Topics in Quantum Foundations: Ontological Models, and Distinguishability as a Resource

by

Ryan Morris

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Applied Mathematics

Waterloo, Ontario, Canada, 2009

© Ryan Morris 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis covers research in two disjoint research areas:

The ontological model program (formerly hidden-variables program) for quantum theory has a long and noble tradition in the quantum foundations literature. By postulating a physical reality beyond the quantum state, we gain intuition on quantum phenomena and also come to understand constraints on realist interpretations of quantum theory. Bell's theorem tells us that such an underlying reality must be non-local, while the Kochen-Specker contextuality theorem abuses the classical notion that measurement should simply reveal pre-existing properties of reality. Recent research programs suggest that it is beneficial to view the quantum state as representing purely information. We show that the only current model which does this in a satisfactory manner is unable to reproduce all the statistics of quantum measurements. A recent generalization of the notion of contextuality has allowed for proofs of contextuality which differ from the original Kochen-Specker notion. We add a new result which shows that measurements in a model where the quantum state represents information must be contextual. Additionally, we refine the generalized notion of contextuality into strong and weak forms in order to parse the relationship between new and old results.

Entanglement resource theory is a highly successful investigation of the usefulness of entanglement for information processing tasks. In this thesis we apply the ideas from entanglement resource theory to another resource: state distinguishability. We show analogies between distinguishability resource theory and entanglement resource theory. In particular, the analogy includes: measures which are monotonic under a class of transformations; units of a resource; and bounds on measures in terms of the amount of the unit resource needed to form states and the amount of unit resource that can be extracted from states. We show that the pairs of states which can be reversibly converted into *classical states* are exactly the pairs of simultaneously diagonalizable states. Lastly, we characterize the trace-distance distinguishability of formation on a qubit system.

Acknowledgements

I would like to thank my supervisor Joseph Emerson for taking me on as his student, for his guidance and reassurance throughout my stint as a Masters student, and also for the provision of very interesting and mentally taxing research topics. His persistence and inquisitiveness have led us to ask and attempt to answer many questions I never would have thought to consider.

With regard to support through the mental taxation, thanks very much to my research group: Cozmin Ududec, Balzak Magesan, and Chris Ferrie, for sharing their expertise and ideas affably and often comically. I'll never forget that time when Easwar said ———————————— and then Coz said ———— ———————— and suggested we all go to Starlight. That was rude and hilarious.

Thanks very much to Matt Leifer for intimidating me with his intensely vast knowledge of quantum information and foundations early on in my program. In actuality, he has been ultimately generous and patient in sharing his knowledge and supervising me on the distinguishability project.

Thanks to Chris Fuchs for taking me under his wing, sharing with me his passion for the SIC-POVM and providing me with many pretty mathematical problems to sink my teeth into. Time constraints have prevented me from adequately completing a chapter on our work together, which I deeply regret.

Thanks to my committee: Joseph Emerson, Ed Vrscay and Achim Kempf, for agreeing to read my thesis and for going easy on me during my defense.

A special thank you to those who reviewed my thesis before submission and provided invaluable feedback. These people include Joseph Emerson, Matt Leifer and Chanda Prescod-Weinstein.

I also greatly appreciate the support of my friends and housemates for getting me through good times and bad times smoothly. In particular thanks to Rob Huneault and Mark Ilton for building a fort with me and in general making the sweetest home sitch. Special thanks to Chanda Prescod-Weinstein for her unfailing support and encouragement.

My family is pretty cool also. Thanks Mom and Dad for having me, without you none of this is possible. Thanks Drew for being brotherly.

Lastly, I would like acknowledge and say thanks to a couple members of the Applied Math department: to Professor John Wainwright, who was certainly instrumental in nudging me towards a Masters in Mathematics, and will likely remain the best professor of all time; and to Helen Warren for guiding me through the network of administrative aspects of the program and obtaining funding. Word is that she is the best in the biz.

Absolutely lastly, this research was funded by the Natural Sciences and Engineering Research Council of Canada.

Thank you.

Contents

List of Figures

Chapter 1

Introduction to Quantum Theory

Quantum theory is a theory of physical systems, typically thought to apply to systems at the atomic or subatomic scales. More ambitiously, it is thought to be a physical theory applicable a system of any scale, including the entire universe. In its most abstract form, is a mathematical tool which describes probabilities of outcomes for experiments. In this form, it is not a physical theory in the sense that it does not apply to a specific physical scenario. Classical mechanics can also be presented in an abstract mathematical form [3]. When an abstract theory of classical mechanics is applied to a particular physical system, the connection between abstract outcomes and actual physical properties becomes apparent. Arguably, this is due to the fact that classical theories allow us to predict the outcomes of all experiments with certainty. In their famous paper refuting the 'completeness' of quantum theory [19], Einstein, Podolsky and Rosen, define an element of reality as any variable whose value can be predicted with certainty. Thus by the EPR criterion, classical mechanics, when applied to a physical experiment, describes real properties. Quantum theory, in contrast, does not allow for the outcomes of all experiments to be predicted with certainty. Thus despite its unassailable applicability to physical systems, the theory does not make a clear connection between its abstract outcomes and real properties of the systems under study.

Research in Quantum Foundations occupies itself primarily with trying to understand the differences and similarities between quantum theory and classical mechanics. A well-established project in Quantum Foundations research is to investigate the consequences of postulating underlying physical properties, typically in addition or in lieu of the state-vector ψ, of a system described by quantum theory. Historically, such a research program has been called a hidden-variables program. In order to accommodate viewpoints wherein the quantum state-vector is fundamental and complete, as well as models such as Bohmian mechanics where the 'hidden-variables' are not really hidden, this research path has been re-dubbed as *the ontological model program*. Famous results from this program include theorems by Bell [6, 8] on the non-locality of quantum theory and a theorem by Kochen and Specker [40] on the contextuality of quantum theory. Recent work by Spekkens [58]

and others [31] has generalized the notion of contextuality and rendered it a much more understandable phenomenon.

A growing trend [4, 20, 22, 23, 58, 33] in Quantum Foundations research is to investigate the consequences of viewing the quantum state as a state of incomplete information. It has been found that this view point assuages much of the mystery and paradox typically associated with quantum theory. However, a fully satisfactory ontological model for quantum theory with such a character has yet to be fully developed.

The first chapter of this thesis gives an introduction to quantum theory and the fundamental concepts within that are necessary for the subsequent chapters. Chapters 2 through 4 constitute, for the most part, a literature review of the ontological model program. Chapter 2 introduces the ontological model framework and some commonly considered features of ontological models. In particular we discuss the results on non-locality and contextuality, the notion of outcome determinism, and a simple but new distinction of convex vs. non-convex models. This chapter presents several example ontological models, including the most fully developed model, Bohmian mechanics. In Chapter 3 the characterization of ψ-epistemic [33] models is presented and we investigate the problem of finding models in which the quantum state represents incomplete information. We present a recent model by Rudolph, and a novel proof of why the only other known ψ-epistemic model, the Kochen-Specker model [40] is not sufficient. Chapter 4 describes the generalization of contextuality put forth in [58]. Several proofs of the necessity of contextuality for quantum theory are presented, as well as a new proof that ψ-epistemic theories must always possess *measurement contextuality*. This chapter concludes with a discussion on advantages that contextual theories have over non-contextual theories for certain information and communication tasks, and whether or not the generalized notion of *preparation contextuality* should be viewed as non-classical resource or phenomenon.

Chapter 5 deals with a disjoint research project, in collaboration with Matt Leifer, which deals with quantum information theory. In the spirit of studying entanglement as a resource theory, we develop a framework for viewing distinguishability as a resource. The main contribution of the research comes with the characterization of the *trace distance of formation* on a qubit system. We also present a characterization of states reversibly convertible to classical states.

1.1 The Postulates of Quantum Theory

This thesis concerns itself entirely with a branch of quantum theory which applies only to experiments for which there are a finite number of outcomes. We call this finite dimensional quantum theory. In the following presentation of the postulates of quantum theory, we follow the exposition in Nielsen and Chuang [49][1]. Their presentation focuses on quantum theory as a computational model and thus gives

an operational formulation of the postulates, which is well-suited to the topics of this thesis. We begin, as they do, with the postulates for *state vectors*, and then generalize to the case of *density matrices*.

The first postulate dictates the mathematical objects that are associated with the possible states of the quantum system.

Postulate 1. *Associated to any isolated physical system is a complex, finite dimensional vector space with an inner product (that is, a Hilbert space) known as the state space of the system. The system is completely described by its state vector, which is a unit vector in the system's state space.*

A general complex Hilbert space will be referred to as $\mathcal{H}$, whereas a Hilbert space of a specific finite dimension d will be referred to as $\mathbb{C}^d$. The canonical notation for a unit-norm vector in a Hilbert space is the Dirac 'ket': $|\, x \,\rangle$. The Dirac 'bra' notation: $\langle\, x \,|$, is used to denote a linear functional on $\mathcal{H}$, which acts via the standard inner product between complex vectors: $\langle\, x \,|\, y \,\rangle$, to produce a complex number. As a Hilbert space is self-dual, the elements of $\mathcal{H}$ are in one-to-one correspondence with the linear functionals on $\mathcal{H}$, and we write the linear functional (the 'bra') which identifies with $|\, x \,\rangle$ as $\langle\, x \,|$ i.e. the functional $\langle\, x \,|$ is the unique unit-norm functional such that $\langle\, x \,|\, x \,\rangle = 1$. The transformation between the unit-vector $|\, x \,\rangle$ (a column vector) and its corresponding functional (a row vector) is the complex-transpose, which is denoted as $|\, x \,\rangle^\dagger = \langle\, x \,|$.

The second postulate concerns the dynamics of a quantum system.

Postulate 2. *The evolution of a* closed *quantum system is described by a* unitary transformation. *That is, the state ρ of the system at time t_1 is related to the state ρ' of the system at time t_2 by a unitary operator U which depends only on the times t_1 and t_2,*

$$|\, \psi' \,\rangle = U \,|\, \psi \,\rangle\,. \tag{1.1}$$

The word 'closed' in the above statement indicates that the system is completely isolated, hence it is not interacting with any other system. The evolution of an isolated quantum system is always expressible as a unitary operator acting on the state vector. If the quantum system is interacting strongly with another, then the representation of the dynamics on the original system can no longer necessarily be described as a unitary transformation. We will touch more on this after the 4^{th} postulate is presented.

In classical mechanics a space of real properties is posited, known as a configuration space. Classical mechanics also dictates a set of possible evolutions of particles through configuration space. It is tacitly assumed that at any point an experimenter or observer could observe the exact configuration or physical properties of the system without affecting the evolution. This observation could be called a measurement. In quantum theory, an explicit statement about how measurements

[1]The mathematical statements in this section are taken from [49], [36] or [14].

can occur is given. Furthermore, this third postulate also indicates that the process of measurement will change the quantum state.

Postulate 3. *Quantum measurements are described by a collection* $\{M_m\}$ *of* measurement operators*. These are operators acting on the state space of the system being measured. The index* m *refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is* $|\,\psi\,\rangle$ *immediately before the measurement then the probability that result* m *occurs is given by*

$$p(m) = \langle\,\psi\,|\,M_m^\dagger M_m\,|\,\psi\,\rangle\,, \tag{1.2}$$

and the state of the system after the measurement is

$$\frac{M_m\,|\,\psi\,\rangle}{\sqrt{\langle\,\psi\,|\,M_m^\dagger M_m\,|\,\psi\,\rangle}}. \tag{1.3}$$

The measurement operators satisfy the completeness equation*,*

$$\sum_m M_m^\dagger M_m = I, \tag{1.4}$$

where I *is identity matrix on* $\mathcal{H}$*. The completeness equation expressed the fact that probabilities sum to one:*

$$1 = \sum_m p(m) = \sum_m \langle\,\psi\,|\,M_m^\dagger M_m\,|\,\psi\,\rangle\,. \tag{1.5}$$

This postulate is stated with maximal generality, and we need to discuss particular special cases of measurements. There exists a set of 'most-precise' measurements from which any general measurement as described in Postulate 3 can be constructed. This set of measurements are the rank-1 Projector-Valued Measures (PVMs). In a PVM (not necessarily rank-1), the collection $B = \{M_m\} = \{P_m\}_{m=1}^n$ is comprised of projectors which sum to identity. If the projectors are all rank-1, then it is a rank-1 PVM, and the set of projectors corresponds to an orthonormal basis for $\mathcal{H}$. In this case, we may refer to the measurement as a measurement in the B basis. Projectors have two important properties: they are self-adjoint and idempotent:

$$P_m^\dagger = P_m, \qquad P_m^2 = P_m. \tag{1.6}$$

These properties can be seen quite easily when we write a rank-1 projector in Dirac notation: $P_m = |\,m\,\rangle\!\langle\,m\,|$,

$$\begin{aligned}(|\,m\,\rangle\!\langle\,m\,|)^\dagger &= (\langle\,m\,|)^\dagger(|\,m\,\rangle)^\dagger = |\,m\,\rangle\!\langle\,m\,|\,,\\ (|\,m\,\rangle\!\langle\,m\,|)^2 &= |\,m\,\rangle\,\langle\,m|\,m\,\rangle\,\langle\,m\,| = |\,m\,\rangle\!\langle\,m\,|\,.\end{aligned} \tag{1.7}$$

Given these properties and the fact that a set of d mutually orthogonal projectors acting on $\mathbb{C}^d$ form a resolution of the identity, a PVM will satisfy the completeness

equation (1.24). These properties also simplify equations (1.22) and (1.23). The probability of outcome m becomes:

$$\begin{aligned} p(m) &= \langle\, \psi \,|\, P_m^\dagger P_m \,|\, \psi \,\rangle = \langle\, \psi \,|\, P_m^2 \,|\, \psi \,\rangle \\ &= \langle\, \psi \,|\, P_m \,|\, \psi \,\rangle = \langle\, \psi \,|\, |\, m \,\rangle\langle\, m \,|\, |\, \psi \,\rangle = |\, \langle\, \psi \,|\, m \,\rangle\, |^2. \end{aligned} \tag{1.8}$$

The update rule simplifies to:

$$\frac{P_m \,|\, \psi \,\rangle}{\sqrt{\langle\, \psi \,|\, P_m \,|\, \psi \,\rangle}} = \frac{|\, m \,\rangle\langle\, m \,|\, |\, \psi \,\rangle}{\langle\, m \,|\, \psi \,\rangle} = |\, m \,\rangle\,. \tag{1.9}$$

Regardless of the quantum state $|\, \psi \,\rangle$, unless $\langle\, m \,|\, \psi \,\rangle = 0$, the state will update to the pure state corresponding to outcome m. The set of PVM measurements is a strictly smaller set than the set of measurements described in Postulate 3. The relationship between the two formalisms will be discussed in the next section.

Some statements of the measurement postulate will associate a measurement process with any self-adjoint operator A. A, being self-adjoint, will have a spectral decomposition

$$A = \sum_{j=1}^{d} \lambda_j \,|\, j \,\rangle\langle\, j \,|\, \tag{1.10}$$

with respect to a unique set of eigenvalues $\{\lambda_j\}$ and a complete set of orthonormal eigenvectors $\{|\, j \,\rangle\langle\, j \,|\}$. Thus a self-adjoint operator A dictates a PVM with a bit of added structure. The eigenvectors form the PVM as presented above; however, the eigenvalues are interpreted to be numerical values associated to each outcome. An observable A is intended to correspond to a 'property' of a quantum state, and if the outcome j of the associated PVM occurs, then the quantum state is considered to 'have value λ_j for observable A'. For example, suppose one were presented with three closed numbered boxes (1 to 3), each containing a specified number of marbles, and randomly choose to open one of the boxes. Then the analogy would be that each box corresponds to an eigenvector, and the number of marbles in each box would correspond to the associated eigenvalue. The experiment of simply choosing a box to open, and ignoring the contents, is analogous to the PVM.

It is often the case that we are not concerned with how the quantum state is updated after a measurement. If this is the case, then a more simplified formalism than presented in Postulate 3 is employed. A Positive Operator Valued Measure (POVM) is a collection of positive operators $\{E_m\}$ satisfying a completeness condition,

$$\sum_m E_m = I. \tag{1.11}$$

Being unconcerned with state update, we simply state that the probability of outcome m given that the system has state $|\, \psi \,\rangle$ is given by

$$p(m) = \langle\, \psi \,|\, E_m \,|\, \psi \,\rangle\,. \tag{1.12}$$

A matrix E is positive if and only if it can be decomposed as the product of a matrix and its complex-conjugate, i.e., if and only if E is a Gram matrix of a set of vectors. Thus any measurement as presented in Postulate 3 can be expressed as a POVM since $M_m^\dagger M_m$ is positive. Conversely, any positive matrix E_m can be decomposed into the product of two operators $E_m = M_m^\dagger M_m$. Hence the POVM formalism describes the same set of measurements as Postulate 3. We often refer to the elements of a POVM as *quantum effects*.

The last postulate describes how two separate quantum systems can interact with each other. When two separate systems are brought together and considered as one whole system, this is called a *composite system* or a *bipartite system*.

Postulate 4. *The state space of a composite system is the tensor product of the state spaces of the component systems. Moreover, if we have systems* 1 *through* n*, each independently prepared in the respective states* $|\,\psi_i\,\rangle$*, then the joint state of these* n *systems is* $|\,\psi_1\,\rangle \otimes |\,\psi_2\,\rangle \otimes \ldots |\,\psi_n\,\rangle$.

If $|\,x\,\rangle_M \in \mathbb{C}^M$ and $|\,y\,\rangle_N \in \mathbb{C}^N$, then $|\,x\,\rangle_M \otimes |\,y\,\rangle_N \in \mathbb{C}^{MN}$. Such a state is also typically abbreviated as $|\,x_M y_N\,\rangle$ or just $|\,xy\,\rangle$. The postulates 1 through 3 apply to an isolated composite system just as well as they do to an isolated non-composite system. If $\{|\,m\,\rangle\}_{m=1}^M$ is an orthonormal basis for $\mathbb{C}^M$ and $\{|\,n\,\rangle\}_{n=1}^N$ is an orthonormal basis for $\mathbb{C}^N$, then $\{|\,mn\,\rangle\}_{m,n=1}^{M,N}$ is an orthonormal basis for $\mathbb{C}^{MN}$.

As indicated by Postulate 1, the full set of state vectors for the composite system is given by the unit vectors on $\mathbb{C}^{MN}$. However, this set is larger than the set of states that can be achieved through the independent preparation of a state on $\mathbb{C}^M$ and a state on $\mathbb{C}^N$. Mathematically, the set of *product states*,

$$\left\{|\,xy\,\rangle \;|\; |\,x\,\rangle \in \mathbb{C}^M, |\,y\,\rangle \in \mathbb{C}^N\right\},$$

is strictly smaller than the set of unit vectors on $\mathbb{C}^{MN-1}$. For example, the state

$$\frac{1}{\sqrt{2}}(|\,00\,\rangle + |\,11\,\rangle)$$

in $\mathbb{C}^4$, where $\{|\,m\,\rangle\}_{m=0}^1$ and $\{|\,n\,\rangle\}_{n=0}^1$ are orthonormal bases for separate two-dimensional quantum systems, cannot be expressed as a product state. However, this state can be realized by independently preparing two separate systems and then performing appropriate transformations on the two systems together. A quantum state comprised of two separate quantum states is called a *bipartite* state.

In the next section, we discuss how state vectors and rank-1 PVMs, together with probability and interactions with other systems, are enough to derive a more generalized formulation of states and measurements than given in the above postulates. The generalized states are called *density operators* and the generalized measurements were mentioned in Postulate 2. We also discuss a generalization of the possible evolutions or operations that can be performed on quantum states. Such general operations are called called Trace-Preserving Completely Positive maps.

1.2 Generalized States, Measurements and Operations

Suppose we wish to describe a quantum system whose state is only known probabilistically, or suppose that we wish to describe a system A as it interacts with another system (often called an *ancilla system*) which we do not explicitly describe. The postulates presented in the previous section do not account for such situations. However, this can be easily rectified by switching to the *density operator* formalism.

A *density operator or matrix* is a positive linear operator ρ with trace one, acting on the state space of the system[2]. The space of density matrices, which we denote as $K(\mathcal{H})$, is a convex set whose extreme points are the rank-1 projectors, i.e. operators (matrices) of the form $|\, x \,\rangle\langle\, x \,|$, whose action on vector $|\, y \,\rangle$ is given by $|\, x \,\rangle\langle\, x \,|\,|\, y \,\rangle = \langle\, x \,|\, y \,\rangle\,|\, x \,\rangle$. The rank-1 projectors are called pure states. The set of density matrices is the convex hull of the pure states, so any density matrix can be decomposed as:

$$\rho = \sum_{j=1}^{n} p_j \,|\, x_j \,\rangle\langle\, x_j \,| \quad \text{where} \quad p_j \geq 0,\ \sum_{j=1}^{n} p_j = 1. \tag{1.13}$$

for some set of pure states $\{|\, x_j \,\rangle\langle\, x_j \,|\}_{j=1}^{n}$. As the density matrices are positive, they are necessarily self-adjoint, and thus have a spectral decomposition. Hence every density matrix has a convex decomposition in terms of an orthonormal basis $\{|\, i \,\rangle\}_{i=1}^{d}$ for $\mathbb{C}^d$:

$$\rho = \sum_{j=1}^{d} p_j \,|\, j \,\rangle\langle\, j \,| \quad \text{where} \quad p_j \geq 0,\ \sum_{j=1}^{d} p_j = 1,\ \langle\, i \,|\, j \,\rangle = \delta_{i,j}. \tag{1.14}$$

As a rank-1 projector is an outer product of a unit vector in $\mathcal{H}$, there is a simple relationship between the set of state vectors on $\mathcal{H}$ and the set of pure states. Specifically, if two unit vectors are related to each other by a *global phase*, i.e. $|\, x \,\rangle = \exp(i\phi)\,|\, y \,\rangle$, then $|\, x \,\rangle\langle\, x \,| = |\, y \,\rangle\langle\, y \,|$ since the global phase cancels with itself. Thus there is a bijection between the pure states and the *projective Hilbert space* $\mathbb{P}\mathcal{H}$, which is the space of unit vectors under the equivalence class of global phase multiplication. For a particular finite dimension Hilbert space $\mathbb{C}^d$, the projective Hilbert space is denoted as $\mathbb{CP}^{d-1}$. Thus a pure state may be written as a projector on $\mathcal{H}$ ($|\, x \,\rangle\langle\, x \,|$), or as a unit vector in $\mathcal{H}$ ($|\, x \,\rangle$).

For a given density matrix ρ, there are generally two mechanisms by which a system could be prepared to correspond to ρ. The first is probability based. A density matrix always has a decomposition as a convex combination of pure states (1.13). Experimentally, this state can come about purely through the ignorance of the experimenter. The experimenter could sample the value j from the probability distribution p_j and subsequently prepare the state $|\, x_j \,\rangle\langle\, x_j \,|$. Every non-pure state

[2] Positive will be taken to mean the more precise term positive-semidefinite.

has, in fact, an infinite number of such decompositions [49], and so there are an infinite number of ways to prepare any non-pure state.

However, ρ can also arise as the state describing a system which is part of a larger composite system. Suppose two systems, represented by Hilbert spaces $\mathcal{H}_A = \mathbb{C}_A^{d_1}$ and $\mathcal{H}_B = \mathbb{C}_B^{d_2}$ are interacting. Postulate 4 tells us that the state space for the composite system is given by the tensor product: $\mathcal{H}_{AB} = \mathbb{C}_{AB}^{d_1 d_2} = \mathbb{C}_A^{d_1} \otimes \mathbb{C}_B^{d_2}$. Denote the state of the composite system as ρ_{AB}. Consider the set of measurements on the system $\mathcal{H}_{AB}$ which are completely ignorant of the subsystem B. These measurements are of the form $\{E_m \otimes I_B\}$ where I_B is the identity matrix on $\mathcal{H}_B$ and $\{E_m\}$ is a POVM on $\mathcal{H}_A$. There is a unique density matrix ρ_A that reproduces the statistics of ρ_{AB} for all such measurements. This is called the *reduced density matrix on* A, and it can be calculated via the partial-trace operation on ρ_{AB},

$$\rho_A = \mathrm{tr}_B(\rho_{AB}) = \sum_{j=1}^{d_2} (I_A \otimes \langle\, j_B \,|)\rho_{AB}(I_A \otimes |\, j_B \,\rangle), \tag{1.15}$$

where $\{|\, j_B \,\rangle\}$ is any orthonormal basis for $\mathcal{H}_B$. The density matrix ρ_A arising as a subsystem cannot be viewed as a probabilistic combination of pure states on $\mathcal{H}_A$, as it is in fact an element of a known state ρ_{AB} on $\mathcal{H}_{AB}$.

Despite the fact that these two processes can give rise to the exact same set of density matrices, there have been historical and philosophical reasons for distinguishing between the two types. A non-pure density matrix prepared via ignorance is referred to as a *proper mixture*, while a non-pure density matrix representing a subsystem of a larger system is referred to as an *improper mixture*. The proper mixture captures the states of a system that can be produced if the experimenter has access only to that isolated system, and a source of randomness or ignorance.

Having explained the density matrix, we can now restate Posulate 1.

Postulate 1. *Associated to any isolated physical system is a complex, finite-dimensional vector space with an inner product (that is, a Hilbert space) known as the state space of the system. The system is completely described by its density operator, which is a positive linear operator ρ with trace one, acting on the state space of the system.*

The remaining Postulates easily adapt to this new description of the system state.

First, consider evolution. If a state vector evolves as $|\, \psi \,\rangle \to U\,|\, \psi \,\rangle$, then it is clear that a density matrix will evolve as:

$$\rho = \sum_i p_i\,|\, \psi_i \,\rangle\langle\, \psi_i \,| \to \sum_i p_i U\,|\, \psi_i \,\rangle\langle\, \psi_i \,| U^\dagger = U\rho U^\dagger. \tag{1.16}$$

Thus a restated Postulate 2 is:

Postulate 2. *The evolution of a* closed *quantum system is described by a* unitary transformation. *That is, the state ρ of the system at time t_1 is related to the state*

ρ' of the system at time t_2 by a unitary operator U which depends only on the times t_1 and t_2,

$$\rho' = U\rho. \tag{1.17}$$

Consider measurements. If the quantum state is $|\,\psi\,\rangle$, then the probability of a measurement outcome M_m is given by $\langle\,\psi\,|\,M_m^\dagger M_m\,|\,\psi\,\rangle = \mathrm{tr}(|\,\psi\,\rangle\langle\,\psi\,|\,M_m^\dagger M_m)$. Thus if the density operator is $\rho = \sum_i p_i\,|\,\psi_i\,\rangle\langle\,\psi_i\,|$, then the probability of outcome M_m is given by

$$p_m = \sum_i p_i\,\mathrm{tr}(|\,\psi_i\,\rangle\langle\,\psi_i\,|\,M_m^\dagger M_m) = \mathrm{tr}(\rho M_m^\dagger M_m). \tag{1.18}$$

In the POVM formalism, this is simply $\mathrm{tr}(\rho E_m)$. If the quantum state was $|\,\psi_i\,\rangle$ then it updates to

$$|\,\psi_i^m\,\rangle = \frac{M_m\,|\,\psi_i\,\rangle}{\sqrt{\langle\,\psi_i\,|\,M_m^\dagger M_m\,|\,\psi_i\,\rangle}}.$$

Thus if the system is in state $|\,\psi_i\,\rangle$ with probability p_i (i.e. in the state $\rho = \sum_i p_i\,|\,\psi_i\,\rangle\langle\,\psi_i\,|$) then the density operator updates to

$$\rho_m = \sum_i p_{i|m}\,|\,\psi_i^m\,\rangle\langle\,\psi_i^m\,| \tag{1.19}$$

where $p_{i|m}$ is the probability that the state *was* $|\,\psi_i\,\rangle$ given that outcome m occured. By Bayes theorem,

$$p_{i|m} = \frac{p_{m|i}p_i}{p_m} = \frac{p_i\,\mathrm{tr}(|\,\psi_i\,\rangle\langle\,\psi_i\,|\,M_m^\dagger M_m)}{\mathrm{tr}(\rho M_m^\dagger M_m)}. \tag{1.20}$$

Thus the updated density matrix is

$$\rho_m = \sum_i p_i \frac{M_m\,|\,\psi_i\,\rangle\langle\,\psi_i\,|\,M_m^\dagger}{\mathrm{tr}(\rho M_m^\dagger M_m)} = \frac{M_m\rho M_m^\dagger}{\mathrm{tr}(\rho M_m^\dagger M_m)}. \tag{1.21}$$

Thus the restated third postulate reads:

Postulate 3. *Quantum measurements are described by a collection $\{M_m\}$ of* measurement operators*. These are operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is ρ immediately before the measurement then the probability that result m occurs is given by*

$$p(m) = \mathrm{tr}(M_m^\dagger M_m \rho), \tag{1.22}$$

and the state of the system after the measurement is

$$\frac{M_m\rho M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m\rho)}. \tag{1.23}$$

The measurement operators satisfy the completeness equation*,*

$$\sum_m M_m^\dagger M_m = I, \tag{1.24}$$

where I is identity matrix on $\mathcal{H}$.

Lastly, the fourth postulate generalizes easily as:

Postulate 4. *The state space of a composite system is the tensor product of the state spaces of the component systems. Moreover, if we have systems 1 through n, each independently prepared in the respective states ρ_i, then the joint state of these n systems is $\rho_1 \otimes \rho_2 \otimes \ldots \rho_n$.*

We now consider how measurements other than rank-1 PVMs arise from ignorance/relabeling or from coupling with an ancilla system. We start by considering measurements arising from ignorance.

It is not hard to see that any PVM (elements not necessarily rank-1) can be implemented with rank-1 PVM. A set of projectors which sums to identity is at most a coarse graining of a set of rank-1 projectors which sum to identity. Thus we can imagine that a general PVM can be realized by performing a rank-1 PVM and then relabeling the outputs. Mathematically, let $B = \{P_j\}_{j=1}^d$ be a rank-1 PVM on $\mathbb{C}^d$. If outcome j gets relabeled as outcome $m(j)$, then we represent this with the vector $\vec{e}^{m(j)}$ which has a 1 in the $m(j)^{th}$ position and is zero elsewhere. Then the PVM arising from performing B, and then relabeling, has effects given by:

$$P_m = \sum_j e_m^{m(j)} P_j.$$

However, we could also probabilistically mix PVMs of this sort. That is, prior to performing an actual measurement, an experimenter could sample a value k randomly, and then perform a PVM associated with k. For example, denoting the k^{th} PVM as $B_k = \left\{P_j^k\right\}$, and $(p_1, \ldots, p_K)$ as a K-dimensional probability distribution, the probabilistic mixture of PVMs B_k has effects given by

$$E_m = \sum_{k,j} p_k e_m^{m(j,k)} P_j^k,$$

where $m(j,k)$ is the label assigned to the j^{th} outcome of the k^{th} PVM. In analogy with the nomenclature for density matrices, we make the following definition.

Definition 1.1 (Proper d-level POVM)**.** A POVM (which is not a PVM) acting on $\mathbb{C}^d$ which can be realized through a combination of randomness and relabeling will be called a *proper d-level POVM*.

Another method of performing POVMs, similarly to an improper mixture, requires an ancilla system. The most general measurement that can be performed on

a system in $\mathcal{H}_A$ is to couple to an ancilla ($\mathcal{H}_B$), perform a unitary transformation on the total system, and then perform a PVM on the B system i.e. a PVM of the form $\{I_A \otimes |\, j_B \rangle\langle\, j_B \,|\}$ for some orthonormal basis $\{|\, j_B \rangle\langle\, j_B \,|\}$ on $\mathcal{H}_B$. It can be shown that any POVM on $\mathcal{H}_A$ can be performed in this fashion [49].

Definition 1.2 (Improper d-level POVM)**.** A POVM on $\mathbb{C}^d$ performed by coupling and measuring on an ancilla system is called *an improper d-level POVM.*

Unlike the situation for density matrices, the set of improper d-POVMs is a strict superset of the proper POVMs. There are POVMs which cannot be implemented as proper d-level POVMs. This will be shown explicitly in a later section (see Section 3.3.1).

Postulate 2 states that the evolution of a closed quantum system is described by a unitary operator. As was the case with pure states and PVMs, unitary operators are not the most general allowed operations on a quantum state. In fact, a quantum measurement, together with the update rule (1.23), could be considered as an operation as it induces a change in the quantum state. Also, similarly to how an improper POVM arises from coupling to an ancilla system and measuring, we can evolve a quantum state by coupling to an ancilla, evolving the entire system by a unitary, and then discarding a portion of the whole system. In such a scenario we say that the original system evolves through interaction with an environment, or undergoes open evolution. If we choose to discard a portion of the system that is more or less than the original ancilla, then we can describe operations that take quantum states between Hilbert spaces of different dimension.

A general quantum operation, $\mathcal{E} : K(\mathbb{C}^M) \to K(\mathbb{C}^N)$, which encompasses unitary evolution, measurements, and environment interactions has the following properties:

- Linearity;
- Trace preservation - $\mathrm{tr}[\mathcal{E}(\rho)] = 1$;
- Complete Positivity - The requirement of positivity for $\mathcal{E}$ is that for any $\rho \in K(\mathbb{C}^M)$, $\mathcal{E}(\rho)$ is positive. The requirement of complete positivity is that for any ancilla system $\mathbb{C}^N$ and any state ρ on the coupled Hilbert space $\mathbb{C}^N \otimes \mathbb{C}^M$, $(\mathcal{I} \otimes \mathcal{E})(\rho)$ is positive, where $\mathcal{I}$ is the identity operator on $\mathbb{C}^N$.[3]

These general operations are typically referred to as Trace-Preserving Completely Positive (TPCP) maps.

An important characterization of the TPCP maps comes from the *Kraus representation.* Specifically, any TPCP completely positive map $\mathcal{E}$ from $K(\mathbb{C}^M)$ to

[3]The tensor product operator $(\mathcal{E}_1 \otimes \mathcal{E}_2)$ is the unique linear operator on $\mathbb{C}^N \otimes \mathbb{C}^M$ that acts as follows on a product state: $(\mathcal{E}_1 \otimes \mathcal{E}_2)(\rho_1 \otimes \rho_2) = \mathcal{E}_1(\rho_1) \otimes \mathcal{E}_2(\rho_2)$

$K(\mathbb{C}^N)$ can be represented by a set $\{E_i\}_{i=1}^n$ of M-by-N matrices such that

$$\sum_{i=1}^{n} E_i^\dagger E_i = \mathbb{I}_N, \tag{1.25}$$

and

$$\mathcal{E}(\rho) = \sum_{i=1}^{n} E_i \rho E_i^\dagger. \tag{1.26}$$

We refer to any set of matrices $\{E_i\}_{i=1}^n$ satisfying (1.25) as set of *Kraus operators*. Conversely, any set of Kraus operators specifies a valid TPCP map.

1.3 The Bloch Sphere Model of the Qubit

The geometry of the quantum states on a 2-dimensional Hilbert space lends itself to a particularly nice geometrical picture, known as the Bloch Sphere. A 2-dimensional quantum system is referred to as a *qubit* in analogy to a bit as being the fundamental element of classical computation. Similarly, the qubit is the fundamental element of quantum computation. Recall that a pure quantum state is a one-dimensional projector, so it has the form $|\, x \,\rangle\langle\, x \,|$, with $|\, x \,\rangle \in \mathbb{C}^d$ and $\langle\, x \,|\, x \,\rangle = 1$. Suppose now that we take the vectors $|\, 0 \,\rangle$ and $|\, 1 \,\rangle$ to be the standard basis vectors for $\mathbb{C}^2$. Then, any unit vector can be written as

$$|\, x \,\rangle = \alpha\, |\, 0 \,\rangle + \beta\, |\, 1 \,\rangle \qquad \text{where } |\alpha|^2 + |\beta|^2 = 1 \tag{1.27}$$

Hence we may choose the following parameterization:

$$\alpha = \cos(\tfrac{\theta}{2}) e^{i\gamma}, \qquad \beta = \sin(\tfrac{\theta}{2}) e^{i(\phi+\gamma)} \tag{1.28}$$

with $\theta \in [0, \pi]$ and $\phi, \gamma \in [0, 2\pi]$. Thus,

$$|\, x \,\rangle = e^{i\gamma} \left(\cos(\tfrac{\theta}{2})\, |\, 0 \,\rangle + \sin(\tfrac{\theta}{2}) e^{i(\phi)}\, |\, 1 \,\rangle \right). \tag{1.29}$$

However, from a statistical point of view, the global phase γ plays no role, as was mentioned in Section 1.1. Thus any pure state projector can be parameterized by $\{(\theta, \phi) \mid \theta \in [0, \pi], \phi \in [0, 2\pi]\}$. The set of pure states for a qubit, $\mathbb{CP}^1$, is then isomorphic to the points on the unit 2-sphere $\mathcal{S}^2$, parameterized in the standard way by θ and ϕ (with θ as the zenith angle, and ϕ as the azimuth angle.). See Figure 1.1.

Thus we can refer to a point on the sphere $\mathcal{S}^2$ and a qubit pure state $|\, x \,\rangle\langle\, x \,|$ interchangeably. In particular the point on the sphere corresponding to coordinates (θ, ψ) is given by the Hilbert space vector

$$|\, \theta, \phi \,\rangle = \cos(\tfrac{\theta}{2})\, |\, 0 \,\rangle + \sin(\tfrac{\theta}{2}) e^{i\phi}\, |\, 1 \,\rangle, \tag{1.30}$$

or the projector $|\, \theta, \phi \,\rangle\langle\, \theta, \phi \,|$.

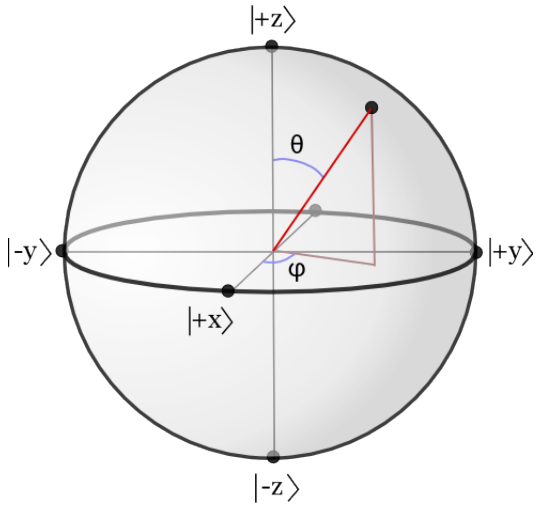

Figure 1.1: In the standard view of the Bloch Sphere, the standard basis states $|0\rangle$ and $|1\rangle$ correspond respectively to $|+z\rangle$ and $|-z\rangle$. In the parameterization we have set up, the states on the other axes are: $|+x\rangle = |\pi/2, 0\rangle$, $|-x\rangle = |-\pi/2, 0\rangle$, $|+y\rangle = |\pi/2, \pi/2\rangle$, $|-y\rangle = |-\pi/2, \pi/2\rangle$.

