigma_i}^1}))
\end{aligned}$$

The partial trace ρ_n is equal to the maximally mixed state if and only if

$$\begin{aligned}
1/2 &= \sum_{\sigma_i} (a^2 a_{\sigma_i}^0 \overline{a_{\sigma_i}^0} + (1 - a^2) b_{\sigma_i}^0 \overline{b_{\sigma_i}^0} + a \sqrt{1 - a^2} (a_{\sigma_i}^0 \overline{b_{\sigma_i}^0} + b_{\sigma_i}^0 \overline{a_{\sigma_i}^0})) \\
0 &= \sum_{\sigma_i} (a^2 a_{\sigma_i}^0 \overline{a_{\sigma_i}^1} + (1 - a^2) b_{\sigma_i}^0 \overline{b_{\sigma_i}^1} + a \sqrt{1 - a^2} (a_{\sigma_i}^0 \overline{b_{\sigma_i}^1} + b_{\sigma_i}^0 \overline{a_{\sigma_i}^1})) \\
0 &= \sum_{\sigma_i} (a^2 a_{\sigma_i}^1 \overline{a_{\sigma_i}^0} + (1 - a^2) b_{\sigma_i}^1 \overline{b_{\sigma_i}^0} + a \sqrt{1 - a^2} (a_{\sigma_i}^1 \overline{b_{\sigma_i}^0} + b_{\sigma_i}^1 \overline{a_{\sigma_i}^0})) \\
1/2 &= \sum_{\sigma_i} (a^2 a_{\sigma_i}^1 \overline{a_{\sigma_i}^1} + (1 - a^2) b_{\sigma_i}^1 \overline{b_{\sigma_i}^1} + a \sqrt{1 - a^2} (a_{\sigma_i}^1 \overline{b_{\sigma_i}^1} + b_{\sigma_i}^1 \overline{a_{\sigma_i}^1}))
\end{aligned}$$

Each of these equations yields a polynomial of degree 4 in a . Indeed, each equation has the form

$$cab + q(a) = d$$

where c and d are constants, $b = \sqrt{1 - a^2}$, and $q(a)$ is a quadratic polynomial in a . We can write this as

$$cab = -q(a) + d$$

and square both sides to obtain

$$c^2 a^2 (1 - a^2) = (-q(a) + d)^2$$

which is a quartic polynomial in a . Hence, there can be at most 4 values of a in $[0, 1]$ for which the partial trace ρ_n is equal to a maximally mixed state.

On the other hand, ρ_n is equal to a pure state if and only if $\text{Tr}\rho_n^2 = 1$. By a similar analysis, this condition turns out to be equivalent to a polynomial equation of degree 16 in a . Unless the polynomial is degenerate, *i.e.* the coefficients cancel so that the equation reduces to $1 = 1$, there can be at most 16 values of a for which the partial trace ρ_n is equal to a pure state.

Therefore, if there are more than $4 + 16 = 20$ values of a in $[0, 1]$ for which the linear combination $\tau(a)$ belongs to $\mathcal{P}_n$, we can conclude that one of the polynomial equations above was degenerate, hence $\tau(a) \in \mathcal{P}_n$ for all values of a .

Remark If $|\psi\rangle$ and $|\phi\rangle$ differ by one qubit, the partial trace of $\tau(a)$ for any value of a will yield a pure state, and hence we will always have $\text{Tr}\rho_n^2 = 1$. Thus the polynomial will indeed be degenerate in this case. This shows the necessity for the Small Difference Lemma.

4.3 The algorithm

We now give an explicit, albeit informal description of the algorithm which follows straightforwardly from our results.

We begin with a subroutine which we will use to test if a state is in $\mathcal{P}_n$.

SUBROUTINE **Test** $\mathcal{P}_n$

Input n -qubit quantum state $|\theta\rangle$

Output Either

Yes, and entanglement type of $|\theta\rangle$, or

No

1. Compute the $n - 1$ partial traces ρ_i over $n - 1$ qubits of $|\theta\rangle$. If any ρ_i is not a maximally mixed state, compute $\text{Tr}\rho_i^2$. If $\text{Tr}\rho_i^2 \neq 1$, return **No**.

We now have the list $\{i_1, \dots, i_k\}$ of indices for which the maximally mixed state was returned.

2. For each i_p in the list, find its “partner” i_q by computing the partial traces ρ_{i_p, i_q} over $n - 2$ qubits, and then testing if $\text{Tr}\rho_{i_p, i_q}^2 = 1$.

If we cannot find the partner for some i_p , return **No**.

3. Otherwise, we return **Yes**. We also have the complete entanglement type of $|\theta\rangle$, and we have computed all the single-qubit components. $\square$

ALGORITHM

Input An n -qubit state $|\omega\rangle$
Output Either
 Yes if $|\omega\rangle$ is logically contextual,
 together with a list of $n + 2$ local observables, or
 No if $|\omega\rangle$ is in $\mathcal{P}_n$.

Base Cases

1. If $n = 1$, output **No**.
2. If $n = 2$, apply the Hardy procedure of the Base Case Lemma to the Schmidt decomposition of $|\omega\rangle$.

Recursive Case: $n + 1, n > 1$

1. We apply $\text{Test}\mathcal{P}_{n+1}$ to $|\omega\rangle$. If $|\omega\rangle$ is in $\mathcal{P}_{n+1}$, return **No**.
2. Otherwise, we write

$$|\omega\rangle = \alpha |\psi\rangle |0\rangle + \beta |\phi\rangle |1\rangle.$$

Explicitly, if $|\omega\rangle$ is represented by a 2^{n+1} -dimensional complex vector

$$\sum_{\sigma \in \{0,1\}^{n+1}} a_{\sigma} |\sigma\rangle$$

in the computational basis, we can define

$$\alpha = \sqrt{\sum_{\sigma \in \{0,1\}^n} |a_{\sigma 0}|^2}, \quad \beta = \sqrt{\sum_{\sigma \in \{0,1\}^n} |a_{\sigma 1}|^2}$$

$$|\psi\rangle = \frac{1}{\alpha} \sum_{\sigma \in \{0,1\}^n} a_{\sigma 0} |\sigma\rangle, \quad |\phi\rangle = \frac{1}{\beta} \sum_{\sigma \in \{0,1\}^n} a_{\sigma 1} |\sigma\rangle.$$

3. We apply $\text{Test}\mathcal{P}_n$ to $|\psi\rangle$. If $|\psi\rangle$ is not in $\mathcal{P}_n$, we proceed recursively with $|\psi\rangle$, and then extend the observables using the construction of the Going Up Lemma I.
4. Otherwise, we proceed similarly with $|\phi\rangle$.

5. Otherwise, both $|\psi\rangle$ and $|\phi\rangle$ are in $\mathcal{P}_n$.

For a in $(0, 1)$, we define

$$\tau(a) := a|\psi\rangle + \sqrt{1-a^2}|\phi\rangle.$$

For 19 distinct values in $(0, 1)$, we assign these values to a , and apply $\text{Test}\mathcal{P}_n$ to $\tau(a)$.

If we find a value of a for which $\tau(a)$ is not in $\mathcal{P}_n$, we use that value to compute the local observable $B(\frac{\alpha}{a}, \frac{\beta}{\sqrt{1-a^2}})$ for the $n+1^{\text{th}}$ party, as specified in the Going Up Lemma II, and continue the recursion with the n -qubit state $\tau(a)$.

6. Otherwise, by the 21 Lemma and the Small Difference Lemma, the only remaining case is where $|\psi\rangle$ and $|\phi\rangle$ differ in one qubit. We have these qubits $|\psi_1\rangle, |\phi_1\rangle$ from our previous applications of $\text{Test}\mathcal{P}_n$. In this final case, we can write $|\omega\rangle$ as

$$|\omega\rangle = |\Psi\rangle \otimes |\xi\rangle$$

where $|\Psi\rangle$ is in $\mathcal{P}_{n-1}$, and $|\xi\rangle$ is a 2-qubit state. Moreover, we have

$$|\xi\rangle = \alpha|\psi_1\rangle|0\rangle + \beta|\phi_1\rangle|1\rangle.$$

7. We apply the Base Case procedure to $|\xi\rangle$, which we know cannot be maximally entangled, by Step 1. We output Yes, together with the two local observables for each party produced by the Hardy construction, and the $n-2$ local observables for $|\Psi\rangle$ produced by the Corollary to the Going Up lemmas. $\square$

The above algorithm of course involves computation over the real and complex numbers. More precisely, with the usual coding of complex numbers as pairs of reals, we require the field operations and comparison tests on real numbers. For simplicity, we discuss the complexity of the algorithm in the Blum-Shub-Smale model of computation [21], where we assume that arbitrary real numbers can be stored, and the above operations performed, with unit cost. Thus the input size of an n -qubit state is the dimension $d = 2^n$.

The $\text{Test}\mathcal{P}_n$ subroutine performs $n-1$ partial traces over $n-1$ qubits. Each such partial trace involves computing the 4 entries of a matrix, where each entry is a sum over 2^{n-1} products. It also computes $O(n^2)$ partial traces over $n-2$ qubits, each of which involves computing 16 entries, each a sum over 2^{n-2} products. For an n -qubit input, at each level of the recursion, we call the subroutine a number of times bounded by a constant, and the recursion terminates in $O(n)$ steps. Thus we obtain a complexity bound of $O(d \log^3 d)$ operations, in the input size d .

Of course, in practice the limiting factor is the exponential size of the classical representation of a quantum state.

The algorithm has been implemented in `Mathematica`TM, and has been tested on input states of up to 10 qubits.

Part II

Second part

Chapter 5

Overview of the Topos Approach

In this chapter we shall gradually build up the necessary mathematical machinery needed to understand the main topos theoretic constructions. The abstract mathematics is presented first, and we have limited our exposition of category theoretical abstractions to those notions which are absolutely indispensable for understanding the concept of an elementary topos and the categorial definition of Gelfand duality, plus a few illuminating examples and comments. The basic category theoretical concepts introduced here can be found in any standard textbook, such as [77, 99] and, for the sake of clarity, we have chosen to follow the slightly more simplified presentation in [77].

We then proceed to give the technical definitions of those physical concepts which the topos approach seeks to model, again restricting ourselves to those which are absolutely necessary for the understanding of the following chapters.

5.1 The Basics

5.1.1 Categories and functors

A category may be thought of in the first instance as a universe for a particular kind of mathematical discourse. In axiomatic terms a category $\mathcal{C}$ consists of

- a) a class $Ob(\mathcal{C})$ of **objects**
- b) for each $A, B \in Ob(\mathcal{C})$ a class $\mathcal{C}(A, B)$ of **morphisms** from A to B
- c) for each $A, B, D \in Ob(\mathcal{C})$ a binary operation (called **composition of morphisms**)

$$\begin{aligned}\mathcal{C}(B, D) \times \mathcal{C}(A, B) &\longrightarrow \mathcal{C}(A, D) \\ (g, f) &\mapsto g \circ f\end{aligned}$$

which satisfies the following axioms

(i) **identity:** for each $A \in \text{Ob}(\mathcal{C})$ there exists an identity $1_A \in \mathcal{C}(A, A)$ such that

$$\begin{aligned} f \circ 1_A &= f, & \forall f \in \mathcal{C}(A, B) \\ 1_A \circ f &= f, & \forall f \in \mathcal{C}(B, A) \end{aligned}$$

(ii) **functoriality:** for any $A, B, D, E \in \text{Ob}(\mathcal{C})$, $f \in \mathcal{C}(A, B)$, $g \in \mathcal{C}(B, D)$, $h \in \mathcal{C}(D, E)$

$$h \circ (g \circ f) = (h \circ g) \circ f$$

The opposite category of $\mathcal{C}$ is denoted by $\mathcal{C}^{op}$ and it has the same objects as $\mathcal{C}$ but the directions of the arrows are reversed.

Definition 5.1.1. An arrow $m : C \rightarrow D$ is monic if for any pair of arrows $f, g : B \rightrightarrows C$ if $m \circ f = m \circ g$ then $f = g$.

Examples

1. One of the most familiar examples of a category is the category **Set** whose objects are sets and whose arrows are functions between sets.
2. If P is a partially ordered set, we can regard it as a category in which the objects are the elements of P and the arrows correspond to the partial order on P . Thus if $x \leq y$ in the poset P we have an arrow $x \rightarrow y$ in the poset P seen as a category.

Definition 5.1.2. Given two categories $\mathcal{C}$ and $\mathcal{D}$, a covariant functor $F : \mathcal{C} \rightarrow \mathcal{D}$ is given by:

- a) a map $F : \text{Ob}(\mathcal{C}) \rightarrow \text{Ob}(\mathcal{D})$
- b) for all $A, B \in \text{Ob}(\mathcal{C})$, a map $F : \mathcal{C}(A, B) \rightarrow \mathcal{D}(FA, FB)$ satisfying the principles of
 - i) **identity preservation:** $F(1_A) = 1_{FA}$
 - ii) **functoriality preservation:** $F(f \circ g) = Ff \circ Fg$

A contravariant functor $F : \mathcal{C} \rightarrow \mathcal{D}$ is a covariant functor from $\mathcal{C}^{op}$ to $\mathcal{D}$. Functors can be composed in an obvious manner.

A contravariant **Set**-valued functor defined on a category $\mathcal{C}$ is called a presheaf over the base category $\mathcal{C}$. Note that this can also be regarded as a covariant functor defined on $\mathcal{C}^{op}$ with values in **Set**.

Definition 5.1.3. If F and G are (covariant) functors between the categories $\mathcal{C}$ and $\mathcal{D}$, then a natural transformation η from F to G associates to every object X in $\mathcal{C}$ a morphism $\eta_X :$

$F(X) \rightarrow G(X)$ in $\mathcal{D}$ such that for every morphism $f : X \rightarrow Y$ in $\mathcal{C}$ we have $\eta_Y \circ F(f) = G(f) \circ \eta_X$; this means that the following diagram is commutative:

$$\begin{array}{ccc} F(X) & \xrightarrow{F(f)} & F(Y) \\ \eta_X \downarrow & & \downarrow \eta_Y \\ G(X) & \xrightarrow{G(f)} & G(Y) \end{array}$$

For contravariant functors the arrows $F(f)$ and $G(f)$ in the definition above will be reversed. The two functors F and G are called naturally isomorphic if there exists a natural transformation from F to G such that η_X is an isomorphism for every object X in $\mathcal{C}$.

Natural transformations act on functors just like functors act on objects of a category. Hence the collection of functors between two categories together with the natural transformations between them is a category in itself. A notable example is the collection $\mathbf{Set}^{\mathcal{C}^{op}}$ of all presheaves $\mathbf{Set}$ -valued presheaves over $\mathcal{C}$.

Definition 5.1.4. If $\mathcal{C}$ and $\mathcal{D}$ are categories then functors $F : \mathcal{C} \rightarrow \mathcal{D}$ and $G : \mathcal{D} \rightarrow \mathcal{C}$ form an adjunction if for any A in $\mathcal{C}$ and B in $\mathcal{D}$, there an isomorphism

$$\theta_{A,B} : \mathcal{D}(FA, B) \rightarrow \mathcal{C}(A, GB)$$

which is natural in both A and B . In this case we say that F is the left adjoint of G , or equivalently that G is the right adjoint of F .

What is meant by ‘naturality in A ’ in the definition above is that given any A' in $\mathcal{C}$ and any arrow $h : A' \rightarrow A$ the following diagram commutes:

$$\begin{array}{ccc} \mathcal{D}(FA, B) & \xrightarrow{\theta_{A,B}} & \mathcal{C}(A, GB) \\ (Fh)^* \downarrow & & \downarrow h^* \\ \mathcal{D}(FA', B) & \xrightarrow{\theta_{A',B}} & \mathcal{C}(A', GB) \end{array}$$

where h^* denotes composition on the right by h . Similarly, for ‘naturality in B ’.

Definition 5.1.5. Given two categories $\mathcal{C}$ and $\mathcal{D}$, an equivalence of categories consists of a functor $F : \mathcal{C} \rightarrow \mathcal{D}$, a functor $G : \mathcal{D} \rightarrow \mathcal{C}$, and two natural isomorphisms $\epsilon : FG \rightarrow I_{\mathcal{D}}$ and $\eta : I_{\mathcal{C}} \rightarrow GF$. Here $FG : \mathcal{D} \rightarrow \mathcal{D}$ and $GF : \mathcal{C} \rightarrow \mathcal{C}$ denote the respective compositions of F and

G , while $I_{\mathcal{C}} : \mathcal{C} \rightarrow \mathcal{C}$ and $I_{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{D}$ denote the identity functors on $\mathcal{C}$ and $\mathcal{D}$, assigning each object and morphism to itself. If F and G are contravariant functors one speaks of a duality of categories instead.

5.1.2 Limits and colimits

It is possible to define limits and colimits in a category $\mathcal{C}$ by means of diagrams in $\mathcal{C}$. A diagram of type J in $\mathcal{C}$ is a functor from J to $\mathcal{C}$:

$$F : J \rightarrow \mathcal{C}$$

The category J is thought of as index category, and the diagram F is thought of as indexing a collection of objects and morphisms in $\mathcal{C}$ patterned on J . The actual objects and morphisms in J are largely irrelevant, only the way in which they are interrelated matters.

A diagram is said to be finite whenever J is.

Definition 5.1.6. Let $F : J \rightarrow \mathcal{C}$ be a diagram of type J in a category $\mathcal{C}$. A cone to F is an object N of $\mathcal{C}$ together with a family $\psi_X : N \rightarrow F(X)$ of morphisms indexed by the objects of J , such that for every morphism $f : X \rightarrow Y$ in J , we have $F(f) \circ \psi_X = \psi_Y$:

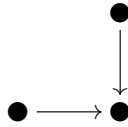
$$\begin{array}{ccc} & N & \\ \psi_X \swarrow & & \searrow \psi_Y \\ F(X) & \xrightarrow{F(f)} & F(Y) \end{array}$$

Definition 5.1.7. A limit of the diagram $F : J \rightarrow \mathcal{C}$ is a cone (L, φ) to F such that for any other cone (N, ψ) to F there exists a unique morphism $!_N : N \rightarrow L$ such that $\varphi_X \circ !_N = \psi_X$ for all X and Y in J :

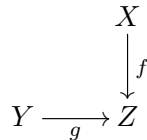
$$\begin{array}{ccc} & N & \\ \psi_X \swarrow & \downarrow !_N & \searrow \psi_Y \\ & L & \\ \varphi_X \swarrow & & \searrow \varphi_Y \\ F(X) & \xrightarrow{F(f)} & F(Y) \end{array}$$

The dual notions of limits and cones are colimits and co-cones. It is straightforward to obtain the definitions of these by inverting all morphisms in Definitions 5.1.6 and 5.1.7.

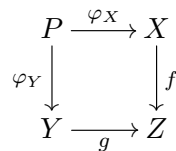
Important examples of limits include pullbacks, products and terminal objects, while examples of colimits include coproducts and initial objects. So, for instance, a pullback (P, φ) is the limit of a diagram F of type



in a category $\mathcal{C}$. More explicitly, if the image of F in $\mathcal{C}$ is



we say that the object P , together with the family of arrows $\varphi = \{\varphi_X, \varphi_Y\}$, giving the limit of this diagram is the pullback of f and g , and we call the diagram below



a pullback square.

A product is the limit (P, φ) of a diagram F of type $\bullet \bullet$, while a coproduct is the colimit of a diagram of the same type. More explicitly, if the image of F in $\mathcal{C}$ is $A \rightarrow B$ we say that the object P , together with the family of arrow $\varphi = \{\varphi_A, \varphi_B\}$, giving the limit of the diagram is the product of A and B . Similarly, the object, and arrows, giving the colimit of the same diagram is called the coproduct of A and B .

A terminal object $\mathbf{1}$ in a category $\mathcal{C}$ is the limit of the empty diagram in $\mathcal{C}$. Thus the terminal object has the property that for any other object N in $\mathcal{C}$ there is a unique arrow $!_N$ going from N to $\mathbf{1}$. Similarly, an initial object $\mathbf{0}$ in a category $\mathcal{C}$ is the colimit of the empty diagram in $\mathcal{C}$ and there is a unique arrow going from $\mathbf{0}$ to each N in $\mathcal{C}$.

Examples

1. The terminal object in **Set** is the one-element set.
2. The terminal object in **Set**^{*op*} is the functor (presheaf) which assigns the one element set $\{*\}$ to all objects in $\mathcal{C}$ and the identity on $\{*\}$ to every $\mathcal{C}$ -arrow.

5.1.3 Lattices and Heyting algebras

A lattice is a partially ordered set in which any two elements have a least upper bound (also called a join) as well as a greatest lower bound (also called a meet). The symbols $\wedge$ and $\vee$ are used to denote these two operations. A lattice can also be defined in categorical language as a poset which, when seen as a category, has all finite limits and colimits. From this point of view, the lattice operations $\wedge$ and $\vee$ correspond to the categorical notions of product and coproduct.

A morphism of lattices is a monotone function which preserves both (binary) meets and (binary) joins.

A lattice is called complete if it has all limits and colimits. A complete lattice has an initial element given by the empty join, which we denote by $\mathbf{0}$, and dually, a terminal element given by the empty meet, which we denote by $\mathbf{1}$.

Definition 5.1.8. *An element x of a lattice L with global minimum $\mathbf{0}$ is called an **atom** if any element y in L which is strictly smaller than x is equal to $\mathbf{0}$. A lattice is **atomic** if for every non-zero element $y \in L$ there is some atom x such that $x \leq y$. An **atomistic** lattice is an atomic lattice such that each element is a join of atoms.*

Definition 5.1.9. *An orthocomplemented lattice or ortholattice is a bounded lattice which is equipped with a function $\perp$ that maps each element x to an orthocomplement $x^\perp$ in such a way that the following axioms are satisfied:*

- (i) *complement law: $x^\perp \vee x = \mathbf{1}$ and $x^\perp \wedge x = \mathbf{0}$*
- (ii) *involution law: $x^{\perp\perp} = x$*
- (iii) *order-reversing: if $x \leq y$ then $y^\perp \leq x^\perp$.*

We call a lattice complemented if it is equipped with a (not necessarily unique) function that satisfies only the first of the three conditions above. We say that a lattice L is distributive if, for all $x, y, z \in L$,

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$$

and hence also

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$$

We say that two elements x and y of L are orthogonal if $x \leq y^\perp$.

Definition 5.1.10. *A Boolean algebra is a complemented distributive lattice (and thereby also orthocomplemented).*

Stone's representation theorem [120] for Boolean algebras states that every Boolean algebra is isomorphic to the Boolean algebra of all clopen subsets of some (compact totally disconnected Hausdorff) topological space called the Stone space.

Boolean algebras are intimately connected with classical propositional logic. The Lindenbaum algebra given by the set of sentences in propositional calculus modulo tautology has the structure of a Boolean algebra [121]. A truth assignment in propositional calculus is then a homomorphism from the Lindenbaum algebra to the two-element Boolean algebra.

Definition 5.1.11. *A Heyting algebra is a lattice H containing a bottom element with the property that for any two elements $x, y \in H$ there exists an exponential $x \Rightarrow y$ characterised by*

$$z \leq (x \Rightarrow y) \iff z \wedge x \leq y$$

A Heyting algebra is called complete if the underlying lattice is complete.

The exponential $x \Rightarrow y$ is also referred to as the pseudo-complement of x relative to y . When y is equal to the bottom element, $x \Rightarrow \mathbf{0}$ is referred to as the pseudo-complement of x . In a Boolean algebra, the pseudo-complement is actually the complement of an element, so every Boolean algebra is also a Heyting algebra. However, while the definition of the exponential implies that the meet of an element and its pseudo-complement is always equal to $\mathbf{0}$, their join may be less than $\mathbf{1}$. This means that there are Heyting algebras which are not Boolean.

It can be shown that the underlying lattice of a Heyting algebra is distributive. Moreover, if the underlying lattice is complete, then for any family $(h_i)_{i \in I}$ of elements of H , the following infinite distributivity law holds:

$$\forall h \in H, \quad h \wedge \bigvee_{i \in I} h_i = \bigvee_{i \in I} (h \wedge h_i)$$

Just like classical propositional logic is modelled by Boolean algebras, the Lindenbaum algebra of intuitionistic logic has the structure of a Heyting algebra. Intuitionistic logic can have more than two truth values, as we will see in the following section.

5.1.4 Topoi

An (elementary) topos is a category which is similar to the category of sets and functions in a sense which will be made precise in this section. The elementary topoi which we will be considering in this thesis are all Grothendieck topoi of **Set**-valued presheaves over a base category given by a partially ordered set. We will pay special attention to topoi of presheaves throughout the remainder of this section.

We start our discussion by extending the notion of subset to a general category. Thus we define, as in [77], a subobject of an object A to be an object S from the same category such that there exists a monic $m : S \rightarrow A$. In **Set** the collection of subobjects of a given set A is equivalent to the power set of A . The set-theoretic operations of union, intersection and complement give the structure of a Boolean algebra to this collection of subobjects.