Notice that the two orthogonal vectors $|0\rangle$ and $|1\rangle$ occur at antipodal points on the Bloch sphere, at coordinates $(0, \phi)$ and (π, ϕ) respectively. Such is the case for all pairs of orthogonal vectors in $\mathbb{CP}^1$. Also of note is the fact that the interior of the Bloch sphere can be used to visualize the mixed qubit states. This is a consequence of the fact that all density matrices on $\mathbb{C}^2$ are convex combinations of the pure states, and the unit ball is the convex hull of the unit sphere. Given that any vector $\vec{x}$ such that $\vec{x} \cdot \vec{x} \leq 1$ represents a quantum state ρ on $\mathbb{C}^2$, we refer to such a vector as *the Bloch vector for* ρ.

The Bloch sphere is also useful for visualizing the possible observables associated with a 2-dimensional quantum system. The set $\mathcal{L}_{sa}(\mathbb{C}^2)$ of hermitian matrices acting on $\mathbb{C}^2$ can be viewed as a real vector space with the following basis:

$$\begin{aligned} I &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \\ \sigma_y &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \end{aligned} \tag{1.31}$$

The three matrices $\{\sigma_x, \sigma_y, \sigma_z\}$, are known as the Pauli matrices, and are in fact also unitary operators. Notice that they are traceless, i.e. $\mathrm{tr}(\sigma_\alpha) = 0$, and I is not. The trace being linear, the Pauli matrices form a basis for the real-vector space V of traceless Hermitian operators acting on $\mathbb{C}^2$. If we assign the three canonical unit vectors $\left\{\hat{i}, \hat{j}, \hat{k}\right\}$ respectively to the Pauli matrices, then we derive a linear isomorphism P between V and $\mathbb{R}^3$ [40]:

$$P(x, y, z) = xP(\hat{i}) + yP(\hat{j}) + zP(\hat{k}) = x\sigma_x + y\sigma_y + z\sigma_z = \sigma. \tag{1.32}$$

Let $\mathcal{C}$ be the change in coordinates between the Cartesian coordinates (x, y, z) where $x^2 + y^2 + z^2 = 1$ and the spherical coordinates (θ, ϕ):

$$\mathcal{C}(x, y, z) = \left(\arctan\left(\frac{\sqrt{(x^2+y^2)}}{z} \right), \arctan\left(\tfrac{y}{z}\right) \right), \tag{1.33}$$

$$\mathcal{C}^{-1}(\theta, \phi) = (\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)). \tag{1.34}$$

Then if (x, y, z) is a unit vector (a point on $\mathcal{S}^2$), then the Hilbert space vector $|\, \mathcal{C}(x, y, z)\, \rangle$ is the $+1$ eigenvector of the of the hermitian matrix $P(x, y, z)$. Thus a point on the Bloch sphere corresponds to a pure state and also to the unique PVM (together with its antipodal point) for which the outcome corresponding to that pure state is certain.

In the case where the quantum system described by $\mathbb{C}^2$ corresponds to an electron spin, the Hermitian matrix $P(x, y, z)$ is interpreted as the observable corresponding to measuring the electron spin component in the direction (x, y, z). Such an experiment is typically thought to be done by having the electron pass through with a Stern-Gerlach magnet oriented and polarized in the direction (x, y, z).

Any other observable $A \in \mathcal{L}_{sa}(\mathbb{C}^2)$ can be thought of as an observable in V with its outcomes relabeled. For example, if A has two distinct eigenvalues λ_1 and λ_2, then [40]

$$\sigma(A) = \frac{2}{\lambda_1 - \lambda_2} A - \frac{\lambda_1 + \lambda_2}{\lambda_1 - \lambda_2} I \tag{1.35}$$

is a traceless hermitian operator. Moreover, $\sigma(A)$ has the same eigenvectors as A, with the λ_1 (λ_2) eigenvector having eigenvalue $+1$ (-1). If A has identical eigenvalues λ, then this just corresponds to the experiment where one checks if the system is there and assigns outcome λ to that outcome.

Thus we see that the Bloch sphere is a useful tool for visualizing the states and measurements that are possible on a qubit system. As such it also provides a simple way to calculate the statistics of outcomes. Suppose that the system is in the quantum state $|\, 0, 0 \,\rangle\langle\, 0, 0\, |$, and we perform the PVM involving the observable $P\mathcal{C}^{-1}(\theta, \phi)$ i.e the PVM containing $|\, \theta, \phi \,\rangle\langle\, \theta, \phi\, |$ as one of its rank-1 projectors. The probability that getting the outcome $+1$ (corresponding to $|\, \theta, \phi \,\rangle\langle\, \theta, \phi\, |$) when in state $|\, 0, 0 \,\rangle\langle\, 0, 0\, |$ is

$$\mathrm{tr}\left(|\, \theta, \phi \,\rangle\langle\, \theta, \phi\, |\, |\, 0, 0 \,\rangle\langle\, 0, 0\, | \right) = \langle\, 0\, |\, |\, \theta, \phi \,\rangle\langle\, \theta, \phi\, |\, |\, 0\, \rangle = \cos^2\left(\tfrac{\theta}{2}\right). \tag{1.36}$$

Due to the choice of $|\, 0, 0 \,\rangle\langle\, 0, 0\, |$ as the system state, the value θ appearing in $\cos^2\left(\frac{\theta}{2}\right)$ is the angular separation of the Bloch sphere points corresponding to the state and the observable $P\mathcal{C}^{-1}(\theta, \phi)$.

Given any pure state $|\, \alpha, \beta \,\rangle\langle\, \alpha, \beta\, |$ there is always a unitary matrix U such that $U\, |\, \alpha, \beta \,\rangle\langle\, \alpha, \beta\, |\, U^\dagger = |\, 0 \,\rangle\langle\, 0\, |$. If we transform another projector $|\, x \,\rangle\langle\, x\, |$ by the same unitary, then

$$\begin{aligned} \mathrm{tr}\left(U\, |\, x \,\rangle\langle\, x\, |\, U^\dagger\, |\, 0 \,\rangle\langle\, 0\, | \right) &= \mathrm{tr}\left(U\, |\, x \,\rangle\langle\, x\, |\, U^\dagger U\, |\, \alpha, \beta \,\rangle\langle\, \alpha, \beta\, |\, U^\dagger \right) \\ &= \mathrm{tr}\left(|\, x \,\rangle\langle\, x\, |\, |\, \alpha, \beta \,\rangle\langle\, \alpha, \beta\, | \right), \end{aligned} \tag{1.37}$$

where we used the cyclic property of the trace and the fact that $UU^\dagger = U^\dagger U = I$ in the last line. Thus the probability of outcome $+1$ for the observable with $+1$ eigenvector $|\, x \,\rangle\langle\, x \,|$ when in the state $|\, \alpha, \beta \,\rangle\langle\, \alpha, \beta \,|$, is $\cos\left(\frac{\theta}{2}\right)$, where θ is the angular separation between the points corresponding to $|\, 0 \,\rangle\langle\, 0 \,|$ and $U \,|\, \alpha, \beta \,\rangle\langle\, \alpha, \beta \,|\, U^\dagger$. However, the action of the unitary matrices, translated to the Bloch sphere, amounts to a smooth rotation of the sphere. Thus angular separations are preserved, and we can always calculate probabilities of outcomes in terms of the angular separation between representations of pure states and PVM projectors on the Bloch sphere.

Chapter 2

Ontological Models for Operational Theories

In order to understand the differences and similarities between quantum theory and classical theory, it is enlightening to place them both within the context of a general framework which can accommodate both theories. This approach is very typical of much research in quantum foundations. For example, there have been efforts to provide axiomatizations for quantum theory in which the change or removal of one axiom gives rise to classical theory [30, 53, 54]. The study of quantum logic associates the projectors of quantum theory with propositions and examines how the logical structure differs from classical logic [11, 65].

In this chapter we describe a program of study in quantum foundations called the *ontological model program*, where the word 'ontological' implies pertinence to physical reality. The motivation is that clarity can be brought to the understanding of quantum theory through the positing of ontological models. Consider the situation before the arrival of Bell's theorem (see Section 2.3.2). Physicists and philosophers argued about whether or not quantum entanglement (see Section 5.1) was a non-classical phenomenon. By positing the existence of an ontological model for quantum theory, and imposing classical restrictions upon it (i.e. local causality in a separable universe), he was able to generate a restriction on possible measurement correlations. Entangled states are able to violate this restriction both theoretically and experimentally, thus concluding the debate on the classicality of entanglement. The study of ontological models should help us to further understand and classify the various features of quantum theory and how they differ from features of classical theories.

In Section 2.1 we present the concept of an *operational theory*, which is thought to capture the notion of any conceivable experimental situation, and we discuss in particular how it captures quantum theory. Section 2.2 presents the ontological model framework, introduces the notion of *outcome determinism* and makes a simple but new distinction between *convex* and *non-convex* ontological models. Section 2.2.3 describes two 'extreme case' ontological models which are adaptable

to any operational theory, and discusses the shortcomings of both approaches. In Sections 2.3.1 and 2.3.2 we present the theorems of Kochen/Specker and Bell and the ramifications of their discovery for ontological models describing quantum theory. In Section 2.4 we describe Bohmian mechanics as an ontological model and discuss its features. We also show a trivial model attributed to Bell. In light of the distinction between convex and non-convex models, we present an example of a non-convex model and analyze why perhaps non-convex models should be avoided.

2.1 Operational Theories

An operational theory is a procedural abstraction of statistics for outcomes associated with various ways of preparing and measuring some system. That is, it is simply a description of a set of experiments that one could perform and the statistics of the outcomes of these experiments. The theories we will be discussing are prepare-and-measure theories. We will be discussing one-shot measurements wherein once a measurement outcome has been obtained, the experiment is concluded. We are generally not interested in state update and hence in the case of quantum theory we focus on the POVM formalism over the more general formalism introduced by Postulate 3. The present discussion of operational theories is derived from recent work [29, 32, 31, 58], but mention of operational theories for studying quantum theory has also been found in older publications [34, 44].

The operational theory begins with a notion of a system $\mathcal{S}$ upon which the experiment is performed. At the level of the operational theory, there is no specification of what $\mathcal{S}$ is or could be. To begin an experiment, $\mathcal{S}$ must be manipulated to be put into any one of a number of possible states. Since the operational theory does not deal with physical reality necessarily, it provides not a listing of possible states $\mathcal{S}$ could be placed in, but a set $\mathcal{P}$ of possible preparation procedures that could be performed on $\mathcal{S}$. Any element $P \in \mathcal{P}$ could be viewed as a list of instructions to be performed on $\mathcal{S}$. Alternatively, it is common to refer to a *preparation device* and view P as a possible setting for the preparation device.

Together with $\mathcal{P}$ comes a set of possible measurements $\mathcal{M}$ that can be performed on $\mathcal{S}$ (perhaps via a *measurement device*). It is assumed that any measurement in $\mathcal{M}$ can be performed on $\mathcal{S}$ no matter what preparation in $\mathcal{P}$ has been performed. Each measurement $M \in \mathcal{M}$ has a possible set of outcomes. We make another simplifying assumption that all measurements in M have as outcomes elements of a global indexing set I. Depending on the operational theory, I could be a finite set, the set of integers or the set of real numbers, for example. The pair $(M, k) \in \mathcal{M} \times I$ is referred to as an effect. To avoid any confusion with the effects of quantum theory, the effects in an operational theory will be referred to as operational effects if the distinction is necessary.

The last piece of an operational theory is a prescription of statistics for the various combinations of preparations, measurements and outcomes. The value

$\Pr(k|P,M)$ is the probability of outcome $k \in I$ given that $\mathcal{S}$ was prepared via $P \in \mathcal{P}$ and measured according to $M \in \mathcal{M}$. Most generally, if I is not a discrete set, we may instead have $\Pr(B|P,M)$ as the probability of an outcome occurring in the measurable set $B \subseteq I$ given preparation P and measurement M. However, we will typically be dealing with discrete outcome sets and our notation will reflect this. One further assumption imposed on the operational theory is a probabilistic consistency requirement:

$$\sum_{k \in I} \Pr(k|M,P) = 1. \tag{2.1}$$

This requirement states that for any preparation and measurement, some outcome $k \in I$ must occur. For every M and P we can view $\Pr(\cdot|M,P)$ as a probability distribution on I.

Definition 2.1 (Operational Theory)**.** An *Operational Theory* is a quadruplet of possible preparations, measurements, outcomes, and probabilities of outcomes for all combinations of preparations and measurements on an abstract system: $(\mathcal{P}, \mathcal{M}, I, \Pr)$.

2.1.1 Convex Operational Theories

If we allow for the experimenter to introduce randomness into the preparations, then the set $\mathcal{P}$ takes on a convex structure.

Definition 2.2 (Mixed Preparations)**.** Suppose that $\vec{p} = (p_1, \ldots, p_n)$ is a probability vector, and $P_1, \ldots, P_n \in \mathcal{P}$. We will denote the preparation whereby a value $i \in \{1, \ldots, n\}$ is sampled from the probability vector $\vec{p}$ and preparation P_i is performed, as the convex combination $p_1 P_1 + \ldots + p_n P_n$ or as the ensemble $\{p_i; P_i\}_{i=1}^n$. For a non-trivial probability vector $\vec{p}$, this is called a *mixed preparation.*

It is clear that given the specification of statistics for the preparations $\{P_1, \ldots, P_n\}$, the statistics for the probabilistic preparation $P' = \{p_i; P_i\}_{i=1}^n$ are already known to be

$$\Pr(k|P',M) = \sum_{i=1}^{n} p_i \Pr(k|P_i,M) \qquad \forall\, (M,k) \in \mathcal{M} \times I. \tag{2.2}$$

If the set $\mathcal{P}$ is convex, that is the specification of preparation procedures allows for all possible mixed preparations, then we make the assumption that the set of extreme points of $\mathcal{P}$ is included in $\mathcal{P}$. The extreme elements of $\mathcal{P}$, denoted $\operatorname{ext} \mathcal{P}$, are the preparations whose statistics cannot be expressed as a convex combination of statistics for other preparations, as in (2.2). We refer to them as *pure preparations.* Note that the mixed preparations are exactly analogous to the proper mixtures of density matrices discussed in Section 1.2.

We speculate that this is not a strong assumption. In particular, we speculate that for any or most convex operational theories, the extreme preparations

are precisely the preparations that would remain if mixed preparations were not allowed.

Similarly to the proper POVM defined for quantum measurements, we can define a concept of a mixture of measurements in an operational theory.

Definition 2.3 (Mixed Measurements)**.** To define a measurement with possible outcomes in $M \subseteq I$, let $\vec{p} = (p_1, \ldots, p_K)$ be a K-dimensional probability vector, $\{M_k\}_{k=1}^K \subset \mathcal{M}$ and let $m(j,k) \in M$ for all $j \in I$ and $k \in 1 \ldots K$. Consider the measurement whereby $k \in \{1, \ldots, K\}$ is chosen according to $\vec{p}$, measurement M_k is then performed, and if outcome j occurs then the measurement outcome is labeled $m(j,k)$. We denote this measurement as $\{p_k, m(j,k); M_k\}_{k=1}^K$. If $\vec{p}$ is non-trivial, we refer to this as a *mixed measurement*.

If such mixed measurements are allowed then we see that this induces a linear structure on the set of effects $\mathcal{M} \times I$. Indeed, for a measurement M' as defined in Definition 2.3, the probability of outcome m is determined to be

$$\begin{aligned} \Pr(m|P, M') &= \sum_{j,k} e_m^{m(j,k)} \Pr(\text{outcome } j|M_k) \Pr(M_k) \\ &= \sum_{j,k} e_m^{m(j,k)} p_k \Pr(j|P, M_k) \qquad \forall\, P \in \mathcal{P}. \end{aligned} \tag{2.3}$$

Thus we could write the effect (M', m) as

$$(M', m) = \sum_{k,j} e_m^{m(j,k)} p_k \, (M_k, j). \tag{2.4}$$

As with the preparations, if $\mathcal{M}$ contains all possible mixed measurements, then we assume that $\mathcal{M} \times I$ contains its extreme points $\text{ext}(\mathcal{M} \times I)$. These are the effects which cannot arise in a mixed measurement. We also define the set of extreme measurements $\text{ext}\,\mathcal{M}$ to be the measurements comprised of extreme effects.

Definition 2.4 (Convex Operational Model)**.** An operational theory is called convex if both the set of preparations and measurements include all possible mixed preparations and measurements, as well as their extreme elements.

2.1.2 A Quantum Example

The most pertinent example of an operational theory is quantum theory itself. In quantum theory, every density matrix ρ corresponds to a preparation, and every positive operator valued measure (POVM) $B = \{E_k\}$ corresponds to a measurement. The quantum effects comprising a POVM are the effects of the operational theory. The statistics are given by the Born rule (1.12):

$$\Pr(k|P_\rho, M_B) = \text{tr}(\rho E_k). \tag{2.5}$$

However, given that operational theories are statistical abstractions of physical experiments, we wish to define operational theories which include only physically implementable subsets of the full set of quantum preparations and measurements. In any particular experiment, one may be restricted in what sorts of quantum systems they have access to.

Definition 2.5. We will define $\mathbb{Q}_d$ to be the convex operational theory representing the preparations and measurements one can perform with access to a single d-level quantum system. Thus the set $\mathcal{P}$ corresponds to the d-by-d density matrices, and $\operatorname{ext}\mathcal{P}$ corresponds to pure states on $\mathbb{C}^d$. The set $\mathcal{M}$ includes all proper d-level POVMs, but *not* the improper d-level POVMs. The extreme effects are the rank-1 projectors and the extreme measurements are the rank-1 PVMs.

It will also be convenient to consider the operational theory representing only the extreme preparations and measurements of $\mathbb{Q}_d$.

Definition 2.6. We define $\partial\mathbb{Q}_d$ to be the non-convex operational theory representing only the pure states and PVMs of a single d-level quantum system.

It is clear that $\mathbb{Q}_d$ is the theory obtained by allowing all mixed preparations and measurements in $\partial\mathbb{Q}_d$.

2.2 Ontological Models for Operational Theories

The operational theory is a very abstract concept. On its own, it provides only the ability to make statistical predictions. The goal of the *ontological model* formalism is to provide an underlying framework which roots an operational theory in a physical reality. This discussion is motivated by the same recent work as our discussion of operational theories, except that we have slightly generalized the exposition to involve probability *measures* instead of *distributions* (see A for a brief exposition of measure theory and integration).

Each element of an operational theory has a corresponding representation in an ontological model. To the abstract system $\mathcal{S}$ of an operational theory an ontological model associates a space of actual physical properties Λ[1]. We will always require Λ to be a measurable space, and so it must come equipped with a σ-algebra Σ, making (Λ, Σ) a measurable space. The set Λ will be referred to as an *ontic space* and the elements $\lambda \in \Lambda$ will be referred to as *ontic states*. The word *ontic* is a philosophical term which pertains to existence or reality. In an ontological model, a preparation P from an operational theory is considered to set the ontic state to a specific element of Λ. However, in general, it is not assumed that P determines an

[1]In a more complete discussion, we would impose certain restrictions on what λcould be. For example, we may require that λbe configuration space or phase space for a classical system, or that it possess certain symmetries.

ontic state with certainty. Most generally, we assume that P induces a probability measure over Λ. Thus we posit a map:

$$\begin{aligned} \mu : \mathcal{P} &\longrightarrow \mathcal{M}_1^+(\Lambda) \\ P &\longrightarrow \mu_P, \end{aligned} \tag{2.6}$$

where $\mathcal{M}_1^+(\Lambda)$ denotes the set of probability measures on (Λ, Σ). For any set $B \in \Sigma$, $\mu_P(B)$ is the probability that the actual ontic state is contained in the set B given that preparation P was performed. Thus any preparation is associated with a set of possible real physical situations, and relative probabilities of occurrence. The measure μ_P is often referred to as an *epistemic state*. The word epistemic is a philosophical term which means 'pertaining to knowledge'. The state μ_P represents our knowledge, possibly imprecise, of the ontic state λ. The support, $\operatorname{supp} \mu_P$ indicates the set of ontic states that are consistent with the preparation P.

A probabilistic interpretation is also given to a measurement in an ontological model. In particular, any effect (M, k) is associated with a fuzzy indicator function[2]over Λ:

$$\begin{aligned} \xi : \mathcal{M} \times I &\longrightarrow [0, \chi_\Lambda] \\ (M, k) &\longrightarrow \xi_{(M,k)}. \end{aligned} \tag{2.7}$$

The notation $[0, \chi_\Lambda]$ denotes the set of functions on Λ taking values in $[0, 1]$ and which are measurable with respect to the measurable space (Λ, Σ). If the ontic state is λ, then $\xi_{(M,k)}$ is the probability of outcome k when measurement M is performed. Thus ξ can be viewed as a conditional probability. Recall that equation (2.1) requires that for any measurement and any preparation, some outcome must occur. This requirement is passed down to the level of the ontic state by requiring that an outcome occurs for every measurement and every ontic state:

$$\forall\, \lambda \in \Lambda, \quad \sum_k \xi_{(M,k)}(\lambda) = 1. \tag{2.8}$$

Given the interpretation of μ and ξ, it is clear how the statistics of the operational theory must be reproduced in an ontological model:

$$\Pr(k|M, P) = \int_\lambda \xi_{(M,k)}(\lambda) d\mu_P(\lambda) \quad \forall (M, k) \in \mathcal{M} \times I, P \in \mathcal{P}. \tag{2.9}$$

Definition 2.7 (Ontological Model for an Operational Theory)**.** An ontological model for an operational theory, $(\mathcal{P}, \mathcal{M}, I, \Pr)$, is a triplet (Λ, μ, ξ) satisfying equations (2.6)(2.7)(2.8) and (2.9). We label the set of all ontological models for operational theories as $\mathbb{O}$.

The above is a minimal definition of an ontological model. As these general models are applied to physical theories we may wish or be forced to discuss further properties of such models, such as convexity.

[2]See Appendix B on classical and fuzzy probability theories.

2.2.1 Convex Ontological Models

An operational theory may possess a convex structure in its preparations and effects 2.1. This structure is motivated by the possibility of performing preparations and measurements involving uncertainty. It seems reasonable that the probabilistic interpretation of a probabilistic preparation or measurement should be passed onto the ontic level. Hence we define the following natural class of ontological models.

Definition 2.8 (Convex Ontological Model)**.** An ontological model for a convex operational theory is convex itself if the maps μ and ξ preserve the structure of mixed preparations and measurements respectively. That is, if the epistemic state for the preparation $\{p_i; P_i\}_{i=1}^n$ is

$$\mu_{\{p_i;P_i\}_{i=1}^n} = \sum_{i=1}^{n} p_i \mu_{P_i} \tag{2.10}$$

and the fuzzy indicator functions for the measurement $M' = \{p_k, m(j,k); M_k\}_{k=1}^K$ are given by

$$\xi_{M',m} = \sum_{k=1}^{K} \sum_{j \in I} e_m^{m(j,k)} p_k \xi_{M_k,j}. \tag{2.11}$$

The set of convex ontological models is denoted as $\mathbb{O}_{conv}$.

A non-convex ontological model for a convex operational theory results in seemingly strange physical and causal properties. Nevertheless, non-convexity is mathematically consistent with the above Definition 2.7 of an ontological model. We will demonstrate and discuss a specific example of a non-convex model for quantum theory at the end of this chapter (see Section 2.4.3).

2.2.2 Outcome Determinism

The 'measurement problem' [43] of quantum theory is rooted in the inherent indeterminism of outcomes that comes with the assumption that quantum theory is complete and describes an objective physical reality. Such a viewpoint inexorably leads one to consider the possibility that macroscopic indicators of outcomes may actually exist as some physical mixture (see the start of Chapter 3 for a more formal discussion). The most common example of such a thought experiment is Schrödinger's cat who at some point in an experiment must be both alive and dead. In much of the literature on ontological models, it is assumed that a successful ontological model should be fundamentally deterministic. Given the ontological model framework defined in Section 2.2, the only way to account for indeterminism while staying rooted in a deterministic reality is to require that the indicator functions be idempotent ($\xi^2 = \xi$). Such a property assures us that if the actual state of the system ($\lambda \in \Lambda$) were to be known, then the outcome of any measurement would be known.

However, we can quickly prove that it is generally nonsensical to require this property for every measurement effect.

Proposition 2.1. *In $(\Lambda, \mu, \xi) \in \mathbb{O}_{conv}$, there exists an effect $(M', k') \in (M, k)$ of a mixed measurement for which $\xi_{M',k'}$ is not idempotent.*

Proof. By (2.11) the effects for mixed measurement $\{p_k, m(j,k); M_k\}_{k=1}^K$ are given by $\xi_{M',m} = \sum_{j\in I, k=1}^K e_m^{m(j,k)} p_k \xi_{M_k,j}$. If say $m(j,k) = k$, then $\xi_{M',m} = \sum_{j\in I} p_m \xi_{M_m,j} = p_m$. As long $\vec{p}$ is non-trivial, then $\xi_{M',m}$ is not idempotent. □

This suggests the definition:

Definition 2.9 (Outcome Determinism)**.** We say that an ontological model is *outcome deterministic* if when $(M,k) \in \text{ext}(\mathcal{M} \times I)$ then the indicator function $\xi_{(M,k)}$ is idempotent.

2.2.3 Fuzzy and Unfuzzy Quantum Models

Appendix B describes the frameworks of classical and fuzzy probability theory. Any operational theory may be represented in either of these frameworks, although these representations will have different characteristics. If an operational theory is always indeterministic, then a fuzzy probability representation is necessary if one wishes the ontic state to be precisely analogous to a *pure preparation* (e.g. the pure states in quantum theory). Also, a classical probability theory is equivalent to an outcome deterministic ontological model. Thus, in general, outcome determinism can only be achieved in a model where the pure preparations and the ontic state are not identified with each other.

Kochen and Specker's Unsatisfactory Model

Kochen and Specker provide an abstract but fully general demonstration of an outcome deterministic model for any operational theory [40]. Firstly, the ontic space is taken to be the set of all I-valued functions on the space of measurements,

$$\Lambda = I^{\mathcal{M}} = \{\lambda \mid \lambda : \mathcal{M} \to I\}.$$

For a given effect (M,k), we define the indicator function as:

$$\xi_{M,k}(\lambda) = \chi_{S_{M,k}}(\lambda), \quad \text{where} \quad S_{M,k} = \{\lambda \mid \lambda(M) = k\}.$$

This indicator function picks out all the states (which are also functions) which assign value k to measurement M. We now need to define the preparation measures such that the measure given to the set $S_{M,k}$ when the preparation is P is precisely $\Pr(k|P,M)$. This can be accomplished by taking μ_P to be a product measure,

$$\prod_{M'\in\mathcal{M}} \Pr(\cdot|P,M').$$

With this measure, the set $S_{M,k}$ is assigned the value $\Pr(k|P,M)$. To see this explicitly, we ask: for each $M' \in \mathcal{M}$, what values in I do the functions in $S_{M,k}$ take on? Call this set $U_{M'}$. For any $M' \neq M$, we have $U_{M'} = I$, and for M we have $U_M = k$. Thus μ_P assigns the value

$$\prod_{M' \in \mathcal{M}} \Pr(U_{M'}|P,M) = \prod_{M' \neq M} \Pr(I|M,P) \times \Pr(k|M,P) = \Pr(k|M,P). \tag{2.12}$$

The problem with this model, as is pointed out by Kochen and Specker [40], is that all the effects from distinct measurements are statistically independent.

Suppose we fix the preparation P and consider calculating the probability $\Pr(k_1|P, M_1 \,\&\, k_2|P, M_2)$. We simply have to determine the measure of the set $S = \{\lambda \mid \lambda(M_1) = k_1 \;\&\; \lambda(M_2) = k_2\} = S_{M_1,k_1} \cap S_{M_2,k_2}$. But similarly to how we determined (2.12), we see that the probability assigned to S by μ_P is $\Pr(k_1|P,M_1) \times \Pr(k_2|P,M_2)$. Thus all effects are independent. The issue with independent effects is that some structure or intuition of a physical theory becomes irrelevant.

For example, in quantum theory consider the positive observable A and the observable A^2. Given a quantum state ψ, the distribution of outcomes for A and A^2 are identical, except that the outcomes for A^2 are the squares of the outcomes for A. In fact, A^2 can be implemented by performing the experiment corresponding to observable A and then squaring the outcome. But, if we model quantum theory using the above model, the implication is that if the system is in the ontic state λ, which assigns value a to observable A, then since effects for A and A^2 are all completely independent, the ontic state λ could easily result in a value for A^2 which is not a^2. In Section 2.3.1 we will discuss how Kochen and Specker hoped the functional relationships between observables of quantum theory would be respected by an ontological model.

Aside from the independence issue, the above model also appears to be much too abstract to explain a physical theory with any level of satisfaction. We will find that this is not necessarily the case for all outcome deterministic models when we discuss the de Broglie-Bohm model for quantum theory (see Section 2.4.1).

The Beltrametti-Bugajski Model

The following discussion is in the context of quantum theory but can be easily adapted to accommodate any operational theory.

The notion of fuzzy probability theory (see B.2) is necessary for anyone who wants to take seriously the notion that the quantum pure state ψ is real and completely describes everything physical about the system under investigation. For those with this view, the ontic state space would have to be $\Lambda = \mathbb{P}\mathcal{H}$, and the epistemic states correspond to delta distributions on $\mathbb{P}\mathcal{H}$:

$$\mu_\psi(\lambda) = \delta(\lambda - \psi).$$

Consequently, the measurements of the theory are fuzzy random variables (i.e. the effects are represented by fuzzy indicator functions). Given the POVM $E = \{E_m\}$, the corresponding indicator functions (Markov kernels in the fuzzy probability language) are given by

$$\xi_{E,m}(\lambda) = \mathrm{tr}(\lambda E_m). \tag{2.13}$$

Then the statistics of quantum theory are trivially reproduced:

$$\int_{\mathbb{P}\mathcal{H}} \xi_{E,m}(\lambda)\delta(\lambda - \psi)d\lambda = \xi_{E,m}(\psi) = \mathrm{tr}(\psi E_m). \tag{2.14}$$

In quantum theory, for every effect E_m aside from the identity matrix, there are quantum states such that $0 \leq \mathrm{tr}(\psi E_m) \leq 1$, hence $\xi_{E,m}(\lambda)$ is always a fuzzy indicator function. This model is often referred to as the Beltrametti-Bugajski model [9]. Note that in the fuzzy probability language, the indicator functions $\xi_{E,m}(\lambda)$ for measurement E are equivalent to the *Markov kernel* $K_E(\lambda, m)$.

The problems with such a model are the problems that have been plaguing all who have attempted to view the quantum state as 'the whole story'. Most prominent is the measurement problem or Schrödinger's cat paradox [43] (see the opening of Chapter 3).

2.3 Two Major Restrictions on Models for Quantum Theory

The theorems of Kochen/Specker and Bell have had a significant impact on the allowed properties of any ontological model for quantum theory. In Section 2.3.1 we describe the Kochen-Specker theorem and how it restricts the mathematical objects in quantum theory that we can possibly view as representing pre-existing properties of a system. In Section 2.3.2 we present Bell's theorem and the necessity of non-locality. Section 2.3.3 discusses the relationship between these two results.

2.3.1 Kochen-Specker Theorem and Contextuality

In Section 2.2.3 we presented a general ontological model for any given operational theory. Kochen and Specker found this model in that all effects were statistically independent of each other. It is argued in [40] that an ontological model should preserve the functional relationships between the observables of quantum theory. It is important to note that in their discussion of quantum mechanics and ontological models, Kochen and Specker consider only outcome deterministic models for $\partial\mathbb{Q}_d$ (pure states and PVMs on $\mathbb{C}^d$).

The basic requirements that Kochen and Specker impose on an ontological model for quantum theory are:

- a pure quantum state ψ should be mapped to a probability measure μ_ψ on a measurable space Λ;
- every observable A with eigenvalues $\{a_i\}_{i=1}^d$ and corresponding eigenspace projectors $\{P_i\}_{i=1}^d$ should be associated with a function $f_A : \Lambda \to \mathbb{R}$ where $f_A(\Lambda) = \{a_i\}_{i=1}^d$;
- μ_ψ and f_A should reproduce the statistics of the PVM corresponding to A when the quantum state is ψ:

$$\mu_\psi(f_A^{-1}(a_i)) = \Pr(a_i|\psi, A) = \mathrm{tr}(\psi P_i). \tag{2.15}$$

Given that the codomain of f_A is the spectrum of A, we can interpret the function f_A as assigning values or properties to the ontic states. We interpret $f_A(\lambda) = a_i$ to mean that the ontic state λ has value a_i for observable A. We can show the above set of requirements is equivalent to the requirement of an ontological model which is outcome deterministic (see Definition 2.5 and Definition 2.9).

Define the set $\Lambda_{A,a_i} = f_A^{-1}(a_i) = \{\lambda \in \Lambda | f_A(\lambda) = a_i\}$ i.e the set of ontic states which have value a_i for the observable A. Then disjoint idempotent indicator functions can be defined:

$$\chi_{A,a_i}(\lambda) = \begin{cases} 1 & \text{if } \lambda \in \Lambda_{A,a_i} \\ 0 & \text{else} \end{cases}. \tag{2.16}$$

Since

$$f_A = \sum_{i=1}^d a_i \chi_{A,a_i}, \tag{2.17}$$

a set of disjoint idempotent indicator functions for the projectors $\{P_k\}_{k=1}^d$ exists if and only if f_A exists. Furthermore, we have that

$$\mu_\psi(f_A^{-1}(a_i)) = \int_\Lambda \chi_{A,a_i}(\lambda) d\mu_\psi(\lambda). \tag{2.18}$$

Therefore f_A will reproduce the statistics in (2.15) if and only if χ_{A,a_i} reproduces the probability for outcome i for the measurement corresponding to A. Thus we have our stated equivalence.

In addition to the above basic conditions, Kochen and Specker also require that the functions f_A maintain the functional relationships between observables:

- for all observables A and all Borel functions g[3],

$$f_{g(A)} = g(f_A). \tag{2.19}$$

For two observables A and B, there exists a g such that $B = g(A)$ if and only if A and B commute. We can understand the action of g as the action of a function on the eigenvalues. That is, the function g on $\mathcal{L}_{sa}(\mathbb{C}^d)$ is given by the function g on $\mathbb{R}$ such that:

$$g(A) = g(\sum_{i=1}^{d} a_i \,|\, \psi_i \rangle\langle\, \psi_i \,|) = \sum_{i=1}^{d} g(a_i) \,|\, \psi_i \rangle\langle\, \psi_i \,| \,. \tag{2.20}$$

We can imagine the ramifications of requirement (2.19) with respect to physical experiments. Suppose that $B = g(A)$, and the state ψ is prepared, inducing the probability measure μ_ψ over the ontic space, and putting the system into some actual ontic state λ. A measurement of observable B will yield $f_B(\lambda)$, and a measurement of observable A will yield $f_A(\lambda)$. However, if (2.19) holds, we have that $g(f_A(\lambda)) = f_B$. Thus we can measure A, plug the output into g, and the result will be identical to that of having measured B instead.

To give an intuitive idea of the term *contextuality*, the observable B could be measured by measuring A, or by measuring any other observable C such that $g'(C) = B$ for some Borel function g'. Thus A and C are two *contexts* for measuring B. Moreover, it may even be that A and C do not commute, such that one could not measure A by measuring C, or vice-versa. The assumption (2.19) implies that the measurement outcomes are independent of the context.

Kochen and Specker were able to prove that outcome deterministic ontological models for $\partial\mathbb{Q}_d$ satisfying (2.19) are impossible by first reducing (2.19) to a requirement on *prediction functions*. A prediction function is a mapping $h : \mathcal{L}_{sa}(\mathbb{C}^d) \to \mathbb{R}$ predicting the outcome for any given observable. If an ontological model satisfying (2.15) exists, then each ontic state λ induces a prediction function via $h_\lambda(A) = f_A(\lambda)$.

Proposition 2.2. *In an outcome deterministic ontological model, the satisfaction of* (2.19) *is equivalent to the existence of functions $f_P : \Lambda \to \{0, 1\}$ for every projector $P \in \mathbb{CP}^{d-1}$ such that*

$$f_A = \sum_{i=1}^{d} a_i f_{P_i}. \tag{2.21}$$

where

$$A = \sum_{i=1}^{d} a_i P_i. \tag{2.22}$$

Furthermore, for any complete set of orthogonal projectors $\{P_i\}_{i=1}^{d}$ and for all $\lambda \in \Lambda$ it must be the case that $f_{P_i}(\lambda) = 1$ for exactly one $i \in \{1, \ldots, d\}$.

Proof. Suppose (2.21) is true. Then for any Borel function g,

$$g(f_A) = g\left(\sum_{i=1}^{d} a_i f_{P_i}\right) = \sum_{i=1}^{d} g(a_i) f_{P_i} = f_{g(A)}. \tag{2.23}$$

[3]See Appendix A

Conversely, let $A, B \in \mathcal{L}_{sa}(\mathbb{C}^d)$ be any two observables which both contain the projector P in their spectral decomposition, with eigenvalues a and b for P respectively. We can consider P itself as an observable with eigenvalues 0 and 1. Given the ability to measure the observable A, one could measure P by measuring A and outputting 1 if outcome a is obtained and 0 otherwise. Similarly, one could measure P by measuring B. This indicates that there are Borel functions g_1 and g_2 such that $g_1(A) = g_2(B) = P$. In particular

$$g_1(x) = \begin{cases} 1 & x = a \\ 0 & x \in Sp(A) \backslash a \end{cases} \qquad g_2(x) = \begin{cases} 1 & x = b \\ 0 & x \in Sp(B) \backslash b \end{cases} \tag{2.24}$$

where $Sp()$ denotes the spectrum of an operator. By (2.19) it must be the case that $f_P = g_1(f_A) = g_2(f_B)$. Therefore (2.24) implies that $f_P = \chi_{A,a} = \chi_{B,b}$. Finally, (2.17) implies that f_A and f_B must have the form (2.21).

Now let $\mathcal{B} = \{P_i\}_{i=1}^d$ be a complete set of orthogonal projectors, and let A be an observable having $\mathcal{B}$ forming its spectral decomposition and its eigenvalues $\{a_i\}_{i=1}^d$ having the property that no one eigenvalue is the sum of any other set of eigenvalues. Since $f_A = \sum_{i=1}^d a_i f_{P_i}$, given the stated property of the eigenvalues, it must be the case that $f_{P_i}(\lambda) = 1$ for exactly one $i \in \{1, \ldots, d\}$ for every $\lambda \in \Lambda$. □

Consider a predictor function $h : \mathcal{L}_{sa}(\mathbb{C}^d) \to \mathbb{R}$ restricted to the set of projectors. The above theorem indicates that h must be such that

$$\begin{gathered} \forall\ \{P_i\}_{i=1}^d \subset \mathcal{L}_{sa}(\mathbb{C}^d) \text{ such that } P_i P_j = P_i \delta_{i,j} \text{ and } \sum_{i=1}^d P_i = I, \\ h(P_i) \in \{0, 1\} \text{ and } \sum_{i=1}^d h(P_i) = 1. \end{gathered} \tag{2.25}$$

In words, only one of any complete orthogonal set of projectors may be assigned the value 1 by a prediction function h. All others must be assigned the value 0.

Theorem 2.1 (Kochen-Specker Theorem)**.** *For $d > 2$, there exists a finite set of orthonormal bases $\{B_i\}$ for $\mathbb{C}^3$ such that no prediction function $h : \mathbb{CP}^{d-1} \to \{0, 1\}$ exists satisfying* (2.25) *for all B_i.*

Proof Sketch. The original proof by Kochen and Specker [40] involves a careful selection of 117 unit vectors in $\mathbb{R}^3$ (and hence in $\mathbb{C}^3$ and hence corresponding to elements of $\mathbb{CP}^2$). Various triplets of these vectors are orthogonal to each other and hence correspond to complete sets of orthogonal projectors. It is then shown that 1s and 0s cannot be assigned to the 117 vectors without violating (2.25). □

Thus the hope of a non-contextual assignment of outcomes for all measurements is dashed, as long as the Hilbert space is of dimension 3 or higher. Any outcome deterministic ontological model for quantum theory ($d \geq 3$) will be forced to assign

different indicator functions to projectors, depending on which PVM the projectors are in.

To explicitly see the consequences of this result, suppose that we have two orthonormal bases for $\mathbb{C}^3$:

$$\begin{aligned} B_1 &= \{|\,\psi_1\,\rangle\,, |\,\psi_2\,\rangle\,, |\,\psi_3\,\rangle\} \\ B_2 &= \{|\,\psi_1\,\rangle\,, |\,\psi_{2'}\,\rangle\,, |\,\psi_{3'}\,\rangle\}\,. \end{aligned} \tag{2.26}$$

These specify two different PVMs, both of which contain the projector $|\,\psi_1\,\rangle\langle\,\psi_1\,|$. Now let (Λ, μ, ξ) be any outcome deterministic ontological model for $\partial\mathbb{Q}_3$. The Kochen-Specker result states that for an appropriately chosen $\psi_{2'}$ and $\psi_{3'}$, it must be the case that $\xi_{B_1,1} \neq \xi_{B_2,1}$, despite the fact that they both represent the same effect.

On a finer level, there must be some ontic state λ such that $\xi_{B_1,1}(\lambda) = 1$ and $\xi_{B_2,1}(\lambda) = 0$. Thus we conclude that the effect $|\,\psi_1\,\rangle\langle\,\psi_1\,|$ does not represent a pre-existing property[4]of the system in state λ. At most, what can be said is that the effect $|\,\psi_1\,\rangle\langle\,\psi_1\,|$ *in the context of* the PVM B_1 *might be* a pre-existing property of the system, and the effect $|\,\psi_1\,\rangle\langle\,\psi_1\,|$ *in the context of* the PVM B_2 *might be* a pre-existing property of the system.

We will see that in Bohmian mechanics, even this is not true. An effect within the context of a certain PVM does not represent a pre-existing property, as in Bohmian mechanics, the physical implementation of a PVM also has an effect on the indicator functions for the measurements.

Improved proofs of contextuality for quantum theory, involving a smaller numbers of vectors and bases, exist and are presented in [50].

2.3.2 Bell's Theorem

One of the two postulates of Einstein's theory of special relativity is the constancy of the speed of light. The speed of light is considered to be an upper bound on the propagation speed of anything physical, and thus stipulates a limitation on the regions of possible future and past influence (light cones) for any point in space and time. A famous theorem due to Bell [6, 7] implies that quantum theory can have influences that act instantaneously in regions that lay outside of these light cones, and hence implies that any realistic elements (ie ontic properties) attributed to quantum theory have causal influence outside of these limits. This flies in the face of classical intuition.

The EPR-Bell Experiment

In their 1935 paper [19], Einstein, Podolsky and Rosen (EPR) discussed a thought experiment of the following sort as a motivation for the belief that quantum theory

[4]We use the term *pre-existing property* to mean a property which is definitely possessed by the system.

is 'incomplete'. Bell revised the thought experiment into the form presented here.

A bipartite quantum system is prepared in the 'singlet state'

$$|\,\psi_{AB}\,\rangle = \frac{1}{\sqrt{2}}\left(|\,01\,\rangle - |\,10\,\rangle\right) \tag{2.27}$$

and the two qubit subsystems, labeled A and B, are then separated and propagated in opposite directions, while maintaining the state (2.27) between them. At some unspecified distance, each subsystem is subjected to a Stern-Gerlach measurement. The A magnet is oriented in the $\vec{a}$ direction while the B magnet is oriented in the $\vec{b}$ direction. Recall that for a qubit system a directional choice is equivalent to choosing a PVM, or a measurement in certain basis (see Section 1.3).

If a PVM is performed on subsystem A, leaving it in a particular pure state $|\,\psi\,\rangle$, then subsystem B is definitely in the pure state $\left|\,\psi^{\perp}\,\right\rangle$ orthogonal to $|\,\psi\,\rangle$. To see this, let $\left\{|\,\psi\,\rangle\,, \left|\,\psi^{\perp}\,\right\rangle\right\}$ be any orthonormal basis for $\mathbb{C}^2$. Thus there exist complex numbers a and b such that

$$|\,0\,\rangle = a\,|\,\psi\,\rangle + b\left|\,\psi^{\perp}\,\right\rangle \quad \text{and} \quad |\,1\,\rangle = \bar{b}\,|\,\psi\,\rangle - \bar{a}\left|\,\psi^{\perp}\,\right\rangle, \tag{2.28}$$

where $\bar{a}$ denotes the complex conjugate of a. Substituting these decompositions into (2.27) gives

$$|\,\psi_{AB}\,\rangle = \frac{1}{\sqrt{2}}\left(\left|\,\psi\psi^{\perp}\,\right\rangle - \left|\,\psi^{\perp}\psi\,\right\rangle\right). \tag{2.29}$$

Then suppose that the A subsystem of $|\,\psi_{AB}\,\rangle$ is measured with respect to the PVM corresponding to the states ψ and $\psi^{\perp}$. The measurement update rule (1.23) dictates that if the outcome of the measurement is ψ, then the bipartite system will be left in the state $\left|\,\psi\psi^{\perp}\,\right\rangle$ and if the outcome is $\psi^{\perp}$, then the bipartite system will be left in the state $\left|\,\psi^{\perp}\psi\,\right\rangle$.

For one who take seriously the idea that the quantum state is real and is a complete description of a physical system, the discussion of the EPR experiment leads one to conclude that quantum theory must be non-local: if measurement A is performed before measurement B, we see that the state is updated instantaneously at a space-like separated distance. For those wishing to hold onto their classical intuition of locality, the logical conclusion is that quantum theory must be incomplete, i.e. it is an operational theory for which an ontological model over a more detailed ontology exists.

Any Ontological Model for Quantum Theory is Non-Local

Despite the hope that a more detailed ontology for quantum theory would bypass the apparent non-locality issue, Bell proved that this is not the case [6, 7].

Firstly, if measurement directions $\vec{a}$ and $\vec{b}$ are equal (both systems are measured in the same basis), then the measurement outcomes will be perfectly anti-correlated.

Moreover, if we assign the values ± 1 to the outcomes of Stern-Gerlach measurements, this implies that the expectation value of the joint experiment where A is measured in the direction $\vec{a}$ and B in the direction $\vec{b}$ is [6]

$$\langle \vec{a}\vec{b} \rangle = -\vec{a} \cdot \vec{b}. \tag{2.30}$$

Consider an experiment of the EPR sort, where the bipartite state (2.29) is prepared, and the two subsystems are propagated in opposite directions. We will posit an ontological model (Λ, μ, ξ) for $\mathbb{Q}_2$. Suppose that the preparation of the state (2.29) induces the measure μ on Λ. If a measurement in direction $\vec{a}$ is performed on subsystem A and a measurement in direction $\vec{b}$ is performed on subsystem B, then this constitutes a rank-1 PVM performed on the state of the whole system. Thus we have the four indicator functions $\xi_{\vec{a},\vec{b},a,b}$ where a and b denote the outcomes of the measurements on each subsystem, which we label as ± 1. In general we have

$$\Pr(a, b|\vec{a}, \vec{b}, |\, \psi_{AB}\, \rangle) = \int_\Lambda \xi_{\vec{a},\vec{b},a,b}(\lambda) d\mu(\lambda), \tag{2.31}$$

and hence the expectation value of the product of the two outcomes is

$$\langle \vec{a}\vec{b} \rangle == \int_\Lambda \Big(\xi_{\vec{a},\vec{b},1,1}(\lambda) + \xi_{\vec{a},\vec{b},-1,-1}(\lambda) - \xi_{\vec{a},\vec{b},1,-1}(\lambda) - \xi_{\vec{a},\vec{b},-1,1}(\lambda) \Big)\, d\mu(\lambda). \tag{2.32}$$

The assumption of locality implies that the outcome of the measurement for subsystem A is independent of the measurement choice at subsystem B and vice-versa. Recall that the indicator function $\xi_{\vec{a},\vec{b},a,b}(\lambda)$ can be interpreted as the probability of outcome (a, b) given the measurement $(\vec{a}, \vec{b})$ and the ontic state λ:

$$\xi_{\vec{a},\vec{b},a,b}(\lambda) = \Pr(a, b \mid \vec{a}, \vec{b}, \lambda). \tag{2.33}$$

Thus the locality assumption implies:

$$\Pr(a, b|\vec{a}, \vec{b}, \lambda) = \Pr(a \mid \vec{a}, \lambda) \Pr(b \mid \vec{b}, \lambda) = \xi_{\vec{a},a}(\lambda) \xi_{\vec{b},b}(\lambda) \tag{2.34}$$

for some indicator functions $\xi_{\vec{a},a}$ and $\xi_{\vec{b},b}$.

Now suppose that experimenters at the two ends of the experiments each independently choose two measurement directions: $\vec{a}_1, \vec{a}_2, \vec{b}_1, \vec{b}_2$. As before, we denote, for example, the expectation value of measurement $\vec{a}_1$ on subsystem A and measurement $\vec{b}_1$ on subsystem B by $\langle \vec{a}_1 \vec{b}_1 \rangle$. In the ontological model framework, these expectations are again given by an equation of the form (2.32). If the indicator functions corresponding to these measurements are *local*, i.e. given in the form (2.34), then one can show the following inequality [7, 16]:

$$\Big| \langle \vec{a}_1 \vec{b}_1 \rangle + \langle \vec{a}_2 \vec{b}_1 \rangle + \langle \vec{a}_2 \vec{b}_2 \rangle - \vec{a}_1 \vec{b}_2 \rangle \Big| \leq 2. \tag{2.35}$$

This is called the CHSH inequality, named after those who derived it: Clauser, Horne, Shimony and Holt [16]. It is a generalization of Bell's original inequality,

presented in [6]. One can then simply show that quantum theory can violate this inequality in the case of the EPR singlet experiment.

Suppose A measures in directions $\vec{a}_1 = (0, 0, 1)$, $\vec{a}_2 = (0, 1, 0)$ and B measures in directions $\vec{b}_1 = \frac{1}{\sqrt{2}}(0, 1, 1)$, $\vec{b}_2 = \frac{1}{\sqrt{2}}(0, 1, -1)$. Then by (2.30), the left-hand side of (2.35) evaluates to $2\sqrt{2} > 2$. Thus quantum theory violates the locality inequality (2.35).

There are two ways in which an ontological model can achieve non-locality [38, 58]. Firstly, it may have an ontology which is non-separable. Loosely speaking, this means that the ontic state λ cannot be localized to some particular space-time region. This is precisely why orthodox quantum theory i.e. the Beltrametti-Bugajski model (Section 2.2.3), is non-local. If two particles are jointly in the singlet state (2.29) and space-like separated, there are no two quantum states for the subsystems which combine to give the same description that (2.29) does. If an ontological model is in fact separable, then, as will be discussed subsequently, it can achieve non-locality by being contextual.

2.3.3 Non-Locality as Contextuality

We saw in Section 2.3.1 how contextuality implies the inability to assign outcomes to projector P independently of other projectors measured with P in a PVM. The proof of contextuality involved an explicit attempt to assign outcomes, and a demonstration of the impossibility of the task.

Another flavour of contextuality proof considers two subsystems of a larger system, and then assumes that a choice of measurement on one subsystem does not affect the outcome of a measurement on the other subsystem, and vice-versa. In particular, a PVM for the whole system would be comprised of two PVMs, one for each subsystem. If we assume the outcomes of the PVM for one subsystem are independent of the choice of PVM on the other subsystem, this is again the non-contextuality assumption. However, in this case it can be motivated by locality. Belief in locality would allow one to assume that it was *impossible* for the outcomes of one PVM to be dependent on the choice of a distant PVM.

Adhering to such an assumption and drawing it to a logical conclusion leads to the Bell/CHSH inequalities. Thus derivation of the Bell/CHSH inequality and demonstrating that quantum theory can violate these is another proof of contextuality. However, the Bell/CHSH inequality does not require an assumption of outcome determinism. Thus this proof of contextuality implies that even in an outcome indeterministic model, if the model is separable, then the indicator functions associated to a projector are dependent on the PVM. This is a different sort of result than the Kochen-Specker result which speaks only to outcome deterministic results. The *generalized contextuality* presented in Section 4.1 will define a framework such that both the Kochen-Specker Theorem and Bell's Theorem imply a form of contextuality.

2.4 Example Ontological Models

To conclude this chapter, we briefly discuss three ontological models for quantum theory and their properties.

2.4.1 Bohmian Mechanics

The most successful (arguably) ontological model to date was presented by Louis de Broglie at the 1927 Solvay conference [17] and was further developed by David Bohm in 1952 [12, 13]. The theory, referred to as Bohmian mechanics, de Broglie-Bohm theory, or the causal interpretation, grounds quantum mechanics in a physical, dynamic and deterministic reality by positing as a 'hidden variable' the actual positions of the particles that quantum mechanics predicts the statistics of outcomes for.

Postulates of Bohmian Mechanics

The following discussion on Bohmian mechanics comes largely from the compreshensive text by Holland [35], titled 'The Quantum Theory of Motion'. This presentation is largely a pedagogical exposition on Bohmian mechanics where all the mathematical statements are backed up by rigour and calculation presented in [35].

Holland lists the following as the postulates of Bohmian mechanics for a single system:

1. An individual physical system comprises a wave propagating in space and time together with a point particle which moves continuously under the guidance of the wave.

2. The wave is mathematically described by $\psi(\mathbf{x}, t)$, a solution to Schrödinger's wave equation.

3. The particle motion is obtained as the solution $\mathbf{x}(t)$ to the equation

$$\dot{\mathbf{x}} = (1/m)\nabla S(\mathbf{x}, t)|_{\mathbf{x}=\mathbf{x}(t)} \tag{2.36}$$

   where S is the phase of ψ [and m is the mass of the particle]. To solve this equation we have to specify the initial condition $\mathbf{x}(0) = \mathbf{x}_0$. This specification constitutes the only extra information introduced by the theory that is not contained in $\psi(\mathbf{x}, t)$ (the initial velocity is fixed once we know S). An ensemble of possible motions associated with same wave is generated by varying $\mathbf{x}_0$.

In the above second postulate, the wave ψ is essentially the same object as was defined in Postulate 1. The statement of Postulate 1 presented in this thesis suppresses the possible spatial dependence of ψ. The second postulate above

also makes reference to Schrödinger's equation, which was not mentioned in Section 1.1. The Schrödinger equation is an equivalent way of stating Postulate 2 for the evolution of the quantum system, and takes the form

$$i\hbar\frac{d\psi}{dt} = H\psi, \tag{2.37}$$

where H is the Hamiltonian of the isolated system.

The phase S of the wave-function, mentioned in the third postulate, is simply the phase of ψ which comes about when writing the complex wave-function in phase-amplitude form: $\psi = Re^{iS}$. The third postulate stipulates that Bohmian mechanics is not simply a prescription for calculating probabilities of outcomes of future measurements. Given a specified wave-function and initial position, (2.36) gives the particle a definite and deterministic trajectory. Since all measurements in Bohmian mechanics are essentially position measurements, the initial position and the wave-function ψ are enough to predict with certainty the outcome of any future measurement. Thus Bohmian mechanics is able to provide a causal and dynamical explanation for the statistics of quantum theory. Through (2.36) ∇S associates with each space-time coordinate a unique tangent vector. Thus the trajectories in space-time are non-intersecting.

Bohmian Mechanics as an Ontological Model

In Bohmian mechanics, it is clear that the position of the particle is treated as a real property, which has a definite and deterministic value at all times. Thus $\mathbb{R}^3$ is taken as part of the ontic space. However, the wave-function ψ plays a physical role as well in that it determines the trajectory of the particle through its phase. All measurements at some point boil down to a measurement of position, and thus anything that 'guides' the position of particles must be taken to be ontic. Thus the ontic space is $\Lambda = \mathbb{R}^3 \times \mathbb{P}\mathcal{H}$. We will denote the ontic variable as $\lambda = (\lambda_x, \lambda_\psi)$[5]. The postulates above list the initial particle position $\mathbf{x}_0$ as an extra parameter which needs to be specified for the theory to be deterministic, once ψ is given. Quantum theory is indeterministic, and thus an initial distribution of particle positions needs to be specified in order for Bohmian mechanics to agree with quantum theory, and moreover, to be put into the ontological model framework. Thus an additional assumption [35] is that an ensemble of particles prepared with quantum state $\psi(x)$ will be distributed according to $|\psi(\mathbf{x}_0, t)|^2$.

Generally, if the ignorance of the wave-function can be described by the ensemble $\sigma = \{p_i; \psi_i\}$, then the distribution over the ontic space is

$$\mu_\sigma(\lambda|\sigma) = \sum_i p_i |\psi(\lambda_x, t)|^2 \delta(\psi - \lambda_\psi). \tag{2.38}$$

[5]Not all who study Bohmian mechanics consider ψ to be ontological, but more ''law-like' in the sense that it dictates evolution, much like Newton's laws dictate evolution [18].

In Bohmian mechanics, the theory of measurement is incorporated directly into the formalism[6]. A measurement is carried out by having the system of interest interact with a measurement 'device'. The measurement device is idealized to be a particle localized around some point in its own configuration space. The interaction correlates the value of the observable in question with the position of the measurement particle, thus the value can be inferred by determining the position of the measurement particle after the interaction is completed.

Consider $\hat{A}$, an observable which is a function of the position and momentum observables $\hat{\mathbf{x}}$ and $\hat{\mathbf{p}}$. Since Bohmian mechanics postulates that the particle does have a definite position and momentum at all times, the particle does indeed possess a value for $\hat{A}$ at all times. This value can be determined by evaluating

$$A(\mathbf{x}, t) = \frac{Re(\psi^*(\hat{A}\psi)(\mathbf{x}, t))}{|\psi(\mathbf{x}, t)|^2}. \tag{2.39}$$

at the position and time of interest.

Suppose we are interested in measuring the observable $\hat{A}$, beginning at time t_0. Suppose further that at t_0, $\mathbf{x} = \mathbf{x}_0$ and $\psi = \psi_0 = \sum_a c_a \psi_a$, where the ψ_a are eigenvectors of the observable $\hat{A}$. We consider the measurement system wave-function to be described by a probability distribution $\phi(y, t_0)$ localized around a point y_0. Thus the wave-function of the whole system at time t_0 is $\psi_0(\mathbf{x}, t_0)\phi(y, t_0)$.

We then evolve the whole system under the Hamiltonian $H = g\hat{A}\hat{p}_y$, where p_y is the momentum conjugate to the position y of the measurement system. It is assumed that the interaction is an impulse, and will be sufficiently strong and short that the free evolution of either system is negligible during the interaction. If this is done successfully, then after an interaction of duration T, the wave-function of the whole system will have evolved into $\psi_f = \sum_a c_a \psi_a \phi(y - gaT)$. If the initial ϕ is sufficiently localized, then the position of the measurement system will be very highly correlated with a specific eigenfunction of the observable $\hat{A}$. Thus if we find the measurement system to be located near $y_0 + gaT$, then we infer that ψ_f is comprised (nearly) entirely of ψ_a, and thus the particle itself must possess the value a for observable A. However, this does not imply that the particle possessed the value a prior to measurement, but the process of the measurement has caused the particle to evolve such that it now has the value a.

If we know the initial ontic states of the system ($\lambda_x = \mathbf{x}_0, \lambda_\psi = \psi_0$), then we can determine how they will evolve under the interaction H with the measurement system and hence we can calculate not only ψ_f but $\mathbf{x}_f$ as well. We can then plug both into (2.39) to determine the value of $\hat{A}$. Thus it is clear that we can write down idempotent indicator functions over Λ for the measurement of observable $\hat{A}$. In particular, let $\mathbf{x}_f(\lambda_x)$ and $\psi_f(\lambda_\psi)$ be the final position and wave-function of the system having started in the state $(\lambda_x, \lambda_\psi)$ and evolved under H with a

[6]The following discussion of measurement in Bohmian mechanics also comes from Holland [35].

measurement system. Then the value of $\hat{A}$ for the system will be

$$A(\lambda_x, \lambda_\psi) = \Big[Re(\psi_f^*(\hat{A}\psi_f)/|\psi_f|^2 \Big]_{\mathbf{x}=\mathbf{x}_f}. \tag{2.40}$$

Thus the indicator function for the measurement for outcome a of observable $\hat{A}$ at time t_0 is given by

$$\xi(a|\lambda_x, \lambda_\psi, \hat{A}) = \delta_{a,A(\lambda_x,\lambda_\psi)}, \tag{2.41}$$

and hence we see that Bohmian mechanics is outcome deterministic. The discussion above applies to all possible PVMs, but can be extended to accommodate proper POVMs. The above discussion is catered to the case of a single particle, but the formalism is extendible to multiple system [35]. Thus for any given d, improper d-level POVMs are also accommodated by the theory.

Revealed Properties, Non-Locality, and Contextuality

Despite the fact that Bohmian mechanics is a fundamentally deterministic theory, there are still an abundance of unusual features worth discussing.

Measurements in classical mechanics are always thought to passively reveal pre-existing properties of a system. For example, a police officer can measure the speed (and hence momentum) of a car and then correctly deduce whether or not the driver was speeding. In Bohmian mechanics, a measurement is an interference into the system which affects the property that is being measured. Since the Bohmian particle is travelling along a specific trajectory, it always has an actual value of momentum. However, a measurement of momentum will not reveal what the momentum of the particle was before the measurement. The momentum measurement will influence the momentum of the particle, and then report back the new, disturbed value of the momentum. In fact, the only property of the particle which is faithfully revealed or preserved by measurement in Bohmian mechanics is the particle position [1].

Bohmian mechanics is also non-local, as the wave-function ψ is still a part of the ontology. One can affect the values of ψ in space-like separated regions by acting on ψ locally. In the EPR experiment for example, the measurement of spin in region A involves a coupling of ψ to a measurement apparatus, both of which then evolve together. This evolution will effect ψ in the space-like separated region B, in some-sense 'communicating' what measurement was performed in region A. However, the non-locality is to be expected, as Bell's theorem tells us.

The contextuality of quantum theory is rendered completely understandable by Bohmian mechanics [25, 28, 1]. Bohmian mechanics incorporates measurement directly into the theory. Thus the ontological representation of different measurements are explicitly different in the Bohmian model. We can illustrate the point

with a rather simple example, which simultaneously shows that Bohmian mechanics is in a sense *more* contextual than is required by the Kochen-Specker theorem.

Consider a Stern-Gerlach experiment on a qubit system and suppose we restrict the spatial extent of the prepared wave-function to two dimensions: $\psi(\mathbf{x}, t) = \psi(x, y, t)$. Suppose that at time t_0, ψ localizes the ontic state λ_x to an exact value x_0, and to an extended region in the y dimension. The system is propagated in the $+x$ direction towards the Stern-Gerlach apparatus. The wave-function and apparatus interact such that the wave-function splits and some of the possible trajectories head in the $+y$ direction and others in the $-y$ direction with no trajectories crossing (see Figure 2.1). The proportions of trajectories that deviate upwards will be exactly what is needed in order for repeated trials to confirm the statistics of quantum theory. In the typical language that we use with regards to properties and measurements, we would say that the particles whose trajectories go upwards have a $+1$ component of spin in the $+y$ direction. But now consider what happens if the polarization of the Stern-Gerlach magnet is reversed, such that a trajectory travelling downwards corresponds to a $+1$ component of spin in the $+y$ direction. The proportion of trajectories that travel downwards must be equal to the proportion of trajectories that traveled upwards in the previous arrangement. However, trajectories cannot intersect each other in space-time. Thus some trajectories which deviated upwards in the previous set-up must still deviate upwards in order to preserve the quantum statistics, and thus will receive a different outcome assignment (see Figure 2.2).

Thus we see that the Bohmian model is contextual even for the qubit, which is not required by the Kochen-Specker theorem. Moreover, it is a sort of contextuality that is not required by the Kochen-Specker theorem. The outcome assigned to a particle is dependent on precisely how a PVM is measured, whereas the Kochen-Specker theorem only requires that a particle may change its outcome with respect to a projector depending on what PVM the projector appears in. This example makes explicit the fact that it is meaningless to say that the Bohmian particle has a particular value for a given quantum effect, or even a full PVM. However, if one knows the Bohmian ontic state, and the precise physical setup, then one can predict what the outcome of the experiment will be.

2.4.2 Bell's First Model

Prior to Bell's paper concerning his famous inequality, he published a paper [8] concerning the viability of hidden variable models in light of many arguments and supposed proofs against their existence [61, 39]. In it he demonstrates a very simple, albeit very cheap, ontological model for $\mathbb{Q}_2$. The model is easily extendible to a model for quantum theory of all dimensions, and is also able to handle POVMs. We present such an extension below. One can understand this ontological model as an actualizing of the probabilities of outcomes; as a model in which the quantum state comes coupled with its own set of dice to roll. This model is used as basis

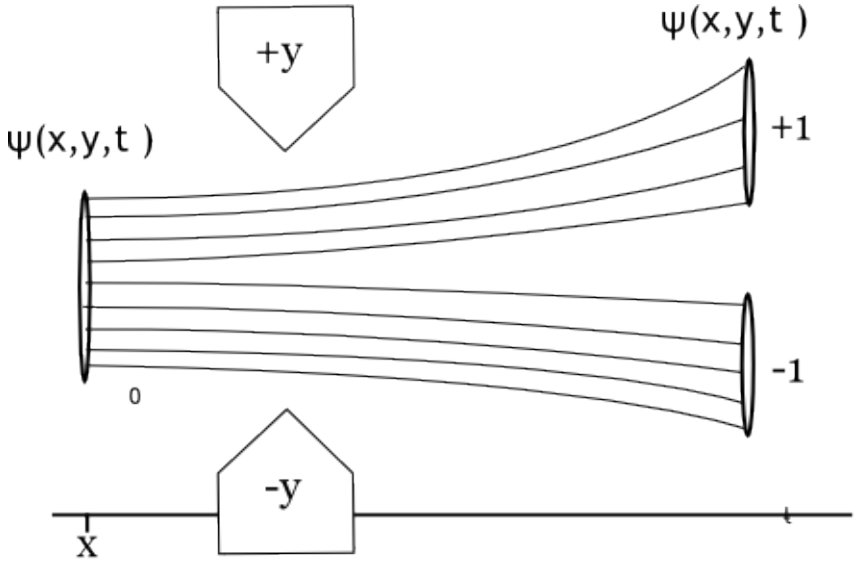

Figure 2.1: A simplified depiction of Bohmian particle trajectories in a Stern-Gerlach experiment. The trajectory of the particle, and hence the assignment of ± 1 as an outcome depends on the measurement and the initial particle position.

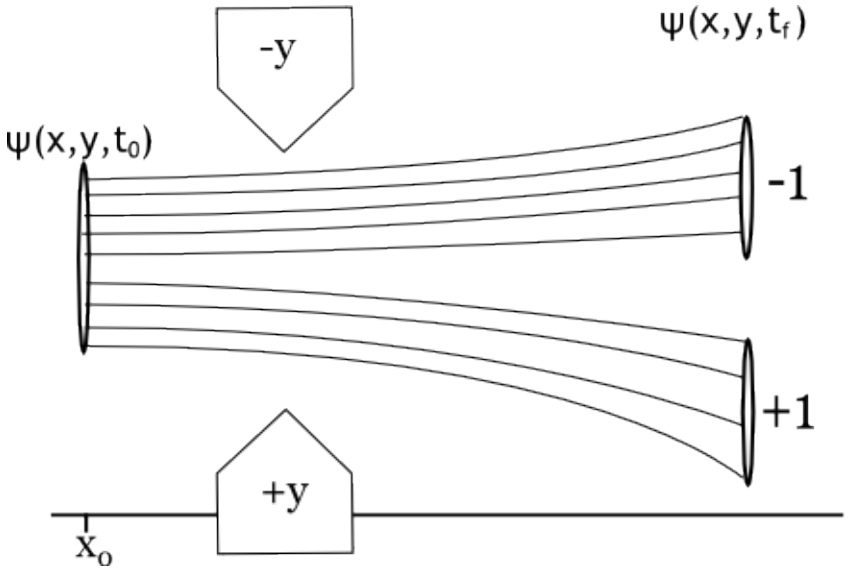

Figure 2.2: The Stern-Gerlach magnet is flipped. Some particles that would have traveled downwards in the previous setup before now travel upwards. However, some particles which would have registered $+1$ in the previous experiment register -1 in this one.

for another model presented later (see Section 3.3.2) which possesses interesting properties.

Consider as an ontic space $\Lambda = \Lambda_1 \times \Lambda_2 = \mathbb{P}\mathcal{H} \times [0, 1]$, the projective Hilbert space coupled with the unit interval. When the pure state ψ is prepared, the ontic

state of Λ_1 is ψ itself, and the ontic state of $[0,1]$ is drawn uniformly:

$$\mu_\psi(\lambda) = \delta(\lambda_1 - \psi) d\lambda_1 d\lambda_2. \tag{2.42}$$

The indicator functions in this model amount to appropriately slicing up the unit interval so that the probability of λ_2 being in each bin corresponds to the probability of a certain outcome. In particular, let $\{E_m\}_{m=1}^n$ be a POVM. Then

$$\xi_{E_m}(\lambda) = \begin{cases} 1 & \text{if } \sum_{j=1}^{m-1} \mathrm{tr}(E_m\psi) \leq \lambda_2 < \sum_{j=1}^{m} \mathrm{tr}(E_m\psi) \\ 0 & \text{else} \end{cases}. \tag{2.43}$$

It should be clear that the probability of outcome m when the quantum state is ψ is exactly $\mathrm{tr}(E_m\psi)$.

This model is clearly outcome deterministic, and in fact represents any quantum effect via an idempotent indicator function. However, in light of the fact that a convex ontological model must represent some mixed measurements via non-idempotent indicator functions (Lemma 2.1), it must be the case that this model is non-convex.

Non-locality manifests itself in this model for the same reason as in the Bohmian model: ψ is part of the ontology, and hence the model in inseparable.

2.4.3 Non-Convex Model

In Section 2.1 we briefly discussed the motivation behind considering the class $\mathbb{O}_{conv}$ of convex ontological models as candidates for rooting an operational theory in a realistic and physical framework. Here we demonstrate in a straight-forward manner that convexity is not a necessary feature of an ontological model representing a operational theory. As a counter-example, we construct here a simple non-convex ontological model for $\mathbb{Q}_2$ which is based upon the Beltrametti-Bugajski model (see Section 2.2.3).

Recall that the Beltrametti-Bugajski model is non-local by virtue of being non-separable. It is also not outcome deterministic. Another interesting property of the Beltrametti-Bugajski model is that it is not contextual in the Kochen-Specker sense because the model is not outcome deterministic. It is also not contextual in the sense of Bell's Theorem, since even the indicator functions for projectors, outcome deterministic or not, depend *only* on the projector they are representing, without reference to the PVM.

We now construct the non-convex model. For every preparation P, consider the density matrix ρ_P associated with P. Unless $\rho_P = \frac{\mathbb{I}}{2}$, then ρ_P is a non-degenerate matrix with a unique spectral decomposition:

$$\rho_P = e_1 \,|\, e_1 \rangle\langle\, e_1 \,| + e_2 \,|\, e_2 \rangle\langle\, e_2 \,|\,.$$

For such ρ_P, we make the definition for μ,

$$\mu_P = \mu_\rho = e_1\,\delta\left(\lambda - |\, e_1 \,\rangle\langle\, e_1\,|\right) + e_2\,\delta\left(\lambda - |\, e_2 \,\rangle\langle\, e_2\,|\right),$$

which is simply the linear combination of epistemic states of the Beltrametti-Bugajski model corresponding to the spectral decomposition of ρ_P. If $\rho_P = \frac{\mathbb{I}}{2}$ then there is no unique spectral decomposition, since there are degenerate eigenvalues. However, we can choose a preferred basis in which to decompose the degenerate eigenspace:

$$\mu_{\frac{\mathbb{I}}{2}} = \frac{1}{2}\,\delta\left(\lambda - |\, 0 \,\rangle\langle\, 0\,|\right) + \frac{1}{2}\,\delta\left(\lambda - |\, 1 \,\rangle\langle\, 1\,|\right). \tag{2.44}$$

Defined in this way, μ is not convex. For any non-pure density matrix, ρ has an infinite number of decompositions into pure states [49]. Suppose that $\sum_i p_i\psi_i$ is a decomposition of ρ_p other than the spectral decomposition. Then

$$\sum_i p_i\delta(\lambda - \psi_i) \neq e_1\delta\left(\lambda - |\, e_1 \,\rangle\langle\, e_1\,|\right) + e_2\,\delta(\lambda - |\, e_2 \,\rangle\langle\, e_2\,|), \tag{2.45}$$

since the supports of these two measures will differ. However, the above two measures would have to be equal if μ were a convex function. This model, together with the indicator functions of the Beltrametti-Bugajski model will reproduce the statistics of $\mathbb{Q}_2$ since for each density matrix ρ, μ_ρ is a convex decomposition into the measures corresponding to *some* valid convex decomposition of ρ into pure states.

While this model is mathematically consistent with quantum statistics and the ontological model framework, it presents interpretive problems. Let $\{|\, e_1 \,\rangle, |\, e_1 \,\rangle\}$ be any basis for $\mathbb{C}^2$ which is not the standard basis $\{|\, 0 \,\rangle, |\, 1 \,\rangle\}$. Suppose a sequence of experiments is performed wherein $|\, e_1 \,\rangle$ or $|\, e_2 \,\rangle$ is prepared with equal probability. Considering any one run, if the state $|\, e_1 \,\rangle$ is prepared, the model implies that the value of the ontic state is $\lambda = |\, e_1 \,\rangle\langle\, e_1\,|$, and if the state $|\, e_2 \,\rangle$ is prepared, then the ontic state is $\lambda = |\, e_2 \,\rangle\langle\, e_2\,|$. However, since these individual runs take place in a sequence of probabilistic experiments, the quantum state being prepared is

$$\frac{1}{2}\,|\, e_1 \,\rangle\langle\, e_1\,| + \frac{1}{2}\,|\, e_2 \,\rangle\langle\, e_2\,| = \tfrac{\mathbb{I}}{2},$$

which the model tells us should be associated with the epistemic state

$$\frac{1}{2}\delta(\lambda - |\, 0 \,\rangle\langle\, 0\,|) + \frac{1}{2}\delta(\lambda - |\, 1 \,\rangle\langle\, 1\,|).$$

If this is the case, then the possible ontic values for λ are $|\, 0 \,\rangle\langle\, 0\,|$ and $|\, 1 \,\rangle\langle\, 1\,|$, which is contradictory with the assumption that in any single run that the possible values for λ are $|\, e_1 \,\rangle\langle\, e_1\,|$ or $|\, e_2 \,\rangle\langle\, e_2\,|$. Thus a non-convex model could potentially imply a dependence of the ontic state on whether or not a preparation takes place within a probabilistic ensemble of preparations or not. This does not seem like a desirable property for an ontological model.

We can define similar models for higher dimensional quantum systems. The problem that arises is that in higher dimensions there are an infinite number of degenerate density matrices. Thus we need a rule, as in (2.44), for choosing a convex decomposition for any degenerate density matrix. Such a choice amounts to choosing an orthogonal basis for any possible subspace of the Hilbert space $\mathbb{C}^d$.

For example, suppose a density matrix ρ has eigenvalues $\{a_i\}$. Associated to each eigenvalue will be a projector P_i onto the eigenspace corresponding to a_i such that

$$\rho = \sum_i a_i P_i. \tag{2.46}$$

However, if any eigenvalue a_i is degenerate, then the projector P_i will have rank larger than 1, and hence (2.46) will not be a convex decomposition in terms of pure states. Thus for each P_i of rank larger than 1, we need to find pure states $\left\{P_i^j\right\}_j$ such that $P_i = \sum_j P_i^j$. We can make the choice in the following way.

Suppose S is subspace of dimension 2 or greater. Starting with the standard basis $B = \{|\, i \,\rangle\}_{i=1}^d$ for the entire Hilbert space, we first project each element onto S. This gives B' such that $\operatorname{span} B' = S$. Then simply perform a Gram-Schmidt procedure on B' in order to get an orthonormal basis for S.

Chapter 3

Epistemics

The epistemic view of the quantum state is one in which the quantum state is not taken to be a complete descriptor of a physical system. The quantum state, in this view, should at most specify a distribution or measure over other properties, and thus the quantum state corresponds to an *ensemble* of systems, rather than any one system. Einstein was one of the first physicists to espouse the *epistemic* view of the quantum state. However, his motivations for this view were derived from an obligation to locality. Bell's Theorem quashes the validity of this motivation, but other arguments have been made for this viewpoint.