In $\mathbf{Set}^{\mathcal{C}^{op}}$, a subobject of a presheaf $\underline{A}$ is another presheaf $\underline{S}$ which associates a subset $\underline{S}_C$ of $\underline{A}_C$ to every object of $\mathcal{C}$ in a way which is compatible with the restrictions $\underline{A}(f)|_{\underline{S}_C}$ for all arrows $C \xrightarrow{f} D$ in $\mathcal{C}^{op}$. In other words, there is a natural transformation between the presheaves $\underline{A}$ and $\underline{S}$ given by the collection of monics:

$$\begin{array}{ccc} \underline{S}_C & \xrightarrow{\underline{A}(f)|_{\underline{S}_C}} & \underline{S}_D \\ \downarrow & & \downarrow \\ \underline{A}_C & \xrightarrow{\underline{A}(f)} & \underline{A}_D \end{array}$$

A point of an object A in a category $\mathcal{C}$ is an arrow (which is necessarily monic) from the terminal object in $\mathcal{C}$ to A . For the category **Set** the terminal object is the one object set $\{*\}$ and so the categorical notion of point coincides with the familiar notion of an element of a given set. In $\mathbf{Set}^{\mathcal{C}^{op}}$ a point of a presheaf is a collection of set-theoretic points which is compatible with the restriction maps. It is also called a global section.

It is a well-known fact that in **Set**, a subset $S \subseteq A$ is uniquely determined by its characteristic function ξ_S . This connection can be generalised to arbitrary categories, as long as they contain what is known as a subobject classifier.

Definition 5.1.12. *In a category $\mathcal{C}$ with finite limits, a subobject classifier is an object Ω and a monic $true : \mathbf{1} \rightarrow \Omega$ going from the terminal object in $\mathcal{C}$ to Ω such that for any subobject $m : S \rightarrow A$ there is a unique arrow $\chi_S : A \rightarrow \Omega$ making the following square a pullback:*

$$\begin{array}{ccc}
S & \xrightarrow{!_S} & \mathbf{1} \\
m \downarrow & & \downarrow true \\
A & \xrightarrow{\chi_S} & \Omega
\end{array}$$

One can readily check that in **Set** the subobject classifier is given by $\Omega = \{0, 1\}$ together with the map $true(*) = 1$, and χ_S above is simply the characteristic function corresponding to S .

Using the fact that the pullback of a monic is itself monic one can prove that, in any category where such constructions are possible, there is a bijective correspondence between the collection of subobjects of an object A and the collection of arrows between A and Ω . In particular

$$Sub(\mathbf{1}) \cong \mathcal{C}(1, \Omega)$$

The collection of subobjects of a given object can be given the structure of a Heyting algebra using a suitable generalisation of the set-theoretic operations of union, intersection and complement (see, for example Ch. 7 in [77]). This means that the category theoretic points of Ω are in bijective correspondence with the subobjects of the terminal object, and as such they form a Heyting algebra. The points of Ω can be thought of as the truth-values of the intuitionistic logic associated with the topos in which Ω is defined.

We now describe the subobject classifier in $\mathbf{Set}^{\mathcal{C}^{op}}$. For this we need to introduce the notion of sieve. For a $\mathcal{C}$ -object A , let $\mathfrak{S}_A$ be the collection of all $\mathcal{C}$ -arrows with co-domain A . Note that $\mathfrak{S}_A$ is “closed under right composition”, i.e. if an arrow $B \xrightarrow{g} C$ in $\mathcal{C}^{op}$ can be composed with $f \in \mathfrak{S}_A$ then $f \circ g \in \mathfrak{S}_A$ since the arrow $f \circ g$ has co-domain A .

Definition 5.1.13. *A sieve on A is a subset $\mathfrak{s}$ of $\mathfrak{S}_A$ which is itself closed under right composition.*

Note that for every object A there are at least two A -sieves $\mathfrak{S}_A$ and $\emptyset$, the empty sieve.

Now we are ready to define the subobject classifier in $\mathbf{Set}^{\mathcal{C}^{op}}$ as the presheaf Ω which associates the set of sieves on A to every $\mathcal{A}$ -object A . To each arrow $A \xrightarrow{f} B$ in $\mathcal{C}$, the

subobject classifier associates a function

$$\Omega(f) : \Omega(B) \longrightarrow \Omega(A)$$

$$\mathfrak{s} \longmapsto \mathfrak{s}_A = \{g \in \mathfrak{S}_A \mid f \circ g \in \mathfrak{s}\}$$

The set $\mathfrak{s}_A$ is closed under right composition because if h can be composed with some $g \in \mathfrak{s}_A$ then $f \circ g \circ h \in \mathfrak{s}$ because sieves on B are closed under right composition, and so $g \circ h \in \mathfrak{s}_A$, as required.

The arrow $true : \mathbf{1} \rightarrow \Omega$ is defined as the natural transformation that has components $true_A : \{*\} \rightarrow \Omega(A)$ given by $true_A(*) = \mathfrak{S}_A$.

Examples

1. In the limiting case when the category $\mathcal{C}$ has only one object A and only one arrow, namely the identity on A , there are precisely two A -sieves: $\mathfrak{S}_A = \{id_A\}$ and $\emptyset$. In this case, the definition of the subobject classifier in $\mathbf{Set}^{\mathcal{C}^{op}}$ reduces to the classical one, as expected.

2. When the category $\mathcal{C}$ is a poset, there is an arrow from A to B only when $A \leq B$. We can identify arrows to B with their domains. A sieve on B is therefore the equivalent of a lower set on B in order-theoretic terms. To see how subobjects are classified in this case, consider a subpresheaf $\underline{X}$ of some presheaf $\underline{Y}$ and, for simplicity, assume that the monic between them is given by inclusions at every stage. For every stage A an element of $\underline{Y}_A$ is classified according to whether that element together with its restrictions to all possible $\underline{Y}_B$ belong to the sets $\underline{X}_A$ or $\underline{X}_B$ respectively. This classification yields a sieve on A for every point because if y restricts to an element of $\underline{X}_B$ for $B \leq A$ then it also restricts to an element of $\underline{X}_C$ for any $C \leq B$. The sieve, which is a lower set on A , possibly empty, tells us from which stage onwards we can expect to find the restriction of $y \in \underline{Y}_A$ within the subpresheaf $\underline{X}$.

We need one more concept: exponentiation, before we are able to state the formal definition of a topos. In the category of sets the exponential B^A is the set of all functions from A to B . To characterise B^A by arrows note that associated with B^A is a special arrow, the evaluation function:

$$ev : B^A \times A \rightarrow B$$

given by the rule $ev(f, x) = f(x)$. The evaluation function is distinguished from the set of all functions of the form $g : C \times A \rightarrow B$ because given any such g , there is exactly one function

$\hat{g} : C \rightarrow B^A$ making the diagram below commute:

$$\begin{array}{ccc}
 B^A \times A & \xrightarrow{ev} & B \\
 \hat{g} \times id_A \uparrow \text{dotted} & & \nearrow g \\
 C \times A & &
 \end{array}$$

By abstraction, we say that a category $\mathcal{C}$ has exponentiation if it has all binary products and if for any $\mathcal{C}$ -objects A and B there is a $\mathcal{C}$ -object B^A and a $\mathcal{C}$ -arrow $ev : B^A \times A \rightarrow B$ which enjoys the universal property we have described in the diagram above. If such an arrow exists, the assignment of $\hat{g}$ to g establishes an adjunction

$$\mathcal{C}(C \times A, B) \cong \mathcal{C}(C, B^A)$$

Definition 5.1.14. *An elementary topos is a category which has*

1. *all limits and colimits taken over finite index categories*
2. *exponentials*
3. *a subobject classifier*

5.2 C*-Algebras and von Neumann Algebras

In this section we shall briefly present a selection of standard algebraic results which will allow the reader to understand the statement and implications Gelfand duality. This important duality will then be used in the following section to define the central construction of the Topos Approach - the spectral presheaf. A more detailed presentation of the concepts introduced in this section can be found in [93] and [43].

Definition 5.2.1. *A Banach algebra B is an associative algebra over the real or complex numbers, whose underlying vector space is a Banach space. The algebra multiplication and the Banach space norm must be related by the following inequality called submultiplicativity:*

$$\forall x, y \in B, \quad \|xy\| \leq \|x\| \|y\|$$

This ensures that the multiplication operation, seen as a function from $B \times B$ to B is continuous with respect to the norm topology.

The prototypical example of a commutative Banach algebra is the space $C_0(X)$ of complex-valued continuous functions on a locally compact Hausdorff space, which vanish at infinity. The algebra operations are pointwise addition and multiplication of functions. When the space X is compact, this is equivalent to the algebra $C(X)$ of continuous complex valued functions on X .

Definition 5.2.2. A C^* -algebra A is a Banach algebra over the field of complex numbers, together with an involution map $*$: $A \rightarrow A$ satisfying the following properties:

- $(x^*)^* = x$
- $(x + y)^* = x^* + y^*$
- $(\lambda x)^* = \bar{\lambda}x^*$
- $(xy)^* = y^*x^*$
- $\|x^*x\| = \|xx^*\| = \|x\| \|x^*\|$

for all elements x, y of the algebra.

A C^* -algebra is called *unital* if it contains a unit element u such that

$$ua = au = a, \quad \forall a \in A$$

The prototypical example of a non-commutative C^* -algebra is the complex algebra of continuous bounded linear operators on a Hilbert space H with involution given by the adjoint, which we shall denote by $\mathcal{B}(H)$. A subset of $\mathcal{B}(H)$ is a C^* -algebra if it is a topologically closed set in the norm topology of operators, and closed under involution, that is under the operation of taking adjoints of operators. Note that if the Hilbert space H is finite-dimensional, then $\mathcal{B}(H) \simeq M_n(\mathbb{C})$.

The Banach algebra $C(X)$ is also an example of a C^* -algebra since the operation of pointwise complex conjugation satisfies the involution properties above.

Definition 5.2.3. A $*$ -homomorphism is a homomorphism between two C^* -algebras which is compatible with the involution maps of the two algebras i.e., $\phi(x^*) = \phi(x)^*$ for all $x \in A$.

Definition 5.2.4. An element x of a C^* -algebra is called *self-adjoint* if it satisfies $x^* = x$.

Definition 5.2.5. If A is a unital Banach algebra we define the *spectrum* of an element $x \in A$ to be the set of scalars λ such that $x - \lambda 1_A$ is not invertible. We denote it by $\sigma(x)$.

Note that the spectrum of a self-adjoint element of a C^* -algebra consists only of real numbers. As we have mentioned in the introduction, self-adjoint operators of C^* -algebras have a special significance from the physical point of view: they represent the observables of the system described by the respective algebra. The spectrum of a self-adjoint operator contains the possible values which the observable represented by that operator can take upon measurement.

Definition 5.2.6. A C^* -algebra B is called a W^* -algebra if as a Banach space it is the dual of a Banach space (i.e. if there exists a Banach space B_* such that $(B_*)^* = B$, where $(B_*)^*$ is the dual of the Banach space B_*). B_* is called the **predual** of B .

A **von Neumann algebra** is a C^* -subalgebra of some $\mathcal{B}(H)$ that is closed in the weak (and equivalently, the strong) operator topology on $\mathcal{B}(H)$.

Von Neumann algebras arise as (faithful) representations of W^* -algebras by linear operators on a Hilbert space, and W^* -algebras can be seen as abstract, representation-free von Neumann algebras.

An important example of a von Neumann algebra is the algebra $\mathcal{B}(H)$ of all bounded operators on a Hilbert space H . Moreover it is known that every von Neumann algebra is isomorphic to a subalgebra of some $\mathcal{B}(H)$ for a suitable Hilbert space H .

Recall that a self-adjoint element of a C^* -algebra or von Neumann algebra is called a projection if it is idempotent. We can define a partial order $\leq$ on the projections of a von Neumann algebra such that $p \leq q$ iff $pq = qp = p$. Note that the projections of a von Neumann algebra form a complete orthomodular lattice.

Definition 5.2.7. An **atom** of a von Neumann algebra N is an atom of the projection lattice $\mathcal{P}(N)$. A von Neumann algebra is called **atomic** if every non-zero projection is greater than an atom of that algebra.

Definition 5.2.8. A linear functional f on a C^* -algebra is called **positive** if $f(x^*x) \geq 0$ for all non-zero elements x of the algebra.

A bounded linear functional f is positive if and only if $\|f\| = f(1)$ (see Theorem 4.3.2 in [93]). If f and g are two linear functionals on an algebra, we say that f is *majorized* by g if $f - g$ is positive.

Note that every positive linear functional defines an inner product on A :

$$\forall x, y \in A \quad \langle x|y \rangle := f(y^*x)$$

The Cauchy-Schwartz inequality (see Proposition 4.3.1 in [93]) then gives $|f(y^*x)|^2 \leq f(x^*x)f(y^*y)$ for all $x, y \in A$.

Definition 5.2.9. *A state of an algebra is a positive linear functional of norm 1.*

Definition 5.2.10. *The positive linear functionals of norm ≤ 1 form a convex subset of the dual space of a C^* -algebra A . The non-zero extremal points of this subset are called pure states.*

In the case of a commutative C^* -algebra the pure states are precisely those positive linear functionals of norm 1 which are multiplicative. For non-commutative algebras this does not hold in general.

5.2.1 Gelfand duality and Gelfand spectra

Definition 5.2.11. *The **Gelfand spectrum** of a commutative C^* -algebra A is the set of pure states of that algebra, equipped with the weak*-topology. We will denote it by $\text{Spec } A$.*

The spectrum of a unital C^* -algebra is a compact Hausdorff space with respect to the weak*-topology. The Gelfand spectrum of a commutative C^* -algebra A and the spectrum of an element $x \in A$ are related in the following way:

$$\sigma(x) = \{f(x) \mid f \in \text{Spec } A\}$$

Theorem 5.2.12 (Gelfand and Naimark, 1943). *For every unital and commutative C^* -algebra A , there exists a *-isomorphism between A and $C(\text{Spec } A)$:*

$$\begin{aligned} \mathcal{G} : A &\longrightarrow C(\text{Spec } A) \\ a &\longmapsto \bar{a} : \text{Spec } A \rightarrow \mathbb{C} \\ f &\mapsto f(a) \end{aligned}$$

*called the **Gelfand representation** of A . The functional $\bar{a}$ is called the Gelfand transform of the element a .*

Let $\mathbf{UComC}^*\mathbf{Alg}$ denote the category of unital commutative C^* -algebras, and let $\mathbf{KHausSp}^{op}$ denote the opposite of the category of compact Hausdorff spaces. A direct consequence of Theorem 5.2.12 is that these are equivalent categories. The equivalence is given by the following pair of functors:

$$\begin{array}{ccc} \mathbf{UComC}^*\mathbf{Alg} & \xrightleftharpoons[C(-)]{\Sigma} & \mathbf{KHausSp}^{op} \\ A & \longrightarrow & \mathrm{Spec} A \\ C(X) & \longleftarrow & X \end{array}$$

Both functors are contravariant and act in the following way: if $\varphi : A \rightarrow B$ is a unital $*$ -homomorphism between two C^* -algebras, A and B , then $\Sigma(\varphi)$ is given by

$$\begin{aligned} \Sigma(\varphi) : \mathrm{Spec} B &\longrightarrow \mathrm{Spec} A \\ f &\longmapsto f \circ \varphi. \end{aligned}$$

Similarly, if $\phi : X \rightarrow Y$ is a morphism between two compact Hausdorff spaces, X and Y , then $C(\phi)$ is given by

$$\begin{aligned} C(\phi) : C(Y) &\longrightarrow C(X) \\ g &\longmapsto g \circ \phi. \end{aligned}$$

5.3 The spectral presheaf

In classical mechanics, Gelfand duality establishes a correspondence between the state space S of a physical system, which at the most basic level can be seen as a set, and the commutative C^* -algebra of observables, which is given by the collection of real-valued functions

$$f_A : S \rightarrow \mathbb{R} \tag{5.3.1}$$

on state space, under the operations of pointwise addition and multiplication. Each function f_A corresponds to an observable of the classical system. In this interpretation of classical mechanics, propositions about the system correspond to measurable subsets of S of the form $f_A^{-1}(D)$. Such

a subset represents the proposition asserting that the observable A takes values in the subset D of the real numbers.

Within the topos approach [28–31, 58–61], the spectral presheaf associated with the non-commutative von Neumann algebra N of observables of a quantum system is the analogue of the state space of a classical system.

We denote the set of all commutative subalgebras (or contexts) of N , minus the trivial one $V_0 = \mathbb{C}1$, by $\mathcal{V}(N)$. This is a partially ordered set under inclusion, and as such it forms a category. We can use Gelfand duality to associate a Hausdorff space to each context, and this collection of Hausdorff spaces forms the spectral presheaf.

Definition 5.3.1. *The **spectral presheaf** $\underline{\Sigma}^N$ of a given von Neumann algebra N is the following contravariant functor from the category $\mathcal{V}(N)$ to the category of sets:*

- a) on objects: for all $V \in \mathcal{V}(N)$, let $\underline{\Sigma}_V^N$ be the Gelfand spectrum of V , i.e. the set of multiplicative positive linear functionals of norm one, or equivalently, the set of pure states on V , equipped with the weak- $*$ topology*
- b) on arrows: for all inclusions $i_{VV'} : V' \hookrightarrow V$, let $\underline{\Sigma}^N(i_{VV'}) : \underline{\Sigma}_{V'}^N \rightarrow \underline{\Sigma}_V^N$ be the function that sends each pure state f to its restriction $f|_{V'}$ to the smaller algebra. This function is well-known to be continuous and surjective.*

When no confusion arises we will simply write $\underline{\Sigma}$ instead of $\underline{\Sigma}^N$.

For the sake of obtaining a better understanding of this abstract construction, we consider a concrete von Neumann algebra $N := M_n(\mathbb{C})$, which corresponds to a quantum system described by the Hilbert space $\mathcal{H} := \mathbb{C}^n$, and we show what the spectral presheaf looks like in this particular case.

Let $(\psi_1, \psi_2, \dots, \psi_n)$ be an orthonormal basis of $\mathcal{H}$, and let $(P_1, P_2, \dots, P_n)$ be the n projections onto the one-dimensional subspaces spanned by each of the basis vectors. These projections are pairwise orthogonal, i.e., $P_i P_j = \delta_{ij} P_i$.

Using von Neumann's double commutant construction (see below), we can define the abelian subalgebra of N generated by the n projections considered above, $V = \{P_1, P_2, \dots, P_n\}''$.

Definition 5.3.2. *If $\mathcal{B}(H)$ is the algebra of bounded operators on some Hilbert space H and $\mathcal{F} \subseteq \mathcal{B}(H)$, the **commutant** of $\mathcal{F}$ is the subset of $\mathcal{B}(H)$ consisting of all elements that commute*

with every element of $\mathcal{F}$, that is

$$\mathcal{F}' = \{T \in \mathcal{B}(H) \mid TS = ST, \forall S \in \mathcal{F}\}$$

The **double commutant** of $\mathcal{F}$ is just $(\mathcal{F}')'$ and is usually denoted by $\mathcal{F}''$.

With respect to the basis $(\psi_1, \psi_2, \dots, \psi_n)$, the abelian algebra V is equal to the algebra D_n of diagonal $n \times n$ matrices with complex entries on the diagonal. In this way, every orthonormal basis of $\mathbb{C}^n$ determines a maximal abelian subalgebra of N , and all maximal abelian subalgebras of N are isomorphic to each other.

The abelian von Neumann subalgebras of $V \simeq D_n$ above are the diagonal matrices with k independent entries. A diagonal matrix with k independent entries corresponds to a collection $\mathcal{S}$ of k pairwise disjoint sets $S_1, S_2, \dots, S_k$ whose union is the set $\{1, 2, \dots, n\}$. Each set in $\mathcal{S}$ contains the indices of those positions on the diagonal which have the same value. For example, the maximal algebra D_n corresponds to the collection

$$\{ \{1\}, \{2\}, \dots, \{n\} \}$$

The minimal algebra $\mathbb{C}I_n$ corresponds to the collection

$$\{ \{1, 2, \dots, n\} \}$$

Recall that this algebra is excluded by convention from the poset $\mathcal{V}(N)$.

Given a collection $\mathcal{S}$ containing k sets $S_1, \dots, S_k$, for every i from 1 to k we can construct the projections

$$Q_i := \sum_{j \in S_i} P_j$$

The algebra corresponding to $\mathcal{S}$ is

$$V_{\mathcal{S}} = \{Q_1, Q_2, \dots, Q_k\}''$$

Of course, there are other projections whose double commutant would give the same algebra, but the advantage of using this particular set of projections is that they form an ‘orthonormal basis’ for $V_{\mathcal{S}}$ in the sense that the projections are pairwise orthogonal and sum up to the identity

1.

The Gelfand spectrum of $V_{\mathcal{S}}$ hence contains k elements and is equipped with the discrete topology. The spectral elements are linear multiplicative functionals, and they are determined by their behaviour on the ‘basis’ $Q_1, \dots, Q_k$, that is

$$f_i(Q_j) = \delta_{ij}, \quad i \in \{1, 2, \dots, k\}$$

If $A \in V_{\mathcal{S}}$ is an arbitrary operator, $A = \sum_{i=1}^k a_i Q_i$ for some (unique) complex coefficients a_i . For each i , linearity gives us $f_i(A) = a_i$. The a_i s are the eigenvalues of A and if A is self-adjoint, they are real numbers.

If $V_{\mathcal{S}} \subseteq V_{\tilde{\mathcal{S}}}$, then $\tilde{\mathcal{S}}$ is a refinement of $\mathcal{S}$ in the following sense: every set $S_i \in \mathcal{S}$ can be written as the union of $t_i \geq 1$ pairwise disjoint sets $\tilde{S}_1^i, \tilde{S}_2^i, \dots, \tilde{S}_{t_i}^i$ such that

$$\tilde{\mathcal{S}} = \left\{ \tilde{S}_1^1, \tilde{S}_2^1, \dots, \tilde{S}_{t_1}^1, \quad \tilde{S}_1^2, \tilde{S}_2^2, \dots, \tilde{S}_{t_2}^2, \quad \dots, \quad \tilde{S}_1^k, \tilde{S}_2^k, \dots, \tilde{S}_{t_k}^k \right\}$$

where all the sets in $\tilde{\mathcal{S}}$ are pairwise disjoint. For each $i \in \{1, \dots, k\}$ and $j \in \{1, \dots, t_i\}$ define the projection

$$Q_j^i := \sum_{x \in \tilde{S}_j^i} P_x$$

The collection of all projections of this form generates the algebra $V_{\tilde{\mathcal{S}}}$ via the double commutant construction. The pure states of $V_{\tilde{\mathcal{S}}}$ are therefore given as

$$\tilde{f}_j^i(Q_v^u) = \delta_{iu} \delta_{jv}$$

The restriction map going from $\Sigma_{V_{\tilde{\mathcal{S}}}}$ to $\Sigma_{V_{\mathcal{S}}}$ takes the spectral element $\tilde{f}_j^i$ to the element f_i .

5.3.1 Daseinisation

We can use the spectral theorem to interpret projections of a von Neumann algebra N as propositions of the form “ $A \in \Delta$ ”, that is propositions of the form “the physical quantity A , which is represented by the self-adjoint operator $A \in N$, has a value in the set Borel set Δ ”. More precisely, each projection corresponds to an equivalence class of such propositions.

If we take a commutative subalgebra V of N , every state $f \in \text{Spec } V$ of V gives us a way to

assign truth values to propositions which involve quantities represented by self-adjoint operators from V . Any such f can take only one of the two values 0, 1 when applied to a projection $P \in V$, since

$$f(P) = f(P^2) = f(P)f(P)$$

So we can assign to those propositions which correspond to the projection P the value true if $f(P) = 1$, and false if $f(P) = 0$. We know from the Kochen-Specker theorem that, under certain natural conditions, it would not be possible to make such truth-value assignments for the projections of the non-commutative algebra N (unless N was a type I_2 -algebra).

The projections in a commutative von Neumann algebra V correspond bijectively to clopen subsets of $\text{Spec } V$:

Proposition 5.3.3. *If $\mathcal{P}(V)$ is the lattice of all projections in V and $Cl(\text{Spec } V)$ is the lattice of clopen subsets of $\text{Spec } V$, then the map*

$$\alpha_V : \mathcal{P}(V) \rightarrow Cl(\text{Spec } V)$$

$$P \mapsto S_P := \{f \in \text{Spec } V \mid f(P) = 1\}$$

is a lattice isomorphism.

Proof: It is easy to check that S_P is indeed a clopen subset of $\text{Spec } V$. We have

$$S = \overline{P}^{-1} \left(\left(\frac{1}{2}, \infty \right) \right)$$

and so S is open. Similarly

$$\text{Spec } V \setminus S = \overline{P}^{-1} \left(\left(-\infty, \frac{1}{2} \right) \right)$$

and so $\text{Spec } V \setminus S$ is open, hence S is closed.

Since the Gelfand representation is a $*$ -isomorphism for unital commutative algebras, α_V must be a bijective map. $\square$

Note that this result is also a consequence of the Stone representation theorem, as shown by de Groote (see Theorem 3.2 in [39]): the projections in V form a complete Boolean algebra $\mathcal{P}(V)$, which is isomorphic to the complete Boolean algebra of clopen subsets of the Stone space of $\mathcal{P}(V)$. For an abelian von Neumann algebra V , the Stone space of $\mathcal{P}(V)$ is homeomorphic to the Gelfand spectrum of V .

We have seen that for a ‘classical part’ of a quantum system described by a commutative algebra V there is a correspondence between propositions, or rather the projections which represent them, and clopen subsets of the Gelfand spectrum of the algebra V . Next we will see that for quantum systems (as a whole) there is an analogous correspondence between propositions and clopen sub-objects of the spectral presheaf.