Ballentine [4] has argued that an epistemic view of the quantum state is vital for doing away with the measurement problem. The measurement problem involves a conflict between the mathematical structure of quantum theory and our classical intuition for macroscopic objects. In particular, suppose a quantum system interacts with a measurement apparatus, which we can also describe with quantum theory. The measurement apparatus starts in a *ready* state $|\, M_{ready} \,\rangle$. The quantum system and the measurement apparatus are considered to be a closed system, and so their evolution (the measurement) can be described by a unitary U. If the measurement interaction and apparatus faithfully measures in the standard basis, then we have

$$U\,|\,0\,\rangle\,|\,M_{ready}\,\rangle = |\,0\,\rangle\,|\,M_0\,\rangle \qquad U\,|\,1\,\rangle\,|\,M_{ready}\,\rangle = |\,1\,\rangle\,|\,M_1\,\rangle\,. \tag{3.1}$$

In order for this to be an effective measurement procedure $|\,M_0\,\rangle$ and $|\,M_1\,\rangle$ must be macroscopically distinguishable states. Now suppose that the initial state of the quantum system is given by a superposition:

$$|\,\psi\,\rangle = a\,|\,0\,\rangle + b\,|\,1\,\rangle \quad \text{such that} \quad |a|^2 + |b^2| = 1. \tag{3.2}$$

Then the measurement interaction yields

$$U\,|\,\psi\,\rangle\,|\,M_{ready}\,\rangle = a\,|\,0\,\rangle\,|\,M_0\,\rangle + b\,|\,1\,\rangle\,|\,M_1\,\rangle\,. \tag{3.3}$$

If this state is to be taken as a full description of the system, then we see that there is no sense in which the macroscopic measurement apparatus is in a macroscopically

distinguishable state. Furthermore, it is not until the measurement apparatus is observed by the experimentalist that the post-measurement state is determined to be $|\,0\,\rangle\,|\,M_0\,\rangle$ or $|\,1\,\rangle\,|\,M_1\,\rangle$. If the quantum state is taken to be a complete description, then this accounting of measurement requires a discontinuous evolution of the quantum state which is dependent on the act of observation. Such odd characteristics of measurement do not occur when one assumes that the quantum state describes a statistical ensemble [4].

However, there is also evidence that such an epistemic view can demystify much of quantum theory. Arguments and work presented over the years by Ballentine, Emerson, Spekkens and Fuchs (to name a few) all demonstrate that many feats and features thought to be unique to quantum theory are in fact reproducible in models where the most precise state of knowledge is restricted to be a non-trivial measure over some ontic space, i.e. a truly epistemic state [4, 59, 20].

In Section 3.1 we present characterizations of ontological models intended to capture the cases where the quantum state represents a truly epistemic state. Such models are called ψ-epistemic [33]. In Section 3.2 we present the historically first model which falls into the ψ-epistemic characterization. The problem of existence of satisfactory ψ-epistemic models is discussed in Section 3.3. Within this section we contribute a new result, namely that the model of Section 3.2 cannot be extended to accommodate all measurements. We also present a recently proposed ψ-epistemic model due to Rudolph[1]. This model ends up not being 'ψ-epistemic enough'. In light of this character of Ruldolph's model, in Section 3.3.3 we suggest alternative characterizations to capture the desired idea that the quantum state is truly epistemic. These suggestions lead to some questions for further research.

3.1 A Characterization of Epistemic Ontological Models

In [33], a new characterization of ontological models was introduced which categorized based on the ontological status the model accorded to a pure quantum state. At one end of the categorization spectrum is the orthodox realist view, responsible for the measurement problem, that the quantum state is the most descriptive one can get about physical reality.

Definition 3.1 (ψ-complete)**.** An ontological model (Λ, μ, ξ) for $\mathbb{Q}_d$ is ψ-complete if the ontic space Λ is isomorphic to the set of pure states $\mathbb{CP}^{d-1}$ and the preparation of a pure-state ψ induces a delta distribution over ψ on Λ i.e. $\mu_\psi = \delta(\lambda - \psi)$.

This idea is precisely that which is encompassed by the Beltrametti-Bugajski model (see Section 2.2.3). However, this is not the only sense in which the quantum state can be viewed as ontological. Consider Bohmian mechanics, in which

[1]This model is not published, but has been reproduced here with permission from Rudolph.

the quantum state of the system dictates the motion of the particle, and hence determines measurement outcomes. The quantum state is *supplemented* with a hidden position variable, and yet it still plays an ontological role. Hence Bohmian mechanics falls into the following category:

Definition 3.2 (ψ-ontic)**.** An ontological model (Λ, μ, ξ) for $\mathbb{Q}_d$ is ψ-ontic if for every pair of distinct pure states ψ_1 and ψ_2, it is the case that

$$\int_\Lambda \mu_{\psi_1}(\lambda)\mu_{\psi_2}(\lambda)d\mu = 0.^2 \tag{3.4}$$

Intuitively, condition (3.4) states that for two distinct pure states, the preparations of these pure states result in ontic states from disjoint sets in Λ. The motivation for this definition is the following. If a change in the quantum state necessarily induces a change in the ontic state, then the quantum state must have some direct correspondence to physical reality. Knowledge of the ontic state allows one to determine the quantum state, thus the quantum state is ontological.

A ψ-complete theory is also ψ-ontic, thus it is useful to make a sub-categorization of ψ-ontic for the case where ψ is not 'the whole story'.

Definition 3.3 (ψ-supplemented)**.** An ontological model (Λ, μ, ξ) for $\mathbb{Q}_d$ is ψ-supplemented if it is ψ-ontic but not ψ-complete.

Lastly, we can define a theory in which not all ontic states can be connect with particular quantum states.

Definition 3.4 (ψ-epistemic)**.** An ontological model (Λ, μ, ξ) for $\mathbb{Q}_d$ is ψ-epistemic if it is not ψ-ontic. In particular, if there exist two quantum states ψ_1 and ψ_2 such that

$$\int_\Lambda \mu_{\psi_1}(\lambda)\mu_{\psi_2}(\lambda)d\mu > 0, \tag{3.5}$$

then the ontological model is ψ-epistemic.

In Section 3.3 we show why the above definition is not entirely satisfactory. We give an explicit example of a model (Section 3.3.2) which satisfies it, and yet does not seem to capture the notion that pure states have 'epistemic character'.

3.2 Kochen-Specker Qubit Model

In their classic paper [40], Kochen and Specker present an ontological model for quantum mechanics on a 2 dimensional Hilbert space. What is particularly interesting about this model is that it is one of the only known ψ-epistemic models.

[2] μ is any measure with respect to which both μ_{ψ_1} and μ_{ψ_2} are absolutely continuous. i.e. $d\mu_{\psi_1} = \mu_{\psi_1}(\lambda)d\mu$, $d\mu_{\psi_2} = \mu_{\psi_2}(\lambda)d\mu$.

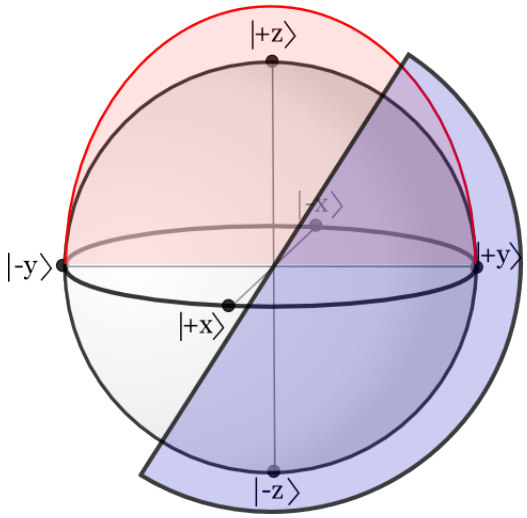

Figure 3.1: The Kochen-Specker Qubit Model. Depicted here is the epistemic state $\mu_{|\,0\,\rangle}$ for the preparation of the $|\,0\,\rangle$ state, and a typical indicator function. The epistemic state is centered around the Bloch sphere point of the quantum state that it represents. It decreases like $\cos\theta$, and has support only on a hemisphere. The indicator function is idempotent and also has support only on the hemisphere centered on the Bloch sphere point of the projector it represents.

The Kochen-Specker (KS) model is presented as a model for $\partial\mathbb{Q}_2$. Consider the ontic space Λ to be the 2-sphere, $\mathcal{S}^2$. Since $\mathcal{S}^2$ is isomorphic to the projective Hilbert space $\mathbb{CP}^1$, there is a bijection between points on the sphere and the pure states (see Section 1.3). Here we will write $\lambda(\psi)$ to represent the point on the Bloch sphere corresponding to pure state ψ. Kochen and Specker then define the maps μ and ξ in the following way. The pure-state projector ψ is represented as

$$\mu_\psi(\lambda) = \begin{cases} \frac{1}{\pi}\cos\theta & 0 \leq \theta \leq \frac{\pi}{2} \\ 0 & \text{else} \end{cases} \tag{3.6}$$

where θ is the angular separation between λ and $\lambda(\psi)$ on the Bloch sphere. In terms of PVMs, the representation of the projector ψ is given by

$$\xi_\psi(\lambda) = \begin{cases} 1 & 0 \leq \theta \leq \frac{\pi}{2} \\ 0 & \text{else} \end{cases}. \tag{3.7}$$

To verify that these choices of distributions and indicator functions reproduce the qubit statistics, we simply have to verify that if $\psi = |\,0,0\,\rangle\langle\,0,0\,|$ and $\psi' = |\,\theta,\frac{\pi}{2}\,\rangle\langle\,\theta,\frac{\pi}{2}\,|$ then $\int_{\mathcal{S}^2}\mu_\psi\xi_{\psi'}d\lambda = \cos^2\left(\frac{\theta}{2}\right)$. We will be parameterizing the sphere in two ways. The first set of coordinates are the standard (θ,ϕ) coordinates the we have been using previously. The second set is (β,α) where β is the zenith angle off of the $+x$-axis and α is the corresponding azimuth angle in the yz-plane off of the $+y$-axis. The two are displayed side-by-side in Fig 3.2. Given the choice of ϕ and

ψ', a sample area of integration is displayed in Fig 3.3, which gives an idea for why we choose to work with the (β, α) coordinate system.

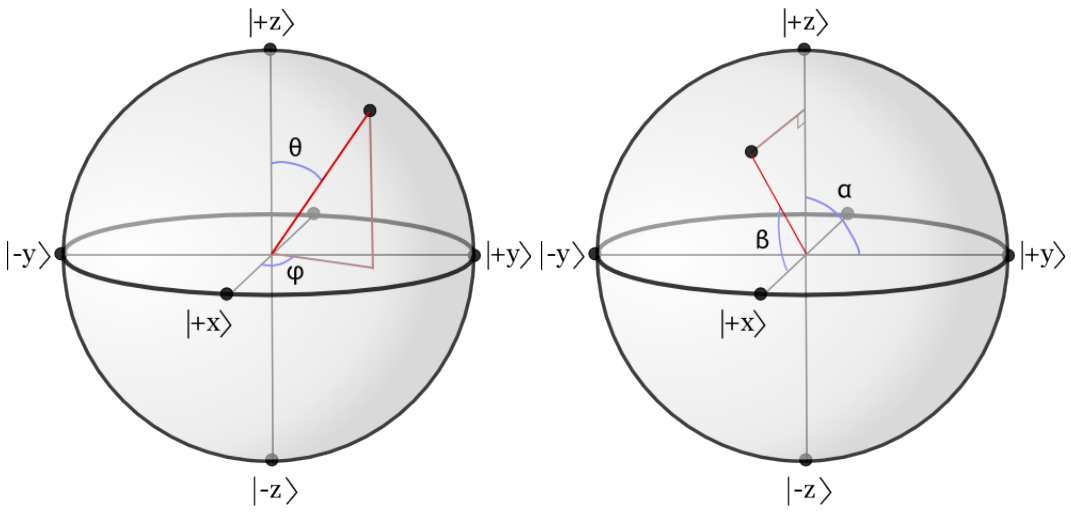

Figure 3.2: The two coordinate systems in use.

One important thing to note is that given a point (β, α), the value of $\cos(\theta)$ for the corresponding point in (θ, ϕ) coordinates is given by $\sin(\alpha)\sin(\beta)$. To see this, we simply note that $\cos(\theta)$ is the length of the projection of the vector corresponding to (β, α) onto the z axis, as is the value $\sin(\alpha)\sin(\beta)$. Thus we have

$$\begin{aligned}\int_{\mathcal{S}^2} \mu_\psi \xi_{\psi'} d\lambda &= \int_0^{\pi} \int_0^{\pi-\theta} \frac{1}{\pi} \sin(\alpha)\sin^2(\beta) d\alpha d\beta \\ &= [1+\cos(\theta)]\frac{1}{\pi}\int_0^{\pi} \sin^2(\beta) d\beta \\ &= [1+\cos(\theta)]\frac{\pi}{2\pi} \\ &= \cos^2(\tfrac{\theta}{2}),\end{aligned} \tag{3.8}$$

which verifies the correctness of the model.

In principle such a model could be implemented physically with a classical angular momentum. We merely have to impose restrictions on how the spin can be prepared and measured. It can be prepared probabilistically in a cos-distribution over any hemisphere of orientation, and the allowed measurements are restricted to learning only which of two disjoint hemispheres the spin is pointing in (see Figure 3.1).

It is easy to extend this model to a convex ontological model for all of $\mathbb{Q}_2$. We simply allow for mixed preparations and extend μ linearly on this set. Similarly we allow for mixed measurements i.e. proper d-level POVMs, and extend ξ linearly.

3.3 Existence of ψ-epistemic Models for $\mathbb{Q}_d$

The vast majority of proposed ontological models for quantum theory to date have not been ψ-epistemic. Until recently, the only known example has been the KS model, which as presented only applies to $\mathbb{Q}_2$. All other proposed models have been ψ-ontic. Given the lack of ψ-epistemic models, one might consider the possibility that none exist which are applicable to all levels of quantum theory.

Additionally, one could validly be concerned that an ontological model satisfying Definition 3.4 may be entirely unsatisfactory. The definition requires that there only exist two quantum states for which their induced probability measures 'overlap'. By that condition, every other pair of quantum states could have non-overlapping measures, and thus most quantum states would still retain an ontic character. In fact, it is possible to construct a model for $\mathbb{Q}_d$ for which most quantum states do not have overlapping preparation measures. Such a model has been suggested by Rudolph, but remains unpublished. We give a summary of his model in Section 3.3.2.

It is reasonable to wonder whether or not the KS model could be extended in some way to incorporate POVMs which are not proper 2-level POVMs (see Section 1.2). If this was the case, it could be seen as evidence that the KS model could be extended to a ψ-epistemic model for $\mathbb{Q}_d$ ($d > 2$). In the next section, we present a new result that the KS model cannot reproduce the statistics for all improper 2-level POVMs. In particular, we show that a measurement referred to as the 'trine' measurement, can not be represented in the KS Model.

3.3.1 No Trine in the KS Model

Theorem 3.1. *The so-called trine POVM on a qubit system has no representation in the KS model.*

Since the KS model is an ontological model for $\mathbb{Q}_2$, it must be able to reproduce all proper 2-level POVMs. Thus a direct corollary of this result is the separation of the proper and improper POVMs. Since the trine is not reproducible in the KS model, it must not be a proper 2-level POVM.

Proof. The trine measurement has effects

$$E_1 = \frac{2}{3}\left|\, 0, \tfrac{\pi}{2} \right\rangle\!\!\left\langle 0, \tfrac{\pi}{2} \,\right| \quad E_2 = \frac{2}{3}\left|\, \tfrac{2\pi}{3}, \tfrac{\pi}{2} \right\rangle\!\!\left\langle \tfrac{2\pi}{3}, \tfrac{\pi}{2} \,\right| \quad E_3 = \frac{2}{3}\left|\, -\tfrac{2\pi}{3}, \tfrac{\pi}{2} \right\rangle\!\!\left\langle -\tfrac{2\pi}{3}, \tfrac{\pi}{2} \,\right|.$$

We will show that there are no indicator functions $\{\xi_1, \xi_2, \xi_3\}$ on S^2 such that $\sum \xi_i(\lambda) = 1$ and $tr(\rho E_i) = \int \mu_\rho(\lambda)\xi_i(\lambda)d\lambda$, where μ_ρ is given by the distribution in the KS model (3.6).

To simplify the argument, we will only be considering pure states that lie in the yz-plane (i.e. pure states with Bloch vectors at $(\theta, \frac{\pi}{2})$ and $(\frac{\pi}{2}, \alpha)$ in the above

coordinate systems, with $\alpha = \theta$). In general, we will refer to such states as $|\,\theta\,\rangle$, or ρ_θ for the associated density matrix. The point on the Bloch sphere corresponding to ρ_θ will be referred to as p_θ.

The first thing to note is that ξ_1 must be zero outside of the hemisphere centered at p_0. Indeed, if the state $|\,\pi\,\rangle\langle\,\pi\,|$ is prepared, which has support on the opposite hemisphere, the trine measurement will never give outcome E_1, as $\langle\,\pi\,|\,0\,\rangle = 0$.

For $0 \leq \theta \leq \pi$ the quantum statistics give

$$tr(\rho_{\pi-\theta}E_1) = \tfrac{2}{3}\cos^2\frac{\pi-\theta}{2} = \tfrac{2}{3}\sin^2\tfrac{\theta}{2}.$$

Thus ξ_1 must be such that

$$\int \mu_{\rho_{\pi-\theta}}(\lambda)\xi_1(\lambda)d\lambda = \tfrac{2}{3}\sin^2\tfrac{\theta}{2}. \tag{3.9}$$

Figure 3.3 depicts the area of integration for (3.9).

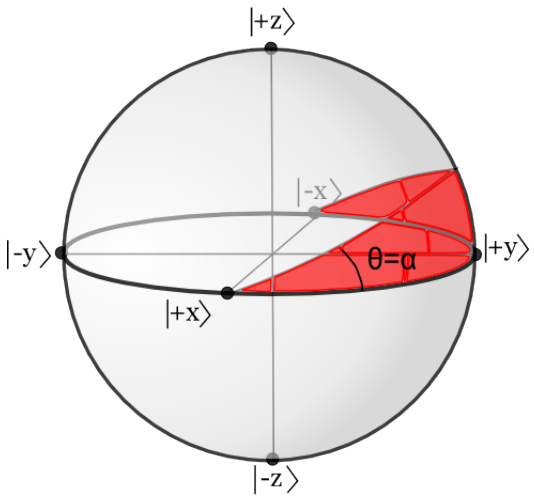

Figure 3.3: Area of overlap between the support of ξ_1 and $\mu_{\rho_{\pi-\theta}}$

Using this as a visual guide and parameterizing with (α, β) gives

$$\begin{aligned}\tfrac{2}{3}\sin^2\tfrac{\theta}{2} &= \int_0^{2\pi}\int_0^{\pi} \mu_{\rho_{\pi-\theta}}(\beta,\alpha)\xi_1(\beta,\alpha)\sin\beta d\beta d\alpha \\ &= \int_0^{\theta}\int_0^{\pi} \tfrac{1}{\pi}\cos(\gamma)\xi_1(\lambda)\sin\beta d\beta d\alpha,\end{aligned} \tag{3.10}$$

where γ denotes the angular separation between the point (β, α) and the point $p_{\pi-\theta}$ at the center of the distribution $\mu_{\rho_{\pi-\theta}}$. We make a substitution for $\cos\gamma$ in terms

of (β, α) and continue:

$$\begin{aligned}
&= \int_0^\theta \int_0^\pi \tfrac{1}{\pi} \sin\beta \cos(\theta - \pi/2 - \alpha)\xi_1(\alpha,\beta) d\beta d\alpha \\
&= \int_0^\theta \int_0^\pi \tfrac{1}{\pi} \sin^2\beta \sin(\theta - \alpha)\xi_1(\alpha,\beta) d\beta d\alpha \\
&= \int_0^\theta \sin(\theta-\alpha)\tilde{\xi}_1(\alpha) d\alpha.
\end{aligned}$$

In the second line, we have used the fact that ξ_1 is zero outside of the hemisphere centered at $|\,0\rangle\langle\,0\,|$ and that $\mu_{\rho_{\pi-\theta}}$ is zero outside of the hemisphere centered at $\rho_{\pi-\theta}$ (see. Figure 3.3). In the last line we have performed the integration over the β variable and defined

$$\tilde{\xi}_i \equiv \int_0^\pi \tfrac{1}{\pi} \sin^2(\beta)\, \xi_i(\alpha,\beta) d\beta. \tag{3.11}$$

Now, since

$$\int_0^\theta \tfrac{1}{3} \sin(\theta-\alpha) d\alpha = \tfrac{2}{3} \sin^2 \tfrac{\theta}{2},$$

using (3.10) gives

$$0 = \int_0^\theta (\tfrac{1}{3} - \tilde{\xi}_1(\alpha)) \sin(\theta-\alpha) d\alpha.$$

This implies that $1/3 - \tilde{\xi}_1(\alpha) = 0$, by a zero lemma (see C)[3].

Similarly, we can show that each $\tilde{\xi}_i$ must take on values 0 or $\frac{1}{3}$ over S^2. But since we require $\sum_{i=1}^3 \xi_i = 1$, the definition of $\tilde{\xi}$ yields

$$\sum_{i=1}^3 \tilde{\xi}_i = \int_0^\pi \tfrac{1}{\pi} \sin^2\beta = \tfrac{1}{2},$$

which contradicts $\tilde{\xi}_i(\theta) \in (\frac{1}{3}, 0)$. □

3.3.2 Rudolph's ψ-epistemic Model

Rudolph's model is based on the idea of Bell's first model (see Section 2.4.2). Thus, consider again the ontic space $\Lambda = \mathbb{CP}^{d-1} \times [0,1]$. The measurement procedure is exactly the same as in Bell's first model, with the one important difference that the outcomes are ordered in a specific way.

Suppose that there is a fixed special quantum state $\psi_F \in \mathbb{CP}^{d-1}$ known to every preparation and measurement device in the universe, and that a measurement device is set to measure a PVM $M = \{P_i\}_{i=1}^d$. The device knows the probabilities of the outcomes for this PVM when the quantum state is ψ_F ($p_1 = tr(\psi_F P_1), \ldots p_d =$

[3]For the zero lemma to go through, we assume that $\tilde{\xi}$ must be piece-wise continuous

$tr(\psi_F P_d)$). The measurement device in Rudolph's model always indexes the effects such that $p_1 \geq p_2 \ldots \geq p_n$. That is the 'first outcome' is always the outcome which is most likely for the quantum state ψ_F. When we write P_1, it is implicit that this is the effect (in any PVM) which is most likely for ψ_F.

The preparation device will be aware of this fact and is able to take advantage of it in the following way. For any state ψ we can calculate a lower bound G_ψ on the probability that ψ will have the 'first' outcome (by the ordering of the measurement device) for *any* PVM that may be performed. Intuitively, this lower bound should increase as ψ gets closer to ψ_F. The preparation device will begin the preparation procedure for ψ by calculating G_ψ. Then $\lambda_p \in [0, 1]$ is drawn uniformly at random. If $\lambda_p < G_\psi$, then the preparation device knows that no matter what PVM is performed, the outcome will be whichever effect is *most likely for the quantum state* ψ_F, since the measurement device will allocate the first block of $[0, 1]$ to this outcome (this is what is done in Bell's first model, except that here the outcome ordering is specific). The preparation device then has a freedom of which value to pick for λ_q. Any ψ' for which $G_{\psi'} \geq \lambda_p$ will in fact work. Why? Suppose the PVM is given by $M = \{P_i\}_{i=1}^d$. The measurement device, set to measure M, will receive the ontic state $\lambda = \psi' \times \lambda_p$ and will calculate the probabilities of outcomes for M and ψ'. It will determine that $p_1 = tr(\psi' P_1) \geq G_{\psi'} \geq \lambda_p$, and so will register outcome 1, as it would have done if the preparation device set $\lambda_q = \psi$. Thus, having drawn $\lambda_p \leq G_\psi$, the preparation device then chooses randomly any ψ' such that $G_{\psi'} \geq \lambda_p$ and sets $\lambda_q = \psi'$.

This model reproduces the statistics of quantum theory: in the case where $\lambda_p > G_\psi$ (normal operation), it performs in the exact same fashion as Bell's first model; in the case where $\lambda_p \leq G_\psi$, the measurement outcome (for any measurement!) under normal operation is known at the time preparation, and λ_q is chosen to ensure this outcome persists.

We now discuss the derivation of the bound G_ψ.

Claim 3.1. *For any PVM $M = \{P_i\}$ and any pure state ψ,*

$$\min_M \max_i tr(\psi P_i) = \tfrac{1}{d}. \tag{3.12}$$

Proof. It is clear that if $\min_M \max_i tr(\psi P_i) < \frac{1}{d}$, then the sum of all d probabilities will be less than 1, thus the minimum probability must be bounded below by $\frac{1}{d}$. However, the value $\frac{1}{d}$ can be achieved. Let ψ be a uniform superposition of the elements of any basis and let M be the measurement of the same basis. Then all outcomes have probability $\frac{1}{d}$. □

As mentioned, if we fix a PVM $M = \{P_i\}$, the measurement device orders the outcomes such that $p_1 \geq p_2 \ldots \geq p_d$. By the above claim, p_1 is bounded below by $\frac{1}{d}$. If $p = \mathrm{tr}(\psi_1 \psi_2)$ for two pure states ψ_1 and ψ_2, then $\arccos p$ defines an angle between these states, since $\mathrm{tr}(A^\dagger B)$ defines an inner product on the space of linear

operators on $\mathbb{C}^d$. Since $0 \leq p \leq 1$, taking the inverse cosine gives $0 \leq \arccos p \leq \frac{\pi}{2}$. Now define α to be the angle between ψ and P_1, θ to be the angle between ψ and ψ_F. It follows from a triangle inequality on angles that $\alpha \leq \theta + \arccos(\frac{1}{d})$.

We then hope to take the cosine of this expression to achieve a lower bound on $\cos\alpha$. However, a lower bound is only achieved if $\theta + \arccos(\frac{1}{d})$ is in the range of arccos, as cos is only guaranteed to be decreasing on this domain. However, we have $0 \leq \theta, \arccos(\frac{1}{d}) \leq \frac{\pi}{2}$, thus $0 \leq \theta + \arccos(\frac{1}{d}) \leq \pi$. Hence we do in fact derive a lower bound for $\cos\alpha = tr(\psi P_1)$:

$$G(d, \psi) = \cos(\theta + \arccos(\tfrac{1}{d})). \tag{3.13}$$

Notice though that since it is possible that $\theta + \arccos(\frac{1}{d}) \geq \frac{\pi}{2}$, then we could have $G(d, \psi) \leq 0$ (see Fig. 3.3.2). In this case, the bound on $tr(\psi P_1)$ is trivial, and we can never have $\lambda_p \leq G(d, \psi)$.

Without explicitly writing down the formula for μ_ψ,we can show that this model is ψ-epistemic. If it is possible that the ontic state $\lambda_q \times \lambda_p$ could be sent to represent two different quantum states, ψ and ψ', then the model is ψ-epistemic by Definition 3.4. In particular, if $\lambda_p \leq G_\psi$, then an ontic state $\psi' \times \lambda_p$, with $\psi' \neq \psi$ could be prepared. However, it is clear that $\psi' \times \lambda_p$ must also be consistent with a preparation of ψ'.

This model, although very clever, is unsatisfying as a ψ-epistemic model since $G(d, \psi) > 0$ for a limited set of ψ. If preparing a ψ such that $G(d, \psi) < 0$, the preparation device will always set $\lambda_q = \psi$. Similarly, $\lambda_q = \psi$ could never be prepared in lieu of another ontic state. Thus the probability measure for ψ has no overlap with any other preparation. Thus this particular ψ could be viewed as having an ontic character, since it corresponds to specific ontic states. Alternatively we say that it does not have epistemic character. Moreover, as d gets larger, $\arccos\frac{1}{d}$ gets ever closer to $\frac{\pi}{2}$. Thus the angular separation between ψ and ψ_F needed for $G(d, \psi) > 0$ becomes smaller, and the proportion of quantum states which can share ontic states decreases.

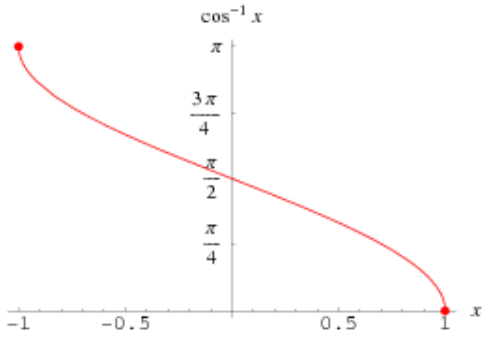

Figure 3.4: Arccos

3.3.3 Proposed Alternative Definition of ψ-epistemic and Discussion

The characterization in Section 3.1 would label Rudolph's model as ψ-epistemic. However, given that most pure states in this model do not have epistemic character, one should modify the epistemic characterization to exclude such models. A clear alternative which requires all pure states to have epistemic character is the following:

Definition 3.5 (Completely ψ-epistemic)**.** An ontological model (Λ, μ, ξ) for $\mathbb{Q}_d$ is completely ψ-epistemic if for every pure state ψ, there exists a pure state ψ' such that

$$\int_\Lambda \mu_\psi(\lambda)\mu_{\psi'}(\lambda)d\mu > 0 \tag{3.14}$$

holds.

This definition requires that every pure state has overlap with at least one other state. As such, we might say that all pure states have an epistemic character if an ontological model for $\mathbb{Q}_d$ is completely ψ-epistemic.

However, for every pair of non-orthogonal pure states $\{\psi_1, \psi_2\}$, the quantum projector effect E_{ψ_1} corresponding to ψ has a non-zero inner product with ψ_2, and vice-versa. That is, in a PVM which includes E_{ψ_1}, if the quantum state is ψ_2, there is a non-zero probability that the outcome will be E_{ψ_1}. Thus one may suppose that the epistemic state μ_{ψ_2} should have overlap with μ_{ψ_2}. Consequently, we have the following stronger notion of an epistemic model.

Definition 3.6 (Super Completely ψ-epistemic)**.** An ontological model (Λ, μ, ξ) for $\mathbb{Q}_d$ is super completely ψ-epistemic if for every pair of non-orthogonal pure states ψ and ψ',

$$\int_\Lambda \mu_\psi(\lambda)\mu_{\psi'}(\lambda)d\mu > 0 \tag{3.15}$$

holds.

It is shown in Section 4.1 that two orthogonal pure states must have non-overlapping support, thus the above definition requires that each pure state shares ontic states with every other pure state that it can. The Kochen-Specker model satisfies this definition, as can be seen by the fact that each epistemic state has support on an entire hemisphere. This notion of an epistemic model was actually suggested by Spekkens [58] in a paper previous to [33] in which the stated definition of ψ-epistemic is presented.

The following argument shows that an ontological model may fail to satisfy any definition of ψ-epistemic, while still essentially obeying the spirit of these definitions.

Let (Λ, μ, ξ) be a ((super) completely) ψ-epistemic model for $\mathbb{Q}_d$. We can define a new ontological model in the following way:

- $\hat{\Lambda} = \Lambda \times P\mathcal{H}$;

- $\hat{\mu}_\psi(\lambda, \lambda') = \mu_\psi(\lambda)\delta(\lambda' - \psi)$;

- $\hat{\xi}_M(k|\lambda, \lambda') = \xi_M(k|\lambda)$.

This model reproduces the statistics of $\mathbb{Q}_d$:

$$\begin{aligned} &\int_{\hat{\Lambda}} \hat{\xi}_M(k|\lambda, \lambda') d\hat{\mu}_\psi(\lambda, \lambda') \\ &= \int_\Lambda \xi_M(k|\lambda) d\mu_\psi(\lambda) \int_{\mathbb{CP}^{d-1}} \delta(\lambda' - \psi) d\lambda' \\ &= \int_\Lambda \xi_M(k|\lambda) d\mu_\psi(\lambda) = tr(\psi M_k). \end{aligned} \tag{3.16}$$

However, for $\psi \neq \psi'$, we have

$$\int_{\hat{\Lambda}} d(\hat{\mu}_\psi \hat{\mu}_{\psi'})(\lambda, \lambda') = \int_{\hat{\Lambda}} \delta(\lambda' - \psi)\delta(\lambda' - \psi') d\lambda' d(\mu_\psi \mu_{\psi'})(\lambda) = 0. \tag{3.17}$$

As such, this new model is ψ-ontic, despite the fact that the extra space contributes nothing to the measurement. Indeed, $\hat{\xi}$ is independent of λ', and without the extra ontology the model is ψ-epistemic.

The above discussion suggests a number of questions worth exploring. Can we rigourously define a notion of a reduced ontological model which removes any possible 'junk' ontology which might be obstructing a true epistemic character? Can we find a reduced ontological for $\mathbb{Q}_d$, $d \geq 2$ which is (super) completely ψ-epistemic? In light of the fact that ψ-epistemic models seem to mitigate the measurement problem, can we make a precise characterization of ontological models which successfully do so? In what way would such a characterization be related to the above ψ-epistemic characterizations?

Undoubtedly, to fully consider any question about the existence of a satisfactory model for quantum theory, we must ensure that the model handles all aspects of quantum theory. The majority of the models considered thus far are clearly lacking. Some fail to accommodate quantum systems of all dimensions. Most do not give an adequate account of dynamics, or how separate systems couple with each other. The only model which accomplishes either of these is Bohmian mechanics, which is certainly not ψ-epistemic. A very wide open problem then is to consider how incorporation of dynamics and system coupling restrict the possibilities for an ontological model, and whether or not these restrictions rule out the possibility of a ((super) completely) ψ-epistemic model.

Chapter 4

Generalized Contextuality

In Section 2.3.1 we discussed the contextuality result of Kochen and Specker for quantum systems of dimension greater than 2. We found that an ontological model, under the assumption of outcome determinism, is forced to distinguish between *measurement contexts* when assigning indicator functions to effects in a PVM. In particular, the indicator function assigned to a projector must depend on the PVM in which the projector is included.

Spekkens [58] has generalized the notion of contextuality such that it can be stated as a property of any ontological model, not just for quantum theory. Moreover, the idea of the generalization naturally allows notions of *preparation and measurement contextuality*, while fully incorporating the contextuality shown by Kochen and Specker as a special case. To make the distinction, the contextuality of Kochen and Specker will be referred to as *traditional contextuality*. With this revised definition of contextuality, one can in fact show quantum theory is contextual in ways other than the traditional contextuality.

In Section 4.1 we present Spekkens' notions of generalized contextuality, along with his results on the necessity of this contextuality for quantum theory. The proofs are adapted to a more general notion of ontological model (wherein preparations map to measures) than presented in the original paper. In Section 4.2 we present a proof of contextuality for ψ-epistemic ontological models, which as far is we know is new and is not directly implied by any other results.

The framework presented by Spekkens has the effect of equalizing the contextuality put forth by Kochen and Specker and the new contextualities proven in [58]. There is some suspicion [27, 48] that the contextuality proven by Spekkens has a different quality than traditional contextuality. Thus in Section 4.3 we present new refinements of generalized contextuality which explicitly places traditional contextuality in a stronger position than the results on generalized contextuality.

In the discussion and future work section 4.4 we mention two communication tasks (Appendix D) for which quantum implementations have advantage over classical implementations. It has been shown that contextuality (traditional and gen-

eralized) can be considered as the source of these advantages. We discuss these results and the consequences for viewing contextuality as a quantum resource.

4.1 Spekkens' Generalized Contextuality and Results

This section describes the contextuality framework and results in [58].

The probabilities of an operational theory give rise to a definition of equivalence classes on the spaces of preparations and effects.

Definition 4.1. Given an operational theory $(\mathcal{P}, \mathcal{M}, I, \Pr)$, two preparations $P_1, P_2 \in \mathcal{P}$ are said to be equivalent if

$$\Pr(k|P_1, M) = \Pr(k|P_2, M) \quad \forall (M, k) \in \mathcal{M} \times I.$$

The set of preparations equivalent with a preparation P is denoted as $[P]$.

Two effects $(M_1, k_1), (M_2, k_2) \in \mathcal{M} \times I$ are said to be equivalent if

$$\Pr(k_1|P, M_1) = \Pr(k_2|P, M_2) \quad \forall P \in \mathcal{P}.$$

The set of effects equivalent to an effect (M, k) is denoted as $[(M, k)]$.

Statistically, it is clear that the equivalence class of a preparation or effect is all that is relevant. In quantum theory, a density matrix specifies the statistics corresponding to a preparation and a quantum effect specifies the statistics corresponding to a measurement outcome. Thus the equivalence class of a preparation is represented by a density matrix and the equivalence class of an operational effect is represented by a quantum effect. Any change in a preparation or effect which preserves the equivalence class is then a change in *context*.

Definition 4.2. The *context* of a preparation or effect is any specification, additional to the equivalence class, which determines the preparation or effect.

We can view the reference frame for an experiment with classical physics to be an example of a context. The reference frame of a laboratory has no effect on outcomes of experiments, thus if P_1 and P_2 are experiments which differ only in the reference frame of the experiment then they are in the same equivalence class. Thus the reference frame is an example of a preparation context.

With this definition of context in hand we can define preparation and measurement *non-contextuality* for an ontological model.

Definition 4.3 (Non-Contextuality of Ontological Models)**.** An ontological model $(\mathcal{P}, \mathcal{M}, I, \Pr)$ for an operational theory is said to be *preparation non-contextual* if

$$P_1, P_2 \in [P_1] \Rightarrow \mu_{P_1} = \mu_{P_2}. \tag{4.1}$$

The ontological model is said to be *measurement non-contextual* if

$$(M_1, k_1), (M_2, k_2) \in [(M_1, k_1)] \Rightarrow \xi_{M_1,k_1} = \xi_{M_2,k_2}. \tag{4.2}$$

Explicitly, an ontological model is preparation (measurement) non-contextual if a change in the context of the preparation (effect) has no effect on the representation in the ontological model. The property of non-contextuality is a bijection between statistically equivalent operational objects and the physical objects of the ontological model. The property of contextuality is then the negation of non-contextuality. An ontological model is contextual if the ontological representation of an operational object is dependent on the context of the operational object.

We can make a definition for contextuality of operational theories based on the necessity or non-necessity of contextuality for an ontological model.

Definition 4.4 (Contextuality of Operational Theories)**.** An operational theory is said to be *preparation (measurement) contextual* if any ontological model for the operational theory must be preparation (measurement) contextual.

We can see that the result of Kochen and Specker fits cleanly into this framework. Their result states that an ontological model for $\partial\mathbb{Q}_d$ which is outcome deterministic must be measurement contextual for $d \geq 3$. Indeed, their discussion is limited to pure states and PVMs and so theirs is a statement about models for $\partial\mathbb{Q}_d$. Discussion in Section 2.3.1 together with Lemma 2.2 showed the equivalence between the assumptions of Kochen and Specker and the framework of an outcome deterministic ontological model where a unique idempotent indicator function could be associated to every projector, independent of the PVM i.e. a non-contextual assignment of indicator functions to projectors.

Next we state and prove a lemma needed to demonstrate Spekkens' results concerning the necessity of preparation and measurement contextuality for $\mathbb{Q}_d$.

Definition 4.5. If an experimenter knows that one of two preparations, P_1 or P_2, has been prepared, and there is a measurement M which allows her to perfectly retrodict which one it was, then P_1 and P_2 are *perfectly distinguishable*.

As a quantum example, consider preparations corresponding to orthogonal pure quantum states ψ and $\psi^\perp$. If we perform any PVM which contains $|\,\psi\rangle\langle\,\psi\,|$ and $|\,\psi^\perp\rangle\langle\,\psi^\perp\,|$ then that PVM will distinguish between ψ and $\psi^\perp$ with a perfect success rate.

Lemma 4.1. *If two preparations are perfectly distinguishable then the intersection of the supports of their measures in an ontological model must be empty.*

Proof. Let P_1 and P_2 be perfectly distinguishable preparations, and let μ_1 and μ_2 be their ontological measures. We have $\operatorname{supp}\mu_1 \in \Sigma$ and $\operatorname{supp}\mu_2 \in \Sigma$. Thus $E = \operatorname{supp}\mu_1 \cap \operatorname{supp}\mu_2 \in \Sigma$ by the definition of a σ-algebra (Definition A.1). Suppose $E \neq$

$\emptyset$. Since $E \in \Sigma$ and $E \subseteq \operatorname{supp} \mu_1$ it must be that $\mu_1(E) > 0$ by the definition of supp (Definition A.7). Similarly we have $\mu_2(E) > 0$. Thus with non-zero probability, a preparation of P_1 puts the system in an ontic state $\lambda \in E$ and similarly for P_2. However, if the ontic state is in E, no measurement could determine whether or not the ontic state came about from a preparation of P_1 or from P_2. This contradicts the perfect distinguishability of P_1 and P_2, thus we must have $E = \emptyset$. □

4.1.1 The Necessity of Preparation Contextuality for Quantum Theory

Theorem 4.1. *Any convex operational theory for $\mathbb{Q}_d$ is preparation contextual for all $d \geq 2$.*

Proof. Consider first $\mathbb{Q}_2$. Let

$$\begin{aligned}
\psi_a &= \left| 0, \tfrac{\pi}{2} \right\rangle\!\!\left\langle 0, \tfrac{\pi}{2} \right| \\
\psi_A &= \left| \pi, \tfrac{\pi}{2} \right\rangle\!\!\left\langle \pi, \tfrac{\pi}{2} \right| \\
\psi_b &= \left| \tfrac{2}{3}\pi, \tfrac{\pi}{2} \right\rangle\!\!\left\langle \tfrac{2}{3}\pi, \tfrac{\pi}{2} \right| \\
\psi_B &= \left| -\tfrac{1}{3}\pi, \tfrac{\pi}{2} \right\rangle\!\!\left\langle -\tfrac{1}{3}\pi, \tfrac{\pi}{2} \right| \\
\psi_c &= \left| -\tfrac{2}{3}\pi, \tfrac{\pi}{2} \right\rangle\!\!\left\langle -\tfrac{2}{3}\pi, \tfrac{\pi}{2} \right| \\
\psi_C &= \left| \tfrac{1}{3}\pi, \tfrac{\pi}{2} \right\rangle\!\!\left\langle \tfrac{1}{3}\pi, \tfrac{\pi}{2} \right|,
\end{aligned} \tag{4.3}$$

and denote preparations which induce each state respectively as $P_a, P_A, P_b, P_B, P_c, P_C$. The states are depicted on a plane of the Bloch sphere in Figure 4.1. Let (Λ, μ, ξ) be an ontological model for $\mathbb{Q}_2$ and denote the measures corresponding to the listed preparations as $\mu_a, \mu_A, \mu_b, \mu_B, \mu_c$ and μ_C. Consider the following convex preparations:

$$\begin{aligned}
P_{aA} &= \frac{1}{2}P_a + \frac{1}{2}P_A \\
P_{bB} &= \frac{1}{2}P_b + \frac{1}{2}P_B \\
P_{cC} &= \frac{1}{2}P_c + \frac{1}{2}P_C \\
P_{abc} &= \frac{1}{3}P_a + \frac{1}{3}P_a + \frac{1}{3}P_c \\
P_{ABC} &= \frac{1}{3}P_A + \frac{1}{3}P_B + \frac{1}{3}P_C.
\end{aligned} \tag{4.4}$$

The quantum states corresponding to these preparations are all the maximally mixed state $\frac{\mathbb{I}}{2}$. We could verify this algebraically, or we can use the isomorphism between the qubit density matrices and the Bloch sphere to verify it visually in

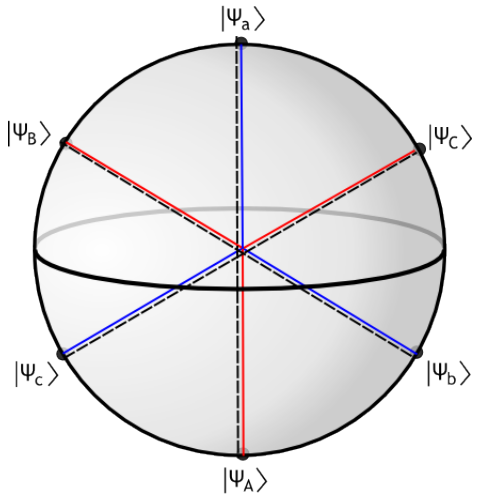

Figure 4.1: A Depiction of the six pure states used in the proofs of preparation and measurement contextuality. The ways in which the pure states mix to create the maximally mixed state $\frac{\mathbb{I}}{2}$ are depicted. The coloured triples form decompositions, as do each of the three dashed pairs.

Figure 4.1. The assumption of convexity gives us:

$$\begin{aligned}
\mu_{aA} &= \frac{1}{2}\mu_a + \frac{1}{2}\mu_A \\
\mu_{bB} &= \frac{1}{2}\mu_b + \frac{1}{2}\mu_B \\
\mu_{cC} &= \frac{1}{2}\mu_c + \frac{1}{2}\mu_C \\
\mu_{abc} &= \frac{1}{3}\mu_a + \frac{1}{3}\mu_a + \frac{1}{3}\mu_c \\
\mu_{ABC} &= \frac{1}{3}\mu_A + \frac{1}{3}\mu_B + \frac{1}{3}\mu_C.
\end{aligned} \tag{4.5}$$

Now assume that (Λ, μ, ξ) is preparation non-contextual. Since each preparation in (4.4) represents the same density matrix, namely $\frac{\mathbb{I}}{2}$, the assumption of preparation non-contextuality dictates that the measures in (4.5) are identical to each other. Denote this measure as μ. We will derive a contradiction by showing that μ must in fact be the zero measure.

Consider any $E \in \Sigma$. Define

$$E_{ijk} \equiv E \cap \operatorname{supp} \mu_i \cap \operatorname{supp} \mu_j \cap \operatorname{supp} \mu_k$$

where $i \in \{a, A\}$, $j \in \{b, B\}$, $k \in \{c, C\}$. Let

$$E' = E \backslash \bigcup_{ijk} E_{ijk} \quad \text{such that} \quad E = E' \cup \bigcup_{ijk} E_{ijk}.$$

Since E' is not in the support of μ_a or μ_A, we have that $\mu_{aA}(E') = 0$, implying $\mu(E') = 0$. Now we consider the case of E_{ijk} for the case when all three of i, j, k are lower case, and the case where one is upper and the other two are lower case. The other cases follow by symmetry. For E_{abc}, we have that $\mu_A(E_{abc}) = 0$, $\mu_B(E_{abc}) = 0$, and $\mu_C(E_{abc}) = 0$. Therefore $\mu_{ABC}(E_{abc}) = 0$ and so $\mu(E_{abc}) = 0$. Next, for E_{Abc} we have $\mu_a(E_{Abc}) = 0$, $\mu_B(E_{Abc}) = 0$, and $\mu_C(E_{Abc}) = 0$. Thus

$$\tfrac{1}{2}\mu_A(E_{Abc}) = \mu_{aA}(E_{Abc}) = \mu(E_{Abc}) = \mu_{ABC}(E_{Abc}) = \tfrac{1}{3}\mu_A(E_{Abc})$$

which implies that $\mu(E_{Abc}) = 0$. Thus

$$\mu(E) = \mu(E' \cup \bigcup_{ijk} E_{ijk}) = \mu(E') + \sum_{ijk} \mu(E_{ijk}) = 0.$$

Since E was arbitrary, we have $\mu = 0$, contradicting that it must be a probability measure.

Lastly, since any quantum system of dimension $d > 2$ has a 2-dimensional subsystem on which the above argument can be made, the theorem holds. □

4.1.2 The Necessity of Measurement Contextuality for Quantum Theory

The following are the results of Spekkens proving the necessity of measurement contextuality in outcome deterministic ontological models of quantum theory, as well as a result showing that preparation non-contextuality implies outcome determinism.

We begin with a small lemma concerning restrictions on indicator functions corresponding to orthogonal projectors.

Lemma 4.2. *Let ψ and $\psi^\perp$ be pure orthogonal quantum states, with μ_ψ and $\mu_{\psi^\perp}$ their representations in an ontological model. Let ξ_ψ be any indicator function reproducing the statistics of the effect associated with $|\,\psi\rangle\langle\psi\,|$, and let $\xi_{\psi^\perp}$ be any indicator function reproducing the statistics of the effect associated with $\left|\,\psi^\perp\right\rangle\!\left\langle\psi^\perp\,\right|$. Then*

$$\xi_\psi(\lambda) = \begin{cases} 1 & \lambda \in \operatorname{supp}\mu_\psi \\ 0 & \lambda \in \operatorname{supp}\mu_{\psi^\perp} \\ ? & \textit{else} \end{cases}, \qquad \xi_{\psi^\perp}(\lambda) = \begin{cases} 1 & \lambda \in \operatorname{supp}\mu_{\psi^\perp} \\ 0 & \lambda \in \operatorname{supp}\mu_\psi \\ ? & \textit{else} \end{cases}. \tag{4.6}$$

Proof. In order for μ_ψ, $\mu_{\psi^\perp}$, ξ_ψ, and $\xi_{\psi^\perp}$ to reproduce the perfect distinguishability of ψ and $\psi^\perp$, they need to satisfy

$$\begin{aligned} \int_\Lambda \xi_\psi(\lambda) d\mu_\psi(\lambda) &= 1 & \int_\Lambda \xi_\psi(\lambda) d\mu_{\psi^\perp}(\lambda) &= 0 \\ \int_\Lambda \xi_{\psi^\perp}(\lambda) d\mu_\psi(\lambda) &= 0 & \int_\Lambda \xi_{\psi^\perp}(\lambda) d\mu_{\psi^\perp}(\lambda) &= 1. \end{aligned} \tag{4.7}$$

Since each integral above can be reduced to an integral over the support of the measure, and by Lemma 4.1 we know that $\operatorname{supp}\mu_\psi$ and $\operatorname{supp}\mu_{\psi^\perp}$ are disjoint, we arrive immediately at the conditions (4.6). □

This is not a proof of outcome determinism. In order for this result to imply outcome determinism, we would need that for any orthonormal basis $\{\psi_i\}_{i=1}^d$ for $\mathbb{C}^d$, $\bigcup_i \operatorname{supp}\mu_{\psi_i} = \Lambda$ (this would erase the 'else' conditions in (4.6)). However, this is true if we assume preparation non-contextuality.

Lemma 4.3. *If a convex ontological model for $\mathbb{Q}_d$ is preparation non-contextual, then for any orthonormal basis $\{\psi_i\}_{i=1}^d$ for $\mathbb{C}^d$, $\bigcup_i \operatorname{supp}\mu_{\psi_i} = \Lambda$.*

Proof. It is possible that we could define Λ such that there exist elements or subsets of Λ that lie in the support of no measures corresponding to preparations. For simplicity, we will assume that $\Lambda = \cup_{P\in\mathcal{P}} \operatorname{supp}\mu_P$.

In quantum theory, if ρ is any density matrix, then ρ appears in a convex decomposition of the maximally mixed state $\frac{\mathbb{I}}{d}$. For example, suppose $\rho = \sum_{i=1}^d a_i \,|\, i \,\rangle\langle\, i \,|$ for some orthonormal basis $\{|\, i \,\rangle\langle\, i \,|\}_{i=1}^d$. Without loss of generality, suppose $a_1 = \max_i a_i$. Then the decomposition

$$\tfrac{1}{da_1}\rho + \sum_{i=2}^{d} \tfrac{a_1 - a_i}{da_1} \,|\, i \,\rangle\langle\, i \,| = \sum_{i=1}^{d} \tfrac{1}{d} \,|\, i \,\rangle\langle\, i \,| = \tfrac{\mathbb{I}}{d},$$

is a convex decomposition of $\frac{\mathbb{I}}{d}$. Therefore, in a convex model, it must be the case that $\Lambda = \operatorname{supp}\frac{\mathbb{I}}{d}$. However, if $\{\psi_i\}_{i=1}^d$ is an orthonormal basis for $\mathbb{C}^d$, then

$$\sum_{i=1}^{d} \frac{1}{d}\psi_i = \tfrac{\mathbb{I}}{d}.$$

Thus if the ontological model is convex and preparation non-contextual, we have

$$\mu_{\frac{\mathbb{I}}{d}} = \sum_{i=1}^{d} \frac{1}{d}\mu_{\psi_i}$$

and hence $\bigcup_i \operatorname{supp}\mu_{\psi_i} = \Lambda$. □

These two lemmas combine to give

Theorem 4.2. *Preparation non-contextuality in a convex ontological model for $\mathbb{Q}_d$ imply outcome determinism.*

Given that the preparation non-contextuality implies outcome determinism, the following theorem can be considered as a proof that quantum theory is preparation contextual, if not measurement contextual.

Theorem 4.3. *Any convex and outcome deterministic ontological model for $\mathbb{Q}_d$ is measurement contextual.*

Proof. As with the proof of preparation contextuality, we give a proof for $\mathbb{Q}_2$ which can be used in any quantum system of higher finite dimension.

We again use the six pure states given in the proof of preparation contextuality (4.3). Denote the PVM comprised of projectors $\{\psi_a, \psi_A\}$ as M_a, and similarly define M_b and M_c. Denote the indicator functions corresponding to the relevant effects as $\{\xi_a, \xi_A\}$ and so forth.

Define the mixed measurement M in the following way. With uniform probability we choose to perform M_a, M_b or M_c, and record only whether or not the outcome corresponds to an upper or lower case letter. The quantum effects for this measurement are

$$\left\{\tfrac{1}{3}\psi_a + \tfrac{1}{3}\psi_b + \tfrac{1}{3}\psi_c, \tfrac{1}{3}\psi_A + \tfrac{1}{3}\psi_B + \tfrac{1}{3}\psi_C\right\} = \left\{\tfrac{\mathbb{I}}{2}, \tfrac{\mathbb{I}}{2}\right\}. \tag{4.8}$$

Under the assumption of convexity, the indicator functions for these effects are

$$\left\{\tfrac{1}{3}\xi_a + \tfrac{1}{3}\xi_b + \tfrac{1}{3}\xi_c, \tfrac{1}{3}\xi_A + \tfrac{1}{3}\xi_B + \tfrac{1}{3}\xi_C\right\}.$$

At this point, we argue simply that the indicator functions for the measurement with quantum effects $\left\{\frac{\mathbb{I}}{2}, \frac{\mathbb{I}}{2}\right\}$ must both be equal to $\frac{1}{2}$ everywhere. Indeed, for any state ρ, we have that $\mathrm{tr}(\rho\frac{\mathbb{I}}{2}) = \frac{1}{2}$. Thus we can perform this measurement by completely ignoring the system and simply choosing an outcome from a uniform binary distribution. Then for any ontic state λ of the system, it must be that the probability of the first outcome is $\frac{1}{2}$ and the probability of the second outcome is $\frac{1}{2}$. Thus the indicator functions for $\left\{\frac{\mathbb{I}}{2}, \frac{\mathbb{I}}{2}\right\}$ are $\left\{\frac{1}{2}, \frac{1}{2}\right\}$.

Assume that the ontological model is measurement non-contextual. Thus by (4.8) and the fact that the indicator function for the effect $\frac{\mathbb{I}}{2}$ is the constant function $\frac{1}{2}$, we get

$$\tfrac{1}{3}\xi_a + \tfrac{1}{3}\xi_b + \tfrac{1}{3}\xi_c = \frac{1}{2}$$
$$\tfrac{1}{3}\xi_A + \tfrac{1}{3}\xi_B + \tfrac{1}{3}\xi_C = \frac{1}{2}.$$

However, it is impossible to satisfy these equations if each of the $\xi_a, \xi_A, \xi_b, \xi_B, \xi_c, \xi_C$ are idempotent. □

The results of Spekkens [58] show that generalized measurement contextuality is a necessary feature of an outcome deterministic model for $\mathbb{Q}_d$, $d \geq 2$. We know that outcome determinism is necessary for this result, as the Beltrametti-Bugajski model is a clear counter-example of an outcome in-deterministic model which is measurement non-contextual. To see this, note equation (2.13), which specifies the indicator functions of the Beltrametti-Bugajski model. Indeed, the indicator function is dependent only on *the quantum effect* E_m, and thus can have no dependence on any other context. The conclusion is that while an ontological model for quantum theory must always be preparation contextual, it can avoid being measurement contextual by not being outcome deterministic.

4.2 The Necessity of Measurement Contextuality for ψ-epistemic Theories

Here we demonstrate that by dropping the requirement of outcome determinism, and adding another desirable property, namely ψ-epistemicity, we can prove the necessity of measurement contextuality.

Theorem 4.4. *A convex ψ-epistemic model for $\mathbb{Q}_d$ is measurement contextual for all $d \geq 2$.*

Proof. Consider the Hilbert space $\mathbb{C}^2$, and let $\psi_1 = |\,\theta, \phi\,\rangle\langle\,\theta, \phi\,|$ be any pure state. By the definition of ψ-epistemic, we know that there exists a ψ_2 such that the measures μ_{ψ_1} and μ_{ψ_2} are overlapping (3.5).

Let P_{ψ_1} and P_{ψ_2} be the points on the Bloch sphere corresponding to ψ_1 and ψ_2. Let P_{ψ_3} be any point on the Bloch sphere such that the center (representing the density matrix $\frac{\mathbb{I}}{2}$) is in the convex hull of $\{P_{\psi_1}, P_{\psi_2}, P_{\psi_3}\}$. For simplicity, and without loss of generality, let us suppose that $\psi_1 = \left|\,\frac{\theta}{2}, \frac{\pi}{2}\,\right\rangle$, $\psi_2 = \left|\,-\frac{\theta}{2}, \frac{\pi}{2}\,\right\rangle$. Then the possible pure states ψ_3 which give rise to a valid P_{ψ_3} are $\psi_3 = \left|\,\gamma, \frac{\pi}{2}\,\right\rangle$, for $\pi - \frac{\theta}{2} < \gamma < \pi + \frac{\theta}{2}$. The situation is depicted in Fig 4.2.

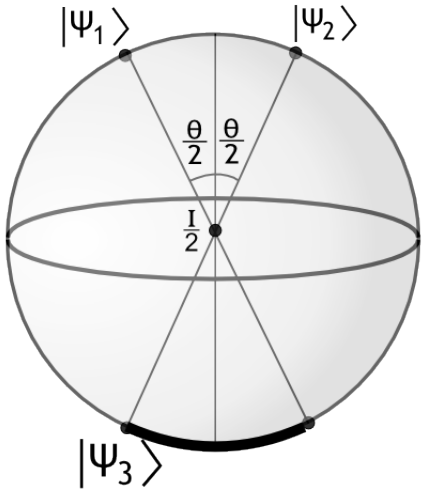

Figure 4.2: A depiction of the states ψ_1 and ψ_2 and the range of possible states for ψ_3.

A straight-forward use of trigonometry yields the unique convex decomposition

for $\frac{\mathbb{I}}{2}$ in terms of $\{P_{\psi_1}, P_{\psi_2}, P_{\psi_3}\}$:

$$\begin{aligned}
\tfrac{\mathbb{I}}{2} &= a_1 P_{\psi_1} + a_2 P_{\psi_2} + a_3 P_{\psi_3} \\
a_1 &= \frac{\sin\left(\frac{\theta}{2} - \gamma\right)}{\sin(\theta) + \sin\left(\frac{\theta}{2} + \gamma\right) + \sin\left(\frac{\theta}{2} - \gamma\right)} \\
a_2 &= \frac{\sin\left(\frac{\theta}{2} + \gamma\right)}{\sin(\theta) + \sin\left(\frac{\theta}{2} + \gamma\right) + \sin\left(\frac{\theta}{2} - \gamma\right)} \\
a_3 &= \frac{\sin(\theta)}{\sin(\theta) + \sin\left(\frac{\theta}{2} + \gamma\right) + \sin\left(\frac{\theta}{2} - \gamma\right)}.
\end{aligned} \tag{4.9}$$

Now consider the three PVMs corresponding to the unique bases defined by ψ_1, ψ_2 and ψ_3 (and their orthogonal partners $\bar{\psi}_1$, $\bar{\psi}_2$ and $\bar{\psi}_3$). We denote these measurements as

$$\begin{aligned}
M_1 &= \left\{E_{\psi_1}, E_{\bar{\psi}_1}\right\} \\
M_2 &= \left\{E_{\psi_2}, E_{\bar{\psi}_2}\right\} \\
M_3 &= \left\{E_{\psi_3}, E_{\bar{\psi}_3}\right\}.
\end{aligned} \tag{4.10}$$

The E_{ψ_i} outcome for each will be considered as outcome $k = 0$, and the $E_{\bar{\psi}_i}$ outcome as $k = 1$. Consider the measurement M whereby measurement M_i is performed with probability a_i, and the outcome 0 or 1 is registered. The effects for these outcomes are

$$\begin{aligned}
E_0 &= a_1 E_{\psi_1} + a_2 E_{\psi_2} + a_3 E_{\psi_3} = \tfrac{\mathbb{I}}{2} \\
E_1 &= a_1 E_{\bar{\psi}_1} + a_2 E_{\bar{\psi}_2} + a_3 E_{\bar{\psi}_3} = \tfrac{\mathbb{I}}{2}.
\end{aligned} \tag{4.11}$$

The assumption of convexity gives

$$\begin{aligned}
\xi_{M,0} &= a_1 \xi_{M_1,0} + a_2 \xi_{M_2,0} + a_3 \xi_{M_3,0} \\
\xi_{M,1} &= a_1 \xi_{M_1,1} + a_2 \xi_{M_2,1} + a_3 \xi_{M_3,1}.
\end{aligned} \tag{4.12}$$

However, since $E_0 = E_1 = \frac{\mathbb{I}}{2}$, the statistics for M are equivalent to those of the experiment M' where we simply randomly pick between outcome 0 and outcome 1. The indicator functions for this experiment must be $\xi_{M',0} = \xi_{M',1} = \frac{1}{2}$. Therefore, the assumption of non-contextuality dictates that for all $\lambda \in \Lambda$

$$\begin{aligned}
\xi_{M,0}(\lambda) &= a_1 \xi_{M_1,0}(\lambda) + a_2 \xi_{M_2,0}(\lambda) + a_3 \xi_{M_3,0}(\lambda) = \frac{1}{2} \\
\xi_{M,1}(\lambda) &= a_1 \xi_{M_1,1}(\lambda) + a_2 \xi_{M_2,1}(\lambda) + a_3 \xi_{M_3,1}(\lambda) = \frac{1}{2}.
\end{aligned} \tag{4.13}$$

Now, by the definition of ψ-epistemic, we know that there exists a measure μ such that μ_{ψ_1} and μ_{ψ_2} are absolutely continuous with respect to μ. Let $\mu_{\psi_i}(\lambda)$ be functions such that $d\mu_{\psi_i} = \mu_{\psi_i}(\lambda) d\mu$. Now let S be a subset of Λ such that the $\mu_{\psi_1}(\lambda^*) > 0$ and $mu_{\psi_2}(\lambda^*) > 0$, for all $\lambda^* \in S$. Since $\lambda^* \in S$, we must have

$\xi_{M_1,0}(\lambda^*) = \xi_{M_2,0}(\lambda^*) = 1$, and also $\xi_{M_1,1}(\lambda^*) = \xi_{M_2,1}(\lambda^*) = 0$. Subbing these equalities into (4.13) gives

$$\begin{aligned} \xi_{M,0}(\lambda^*) &= a_1 + a_2 + \quad a_3\xi_{M_3,0}(\lambda^*) = \frac{1}{2} \\ \xi_{M,1}(\lambda^*) &= \qquad\qquad a_3\xi_{M_3,1}(\lambda^*) = \frac{1}{2}. \end{aligned} \tag{4.14}$$

Invoking $\xi_{M_3,0} + \xi_{M_3,1} = 1$ (2.8) and subbing in values for the a_i (4.9) allows us to solve for $\xi_{M_3,0}(\lambda)$:

$$\xi_{M_3,0}(\lambda^*) = \frac{\sin(\theta) - \sin(\frac{\theta}{2} - \gamma) - \sin(\frac{\theta}{2} + \gamma)}{2\sin(\theta)}. \tag{4.15}$$

However, when $0 < \theta < \pi$ and $-\frac{\theta}{2} < \gamma < \frac{\theta}{2}$,

$$\begin{aligned} &\sin(\theta) - \sin(\tfrac{\theta}{2} - \gamma) - \sin(\tfrac{\theta}{2} + \gamma) \\ =&\sin(\theta) - \sin(\tfrac{\theta}{2})\cos(\gamma) + \cos(\tfrac{\theta}{2})\sin(\gamma) - \sin(\tfrac{\theta}{2})\cos(\gamma) - \cos(\tfrac{\theta}{2})\sin(\gamma) \\ =&\sin(\theta) - 2\sin(\tfrac{\theta}{2})\cos(\gamma) \\ <&\sin(\theta) - 2\sin(\tfrac{\theta}{2})\cos(\tfrac{\theta}{2}) \\ =&\sin(\theta) - \sin(\theta) \\ =&0. \end{aligned} \tag{4.16}$$

Thus $\xi_{M_3,0}(\lambda^*)$ is negative, contradicting the fact that $\xi_{M_3,0}$ is assumed to be an indicator function. □

4.3 Strong and Weak Notions of Contextuality

The generalized notion of contextuality is useful indeed for capturing traditional contextuality within a framework which allows for other forms of contextuality. However, we wish to present a refinement of this framework which makes more explicit the relationship between traditional contextuality (Section 2.3.1) and the contextuality proven by Spekkens (Section 4.1).

To start, we redefine the generalized notions of contextuality as *weak contextuality*.

Definition 4.6 (Weak Preparation and Measurement Contextuality)**.** Let (Λ, μ, ξ) be a convex ontological model. If there exist two preparations $P_1, P_2 \in [P_1]$ such that $\mu_{P_1} \neq \mu_{P_2}$, then we say the model is *weakly contextual for preparations*.

If there exist two effects $(M_1, k_1), (M_2, k_2) \in [(M_1, k_1)]$ and $\xi_{M_1,k_1} \neq \xi_{M_2,k_2}$, then we say the model is *weakly measurement contextual*.

Notice that we have explicitly defined weak contextuality to occur in convex ontological models. The idea is that weak contextuality can occur for mixed preparations and within mixed measurements, as well as possibly for the extreme preparations and effects. Our definition of strong contextuality only allows for contextuality that occurs for the extreme preparations and effects.

Definition 4.7 (Strong Preparation and Measurement Contextuality)**.** If there exists two extreme preparations $P_1, P_2 \in \operatorname{ext} \mathcal{P}$ in the same equivalence class, and $\mu_{P_1} \neq \mu_{P_2}$, then we say the model is *strongly contextual for preparations*.

If $(M_1, k_1), (M_2, k_2) \in \operatorname{ext}(\mathcal{M} \times I)$ are in the same equivalence class, and $\xi_{M_1,k_1} \neq \xi_{M_2,k_2}$, then we say the model is *strongly measurement contextual*.

The intention of this distinction is to capture the relationship between traditional contextuality and the preparation and measurement contextuality proven by Spekkens and in Section 4.2. Indeed, traditional contextuality is a strong contextuality, as it requires the indicator functions for *projectors* to be context-dependent in an outcome deterministic model. The proofs of Spekkens as well as the new proof in Section 4.2 make explicit use of mixed preparations and measurements in order to prove contextuality. Thus it is clear that these latter results show a necessity of *weak* contextuality in quantum theory, but not *strong* contextuality.

Another important reason for this distinction is the existence of the Kochen-Specker quibit model (see Section 3.2), which is weakly measurement contextual, but not strongly measurement contextual. Recall that we can extend it to accomodate mixed measurements. We know from Section 4.1.2 that any model which accounts for the mixed measurements of quantum theory must exhibit weak measurement contextuality. However, it is not strongly measurement contextual. In particular, the indicator functions for the PVM measurements depend only on projector they are representing (3.7).

We can explicitly state the discussed contextuality results within this refined framework.

- Spekkens: Any convex ontological model for $\mathbb{Q}_d$, $d \geq 2$ is weakly preparation contextual .
- Spekkens: Any convex outcome deterministic model for $\mathbb{Q}_d$, $d \geq 2$ is weakly measurement contextual.
- Kochen-Specker: Any outcome deterministic model for $\mathbb{Q}_d$, $d \geq 3$ is strongly measurement contextual.
- Section 4.2: Any ψ-epistemic convex model for $\mathbb{Q}_d$, $d \geq 2$ is weakly measurement contextual.

The fact that strong (preparation/measurement) contextuality implies weak (preparation/measurement) contextuality is trivial from the definitions, and thus

the nomenclature seems appropriate. However, we can show explicitly that strong contextuality implies contextuality for mixed preparations or measurements (i.e contextuality on ext $\mathcal{P}$ or ext$(\mathcal{M} \times I)$ implies contextuality on the interior of $\mathcal{P}$ or (M, k)).

Proposition 4.1. *In a convex ontological model, strong preparation contextuality arising on the extreme preparations implies contextuality for mixed preparations.*

Proof. Suppose (Λ, μ, ξ) is convex and strongly preparation contextual. Thus $\exists P_1, P_2 \in [P] \in \text{ext}\,\mathcal{P}$ such that $\mu_{P_1} \neq \mu_{P_2}$. Take $P \notin [P]$ and consider two non-extreme preparations:

$$P_1' = \tfrac{1}{2}P_1 + \tfrac{1}{2}P \qquad P_2' = \tfrac{1}{2}P_2 + \tfrac{1}{2}P. \tag{4.17}$$

By the convexity assumption,

$$\begin{aligned} \mu_{P_1'} &= \tfrac{1}{2}\mu_{P_1} + \tfrac{1}{2}\mu_P \\ \mu_{P_2'} &= \tfrac{1}{2}\mu_{P_2'} + \tfrac{1}{2}\mu_P. \end{aligned} \tag{4.18}$$

But since $\mu_{P_1} \neq \mu_{P_2}$ by assumption, we have $\mu_{P_1'} \neq \mu_{P_2'}$. □

Proposition 4.2. *In a convex ontological model, strong measurement contextuality implies contextuality for the effects in mixed measurements.*

Proof. Suppose $(\Lambda, \mu, \xi) \in \mathbb{O}_{conv}$ is strongly measurement contextual. Thus

$$\begin{aligned} &\exists (M_1, k_1), (M_2, k_2) \in [M, k] \in \text{ext}(\mathcal{M} \times I) \\ &\text{such that} \quad \xi_{M_1,k_1} \neq \xi_{M_2,k_2} \ \text{ and } M_1 \neq M_2. \end{aligned} \tag{4.19}$$

Let M_3 be any third measurement distinct from M_1 and M_2. Consider the measurement M_1' (and M_2') where the first step is to pick between between measurement M_3 and measurement M_1 (respectively M_2). If measurement M_3 is performed, any outcome is relabeled as outcome k_1 (respectively k_2). If measurement M_1 (respectively M_2) is performed, the outcome is stated as is. The probability of outcome k_1 for M_1' is

$$\Pr(k_1|M_1', P) = \tfrac{1}{2}\Pr(k_1|M_3, P) + \tfrac{1}{2}\Pr(k_1|M_1, P) = \tfrac{1}{2}(1 + \Pr(k_1|M_1, P)) \quad \forall P \in \mathcal{P}. \tag{4.20}$$

Since $\Pr(k_2|M_2, P) = \Pr(k_1|M_1, P)$, the probability of outcome k_2 for M_2' will be the same, thus the effects (M_1', k_1) and (M_2', k_1) are equivalent. Recall that condition (2.8) states effectively that for any ontic state, an outcome must occur for any measurement. Thus if we have a measurement with only one outcome, the representation of this effect in the ontological model must be the identity function $1(\lambda) = 1$. Thus by the assumption of convexity we have

$$\begin{aligned} \xi_{M_1',k_1} &= \tfrac{1}{2}1(\lambda) + \tfrac{1}{2}\xi_{M_1,k_1}, \\ \xi_{M_2',k_2} &= \tfrac{1}{2}1(\lambda) + \tfrac{1}{2}\xi_{M_2,k_2}. \end{aligned} \tag{4.21}$$

Hence $\xi_{M_1',k_1} \neq \xi_{M_2',k_2}$ since $\xi_{M_1,k_1} \neq \xi_{M_2,k_2}$. □

4.3.1 Discussion and Future Work

An important point to make is that strong preparation contextuality may possibly be a feature of an ontological model, but it is never the case that an operational theory *must* have an strongly preparation contextual ontological model, given the ontological model framework presented in Section 2.2. The proof of this is elementary.

Proposition 4.3. *Any operational theory has an ontological model which is not strongly preparation contextual.*

Proof. Suppose that (Λ, μ, ξ) is a strongly preparation contextual ontological model for an operational theory. For each $[P] \in [\mathcal{P}]$, choose a preferred representative $P' \in [P]$. For every $P \in [P]$, define $\tilde{\mu}_P = \mu_{P'}$. Replacing μ with $\tilde{\mu}$ in (Λ, μ, ξ) gives an ontological model for operational theory which is not strongly preparation contextual. □

The corresponding statement for effects is not true, as is evidenced by the Kochen-Specker theorem. We can see this asymmetry as arising from the fact that effects occur within the structure of a measurement, which imposes the restriction (2.8) in an ontological model. If an extreme effect can occur within the context of two statistically distinct measurements M_1 and M_2, then the functions ξ_{M_1,k_1} and ξ_{M_2,k_2} will have to satisfy (2.8) with distinct sets of other indicator functions. This requirement does indeed then dictate the necessity of strong measurement contextuality in an outcome deterministic ontological model for quantum theory. Without this requirement, one could easily present a result analogous to Proposition 4.3 for measurements.

We may find that if we were to extend the idea of an operational theory/ontological model to account for transformations and compositions of systems, this may impose additional restrictions on the representations of extreme preparations and effects. This may allow us to find other ways for strong contextuality to arise in models for quantum theory. It may even allow for a form of strong preparation contextuality. As an example of another type of strong contextuality that arises from extending the scope of operational theories and ontological models, Westman [64] has shown that if we consider sequences of measurements, then there is a strong contextuality based on the ordering of measurements for commuting observables. The general problem of extending the ontological model framework and seeking other forms of contextuality constitutes a possible avenue of future research.

4.4 Quantum Advantages Derived from Contextuality

[1]It has been argued [58, 59, 24] that contextuality is at the heart of some information theoretical advantages that quantum theory has over classical theory. In particular,

Galvão [24] has suggested that a *quantum random access code* (QRAC) protocol performs better than its classical counterpart due to the contextuality demonstrated by Kochen and Specker. Spekkens and collaborators [59] have proven that preparation contextuality is at the heart of a quantum advantage for the *Parity-Oblivious MultiPlexing* (POMP) protocol. See Appendix D for a description of these protocols and arguments for the role of contextuality in the advantage for quantum implementations.

The analysis of the POMP protocol involves the derivation of a *contextuality* upper bound on the success rate of a preparation non-contextual operational theory used to implement the protocol[2]. This contextuality inequality is shown to be identical to an upper bound on the success rate of any *classical operational theory*. It is then shown that there exists a quantum protocol involving a qubit system which beats these bounds.

Thus if an operational protocol beats a 'classical' bound for POMP, then any ontological model for it must be preparation contextual. This logically leads one immediately to conclude that preparation contextuality is a 'non-classical' feature. However, there is confusion if one considers the fact that the quantum protocol for POMP (as well as for QRAC) is completely implementable with $\mathbb{Q}_2$, as is described in Appendix D. As we have seen in Section 3.2, $\mathbb{Q}_2$ is implementable via the Kochen-Specker (KS) model for a qubit, which we have argued is entirely implementable with classical angular momenta coupled with a source of randomness. Thus, theoretically, one could devise an experiment with classical physics which reproduces the improved success rates of both QRAC and POMP.

This thought process leads us to conclude that although a protocol could be carried out with classical physics, this does not imply that a quantum system is not advantageous. Why is this? Suppose we construct devices which prepare and measure a classical angular momentum according to the KS model for a qubit. We could give these *classical* devices to A and B and they could proceed to perform the optimal *quantum* POMP protocol. However, this does not quite capture the entire point of POMP. The idea behind POMP is that even if B *wanted* to be able to learn parity information, he should not be able to. Realistically, if B receives a system from A prepared by her classical device, he could choose not to feed it into his device, and instead just measure the orientation of the angular momentum, and thus gain information on the parity of A's message. In a quantum system, the parity-oblivious restriction is *built in*. This fundamental restriction is what, in this example, makes the qubit more powerful than the classical KS system.