The collection of all contexts of a non-commutative von Neumann algebra N can be understood as the collection of all classical perspectives on a quantum system. As we have mentioned before, the idea behind the spectral presheaf is to characterise a quantum system by taking into account all the classical perspectives at the same time. In order to do this, we need to adapt every proposition about the whole quantum system to each possible classical context. That is, given a proposition “ $A\varepsilon\Delta$ ” and its representing projection P , we want to choose for every context V the strongest proposition implied by “ $A\varepsilon\Delta$ ” which can be made from the perspective of that context. For projections, this is equivalent to taking the smallest projection in any context V that is larger or equal to P :

$$\delta^o(P)_V := \bigwedge \{Q \in \mathcal{P}(V) \mid Q \geq P\}$$

If $P \in \mathcal{P}(V)$, the above approximation will simply be equal to P . We will call the original proposition “ $A\varepsilon\Delta$ ” the *global proposition*, while a proposition “ $B\varepsilon\Gamma$ ” corresponding to the projection $\delta^o(P)_V$ will be called a *local proposition*.

From the family of projections $(\delta^o(P)_V)_{V \in \mathcal{V}(N)}$ we can obtain a family of clopen subsets of the Gelfand spectra $(\text{Spec } V)_{V \in \mathcal{V}(N)}$ by choosing for every V the subset

$$S_{\delta^o(P)_V} = \alpha_V(\delta^o(P)_V) \subseteq \text{Spec } V$$

These subsets form a subobject under the restriction mappings of the spectral presheaf $\underline{\Sigma}$ and so we can give the following definition, as in [46].

Definition 5.3.4. *The **daseinisation of a projection** P is the subobject (or equivalently, the subpresheaf) $\underline{\delta}(P)$ of the spectral presheaf $\underline{\Sigma}$ given by the collection of clopen subsets $(S_{\delta^o(P)_V})_{V \in \mathcal{V}(N)}$, together with the restriction mappings between them.*

We denote the collection of all sub-objects of the spectral presheaf by $\text{Sub}(\underline{\Sigma})$. It can be seen as the analogue of the lattice of subsets of the state space of a classical system.

Definition 5.3.5. A subobject $\underline{S}$ of the spectral presheaf $\underline{\Sigma}$ such that for each $V \in \mathcal{V}(N)$ the component $\underline{S}_V$ is a clopen subset of $\underline{\Sigma}_V$ is called a ***clopen subobject***.

Daseinisation thus gives us a map between the projections of a von Neumann algebra and the collection of clopen subobjects of the spectral presheaf

$$\underline{\delta} : \mathcal{P}(N) \rightarrow \text{Sub}_{cl}(\underline{\Sigma})$$

since all sub-objects obtained from the daseinisation of projections are clopen. The collection $\text{Sub}_{cl}(\underline{\Sigma})$ can be seen as the analogue of the collection of measurable subsets of the state space of a classical system.

The daseinisation $\underline{\delta}(P)$ of a projection P representing the proposition “ $A \varepsilon \Delta$ ” can be seen as quantum the analogue of the measurable subset $f_A^{-1}(D)$ of the state space of a classical system, where f_A represents a classical observable, as in Equation 5.3.1. We interpret $\underline{\delta}(P)$ as the representative of the global proposition “ $A \varepsilon \Delta$ ”.

The daseinisation $\underline{\delta}(P_\psi)$ of a projection P_ψ which projects onto the ray spanned by the vector ψ is called the *pseudo-state* associated to ψ . It can be regarded as the analogue of a point in the state space of a classical system. It is important to note however that the pseudo-states are not global elements of $\underline{\Sigma}$. In fact, global elements of a presheaf are the category-theoretical analogues of points. Isham and Butterfield have observed [28] that the Kochen-Specker theorem is equivalent to the fact that the spectral presheaf has no global elements. A global element γ of $\underline{\Sigma}$ would pick one $\gamma_V \in \underline{\Sigma}_V$ for each context V such that, whenever $V' \subset V$, one would have $\gamma_V|_{V'} = \gamma_{V'}$. Each γ_V assigns values to all physical quantities described by self-adjoint operators A in V by evaluation, i.e., by simply forming $\gamma_V(A)$. If A is contained in different commutative subalgebras $V, \tilde{V}$, then it is also contained in $V' := V \cap \tilde{V}$, and $\gamma_V(A) = \gamma_{V'}(A) = \gamma_{\tilde{V}}(A)$, so the defining condition of the global element γ guarantees that A is assigned the same value in every context. Since every self-adjoint operator is contained in some commutative subalgebra V , a global element γ of $\underline{\Sigma}$ would provide a consistent assignment of values to all self-adjoint operators. But the Kochen-Specker theorem precisely shows that this is impossible, hence such global elements γ cannot exist.

Pseudo-states on the other hand are minimal sub-objects in a suitable sense: they come from rank-1 projections, the smallest non-trivial projections, and daseinisation is order-preserving,

so pseudo-states are the smallest non-trivial sub-objects of $\underline{\Sigma}$ that can be obtained from daseinisation. Hence, pseudo-states are ‘as close to points as possible’.

Note also that, in analogy with the classical case where the powerset of the state space formed a Boolean algebra, the collection of all sub-objects of the spectral presheaf can be turned (see [47]) into a complete Heyting algebra by defining suitable meet and join operations.

Definition 5.3.6. *If $\underline{S}_1$ and $\underline{S}_2$ are two sub-objects of the spectral presheaf, their join is defined by stagewise unions in the following way:*

$$(\underline{S}_1 \vee \underline{S}_2)_V = \underline{S}_{1V} \cup \underline{S}_{2V}$$

Similarly, their meet is given by stagewise intersections:

$$(\underline{S}_1 \wedge \underline{S}_2)_V = \underline{S}_{1V} \cap \underline{S}_{2V}$$

In [47, 52], Döring has also shown that the clopen sub-objects of the spectral presheaf $\underline{\Sigma}$ form a complete Heyting algebra under stagewise meet and join operations. For more information about the daseinisation map, its physical interpretation, its properties, and in particular its relation to the logic of the Topos Approach and to the lattice of clopen subobjects of the spectral presheaf the interested reader is directed to the clear and concise presentation exposed in [47].

5.3.2 States as measures on the spectral presheaf

In general, a state on a von Neumann algebra is a positive linear functional of norm one on that algebra. Given such a state, we can associate to it a certain measure on the corresponding spectral presheaf. This construction was explored in detail by Döring, who also showed that measures on the spectral presheaf can be defined without reference to states and moreover that from each abstractly defined measure a unique state can be reconstructed [44]. We give a brief overview of these ideas below.

In classical physics states are represented by probability measures on state space, and pure states are represented by Dirac measures. A probability measure assigns a number between 0 and 1 to each measurable subset of state space. Within the topos approach the role of the state space is played by the spectral presheaf, and so in analogy with classical mechanics we would like

states to be represented by probability measures on (clopen subobjects of) the spectral presheaf. However, since subobjects of the spectral presheaf are not simply sets, but collections of sets, we can not expect the values taken by the measure to be given by single numbers. Instead we would expect to obtain a collection of such numbers, one for each context of the algebra which represents our system. With this in mind we give the following definitions of presheaves of real values and their global sections, which will be essential to our discussion.

Definition 5.3.7. *Given a von Neumann algebra N and its associated poset of abelian subalgebras $\mathcal{V}(N)$, let $\downarrow V := \{W \in \mathcal{V}(N) \mid W \subseteq V\}$ denote the down-set of a context $V \in \mathcal{V}(N)$. The presheaf $\underline{\mathbb{R}}^\succeq$ is defined*

- on objects: $\underline{\mathbb{R}}^\succeq_V = \{f : \downarrow V \rightarrow \mathbb{R} \mid f \text{ is order reversing}\}$
- on arrows: for $i_{V'V} : V' \hookrightarrow V$, $\underline{\mathbb{R}}^\succeq(i_{V'V}) : \mathbb{R} \rightarrow \mathbb{R}$ is given by

$$\underline{\mathbb{R}}^\succeq(i_{V'V})(f) := f|_{\downarrow V'}$$

Note that this presheaf lives in the same topos as $\underline{\Sigma}^N$. However, we do not explicitly specify this topos, by indicating the base category, when discussing this and similar presheaves of real values. It is usually clear from the context, which base category we are using.

A global section of this presheaf can be regarded as an order-reversing function from the partially ordered set $\mathcal{V}(N)$ to the real numbers equipped with the usual ordering.

The presheaf defined above plays an important role within the topos approach, and is discussed extensively in [60, 62]. However, when defining measures we will only use a sub-presheaf of this presheaf of real numbers, which we denote by $\underline{[0, 1]}^\succeq$. Later on, when we will introduce the notion of entropy we will encounter a closely related presheaf, $\underline{[0, \ln n]}^\preceq$, where n denotes the dimension of the algebra corresponding to our system. In this case global sections will be equivalent to order-preserving functions from $\mathcal{V}(N)$ to the real interval $[0, \ln n]$.

Definition 5.3.8. *Given a von Neumann algebra N , a measure on its associated spectral*

presheaf $\underline{\Sigma}$ is a mapping

$$\begin{aligned}\mu : \text{Sub}_{\text{cl}}(\underline{\Sigma}) &\longrightarrow \Gamma[0, 1]_{\leq}^{\leq} \\ \underline{S} = (\underline{S}_V)_{V \in \mathcal{V}(N)} &\longmapsto \mu(\underline{S}) : \mathcal{V}(N) \rightarrow [0, 1] \\ V &\longmapsto \mu(\underline{S}_V)\end{aligned}$$

which satisfies the following conditions:

1. $\mu(\underline{\Sigma}) = 1_{\mathcal{V}(N)}$
2. for all $\underline{S}_1, \underline{S}_2 \in \text{Sub}_{\text{cl}}(\underline{\Sigma})$, it holds that

$$\mu(\underline{S}_1 \vee \underline{S}_2) + \mu(\underline{S}_1 \wedge \underline{S}_2) = \mu(\underline{S}_1) + \mu(\underline{S}_2)$$

where the addition, just like the meet and the join for sub-objects, is defined as a stagewise operation.

The conditions above also imply that $\mu(\underline{0}) = 0$, where $\underline{0}$ is the subobject of $\underline{\Sigma}$ which assigns the empty set to each context. Note also that we have slightly abused notation by writing μ both for the measure and for its contextual components.

In this text we will mostly be concerned with a particular type of von Neumann algebras, the algebras of bounded linear operators on finite dimensional Hilbert spaces (i.e. matrix algebras). For these algebras the states can be identified with the density matrices: to each density matrix $\rho \in M_n$, we can associate the functional

$$A \longmapsto \text{Tr}(\rho A), \quad \forall A \in M_n$$

and moreover every positive linear functional of unit norm is of this form in the finite dimensional setting. With this in mind, when talking about matrix algebras we shall refer to the density matrices as states on those algebras.

Definition 5.3.9. *Given a state ρ on the matrix algebra M_n , it is straightforward to define its*

associated measure:

$$\begin{aligned}\mu_\rho : \text{Sub}_{\text{cl}}(\underline{\Sigma}^{M_n}) &\longrightarrow \Gamma[0, 1]^\succeq \\ \underline{S} = (\underline{S}_V)_{V \in \mathcal{V}(M_n)} &\longmapsto \mu_\rho(\underline{S}) : \mathcal{V}(M_n) \rightarrow [0, 1] \\ V &\mapsto \text{Tr}(\rho P_{\underline{S}_V})\end{aligned}$$

where $P_{\underline{S}_V} = \alpha_V^{-1}(\underline{S}_V)$.

One can easily check that the function $\mu_\rho(\underline{S})$ is order reversing and that μ_ρ satisfies the two properties required in the definition of a measure. This is explicitly done in [44] for arbitrary von Neumann algebras without type I_2 summands.

On the other hand, an abstract measure on the spectral presheaf associated to any given algebra N , determines a unique state of N , provided N contains no direct summand of type I_2 . The proof of this rather surprising result uses a generalized version of Gleason's theorem, a review of which can be found in [100].

Definition 5.3.10. *A finitely additive probability measure m on the projections of a von Neumann algebra N is a map*

$$m : \mathcal{P}(N) \rightarrow [0, 1]$$

such that $m(I) = 1$ and if P and Q are orthogonal projections then

$$m(P \vee Q) = m(P + Q) = m(P) + m(Q)$$

Theorem 5.3.11 (Generalized Gleason's Theorem). *Each finitely additive probability measure on the projections of a von Neumann algebra without type I_2 summands can be uniquely extended to a state on that algebra.*

Using this powerful result we can show that each measure on the spectral presheaf uniquely determines a state on the corresponding algebra by showing that such a measure determines a unique finitely additive probability measure on the projections of the respective algebra. This has been done by Döring in [44], and we will reproduce his proof in the remainder of this section.

Given a measure μ on the spectral presheaf $\underline{\Sigma}$ associated to some von Neumann algebra N , let $\underline{S}$ be a clopen subobject of $\underline{\Sigma}$. From Proposition 5.3.3 we know that for each context V

there exists an isomorphism α_V between $\mathcal{P}(V)$ and $Cl(\underline{\Sigma}_V)$. If $P = \alpha_V^{-1}(\underline{S}_V)$ we define

$$m(P) = \mu(\underline{S})(V) = \mu(\underline{S}_V)$$

First note that for every projection P , there is some subobject S and context V such that $\alpha_V^{-1}(\underline{S}_V) = P$: for example, the daseinisation of P has this property, if we set $V = P, 1 - P''$. But we also have to show that this does not depend on the choice of the subobject $\underline{S}$ and the context V , i.e. we must show that if $\tilde{S}$ is another subobject of $\underline{\Sigma}$ and $\tilde{V}$ is a context such that $\alpha_{\tilde{V}}^{-1}(\underline{S}_{\tilde{V}}) = \alpha_V^{-1}(\underline{S}_V)$ then $\mu(\underline{S}_V) = \mu(\underline{S}_{\tilde{V}})$. For this we will need two intermediate results.

Lemma 5.3.12. *If $\underline{S}$ is a clopen subobject of $\underline{\Sigma}$ and $V' \subseteq V$ are two contexts such that P is contained in both V and V' and $\alpha_V^{-1}(\underline{S}_V) = \alpha_{V'}^{-1}(\underline{S}_{V'}) = P$, then $\mu(\underline{S}_V) = \mu(\underline{S}_{V'})$.*

Proof: Since the maximal projection I is contained in every context, it follows that $I - P \in V', V$. Let $\underline{S}^c$ be another clopen subobject such that

$$\alpha_V^{-1}(\underline{S}_V^c) = \alpha_{V'}^{-1}(\underline{S}_{V'}^c) = I - P$$

Such a subobject certainly exists: $\underline{\delta(I - P)}$, for example, satisfies the above property.

Since every α is a lattice isomorphism, we have

$$\begin{aligned} (\underline{S} \wedge \underline{S}^c)_V &= 0_V = \emptyset, & (\underline{S} \wedge \underline{S}^c)_{V'} &= 0_{V'} = \emptyset \\ (\underline{S} \vee \underline{S}^c)_V &= \underline{\Sigma}_V, & (\underline{S} \vee \underline{S}^c)_{V'} &= \underline{\Sigma}_{V'} \end{aligned}$$

Using the two defining properties of a measure μ we obtain

$$\begin{aligned} 1 &= \mu(\underline{\Sigma})(V) \\ &= \mu(\underline{S} \vee \underline{S}^c)(V) \\ &= \mu(\underline{S})(V) + \mu(\underline{S}^c)(V) - \mu(\underline{S} \wedge \underline{S}^c)(V) \end{aligned}$$

Since the last term vanishes we obtain that $\mu(\underline{S})(V) + \mu(\underline{S}^c)(V) = 1$. Similarly, we can also deduce that $\mu(\underline{S})(V') + \mu(\underline{S}^c)(V') = 1$. But $\mu(S) : \mathcal{V}(N) \rightarrow [0, 1]$ is an order-reversing function,

hence

$$\begin{aligned}\mu(\underline{S})(V') &\geq \mu(\underline{S})(V) \\ \mu(\underline{S}^c)(V') &\geq \mu(\underline{S}^c)(V)\end{aligned}$$

This implies that in fact $\mu(\underline{S})(V') = \mu(\underline{S})(V)$ and $\mu(\underline{S}^c)(V') = \mu(\underline{S}^c)(V)$, which completes our proof. $\square$

Lemma 5.3.13. *If $\underline{S}$ and $\tilde{\underline{S}}$ are two subobjects which coincide at V , i.e. if $\underline{S}_V = \tilde{\underline{S}}_V$, then $\mu(\underline{S})(V) = \mu(\tilde{\underline{S}})(V)$.*

Proof: From the second defining property of a measure μ we obtain that

$$\begin{aligned}\mu(\underline{S})(V) + \mu(\tilde{\underline{S}})(V) &= \mu(\underline{S} \vee \tilde{\underline{S}})(V) + \mu(\underline{S} \wedge \tilde{\underline{S}})(V) \\ &= \mu((\underline{S} \vee \tilde{\underline{S}})_V) + \mu((\underline{S} \wedge \tilde{\underline{S}})_V) \\ &= \mu(\underline{S}_V \cup \tilde{\underline{S}}_V) + \mu(\underline{S}_V \cap \tilde{\underline{S}}_V) \\ &= \mu(\underline{S}_V) + \mu(\tilde{\underline{S}}_V) \\ &= \mu(\underline{S})(V) + \mu(\tilde{\underline{S}})(V)\end{aligned}$$

which implies that $\mu(\underline{S})(V) = \mu(\tilde{\underline{S}})(V)$. $\square$

Now assume that $\underline{S}$ and $\tilde{\underline{S}}$ are two clopen subobjects of $\underline{\Sigma}$ and V and $\tilde{V}$ are two contexts such that $\underline{S}_V$ and $\tilde{\underline{S}}_{\tilde{V}}$ correspond to the same projection $P \in V, \tilde{V}$. Then we must have that P also belongs to $V \cap \tilde{V}$. We know that the clopen subobject $\underline{\delta(P)}$ coincides with $\underline{S}$ at V and it also coincides with $\tilde{\underline{S}}$ at $\tilde{V}$. Moreover, $\underline{\delta(P)}_{V \cap \tilde{V}} \subseteq \underline{\Sigma}_{V \cap \tilde{V}}$ and $\alpha_{V \cap \tilde{V}}^{-1}(\underline{\delta(P)}_{V \cap \tilde{V}}) = P$. From the previous two lemmas we obtain that

$$\begin{aligned}\mu(\underline{S})(V) &= \mu(\underline{\delta(P)})(V) \\ &= \mu(\underline{\delta(P)})(V \cap \tilde{V}) \\ &= \mu(\underline{\delta(P)})(\tilde{V}) \\ &= \mu(\tilde{\underline{S}})(\tilde{V})\end{aligned}$$

This shows that the value $m(P) = \mu(\underline{S})(V)$ is well defined. For any V , the projection corresponding to $\underline{\Sigma}_V$ is the maximal projection, I . So from the first defining property of a

measure μ , we must have

$$m(I) = \mu(\underline{\Sigma})(V) = 1$$

Finally, let P and Q be two orthogonal projections and let V be a context that contains both P and Q . Let $\underline{S}^P$ and $\underline{S}^Q$ be two subobjects such that $\alpha_V^{-1}(\underline{S}^P_V) = P$ and $\alpha_V^{-1}(\underline{S}^Q_V) = Q$. Then $(\underline{S}^P \vee \underline{S}^Q)_V$ corresponds to $P \vee Q$ and we obtain

$$m(P \vee Q) = \mu(\underline{S}^P \vee \underline{S}^Q)(V) \tag{5.3.2}$$

$$= \mu(\underline{S}^P)(V) + \mu(\underline{S}^Q)(V) - \mu(\underline{S}^P \wedge \underline{S}^Q)(V) \tag{5.3.3}$$

$$= \mu(\underline{S}^P)(V) + \mu(\underline{S}^Q)(V) \tag{5.3.4}$$

$$= m(P) + m(Q) \tag{5.3.5}$$

This shows that the map $m : \mathcal{P} \rightarrow [0, 1]$ is indeed a finitely additive probability measure, and so from the generalised version of Gleason's theorem we know that m extends to a unique state ρ_m of the algebra N .

In particular this implies that when the algebra N is a finite dimensional matrix algebra, of dimension greater than 2, there is a bijective correspondence between density matrices and measures on the corresponding spectral presheaf.

Chapter 6

Quantum Logic vs. the Logic of the Topos Approach

Quantum logic was initially introduced in the '30s by Birkhoff and von Neumann [20]. At that time, the mathematical apparatus of quantum mechanics, mostly developed by von Neumann, was already in place. Birkhoff and von Neumann's goal was “to discover what logical structure one may hope to find in physical theories which, like quantum mechanics, do not conform to classical logic”. Their paper discusses the mathematical structures which may be used to describe propositions about the values of physical quantities, and how states assign truth values to these propositions.

Orthomodular lattices play a prominent role in quantum logic [33,123]. Often, existence of atoms is required for conceptual reasons: it corresponds to the assumption that every pure state represents a ‘physical property’. The prototypical example is $\mathcal{P}(\mathcal{H})$, the lattice of projections on a Hilbert space $\mathcal{H}$, which is an atomic complete orthomodular lattice.

In Chapter 5 we have seen that the topos-based form of logic for quantum systems uses certain presheaf constructions over the poset $\mathcal{V}(\mathcal{N})$ of abelian von Neumann subalgebras of a von Neumann algebra $\mathcal{N}$. For clarity and simplicity, we focus here on the case $\mathcal{N} = \mathcal{B}(\mathcal{H})$, the algebra of bounded linear operators on the Hilbert space $\mathcal{H}$. Its lattice of projections is given by $\mathcal{P}(\mathcal{H})$.

An abelian von Neumann subalgebra $V \subset \mathcal{B}(\mathcal{H})$ has a lattice of projections $\mathcal{P}(V)$ that is a complete Boolean algebra. The abelian subalgebras $V \in \mathcal{V}(\mathcal{B}(\mathcal{H}))$ and their corresponding complete Boolean sublattices $\mathcal{P}(V) \in \mathcal{B}(\mathcal{P}(\mathcal{H}))$ are called *contexts*. Conceptually, they can be thought of as ‘classical perspectives’ on the quantum system, as discussed in Section 5.3.1.

We now turn our attention to an interesting question: can the orthomodular lattice $\mathcal{P}(\mathcal{H})$, which is traditionally used in quantum logic, be reconstructed from the poset of contexts $\mathcal{B}(\mathcal{P}(\mathcal{H}))$ that underlies the constructions in the topos approach? The answer is affirmative, as was shown by Harding and Navara in [80]. Their main result is that if L and M are OMLs and $\phi : \mathcal{B}(L) \rightarrow \mathcal{B}(M)$ is an isomorphism of posets, then there is an isomorphism $\phi^* : L \rightarrow M$ with $\phi(B) = \phi^*[B]$ for each Boolean subalgebra B of L .

Conceptually, this means that by considering the partially ordered set of contexts, one does not lose information compared to considering the whole orthomodular lattice. This also implies that the new form of presheaf- and topos-based form of logic for quantum systems is (at least) as rich as traditional quantum logic.

One motivation for the work presented in this chapter comes from a question posed in the concluding section of [80]: let $\mathcal{S}(L)$ denote the lattice of all subalgebras of an OML L . Give an order-theoretic construction of $\mathcal{S}(L)$ from $\mathcal{B}(L)$. As a partial solution to this problem, we present here a reasonably direct order-theoretic way of reconstructing an atomic OML L from $\mathcal{B}(L)$.

We are concerned only with a single atomic OML L , not with morphisms between two OMLs, so our contribution can moreover be seen as the ‘object counterpart’ to the result by Harding and Navara. The atoms of $\mathcal{B}(L)$ have the form $\{0, P, P^\perp, 1\}$, where P and $P^\perp$ are elements of L . The main task is to show how the order relations between the elements of L arise from the order relations between the elements of $\mathcal{B}(L)$. Of course, L determines $\mathcal{S}(L)$ in a straightforward way.

6.1 Grouping and splitting

We first remark that if $L = \mathcal{P}(\mathcal{H})$, the projection lattice on a Hilbert space $\mathcal{H}$, then the height of the poset $\mathcal{V}(\mathcal{H})$ of abelian subalgebras equals the dimension of $\mathcal{H}$ (if we include the trivial subalgebra $V_0 = \mathbb{C}\hat{1}$ in $\mathcal{V}(\mathcal{H})$; otherwise, the dimension of $\mathcal{H}$ equals the height of $\mathcal{V}(\mathcal{H})$ plus 1). The dimension of $\mathcal{H}$ determines the Hilbert space $\mathcal{H}$ up to isomorphism, and hence determines $\mathcal{P}(\mathcal{H})$ up to isomorphism. This is a cheap (and rather indirect) way of ‘reconstructing’ $\mathcal{P}(\mathcal{H})$ from $\mathcal{V}(\mathcal{H})$.

In this chapter, we will present a more explicit and generally applicable reconstruction.

Remark 6.1.1. *It would be conceivable to have a reconstruction of an atomic OML L from the poset $\mathcal{B}(L)$ of its Boolean subalgebras along the following lines:¹*

¹We thank the anonymous referee for suggesting this alternative.

- a) First, one identifies the atoms of L from $\mathcal{B}(L)$. Of course, this is not entirely straightforward, since atoms in $\mathcal{B}(L)$ are of the form $\{0, P, P^\perp, 1\}$ for arbitrary elements of L , not just atom - co-atom pairs, but the identification can be made using only the order-theoretic information encoded within $\mathcal{B}(L)$.
- b) Then, by considering which atoms of L jointly lie in which elements of $\mathcal{B}(L)$, one identifies the compatibility (i.e., orthogonality) relations between atoms.
- c) Finally, one attempts to reconstruct L in a bottom-up fashion, starting from the sets of pairwise orthogonal atoms, using the additional information about relations between elements of L and atoms which is encoded within $\mathcal{B}(L)$

If one uses a bottom-up reconstruction then, even when L is a Boolean algebra, one must pay attention to the fact that infinite atomic Boolean algebras may have the same set of atoms without being isomorphic. A well-known example is given by the following two Boolean algebras: on the one hand, $B_1 = P\mathbb{N}$, the power set of the set of natural numbers, with intersections as meets, unions as joins, and complements (with respect to $\mathbb{N}$) as complement operation; on the other hand the Boolean algebra B_2 , whose elements are the finite and the cofinite subsets of $\mathbb{N}$ (a subset S of $\mathbb{N}$ is cofinite if $S' = \mathbb{N} \setminus S$ is finite), with intersections, unions and complements as operations.

Clearly, the atoms in both B_1 and B_2 are the singleton subsets of $\mathbb{N}$, and both algebras are atomic and infinite. Yet, B_1 contains more elements than B_2 (uncountably many, vs. countably many), so the algebras are not isomorphic. This relates to the fact that B_1 is complete, while B_2 is not complete.

Our reconstruction algorithm, to be presented below, is also based on the poset $\mathcal{B}(L)$ of Boolean subalgebras of a given atomic OML L . However, in our reconstruction we will follow a top-down route which can be summarized as follows: if $V = \{0, P, P^\perp, 1\}$ is an atom of $\mathcal{B}(L)$, we use the order structure of $\mathcal{B}(L)$ to identify Boolean sup-algebras of V that contain only P , or respectively $P^\perp$, and atoms of L as generating elements (such algebras are called minimal spiked Boolean super-algebras below). We then form equivalence classes, corresponding to P , or respectively $P^\perp$, and show that the order on generic elements P, Q of L can be deduced from the structure of these equivalence classes (which is determined by the order on $\mathcal{B}(L)$).

Thus, instead of singling out atoms of L and proceeding from them, we use the order-theoretic information encoded within $\mathcal{B}(L)$ to identify the elements of L directly – they are in a two-to-one correspondence with the atoms of $\mathcal{B}(L)$. We then show that the order between generic elements of L can also be deduced from the order structure of $\mathcal{B}(L)$.

We now briefly consider the special case of atomic Boolean algebras in order to make clear how, for example, the two infinite atomic Boolean algebras B_1, B_2 are distinguished by our reconstruction procedure:² the poset (in fact, lattice) $\mathcal{B}(B)$ of Boolean subalgebras of a Boolean algebra B has atoms itself, namely the Boolean subalgebras of the form $\{0, P, P^\perp, 1\}$, where P is an arbitrary element of B (not necessarily an atom). Let B_1, B_2 be the two atomic Boolean algebras presented above. Since B_1 has more elements than B_2 , the poset $\mathcal{B}(B_1)$ has more atoms than the poset $\mathcal{B}(B_2)$, so B_1 and B_2 can be distinguished at the level of their posets of Boolean subalgebras, and our reconstruction makes use of this.

Let L be an atomic orthomodular lattice with 0 and 1, and let $\mathcal{B}(L)$ be the set of Boolean subalgebras (BSAs) of L , partially ordered under inclusion. Two elements P and Q of L are orthogonal if $P \leq Q^\perp$, where $P^\perp$ denotes the orthocomplement of P . Orthogonality also implies that the meet of P and Q is equal to 0. It is clear that every element P of L is contained in at least one Boolean subalgebra V of L (for example in $V_P = \{0, P, P^\perp, 1\}$).

Let $\mathcal{F} = \{P_1, P_2, \dots, P_n \dots\}$ be a (possibly infinite) family of pairwise orthogonal elements in L with join 1. Then $\mathcal{F}$ generates an atomistic BSA $V \subseteq L$. The elements in $\mathcal{F}$ are the atoms of the V and since each element of V is a join of elements in $\mathcal{F}$, V is an atomistic BSA.

We say that a BSA V generated by a family $\mathcal{F}$ as above has **dimension** $n = \#\mathcal{F}$, the cardinality of $\mathcal{F}$. In general, not every BSA V of an atomic orthomodular lattice L is generated by a family $\mathcal{F}$ of pairwise orthogonal elements,³ but each element of L is contained in some BSA, as we have remarked in the previous paragraph.

From now on, we will only consider those BSAs in $\mathcal{B}(L)$ which are generated by families of pairwise orthogonal elements. This allows us to describe inclusion relations within $\mathcal{B}(L)$ in terms of grouping and splitting actions. We will write $\mathcal{F}_V$ for the family of join 1, pairwise orthogonal elements generating a BSA V .

Definition 6.1.2. If $\mathcal{F}$ and $\mathcal{G}$ are two families of pairwise orthogonal elements with join 1, we say that $\mathcal{G}$ is obtained by **grouping** the elements in $\mathcal{F}$ if any $Q \in \mathcal{G}$ can be written as a join of elements in $\mathcal{F}$. Let S_Q denote the set of elements in $\mathcal{F}$ that have join Q . The fact that the elements in $\mathcal{G}$ are pairwise orthogonal implies that the sets S_Q , $Q \in \mathcal{G}$, are pairwise disjoint. If $\mathcal{G}$ is obtained by grouping the elements in $\mathcal{F}$, we say that $\mathcal{F}$ is obtained by **splitting** the elements in $\mathcal{G}$.

²In general, by the result by Sachs [113], two Boolean algebras have order-isomorphic posets of Boolean subalgebras if and only if they are isomorphic.

³An example is $\mathcal{P}(\mathcal{H})$, the projection lattice on an infinite-dimensional Hilbert space, which has complete Boolean sublattices that have no atoms at all, e.g. the projection lattice of the abelian von Neumann algebra generated by the position operator. There also are Boolean sublattices of $\mathcal{P}(\mathcal{H})$ that have some atoms, but are not generated by them.

The BSAs contained in a BSA V are obtained from grouping the elements in $\mathcal{F}_V$ while the algebras which contain V , if they exist, are obtained from V by splitting the elements in $\mathcal{F}_V$.

A 2-dimensional BSA $V \subseteq L$ is generated by two complementary elements. We can find out from the order relations within $\mathcal{B}(L)$ when one (or both) of these elements are atoms. This result will be useful later in our reconstruction of the lattice L .

Lemma 6.1.3. *Given an atomistic ortholattice L and a 2-dimensional BSA V of L , we have three possible scenarios:*

- i) if V is maximal in $\mathcal{B}(L)$ then its generating elements are complementary atoms.*
- ii) if V is included in a 3-dimensional BSA W which is maximal in $\mathcal{B}(L)$ then V is generated by an atom of L together with its complement which is a join of two atoms in L . Moreover, W contains precisely two other 2-dimensional BSAs, apart from V itself.*
- iii) if V is neither maximal, nor included in a maximal BSA, then V contains an atom of L if and only if all 4-dimensional BSAs $W \subseteq L$ which contain V also contain precisely three 3-dimensional BSAs, V_1, V_2 and V_3 , such that $V \subset V_i, i \in \{1, 2, 3\}$.*

Proof: For the first two statements, it is sufficient to observe that a BSA is maximal in $\mathcal{B}(L)$ if neither of its generating elements can be split. Since L is atomistic, this implies that the generating elements of a maximal BSA must be atoms of L . This proves the first statement.

For the second statement note that, W being maximal, must be generated by three pairwise orthogonal atoms, call them P, Q and R which add up to the identity. The only BSAs included in W are those generated either by $\{P, Q \vee R\}$ or $\{Q, P \vee R\}$ or $\{R, P \vee Q\}$, so V must also be generated by one of these three families.

For the third statement, let $\mathcal{F}_V = \{P, P^\perp\}$ denote the generating family of the 2-dimensional BSA V . If P is an atom of L then any 4-dimensional algebra W which contains V is obtained by splitting $P^\perp$ into three elements, since P is an atom and cannot be split. Hence, W has generating family $\mathcal{F}_W = \{P, Q_2, Q_3, Q_4\}$. There are precisely three sub-BSAs of W which contain V . These are given by

$$\mathcal{F}_{V_1} = \{P, Q_2 \vee Q_3, Q_4\},$$

$$\mathcal{F}_{V_2} = \{P, Q_2 \vee Q_4, Q_3\},$$

$$\mathcal{F}_{V_3} = \{P, Q_3 \vee Q_4, Q_2\}.$$

Note that there are three other ways of grouping the elements in W to obtain a 3-dimensional BSA. The resulting 3-dimensional BSAs V_i , $i = 4, 5, 6$ do not contain V , since it is not possible to obtain the element P_1 by grouping the elements generating these other algebras.

On the other hand, consider a 2-dimensional BSA given by $\mathcal{F}_{\tilde{V}} = \{Q, Q^\perp\}$ generated by two orthogonal elements which are not atoms. Since Q is not an atom, it is possible to write it as a join of two orthogonal non-zero elements (in L), that is $Q = Q_1 \vee Q_2$. Similarly, it is possible to express $Q^\perp$ as the join of some orthogonal Q_3 and Q_4 . The BSA $\mathcal{F}_{\tilde{W}} = \{Q_1, Q_2, Q_3, Q_4\}$ is a 4-dimensional algebra which includes $\tilde{V}$, but only two of its sub-BSAs also contain $\tilde{V}$, namely

$$\begin{aligned}\mathcal{F}_{\tilde{V}_1} &= \{Q_1 \vee Q_2, Q_3, Q_4\}, \\ \mathcal{F}_{\tilde{V}_2} &= \{Q_1, Q_2, Q_3 \vee Q_4\}.\end{aligned}$$

□

6.2 Spiked BSAs

Note that a family of pairwise orthogonal atoms of L with join 1 generates a **mBSA** (maximal Boolean subalgebra) of L .

Definition 6.2.1. A *sub-mBSA* of L is a BSA of L generated by a family $\mathcal{F}$ of pairwise orthogonal elements with join 1 with the property that only one element in $\mathcal{F}$ is the join of two atoms in L , while all others are atoms in L .

Definition 6.2.2. An algebra is *spiked* if it is either a mBSA, or is generated by a family $\mathcal{F}$ of pairwise orthogonal elements with join 1 which contains precisely one non-atom of the lattice L (we call this the **leading** element), while all other elements of $\mathcal{F}$ are atoms of L . If an algebra is spiked, we will say that its family $\mathcal{F}$ of generating elements is also spiked.

Remark 6.2.3. The concept of a spiked Boolean subalgebra of an orthomodular lattice is related to the concept of a principal dual subalgebra of a Boolean algebra, introduced by Sachs [113].⁴ Indeed, since all the atoms of a Boolean algebra B must be pairwise orthogonal, the collection of atoms in any spiked family of elements generates a principal ideal of B , while the leading element corresponds to the dual of this ideal [75]. Their union is a principal dual subalgebra of B according to the definition by Sachs [113]. For the general case when L is an orthomodular lattice, every spiked (non-maximal) BSA will also be a principal dual subalgebra of every Boolean subalgebra B of L that contains the given spiked BSA. However, there is no canonical choice for

⁴We would like to thank the anonymous referee for pointing out this connection.

the Boolean algebra B of which a spiked BSA of L is a principal dual subalgebra, so we have used a non-relational term for denoting those BSAs of L which enjoy the property described in Definition 6.2.2.

Definition 6.2.4. Given a BSA V which is not a mBSA, we say that a BSA W is a **successor** of V if $V \subsetneq W$ and there is no BSA W' such that $V \subsetneq W' \subsetneq W$. We call a successor of a BSA V , if it exists, a double successor of V .

Note that in terms of generating elements, if W is a successor of V then the family of elements generating W is obtained from the family of elements generating V by splitting precisely one element into two pairwise orthogonal elements.

Since L is atomic and orthomodular, such a splitting is possible whenever V is not a mBSA (i.e. when its generating family contains at least one non-atom of L). This is because any non-atomic element P of an atomic lattice must be larger than some atom Q and the orthomodularity condition then allows us to write P as the join of Q and $Q^\perp \wedge P$, which are easily seen to be pairwise orthogonal. Moreover, note that any element of L can be written as a join of pairwise orthogonal atoms of L .

Proposition 6.2.5. Let $V \in \mathcal{B}(L)$ be a BSA generated by the (possibly infinite) family of elements $\mathcal{F}_V = \{P_1, P_2, \dots, P_k, \dots\}$. If we assume that V is neither a mBSA, nor a sub-mBSA, then V is spiked if and only if all double successors of V contain precisely three successors of V .

Proof: Completely analogous to the proof of the third statement of Lemma 6.1.3. $\square$

This result is important because it shows that the order structure of $\mathcal{B}(L)$ allows us to decide whether a given BSA $V \in \mathcal{B}(L)$ is spiked or not. Since every spiked BSA has a distinguished leading element, one can guess that we want to somehow link the elements of the lattice L to the spiked BSAs of $\mathcal{B}(L)$ using the information encoded within the order structure of $\mathcal{B}(L)$, which is what we will do in the following section.

Let V be a 2-dimensional BSA generated by $\mathcal{F}_V = \{P, P^\perp\}$, and let $\mathcal{S}_V$ be the set of spiked BSAs which contain V . The generating family $\mathcal{F}$ of an element $\tilde{V}$ of $\mathcal{S}_V$ is obtained either by completely splitting P into pairwise orthogonal atoms and splitting $P^\perp$ into a spiked family of elements, or by completely splitting $P^\perp$ into pairwise orthogonal atoms and splitting P into a spiked family of elements.

The set $\mathcal{S}_V$ of spiked BSAs which contain V is partially ordered under inclusion.

- (a) If V is not spiked, the generating family $\mathcal{F}$ of a *minimal* element in $\mathcal{S}_V$ with respect to this partial order is obtained by either taking P as the leading element and splitting $P^\perp$

into atoms, or by taking $P^\perp$ as leading projection and splitting $P^\perp$ into atoms. Let $\mathcal{M}_V$ denote the set of minimal elements in $\mathcal{S}_V$. We call $\mathcal{M}_V$ the set of minimal spiked sup-BSAs of V in $\mathcal{B}(L)$.

- (b) If V is spiked, the minimal element of $\mathcal{S}_V$ which contains V will of course be V itself. Hence for a spiked 2-dimensional BSA V we establish by convention the set $\mathcal{M}_V$ to be the set of all mBSAs which contain V , as these algebras correspond to keeping the atom fixed and completely splitting the co-atom, together with V itself which corresponds to keeping the co-atom fixed.
- (c) if V is spiked and submaximal, we again define $\mathcal{M}_V$ to be the set of all mBSAs which contain V together with V itself.
- (d) If V is spiked and maximal then it is generated by a pair of orthocomplementary atoms. These two atoms are not comparable to any other elements in the lattice L except for the top and bottom elements. The set $\mathcal{M}_V$ contains only one element, namely V itself.

6.3 Reconstructing L from $\mathcal{B}(L)$

As we have already remarked, every 2-dimensional BSA V with $\mathcal{F}_V = \{P, P^\perp\}$ is generated by two complementary elements. Hence there is an obvious two-to-one mapping from L to the 2-dimensional elements of $\mathcal{B}(L)$, which of course are the atoms of the poset $\mathcal{B}(L)$. Therefore, in order to generate all the elements of the atomic orthomodular lattice L from the poset $\mathcal{B}(L)$, we need to assign two elements (corresponding to the two elements $P, P^\perp$) to each 2-dimensional BSA V with $\mathcal{F}_V = \{P, P^\perp\}$.

The minimal spiked sup-BSAs of a given 2-dimensional BSA V make good candidates for this assignment. On the one hand, they can be characterised using only information derived from the poset structure of $\mathcal{B}(L)$, on the other hand, a minimal spiked sup-BSA of V can be identified with its leading element, which is one of the two generating elements of V . Yet, this would give us a many-to-one mapping in general, since there are many (e.g. in $\mathcal{B}(\mathcal{H})$ continuously many) minimal spiked sup-BSAs of V with the same leading element, corresponding to the many possible ways of splitting its complement. Therefore, it will make sense to define two equivalence classes of algebras within $\mathcal{M}_V$ consisting of those algebras whose generating families of elements have the same leading element.

In the non-degenerate cases (a-c) above, our task is to identify these two equivalence classes using the information encoded within the order structure of $\mathcal{B}(L)$. By partitioning the

sets $\mathcal{M}_V$ into two equivalence classes, we are in effect identifying all pairs of complementary elements of the lattice L . Later we will see how the order relations between non-complementary elements can be replicated using the corresponding equivalence classes.

Of course, in the degenerate case (d) when V is also maximal, we already know that V is generated by two orthocomplementary atoms, and since these are not comparable to any other elements of L , the set $\mathcal{M}_V$ does not need any further analysis.

For a spiked 2-dimensional BSA V with $\mathcal{F}_V = \{P, P^\perp\}$, where P is an atom, it is easy to establish what the two equivalence classes should be. One of them, call it $\mathcal{R}_V$, ought to contain the mBSAs which contain V – this corresponds to keeping the atom P as the ‘leading’ element and completely splitting its complement $P^\perp$ into atoms (this is a slight abuse of terminology, since there is no leading element in a mBSA). The other equivalence class, call it $\mathcal{S}_V$, ought to contain only V itself – this corresponds to keeping the co-atom $P^\perp$ as the leading element.

Similarly, for a 2-dimensional BSA W generated by an element P that is the join of two atoms, together with its complement $P^\perp$, we define one equivalence class to contain all the sub-mBSAs in $\mathcal{M}_W$ and the other one to contain all the 3-dimensional BSAs in $\mathcal{M}_W$.

For non-spiked 2-dimensional BSA whose (minimal) generating elements are joins of 3 or more atoms, the two equivalence classes can be determined by considering the inclusion relations between elements belonging to different sets of minimal spiked sup-BSAs, as we will show now.

Lemma 6.3.1. *If V is a non-spiked 2-dimensional BSA whose generating elements are joins of 3 or more atoms, and if $A, B \in \mathcal{M}_V$, then A and B have the same leading element if and only if there exists some non-spiked 2-dimensional $W \neq V$ and $C, D \in \mathcal{M}_W$ such that $A \subseteq C$ and $B \subseteq D$.*

Proof: Assume that $\mathcal{F}_W = \{Q, Q^\perp\}$ and $\mathcal{F}_V = \{P, P^\perp\}$ and that $A, B \in \mathcal{M}_V$ and $C, D \in \mathcal{M}_W$ such that $A \subseteq C$ and $B \subseteq D$. If P_A, P_B, P_C and P_D are the respective leading elements of A, B, C and D (it makes sense to speak about leading elements, since $\mathcal{M}_V$ and $\mathcal{M}_W$ do not contain any mBSAs, as neither V nor W are spiked BSAs), the inclusion relations imply that $P_C \leq P_A$ and $P_D \leq P_B$.

Note at this point that the leading element of a minimal spiked sup-BSA of V must be equal to either P or $P^\perp$ (hence $P_A, P_B \in \{P, P^\perp\}$), while the leading element of a minimal spiked sup-BSA of W must be equal to either Q or $Q^\perp$ (hence $P_C, P_D \in \{Q, Q^\perp\}$). Assume towards a contradiction that $P_A \neq P_B$. Then P_A and P_B must be complementary elements. But if $P_A^\perp = P_B$, then P_C and P_D must also be complementary elements, otherwise the inclusion relations would imply that $P_A \geq P_C$ and $P_A^\perp \geq P_D = P_C$, which is impossible. This means that

$P_C^\perp = P_D$. However, this leads to a contradiction, since

$$P_C^\perp = P_D \leq P_B = P_A^\perp \implies P_C \geq P_A,$$

but $P_C = P_A$ is not possible since $W \neq V$. Hence P_A must be equal to P_B .

On the other hand, if A and B have the same leading element, then their generating families are of the form $\mathcal{F}_A = \{P, R_1, R_2, \dots\}$ and $\mathcal{F}_B = \{P, S_1, S_2, \dots\}$, and it is possible to write P as the join of two orthogonal elements Q and Z , where Q is an atom, and Z is not an atom. The BSAs C and D given by

$$\mathcal{F}_C := \{Q, Z, S_1, S_2, \dots\} \text{ and } \mathcal{F}_D := \{Q, Z, R_1, R_2, \dots\}$$

are sup-BSAs of A and B , respectively, and they belong to the set of minimal spiked sup-BSAs of the non-spiked 2-dimensional BSA W with $\mathcal{F}_W = \{Q, Q^\perp\}$. $\square$

Once the equivalence classes on the sets of minimal spiked sup-BSAs have been established, it is possible to define an order $\preceq$ on them which replicates the order within the lattice of elements.

Definition 6.3.2. *If $[X]$ and $[Y]$ are two equivalence classes corresponding to non-spiked BSAs, we say that $[X] \preceq [Y]$ if there exists $A \in [X]$ and $B \in [Y]$ such that $A \supseteq B$.*

If $A \supseteq B$ as above and $\mathcal{F}_A = \{P, R_1, R_2, \dots\}$ with all the R_i atoms while $\mathcal{F}_B = \{Q, S_1, S_2, \dots\}$ with all the S_i atoms, then the generating elements of B are obtained by grouping the generating elements in A . This implies that the leading element of B (which is the only non-atom) must be equal to a join of generating elements of A . This join must include the leading element of A among its terms, as this is the only possible way of grouping the generating elements of A into a spiked family. Hence there is some index set I such that

$$Q = P \vee \bigvee_{i \in I} S_i$$

and hence $Q \geq P$.

Since all elements of $[X]$ have the same leading element and similarly, all elements of $[Y]$ have the same leading element, the order relation introduced in Definition 6.3.2 is well defined.

Moreover, given any two elements Q and P of an atomic orthomodular lattice which are neither atoms nor co-atoms, and which satisfy the order relation $Q \geq P$ within L , one has $Q = P \vee (P^\perp \wedge Q)$, and we know that $P^\perp \wedge Q$ can be expressed as a join of pairwise orthogonal

atoms. Hence the equivalence classes corresponding to elements of L which are neither atoms nor co-atoms, will always be related by $\preceq$ whenever their corresponding leading elements are related within the lattice L .

We have to use a different approach for defining the order relations which involve the equivalence classes corresponding to atoms and co-atoms of L .

Recall first that for a spiked 2-dimensional BSA V , the set $\mathcal{R}_V$ denotes the mBSAs which contain V , and this is the equivalence class which corresponds to the atom of V , while $\mathcal{S}_V$ denotes the one member equivalence class (containing only V itself) which corresponds to the co-atom of V .

Definition 6.3.3. *If V and W are 2-dimensional BSAs such that V is spiked and W is not spiked, and $[X] \subset \mathcal{M}_W$, then $\mathcal{R}_V \prec [X]$ if there exists $A \in \mathcal{M}_W - [X]$ such that $V \subseteq A$. If this is the case, then also $\mathcal{S}_V \succ \mathcal{M}_W - [X]$.*

Note that if V with $\mathcal{F}_V = \{P, P^\perp\}$ is a spiked BSA with P an atom, and W with $\mathcal{F}_W = \{Q, Q^\perp\}$ is a non-spiked BSA. And if moreover $P < Q$, then $P^\perp > Q^\perp$ and there is some way of decomposing Q into a join over a set of atoms which includes P . Hence V will be contained in some minimal spiked sup-BSA of W which has $Q^\perp$ as its leading element. So $\preceq$ is again well-defined and it captures all the relations between atoms (or co-atoms) and other elements of L .

The only relations from L we have not yet captured are those between the atoms and co-atoms themselves. We do this with the following definition.

Definition 6.3.4. *If both V and W are spiked, their generating elements will contain either equal atoms (if $V = W$) or pairwise orthogonal atoms (if V and W are both contained in some maximal BSA of L) or incomparable atoms. Between equivalence classes we then either have $\mathcal{R}_V = \mathcal{R}_W$ and $\mathcal{S}_V = \mathcal{S}_W$, or we define $\mathcal{R}_V \preceq \mathcal{S}_W$ and $\mathcal{R}_W \preceq \mathcal{S}_V$, or they are incomparable.*

Theorem 6.3.5. *Let $\mathcal{B}_2(L)$ denote the set of 2-dimensional BSAs of an atomistic ortholattice L which are not mBSAs. Let $\mathcal{M}_2(L)$ denote the set of 2-dimensional mBSAs of L . The set*

$$\mathcal{C}(L) := \{\mathcal{M}_V / \sim\}_{V \in \mathcal{B}_2(L)} \cup \{A_W^1, A_W^2\}_{W \in \mathcal{M}_2(L)} \cup \{0, 1\}$$

together with the order $\preceq$ defined in 6.3.2, 6.3.3 and 6.3.4 above, and the additional conventions that 0 and 1 stand for the top and the bottom elements of $\mathcal{C}(L)$, while the A_W^1 s and A_W^2 s are pairs of orthocomplementary atoms which are only comparable with 0 and 1, is isomorphic to L .

Proof:

The lattice isomorphism can easily be constructed using the results presented so far. It sends the top and bottom elements of L to the top and bottom elements of $\mathcal{C}(L)$. The orthocomplementary atoms of L are identified with the elements of the pairs of the form $\{A_W^1, A_W^2\}$. And for all the other elements $P \in L$, if $\{P, R_1, R_2, \dots\}$ is a spiked family of elements, we have the assignment

$$P \mapsto [\{P, R_1, R_2, \dots, \}] \in \mathcal{M}_{\{P, P^\perp\}}$$

□

Chapter 7

Entropy within the Topos Approach

It has been argued [23, 25, 26, 73] that Quantum Mechanics can be understood in a more natural way as a theory about the possibilities and impossibilities of information transfer and processing as opposed to a theory about the mechanics of nonclassical waves or particles. Understanding the representation and manipulation of information can help us shed light on the fundamental structure of both classical and quantum theories and it can lead to fresh insights about the essential differences between these two.