However, this argument can be leveled at other examples of quantum features which can be seen as arising from epistemic restrictions. For example, in [59] a

[1] Appendix D is required reading for this discussion

[2] Recall that an operational theory involves a set of procedures that can be carried out, and thus any physical implementation of a protocol, whether it be a classical or quantum implementation (or something else!) stipulates an operational theory. Then as per Definition 4.4, the operational theory is preparation non-contextual if there exists an ontological model for the theory which is preparation non-contextual.

toy model is described for which the elementary system is a ball in one of four boxes. The maximal state of knowledge is to know that the ball is in one of two boxes (this makes the theory epistemic in nature). Within this toy theory, one can prove a no-cloning theorem in analogy to the no-cloning theorem in quantum theory [49][3]. However, since the epistemic restriction on ball location is ad-hoc, in an actual physical implementation of this theory, one could easily clone a state by determining the exact location of the ball and then creating another system with the ball in that location.

Thus it would seem as though the advantage gained by a quantum system for POMP is analogous to the property of no-cloning: both can be seen as arising from an appropriate epistemic restriction on a classical model. One goal of the toy model paper [59] was to categorize quantum phenomena based whether or not they can arise from an epistemic restriction. It would seem that quantum advantages for POMP and QRAC should fall into this category, despite the fact that they are viewed [60] as arising from contextuality, where contextuality has been considered to be independent of epistemic restrictions. The question of the full relationship between epistemic restrictions and instances of contextuality looks to be open, and could be the focus of future research.

[3]The no-cloning theorem for quantum theory states that given an arbitrary unknown pure state ψ on one system and a known state ψ' on another, there is no operation on the joint system which is guaranteed to result in the state $\psi' \otimes \psi'$ for all ψ. This has been generally thought to be a uniquely quantum result.

Chapter 5

Distinguishability as a Resource

In information theory it is useful to ask how similar or different two states are. Suppose party A wishes to convey one of two messages, m_1 or m_2, to party B, but along the way the message is subject to a known noise function, Φ. This induces a state of uncertainty on the true value of the received message. In effect, party B receives $\Phi(m_1)$ or $\Phi(m_2)$, probability distributions over possible messages. B's ability to determine the intended message is related to B's ability to distinguish the probability distributions $\Phi(m_1)$ and $\Phi(m_2)$. As quantum theory is a fundamentally probabilistic theory, questions about distinguishing quantum states are natural extensions of questions about distinguishing probability distributions.

The goal of this research is to study distinguishability as a *resource theory*. A study of resources is a common feature of information theory, both classical and quantum. In general, we can think of a resource as anything that might be useful for some sort of elementary task. In classical information theory a prime example is Shannon theory, and in quantum information theory, entanglement is the most studied resource.

A major theorem in classical information theory is *Shannon's noiseless coding theorem* [57]. Suppose we have a source of text, m, and the relative frequency of the alphabet in m is given by a probability distribution $\vec{p}$. Shannon's theorem states that source text can be compressed and decompressed without losing any data in m, and the average number bits needed to encode a character in m is given by $H(\vec{p}) = -\sum_j p_j \log p_j$, which is called the Shannon entropy of $\vec{p}$.

This example highlights key components of the study of resources. Two resources exist: the source message m, and classical bits. The utility of a message is obvious, and the bit is clearly useful as a mode for conveying or communicating the message, or other information. Shannon's theorem is ultimately a statement about inter-conversion between resources, a key component of resource study. It dictates the number of classical bits needed such that m can be converted to classical bits and back again, without losing any of m. This question of reversible convertibility is common in resource theories, and we will examine it in our study of distinguishability. Implicit in the discussion of inter-convertibility is a quantification of *how*

much of each resource is present. In particular, the measure on classical bits is the *average number of classical bits per character* needed. There is even an implicit measure of *how much* with regards to m. The theorem states that m remains *unaltered* by the compression/decompression process, and thus the process should leave any measure of m invariant. Suppose, however, that the number of bits required to compress/decompress m flawlessly is not available. Then one would want to quantify in what way the compression/decompression procedure alters the utility of m. Studying quantifications of distinguishability will also be a major component of our analysis.

Many other examples of resource theories exist. For example, Spekkens and Gour [26] make note of the fact that any restriction on performable quantum operations gives rise to a resource theory in which the resource is any state that could be prepared were the restriction not in place. There is also a quantum equivalent of Shannon's noiseless coding theorem [56].

We will be using entanglement as our main example of a resource theory. We begin with an exposition on entanglement theory, in which we will make note of the main structures and features. Having done this, we will begin a development of distinguishability as a resource theory and point out the many analogies to entanglement theory. In the course of this development, we pick out two problems for analysis: a question of inter-convertibility between quantum and classical states; and a calculation of a particular distinguishability measure, the trace distance of formation.

5.1 Entanglement - A Resource Theory

In this section we will be describing entanglement on bipartite systems. In general, this exposition follows from the summary of entanglement theory given by Plenio and Virmani [51]. We begin with a definition of an entangled state:

Definition 5.1. If a bipartite quantum state on the Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ has the form

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i \tag{5.1}$$

where $\rho_A^i \in K(\mathcal{H}_A)$ and $\rho_B^i \in K(\mathcal{H}_B)$ and the p_i form a probability vector, then we call the state *separable*. Any state on $\mathcal{H}_{AB}$ which is not separable is called *entangled*. In particular, if a pure state cannot be expressed as a product state $|\,\psi_A\,\rangle \otimes |\,\psi_B\,\rangle$, then it is entangled.

Entanglement is viewed as a resource which is beneficial for the performance of various quantum information and communication tasks. The most common benefit is derived from two separated parties sharing an entangled state between them. One of the most elementary examples is the teleportation protocol [10], which we now describe.

5.1.1 Teleportation, LOCC and ebits

Suppose a party A wishes to send an unknown qubit pure state $|\psi\rangle$ to party B, but no quantum channel exists between them for communicating quantum states. The teleportation protocol allows for B to 'receive' the state $|\psi\rangle$ as long as A and B share an entangled 'Bell state' between them, and A is allowed to communicate two classical bits to B. The Bell states are the following four bipartite pure states:

$$\left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{5.2}$$

$$\left|\Phi^-\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \tag{5.3}$$

$$\left|\Psi^+\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \tag{5.4}$$

$$\left|\Psi^-\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \tag{5.5}$$

These four states form an orthonormal basis for $\mathbb{C}^4$, and as such they correspond to a PVM. Suppose that A and B share the state $|\Phi^+\rangle$, and A possesses the state $|\psi\rangle$ which she wishes B to have. One can show that the state of $|\psi\rangle$ possessed by A, and $|\Phi^+\rangle$ shared by A and B can be expressed as

$$|\psi\rangle_A \otimes \left|\Phi^+\right\rangle_{AB} = \\ \frac{1}{2}\left(\left|\Phi^+\right\rangle_A \otimes |\psi\rangle_B + \left|\Phi^-\right\rangle_A \otimes \sigma_z|\psi\rangle_B + \left|\Psi^+\right\rangle_A \otimes \sigma_x|\psi\rangle_B + \left|\Psi^-\right\rangle_A \otimes \sigma_y|\psi\rangle_B\right) \tag{5.6}$$

where σ_x, σ_y and σ_z are qubit unitary operators defined in Section 1.3, equation (1.31). Notice we have added the subscripts A and B to explicitly denote the portions of the state that are possessed by A and B respectively.

Given that the Bell states form an orthonormal basis, A can perform the corresponding PVM on her system. Equation (5.6) guarantees that no matter what the outcome of A's measurement, B's system will be left in a state which is one of four possible unitary rotations away from the state $|\psi\rangle$. A can communicate which of these four states by sending two classical bits to B. For example, if A measures in the Bell basis, and receives the outcome corresponding to $|\Phi^-\rangle$, then (5.6) and the measurement update rule indicate that the joint system is left in the state $|\Phi^-\rangle_A \otimes \sigma_z|\psi\rangle_B$. Given this outcome, A will communicate to B that he needs to perform the inverse of σ_z on his system. Doing so will leave B's system in the state $|\psi\rangle$, and the unknown state has been teleported.

In a two party protocol, such as teleportation, where two parties are spatially separated and can only communicate classical bits, there is an implicit restriction on the operations that can be performed on a shared bipartite state. In particular, they can only perform local qubit operations on their own subsystems and

communicate classically. We say they can perform Local Operations and Classical Communication - LOCC. The LOCC class of operations is fundamental to the theory of entanglement. In particular, entangled states are precisely the quantum states that allow a bypass of an LOCC restriction. This follows from two facts:

- starting with no entanglement, LOCC between two parties can create only separable states [63];
- starting with a Bell state, LOCC between two parties can deterministically create any bipartite state [51].

The latter point also motivates the labeling of a Bell state as the fundamental unit of entanglement on bipartite states, which is often called an *ebit*. Naturally, any state which can be used to create a Bell state through LOCC should also be viewed as an ebit, since such a state is just as useful as a Bell state. With respect to resource accounting, we say that the teleportation protocol converts one ebit and two communicated classical bits into one communicated qubit.

It should be noted that an ebit is a *maximally entangled state* on $\mathbb{C}^2 \otimes \mathbb{C}^2$, but in general, the maximally mixed state on $\mathbb{C}^d \otimes \mathbb{C}^d$ is:

$$|\,\Phi_d\,\rangle = \frac{|\,00\,\rangle + |\,11\,\rangle + \ldots + |\,(d-1)(d-1)\,\rangle}{\sqrt{d}}, \tag{5.7}$$

or any state which is equivalent via local unitary operations. The state (5.7) is also the state used in the d-dimensional generalization of the teleportation protocol [55].

The teleportation protocol indicates that ebits and other maximally entangled states have utility, but what about other entangled states? Recent work by Masanes has shown that the presence of any entangled states ρ allows for improved efficiency in the teleportation of some other quantum state σ between two parties [45, 46]. This example demonstrates that any entanglement is useful. However, if an entangled state is not an ebit, then how do we characterize how much entanglement is used? This question brings us to the important study of entanglement measures.

5.1.2 Entanglement Measures

Ideally, an entanglement measure should have some operational meaning: perhaps pertaining to a relationship to ebits, or perhaps quantifying; the utility of a quantum state for use in protocols where operations are restricted to LOCC. Two such measures are the *entanglement of distillation* and the *entanglement cost*. Operationally, these are defined as [51]:

- $E_D(\rho)$ - *the entanglement of distillation* - is the maximal ratio of singlet states (maximally entangled states on $\mathbb{C}^4$), that can be produced per n copies of ρ, as $n \to \infty$, by LOCC.

- $E_C(\rho)$ - *the entanglement cost* - is the minimal ratio of singlet states needed to produce n copies of ρ, as $n \to \infty$, by LOCC.

We denote n copies of ρ as $\rho \otimes \ldots \otimes \rho = \rho^{\otimes n}$. Mathematically, we can write these entanglement measures as:

$$E_D(\rho) = \sup \left\{ r : \lim_{n\to\infty} \left[\inf_{\mathcal{T}\in LOCC} \operatorname{tr} \left| \mathcal{T}(\rho^{\otimes n}) - (\left| \Phi^+ \right\rangle\!\!\left\langle \Phi^+ \right|)^{\otimes rn} \right| \right] = 0 \right\}, \tag{5.8}$$

$$E_C(\rho) = \inf \left\{ r : \lim_{n\to\infty} \left[\inf_{\mathcal{T}\in LOCC} \operatorname{tr} \left| \rho^{\otimes n} - \mathcal{T}\left((\left| \Phi^+ \right\rangle\!\!\left\langle \Phi^+ \right|)^{\otimes rn} \right) \right| \right] = 0 \right\}. \tag{5.9}$$

Note that $\operatorname{tr}|\cdot|$ is a measure of *distance* or *distinguishability* between quantum states which we talk about extensively in the next section. These definitions allow for the possibility that the ratio for cost or distillation can only be approached arbitrarily closely, and even then, only asymptotically.

It is worth discussing what properties of these functions make them reasonable measures of entanglement. Both these measures will assign value 1 to any ebit, and in general assign the value $\log d$ to any maximally entangled state. Also, we know that no separable state can create entanglement via LOCC, so $E_D = 0$ for any separable state. Similarly, no entanglement is needed to create a separable state and so $E_C = 0$ for any separable state.

Another interesting property of E_D and E_C is that they are both monotonically decreasing under LOCC:

$$\begin{aligned} E_D(\mathcal{T}(\rho)) &\le E_D(\rho) \qquad \forall\, \mathcal{T} \in LOCC \\ E_C(\mathcal{T}(\rho)) &\le E_C(\rho) \qquad \forall\, \mathcal{T} \in LOCC. \end{aligned} \tag{5.10}$$

One would expect this from any entanglement measure: given that LOCC cannot create entanglement from no entanglement, LOCC should not be able to produce more entanglement from less. Additionally, any state ρ which can be converted to ρ' through LOCC must have at least the same utility as ρ' in any LOCC protocol.

Other important properties possessed by E_D and E_C, relating to their asymptotic nature, are:

- Additivity:
$$E(\rho^{\otimes n}) = nE(\rho) \tag{5.11}$$
- Continuity: if $\rho_n \in K(\mathbb{C}^{2n} \otimes \mathbb{C}^{2n})$ then
$$\left\langle \psi^{\otimes n} \right| \rho_n \left| \psi^{\otimes n} \right\rangle \to 1 \Rightarrow \frac{1}{n} \left| E\left(\left| \psi^{\otimes n} \right\rangle\!\!\left\langle \psi^{\otimes n} \right| \right) - E(\rho_n) \right| \to 0. \tag{5.12}$$

It turns out that both E_D and E_C are equivalent on bipartite pure states. In particular, they are equal to an easy-to-calculate quantity called the *entropy of entanglement*. It is unfortunately the case that this equivalence does not hold

for mixed states, and in general E_C and E_D become difficult to calculate. This motivates the study of entanglement measures which may be easier to calculate, but at the cost of losing any clear operational significance.

The examples of E_D and E_C do suggest a number of desirable properties that any entanglement measure should have. Thus they are taken as axioms in a more abstract study of entanglement measures. In particular:

Definition 5.2 (Entanglement Monotone)**.** An *entanglement monotone* is a function $E : K(\mathcal{H}) \to \mathbb{R}^+$ which satisfies:

1. $E(\rho) = 0$ for any separable ρ

2. $E(|\,\Phi_d\,\rangle\langle\,\Phi_d\,|) = \log d$

3. If $\{p_i; \rho_i\}$ is an ensemble of possible quantum states resulting from an indeterministic LOCC operation on ρ, then
$$E(\rho) \geq \sum_i p_i E(\rho_i) \tag{5.13}$$

Indeed, these axioms state that an entanglement monotone is normalized to $\log d$[1]on maximally entangled states, is 0 on non-entangled states, and does not increase unded LOCC.

A major result in the study of axiomatic entanglement measures was presented in [37]:

Theorem 5.1. *An entanglement monotone E also satisfying additivity* (5.11) *and continuity* (5.12) *is bounded below and above by E_D and E_C respectively:*
$$E_D(\rho) \leq E(\rho) \leq E_C(\rho), \tag{5.14}$$
for any bipartite ρ.

Given that $E_D = E_C$ for pure bipartite states, this theorem additionally proves the existence of a unique entanglement measure (satisfying the given axioms) on pure states. Additionally, we can think of this theorem as a statement on resources consumed (ebits) during a conversion between ebits and some state ρ, and back again.

Aside from the basic introduction given here, entanglement theory is a rather complex subject of typically incalculable quantities and few general results. Perhaps insight can be gained on approaches to take in entanglement theory by studying a resource theory with a simpler structure, namely, distinguishability.

[1]$\log d$ being the asymptotic number of Bell states that can be created from one maximally entangled state on a d-level system.

5.2 Distinguishability - A Resource Theory

In entanglement theory, we discussed the definition, quantification, and utility of entanglement for bipartite states. In our study of distinguishability, we attempt to do the same. In this case, the fundamental objects we consider are not bipartite states, but *pairs* of quantum states, typically ρ and σ, belonging to the same Hilbert space $\mathcal{H}$.

5.2.1 One-Shot Distinguishability, d-bits, and TPCP maps

We begin by motivating one of the simplest known distinguishability measures, the *trace distance*, by considering a fundamental question: given either the state ρ or the state σ with equal likelihood, what is the maximum probability of successfully distinguishing which state one has with a single measurement?

Theorem 5.2. *Let $p_s(\rho, \sigma)$ be the maximal probability of distinguishing ρ and σ with a single measurement. Then*

$$p_s = \tfrac{1}{4}(\operatorname{tr}|\rho - \sigma| + 2), \tag{5.15}$$

where $|A| = \sqrt{A^\dagger A}$ is the unique positive square root of $A^\dagger A$.

Of importance in (5.15) is the value

$$\mathcal{D}^{\mathrm{tr}}(\rho, \sigma) \equiv \frac{1}{2}\operatorname{tr}|\rho - \sigma| \tag{5.16}$$

which we call the *trace distance*.

We prove this theorem with two lemmas.

Lemma 5.1. *A Hermitian matrix A can be written in the form $Q - S$ where Q and S are positive definite, and $QS = SQ = 0$.*

Proof. Since A is Hermitian, it has a spectral decomposition:

$$A = \sum_{i=1}^{d} a_i \,|\, \psi_i \rangle\langle\, \psi_i \,|\,,$$

for some eigenvalues $\{a_i\}_{i=1}^d$ and orthonormal projectors $\{|\, \psi_i \rangle\langle\, \psi_i \,|\}_{i=1}^d$. Suppose the eigenvalues are indexed in increasing order, and let a_n be the last negative eigenvalue, and a_p be the first positive eigenvalue. Then define:

$$S = \sum_{i=1}^{n} |a_i| \,|\, \psi_i \rangle\langle\, \psi_i \,| \qquad Q = \sum_{i=p}^{d} a_i \,|\, \psi_i \rangle\langle\, \psi_i \,|\,.$$

Then since $-S = \sum_{i=1}^{n} a_i \,|\, \psi_i \rangle\langle\, \psi_i \,|$, we have $A = Q - S$, and since Q and S are supported on orthogonal sets of projectors, $QS = SQ = 0$. □

Lemma 5.2.

$$\mathcal{D}^{\mathrm{tr}}(\rho,\sigma) = \max_E \mathrm{tr}(E(\rho-\sigma))$$

for all positive operators $E \leq \mathbb{I}$.

Proof. From [49]. Since $\rho - \sigma$ is hermitian, it can be written as $\rho - \sigma = Q - S$ as in Lemma 5.1. Thus since $Q + S$ is positive, and $(Q+S)^2 = (Q-S)^2 = Q^2 + S^2$, we have that $|\rho - \sigma| = Q + S$. Therefore $\mathcal{D}^{\mathrm{tr}}(\rho,\sigma) = \frac{1}{2}\,\mathrm{tr}\,Q + \frac{1}{2}\,\mathrm{tr}\,S$. But $\mathrm{tr}\,Q - \mathrm{tr}\,S = \mathrm{tr}(Q-S) = \mathrm{tr}(\rho-\sigma) = 0$, since ρ and σ are both of unit trace. Thus $\mathrm{tr}\,Q = \mathrm{tr}\,S = \mathcal{D}^{\mathrm{tr}}(\rho,\sigma)$.

Let $E \leq \mathbb{I}$ be any positive operator. Then $\mathrm{tr}(E(\rho-\sigma)) = \mathrm{tr}(E(Q-S)) \leq \mathrm{tr}\,EQ \leq \mathrm{tr}\,\mathbb{I}Q = \mathcal{D}^{\mathrm{tr}}(\rho,\sigma)$, establishing $\mathrm{tr}(E(\rho-\sigma)) \leq \mathcal{D}^{\mathrm{tr}}(\rho,\sigma)$. To show the maximum can be achieved, let E be the projector onto the support of Q. Then $\mathrm{tr}(E(\rho-\sigma)) = \mathrm{tr}(E(Q-S)) = \mathrm{tr}\,EQ - 0 = \mathrm{tr}\,Q = \mathcal{D}^{\mathrm{tr}}(\rho,\sigma)$. □

Proof of Theorem 5.2. Any measurement procedure ending with a choice between two outcomes can be formulated as a two outcome POVM in which the first label corresponds to guessing ρ and the second label corresponds to guessing σ. So given a two outcome measurement $M = \{E_1, E_2\} = \{E_1, \mathbb{I} - E_1\}$, the optimal procedure for distinguishing known possible states ρ and σ with one measurement is the following. The measurement M induces probability distributions $p_1 = \mathrm{tr}(\rho E_1), p_2 = \mathrm{tr}(\rho E_2)$ and $q_1 = \mathrm{tr}(\sigma E_1), q_2 = \mathrm{tr}(\sigma E_2)$. Without loss of generality, suppose that $p_1 \geq q_1$, which implies that $q_2 \geq p_2$. Then, supposing that one knows they have either quantum state ρ or σ with equal probability, in the event of measuring M and receiving output 1, it is most likely that the state was ρ. Upon receiving 2, it is most likely that the state was σ. Guessing in this way produces the following optimal probability of success:

$$\Pr(\textit{success}) = \Pr(\text{outcome } 1|\rho)\Pr(\rho) + \Pr(\text{outcome } 2|\sigma)\Pr(\sigma) = \tfrac{1}{2}p_1 + \tfrac{1}{2}q_2$$

Now we wish to show that picking the measurement which maximizes Pr(success) gives the desired linear relation to $\mathcal{D}^{\mathrm{tr}}(\rho,\sigma)$.

Since $q_2 = p_1 + p_2 - q_1$, we have

$$\Pr(\textit{success}) = \tfrac{1}{2}(p_1 - q_1) + \tfrac{1}{2} = \tfrac{1}{2}\,\mathrm{tr}(E_1(\rho-\sigma)) + \tfrac{1}{2}.$$

Thus applying Lemma 5.2, and picking M to maximize $\frac{1}{2}\,\mathrm{tr}(E_1(\rho-\sigma))$ gives the relation:

$$\mathcal{D}^{\mathrm{tr}}(\rho,\sigma) = 2\,\{\text{maximum probability of one-shot distinguishing } \rho \text{ and } \sigma\} - 1$$

□

Thus the quantity $\mathcal{D}^{\mathrm{tr}}(\rho,\sigma)$ is linearly related to the maximum probability of distinguishing between ρ and σ with a single measurement. Moreover, notice that if we can never distinguish between ρ and σ, then $\mathcal{D}^{\mathrm{tr}}(\rho,\sigma) = 0$ and if we can always

distinguish between ρ and σ, then $\mathcal{D}^{\mathrm{tr}}(\rho,\sigma) = 1$. Thus $\mathcal{D}^{\mathrm{tr}}(\rho,\sigma)$ appears to be a good operational indicator of the distinguishability of ρ and σ. Moreover, we notice that $\mathcal{D}^{\mathrm{tr}}(\rho,\sigma) = 0$ if and only if $\rho = \sigma$.

Another thing to notice about $\mathcal{D}^{\mathrm{tr}}$ is that it is unitarily invariant. Indeed, since $\mathcal{D}^{\mathrm{tr}}(\rho,\sigma)$ is the sum of the absolute values of the eigenvalues of $\rho - \sigma$, and unitary action on a matrix does not change eigenvalues, we must have $\mathcal{D}^{\mathrm{tr}}(U\rho U^\dagger, U\sigma U^\dagger) = \mathcal{D}^{\mathrm{tr}}(\rho,\sigma)$.

The trace distance is also related to a common metric on probability distributions. Indeed, suppose that ρ and σ are diagonal in the same basis:

$$\rho = \sum_{i=1}^{d} p_i \,|\, i \,\rangle\langle\, i \,| \qquad \sigma = \sum_{i=1}^{d} q_i \,|\, i \,\rangle\langle\, i \,| \,.$$

Then the trace distance simplifies to

$$\mathcal{D}^{\mathrm{tr}}(\rho,\sigma) = \mathcal{D}_c^{\mathrm{tr}}(\vec{p},\vec{q}) = \frac{1}{2}\sum_{i=1}^{d} |p_i - q_i|, \tag{5.17}$$

which is known as the *Kolmogorov distance* or L^1 *distance* on probability distributions [49].

Just as we saw that entanglement was a beneficial resource in quantum information theory, we suggest that distinguishability is also a useful resource. In any protocol where B needs to receive and interpret a message from A, which could be one of a number of messages, the distinguishability of the received messages is important. In quantum information theory, messages can come in the form of quantum states, and hence we study distinguishability as a quantum information resource.

In our discussion of entanglement theory, we defined the ebit as the fundamental unit of entanglement. In particular, possession of an ebit is more beneficial than possessing any other entangled state. Similarly, we define a *dbit* as any pair of quantum states which are perfectly distinguishable in a single-shot measurement. Thus, for example, two orthogonal quantum states ρ and σ, pure or mixed, constitute a dbit. In order to distinguish two orthogonal states, we simply perform a measurement in which the projector E onto the support of ρ is one of the effects. If the outcome corresponding to E comes up, the state must have been ρ, otherwise σ. With respect to the classical trace distance on probability vectors (the Kolmogorov distance), it is clear that two probability vectors are perfectly distinguishable if and only if $\vec{p}\cdot\vec{q} = 0$. That is, if the supports of the two probability vectors are disjoint. It can be shown that any non-orthogonal quantum states are not perfectly distinguishable [49], thus orthogonal quantum state pairs constitute all dbits.

Entanglement theory is heavily entwined with LOCC. This is due to the fact that entanglement could not increase under LOCC, and consequently, entanglement measures were required to be monotonically non-increasing under LOCC. Also, an

ebit could be transformed into any other state via LOCC. We argue now that the analogous set of operations for distinguishability theory is in fact all trace-preserving completely-positive maps.

First notice that any procedure that one can use to distinguish between states ρ and σ may very well start with a transformation on ρ and σ, i.e. any TPCP map. Thus a TPCP map should not create distinguishability. Moreover, we can show that under TPCP maps, a dbit can be transformed into any possible pair of states.

Proposition 5.1. *If $\rho, \sigma \in \mathbb{C}^{d_1}$ constitute a dbit, then for any pair of states $\rho', \sigma' \in \mathbb{C}^{d_2}$, there exists a TPCP map $\mathcal{T}$ such that $\rho' = \mathcal{T}(\rho)$ and $\sigma' = \mathcal{T}(\sigma)$.*

Proof. Suppose that ρ and σ are orthogonal. Let $M = \{E_1, E_2\}$ be the measurement that distinguishes ρ and σ with certainty, where outcome 1 corresponds to choosing ρ. Consider the following procedure performed on any state in $\mathcal{H}_A$. The measurement M is performed, and if the outcome 1 occurs, then the state ρ' is prepared in $\mathcal{H}_B$ and if outcome 2 occurs then the state σ' is prepared in $\mathcal{H}_B$.

To show that this procedure is a TPCP map, we will give a Kraus representation for it. Without loss of generality, let

$$\begin{aligned} \rho &= \sum_{k=1}^{n} p_k \,|\, k \,\rangle\!\langle\, k \,| & \sigma &= \sum_{k=n+1}^{d_1} q_k \,|\, k \,\rangle\!\langle\, k \,| \\ \rho' &= \sum_{i=1}^{d_2} p'_i \,|\, \psi_i \,\rangle\!\langle\, \psi_i \,| & \sigma' &= \sum_{j=1}^{d_2} q'_j \,|\, \phi_j \,\rangle\!\langle\, \phi_j \,| \,. \end{aligned} \tag{5.18}$$

where the coefficients in each decomposition form a probability vector. Define the Kraus operators $E_{ik} = \sqrt{p'_i} \,|\, \psi_i \,\rangle\!\langle\, k \,|$ for $k \in 1 \ldots n$, $i \in 1 \ldots d_2$ and $E_{jk} = \sqrt{q'_j} \,|\, \phi_j \,\rangle\!\langle\, k \,|$, $j \in 1 \ldots d_2$, $k \in n+1 \ldots d_1$.

Then indeed:

$$\begin{aligned} \sum_{ik} E_{ik}^\dagger E_{ik} + \sum_{jk} E_{jk}^\dagger E_{jk} &= \sum_{ik} p'_i \,|\, k \,\rangle \langle\, \psi_i \,|\, \psi_i \,\rangle \langle\, k \,| + \sum_{jk} q'_j \,|\, k \,\rangle \langle\, \phi_j \,|\, \phi_j \,\rangle \langle\, k \,| \\ &= \sum_{k=1}^{n} |\, k \,\rangle\!\langle\, k \,| + \sum_{k=n+1}^{d_1} |\, k \,\rangle\!\langle\, k \,| = \mathbb{I}, \end{aligned} \tag{5.19}$$

so these operators define a TPCP map. Lastly we verify that this map takes ρ to ρ'.

$$\begin{aligned}\sum_{ik} E_{ik}\rho E_{ik}^\dagger + \sum_{jk} E_{jk}\rho E_{jk}^\dagger &= \sum_{ik} p_i' \,|\, \psi_i \,\rangle \,\langle\, k \,|\, \rho \,|\, k \,\rangle \,\langle\, \psi_i \,|\, + 0 \\ &= \sum_{ik} p_i' \,|\, \psi_i \,\rangle \,(p_k \,\langle\, k \,|\, k \,\rangle)\,\langle\, \psi_i \,|\, \\ &= \sum_{i} p_i' \,|\, \psi_i \,\rangle\langle\, \psi_i \,|\, \\ &= \rho'.\end{aligned} \tag{5.20}$$

Similarly we can show this map takes σ to σ'. □

In the discussion of entanglement, we considered the problem of converting states to ebits and back again. In general, this problem was considered in the asymptotic regime. Our exposition on distinguishability theory has thus far been concerned with single-shot operations, rather than distinguishability in an asymptotic regime. However, we can still address a similar question of conversion. We know that in general, no TPCP map can create a dbit out of two states that do not constitute a dbit. However, in light of the fact that commuting states can be viewed as a pair of classical probability distributions, we can consider the question of converting pairs of states to classical probability distributions.

5.2.2 Reversible Distinguishability

In $K(\mathbb{C}^d)$, we will define as state as 'classical' if it is diagonal in some preferred eigenbasis and denote this set as $K_c(\mathbb{C}^d)$. The elements of $K_c(\mathbb{C}^d)$ are in one-to-one correspondence with the elements of Δ_d, the d-dimensional probability vectors. In particular, the probability vector $\vec{p} = (p_1, \ldots, p_n) \in \Delta_d$ associated with $\rho \in K_c(\mathbb{C}^d)$ is the vector of eigenvalues of ρ.

We now define two sets of TPCP maps which represent conversions from quantum to classical states, and vice versa:

$$\mathcal{QC} = \{\Phi : K(\mathbb{C}^N) \to K_c(\mathbb{C}^M) | M, N \in 2, 3, \ldots\} \tag{5.21}$$

$$\mathcal{CQ} = \{\Phi : K_c(\mathbb{C}^M) \to K(\mathbb{C}^N) | M, N \in 2, 3, \ldots\} \tag{5.22}$$

Maps in $\mathcal{QC}$ and $\mathcal{CQ}$ are intimately related to the processes of measurement and preparation respectively. In particular, any $\mathcal{QC}$ map can be viewed as a measurement procedure, and any $\mathcal{CQ}$ map can be viewed as a preparation procedure.

Proposition 5.2. *Any map $\Phi \in \mathcal{QC}$ mapping $K(\mathbb{C}^N)$ to $K_c(\mathbb{C}^M)$ is equivalent to an M-outcome POVM $\{E_m\}_{m=1}^M$ acting on $K(\mathbb{C}^N)$, such that $\vec{p} = (p_1, \ldots, p_M) = (\mathrm{tr}(\rho E_1), \ldots, \mathrm{tr}(\rho E_M))$ is the vector of eigenvalues for $\Phi(\rho)$.*

Proof. Given $\Phi \in \mathcal{QC}$ mapping $K(\mathbb{C}^N)$ to $K_c(\mathbb{C}^M)$, consider the map $\Phi_m : K(\mathbb{C}^N) \to \mathbb{R}$, which picks the m^{th} eigenvalue of $\Phi(\rho) = \sum_{m=1}^{M} p_m \,|\, m \,\rangle\langle\, m \,|$, i.e. $\Phi_m(\rho) = p_m$. Since Φ is linear, so must be Φ_m. Thus Φ_m is a linear functional on $K(\mathbb{C}^N)$ which can be extended to a linear function on $\mathcal{L}(\mathbb{C}^N)$. Since $\Phi_m(\rho) = p_m \geq 0$ for all $\rho \in K(\mathbb{C}^N)$, Φ_m is a positive map. Thus, by the Riesz representation theorem [14], there exists a unique positive operator $A_m \in \mathcal{L}^+(\mathbb{C}^N)$ such that $\Phi_m(\rho) = \mathrm{tr}(A\rho)$. Since Φ is trace-preserving, $1 = \mathrm{tr}\,\Phi(\rho) = \sum_{m=1}^{1} p_m = \sum_{m=1}^{M} \mathrm{tr}(A_m\rho)$, for all $\rho \in K(\mathbb{C}^N)$. Thus it must be the case that $\sum_{m=1}^{M} A_m = \mathbb{I}$, and so Φ specifies a unique M-outcome POVM.

Conversely, given an M-outcome POVM $\{A_m\}_{m=1}^{M}$, we can perform the POVM and then on the condition of receiving outcome m, prepare the state $|\, m \,\rangle\langle\, m \,|$. Such a procedure will map the state ρ to $\sum_{m=1}^{M} p_m \,|\, m \,\rangle\langle\, m \,|$ where $p_m = \mathrm{tr}(\rho A_m)$. We know from Lemma 5.1 that such a procedure indeed specifies a TPCP map. □

Proposition 5.3. *Any map $\Phi \in \mathcal{CQ}$ mapping $K_c(\mathbb{C}^M)$ to $K(\mathbb{C}^N)$ is equivalent a to a set of M states $\mathcal{P} = \{\sigma_m\}_{m=1}^{M} \subset K(\mathbb{C}^N)$ such that $\Phi(\rho) = \sum_{m=1}^{M} p_m\sigma_m$ where $\vec{p}$ is the vector of eigenvalues of $\rho \in K_c(\mathbb{C}^M)$. We denote this map induced by $\mathcal{P}$ as $\Phi_{\mathcal{P}}$.*

Proof. Let $\Phi \in \mathcal{CQ}$ map $K_c(\mathbb{C}^M)$ to $K(\mathbb{C}^N)$. Since $K_c(\mathbb{C}^M)$ is isomorphic to Δ_M, and Φ is a linear map, it preserves the structure of $K_c(\mathbb{C}^M)$ in the range of Φ. In particular, consider the extreme points of $K_c(\mathbb{C}^M)$, which are the states $\{|\, m \,\rangle\langle\, m \,|\}_{m=1}^{M}$ of the preferred eigenbasis. Denote $\sigma_m = \Phi(|\, m \,\rangle\langle\, m \,|)$. Then for any $\rho \in K_c(\mathbb{C}^M)$, $\rho = \sum_{m=1}^{M} p_m \,|\, m \,\rangle\langle\, m \,|$ for some probability vector $\vec{p}$. Then the linearity of Φ gives $\Phi(\rho) = \sum_{m=1}^{M} p_m\sigma_m$.

Conversely, if $\mathcal{P} = \{\sigma_m\}_{m=1}^{M} \subset K(\mathbb{C}^N)$, then the map $\Phi_{\mathcal{P}}(\rho)$ is clearly in $\mathcal{CQ}$. □

Lastly we wish to characterize the state pairs in any quantum system which are always reversibly convertible to classical states. Such state pairs, in the sense of distinguishability, are 'as good as classical states'. As it turns out, the notion of classicality provided by this characterization is equivalent to the condition for broadcastability [5]. In particular,

Theorem 5.3. *A pair of quantum states ρ, σ are reversibly convertible to classical states if and only if they are simultaneously diagonalizable, i.e they commute.*

Proof. Two quantum states $\rho, \sigma \in K(\mathbb{C}^N)$, are reversibly convertible to classical states if and only if there exists a map $\mathcal{E} \in \mathcal{QC}$ and a map $\mathcal{F} \in \mathcal{CQ}$, such that $\rho = \mathcal{F}(\vec{p}) = \mathcal{F} \circ \mathcal{E}(\rho)$ and $\sigma = \mathcal{F}(\vec{q}) = \mathcal{F} \circ \mathcal{E}(\sigma)$, for some probability vectors $\vec{p}$ and $\vec{q}$ in $K_c(\mathbb{C}^M)$.

First suppose that ρ and σ are simultaneously diagonalizable in the eigenbasis $B = \{|\, \psi_i \,\rangle\langle\, \psi_i \,|\}_{i=1}^{N}$ and the eigenbasis for $K_c(\mathbb{C}^N)$ is $C = \{|\, \phi_i \,\rangle\langle\, \phi_i \,|\}_{i=1}^{N}$. We can choose the map $\mathcal{E} \in \mathcal{QC}$ to be the procedure which measures in the eigenbasis B,

and on receiving outcome i prepares the state $|\,\phi_i\,\rangle\langle\,\phi_i\,|$. Similarly, we can choose the map $\mathcal{F} \in \mathcal{QC}$ to be the reverse procedure.

Let us then suppose that ρ and σ are not simultaneously diagonalizable, and that the $\mathcal{QC}$ and $\mathcal{CQ}$ maps $\mathcal{E}$ and $\mathcal{F}$ exist, such that both ρ and σ are fixed points of $\mathcal{F} \circ \mathcal{E}$. We will derive a contradiction.

A theorem in [41] characterizes all fixed points of a completely positive map. In particular any fixed point χ of a TPCP map acting on $\mathbb{C}^N$ can be written as

$$\chi = \oplus_{k=1}^{K} p_k \mu_k \otimes \tau_k \tag{5.23}$$

where the μ_k is any density matrix on Hilbert spaces $\mathcal{H}_{k1}$, the τ_k are fixed density matrices on a Hilbert spaces $\mathcal{H}_{k2}$, $\mathbb{C}^N = \oplus_{k=1}^{K} \mathcal{H}_{k1} \otimes \mathcal{H}_{k2}$, and the p_k form any probability distribution on $1 \ldots K$. That is, any choices for $\vec{p}$ and μ_k gives a fixed point.

Suppose (5.23) characterizes the fixed points for the map $\mathcal{F} \circ \mathcal{E}$ for which ρ and σ are fixed points. Suppose further that each of the $\mathcal{H}_{k1}$ are trivial (1-dimensional) spaces. Then any fixed points would be the weighted sum of operators acting on orthogonal subspaces, due to the direct sum structure. Hence all fixed points would be simultaneously diagonalizable. Since we are assuming that ρ and σ are not simultaneously diagonalizable, it is necessary that at least one of the $\mathcal{H}_{k1}$ are non-trivial.

Without loss of generality, assume $\mathcal{H}_{11}$ is non-trivial, and that we fix the μ_k for $k \geq 2$, and that we fix the p_k such that $p_1 > 0$. With these fixed choices, picking any operator on $\mathcal{H}_{11}$ specifies a fixed point of the map $\mathcal{F} \circ \mathcal{E}$.