Shannon entropy [117] is used in classical mechanics as a measure of the unpredictability of a physical system. Its analogue in quantum mechanics is the von Neumann entropy [124]. Several generalizations of these entropies have already been considered [13, 119] and it is interesting to ask how much information about a quantum state can be encoded using Shannon and von Neumann entropies. This led to the task of finding a formulation of the notion of entropy within the framework of the topos approach. It turns out that this can be done in a natural way, via the contextual entropy construction, which unifies the concepts of Shannon and von Neumann entropy. It should also be noted that throughout this section we only consider quantum theory on finite-dimensional Hilbert spaces.

In classical physics entropy is a real-valued function defined on the set of probability distributions. In the quantum case, entropy is a real-valued function on the set of density matrices. In this chapter we shall define contextual entropy within the topos approach. There the role of the real numbers is played by a slightly more complicated presheaf based on real numbers, whose ‘points’ are given by its global sections, while the set of states is given by a set of measures on a certain non-commutative space (the spectral presheaf). In analogy with the classical case, our entropy will be a map from this set of measures to the set of global sections of our real-number object.

In particular, in section 7.2 we show how a measure on the spectral presheaf (i.e. a state) gives a canonical probability distribution in each classical context, and how it is therefore possible to associate a Shannon entropy to each classical ‘perspective’ on a state. Contextual entropy is defined in terms of this collection of Shannon entropies, which are shown to form a global section of a certain presheaf of real values (which can differ from context to context). We also show how one can retrieve the von Neumann entropy of a state from such a global section. This confirms our expectation that entropy within the topos approach should ‘look’ like Shannon entropy from each classical perspective, but one can also retrieve the quantum mechanical von Neumann entropy by taking into account all perspectives at the same time.

In fact, one can do even more than this, and we show that contextual entropy encodes enough information to explicitly reconstruct the quantum state from which it originated. This argument relies on a powerful result known as the Schur-Horn Lemma. In Sections 7.3 and 7.4 we show how pure quantum states and general quantum states respectively can be reconstructed from the contextual entropy map.

In Section 7.5 we show that it is possible to adapt other classical entropies within the formalism of the topos approach. In particular, we show how Renyi entropies can be defined within the topos formalism, and moreover we will see that contextual Renyi entropies also encode sufficient information to allow for state reconstruction.

7.1 Entropy in Classical and Quantum Mechanics

7.1.1 Majorization order

In classical physics states can be interpreted as probability measures on the classical state space. In the discrete case, these measures are simply probability distributions over a set of n outcomes. Given two such distributions we would like to tell which one is more ‘uniform’. For this purpose one can compare the two distributions in the so called majorization order (for more details, see [16]). If $\vec{x}$ is a vector with n components representing a probability distribution, let $x^\downarrow$ denote the vector with the same components as $\vec{x}$ but arranged in decreasing order.

Definition 7.1.1. *Given two n -dimensional probability distributions, we say that $\vec{x}$ is majorized by $\vec{y}$, and write $\vec{x} \prec \vec{y}$, if and only if*

$$\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow, \quad \forall k = 1, \dots, n$$

Roughly speaking, this would mean that $\vec{x}$ is a more ‘uniform’ probability distribution. The smallest probability distribution with respect to the majorization order is given by the totally mixed distribution $\vec{x}_* = (1/n, \dots, 1/n)$, while the largest probability distribution with respect to this order is given by any distribution $\vec{y}$ such that $y^\downarrow = (1, 0, \dots, 0)$.

Because the passage of time tends to make things more uniform, many processes in physics occur in the direction of the majorization arrow.

Definition 7.1.2. *A bistochastic matrix is a matrix B with positive entries such that the entries of each row and of each column add up to one.*

Note that a bistochastic matrix preserves positivity, and as the sum of the entries of each column add up to one, it also preserves the l_1 norm, when acting on positive vectors. The fact that its rows also add up to one implies that it leaves the totally mixed distribution, $\vec{x}_*$, invariant. Hence a bistochastic matrix acting on the set of probability distributions will cause some kind of contraction of the probability simplex towards its centre. More precisely, we have the following lemma which we state without proof:

Lemma 7.1.3 (Hardy, Littlewood and Polya [81]). *$\vec{x} \prec \vec{y}$ if and only if there exists a bistochastic matrix B such that $\vec{x} = B\vec{y}$.*

7.1.2 Shannon entropy

Entropy was initially developed within the framework of thermodynamics, where it was introduced in order to explain the loss of energy within thermodynamic systems. Later, Claude Shannon attempted to mathematically quantify the statistical nature of lost information in phone-line signals. To do this, Shannon developed the very general concept of information entropy, a fundamental cornerstone of information theory [117].

Definition 7.1.4. *If $\vec{x}$ is an n -dimensional probability distribution, we define its Shannon entropy to be*

$$\text{Sh}(\vec{x}) = \text{Sh}(x_1, \dots, x_n) = -k \sum_{i=1}^n x_i \ln x_i$$

where k is a positive real number that we usually set equal to 1, and with the convention that $0 \ln 0 = 0$.

The Shannon entropy has many useful properties which follow more or less directly from its definition, and we state some of them below, following the standard presentation made in [16].

- **Positivity:** Clearly, $\text{Sh}(\vec{x}) \geq 0$ for all discrete probability distributions.

- **Continuity:** Shannon entropy is a continuous function of the distribution, where the topology of the probability simplex is the natural one inherited from $\mathbb{R}^n$.
- **Expansibility:** $\text{Sh}(x_1, \dots, x_n) = \text{Sh}(x_1, \dots, x_n, 0)$.
- **Concavity:** $\text{Sh}(p\vec{x} + (1-p)\vec{y}) \geq p\text{Sh}(\vec{x}) + (1-p)\text{Sh}(\vec{y})$ for any $p \in [0, 1]$.
- **Additivity:** If we have a joint probability distribution of two independent random variables described by probability distributions $\vec{x}$ and $\vec{y}$, so that the joint probabilities are products of the individual probabilities, then

$$\text{Sh}(x_1y_1, \dots, x_1y_m, \dots, x_ny_1, \dots, x_ny_m) = \text{Sh}(\vec{x}) + \text{Sh}(\vec{y})$$

- **Subadditivity:** If we have a joint probability distribution

$$\vec{z} = (z_{11}, \dots, z_{1m}, \dots, z_{n1}, \dots, z_{nm})$$

of two random variables given by the probability distributions

$$\vec{x} = \left(\sum_{i=1}^m z_{1i}, \dots, \sum_{i=1}^m z_{ni} \right) \text{ and } \vec{y} = \left(\sum_{j=1}^n z_{j1}, \dots, \sum_{j=1}^n z_{jm} \right)$$

then

$$\text{Sh}(\vec{z}) \leq \text{Sh}(\vec{x}) + \text{Sh}(\vec{y})$$

with equality if and only if the two random variables are independent.

- **Monotonicity:** If $\vec{z}$, $\vec{x}$ and $\vec{y}$ are defined as for the subadditivity property, then the Shannon entropy of the joint probability distribution is larger than the Shannon entropy of each of its parts.
- **Recursion property:** If we coarse grain our probability distribution in the sense that we do not distinguish between all the outcomes then we are dealing with a new probability distribution with components

$$p_1 = \sum_{i=1}^{k_1} x_i, \quad p_2 = \sum_{i=k_1+1}^{k_2} x_i, \quad \dots, \quad p_r = \sum_{i=k_{r-1}+1}^{k_r} x_i$$

for some $0 < k_1 < k_2 < \dots < k_r = n$. One can easily show that

$$\text{Sh}(x_1, \dots, x_n) = \text{Sh}(p_1, \dots, p_r) + \sum_{i=1}^r p_i \text{Sh}\left(\frac{x_1}{p_i}, \dots, \frac{x_{k_i}}{p_i}\right)$$

- **Schur concavity:** Shannon entropy is majorization reversing (or Schur concave). This implies that the maximum is attained for the totally mixed probability distribution $\vec{x}_* = (1/n, \dots, 1/n)$, when $\text{Sh}(\vec{x}_*) = \ln n$, while the minimum, 0, is attained for any of the probability distributions of random variables with one certain outcome.

One can interpret Shannon entropy as a measure of the uncertainty about the outcome of an experiment that is known to occur according to a given probability distribution, or as the expected length of communication needed to specify the outcome that actually occurs. When using the later interpretation we usually set the constant k to be $1/\ln 2$, which simply means that we use logarithms to the base 2 instead of natural logarithms. With this choice the entropy is said to be measured in units of bits: if we have a $n = d^a$ -dimensional distribution then the maximum value of the Shannon entropy is $\log_2 n = a$ bits, which is the length of the string of binary digits one can use to label the outcomes.

To make this interpretation more precise, consider a source that produces outcomes of an infinite sequence of independent and identically distributed random variables. We want to represent each possible outcome by a code word (i.e. a string of binary numbers) such that given any sequence of code words, it can be read in an unambiguous way. The expected length of string needed to code one outcome is defined as

$$L = \sum_i p_i l_i$$

where p_i is the probability that the i^{th} possible outcome will occur and l_i is the length (in bits) of the code word used to represent that outcome. Given the probability distribution, we would like to find optimal codes which minimize the expected length. While we will not describe their construction, we state the following important theorem concerning such codes:

Theorem 7.1.5 (Shannon's noiseless coding theorem). *Given a source distribution $\vec{x}$, let L_* denote the expected length of a code word used in an optimal code. Then*

$$\text{Sh}(\vec{x}) \leq L_* \leq \text{Sh}(\vec{x}) + 1$$

For the proof of this theorem and how to find an optimal code we refer the reader to the literature [37].

7.1.3 Von Neumann entropy

The quantum mechanical analogue of Shannon entropy, which is actually an older concept than Shannon entropy, is called the von Neumann entropy [124]. In finite dimensional quantum mechanics states are usually represented by density matrices. Any given n -dimensional density matrix ρ can be decomposed as

$$\rho = \sum_{i=1}^n \lambda_i |e_i\rangle\langle e_i|$$

where λ_i are the eigenvalues of ρ and $|e_i\rangle$ are the corresponding eigenvectors. This decomposition is unique if ρ has distinct eigenvalues. Since the eigenvalues of a density matrix are non-negative and they sum to 1, we can give the following definition:

Definition 7.1.6. *The von Neumann entropy of the state ρ is defined as the Shannon entropy of the spectrum of ρ :*

$$\text{VN}(\rho) = -\text{Tr}(\rho \ln \rho) = -\sum_{i=1}^n \lambda_i \ln \lambda_i$$

Note that unitarily equivalent states have the same von Neumann entropy.

We shall end this section by stating and proving some of the most important properties of the von Neumann entropy. The proofs will rely on several well-known results which extend inequalities that hold for functions defined on $\mathbb{R}$ to functions of operators. We summarize these results below and direct the reader to [16] for more details.

Definition 7.1.7. *Hermitian operators admit a partial order: $B \geq A$ if and only if $B - A$ is a positive operator. An operator monotone function is a function f defined on Hermitian operators such that $f(A) \leq f(B)$ whenever $A \leq B$.*

Theorem 7.1.8 (Löwner). *A function $f(t)$ on an open interval is operator monotone if and only if it can be extended analytically to the upper half plane and transforms the upper half plane into itself.*

In particular, $f(t) = -t \ln t$ is operator monotone.

Definition 7.1.9. *An operator concave function is a function f such that*

$$f(pA + (1-p)B) \leq pf(A) + (1-p)f(B), \quad \forall p \in [0, 1]$$

A continuous function mapping $[0, \infty)$ to itself is operator concave if and only if f is operator monotone. An operator convex function f is a function such that $-f$ is operator concave.

Lemma 7.1.10 (Klein's inequality). *If f is an operator convex function and A and B are Hermitian operators then*

$$\mathrm{Tr}[f(A) - f(B)] \geq \mathrm{Tr}[(A - B)f'(B)]$$

with equality if and only if $A = B$.

In particular, if we restrict ourselves to $f(t) = t \ln t$ we have

$$\mathrm{Tr}(A \ln A - A \ln B) \geq \mathrm{Tr}(A - B)$$

with equality if and only if $A = B$.

Now we are ready to have a look at some of the properties of von Neumann entropy.

- **Positivity:** It is clear that von Neumann entropy is positive. Moreover $\mathrm{VN}(\rho)$ vanishes if and only if $\rho = |\psi\rangle\langle\psi|$, for some unit vector $|\psi\rangle$ (i.e. if and only if ρ is a pure state).
- **Continuity:** Von Neumann entropy is a continuous function of the eigenvalues of ρ seen as a vector in $\mathbb{R}^n$ with its standard topology..
- **Concavity:** Is a direct consequence of Klein's inequality: let $\rho = p\sigma + (1-p)\omega$, $0 \leq p \leq 1$. We can use the particular case of the inequality, with $B = \rho$ and $A = \sigma$ and then $A = \omega$. This gives us two sets of inequalities:

$$\mathrm{Tr}(\sigma \ln \rho) \leq \mathrm{Tr}(\sigma \ln \sigma + \rho - \sigma)$$

$$\mathrm{Tr}(\omega \ln \rho) \leq \mathrm{Tr}(\omega \ln \omega + \rho - \omega)$$

Multiplying the first inequality by p , the second one by $(1-p)$ and adding, we obtain

$$\mathrm{Tr}(\rho \ln \rho) = p\mathrm{Tr}(\sigma \ln \rho) + (1-p)\mathrm{Tr}(\omega \ln \rho) \leq p\mathrm{Tr}(\sigma \ln \sigma) + (1-p)\mathrm{Tr}(\omega \ln \omega) + \rho - \rho$$

Reversing the sign gives us $\mathrm{VN}(\rho) \geq p\mathrm{VN}(\sigma) + (1-p)\mathrm{VN}(\omega)$.

- **Subadditivity:** This is a property concerning composite systems. Let ρ be a state defined on a Hilbert space which is isomorphic to a tensor product of two Hilbert spaces. If ρ_1 and ρ_2 are the reduced density matrices obtained by taking the partial trace of ρ over the second and the first subsystem respectively, then subadditivity is expressed by the following equation

$$\mathrm{VN}(\rho) \leq \mathrm{VN}(\rho_1) + \mathrm{VN}(\rho_2)$$

Again, this can be proved using Klein's inequality, this time with $A = \rho$ and $B = \rho_1 \otimes \rho_2 = (\rho_1 \otimes I)(I \otimes \rho_2)$, with the observation that $\text{Tr}(A - B) = 0$ since both are density matrices. We then have

$$\begin{aligned}\text{Tr}(\rho \ln \rho) &\geq \text{Tr}(\rho \ln \rho_1 \otimes \rho_2) \\ &= \text{Tr}(\rho(\ln \rho_1 \otimes I + \ln I \otimes \rho_2)) \\ &= \text{Tr}(\rho_1 \ln \rho_1) + \text{Tr}(\rho_2 \ln \rho_2)\end{aligned}$$

which becomes subadditivity when we reverse the sign.

- **Additivity:** It is not hard to see that if (and only if) $\rho = \rho_1 \otimes \rho_2$ we have equality in the above and

$$\text{VN}(\rho) = \text{VN}(\rho_1) + \text{VN}(\rho_2)$$

- **Recursion:** If the density matrices ρ_i are defined on orthogonal subspaces H_i of a Hilbert space $H = \oplus_i H_i$, then the density matrix $\rho = \sum_i p_i \rho_i$ has the von Neumann entropy

$$\text{VN}(\rho) = \text{Sh}(\vec{p}) + \sum_i p_i \text{VN}(\rho_i)$$

This follows from the recursion property of Shannon entropy and the fact that if the matrix ρ_i has eigenvalues λ_j^i then the eigenvalues of ρ will be of the form $p_i \lambda_j^i$.

Monotonicity: This is a property concerning composite systems, and it can be expressed by the following equation: $\text{VN}(\rho) \geq \text{VN}(\rho_1)$. However, this does not hold in general for von Neumann entropy. It is known that a composite system can be in a pure state (and so its von Neumann entropy will vanish) while its subsystems can be mixed (and so have positive von Neumann entropy).

7.1.4 Renyi entropy

Renyi entropies form a one parameter family of Schur concave, additive entropies defined by

$$R_q(p_1, \dots, p_n) = \frac{1}{1-q} \ln \left[\sum_{i=1}^n p_i^q \right], \quad \forall q \geq 0$$

Special cases of the Renyi entropies include $q = 0$, which is the logarithm of the number of non-zero components of the distribution and is known as the *Hartley entropy*. When $q \rightarrow 1$, we

have the Shannon entropy, and when $q \rightarrow \infty$ the *Chebyshev entropy* $R_\infty = -\ln p_{\max}$, a function of the largest component $p_{\max}$.

For any given probability vector $\vec{p}$ the Renyi entropy is a continuous, non-increasing function of its parameter:

$$R_t(\vec{p}) \leq R_q(\vec{p}), \quad \forall t > q$$

To illustrate this, we have plotted in Figure 1 several Renyi entropies as functions of a probability distribution with two variables. Note that since Renyi entropies are Schur concave, their maximum value is attained for the totally mixed probability distribution, in which case $R_q(1/n, \dots, 1/n) = \ln n$.

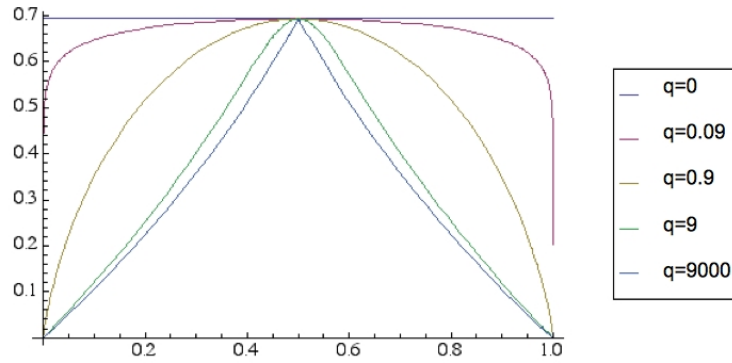


Figure 1: $R_q(x, 1-x)$ for $q = 0, 0.5, 3, q \rightarrow 1$ and $q \rightarrow \infty$

At each parameter q , the quantum Renyi entropy can be defined on the set of density matrices as the classical Renyi entropy of the corresponding spectra:

$$R_q(\rho) = \frac{1}{1-q} \ln \text{Tr}(\rho^q) = \frac{1}{1-q} \ln \left[\sum_{i=1}^n \lambda_i^q \right] = R(\lambda_1, \dots, \lambda_n)$$

Quantum Renyi entropy assigns the value 0 to pure states exclusively, and $\ln n$ to the maximally mixed state $\rho_* = \frac{1}{n}I$.

Note that Renyi entropies are not concave in general, nor subadditive, but they are Schur-concave, additive and monotone [16].

7.2 Contextual entropy

7.2.1 Measures and partial traces

We saw that, given a probability measure μ on the clopen subobjects of a spectral presheaf, if we fix a subobject $\underline{S}$ of $\underline{\Sigma}$ we obtain a map from $\mathcal{V}(N)$ to $[0, 1]$. We assume from now on that we

are given a certain probability measure μ , representing a quantum state. Then we can adopt a different perspective and instead of looking at a fixed subobject we can look at a fixed context V . There is a lattice isomorphism α_V between the projections in V and the clopen subsets of $\underline{\Sigma}_V$. Hence from μ we can also obtain a map

$$\begin{aligned}\mu|_V : \mathcal{P}(V) &\longrightarrow [0, 1] \\ P &\longmapsto \mu(S_P)\end{aligned}$$

where $S_P = \alpha_V(P) \subseteq \underline{\Sigma}_V$.

Using this new perspective, we can show that measures on the spectral presheaf associated to a matrix algebra behave well with respect to the partial trace. This result has a certain physical significance. We have already seen that there is a bijective correspondence between states and probability measures, and we now show that moreover these measures capture the essential information theoretic property of the partial trace in a natural way. Thus, if we are given a measure corresponding to a composite state, we can obtain its partial traces in a direct way by simply considering its restrictions to contexts of a particular form. Intuitively, we would expect these contexts to be precisely those which only encode information related to the first subsystem (if we want to trace out the second one) or vice versa, and we will see that this will indeed be the case.

Note also that this result will be useful for us later on, when discussing the subadditivity property of our contextual entropy.

Proposition 7.2.1. *Consider a state ρ on the matrix algebra $\mathcal{M}_{nm} \simeq \mathcal{M}_n \otimes \mathcal{M}_m$. Let $\rho_1 = \text{Tr}_2(\rho) \in \mathcal{M}_n$ and $\rho_2 = \text{Tr}_1(\rho) \in \mathcal{M}_m$ be the partial traces of ρ . Then if $V \in \mathcal{V}(\mathcal{M}_n)$ and $\mathbb{C}I_m$ denotes the trivial subalgebra of $\mathcal{M}_m$ we have*

$$\mu_\rho|_{V \otimes \mathbb{C}I_m} = \mu_{\rho_1}|_V$$

Similarly, if $W \in \mathcal{V}(\mathcal{M}_m)$ and $\mathbb{C}I_n$ denotes the trivial subalgebra of $\mathcal{M}_n$ we have

$$\mu_\rho|_{\mathbb{C}I_n \otimes W} = \mu_{\rho_2}|_W$$

Proof: Note first that there is a lattice isomorphism between the domains of definition of $\mu_\rho|_{V \otimes \mathbb{C}I_m}$ and $\mu_{\rho_1}|_V$ which takes $P \in \mathcal{P}(V)$ to $P \otimes I_m \in \mathcal{P}(V \otimes \mathbb{C}I_m)$. Then using the definition of measures for states on matrix algebras and the defining property of the partial trace, we have

that

$$\mu_\rho|_{V \otimes \mathbb{C}I_m}(P \otimes I_m) = \text{Tr}(\rho \cdot P \otimes I_m) = \text{Tr}(\rho_1 \cdot P) = \mu_{\rho_1}|_V(P), \quad \forall P \in \mathcal{P}(V \otimes \mathbb{C}I_m)$$

and similarly for the second statement. $\square$

Finally, the fact that μ is a measure implies several properties for $\mu|_V$ which hold for all contexts $V \in \mathcal{V}(N)$, and which we shall state below:

1. $\mu|_V(I) = 1$ and $\mu|_V(0) = 0$
2. $\mu|_V(P \vee Q) + \mu|_V(P \wedge Q) = \mu|_V(P) + \mu|_V(Q)$
3. in particular, if P and Q are orthogonal then $P \wedge Q = 0$ and $P \vee Q = P + Q$ and hence

$$\mu|_V(P + Q) = \mu|_V(P) + \mu|_V(Q)$$

4. if $P \leq Q$ then $\mu|_V(P) \leq \mu|_V(Q)$

These properties show that $\mu|_V$ is a finitely additive probability measure on the lattice of projections of V .

7.2.2 The entropy of a measure

We saw that in classical probability theory Shannon entropy assigns a real number to every discrete probability distribution. It is known that a von Neumann algebra on a finite-dimensional Hilbert space is simply a matrix algebra, or a finite direct sum of matrix algebras. We will see further on how to associate a distinguished probability distribution to each commutative subalgebra (or context) of a von Neumann algebra of bounded operators on finite dimensional Hilbert space, given a state on the system described by that algebra in the form of a measure on its associated spectral presheaf. Once this is done, we will be able to associate to each context its corresponding Shannon entropy, and moreover we will see that this collection of Shannon entropies fits together in a nice way and gives a global section of a certain real-number presheaf. This is consistent with the basic idea of the topos approach, that of putting together the information obtained from each classical perspective on a quantum system. We will see in later sections that by keeping track of all classical entropies associated to a quantum state we can not only retrieve that state's von Neumann entropy, but also reconstruct the state itself.

Definition 7.2.2. *Let H be an Hilbert Space, $\mathcal{B}(H)$ the algebra of bounded operators in H and $\mathcal{F} \subseteq \mathcal{B}(H)$. The von Neumann commutant of $\mathcal{F}$, usually denoted by $\mathcal{F}'$, is the subset of $\mathcal{B}(H)$*

consisting of all elements that commute with every element of $\mathcal{F}$, that is

$$\mathcal{F}' = \{T \in \mathcal{B}(H) \mid TS = ST, \forall S \in \mathcal{F}\}$$

The von Neumann double commutant $\mathcal{F}$ of is just $(\mathcal{F}')'$ and is usually denoted by $\mathcal{F}''$.

If we consider a set of orthogonal rank-one projections $\{P_1, \dots, P_n\}''$, their double commutant can be shown to be simply $\mathbb{C}P_1 + \dots + \mathbb{C}P_n$.

This shows that in finite dimensions each context V can be generated via the von Neumann double commutant construction in a unique way from a set of pairwise orthogonal projections which add up to the identity. If we denote this canonical set of projections by $\{P_1, P_2, \dots, P_k\}$ then $(\mu|_V(P_1), \mu|_V(P_2), \dots, \mu|_V(P_k))$ is a probability distribution. Hence to each context V we can assign the Shannon entropy of its associated probability distribution:

$$\text{Sh}(\mu|_V(P_1), \mu|_V(P_2), \dots, \mu|_V(P_k)) = - \sum_{i=1}^k \mu|_V(P_i) \ln \mu|_V(P_i)$$

If $V' \supseteq V$ then $V' = \{Q_1^1, \dots, Q_{l_1}^1, Q_1^2, \dots, Q_{l_2}^2, \dots, Q_1^k, \dots, Q_{l_k}^k\}''$, where the Q_i^j s are pairwise orthogonal and

$$\sum_{i=1}^{k_j} Q_i^j = P_j$$

The Shannon entropy associated to V' is related to the Shannon entropy associated to V via the recursion formula:

$$\text{Sh}(V') = \text{Sh}(V) + \sum_{i=1}^k \mu|_V(P_i) \cdot \text{Sh}\left(\frac{\mu|_{V'}(Q_1^i)}{\mu|_V(P_i)}, \frac{\mu|_{V'}(Q_2^i)}{\mu|_V(P_i)}, \dots, \frac{\mu|_{V'}(Q_{l_i}^i)}{\mu|_V(P_i)}\right)$$

Since Shannon entropy is non-negative, it follows that $\text{Sh}(V') \geq \text{Sh}(V)$ and this enables us to give the following definition for the entropy of a measure (and hence of a quantum state).

We call a context k -dimensional if it is generated by k pairwise orthogonal projections which add up to the identity.

Definition 7.2.3. If μ is a measure on the clopen subobjects of the spectral presheaf $\underline{\Sigma}$ then the entropy $E(\mu)$ associated to μ is a global section of the presheaf $[0, \ln n]^{\preceq}$ which at a context $V = \{P_1, P_2, \dots, P_k\}''$ has the value

$$E(\mu)|_V = \text{Sh}(\mu|_V(P_1), \mu|_V(P_2), \dots, \mu|_V(P_k)) = - \sum_{i=1}^k \mu|_V(P_i) \ln \mu|_V(P_i)$$

Note that if the V is a k -dimensional context then the value taken by $E(\mu)$ at V is less than or equal to $\ln k$, and hence for an n -dimensional matrix algebra, the maximal value taken by $E(\mu)$ at any context is $\ln n$. Therefore contextual entropy can be seen as a mapping defined on the set of measures associated to a spectral presheaf:

$$E : \mathcal{M}(\underline{\Sigma}) \longrightarrow \Gamma[0, \ln n]_{\leq}^{\preceq} .$$

Notice that although there is a bijective correspondence between states of a von Neumann algebra and measures on the spectral presheaf associated to it, the above definition does not make any direct reference to the quantum state which the measure corresponds to.

7.2.3 Properties of the contextual entropy

Retrieving the von Neumann entropy

Given a density matrix ρ , there exists at least one orthonormal basis of Hilbert space with respect to which ρ is diagonal. Such a basis corresponds to a set of one-dimensional pairwise orthogonal projections $\{P_1, \dots, P_n\}$, which in turn determine a maximal context V_ρ via the double commutant construction. It is easy to check that the eigenvalues $\{\lambda_i\}_{i=1}^n$ of ρ satisfy $\lambda_i = \text{Tr}(\rho P_i)$. Hence the value assigned to the entropy of the measure μ_ρ at any context V_ρ obtained through the above procedure, is just the von Neumann entropy of the state ρ :

$$\begin{aligned} E(\mu_\rho)_{V_\rho} &= - \sum_{i=1}^n \mu_\rho|_{V_\rho}(P_i) \ln \mu_\rho|_{V_\rho}(P_i) \\ &= - \sum_{i=1}^n \text{Tr}(\rho P_i) \ln \text{Tr}(\rho P_i) \\ &= - \sum_{i=1}^n \lambda_i \ln \lambda_i = \text{VN}(\rho) \end{aligned}$$

The natural question to ask at this stage is whether there is any way of determining the von Neumann entropy of a state if we are given an arbitrarily defined measure μ without being explicitly told which state it corresponds to. It turns out that the answer is yes, since it can be showed that the von Neumann entropy is the minimal value amongst the numbers assigned to the maximal contexts of a von Neumann algebra by our generalized notion of entropy. Proving this result requires the Schur-Horn Lemma, which we state below.

Theorem 7.2.4 (Schur-Horn Lemma). *Let ρ be a Hermitian matrix, and let the vector $(\lambda_1, \dots, \lambda_n)$*

denote its spectrum. Let $(\delta_1, \dots, \delta_n)$ denote its diagonal elements in a given basis. Then

$$(\delta_1, \dots, \delta_n) \preceq (\lambda_1, \dots, \lambda_n) \quad .$$

Conversely, if this equation holds, there exists a Hermitian matrix with spectrum $(\lambda_1, \dots, \lambda_n)$ whose diagonal elements are given by $(\delta_1, \dots, \delta_n)$.

The proof of this useful result can be found in [87].

Theorem 7.2.5. *Given a state ρ , a maximal context V_ρ in which ρ is diagonal, and any other maximal context V , we have*

$$\text{Sh}(\delta_1, \dots, \delta_n) = E(\mu_\rho)_V \geq E(\mu_\rho)_{V_\rho} = \text{Sh}(\lambda_1, \dots, \lambda_n) = \text{VN}(\rho)$$

where we have denoted by δ_i the diagonal entries of ρ in the basis (unique up to phases) in which the elements of V are diagonal matrices.

Proof: The key part of this proof is showing that $E(\mu_\rho)_V$ is equal to the Shannon entropy of the diagonal elements of ρ in some given basis.

Let $V = \{Q_1, \dots, Q_n\}''$. Then

$$E(\mu_\rho)_V = - \sum_{i=1}^n \text{Tr}(\rho Q_i) \ln \text{Tr}(\rho Q_i)$$

Since the Q_i 's are pairwise orthogonal, there exists a unitary U which simultaneously diagonalises them. Using the fact that $\text{Tr}(AB) = \text{Tr}(BA)$ for any two matrices A and B , we can write

$$E(\mu_\rho)_V = - \sum_{i=1}^n \text{Tr}(U \rho U^{-1} U Q_i U^{-1}) \ln \text{Tr}(U \rho U^{-1} U Q_i U^{-1})$$

Since for all i , $U Q_i U^{-1}$ is a diagonal rank-one projection, the collection of numbers given by the traces $\text{Tr}(U \rho U^{-1} U Q_i U^{-1})$ is the same as that which consists of the diagonal entries of $U \rho U^{-1}$, i.e. it is simply the collection of diagonal entries of ρ when expressed in the basis determined by the column vectors of the unitary matrix U . Since these entries are positive and add up to unity they form a probability distribution, and we can say that

$$E(\mu_\rho)_V = \text{Sh}(\delta_1, \dots, \delta_n)$$

The Schur-Horn Lemma together with the fact that Shannon entropy is majorization

reversing give us

$$E(\mu_\rho)_V = \text{Sh}(\delta_1, \dots, \delta_n) \geq \text{Sh}(\lambda_1, \dots, \lambda_n) = \text{VN}(\rho)$$

and we have already seen that $E(\mu_\rho)_{V_\rho} = \text{VN}(\rho)$. $\square$

Remark 7.2.6. *From the proof of the last theorem we can extract an important observation: if we evaluate the contextual entropy of a state ρ at some context V (not necessarily maximal), this will be equal to the contextual entropy of any unitarily equivalent state as long as we evaluate it at a context which is obtained from V through rotation by the same unitary. That is,*

$$E(\mu_\rho)_V = E(\mu_{U\rho U^{-1}})_{UVU^{-1}}$$

This equation expresses a covariance property: it does not matter if we consider the contextual entropy of a state ρ (expressed in some basis) or of the same state expressed with respect to some other basis, $U\rho U^{-1}$, as long as we also adapt the context V that we are considering accordingly, i.e. to UVU^{-1} . As functions, $E(\mu_\rho)$ and $E(\mu_{U\rho U^{-1}})$ are not the same, but they are the same up to a ‘rotation by U of contexts’.

Given the contextual entropy map, the problem of finding a maximal context for which the minimum discussed above is attained is equivalent to the problem of finding the point at which a real-valued function on the group of unitaries $\mathcal{U}(n)$ attains its minimal value. To see why this is the case, let $C := (E_1, \dots, E_n)$ denote the maximal context determined by projections which are diagonal with respect to the computational basis. Any other maximal context $V = (P_1, \dots, P_n)$ can be written as $UCU^{-1} := (UE_1U^{-1}, \dots, UE_nU^{-1})$ for some unitary U . Hence we can construct a real-valued function on the group of unitaries by considering the values which the contextual entropy map takes when it is evaluated on the set of maximal contexts. Explicitly, this map is

$$\begin{aligned} \mathcal{W}_\mu : \mathcal{U}(n) &\longrightarrow \mathbb{R} \\ U &\longmapsto E(\mu)_{UCU^{-1}} \end{aligned}$$

It is possible to use existing optimization algorithms [9,10] in order to determine the point at which this function attains its global minimum.

Contextual vs. Shannon and von Neumann entropies

We will now consider which of the properties of Shannon and von Neumann entropies have counterparts for contextual entropy. An immediate difficulty is posed by the fact that the values of the contextual entropies are not real numbers but global sections of certain presheaves of real numbers, which may live in different topoi, i.e. they may be defined over different base categories. In some cases it is possible to work around this difficulty by adapting the definitions of order relations and algebraic operations on $\mathbb{R}$ to suit our more general framework.

1) Positivity

Both von Neumann and Shannon entropies are positive. Shannon entropy is zero for any probability distribution in which one outcome occurs with 100% certainty and strictly positive otherwise. Similarly, von Neumann entropy is zero for all pure states, and strictly positive for the others.

The contextual entropy does assign non-negative values to all contexts, hence the resulting global section can be thought of as non-negative. The minimum of contextual entropy on maximal contexts is 0 if and only the state is pure. This is entirely analogous to Shannon and von Neumann entropy.

2) Concavity

Shannon entropy is concave: if $\vec{p}$ and $\vec{q}$ are two probability distributions then

$$\text{Sh}(r \cdot \vec{p} + (1-r) \cdot \vec{q}) \geq r\text{Sh}(\vec{p}) + (1-r)\text{Sh}(\vec{q})$$

For von Neumann entropy concavity is defined by a similar formula:

$$\text{VN}(r\rho + (1-r)\sigma) \geq r\text{VN}(\rho) + (1-r)\text{VN}(\sigma)$$

The contextual entropy satisfies a similar property. If ρ and σ are defined on the same Hilbert space $\mathcal{H}$ then for every context $V \in \mathcal{B}(\mathcal{H})$, if V is generated by the projections $\{P_1, \dots, P_k\}$, we have

$$\begin{aligned} E(\mu_{r\rho+(1-r)\sigma})_V &= \text{Sh}(\text{Tr}[(r\rho + (1-r)\sigma)P_1], \dots, \text{Tr}[(r\rho + (1-r)\sigma)P_k]) \\ &= \text{Sh}([r\text{Tr}(\rho P_1) + (1-r)\text{Tr}(\sigma P_1)], \dots, [r\text{Tr}(\rho P_k) + (1-r)\text{Tr}(\sigma P_k)]) \\ &\geq r\text{Sh}(\text{Tr}(\rho P_1), \dots, \text{Tr}(\rho P_k)) + (1-r)\text{Sh}(\text{Tr}(\sigma P_1), \dots, \text{Tr}(\sigma P_k)) \\ &= r \cdot E(\mu_\rho)_V + (1-r)E(\mu_\sigma)_V \end{aligned}$$

Hence contextual entropy is globally concave:

$$E(\mu_{r\rho+(1-r)\sigma}) \geq r \cdot E(\mu_\rho) + (1-r)E(\mu_\sigma), \quad \forall r \in [0, 1]$$

3) Additivity and Subadditivity

Subadditivity a property concerning composite systems. Recall that an entropy is called subadditive if the entropy of a composite system is smaller than the sum of the entropies of its parts. Both von Neumann and Shannon entropies are subadditive. We would like to obtain an inequality of the form

$$E(\mu_\rho) \leq E(\mu_{\rho_1}) + E(\mu_{\rho_2})$$

where ρ is the density matrix representing a composite state and ρ_1 and ρ_2 are the partial traces of ρ . It is not immediately clear how one could define such an inequality, since this time the terms involved are global sections of presheaves over three different base categories. Hence in order to talk about subadditivity in a meaningful way, we must first define a suitable notion of addition between the global sections $E(\mu_{\rho_1})$ and $E(\mu_{\rho_2})$.

In order to see how this might be done, we start by considering some context V of the first subsystem and some other context W of the second subsystem. If $V = \{P_1, \dots, P_k\}''$ and $W = \{Q_1, \dots, Q_r\}''$, from the definition of the entropy we have

$$E(\mu_{\rho_1})|_V = \sum_{i=1}^k \text{Tr}(\rho_1 P_i) \ln \text{Tr}(\rho_1 P_i),$$

$$E(\mu_{\rho_2})|_W = \sum_{j=1}^r \text{Tr}(\rho_2 Q_j) \ln \text{Tr}(\rho_2 Q_j)$$

We can add these two numbers together, and we can use the fact that Shannon entropy is additive for independent probability distributions (i.e. $\sum_{i=1}^k p_i \ln p_i + \sum_{j=1}^r q_j \ln q_j = \sum_{i,j} p_i q_j \ln p_i q_j$) and the fact that $\text{Tr}(\rho_1 P_i) \text{Tr}(\rho_2 Q_j) = \text{Tr}(\rho_1 \otimes \rho_2 P_i \otimes Q_j)$ to obtain

$$E(\mu_{\rho_1})|_V + E(\mu_{\rho_2})|_W = \sum_{i=1, j} \text{Tr}(\rho_1 \otimes \rho_2 P_i \otimes Q_j) \ln \text{Tr}(\rho_1 \otimes \rho_2 P_i \otimes Q_j) = E(\mu_{\rho_1 \otimes \rho_2})|_{V \otimes W}$$

Hence we could use the following requirement for the definition of subadditivity: $E(\mu_\rho)$ should be less than or equal to $E(\mu_{\rho_1 \otimes \rho_2})$ at each context $\tilde{V}$ of the composite system. This definition enables us to say, for instance, that the contextual entropy is additive when $\rho = \rho_1 \otimes \rho_2$. Note that this is a direct consequence of the additivity property of Shannon entropy.

Even when ρ is not equal to $\rho_1 \otimes \rho_2$ the subadditivity property holds in split contexts (i.e. contexts of the form $V \otimes W$) as a consequence of Shannon subadditivity. Consider $\tilde{V} = V \otimes W$, with V and W as above. We know that

$$\mu_{\rho_1}(P_i) = \mu_{\rho}(P_i \otimes I) = \sum_{j=1}^r \mu_{\rho}(P_i \otimes Q_j)$$

for all $i \in \{1, \dots, k\}$ and

$$\mu_{\rho_2}(Q_j) = \mu_{\rho}(I \otimes Q_j) = \sum_{i=1}^k \mu_{\rho}(P_i \otimes Q_j)$$

for all $j \in \{1, \dots, r\}$.

Using the subadditivity property of Shannon entropy we obtain

$$\begin{aligned} E(\mu_{\rho})|_{V \otimes W} &= \text{Sh}(P_1 \otimes Q_1, \dots, P_1 \otimes Q_r, \dots, P_k \otimes Q_1, \dots, P_k \otimes Q_r) \\ &\leq \text{Sh}\left(\sum_{i=1}^k \mu_{\rho}(P_i \otimes Q_1), \sum_{i=1}^k \mu_{\rho}(P_i \otimes Q_2), \dots, \sum_{i=1}^k \mu_{\rho}(P_i \otimes Q_r)\right) + \\ &\quad \text{Sh}\left(\sum_{j=1}^r \mu_{\rho}(P_1 \otimes Q_j), \sum_{j=1}^r \mu_{\rho}(P_2 \otimes Q_j), \dots, \sum_{j=1}^r \mu_{\rho}(P_k \otimes Q_j)\right) \\ &= \text{Sh}(\mu_{\rho_1}(P_1), \mu_{\rho_1}(P_2), \dots, \mu_{\rho_1}(P_k)) + \text{Sh}(\mu_{\rho_2}(Q_1), \mu_{\rho_2}(Q_2), \dots, \mu_{\rho_2}(Q_r)) \\ &= E(\mu_{\rho_1})|_V + E(\mu_{\rho_2})|_W = E(\mu_{\rho_1 \otimes \rho_2})|_{V \otimes W} \end{aligned}$$

Remark 7.2.7. *The fact that the contextual entropy is subadditive in all split contexts can be used to give a more direct proof of the subadditivity property of von Neumann entropy, which avoids using Klein's inequality: if we choose the split context $\tilde{V}$ such that ρ_1 is diagonal in W and ρ_2 is diagonal in W we have from Theorem 7.2.5 that*

$$\text{VN}(\rho) \leq E(\mu_{\rho})|_{V \otimes W} \leq E(\mu_{\rho_1})|_V + E(\mu_{\rho_2})|_W = \text{VN}(\rho_1) + \text{VN}(\rho_2)$$

For contexts which are not split (which we usually call entangled contexts), the subadditivity property does not necessarily hold. Intuitively, we can understand why this happens: the converse of the Schur-Horn lemma implies that for any density matrix ρ , there is some unitary U for which the diagonal of $U\rho U^{-1}$ is the maximally mixed vector $(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$. Let D_n denote the context generated by the set of projections $\{E_{11}, \dots, E_{nn}\}$, where we have fixed our basis such that E_{ii} is the projection with the i^{th} diagonal entry equal to one and all other entries equal to zero. Then

$$E(\mu_\rho)|_{U^{-1}.D_n.U} = E(\mu_{U\rho U^{-1}})|_{D_n} = \text{Sh}((U\rho U^{-1})_{11}, \dots, (U\rho U^{-1})_{nn}) = \ln n$$

There is however no guarantee that the diagonal of $U\rho_1 \otimes \rho_2 U^{-1}$ will also be the maximally mixed vector. Hence the contextual entropy map will assign a smaller value to the state $U\rho_1 \otimes \rho_2 U^{-1}$ at the context D_n . This in turn implies that

$$E(\mu_\rho)|_{U^{-1}.D_n.U} > E(\mu_{\rho_1 \otimes \rho_2})|_{U^{-1}.D_n.U}$$

Note however that explicitly finding the unitary matrix U for which the diagonal of $U\rho U^{-1}$ is the maximally mixed vector is an example of an inverse eigenvalue problem, and is not at all trivial. Algorithms for finding such unitaries do exist, see [41], but they will not help us find a counter-example of the subadditivity property - we may end up finding a unitary which takes both the state ρ and the state $\rho_1 \otimes \rho_2$ to a basis where each has the maximally mixed vector on the diagonal. Instead we can look at particular density matrices and construct general examples for which the subadditivity property fails to hold. The details of these computations can be found in Appendix A.

The conclusion is that subadditivity is not a global property of our contextual entropy. In particular, it may fail to hold at entangled contexts, a behaviour which is not captured either by the classical Shannon entropy, or by its quantum-mechanical counterpart.

4) Monotonicity

Monotonicity is a property which refers to composite systems. Recall that an entropy is called monotone if the entropy of a composite system is larger than each of the entropies of its parts. Shannon entropy is monotone: let $(p_1, \dots, p_n)$ and $(q_1, \dots, q_m)$ be two probability distributions, and let $(p_1 q_1, \dots, p_1 q_m, \dots, p_n q_1, \dots, p_n q_m)$ be the probability distribution obtained by composing them. The Shannon entropy of the latter is at least as large as the Shannon entropies of the each of the former distributions.

Von Neumann entropy on the other hand is not monotone. Let ρ be defined on a Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, and let $\rho_1 = \text{Tr}_2 \rho$ and $\rho_2 = \text{Tr}_1 \rho$ be defined on the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively. We know that a composite system can be in a pure state, in which case $\text{VN}(\rho) = 0$, but its subsystems might be mixed, and then $\text{VN}(\rho_i) > 0$.

When we consider the contextual entropy, we would like to obtain an equation of the form $E(\mu_\rho) \geq E(\mu_{\rho_i})$ for each subsystem i . However, we are again faced with the problem that the entropy of a composite system and the entropies of its parts are global sections of presheaves

which live in different topoi. We try to solve this problem by giving a general method of comparing such global sections.

Definition 7.2.8. Let $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ and let $N = \mathcal{B}(\mathcal{H})$, $N_i = \mathcal{B}(\mathcal{H}_i)$, for $i = 1, 2$. If $\gamma \in \Gamma[0, \ln n]^\preceq$ is a global section of the presheaf of real values which lives in the topos $\text{Sets}^{\mathcal{V}(N)^{op}}$ and $\gamma_i \in \Gamma[0, \ln n_i]^\preceq$ are global sections of the presheaves living in the topoi $\text{Sets}^{\mathcal{V}(N_i)^{op}}$, we say that $\gamma \geq \gamma_1$ if for all contexts $V \in \mathcal{V}(N_1)$, $W \in \mathcal{V}(N_2)$ and all unitaries U :

$$\gamma_{U \cdot V \otimes W \cdot U^{-1}} \geq \gamma_{1_V}$$

Similarly, one can define a binary order relation between γ and γ_2 .

Note that the order relation we have just defined is not a partial order since it is not reflexive. However, with respect to this order we can see that for split contexts our contextual entropy behaves like Shannon entropy: let $\{P_1, \dots, P_n\}$ be the projections generating V and $\{Q_1, \dots, Q_m\}$ the projections generating W . Then $\{P_1 \otimes Q_1, P_1 \otimes Q_2, \dots, P_n \otimes Q_m\}$ are the canonical projections generating $\tilde{V}$. For any ρ , ρ_1 and ρ_2 defined as above we have

$$p_i := \text{Tr}(\rho_1 P_i) = \text{Tr}(\rho P_i \otimes I) = \text{Tr}(\rho P_i \otimes \sum_{j=1}^m Q_j) = \sum_{j=1}^m \text{Tr}(\rho P_i \otimes Q_j) =: \sum_{j=1}^m r_{ij}$$

If we now use the recursion relation for Shannon entropy we have:

$$\text{Sh}(r_{11}, \dots, r_{nm}) = \text{Sh}(p_1, \dots, p_n) + \sum_{i=1}^n p_i \cdot \text{Sh}\left(\frac{r_{i1}}{p_i}, \frac{r_{i2}}{p_i}, \dots, \frac{r_{im}}{p_i}\right)$$

and since all the terms in the sum on the right hand side of the equation are positive we must have

$$\gamma_{V \otimes W} = \text{Sh}(r_{11}, \dots, r_{nm}) \geq \text{Sh}(p_1, \dots, p_n) = \gamma_{1_V}$$

for all contexts V and W .

For entangled contexts our contextual entropy behaves like von Neumann entropy in the sense that it does not satisfy the monotonicity property. In order to see this, we can look at the same situation in which monotonicity failed for von Neumann entropy: let ρ_1 be a mixed state and ρ a pure one such that $\text{Tr}_2 \rho = \rho_1$. Let V_ρ be a maximal context in which $\gamma_{V_\rho} = VN(\rho) = 0$. We know that V_ρ must be of the form $U \cdot V \otimes W \cdot U^{-1}$ for some maximal contexts V and W and unitary U . From Theorem 7.2.5 we have $\gamma_{1_V} \geq VN(\rho_1) > 0$. Hence it is not true that $\gamma \geq \gamma_1$.

Clearly if this happens, $U \cdot V \otimes W \cdot U^{-1}$ must be an entangled context, if it would be split we would be in the situation analyzed previously.

7.3 Reconstructing pure states from global sections

A direct implication of Remark 7.2.6 is that unlike von Neumann entropy, which gives the same value for unitarily equivalent states, our contextual entropy gives different (though in a sense unitarily equivalent) global sections of the presheaf $[0, \ln n]^{\preceq}$. This enables us not only to distinguish which global sections come from measures associated to pure states but also to explicitly reconstruct those pure states. We explain this method in more detail.

Recall that the von Neumann entropy of a state vanishes if and only if that state is a pure one. Given a global section $\gamma \in \Gamma[0, \ln n]^{\preceq}$ if γ is in the image of the contextual entropy mapping E then it comes from a measure associated to a pure state if and only if there exists a maximal context V such that $\gamma|_V = 0$. This means that if V is generated by the set of rank one projections

$$\{P_1, \dots, P_n\} = \{|\psi_1\rangle\langle\psi_1|, \dots, |\psi_n\rangle\langle\psi_n|\}$$

our state must equal one of these projections and our only task is to determine which one. For this, consider unitaries $U_1, \dots, U_n$ which have the property that $U_i P_i U_i^{-1} = P_i$ and

$$\{U_i P_j U_i^{-1} \mid 1 \leq j \leq n, j \neq i\} \neq \{P_1, \dots, \hat{P}_i, \dots, P_n\}$$

Think of this as taking n rotations in Hilbert space, each of which preserves one axis of the orthonormal basis $\{|\psi_1\rangle, \dots, |\psi_n\rangle\}$ and rotates the others, but without permuting them.

If we consider the contexts $V_i = \{U_i P_1 U_i^{-1}, \dots, U_i P_n U_i^{-1}\}''$ then ρ will be diagonal only in one of the orthonormal bases which correspond to these contexts. This means the contextual entropy will assign the value zero to precisely one of the contexts V_i , and hence our state is

$$\rho = \{U_i P_1 U_i^{-1}, \dots, U_i P_n U_i^{-1}\} \cap \{P_1, \dots, P_n\}$$

7.4 Reconstructing arbitrary quantum states from global sections

Consider a global section $\gamma \in \Gamma[0, \ln n]^{\preceq}$. We present here an algorithm for reconstructing the state ρ for which $E(\mu_\rho) = \gamma$. We assume for now that γ is in the image of the contextual entropy mapping. If our algorithm will fail to find a solution we will know that our initial assumption was false. Otherwise we must perform one final check at the end of our algorithm to make sure that this assumption was correct.