Consider the following map taking elements of $K(\mathcal{H}_{11})$ to fixed points of $\mathcal{F} \circ \mathcal{E}$.

$$\begin{aligned} \mathcal{G} : K(\mathcal{H}_{11}) &\to K(\mathbb{C}^N) \\ \psi &\to p_1 \psi \otimes \tau_1 + \sum_{k=2}^{K} p_k \mu_k \otimes \tau_k \end{aligned} \tag{5.24}$$

This map is reversible, as given $\mathcal{G}(\psi)$, projecting onto the $\mathcal{H}_{11} \otimes \mathcal{H}_{12}$ subspace and then tracing out $\mathcal{H}_{21}$, gives the state ψ.

Consider two distinct and non-orthogonal pure states, ψ and ϕ, on $\mathcal{H}_{11}$. Now $\mathcal{G}(\psi)$ and $\mathcal{G}(\phi)$ are fixed points of $\mathcal{F} \circ \mathcal{E}$, hence ψ and ϕ can be reversibly mapped to classical states $\vec{r} = \mathcal{E} \circ \mathcal{G}(\psi)$, $\vec{s} = \mathcal{E} \circ \mathcal{G}(\phi)$ (the inverse map is $\mathcal{G}^{-1} \circ \mathcal{F}$). However, since this map and its inverse are linear, they must preserve convex structure, and so if ψ and ϕ are pure states, then $\vec{r}$ and $\vec{s}$ must be extreme points of a simplex. Specifically, they must be trivial probability vectors (1 in one location, 0 everywhere else). However, such vectors would be perfectly distinguishable by one-shot sampling. Consequently, we could perfectly distinguish ψ and ϕ by converting them to $\vec{r}$ and $\vec{s}$. But we know that since ψ and ϕ are non-orthogonal, they are not perfectly distinguishable [49]. Thus we have a contradiction[2]. □

5.2.3 Distinguishability Measures

As we did with entanglement measures, we will consider that not all distinguishability measures may have an operational meaning, and as such we take an axiomatic approach. In Section 5.2.1, we introduced the trace distance as measure of distinguishability. We saw that it had value 0 on equal states, and value 1 on dbits. We also argued that a TPCP map should not 'create distinguishability'. We now verify mathematically that indeed a TPCP map does not increase the trace distance of two quantum states.

Proposition 5.4. *Suppose* Φ *is a TPCP map. Then*

$$\mathcal{D}^{\mathrm{tr}}(\Phi(\rho), \Phi(\sigma)) \leq \mathcal{D}^{\mathrm{tr}}(\rho, \sigma) \tag{5.25}$$

Proof. From [49]. Write $\rho - \sigma = Q - S$, as in lemma 5.2, and let P be the projector such that $\mathcal{D}^{\mathrm{tr}}(\Phi(\rho), \Phi(\sigma)) = \mathrm{tr}(P(\Phi(\rho) - \Phi(\sigma)))$. Since, as shown in lemma 5.2, $\mathrm{tr}\, Q = \mathrm{tr}\, S$, the trace-preserving property of Φ gives $\mathrm{tr}\, \Phi(Q) = \mathrm{tr}\, \Phi(S)$. Thus we have the following chain of (in)equalities:

$$\begin{aligned}
\mathcal{D}^{\mathrm{tr}}(\rho, \sigma) &= \tfrac{1}{2} \mathrm{tr}(Q + S) \\
&= \tfrac{1}{2} \mathrm{tr}\, \Phi(Q) + \tfrac{1}{2} \mathrm{tr}\, \Phi(S) \\
&= \mathrm{tr}\, \Phi(Q) \\
&\geq \mathrm{tr}\, P\Phi(Q) \\
&\geq \mathrm{tr}(P(\Phi(Q) - \Phi(S))) \\
&= \mathrm{tr}(P(\Phi(\rho) - \Phi(\sigma))) \\
&= \mathcal{D}^{\mathrm{tr}}(\Phi(\rho), \Phi(\sigma)).
\end{aligned} \tag{5.26}$$

□

For any conceivable measure of distinguishability, it should be the case that it is monotonically non-increasing under TPCP maps. Thus we make the following definition of a *distinguishability monotone*.

Definition 5.3 (Quantum Distinguishability Monotone)**.** A *distinguishability monotone* is a function $\mathcal{D} : K(\mathcal{H}) \times K(\mathcal{H}) \to \mathbb{R}^+$ satisfying

$$\mathcal{D}(\Phi(\rho).\Phi(\sigma)) \leq \mathcal{D}(\rho, \sigma) \tag{5.27}$$

for any TPCP map Φ, and

$$\mathcal{D}(\rho, \sigma) = \begin{cases} 1 & \rho\sigma = 0 \\ 0 & \rho = \sigma \end{cases}. \tag{5.28}$$

An immediate corollary arising from two applications of (5.27) states that a distinguishability monotone is unitarily-invariant, as we already know the trace distance to be.

$$\mathcal{D}(\rho,\sigma) = \mathcal{D}(U^\dagger U\rho U^\dagger U, U^\dagger U\sigma U^\dagger U) \leq \mathcal{D}(U\rho U^\dagger, U\sigma U^\dagger) \leq \mathcal{D}(\rho,\sigma). \tag{5.29}$$

At this point it is worth noting that any quantum distinguishability monotone induces a classical distinguishability monotone, in much the same way that the trace distance induces a classical analogue (5.17). We define a classical probability monotone as:

Definition 5.4 (Classical Distinguishability Monotone)**.** A *classical distinguishability monotone* is a function $\mathcal{D}^c : \Delta \times \Delta \to \mathbb{R}^+$ satisfying

$$\mathcal{D}^c(\Phi(\vec{p}), \Phi(\vec{q})) \leq \mathcal{D}^c(\vec{p}, \vec{q}) \tag{5.30}$$

for any stochastic map Φ, and

$$\mathcal{D}(\vec{p}, \vec{q}) = \begin{cases} 1 & \vec{p}\cdot\vec{q} = 0 \\ 0 & \vec{p} = \vec{q} \end{cases}. \tag{5.31}$$

The stochastic maps are the set of possible operators on probability vectors. In particular, a stochastic map from Δ_M to Δ_N is equivalent to a N-by-M matrix for which the rows sum to 1.

Now if $\mathcal{D}$ is a distinguishability monotone, then consider two states ρ and σ which are diagonal in the same eigenbasis. Given that ρ and σ are diagonal in the same eigenbasis, they are both completely parameterized by their vector of eigenvalues, $\vec{p}$ and $\vec{q}$ respectively. Thus we effectively have:

$$\mathcal{D}(\rho,\sigma) = \mathcal{D}^c(\vec{p}, \vec{q}) \tag{5.32}$$

for some positive function $\mathcal{D}^c$. We can verify that $\mathcal{D}^c$ is in fact a classical distinguishability monotone. In particular, $\vec{p} = \vec{q}$ if and only if $\rho = \sigma$, and $\vec{p}\cdot\vec{q} = 0$ if and only if $\rho\sigma = 0$. Thus (5.31) is satisfied via (5.28). Also, if Φ is a stochastic matrix on Δ, then there is a TPCP map Φ' which preserves the commutativity of ρ and σ such that

$$\mathcal{D}(\Phi'(\rho), \Phi'(\sigma)) = \mathcal{D}^c(\Phi(\vec{p}), \Phi(\vec{q})). \tag{5.33}$$

In particular, this map could arise from the procedure whereby the input state is measured in the eigenbasis $\{|\, m \,\rangle\langle\, m \,|\}$ of ρ and σ, and if outcome m occurs, the state $|\, n \,\rangle\langle\, n \,|$ is prepared with probability $[\Phi]_{n,m}$. On density operators with the same eigenvectors as ρ and σ, this map has the effect of acting the stochastic map Φ on the eigenvalues. Hence this map has the property (5.33). Thus (5.30) is satisfied via (5.27).

Any distinguishability monotone $\mathcal{D}$ gives rise to another distinguishability monotone related to its induced classical distinguishability monotone $\mathcal{D}^c$. In particular,

we can map to classical states with a $\mathcal{QC}$ map and apply the induced classical distinguishability monotone $\mathcal{D}^c$. Moreover, for a given ρ and σ, we can maximize this value over all $\mathcal{QC}$ maps:

$$\mathcal{D}_D(\rho, \sigma) = \sup_{\Phi \in \mathcal{QC}} \mathcal{D}^c(\Phi(\rho), \Phi(\sigma)). \tag{5.34}$$

In analogy to the entanglement of distillation, we call this the *distinguishability of distillation induced by* $\mathcal{D}$. This is the maximum amount classical distinguishability (according to $\mathcal{D}^c$) that can be achieved from ρ and σ through a $\mathcal{QC}$ map. These types of distinguishability monotones have been studied previously. For example, one can verify that $\mathcal{D}^{\text{tr}}$ is actually equal to its induced distinguishability of distillation [49], which we now show:

Proposition 5.5. *For all ρ, σ, $\mathcal{D}^{\text{tr}}(\rho, \sigma) = \mathcal{D}^{\text{tr}}{}_D(\rho, \sigma)$.*

Proof. First, recall that any $\mathcal{QC}$ map Φ corresponds to a POVM $\{E_m\}_{m=1}^M$ such that $\Phi(\rho) = (\text{tr}(\rho E_1), \ldots, \text{tr}(\rho E_M))$. This fact, combined with the form of the induced classical trace distance (5.17), gives

$$\begin{aligned} \mathcal{D}_D(\rho, \sigma) &= \sup_{\Phi \in \mathcal{QC}} \mathcal{D}^{\text{tr}}(\Phi(\rho), \Phi(\sigma)) \\ &= \sup_{POVM \{E_m\}_m} \frac{1}{2} \sum_{m=1}^M |\,\text{tr}(E_m(\rho - \sigma))| \\ &= \sup_{POVM \{E_m\}_m} \frac{1}{2} \sum_{m=1}^M |p_m - q_m| \end{aligned} \tag{5.35}$$

where $p_m = \text{tr}(\rho E_m)$ and $q_m = \text{tr}(\sigma E_m)$.

Now we will show that for any POVM $\{E_m\}_{m=1}^M$, there is a two-outcome POVM $\{E'_1, E'_2\}$ such that

$$\frac{1}{2} \sum_{m=1}^M |p_m - q_m| = \frac{1}{2} \sum_{i=1}^2 |\,\text{tr}(\rho E'_i) - \text{tr}(\sigma E'_i)|. \tag{5.36}$$

Indeed, let P be the set of all indices such that $p_m > q_m$, and let N be the set of all indices such that $p_m \leq q_m$. Then

$$\begin{aligned} \frac{1}{2} \sum_{m=1}^M |p_m - q_m| &= \frac{1}{2} \left(\sum_{m \in P} (p_m - q_m) + \sum_{m \in N} (q_m - p_n) \right) \\ &= \frac{1}{2} \left(|p'_1 - q'_1| + |p'_2 - q'_2| \right), \end{aligned} \tag{5.37}$$

where $p'_1 = \sum_{m \in P} p_m$, $p'_2 = \sum_{m \in N} p_m$ and similarly for q'. Thus if we define $E'_1 = \sum_{m \in P} E_m$ and $E'_2 = \sum_{m \in N}$, we have our desired two outcome POVM.

Now the result is almost immediate from Lemma 5.2. Since there exists a projector E such that $\mathcal{D}^{\mathrm{tr}}(\rho,\sigma) = \mathrm{tr}(E(\rho-\sigma))$, it remains to show that for the two outcome measurement $M = \{E, \mathbb{I} - E\}$,

$$\mathrm{tr}(E(\rho-\sigma)) = \tfrac{1}{2}(|\,\mathrm{tr}(\rho E) - \mathrm{tr}(\sigma E)| + |\,\mathrm{tr}(\rho(\mathbb{I}-E)) - \mathrm{tr}(\sigma(\mathbb{I}-E))|).$$

Without loss of generality, we assume that $\mathrm{tr}(\rho E) \geq \mathrm{tr}(\sigma E)$, implying that $\mathrm{tr}(\rho(\mathbb{I}-E)) \leq \mathrm{tr}(\sigma(\mathbb{I}-E))$ Thus

$$\begin{aligned}
&\tfrac{1}{2}\left(|\,\mathrm{tr}(\rho E) - \mathrm{tr}(\sigma E)| + |\,\mathrm{tr}(\rho(\mathbb{I}-E)) - \mathrm{tr}(\sigma(\mathbb{I}-E))|\right) \\
&= \tfrac{1}{2}(\mathrm{tr}(\rho E) - \mathrm{tr}(\sigma E) - \mathrm{tr}(\rho(\mathbb{I}-E)) + \mathrm{tr}(\sigma(\mathbb{I}-E))) \\
&= \tfrac{1}{2}(2\,\mathrm{tr}(\rho E) - 2\,\mathrm{tr}(\sigma E) - \mathrm{tr}\,\rho + \mathrm{tr}\,\sigma) \\
&= \mathrm{tr}(E(\rho-\sigma))
\end{aligned}$$

□

Any distinguishability monotone also gives rise a monotone we call the induced distinguishability of formation, or reverse distinguishability, which is analogous to the entanglement cost. Essentially, this is the minimum classical distinguishability (according to $\mathcal{D}^c$) required to form ρ and σ through a $\mathcal{CQ}$ map. Fix ρ and σ, and a $\mathcal{CQ}$ map Φ. Define the set $\mathcal{V}^{\Phi}_{\rho\sigma}$ as the set of probability vector pairs $\vec{p}$ and $\vec{q}$ such that $\Phi(\vec{p}) = \rho$ and $\Phi(\vec{q}) = \sigma$. Recall that any $\mathcal{CQ}$ map corresponds to a set of states $\mathcal{P}$, and the range of the map is the convex hull of $\mathcal{P}$. If Φ corresponds to a set $\mathcal{P}$ for which the convex hull does not contain both ρ and σ, then $\mathcal{V}^{\Phi}_{\rho\sigma}$ is empty. Now we define the *induced distinguishability of formation* as

$$\mathcal{D}_F(\rho,\sigma) = \inf_{\substack{\Phi\in\mathcal{CQ} \\ \vec{p},\vec{q}\in\mathcal{V}^{\Phi}_{\rho\sigma}}} \mathcal{D}^c(\vec{p},\vec{q}). \tag{5.38}$$

The induced distinguishabilities of formation have not been studied as much as the induced distinguishabilities of distillation. However, there is recent work by Matsumoto which investigates such a distinguishability measure with respect to the Fischer metric [47].

Theorem 5.1 for entanglement theory states that any entanglement measure is bounded above and below by the particular entanglement measures E_D and E_C. We have a similar theorem for distinguishability monotones.

Theorem 5.4. *For any distinguishability monotone $\mathcal{D}$,*

$$\mathcal{D}_D(\rho,\sigma) \leq \mathcal{D}(\rho,\sigma) \leq \mathcal{D}_F(\rho,\sigma). \tag{5.39}$$

Proof. Suppose $\Phi \in \mathcal{QC}$ maps ρ to $\vec{p}$ and σ to $\vec{q}$. Then by the monotonicity of $\mathcal{D}$,

$$\mathcal{D}(\rho,\sigma) \geq \mathcal{D}(\Phi(\rho),\Phi(\sigma)) = \mathcal{D}^c(\vec{p},\vec{q}). \tag{5.40}$$

Taking the supremum over $\Phi \in \mathcal{QC}$ gives

$$\mathcal{D}_D(\rho,\sigma) \leq \mathcal{D}(\rho,\sigma). \tag{5.41}$$

Suppose $\Phi \in \mathcal{CQ}$, and $(\vec{p}, \vec{q}) \in \mathcal{V}_{\rho\sigma}^{\Phi}$. Then again by the monotonicity of $\mathcal{D}$,

$$\mathcal{D}^c(\vec{p}, \vec{q}) =\geq \mathcal{D}(\Phi(\vec{p}), \Phi(\vec{q})) = \mathcal{D}(\rho, \sigma). \tag{5.42}$$

Minimizing over all possible $\Phi \in \mathcal{CQ}$ and $(\vec{p}, \vec{q}) \in \mathcal{V}_{\rho\sigma}^{\Phi}$ gives

$$\mathcal{D}(\rho, \sigma) \leq \mathcal{D}_F(\rho, \sigma). \tag{5.43}$$

□

The analogous theorem in entanglement theory could be viewed as uniqueness theorem when the bipartite states were restricted to be pure. This followed from the fact that E_D and E_C were in fact equal on pure states. Similarly we may like to know when a classical distinguishability measure and a restriction to a certain set of states induces a unique distinguishability monotone, i.e. when does $\mathcal{D}_F = \mathcal{D}_D$?. Or, for a given distinguishability monotone and a pair of states (ρ, σ), we may wish to know if there is a gap between $\mathcal{D}_F$ and $\mathcal{D}_D$. Such knowledge allows us to quantify the cost in classical distinguishability for conversion to quantum states. In general, both kinds of questions involve calculating or understanding more about $\mathcal{D}_F$ and $\mathcal{D}_D$.

As mentioned, it is known that $\mathcal{D}^{\text{tr}} = \mathcal{D}^{\text{tr}}{}_D$. Thus to answer these sorts of questions with respect to the trace distance, we must understand $\mathcal{D}^{\text{tr}}{}_F$. In the concluding section, we explicitly characterize $\mathcal{D}^{\text{tr}}{}_F$ for qubit states.

5.2.4 Trace-Distance and the Qubit System

The trace distance of formation has the following general form,

$$\mathcal{D}^{\text{tr}}{}_F(\rho, \sigma) = \inf_{\substack{\Phi \in \mathcal{CQ} \\ \vec{p}, \vec{q} \in \mathcal{V}_{\rho\sigma}^{\Phi}}} \frac{1}{2} \sum_{m=1}^{M} |p_m - q_m| \tag{5.44}$$

where M is dependent on Φ, and in particular is the number of quantum states in the set $\mathcal{P}$ corresponding to a $\Phi \in \mathcal{CQ}$. Our goal is to calculate (5.44) for any qubit states ρ and σ.

We start by assuming that the infimum in the definition of $\mathcal{D}^{\text{tr}}{}_F$ is attainable. The analysis will show that this is justified. Then, in general, the problem can be stated as follows: for a pair of quantum states in $\mathbb{C}^2$, find

1. a set of states $\mathcal{P} = \{\sigma_m\}_{m=1}^{M}$ for which ρ and σ are both contained in the in the convex hull,
2. and probability vectors $\vec{p}, \vec{q} \in \Delta_M$ such that $\sum_m p_m \sigma_m = \rho$, $\sum_m q_m \sigma_m = \sigma$ which ...
3. minimize $\frac{1}{2} \sum_{m=1}^{M} |p_m - q_m|$.

4. Then write down a formula for the minimized value in terms of ρ and σ, or parameters which uniquely specify them.

Firstly we show that it suffices to consider any potential set of states $\mathcal{P}$ to be a set of pure states.

Lemma 5.3. *For any $\mathcal{P} = \{\sigma_m\}_{m=1}^M$ and probability vectors $\vec{p}$, $\vec{q}$, define $\rho = \Phi_{\mathcal{P}}(\vec{p})$ and $\sigma = \Phi_{\mathcal{P}}(\vec{q})$ (the map $\Phi_{\mathcal{P}}$ was defined in the statement of Proposition 5.3). Then there exists a set $\mathcal{P}' = \{\Pi_m\}_{m=1}^{M'}$ of pure states, and probability vectors $\vec{p'}$, $\vec{q'}$ with $\rho = \Phi_{\mathcal{P}'}(\vec{p'})$ and $\sigma = \Phi_{\mathcal{P}'}(\vec{q'})$ such that $\mathcal{D}^{\mathrm{tr}}(\vec{p'}, \vec{q'}) = \mathcal{D}^{\mathrm{tr}}(\vec{p}, \vec{q})$*

Proof. We have $\mathcal{D}^{\mathrm{tr}}(\vec{p}, \vec{q}) = \frac{1}{2}\sum_{m=1}^M |p_m - q_m|$. It will be enough to show that we can replace one mixed state in $\mathcal{P}$ with pure states without altering the induced classical trace distance. Without loss of generality, suppose σ_1 is a mixed state. Let $\sum_{j=1}^n z_j \Pi_j$ be any pure state decomposition of σ_1. Then we have $\rho = \sum_{j=1}^n p_1 z_j \Pi_j + \sum_{m=2}^M p_m \sigma_m$ and similarly for σ. Consider the classical trace distance on the probability distributions $\vec{p'} = (p_1 z_j, \ldots, p_1 z_n, p_2 \ldots, p_M)$ and $\vec{q'} = (q_1 z_j, \ldots, q_1 z_n, q_2 \ldots, q_M)$. We have

$$\mathcal{D}^{\mathrm{tr}}(\vec{p'}, \vec{q'}) = \tfrac{1}{2}\sum_{j=1}^n |z_j(p_1 - q_1)| + \sum_{m=2}^M |p_m - q_m| = \sum_{m=1}^M |p_m - q_m| = \mathcal{D}^{\mathrm{tr}}(\vec{p}, \vec{q}).$$

□

In the following we show that the set $\mathcal{P}$ which gives $\mathcal{D}^{\mathrm{tr}}{}_F$ for a given ρ and σ contains at most three pure states, and we discuss how these pure states should be chosen. The analysis will conclude with a formula for $\mathcal{D}^{\mathrm{tr}}{}_F$, as well as a calculable condition for when $\mathcal{D}^{\mathrm{tr}}{}_F = \mathcal{D}^{\mathrm{tr}} = \mathcal{D}^{\mathrm{tr}}{}_D$.

The analysis is heavily rooted in geometric intuition provided by the fact that we can view quantum states on $\mathbb{C}^2$ as points on or in the Bloch sphere $\mathcal{S}^2$. The trace distance itself has a convenient interpretation in terms of points in the Bloch sphere. Specifically, if $\vec{r}$ and $\vec{s}$ are the Bloch vectors (see Section 1.3) of states ρ and σ respectively, then [49]

$$\mathcal{D}^{\mathrm{tr}}(\rho, \sigma) = \frac{1}{2}|\vec{r} - \vec{s}|. \tag{5.45}$$

Thus Lemma 5.5 implies that ρ and σ are such that $\mathcal{D}^{\mathrm{tr}}{}_F = \mathcal{D}^{\mathrm{tr}}{}_D$ if and only if their distinguishability of formation is equal to half the straight-line distance between the Bloch vectors of ρ and σ.

Consider any decomposition of the state ρ in terms of pure states. This involves a set $\mathcal{P} = \{\Pi_m\}_{m=1}^M$ of pure states containing ρ in the convex hull, and vector $\vec{p}$ such that $\rho = \sum_{m=1}^M \vec{p}_m \Pi_m$. In the Bloch sphere picture, we can view this decomposition as a piecewise linear path of unit length, originating at the maximally mixed state $\frac{\mathbb{I}}{2}$

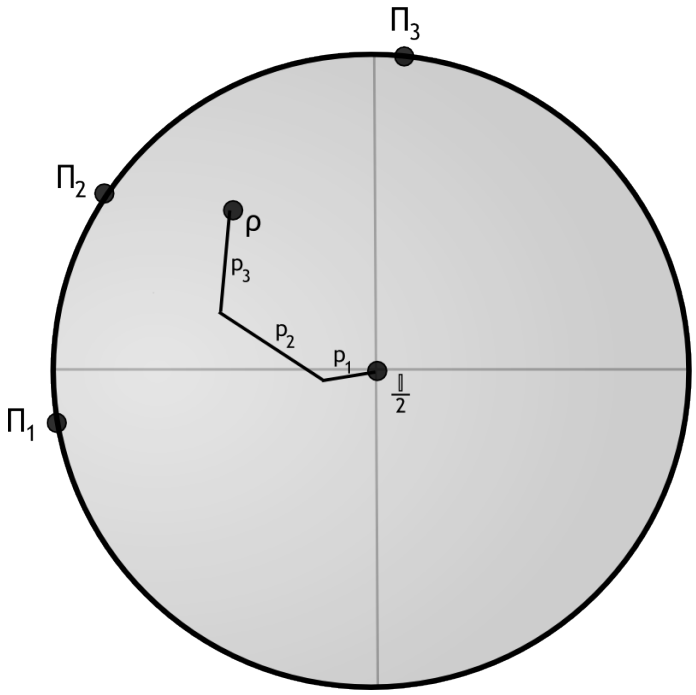

Figure 5.1: The quantum qubit state ρ as a path of length one. Each line segment in the path corresponds to a term in the pure-state decomposition of ρ

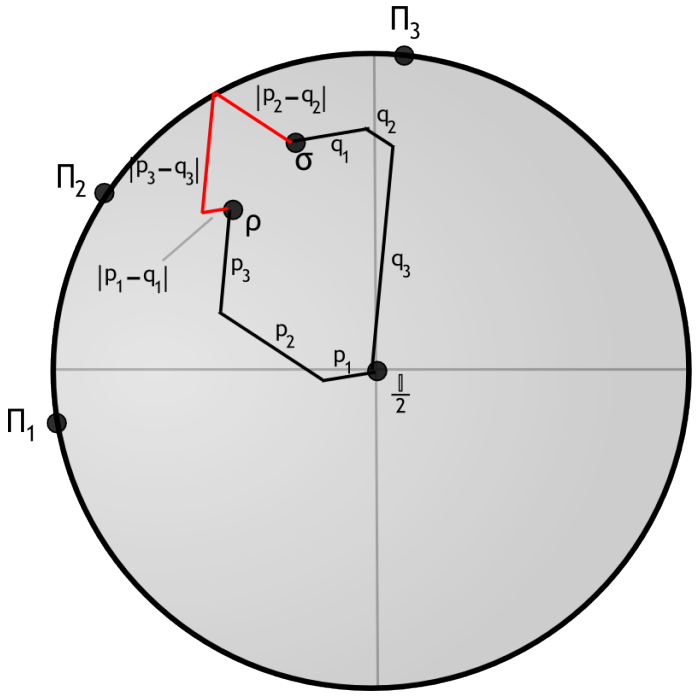

Figure 5.2: A geometric picture of equation (5.46). The difference between path A ($\frac{\mathbb{I}}{2}$ to ρ) and path B ($\frac{\mathbb{I}}{2}$ to σ) is the red path C, which runs from ρ to σ. Minimizing the length of C is equivalent to finding the trace distance of formation.

and concluding at the 'point' ρ. We view each pure state as a unit vector specifying a direction from the origin $\frac{\mathbb{I}}{2}$ (see Figure 5.1). The path commences with a segment of length p_1 in the direction given by the vector Π_1, followed by a segment of length p_2 in the direction given by the vector Π_2, and so on up to the M^{th} segment. However, just as we could reorder the terms in the decomposition, we could reorder the segments of the path. Furthermore, we could cut each segment into smaller pieces, and rearrange all of these into a more fragmented path. Thus, any decomposition of ρ in terms of $\mathcal{P}$, given by probability vector $\vec{p}$, defines a class of unit length piecewise linear paths from $\frac{\mathbb{I}}{2}$ to ρ. For the remainder of the present discussion, we need not specify a set $\mathcal{P}$ containing ρ in its convex hull, nor as a specific decomposition of ρ in terms of $\mathcal{P}$. Instead, we discuss piecewise linear paths of unit length from $\frac{\mathbb{I}}{2}$ to ρ, and the corresponding set $\mathcal{P}$ and decomposition is implied.

We can also include a notion of $\mathcal{D}^{\mathrm{tr}}{}_F(\rho, \sigma)$ into the geometric picture. Suppose that we have a unit length path to ρ, A, and a unit length path to σ, B, which respectively specify pure state sets $\mathcal{P}$ and $\mathcal{Q}$, and probability vectors $\vec{p}$ and $\vec{q}$. If we take the union of these two sets, and call it $\mathcal{P}'$ then we are guaranteed that both ρ and σ lie in the convex hull of $\mathcal{P}'$. For simplicity, we will assume that A and B are as non-fragmented as possible. That is each direction occurs only once in A and B. Subtracting path A piece-wise from path B gives a path from ρ to σ, which we call C. C will have a segment in the Π_m direction of length $|p_m - q_m|$ for all $1 \leq m \leq M$ (see Figure 5.2). The total length of C is

$$\sum_{m=1}^{M} |p_m - q_m|, \tag{5.46}$$

which is proportional to the value we are trying to minimize (5.44). Thus if we can find paths A and B for which the length of the path difference is minimized, then we have found the distinguishability of preparation for the states ρ and σ. Moreover, if we can find the set of state pairs (ρ, σ), such that C can be made to be the straight line between ρ and σ, then we will have classified the states for which there is no gap between $\mathcal{D}^{\mathrm{tr}}{}_F$ and $\mathcal{D}^{\mathrm{tr}}{}_D$.

Valuable geometric insight into the problem can be gained by answering the following question: What sort of bounds exist on any path, A, to ρ? Since the path must be of length 1, it follows that it must lie within the volume of a solid ellipsoid, E_ρ, having $\frac{\mathbb{I}}{2}$ and ρ as its focii, and a major axis length of 1. Any path of length one from $\frac{\mathbb{I}}{2}$ to ρ must lie within such an ellipsoid, since the points reachable by such paths are an exact description of such an ellipsoid (see Figure 5.3).

The problem of minimizing (5.46) in this geometric picture can be viewed equivalently as the problem of maximizing the amount of overlap between the paths A and B. This will leave a minimal difference between the two paths. Suppose we reorder A and B such that all overlap occurs at the beginning of the paths. Then all overlap between A and B must then lie within $E = E_\rho \cap E_\sigma$, the intersection of the bounding ellipsoids for ρ and σ (see Figure 5.4). Otherwise, either A or B would fall outside of its bounding ellipsoid.

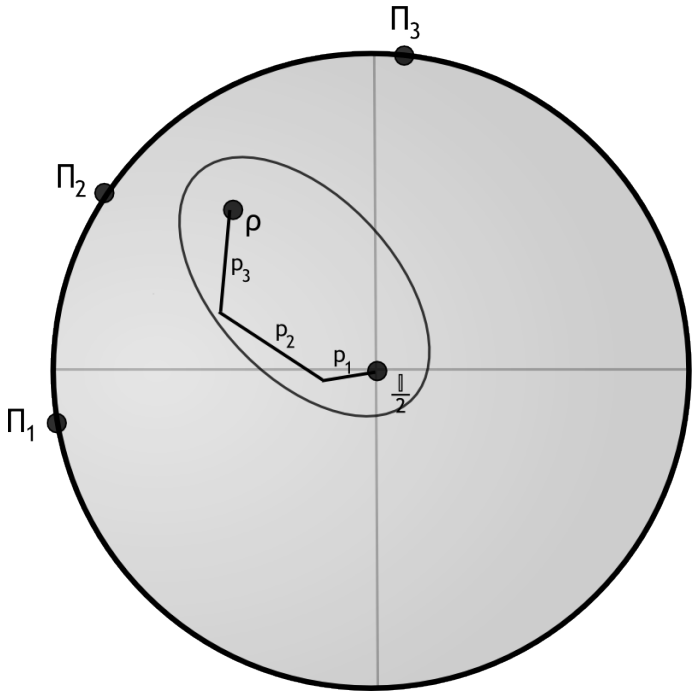

Figure 5.3: The path representations of ρ are bounded by the ellipse E_ρ. Any path extending outside of E_ρ necessarily has length greater than 1.

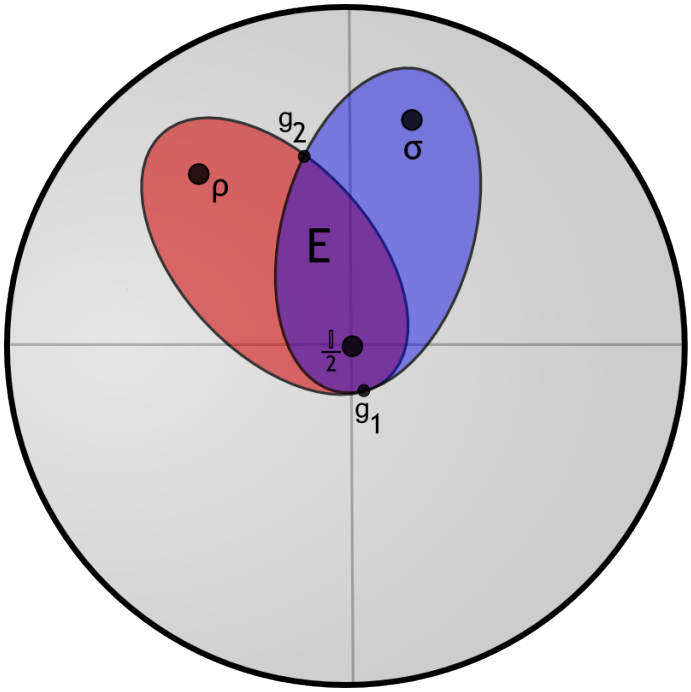

Figure 5.4: All possible overlap between a path to ρ and a path to σ must lie in the intersection of E_ρ and E_σ, denoted E in this 2-dimensional slice. The points g_1 and g_2 play an important role in determining the trace distance of formation.

We can now obtain the following characterization of the states for which $\mathcal{D}^{\text{tr}}{}_D = \mathcal{D}^{\text{tr}}{}_F$. We will call such state pairs *reversibly distinguishable*. The result also gives a construction on how to calculate $\mathcal{D}^{\text{tr}}{}_F$ in general.

Theorem 5.5. *Consider two states ρ and σ, the intersection E of their bounding ellipsoids, and the line segment L joining ρ and σ. Then ρ and σ are reversibly distinguishable if and only if L intersects E. The decompositions of ρ and σ which give rise to the minimization $\mathcal{D}^{\text{tr}}{}_F(\rho, \sigma)$ lie in the same plane as ρ, σ and $\frac{\mathbb{I}}{2}$, which we call F, and contain three pure states in total.*

Proof. Suppose that L passes through E. It is now easy to construct two paths, A and B, from $\frac{\mathbb{I}}{2}$ to ρ and σ respectively, whose difference is the line L (see Figure 5.5). Both paths may start off with a segment from $\frac{\mathbb{I}}{2}$ to a point on L within E. Both can then finish with segments along the line L. Exactly three directions are used in total, so the corresponding decompositions involve three pure states in the plane F. Thus ρ and σ are reversibly distinguishable.

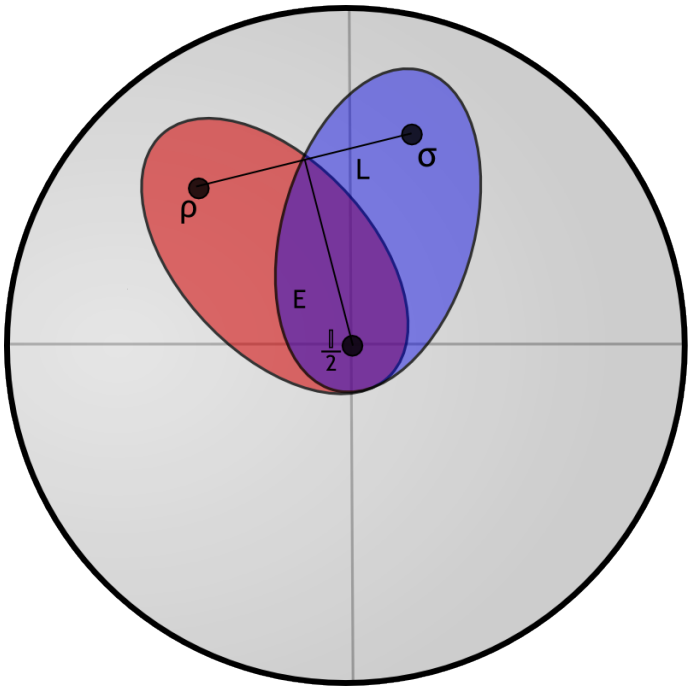

Figure 5.5: Depicted here is the case where L intersects E. Paths A and B are constructible such that all difference between the two lies in L.

Suppose that L does not intersect E. Recall that minimizing C is equivalent to maximizing the overlap of paths A and B. Thus we desire to find the point x in E which minimizes the function

$$D(x) = ||\rho - x|| + ||\sigma - x|| \,.$$

We claim that this point will be one of the two intersection points of the bounding ellipses in the plane F. These points are labeled g_1 and g_2 in Figure 5.4. The

paths A and B will then consists a joint segment from $\frac{\mathbb{I}}{2}$ to one of g_1 or g_2, followed by a concluding segment to ρ or σ. In order to substantiate this claim, we first remark that the desired point x must lie in the plane F. Any point not in F can be projected onto F, and thus will be closer to both ρ and σ. We now argue via two arguments that D is minimized on points which are equidistant to ρ and σ.

Subclaim 1: For any $d \geq 0$, consider the line of points L_d which have perpendicular distance d to L. Of the points on L_d, the unique point x_d which is equidistant to ρ and σ minimizes D. This claim can be verified with a simple calculus exercise.

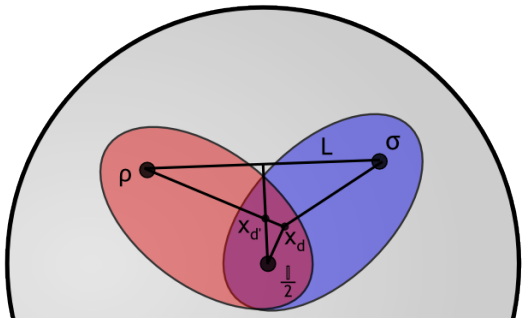

Figure 5.6: Existence of x_d implies existence of $x_{d'}$. The central line indicates the set of points which are equidistant to ρ and σ. If we postulate that the paths $\frac{\mathbb{I}}{2}$-x_d-ρ and $\frac{\mathbb{I}}{2}$-x_d-σ have length at most one, then it is clear that $\frac{\mathbb{I}}{2}$-$x_{d'}$-ρ and $\frac{\mathbb{I}}{2}$-$x_{d'}$-σ will also have length at most one.

Given that equidistant points are better than non-equidistant points, it remains to show that we can in general include such a point in paths A and B to ρ and σ.

Subclaim 2: If there is a point x_d, which is a perpendicular distance d from L and lies in $E \cap F$, then there is a point $x_{d'}$, which is equidistant to ρ and σ, has perpendicular distance d' from L with $d' \leq d$, and lies in $E \cap F$. See Figure 5.6 for a visual aid. Consider paths A (from $\frac{\mathbb{I}}{2}$ to x_d to ρ) and B (from $\frac{\mathbb{I}}{2}$ to x_d to σ). Since x_d is in $E \cap F$, the lengths of both of these paths are no greater than one. Supposing that x_d is not already equidistant to ρ and σ, then at least one of A or B passes through such a point $x_{d'}$, after passing through x_d. Suppose this path is A. The point $x_{d'}$ will necessarily be no further from L than x_d, and thus $D(x_{d'}) \leq D(x_d)$. It is clear that a path of length at most one exists through $x_{d'}$ to ρ, as this is the path that A is taking. However, since $x_{d'}$ is equidistant to ρ and σ, a path of length at most one exists from $\frac{\mathbb{I}}{2}$ to $x_{d'}$ to σ. Therefore x'_d is in $E \cap F$.