Step 1. Start by identifying one maximal context V such that $\gamma|_V \leq \gamma|_W$ for all maximal contexts W . This amounts to retrieving the von Neumann entropy of the state ρ from the contextual entropy. Note that in general, the minimal value of E will be attained in many different maximal contexts, but any of them can be used in our reconstruction.

If the minimal value among maximal contexts equals zero we must have a pure state, and we have already discussed these. Otherwise, we know from Section 7.2.3 that ρ must be diagonal in the context V . If we consider the canonical projections $\{P_1, \dots, P_n\}$ which generate V , the fact that ρ is diagonal at V implies that it is of the form

$$\rho = \lambda_1 P_1 + \dots + \lambda_n P_n$$

where the λ_i 's are the eigenvalues of ρ .

Of course, if we have access to the probability distribution on the spectral presheaf from which E_ρ was constructed, just identifying this context immediately yields the quantum state:

$$\rho = \sum_i \mu|_V(P_i) P_i$$

However, we are seeking to reconstruct the state mathematically using only information encoded by the contextual entropy map.

Step 2. We are now left with the task of determining the eigenvalues of ρ . For this assume that the dimension n of our Hilbert space is greater or equal to 3. For each $i \in \{1, \dots, n\}$ let

$$W_i := \{P_i, I - P_i\}''$$

Then $\text{Sh}(\lambda_i, 1 - \lambda_i)$ must equal $\gamma|_{W_i}$ for all i . If

$$\gamma|_{W_i} > \ln 2$$

then the global section γ cannot be in the image of the contextual entropy mapping, and our algorithm stops. Otherwise, the transcendental equation $\text{Sh}(x_1, x_2) = k$ has two solutions which are symmetric around $\frac{1}{2}$ as indicated in Figure 2.

Let p_i and $1 - p_i$ be the solutions of $\text{Sh}(x_1, x_2) = \gamma|_{W_i}$ and assume without loss of generality that $p_i \leq \frac{1}{2}$. For each i we have at most two choices for the value of the i^{th} eigenvalue of ρ : we

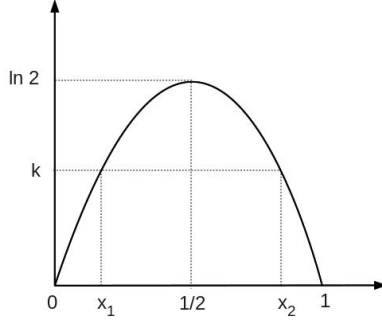


Figure 2: The Shannon entropy of a two-variable probability distribution

can either set $\lambda_i = p_i$ or $\lambda_i = 1 - p_i$. Since

$$\lambda_1 + \dots + \lambda_n = 1$$

there can be at most one j such that $p_j < \frac{1}{2}$ and $\lambda_j = 1 - p_j$, while for all $i \neq j$ we must have $\lambda_i = p_i$. Let

$$S = \sum_{i=1}^n p_i$$

Clearly $\sum_{i=1}^n \lambda_i \geq S$. We are now faced with three possible scenarios:

(a) If $S > 1$ we obtain a contradiction, hence γ can not be in the image of the contextual entropy mapping.

(b) If $S = 1$ then the assignment $\lambda_i = p_i$ gives one possible solution for the set of eigenvalues of our state ρ . This solution is clearly unique: any other choice of values will make the total sum of the eigenvalues of ρ greater than 1.

(c) If $S < 1$ then we must determine the j for which $p_j < \frac{1}{2}$ and $\lambda_j = 1 - p_j$. If such a j exists then

$$1 = \sum_{i=1}^n \lambda_i = S - p_j + (1 - p_j)$$

hence p_j should equal $\frac{S}{2}$. Now

(c1) if the value $\frac{S}{2}$ does not appear amongst $\{p_1, \dots, p_n\}$ then we have no solution

(c2) if $\frac{S}{2}$ appears once, we have a unique solution

(c3) if it appears more than once, let $\{j_1, \dots, j_m\}$ be the set of indices for which $p_{j_k} = \frac{S}{2}$.

If we set $\lambda_{j_k} = 1 - p_{j_k}$ and take another $l \in \{1, \dots, m\}$, $l \neq k$. Then

$$\sum_{i=1}^n \lambda_i \geq \lambda_{j_k} + \lambda_{j_l} = 1 - p_{j_k} + p_{j_k} = 1$$

In order to have equality we must have $m = 2$ and $p_i = 0$ for all $i \notin \{j_1, j_2\}$. Unless this happens we cannot find a solution. On the other hand, for $m = 2$ we have two possible solutions. These correspond to the two states

$$\rho_1 = p_{j_1} P_{j_1} + (1 - p_{j_1}) P_{j_2}$$

and

$$\rho_2 = (1 - p_{j_1}) P_{j_1} + p_{j_1} P_{j_2}$$

In order to decide between the two states in **(c3)** we need to run our algorithm again but with a slight modification: instead of considering two-dimensional subalgebras of V , we take a unitary U which rotates all the canonical projections generating V , except P_{j_1} , which it leaves unchanged, and we consider the two dimensional subalgebras of $U \cdot V \cdot U^{-1}$ of the form

$$\widetilde{W}_i = \{UP_iU^{-1}, I - UP_iU^{-1}\}''$$

We solve the equations $\text{Sh}(x_i, 1-x_i) = \gamma|_{\widetilde{W}_i}$ and choose as before n numbers from these solutions, such that they add up to one. These numbers represent the diagonal entries of the matrix $U^{-1}\rho U$. We will not encounter any problems when retrieving these entries (unless of course, our initial assumption about γ being in the image of the contextual entropy mapping was false) because unlike the eigenvalues of ρ , these diagonal entries must contain more than three non-zero elements, so the **(c3)** branch is not accessible this time. Moreover, the j_1^{th} entry on the diagonal of $U^{-1}\rho U$ will be the same as the j_1^{th} eigenvalue of ρ , which allows us to choose between ρ_1 and ρ_2 .

Step 3. We have now reached the end of our algorithm. If it has failed to retrieve a solution, we conclude that we have considered a global section γ which was not in the image of the contextual entropy mapping. Otherwise, our reconstructed state is

$$\rho = \lambda_1 P_1 + \dots + \lambda_n P_n$$

In order to obtain ρ we have taken into account only a finite number of contexts, and it might happen that when all contexts are taken into account $E(\mu_\rho) \neq \gamma$. In this case we also conclude that γ was not in the image of the contextual entropy mapping, and discard the state ρ .

7.4.1 Two-dimensional Hilbert spaces

For two dimensional Hilbert spaces the contextual entropy is a two-to-one mapping. We will justify this statement below.

First, it is easy to check that for any one dimensional projection P the states $\rho_1 = \lambda P + (1 - \lambda)(I - P)$ and $\rho_2 = (1 - \lambda)P + \lambda(I - P)$ are mapped to the same global section of $[0, \ln 2]^{\preceq}$: note that $\rho_1 = I - \rho_2$. Hence for every context $W = \{Q, I - Q\}''$

$$E(\mu_{\rho_1})|_W = \text{Sh}(\text{Tr}\rho_1 Q, 1 - \text{Tr}\rho_1 Q)$$

while

$$\begin{aligned} E(\mu_{I-\rho_1})|_W &= \text{Sh}(\text{Tr}(I - \rho_1)Q, 1 - \text{Tr}(I - \rho_1)Q) \\ &= \text{Sh}(\text{Tr}(I - \rho_1)(I - Q), 1 - \text{Tr}(I - \rho_1)(I - Q)) \end{aligned}$$

And since every one dimensional projection Q has trace equal to unity,

$$\text{Tr}(I - \rho_1)(I - Q) = \text{Tr}I - \rho_1 - Q + \rho_1 Q = \text{Tr}\rho_1 Q$$

and so also $E(\mu_{\rho_1})|_W = E(\mu_{I-\rho_1})|_W$.

On the other hand, given a global section of $[0, \ln 2]^{\preceq}$, the poset $\mathcal{V}(M_2)$ consists only of two-dimensional subalgebras. We can identify a context $V = \{P, 1 - P\}''$ for which $\gamma|_V$ is minimal, and solve the equation $\text{Sh}(x, 1 - x) = \gamma|_V$ to find the eigenvalues of ρ . Since we have no further information available, we cannot say which eigenvalue corresponds to which of the two projections generating V .

Note however that we are not far from reconstructing ρ : we would need to encode only one extra bit of information in order to fully reconstruct a two-dimensional quantum state.

7.5 Renyi entropy

We have seen how Shannon entropy can be encoded in the topos approach, and how one can afterwards retrieve its quantum analogue, the von Neumann entropy. It is natural to ask at this point whether a similar encoding can be found for other classical entropies, and whether such an encoding would still enable us to retrieve their quantum analogues. We will look here at Renyi entropies, and show that it is possible to obtain their topos theoretic equivalent.

We first define contextual Renyi entropy locally as

$$\mathcal{R}_q(\mu)_V = R_q(\mu|_V(P_1), \dots, \mu|_V(P_n), \forall V = \{P_1, \dots, P_n\}''$$

Of course, we would like these local components to fit together nicely as before, and to form a global section of some presheaf of real numbers. For Shannon entropy, the fact that a global section could be formed was a consequence of the recursion property. Renyi entropies are in general not recursive, but they do satisfy a property which we shall call *weak recursivity*, and we shall see that this is enough for our purposes.

Definition 7.5.1. *Let S be some function defined on the set of all probability distributions. If we coarse grain a probability distribution $(x_1, \dots, x_n)$ by not distinguishing between all the outcomes, we obtain a new probability distribution with components*

$$p_1 = \sum_{i=1}^{k_1} x_i, \dots, p_r = \sum_{i=k_{r-1}+1}^{k_r} x_i$$

for some $0 < k_1 < k_2 < \dots < k_r = n$. We say that S is **weakly recursive** if

$$S(x_1, \dots, x_n) \geq S(p_1, \dots, p_r)$$

One can easily check that Renyi entropies indeed satisfy this property, and hence for any two contexts $V' \supseteq V$

$$\mathcal{R}_q(\mu)_{V'} \geq \mathcal{R}_q(\mu)_V, \forall \mu \in \mathcal{M}(\underline{\Sigma})$$

This means it is possible to define contextual Renyi entropy as a mapping

$$\mathcal{R}_q : \mathcal{M}(\underline{\Sigma}) \longrightarrow \Gamma[\underline{0}, \underline{\ln n}]^{\preceq}$$

Since Renyi entropies are Schur concave, their quantum counterparts can be retrieved from the contextual Renyi entropies by finding the minimum over the set of values assigned to all maximal contexts. This is justified by the Schur-Horn lemma and similar arguments to those that were already used in Section 7.2.3.

7.5.1 Properties

We will now briefly discuss some of the properties of Renyi entropies and their contextual analogues.

Concavity

We saw in Section 7.2.3 that the global concavity of the contextual entropy was expressed as the concavity of each of its local components, and hence it was a direct consequence of the concavity property of Shannon entropy. Renyi entropies however are only concave for $0 < q \leq 1$. In fact, it is known that concavity is lost for $q > q_* > 1$, where q_* depends on the dimension of the probability distribution. Concavity of the contextual Renyi entropies is then going to hold under the same conditions.

Additivity and Subadditivity

Renyi entropies are additive, so we can use the same justification as in Section 7.2.3 to define subadditivity for contextual Renyi entropies as the following condition:

$$\mathcal{R}_q(\mu_\rho)_V \leq \mathcal{R}_q(\mu_{\rho_1 \otimes \rho_2})_V, \quad \forall V \in \mathcal{V}(\mathcal{B}(H))$$

This allows us to say that contextual Renyi entropies are also additive. On the other hand, since neither classical nor quantum Renyi entropies are subadditive (except for $q = 0$ and $q = 1$), contextual Renyi entropy also doesn't have this property.

Monotonicity

Classical Renyi entropies are monotone, just like Shannon entropy. However, their quantum counterparts do not necessarily have this property since, just like von Neumann entropy, they assign the value 0 to pure states and it is possible to have a composite system in a pure state and both its components in mixed states. For contextual Renyi entropies, this means that we will not have a global property of monotonicity. Instead this property is going to hold only for split contexts.

State reconstruction

Finally, recall that the reconstruction algorithms described in Sections 7.3-7.4.1 relied on Gleason's theorem, the Schur-Horn lemma, and two extra ingredients: one was the fact that von Neumann entropy vanished only for pure states, and the second was the fact that for probability distributions with two variables one could find precisely two solutions (symmetric around $1/2$) for which Shannon entropy would take any given value within its image. Both of these ingredients are present when we consider Renyi entropy, for positive parameter $q \neq 0$, as Figure 1 clearly indicates. This means that the reconstruction algorithms can also be applied to contextual Renyi entropies, with the exception of $\mathcal{R}_0$.

Chapter 8

Conclusions and Outlook

In this thesis we have investigated two distinct but related research strands involving sheaf and category theory, within the fields of quantum information theory and foundations of physics. In the following paragraphs we will outline the key results of this work and discuss its potentially interesting extensions to future research.

In the first half of the thesis we have surveyed the landscape of multipartite entangled states through the lens of the contextuality classification developed by Abramsky and Brandenburger in [2].

In Chapter 3 we have made a complete classification of a particular class of non-symmetric multipartite entangled states, the balanced states with functional dependencies. We have showed that each of the three strengths of contextuality distinguished by Abramsky and Brandenburger is exhibited by a certain subclass of these functionally dependent multipartite quantum states. This classification has recently been shown to relate to the study of violations of local realism in quantum hypergraph states conducted by Gühne et al. in [27].

One potential future use of the results of this investigation could fructify the fact that the study carried out by Abramsky and Brandenburger has been conducted at a very high level of generality, without any presupposition of quantum mechanics. Although their methods are readily applicable in quantum mechanical scenarios, there is also the possibility of making further connections between the study of contextuality and non-locality in physics and ideas arising in other fields. Since the functional dependency in the states considered in Chapter 3 is given by Boolean functions, it would be interesting to see if our classification could yield any insights within other areas of research where Boolean functions play a prominent role.

We have also noted in Sections 3.2.2 and 3.3.3 that the states belonging to the subclass with the weakest degree of contextuality, the dictatorship states, are closely related to maximally

entangled two-qubit states, being equivalent to tensor products of pure states and maximally entangled two-qubit states. We have shown that such states can not exhibit a stronger degree of contextuality if the parties which share such a state each have a choice between the same two local observables.

This observation naturally leads to the question of quantifying the maximum degree of contextuality which can be exhibited by a quantum state for some choice of local observables. Given the current scarcity of results related to the classification and quantification of multipartite entanglement, the perspective offered by the question of the strongest degree of contextuality which a state may achieve could become very useful, particularly in the light of recent results [88, 111] showing that contextuality is a key feature enabling quantum computation.

A first step in this direction has already been taken in Chapter 4, where we demonstrate that all n -qubit quantum states, with the exception of tensor products of maximally entangled bipartite states and single quantum states, are at least logically contextual. That is to say, they occupy at least the middle level of the three-level hierarchy of contextuality. This implies that the probability- and inequality-free logical formulation of contextuality and non-locality specific to logically, and in particular also to strongly contextual quantum states, is not a rare occurrence, but in fact arises for almost all states.

If we further consider the notion of strong contextuality, a natural next challenge follows from the results presented in the first half of this thesis. Namely, to characterise those quantum states for which local observables giving rise to a strongly contextual empirical model can be found.

This question remains open, and appears difficult. It has an interesting relation to the question of “All-versus-Nothing” arguments first used by Mermin [103], which have recently been studied in the sheaf-theoretic approach by Abramsky et al. [1], and shown to be related to the cohomological witnesses for contextuality previously introduced by Abramsky, Mansfield and Soares Barbosa in [8]. All currently known examples of strong contextuality arising in quantum mechanics come from All-versus-Nothing arguments. Determining whether this is true in general is another challenging problem, which may hold the key to the main question.

The qualitative arguments of Chapters 3 and 4 can easily be turned into quantitative ones. Abramsky and Hardy have shown in [7] how any instance of logical contextuality gives rise to a Bell inequality based on logical consistency conditions, which allows for quantitative, robust experimental tests. While the structure of the argument in Chapter 4 can be easily modified to yield a lower bound for the violations of the logical Bell inequalities corresponding to each logi-

cally non-local quantum state, it would also be interesting to seek to obtain concrete violations of Bell inequalities for each particular type among the functionally dependent entangled states.

In the second half of our thesis, we have explored two strands of research within the Topos Approach to the formulation of physical theories.

First, in Chapter 6, we have showed that it is possible to give a fairly direct reconstruction of an atomic orthomodular lattice from its associated poset of distributive sub-lattices. Harding and Navara had shown in [80] that an isomorphism of the posets of distributive sub-lattices of two orthomodular lattices implies the existence of an isomorphism between the respective orthomodular lattices themselves. This result has an important implication, as far as the Topos Approach is concerned, since it establishes that the orthomodular lattice $\mathcal{P}(\mathcal{H})$, which is traditionally used in quantum logic, can be reconstructed from the poset of contexts $\mathcal{B}(\mathcal{P}(\mathcal{H}))$ that underlies the constructions in the Topos Approach. In other words, this means that the new form of presheaf- and topos-based form of logic for quantum systems is (at least) as rich as traditional quantum logic.

Our reconstruction result has provided a partial answer to one of the open questions posed by Harding and Navara towards the end of their paper. In future research it would be interesting to investigate whether the reconstruction can be extended to the case of non-atomic orthomodular lattices in order to provide a complete answer to Harding and Navara’s question.

Finally, in Chapter 7, we have considered the information theoretic concept of entropy from the perspective of the Topos Approach. This new perspective allowed us to treat classical and quantum notions of entropy, such as Shannon and von Neumann entropies, or Renyi entropies, in a unified setting, via the innovative construction of contextual entropy. Within this construction a classical Shannon entropy is associated to each commuting subalgebra of observables of the non-commutative algebra corresponding to a given quantum system. This assignment of Shannon entropies is based on the reformulation of quantum states within the Topos Approach described by Döring in [44]. Given a quantum state, the classical Shannon entropies which build up the contextual entropy map are the entropies of the quantum state seen through the ‘classical windows’ given by the commuting subalgebras. We have further analysed how the state’s von Neumann entropy, which is the quantum counterpart of Shannon entropy, is associated to a distinguished maximal commuting subalgebra. In Section 7.5 we have showed that the same principles can be applied to construct the Renyi contextual entropy, which similarly treats classical and quantum Renyi entropies in a unified setting.

Perhaps the most striking feature of the contextual entropy map is the fact that it is rich

enough to allow for a reconstruction algorithm which takes in a contextual entropy map and outputs a quantum state. This implies that contextual entropy can be seen as a mathematical equivalent of quantum states based on the information theoretical concept of entropy. As such it is a step towards an information-theoretic characterisation of quantum states.

In future work we propose checking whether the concept of contextual entropy could also be defined for infinite dimensional quantum systems. Since both Gleason's theorem and the Schur-Horn lemma, which the reconstruction algorithm essentially relies on, can be used in infinite dimensions (as proved by Kaftal and Weiss in [94]), it would be particularly interesting to see if our reconstruction results can also be applied in the infinite dimensional case.

Finding an axiomatic characterisation of the contextual entropy map would allow us to relate our reconstruction results to a generalized version of Gleason's theorem. This theorem states that every finitely additive probability measure m on the projections of a von Neumann algebra with no type I_2 summand can be uniquely extended to a state on that algebra.

Since every m as above uniquely determines a measure μ on the spectral presheaf, we could easily construct its associated contextual entropy map E_μ . Using our reconstruction algorithm, we could in principle retrieve the unique quantum state ρ associated to E_μ , and hence also the probability measure μ . However, the caveat is that in our algorithm we had to assume in the first place that we had started from a probability measure μ on projections. Therefore having an *axiomatic* characterisation of those real-valued maps on contexts which are contextual entropy maps (and hence come from quantum states) would allow us to reconstruct quantum states directly. This would be an important step towards giving a new structural proof of Gleason's theorem, and as such it promises to be both an interesting and non-trivial task.

Last, but not least, perhaps the most advanced and important application of this topos-theoretic construction during this exciting era of Quantum Information Theory would be to obtain new quantitative and qualitative insights into the nature of multipartite entanglement starting from the rich structure of the contextual entropy map.

Appendices

Appendix A

The Subadditivity Property of Contextual Entropy

In this Appendix we seek to find general examples of density matrices and entangled contexts, for which the subadditivity property of contextual entropy fails to hold.

Let us assume we are looking at a composite system of dimension $n = n_1 n_2$, with $n_1, n_2 \geq 2$. Consider the following diagonal density matrix:

$$\rho = \begin{pmatrix} \frac{1}{2} & & \\ & \mathbf{0}_{n-2} & \\ & & \frac{1}{2} \end{pmatrix}$$

Its partial traces will be of the same form, and their tensor product will also be a diagonal matrix:

$$\rho_1 \otimes \rho_2 = \begin{pmatrix} \frac{1}{4} & & & & \\ & \mathbf{0}_{n_2-2} & & & \\ & & \frac{1}{4} & & \\ & & & \mathbf{0}_{(n_1-2)n_2} & \\ & & & & \frac{1}{4} \\ & & & & & \mathbf{0}_{n_2-2} \\ & & & & & & \frac{1}{4} \end{pmatrix}$$

If $U = (U_{i,j})$ is a n -dimensional unitary matrix, then we are only interested in the diagonal entries of $U\rho U^{-1}$ and of $U\rho_1 \otimes \rho_2 U^{-1}$ and these can be easily calculated:

$$U\rho U^{-1} = \frac{1}{2} \begin{pmatrix} |U_{1,1}|^2 + |U_{1,n}|^2 & & & \\ & \ddots & & \\ & & |U_{n-1,1}|^2 + |U_{n-1,n}|^2 & \\ & & & |U_{n,1}|^2 + |U_{n,n}|^2 \end{pmatrix}$$

while

$$U\rho_1 \otimes \rho_2 U^{-1} = \frac{1}{4} \begin{pmatrix} |U_{1,1}|^2 + |U_{1,n}|^2 + |U_{1,n_2}|^2 + |U_{1,n-n_2}|^2 & & & \\ & \ddots & & \\ & & |U_{n-1,1}|^2 + |U_{n-1,n}|^2 + |U_{n-1,n_2}|^2 + |U_{n-1,n-n_2}|^2 & \\ & & & |U_{n,1}|^2 + |U_{n,n}|^2 + |U_{n,n_2}|^2 + |U_{n,n-n_2}|^2 \end{pmatrix}$$

We want to calculate

$$\begin{aligned} A &= E(\mu_\rho)|_{U^{-1} \cdot D_{n_1} \otimes D_{n_2} \cdot U} = E(\mu_\rho)|_{U^{-1} \cdot D_n \cdot U} \\ &= E(\mu_{U\rho U^{-1}})|_{D_n} = \text{Sh}((U\rho U^{-1})_{11}, \dots, (U\rho U^{-1})_{nn}) \end{aligned}$$

and show that it is strictly greater than

$$B = E(\mu_{\rho_1 \otimes \rho_2})|_{U^{-1} \cdot D_{n_1} \otimes D_{n_2} \cdot U} = \text{Sh}((U\rho_1 \otimes \rho_2 U^{-1})_{11}, \dots, (U\rho_1 \otimes \rho_2 U^{-1})_{nn})$$

In order to compute these we need to specify only four columns of the unitary matrix U , namely columns 1, n_2 , $n - n_2$ and n . That is, we must specify four orthonormal vectors of dimension n . We first look at the situation when n is even and greater or equal to six. In this case the four vectors can be taken as follows:

$$\begin{aligned}
U_1^T &= \frac{1}{\sqrt{n}}(1, 1, \dots, 1, 1, 1, \dots, 1) \\
U_n^T &= \frac{1}{\sqrt{n}} \left(\underbrace{1, 1, \dots, 1}_{n/2}, \underbrace{-1, -1, \dots, -1}_{n/2} \right) \\
U_{n_2}^T &= \frac{1}{2} \left(-1, 1, \underbrace{0, \dots, 0}_{n-4}, -1, 1 \right) \\
U_{n-n_2}^T &= \frac{1}{2} \left(1, -1, \underbrace{0, \dots, 0}_{n-4}, -1, 1 \right)
\end{aligned}$$

and so

$$A = \text{Sh}(1/n, \dots, 1/n) = \ln n$$

which is strictly greater than

$$B = \text{Sh} \left(\frac{1}{2n} + \frac{1}{8}, \frac{1}{2n} + \frac{1}{8}, \underbrace{\frac{1}{2n}, \dots, \frac{1}{2n}}_{n-4}, \frac{1}{2n} + \frac{1}{8}, \frac{1}{2n} + \frac{1}{8} \right)$$

as desired. For $n = 4$ we can see that $\frac{1}{2n} + \frac{1}{8} = 1/n$, and this approach will fail to produce a counterexample. However, by randomly generating four-dimensional density matrices and unitaries on a computer, it is easy to come across an example for which subadditivity fails. For instance if we consider the state

$$\rho = \begin{pmatrix}
0.089 & -0.107 - 0.038i & 0.070 - 0.009i & 0.116 - 0.056i \\
-0.107 + 0.038i & 0.328 & -0.150 - 0.048i & -0.226 + 0.053i \\
0.070 + 0.009i & -0.150 + 0.048i & 0.205 & 0.117 - 0.073i \\
0.116 + 0.056i & -0.226 - 0.053i & 0.117 + 0.073i & 0.376
\end{pmatrix}$$

and the unitary matrix

$$U = \begin{pmatrix}
0.662 + 0.163i & 0.027 + 0.130i & 0.411 - 0.532i & 0.234 - 0.098i \\
-0.250 - 0.232i & 0.459 - 0.526i & 0.290 - 0.229i & 0.261 + 0.435i \\
-0.011 + 0.106i & 0.490 - 0.268i & 0.147 + 0.242i & 0.017 - 0.772i \\
0.601 - 0.213i & -0.161 - 0.394i & -0.117 + 0.562i & 0.238 + 0.155i
\end{pmatrix}$$

it is only a matter of straightforward computations to check that

$$E(\mu_\rho)|_{U^{-1} \cdot D_4 \cdot U} = E(\mu_{U\rho U^{-1}})|_{D_4} > E(\mu_{U\rho_1 \otimes \rho_2 U^{-1}})|_{D_4} = E(\mu_{\rho_1 \otimes \rho_2})|_{U^{-1} \cdot D_4 \cdot U}$$

and hence the subadditivity property does not hold at the entangled context $U^{-1} \cdot D_4 \cdot U$.