We have shown that to minimize the length of C, both paths should pass through a point equidistant to ρ and σ and in the plane F. Both g_1 and g_2 are such points, and all such points in $E \cap F$ are convex combinations of g_1 and g_2. Hence, one of g_1 or g_2 is the required point in $E \cap F$, and hence E, which minimizes D. Call that point g. Since g is at an intersection point of E_ρ and E_σ, the most trivial paths to ρ and σ through g already have length one. The length of the difference of these two paths is equal to the minimized quantity $D(g)$. Thus, the decompositions

given by these paths give $\mathcal{D}^{\text{tr}}{}_F(\rho, \sigma)$. This value must be greater than the length of L, since we assumed that g did not lie on the line L. Thus ρ and σ are not reversibly distinguishable. Also, the paths (decompositions) lie in the plane F, and are comprised of three pure states in total. □

Since the proof of the above claim was constructive, we can relatively easily derive formulaic expressions for $\mathcal{D}^{\text{tr}}{}_F(\rho, \sigma)$ and the reversible distinguishability condition. We first parameterize the relevant aspects of ρ and σ in polar coordinates. Viewing ρ and σ by their Bloch vectors, let a and b be the magnitude of these vectors (magnitude 1 implies the state is pure, whereas magnitude 0 implies the state is $\frac{\mathbb{I}}{2}$), and let θ be the angle between the Bloch vectors.

Firstly, the length of the line segment L between ρ and σ is given by:

$$d = \sqrt{a^2 + b^2 - 2ab\cos(\theta)}. \tag{5.47}$$

In the argument for the reversibly distinguishable case, we initiated A and B with a line segment from $\frac{\mathbb{I}}{2}$ to L. The length of this line segment is minimized if we take it to be the path perpendicular to L from $\frac{\mathbb{I}}{2}$. A little bit of trigonometry will reveal that the perpendicular distance from $\frac{\mathbb{I}}{2}$ to L can be written as

$$e = \frac{ab\sin(\theta)}{d}. \tag{5.48}$$

The perpendicular line segment from $\frac{\mathbb{I}}{2}$ to L, which has length e, splits L into two pieces which have lengths

$$f = \frac{a(a - b\cos(\theta))}{d} \quad \text{and} \quad g = \frac{b(b - a\cos(\theta)))}{d}. \tag{5.49}$$

Thus we know the line L lies in the ellipse intersection E if

$$e + f = \frac{ab(\sin(\theta) - \cos(\theta)) + a^2}{d} <= 1 \tag{5.50}$$

and

$$e + g = \frac{ab(\sin(\theta) - \cos(\theta)) + b^2}{d} <= 1. \tag{5.51}$$

If (5.50) and (5.51) are satisfied, then the trace distance of formation is $\frac{d}{2}$ and ρ and σ are reversibly distinguishable in terms of the trace distance.

If (5.50) and (5.51) are not satisfied then we need to calculate the intersection point g of E_ρ and E_σ which lies in the plane F and minimizes D. Then the distinguishability of formation will be $\frac{D(g)}{2}$.

Appealing to ellipse equations [62], we can calculate the distance between the intersection point g and ρ (or σ), and hence the trace distance of formation, as

$$\frac{D(g)}{2} = \sqrt{r_\rho(\phi_I)^2 + a^2 - 2ar(\phi_I)\cos(\pi - \phi_I)}, \tag{5.52}$$

where

$$r_\rho(\phi) = \frac{\frac{1}{2}(1-a^2)}{1+a\cos(\phi)} \tag{5.53}$$

is the radius-valued equation of the planar ellipse E_ρ, and

$$\phi_I = \arccos\left(\frac{b^2-a^2}{d'}\right) + \arctan\left(\frac{b'\sin(\theta)}{a'-b'\cos(\theta)}\right) \tag{5.54}$$

gives the angle of the minimal intersection point g of E_ρ and E_σ, and

$$d' = \sqrt{a'^2 + b'^2 - 2a'b'\cos(\theta)}, \quad a' = a(1-b^2), \text{ and } b' = b(1-a^2).$$

An appealing question is whether or not we can derive simpler expressions for the reversibility condition and for $\mathcal{D}^{\text{tr}}{}_F$ when it is not equal to $\mathcal{D}^{\text{tr}}{}_D$. Such expressions may come about through algebraic rather than geometric considerations.

5.3 Future Work

The study of distinguishability as a resource is a relatively new endeavour, and many open problems remain. A most obvious immediate follow up to this work would be to continue the characterization of the trace distance of formation to arbitrary dimensional quantum systems. One could investigate further the shape of the region of reversible distinguishability for a given distinguishability measure, perhaps by considering ϵ-balls around the maximally mixed state, or investigating whether or not certain classes of states are always reversibly distinguishable or always not.

Another interesting problem would be to study the relationship between distinguishability and compatibility [52]. With compatibility, we consider two parties with different information about the exact same quantum system. Given their differing information, they may describe the system differently, with distinct quantum states ρ and σ. Compatibility attempts to quantify the likelihood that the subjective states ρ and σ could arise from the same true quantum state. In some sense compatibility is inverse to distinguishability, but fully understanding the relationship between the two notions could be mutually beneficial.

APPENDICES

Appendix A

Measure Theory

The following exposition on measure theory is derived from [21].

A.1 σ-algebras

Let $\mathcal{X}$ be a non-empty set.

Definition A.1 (σ-algebra)**.** A *σ-algebra* on $\mathcal{X}$ is a nonempty collection Σ of subsets of $\mathcal{X}$ such that:

- $\emptyset \in \Sigma$ and $\mathcal{X} \in \Sigma$,
- if $\{E_i\}_{i=1}^{\infty} \subset \Sigma$ then $\bigcup_{i=1}^{\infty} E_i \in \Sigma$, and
- if $E \in \Sigma$ then $E^c \in \Sigma$,

where E^c is the complement of the set E.

Proposition A.1. *The intersection of any family of σ-algebras on $\mathcal{X}$ is a σ-algebra on $\mathcal{X}$.*

Since the *power-set* of $\mathcal{X}$, $\mathcal{P}(\mathcal{X})$, is a σ-algebra, if $\mathcal{E}$ is any subset of $\mathcal{P}(\mathcal{X})$, then there exists a unique smallest σ-algebra, $\mathcal{M}(\mathcal{E})$, containing $\mathcal{E}$.

Definition A.2 (Generated σ-algebra)**.** For any $\mathcal{E} \subset \mathcal{P}(\mathcal{X})$, define the *$\sigma$-algebra generated by $\mathcal{E}$* as the intersection of all σ-algebras containing $\mathcal{E}$.

A particularly important example of a generated σ-algebra is the *Borel σ-algebra*.

Definition A.3 (Borel σ-algebra)**.** If $\mathcal{X}$ is a metric space (or any topological space), then the σ-algebra generated by the open sets (the topology) of $\mathcal{X}$ is called the *Borel σ-algebra of $\mathcal{X}$*.

For the purposes of defining a *product measure* later, we now define a *product σ-algebra*.

Definition A.4 (Product σ-algebra)**.** Let $\{\mathcal{X}_a\}_{a\in A}$ be a collection of non-empty sets, with σ-algebras Σ_a. Define the *product set* as

$$X = \prod_{a\in A} \mathcal{X}_a \equiv \{(x_a)_{a\in A} \mid x_a \in \mathcal{X}_a \; \forall a \in A\}, \tag{A.1}$$

and the *coordinate maps* $\pi_a : \mathcal{X} \to \mathcal{X}_a$, $\pi_a((x_b)_{b\in A}) = x_a$. Then the *product σ-algebra* on $\mathcal{X}$ is the σ-algebra generated by the set

$$\left\{\pi_a^{-1}(E_a) \mid E_a \in \Sigma_a,\; a \in A\right\}^{1}. \tag{A.2}$$

We denote this σ-algebra as $\otimes_{a\in A}\Sigma_a$.

A.2 Measures

Let $\mathcal{X}$ be a non-empty set with a σ-algebra Σ.

Definition A.5 (Measure)**.** A *measure* on $\mathcal{X}$ is a function $\mu : \Sigma \to [0,\infty]$ such that

- $\mu(\emptyset) = 0$,
- if $\{E_i\}_{i=1}^{\infty}$ is a sequence of disjoint sets in Σ, then $\mu(\cup_i E_i) = \sum_i \mu(E_i)$.

In the context of a measure μ, the elements of Σ are called *measurable sets*, the pair $(\mathcal{X}, \Sigma)$ is called a *measurable space*, and the triplet $(\mathcal{X}, \Sigma, \mu)$ is called a measure space.

Definition A.6 (Finite / Probability Measures)**.** If $\mu(\mathcal{X}) < \infty$ then μ is a *finite measure*. In particular, if $\mu(\mathcal{X}) = 1$, then μ is a *probability measure*.

In the context of probability theory, a probability measure on a sample space gives the probabilities of an event in some measurable set occurring. The support of a measure in some sense indicates the subset of events that could possibly occur.

Definition A.7 (Support of a Measure)**.** The support of a measure μ is defined as:

$$\operatorname{supp}\mu = \{\lambda | \lambda \in N_\lambda \in \Sigma \Rightarrow \mu(N_\lambda) > 0\}.$$

Definition A.8 (Product Measure)**.** Let $\mathcal{X} = \prod_{a\in A}\mathcal{X}_a$ be a product space with a product σ-algebra $\otimes_{a\in A}\Sigma_a$, and μ_a a measure on $\mathcal{X}_a$ for all $a \in A$. The *product measure on* $\mathcal{X}$ is the unique measure on $\mathcal{X}$ such that if $E \in \otimes_{a\in A}\Sigma_a$ is expressible as $E = \otimes_{a\in A}E_a$, where $E_a = \mathcal{X}_a$ for all but a finite set $J \subset A$ then

$$\mu(B) = \prod_{j\in J} \mu_a(E_j). \tag{A.3}$$

[1]The notation $\pi_a^{-1}(E_a)$ denotes the pullback of the set E_a through π_a. Thus it is the set $\{(x_b)_{b\in A} \mid x_a \in E_a\}$.

A.3 Integration of Non-Negative Functions

Definition A.9 (Measurable Function)**.** Suppose that $(\mathcal{X}_1, \Sigma_1)$ and $(\mathcal{X}_2, \Sigma_2)$ are measurable spaces. Then $f : \mathcal{X}_1 \to \mathcal{X}_2$ is a *measurable function* if $E \in \Sigma_2$ implies $f^{-1}(E) \in \Sigma_1$.

As an important example, suppose that $\mathcal{X}_2 = \mathbb{R}$ and Σ_2 is the Borel σ-algebra on $\mathbb{R}$. Then a measurable function $f : \mathcal{X}_1 \to \mathbb{R}$ is called a *Borel function*.

We now proceed to define integration for Borel functions. Let $(\mathcal{X}, \Sigma, \mu)$ be a measure space.

Definition A.10 (Characteristic and Simple Functions)**.** For any subset $E \subset \mathcal{X}$, the *characteristic function of* E is the function

$$\chi_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}. \tag{A.4}$$

A *simple function* is any finite linear combination of characteristic functions of measurable sets:

$$f = \sum_{i=1}^{n} z_i \chi_{E_i}, \tag{A.5}$$

for $E_i \in \Sigma$, $z_i \in \mathbb{R}$.

Integration on simple functions is defined easily, and the following result allows the generalization to measurable functions.

Proposition A.2. *If $f : \mathcal{X} \to [0, \infty]$ is Borel measurable, then there is a sequence $\{f_n\}$ of simple functions such that $0 \leq f_1 \leq f_2 \ldots \leq f$, and f_n converges to f pointwise and uniformly on any set on which f is bounded.*

Definition A.11 (Lebesgue Integration of Simple Functions)**.** If $f = \sum_{i=1}^{n} z_n \chi_{E_i}$ is a simple function, then the *Lebesgue integral of* f is given by

$$\int_{\mathcal{X}} f \, d\mu = \sum_{i=1}^{n} z_i \mu(E_i). \tag{A.6}$$

Definition A.12 (Lebesgue Integration for Non-Negative Borel Functions)**.** If $f : \mathcal{X} \to [0, \infty]$ is Borel measurable, then we define the *Lebesgue integral of* f as

$$\int_{\mathcal{X}} f \, d\mu = \sup \left\{ \int_{\mathcal{X}} g \, d\mu \mid 0 \leq g \leq f, \, g \text{ simple} \right\}. \tag{A.7}$$

Appendix B

Probability Theory

We present here a simple exposition of some of the basic concepts and structures of classical probability theory as originally given by Kolmogorov [42]. We also present a generalization of classical probability theory called fuzzy probability theory [15], which encapsulates the idea that a knowing a precise physical state may not induce precise knowledge of an experimental outcome.

B.1 Classical Probability Theory

Classical probability theory concerns itself with the probability of occurrence of events within a set Ω of possible events. A particular element $\omega \in \Omega$ is called an *atomic event*. A subset $A \in P(\Omega)$ (the power set of events) is also considered to be an event; the event where any one of the atomic events $\omega \in A$ occurs. The set of possibly occurring events $\mathcal{A} \subseteq P(\Omega)$ is given by a σ-algebra of subsets of Ω. It need not be the case that the singleton sets $\{\omega\}$ are in $\mathcal{A}$. If they are, the algebra is called atomic. Due to the σ-algebra $\mathcal{A}$, it is possible to define a *probability measure* on the space Ω. We can think the elements of Ω as being possible configurations of some physical system, and the value $\mu(A)$ prescribing the probability that the system is in a configuration $\omega \in A$. The measure μ represents ignorance of the true configuration of the system.

We can now imagine performing an experiment on the system which has outcomes occurring in the set I. Suppose the set I also comes equipped with a σ-algebra $\mathcal{B}$. Consider any function $f : \Omega \to I$ from the event space to the outcome space. The function f is called *measurable* if whenever $B \in \mathcal{B}$, then $f^{-1}(B) \in \mathcal{A}$. Here, f^{-1} does not denote the functional inverse, but the pull-back operation i.e. $f^{-1}(B) = \{\omega \in \Omega | f(\omega) \in B\}$. Given a measure μ on Ω, such a function f is called a *random variable*. In the context of physics, f is also known as an *observable*; each observable assigns properties to the configurations of the system. If $f(\omega) = e$, then we say that the configuration, or state, ω has the value e for the observable f.

Notice that μ induces a probability distribution $\hat{\mu}$ on I via f:

$$\hat{\mu}(B) = \mu(f^{-1}(B)). \tag{B.1}$$

The function $\hat{\mu}$ is called the *distribution of the random variable* f. The distribution $\hat{\mu}$ is also known as an *effect*. An effect is typically a function which assigns a probability to an outcome of an experiment given a certain state of the system. In the case of classical probability theory, the state of the system is encompassed by the measure μ, and the experiment is given by f, whose outcomes lie in I. Thus we write the effect as $E_{f,B}(\mu) = \mu(f^{-1}(B))$.

B.2 Fuzzy Probability Theory

Classical probability theory is certainly a useful tool to reason about uncertainty in physics. However, in some senses it is still an idealization. The assumption of applying classical probability theory to physics is that all the uncertainty lies in knowledge, or lack-thereof, of the current configuration of the system. However, if $\omega \in \Omega$ is known precisely, then the outcome of any observable f is known precisely. In a physical experiment, it is reasonable to assume that there may be errors in measurement procedure. If the causes of these errors are not specifically modeled, or if there is simply a fundamental inability to measure precisely, then mapping a specific configuration ω to a specific outcome e is an over-simplification.

Bugajski and collaborators present in [15] a framework to account for such imprecise measurement scenario. Instead of an observable associating configurations in Ω with particular outcomes in I they suggest an association with a probability measure on the outcome space, $K : \Omega \to \mathcal{M}_1^+(I)$. In the context of a specific observable K, each actual configuration is associated with a set of possible outcomes, and a probability distribution over those outcomes. K can also be viewed as a bivariate function $K : \Omega \times \mathcal{B} \to [0, 1]$ called a *Markov kernel*. As a function on Ω, $K_B(\omega)$ is known as an *indicator function*. The set of indicator functions on a set S, $\{f : S \to [0, 1]\}$, is denoted by $[0, \chi_S]$.

The standard classical observables of the previous section can have their counterparts as Markov kernels. If $f : \Omega \to I$ is a classical random variable, then its counter part is $K^f(\omega, B) = \chi_{f^{-1}(B)}(\omega) = \delta_\omega(B) \in \{0, 1\}$. Notice that the Markov kernel for a classical random variable is idempotent (takes on values 0 or 1). These random variables, and the corresponding indicator function / Markov kernels are referred to as *sharp*. If the indicator function of K takes on values between 0 and 1, then it is referred to as *fuzzy*. Under reasonable assumptions about the spaces Ω and I, and assuming that both $\mathcal{A}$ and $\mathcal{B}$ are atomic, then the set of sharp indicator functions are the extreme points of the convex set $[0, \chi_\Omega]$ [15].

In the fuzzy probability framework, we again have a concept of the distribution of a random variable,

$$\hat{\mu}(B) := \int_\Omega K(\omega, B) d\mu(\omega) =: E_{K,B}(\mu) \tag{B.2}$$

where the effect corresponding to fuzzy random variable K and outcome-set B has also been defined.

Appendix C

Zero Lemma

Claim C.1. *If $\int_0^\theta f(\alpha,\theta)g(\alpha) = 0$ for all θ within a domain $D = [0, b]$, $f(\alpha, \theta)$ is positive and continuous within this domain, and $g(\alpha)$ is piece-wise continuous, then $g(\alpha) = 0$ in D.*

Supposing this was not true, then there would be some $\alpha_0 \in (0, b)$ such that $g(\alpha_0) \neq 0$. W.l.o.g, let us assume that $g(\alpha_0) > 0$. Then as long as g is at least piece-wise continuous, then there is a small region $\delta = (a, b') \subset D$ with $\alpha_0 \in \delta$ such that $g(\delta) > 0$. Then

$$\begin{aligned}\int_0^{b'} f(\alpha,\theta)g(\alpha) &= \int_0^{a} f(\alpha,\theta)g(\alpha) + \int_a^{b'} f(\alpha,\theta)g(\alpha) \\ &= 0 + \int_a^{b'} f(\alpha,\theta)g(\alpha) > 0\end{aligned}$$

which is a contradiction.

Appendix D

The Role of Contextuality of Protocols with a Quantum Advantage

D.1 Random Access Codes

A known example of the advantage of quantum information processing over classical information processing is with respect to the random access code task [2, 24]. Suppose that participant A has a string of m uniformly random bits, and wishes to send participant B n bits ($n < m$) such that B can learn any one of the original m bits that he chooses. A has no a priori knowledge of which of the m bits B might be interested in. Of course, this is impossible to do since the m-bit string is uniformly random and A can only send $n < m$ bits. The goal is to find a protocol which maximizes the probability that B is able to correctly learn the value of a single bit that he chooses. Such a protocol as described is called a $m \to n$ random access code.

In [2, 24] it is shown that using classical bits, the optimal $2 \to 1$ random access code protocol has a probability of success of $p_c = \frac{3}{4}$. Note that this is the probability of success averaged uniformly over the 2^m possible inputs to the protocol, and uniformly over the m possible choices B could make. As presented in [2, 24], there exists a quantum protocol in which two bits are encoded into 1 qubit (a $2 \to 1$ quantum random access code (QRAC)) which has a success rate beating the optimal classical success rate. This is reproduced here.

Depending on her input, $b = b_0 b_1$, A prepares one of $2^2 = 4$ states:

$$\begin{aligned} \psi_{00} &= \left| \tfrac{\pi}{4}, \tfrac{\pi}{2} \right\rangle\!\!\left\rangle \tfrac{\pi}{4}, \tfrac{\pi}{2} \right| \\ \psi_{01} &= \left| \tfrac{3\pi}{4}, \tfrac{\pi}{2} \right\rangle\!\!\left\rangle \tfrac{3\pi}{4}, \tfrac{\pi}{2} \right| \\ \psi_{11} &= \left| \tfrac{5\pi}{4}, \tfrac{\pi}{2} \right\rangle\!\!\left\rangle \tfrac{5\pi}{4}, \tfrac{\pi}{2} \right| \\ \psi_{10} &= \left| \tfrac{7\pi}{4}, \tfrac{\pi}{2} \right\rangle\!\!\left\rangle \tfrac{7\pi}{4}, \tfrac{\pi}{2} \right|. \end{aligned} \tag{D.1}$$

She then sends the state ψ_b to B. If B wishes the know the value of first bit, b_0, he

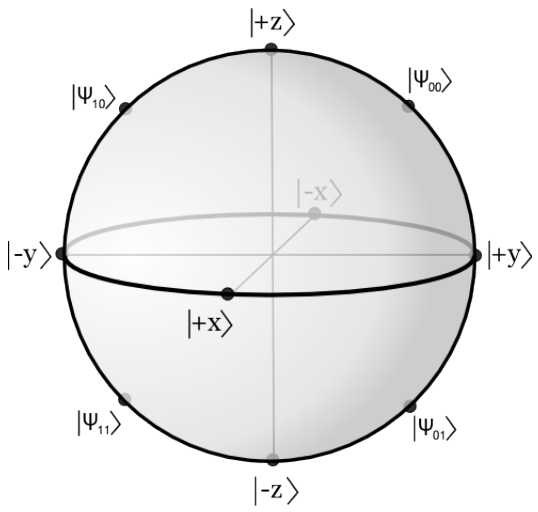

Figure D.1: The four states of the $2 \to 1$ QRAC protocol depicted on the Bloch sphere.

measures the state he receives along the y-axis i.e with the PVM

$$\left\{ \left| \tfrac{\pi}{2}, \tfrac{\pi}{2} \right\rangle\!\!\left\langle \tfrac{\pi}{2}, \tfrac{\pi}{2} \right|, \left| \tfrac{-\pi}{2}, \tfrac{\pi}{2} \right\rangle\!\!\left\langle \tfrac{-\pi}{2}, \tfrac{\pi}{2} \right| \right\}.$$

If the first outcome occurs, B guesses that $b_0 = 0$, and if the second outcome occurs, B guesses $b_0 = 1$. Similarly, if B wishes to know the value of the second bit, he measures along the z-axis. In both cases, given the state that A has sent, B's probability of receiving the measurement outcome which allows him to guess the correct bit-value is $p_q = \cos^2(\frac{\pi}{8}) \approx 0.85 > \frac{3}{4}$.

In Galvão's doctoral thesis [24], he makes an argument for why the above performance can be attributed to the contextuality of quantum theory.

D.1.1 Galvão's Necessity of Contextuality for QRAC

Firstly, recall that an experiment which verifies the non-locality of quantum theory is also an experiment which verifies the measurement contextuality of quantum theory 2.3.3. Galvão takes advantage of this fact by showing that the above protocol can be performed in a manner which is akin to an experiment which would violate a Bell inequality.

Participant A could prepare the states (D.1) for B by initially preparing the state

$$\frac{1}{\sqrt{2}}(|\,00\,\rangle + |\,11\,\rangle) \tag{D.2}$$

on two qubit systems. If she wants to prepare the state corresponding to $b = 00$ or $b = 11$, she measures the first qubit of (D.2) with the PVM given by $\{\psi_{00}, \psi_{11}\}$.

If she gets the first outcome then by properties of the state (D.2), she knows that the second system is in the ψ_{00} state and if she gets the second outcome, she knows that the system is in the ψ_{11} state. If either one of the outcomes is the opposite of what she wants to prepare for B, she performs a rotation (by π with respect to the x-axis) which puts the second qubit in the intended state. She can prepare the states corresponding to $b = 01$ and $b = 10$ strings via a similar procedure i.e measure the PVM $\{\psi_{01}, \psi_{10}\}$ and perform a conditional rotation. The resulting second system is then sent to Bob, who performs the measurement that he wishes.

Galvão argues that the procedure is statistically equivalent to the following. Alice and Bob start with the state (D.2) 50% of the time and start with state

$$\frac{1}{\sqrt{2}}(|\,00\,\rangle - |\,11\,\rangle) \tag{D.3}$$

50% of the time[1]. Alice then performs a measurement on the first subsystem in one of the two bases mentioned above. Bob measures the second subsystem in either the z or the y basis. Notice that these two measurements for A and B each are precisely the measurements required to maximally violate the Bell/CHSH inequality 2.3.2. Thus, if statistics of outcomes are compared for the cases when they started with (D.2)and separately for the cases when they started with (D.3), then they will find that for the separate cases, their statistics will violate a Bell/CHSH inequality, and thus they will have demonstrated contextuality. Since the particular QRAC protocol presented can be viewed as CHSH experiment in this fashion, and the success rate for this protocol mimics the statistics of a Bell/CSHS inequality violation, then Galvão's conclusion is that the success of the QRAC protocol is due to the contextuality of quantum theory.

D.1.2 A Counter-Argument to Galvão's Analysis

The following seemingly straightforward criticism of Galvão's argument has not been found in the literature. However, it is clear that although A *may* prepare the states (D.1) as Galvão indicates, it is not necessary. That is, A is not *forced* to use a bipartite quantum state (a state in $\mathcal{H}_4$) in order to perform the protocol. She may instead simply have access to only qubit states which she can prepare and send to B. In a setup where A is restricted to only single qubits at a time, she and B can still perform the protocol properly. Thus the contextuality of $\mathbb{Q}_4$ is not necessary.

D.2 Parity-Oblivious Multiplexing

In his paper on generalized contextuality [59], Spekkens hypothesizes that contextuality may be responsible for some advantages that quantum information processing

[1]This is because on half of the runs through the random access code protocol Alice will perform the rotation on the second system. This rotation, applied to state (D.2) gives (D.3).

has over classical information processing. In a subsequent paper [60], Spekkens describes an information process task called parity oblivious multiplexing (POMP henceforth). It is proven that an operational theory for which a better-than-'classical' protocol exists must be preparation contextual.

The task is as follows. Two parties A and B wish to pick a protocol which maximizes their success rate for a goal while maintaining a restriction on information available to B. In particular, the first party, A, receives a random n-bit binary string $y = y_1 y_2 \dots y_n \in \{0,1\}^n$. A can then send any message m to a second party B, as long as the message does not allow B to gain any parity information about y. That is for every binary string $s \in \{0,1\}^n$ such that the Hamming weight (number of 1s) of s is greater than 1, B may not gain any information about $s \cdot y$ (the inner product of the binary strings, $\mod 2$). By information, it is meant that having received m and performed any possible measurement on it, B must not have a better than 50% change of correctly guessing the parity of m with respect to any valid s. Having received the message m, B then also receives a random value $i \in 1 \dots n$. The protocol is successful if B correctly outputs the binary value of y_i.

It is shown in [60] that the best classical (involving classical bits) protocol is to have A send one agreed upon bit of y, say $m = y_1$. If B receives the random value $i = 1$, then m is output. Otherwise B outputs 1 or 0 at random. The probability of success is $p_s = \frac{1}{n} + \frac{n-1}{n}\frac{1}{2} = \frac{n+1}{2n}$. In the case of $n = 2$ POMP, this gives $p_c = \frac{3}{4}$.

There exists a quantum protocol for $n = 2$ POMP, which exactly mimics the QRAC protocol. Truly, the only difference in the protocol is that B's choice for which bit to learn is given randomly, not chosen by B. In terms of the goals of the task, the difference is that there is information which B must not learn, namely any parity information of the input string y. Again, the success rate of the quantum protocol for POMP is $p_q = \cos^2(\frac{\pi}{8}) \approx 0.85 > \frac{3}{4}$.

The main result of [60] is a derivation of an upper bound on the success rate of any preparation non-contextual operational theory used to perform a POMP protocol. Suppose we have an operational theory $(\mathcal{P}, \mathcal{M}, I, \Pr)$. With this operational theory, A and B can formulate a POMP protocol. Given an input y, A will perform a preparation P_y on the system $\mathcal{S}$. The system will be sent to B, who upon receiving input integer i will perform measurement M_i on $\mathcal{S}$. Given that B must decide on a value 0 or 1, this implies that M_i is essentially a two-outcome measurement, which we will label $k = 0$ and $k = 1$. The k outcome indicates that B guesses $y_i = k$. Thus the probability of success for this protocol in given operational theory is

$$\sum_{y \in 0,1^n} \sum_{i=1}^{n} \frac{1}{n2^n} \Pr(y_i | P_y, M_i). \tag{D.4}$$

This protocol is parity-oblivious if $\forall s \in \{0,1\}^n$ and $\forall (M,k) \in \mathcal{M} \times I$

$$\sum_{s \cdot y = 0} \Pr(P_y | k, M) = \sum_{s \cdot y = 1} \Pr(P_y | k, M) \tag{D.5}$$

where $\Pr(P_y|k, M)$ is calculated by Bayesian inversion. It is then assumed that $(\mathcal{P}, \mathcal{M}, I, \Pr)$ has a convex and preparation non-contextual ontological model (Λ, μ, ξ). In such a framework, it is proven that the optimal classical bound ($p_c = \frac{3}{4}$) is an upper bound on the success rate. Thus, preparation contextuality must be necessary in order to beat the classical bound. Of course, a quantum protocol exists which beats this bound, thus the analysis of [60] counts as another proof of preparation contextuality.

References

[1] David Albert. *Quantum Mechanics and Experience*. Harvard University Press, 1994. 36

[2] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata. In *Proceedings of the 31st Annual ACM Symposium on the Theory of Computing*, 1999. 104

[3] V. I. Arnold. *Mathematical Methods of Classical Mechanics (Graduate Texts in Mathematics)*. Springer, September 1997. 1

[4] L. E. Ballentine. The statistical interpretation of quantum mechanics. *Rev. Mod. Phys.*, 42(4):358–381, Oct 1970. 2, 42, 43

[5] Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. Noncommuting mixed states cannot be broadcast. *Phys. Rev. Lett.*, 76(15):2818–2821, Apr 1996. 81

[6] J. S. Bell. On the einstein podolsky rosen paradox. *Physics*, 1:195–200, November 1964. 1, 29, 30, 31, 32

[7] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics*, chapter Bertlmann's Socks and the Nature of Reality, pages 139–158. Cambridge University Press, 2004. 29, 30, 31

[8] John S. Bell. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38(3):447–452, Jul 1966. 1, 37

[9] E G Beltrametti and S Bugajski. A classical extension of quantum mechanics. *Journal of Physics A: Mathematical and General*, 28(12):3329–3343, 1995. 25

[10] C. H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. Wooters. Teleporting an unknown quantum state via dual classical and epr channels. *Phys. Rev. Lett.*, 70:1895, 1993. 71

[11] Garrett Birkhoff and John Von Neumann. The logic of quantum mechanics. *The Annals of Mathematics*, 37(4):823–843, 1936. 16

[12] David Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables. i. *Phys. Rev.*, 85(2):166–179, Jan 1952. 33

[13] David Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables. ii. *Phys. Rev.*, 85(2):180–193, Jan 1952. 33

[14] Bela Bollobas. *Linear Analysis*. Cambridge University Press, 1999. 3, 81

[15] S. Bugajski, K. E. Hellwig, and W. Stulpe. On fuzzy random variables and statistical maps. *Reports on Mathematical Physics*, 41(1):1–11, February 1998. 100, 101

[16] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23(15):880–884, Oct 1969. 31

[17] Institut International de Physique Solvay. Electrons et photons. In *Rapport et Discussions de Cinquieme Conseil de Physique tenu a Bruxelles du 24 au 29 Octobre 1927*, 1928. 33

[18] Detlef Durr, Sheldon Goldstein, and Nino Zangh. Bohmian mechanics and the meaning of the wave function. *arXiv:quant-ph/9512031v1*, 2007. 34

[19] A. Einstein, N. Rosen, and B. Podolsky. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777, 1935. 1, 29

[20] Joseph Emerson. *Quantum Chaos and Quantum-Classical Correspondence*. PhD thesis, Simon Fraser University, 2002. 2, 43

[21] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Wiley-Interscience, 1999. 97

[22] Christopher A. Fuchs. Quantum mechanics as quantum information (and only a little more). *arXiv:quant-ph/0205039v1*, 2002. 2

[23] Christopher A. Fuchs and Ruediger Schack. Quantum-bayesian coherence. *arXiv:0906.2187v1*, 2009. 2

[24] Ernesto F. Galvao. *Foundations of quantum theory and quantum information applications*. PhD thesis, University of Oxford, 2007. 67, 68, 104, 105

[25] Sheldon Goldstein. Bohmian mechanics. In *The Stanford Encyclopedia of Philosophy*. Stanford Universitiy, 2006. 36

[26] Gilad Gour and Robert W Spekkens. The resource theory of quantum reference frames: manipulations and monotones. *New Journal of Physics*, 10(3):033023 (63pp), 2008. 71

[27] Andrzej Grudka and PawełKurzyński. Is there contextuality for a single qubit? *Physical Review Letters*, 100(16):160401, 2008. 54

[28] Lucien Hardy. *Bohmian Mechanics and Quantum Theory*, chapter Contextuality in Bohmian Mechanics, pages 67–76. Springer, 1996. 36

[29] Lucien Hardy. Quantum ontological excess baggage. *Studies in History and Philosophy of Modern Physics*, 35:267–276, 2004. 17

[30] Lucien Hardy. Quantum theory from five reasonable axioms, 2007. 16

[31] Nicholas Harrigan and Terry Rudolph. Ontological models and the interpretation of contextuality. *arXiv:0709.4266v1*, 2007. 2, 17

[32] Nicholas Harrigan and Terry Rudolph. Representing probabilistic data via ontological models. *arXiv:0709.1149v2*, 2007. 17

[33] Nicholas Harrigan and Robert W. Spekkens. Einstein, incompleteness, and the epistemic view of quantum states. *arXiv:0706.2661v1*, 2007. 2, 43, 52

[34] K. E. Hellwig. Coexistent effects in quantum mechanics. *International Journal of Theoretical Physics*, 2:147, 1969. 17

[35] Peter R. Holland. *The Quantum Theory of Motion: An Account of the de Broglie-Bohm Casual Interpretation of Quantum Mechanics*. Cambridge University Press, 1993. 33, 34, 35, 36

[36] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2006. 3

[37] Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki. Limits for entanglement measures. *Phys. Rev. Lett.*, 84:2014, 2000. 75

[38] D. Howard. Einstein on locality and separability. *Stud. Hist. Philos. Sci*, 16:171, 1985. 32

[39] J. M. Jauch and C. Piron. *Helv. Phys. Acta.*, page 827, 1963. 37

[40] S. Kochen and E. Specker. The problem of hidden variables in quantum mechanics. *Indiana Univ. Math. J.*, 17:59–87, 1968. 1, 2, 13, 14, 23, 24, 25, 28, 44

[41] R. Blume Kohout, H. Khoon Ng, D. Poulin, and L. Viola. Characterizing the structure of preserved information in quantum processes. *Physical Review Letters*, 100, 2008. 82

[42] A. N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Publishing Company, 1956. 100

[43] Henry Krips. Measurement in quantum theory. In *The Stanford Encyclodepia of Philosophy*. Stanford Universitiy, 2007. 22, 25

[44] Gunther Ludwig. Versuch einer axiomatischen grundlegung der quantenmechanik und allgemeinerer physikalischer theorien. *Zeitschrift fr Physik A Hadrons and Nuclei*, 181:233, 1964. 17

[45] Lluís Masanes. All bipartite entangled states are useful for information processing. *Physical Review Letters*, 96(15):150501, 2006. 73

[46] Lluís Masanes. Useful entanglement can be extracted from all nonseparable states. *J. Math. Phys.*, 49,:022102, 2008. 73

[47] Keiji Matsumoto. Reverse estimation theory, complementality between rld and sld, and monotone distances. *arXiv:quant-ph/0511170v1*, 2007. 86

[48] Ryan Morris and Joseph Emerson. In preparation, 2009. 54

[49] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, October 2000. 2, 3, 8, 11, 40, 69, 77, 78, 82, 83, 85, 88

[50] Asher Peres. *Quantum Theory: Concepts and Methods*. Kluwer Academic Publishers, 2002. 29

[51] Martin B. Plenio and S. Virmani. An introduction to entanglement measures. *Quant. Inf. Comp.*, 7,:1, 2007. 71, 73

[52] David Poulin and Robin Blume-Kohout. Compatibility of quantum states. *Phys. Rev. A*, 67(1):010101, Jan 2003. 95

[53] Jochen Rau. Consistent reasoning about a continuum of hypotheses on the basis of finite evidence. *arXiv:0706.2274v1*, 2007. 16

[54] Jochen Rau. On quantum vs. classical probability. *arXiv:0710.2119v2*, 2007. 16

[55] G. Rigolin. Quantum teleportation of an arbitrary two qubit state and its relation to multipartite entanglement. *Phys. Rev. A*, 71,:032303, 2005. 73

[56] B Schumacher. Quantum coding. *Phys. Rev. A*, 51:446, 1995. 71

[57] N. J. A. Sloane and A. D. Wyner, editors. *Claude Elwood Shannon: Collected Papers*. IEEE Press, 1993. 70

[58] R. W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Physical Review A (Atomic, Molecular, and Optical Physics)*, 71(5):052108, 2005. 1, 2, 17, 32, 52, 54, 55, 61, 67

[59] Robert W. Spekkens. Negativity and contextuality are equivalent notions of nonclassicality, 2007. 43, 67, 68, 69, 106

[60] Robert W. Spekkens, D. H. Buzacott, A. J. Keehn, Ben Toner, and G. J. Pryde. Experimental demonstration of preparation contextuality and parity-oblivious multiplexing, 2008. 69, 107, 108

[61] J. von Neumann. *Methematische Grundlagen der Quantenmechanik*. Princeton University Press, 1955. 37

[62] Eric W. Weisstein. Ellipse. *http://mathworld.wolfram.com/Ellipse.html*. From MathWorld–A Wolfram Web Resource. 94

[63] Reinhard F. Werner. Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40(8):4277–4281, Oct 1989. 73

[64] Hans Westman. Non-locality, contextuality and transition sets. *arXiv:0711.2653v1*, 2007. 67

[65] Alexander Wilce. Quantum logic. In *Stanford Encyclopedia of Philosophy*. Stanford University, 2006. 16