The last case to consider is the one when the dimension of our system is odd. Then we must have both n_1 and n_2 odd and greater or equal to three, so $n \geq 9$. In this case the four vectors can be taken as follows:

$$\begin{aligned} U_1^T &= \frac{1}{\sqrt{n}} (1, 1, \dots, 1, 1, 1, \dots, 1, 3, 0, 0) \\ U_n^T &= \frac{1}{\sqrt{n}} \left(\underbrace{1, 1, \dots, 1}_{(n-3)/2}, \underbrace{-1, -1, \dots, -1}_{(n-3)/2}, 0, 3, 0 \right) \\ U_{n_2}^T &= \frac{1}{2} \left(-1, 1, \underbrace{0, \dots, 0}_{n-7}, -1, 1, 0, 0, 0 \right) \\ U_{n-n_2}^T &= \frac{1}{2} \left(1, -1, \underbrace{0, \dots, 0}_{n-7}, -1, 1, 0, 0, 0 \right) \end{aligned}$$

in which case

$$A = \text{Sh} \left(1/n, \dots, 1/n, \frac{3}{2n}, \frac{3}{2n}, 0 \right)$$

which is less than $\ln n$, but it is still greater than

$$B = \text{Sh} \left(\frac{1}{2n} + \frac{1}{8}, \frac{1}{2n} + \frac{1}{8}, \underbrace{\frac{1}{2n}, \dots, \frac{1}{2n}}_{n-7}, \frac{1}{2n} + \frac{1}{8}, \frac{1}{2n} + \frac{1}{8}, \frac{3}{4n}, \frac{3}{4n}, 0 \right)$$

Bibliography

- [1] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal, and Shane Mansfield, *Contextuality, Cohomology and Paradox*, (2015), [arXiv:1502.03097](#).
- [2] Samson Abramsky and Adam Brandenburger, *The sheaf-theoretic structure of non-locality and contextuality*, New Journal of Physics **13** (2011), [arXiv:1102.0264](#).
- [3] Samson Abramsky and Adam Brandenburger, *An operational interpretation of negative probabilities and no-signalling models*, (2014), [arXiv:1401.2561](#).
- [4] Samson Abramsky and Bob Coecke, *Categorical quantum mechanics*, Handbook of Quantum Logic and Quantum Structures: Quantum Logic (Kurt Engesser, Dov M Gabbay, and Daniel Lehmann, eds.), Elsevier, 2008, pp. 261–324.
- [5] Samson Abramsky and Carmen M. Constantin, *A classification of multipartite states by degree of non-locality*, Electronic Proceedings in Theoretical Computer Science (2014), [arXiv:1412.5213](#).
- [6] Samson Abramsky, Carmen M. Constantin, and Shenggang Ying, *Hardy is (almost) everywhere: nonlocality without inequalities for almost all entangled multipartite states*, Electronic Proceedings in Theoretical Computer Science (2015), [arXiv:1506.01365](#).
- [7] Samson Abramsky and Lucien Hardy, *Logical Bell inequalities*, Physical Review A **85** (2012), [arXiv:1203.1352](#).
- [8] Samson Abramsky, Shane Mansfield, and Rui Soares Barbosa, *The cohomology of non-locality and contextuality*, EPTCS **95** (2012), 1–14, [arXiv:1111.3620](#).
- [9] Traian E. Abrudan, Jan Eriksson, and Visa Koivunen, *Steepest descent algorithm for optimization under unitary matrix constraint*, IEEE Transactions on Signal Processing **56** (2008), no. 3, 1134.

- [10] Traian E. Abrudan, Jan Eriksson, and Visa Koivunen, *Conjugate gradient algorithm for optimization under unitary matrix constraint*, Signal Processing **89** (2009), no. 3, 1704.
- [11] Antonio Acin, Nicolas Gisin, and Lluís Masanes, *From Bell's theorem to secure quantum key distribution*, Physical Review Letters **97** (2006), 120405, [arXiv:quant-ph/0510094](#).
- [12] Charles A. Akemann, *Left ideal structure of C^* -algebras*, Journal of Functional Analysis **6** (1970), 305.
- [13] Howard Barnum, Jonathan Barrett, Lisa O. Clark, Matthew Leifer, Robert Spekkens, Nicholas Stepanik, Alex Wilce, and Robin Wilke, *Entropy and information causality in general probabilistic theories*, New Journal of Physics **12** (2010), 033024, [arXiv:0909.5075](#).
- [14] Jonathan Barrett, Lucien Hardy, and Adrian Kent, *No signaling and quantum key distribution*, Physical Review Letters **95** (2005), 10503, [arXiv:quant-ph/0405101](#).
- [15] John S. Bell, *On the Einstein-Podolsky-Rosen paradox*, Physics **1** (1964), 195.
- [16] Ingemar Bengtsson and Karol Zyczkowski, *Geometry of quantum states: An introduction to quantum entanglement*, Cambridge University Press, 2006.
- [17] Charles H. Bennett and Gilles Brassard, *Quantum cryptography: Public key distribution and coin tossing*, Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing (1984), 175–179, [arXiv:1102.0264](#).
- [18] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters, *Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels*, Physical Review Letters **70** (1993), 1895–1899.
- [19] Charles H. Bennett and Stephen J. Wiesner, *Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states*, Physical Review Letters **69** (1992), 2881–2884.
- [20] Garrett Birkhoff and John von Neumann, *The logic of quantum mechanics*, Annals of Mathematics **37** (1936), no. 4, 823–843.
- [21] Lenore Blum, Mike Shub, and Steve Smale, *On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines*, Bulletin (New Series) of the American Mathematical Society **21** (1989), no. 1, 1–46.

- [22] Francis Borceux and Gilberte van den Bossche, *An essay on noncommutative topology*, Topology and its Applications **31** (1989), 202.
- [23] Gilles Brassard, *Is information the key?*, Nature Physics **1** (2005), no. 2.
- [24] Caslav Brukner, Marek Zukowski, Jian-Wei Pan, and Anton Zeilinger, *Bell's inequalities and quantum communication complexity*, Physical Review Letters **92** (2004), 127901.
- [25] Jeffrey Bub, *Quantum mechanics is about quantum information*, (2004), [arXiv:quant-ph/0408020v2](#).
- [26] Jeffrey Bub, Rob Clifton, and Hans Halvorson, *Characterizing quantum theory in terms of information-theoretic constraints*, Foundations of Physics **33** (2003), 1561, [arXiv:quant-ph/0211089](#).
- [27] Constantino Budroni, Mariami Gachechiladze, and Otfried Gühne, *Extreme violation of local realism in quantum hypergraph states*, (2015), [arXiv:1507.03570](#).
- [28] Jeremy Butterfield and Chris J. Isham, *A topos perspective on the Kochen-Specker theorem: I. Quantum states as generalized valuations.*, International Journal of Theoretical Physics **37** (1998), 2669, [arXiv:quant-ph/9803055](#).
- [29] Jeremy Butterfield and Chris J. Isham, *A topos perspective on the Kochen-Specker theorem: II. Conceptual aspects, and classical analogues.*, International Journal of Theoretical Physics **38** (1999), 827, [arXiv:quant-ph/9808067](#).
- [30] Jeremy Butterfield and Chris J. Isham, *A topos perspective on the Kochen-Specker theorem: III. Von Neumann algebras as the base category.*, International Journal of Theoretical Physics **39** (2000), 1413, [arXiv:quant-ph/9911020](#).
- [31] Jeremy Butterfield and Chris J. Isham, *A topos perspective on the Kochen-Specker theorem: IV. Interval valuations.*, International Journal of Theoretical Physics **41** (2002), 613, [arXiv:quant-ph/0107123](#).
- [32] Adan Cabello, Simone Severini, and Andreas Winter, *(Non-)contextuality of physical theories as an axiom*, (2010), [arXiv:1010.2163](#).
- [33] Maria L. Dalla Chiara and Roberto Giuntini, *Quantum logics*, Handbook of Philosophical Logic (Dov M. Gabbay and Franz Guenther, eds.), vol. 6, Kluwer Academic Publishers, 2002, p. 129.

- [34] John F. Clauser and Michael A. Horne, *Experimental consequences of objective local theories*, Physical Review **D10** (1974), 526.
- [35] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt, *Proposed experiment to test local hidden variable theories*, Physical Review Letters **23** (1969), 880–884.
- [36] Carmen M. Constantin and Andreas Döring, *Reconstructing an atomic orthomodular lattice from the poset of its Boolean sublattices*, Houston Journal of Mathematics (2014), [arXiv:1306.1950](#).
- [37] Thomas M. Cover and Joy A. Thomas, *Elements of information theory*, Wiley Series in Telecommunications and Signal Processing, 1991.
- [38] Yves Crama and Peter L. Hammer, *Boolean functions: Theory, algorithms, and applications*, Encyclopedia of Mathematics and its Applications, vol. 142, Cambridge University Press, 2011.
- [39] Hans F. de Groote, *Observables I: Stone spectra*, (2005), [arXiv:math-ph/0509020](#).
- [40] Nadish de Silva, *From topology to noncommutative geometry: K-theory*, (2014), [arXiv:1408.1170](#).
- [41] Inderjit S. Dhillon, Robert W. Heath, Matyas A. Sustik, and Joel A. Tropp, *Generalized finite algorithms for constructing Hermitian matrices with prescribed diagonal and spectrum*, SIAM Journal on Matrix Analysis and Applications (2003).
- [42] Robert H. Dicke, *Coherence in spontaneous radiation processes*, Physical Review **93** (1954), no. 1, 99.
- [43] Jacques Dixmier, *Les C^* -algèbres et leurs représentations*, Gauthier-Villars, 1969.
- [44] Andreas Döring, *Quantum states and measures on the spectral presheaf*, Advanced Scientific Letters **2** (2009), no. 2, 291, [arXiv:0809.4847](#), Special Issue on ‘Quantum Gravity, Cosmology and Black Holes’, ed. Martin Bojowald.
- [45] Andreas Döring, *Topos theory and ‘neo-realist’ quantum theory*, Quantum Field Theory, Competitive Models (Bertfried Fauser, Jürgen Tolksdorf, and Eberhard Zeidler, eds.), Birkhäuser, 2009.
- [46] Andreas Döring, *The physical interpretation of daseinisation*, Deep Beauty (Hans Halvorson, ed.), Cambridge University Press, 2011, p. 207.

- [47] Andreas Döring, *Topos quantum logic and mixed states*, Electronic Notes in Theoretical Computer Science **270** (2011), no. 2, [arXiv:1004.3561](#), in Proceedings of the 6th International Workshop on Quantum Physics and Logic (QPL 2009).
- [48] Andreas Döring, *Flows on generalised Gelfand spectra of nonabelian unital C^* -algebras and time evolution of quantum systems*, (2012), [arXiv:1212.4882](#).
- [49] Andreas Döring, *Generalised Gelfand spectra of nonabelian unital C^* -algebras*, (2012), [arXiv:1212.2613](#).
- [50] Andreas Döring, *Some remarks on the logic of quantum gravity*, (2013), [arXiv:1306.3076](#).
- [51] Andreas Döring, *Two new complete invariants of von Neumann algebras*, (2014), [arXiv:1411.5558](#).
- [52] Andreas Döring, *Topos-based logic for quantum systems and bi-Heyting algebras*, Logic & Algebra in Quantum Computing, Lecture Notes in Logic, Association for Symbolic Logic and Cambridge University Press, 2015.
- [53] Andreas Döring and Rui Soares Barbosa, *Unsharp values, domains and topoi*, Quantum Field Theory and Gravity, Conceptual and Mathematical Advances in the Search for a Unified Framework (Felix Finster, Olaf Müller, Marc Nardmann, Jürgen Tolksdorf, and Eberhard Zeidler, eds.), Birkhäuser, 2012, p. 65.
- [54] Andreas Döring and Carmen M. Constantin, *Contextual entropy and reconstruction of quantum states*, (2012), [arXiv:1208.2046](#).
- [55] Andreas Döring and Barry Dewitt, *Self-adjoint operators as functions II: Quantum probability*, (2012), [arXiv:1210.5747v2](#).
- [56] Andreas Döring and Barry Dewitt, *Self-adjoint operators as functions I: Lattices, Galois connections, and the spectral order*, Communications in Mathematical Physics **328** (2014), no. 2, 499–525, [arXiv:1208.4724](#).
- [57] Andreas Döring and John Harding, *Abelian subalgebras and the Jordan structure of a von Neumann algebra*, Houston Journal of Mathematics (2013), to appear, [arXiv:1009.4945](#).
- [58] Andreas Döring and Chris J. Isham, *A topos foundation for theories of physics: I. Formal languages for physics*, Journal of Mathematical Physics **49** (2008), no. 053515, [arXiv:quant-ph/0703060](#).

- [59] Andreas Döring and Chris J. Isham, *A topos foundation for theories of physics: II. Daseinisation and the liberation of quantum theory*, Journal of Mathematical Physics **49** (2008), no. 053516, [arXiv:quant-ph/0703062](#).
- [60] Andreas Döring and Chris J. Isham, *A topos foundation for theories of physics: III. The representation of physical quantities with arrows*, Journal of Mathematical Physics **49** (2008), no. 053517, [arXiv:quant-ph/0703064](#).
- [61] Andreas Döring and Chris J. Isham, *A topos foundation for theories of physics: IV. Categories of systems*, Journal of Mathematical Physics **49** (2008), no. 053518, [arXiv:quant-ph/0703066](#).
- [62] Andreas Döring and Chris J. Isham, ‘*What is a thing?*’: *Topos theory in the foundations of physics*, New Structures in Physics (Bob Coecke, ed.), Springer, 2008, p. 753.
- [63] Andreas Döring and Chris J. Isham, *Classical and quantum probabilities as truth values*, Journal of Mathematical Physics **53** (2012), 032101, [arXiv:1102.2213](#).
- [64] Wolfgang Dur, Guifre Vidal, and Ignacio J. Cirac, *Three qubits can be entangled in two inequivalent ways*, Physical Review **A62** (2000), 062314, [arXiv:quant-ph/0005115](#).
- [65] Albert Einstein, *Quanten-Mechanik Und Wirklichkeit*, Dialectica **2** (1948), 320–324.
- [66] Albert Einstein, Boris Podolsky, and Nathan Rosen, *Can quantum mechanical description of physical reality be considered complete?*, Physical Review **47** (1935), 777–780.
- [67] Artur K. Ekert, *Quantum cryptography based on Bell’s theorem*, Physical Review Letters **67** (1991), 661–663.
- [68] Bertfried Fauser, Guillaume Raynaud, and Steven Vickers, *The Born rule as structure of spectral bundles (extended abstract)*, (2012), [arXiv:1210.0615](#).
- [69] Arthur Fine, *Hidden variables, joint probability, and the Bell inequalities*, Physical Review Letters **48** (1982), no. 5, 291–295.
- [70] Cecilia Flori, *Group action in topos quantum physics*, (2011), [arXiv:1110.1650v3](#).
- [71] Cecilia Flori and Wilson Brenna, *Complex numbers, one-parameter of unitary transformations and Stone’s theorem in topos quantum theory*, (2012), [arXiv:1206.0809v2](#).
- [72] Cecilia Flori and Joseph Ben Geloun, *Topos analogues of the KMS state*, (2012), [arXiv:1207.0227v2](#).

- [73] Christopher A. Fuchs, *Quantum mechanics as quantum information (and only a little more)*, (2002), [arXiv:quant-ph/0205039v1](#).
- [74] Robin Giles and Hans Kummer, *A non-commutative generalization of topology*, Indiana University Mathematics Journal **21** (1971), 91.
- [75] Steven Givant and Paul Halmos, *Introduction to Boolean algebras*, Springer, 2009.
- [76] Andrew M. Gleason, *Measures on the closed subspaces of Hilbert space*, Journal of Mathematics and Mechanics **6** (1957), 885.
- [77] Robert Goldblatt, *Topoi - the categorical analysis of logic*, Studies in Logic and the Foundations of Mathematics (J. Barwise, D. Kaplan, H. J. Keisler, P. Suppes, and A. S. Troelstra, eds.), vol. 98, North-Holland Publishing Company, 1979.
- [78] Daniel M. Greenberger, Michael A. Horne, Abner Shimony, and Anton Zeilinger, *Bell's theorem without inequalities*, American Journal of Physics **58** (1990), 1131.
- [79] Daniel M Greenberger, Michael A Horne, and Anton Zeilinger, *Going beyond Bell's theorem*, Bell's theorem, quantum theory and conceptions of the universe, Springer, 1989, pp. 69–72.
- [80] John Harding and Mirko Navara, *Subalgebras of orthomodular lattices*, Order **3** (2011), 549, [arXiv:1009.4433](#).
- [81] Godfrey H. Hardy, John E. Littlewood, and George Polya, *Some simple inequalities satisfied by convex functions*, Messenger of Mathematics **58** (1929), 145–152.
- [82] Lucien Hardy, *Quantum mechanics, local realistic theories and Lorentz-invariant realistic theories*, Physical Review Letters **68** (1992), 2981.
- [83] Lucien Hardy, *Nonlocality for two particles without inequalities for almost all entangled states*, Physical Review Letters **71** (1993), 1665–1668.
- [84] Chris Heunen, Nicolaas P. Landsman, and Bas Spitters, *A topos for algebraic quantum theory*, Communications in Mathematical Physics **291** (2009), 63, [arXiv:0709.4364](#).
- [85] Chris Heunen, Nicolaas P. Landsman, and Bas Spitters, *Bohrification*, Deep Beauty (Hans Halvorson, ed.), Cambridge University Press, 2011, p. 271.
- [86] Chris Heunen, Nicolaas P. Landsman, and Bas Spitters, *Bohrification of von Neumann algebras and quantum logic*, Synthese **186** (2012), no. 3, 719, [arXiv:0905.2275](#).

- [87] Alfred Horn, *Doubly stochastic matrices and the diagonal of a rotation matrix*, American Journal of Mathematics **76** (1954), 620.
- [88] Mark Howard, Joel Wallman, Victor Veitch, and Joseph Emerson, *Contextuality supplies the ‘magic’ for quantum computation*, Nature **510** (2014), no. 7505, 351–355, [arXiv:1401.4174](#).
- [89] Chris J. Isham, *Topos theory and consistent histories: The internal logic of the set of all consistent sets*, International Journal of Theoretical Physics **36** (1997), 785, [arXiv:gr-qc/9607069](#).
- [90] Carmen M. Istrate, *Quantales and locales - comparing two notions of spectra for non-commutative C^* -algebras*, Master’s thesis, University of Oxford, 2010.
- [91] Peter T. Johnstone, *Sketches of an elephant - a topos theory compendium I*, Oxford Logic Guides (Dov M. Gabay, Angus MacIntyre, and Dana Scott, eds.), vol. 44, Oxford University Press, 2002.
- [92] Peter T. Johnstone, *Sketches of an elephant - a topos theory compendium II*, Oxford Logic Guides (Dov M. Gabay, Angus MacIntyre, and Dana Scott, eds.), vol. 44, Oxford University Press, 2003.
- [93] Richard V. Kadison and John R. Ringrose, *Fundamentals of the theory of operator algebras, volume I: Elementary theory*, American Mathematical Society, 2002.
- [94] Victor Kaftal and Gary Weiss, *An infinite dimensional Schur-Horn theorem and majorization theory with application to operator ideals*, (2007), [arXiv:0710.5566v2](#).
- [95] Oskar Benjamin Klein, *Zur Quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre*, Zeitschrift für Physik **72** (1931), no. 11, 767.
- [96] Simon B. Kochen and Ernst P. Specker, *The problem of hidden variables in quantum mechanics*, Journal of Mathematics and Mechanics **17** (1967), 59.
- [97] Vipin Kumar, *Algorithms for constraint-satisfaction problems: A survey.*, AI magazine **13** (1992), no. 1, 32–44.
- [98] Azriel Levy, *Basic Set Theory*, Dover Books, 2012.
- [99] Saunders MacLane and Ieke Moerdijk, *Sheaves in geometry and logic: A first introduction to topos theory*, Springer, 1992.

- [100] Shuichiro Maeda, *Probability measures on projections in von Neumann algebras*, Reviews in Mathematical Physics **1** (1989), no. 2/3, 235.
- [101] Shane Mansfield, *The mathematical structure of non-locality and contextuality*, Ph.D. thesis, University of Oxford, 2013.
- [102] Shane Mansfield and Tobias Fritz, *Hardy’s non-locality paradox and possibilistic conditions for non-locality*, Foundations of Physics **42** (2012), no. 5, 709–719, [arXiv:1105.1819](#).
- [103] David N. Mermin, *Simple unified form for the major no-hidden-variables theorems*, Physical Review Letters **65** (1990), no. 27, 3373–3376.
- [104] Ugo Montanari, *Networks of constraints: Fundamental properties and applications to picture processing.*, Information Sciences **7** (1974), 95–132.
- [105] Cristopher J. Mulvey, ‘&’, *Supplemento ai Rendiconti del Circolo Matematico di Palermo* **12** (1986), 99.
- [106] Kunji Nakayama, *Sheaves in quantum topos induced by quantization*, (2011), [arXiv:1109.1192v2](#).
- [107] Michael A. Nielsen and Isaac L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, 2000.
- [108] Joost Nuiten, *Bohrification of local nets of observables*, (2011), [arXiv:1109.1397](#).
- [109] Martin B. Plenio and Shashank Virmani, *An introduction to entanglement measures*, Quantum Information and Computation **7** (2007), 1–51, [arXiv:quant-ph/0504163](#).
- [110] Sandu Popescu and Daniel Rohrlich, *Quantum nonlocality as an axiom*, Foundations of Physics **24** (1994), no. 3, 379–385.
- [111] Robert Raussendorf, *Contextuality in measurement-based quantum computation*, Physical Review A **88** (2013), no. 2, 022322, [arXiv:0907.5449](#).
- [112] Alfréd Rényi, *On measures of information and entropy*, Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press, 1961, p. 547.
- [113] David Sachs, *The lattice of subalgebras of a Boolean algebra*, Canadian Journal of Mathematics **14** (1962), 451–460.
- [114] Daniel Salart, Augustin Baas, Jeroen A. W. van Houwelingen, Nicolas Gisin, and Hugo Zbinden, *Spacelike separation in a Bell test assuming gravitationally induced collapses*, Physical Review Letters **100** (2008), [arXiv:0803.2425](#).

- [115] Emilio Santos, *Bell's theorem and the experiments: Increasing empirical support to local realism?*, *Studies in History and Philosophy of Modern Physics* **2** (2005), 179–206, [arXiv:quant-ph/0410193](#).
- [116] Benjamin Schumacher and Michael Westmoreland, *Quantum processes, systems and information*, Cambridge University Press, 2010.
- [117] Claude E. Shannon, *A mathematical theory of communication*, *Bell System Technical Journal* **27** (1948), no. 3, 379.
- [118] Peter W. Shor, *Polynomial time algorithms for prime factorization and discrete logarithms on a quantum computer*, *SIAM Journal on Scientific and Statistical Computing* **26** (1997), 1484, [arXiv:quant-ph/9508027](#).
- [119] Anthony J. Short and Stephanie Wehner, *Entropy in general physical theories*, *New Journal of Physics* **12** (2010), 033023, [arXiv:0909.4801](#).
- [120] Marshall H. Stone, *The theory of representations of Boolean algebras*, *Transactions of the American Mathematical Society* **40** (1936), 37.
- [121] Alfred Tarski, *Logic, semantics, and metamathematics - papers from 1923 to 1938*, Hackett Publishing Company, 1983.
- [122] Barbara Terhal, *Detecting quantum entanglement*, *Theoretical Computer Science* **287** (2002), 313, [arXiv:quant-ph/0101032](#).
- [123] Veeravalli S. Varadarajan, *Geometry of quantum theory*, 2 ed., Springer, 2007.
- [124] John von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Springer, 1955.
- [125] Zizhu Wang and Damian Markham, *Nonlocality of symmetric states*, *Physical Review Letters* **108** (2012), no. 21, 210407.
- [126] Sander Wolters, *A comparison of two topos-theoretic approaches to quantum theory*, (2010), [arXiv:1010.2031v2](#).
- [127] Thomas Woodhouse, *Time evolution in quantum theory and quantum information, a topos theoretic perspective*, Master's thesis, University of Oxford, 2011.