An Examination of Quantum Foundations

by

Sonia Markes

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, 2011

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

Quantum foundations is a field of diverse goals and methods. In this thesis, I will present three different approaches to quantum foundations, each emphasizing a different goal or perspective. The causaloid framework has the goal is to use insight from quantum foundations to study quantum gravity. Ontic models are a tool used to study realist theories of quantum mechanics from an operational quantum information perspective. Nelson's mechanics is a derivation of the Schrodinger equation using the machinery of stochastic mechanics.

As each of these approaches has different set of goals, they are suited to different purposes and have different limitations. From the causaloid, I construct the concept of causally unbiased entropy and at the same time, find an emergent idea of causality in the form of a measure of causal connectedness, termed the Q factor. In the ontic models framework, I reproduce the generalization of the concept of contextuality. For Nelson's mechanics, I examine its relationship to Bohmian mechanics - a realist formulation of quantum mechanics.

I will then examine the relationship of these different approaches to one another. From this examination I will introduce the concept of physical contextuality in order to ask whether contextuality could be more than just a mathematical artifact. I also include a discussion of the property of deficiency in ontic models and its relation to contextuality given certain constraints.

Acknowledgements

I would like to thank Joseph Emerson for being my supervisor and for his feedback on my thesis. Lucien Hardy for being my mentor and teaching me so much. Matthew Scott for being on my committee and for being a great professor.

I would like to thank my partner Jonathan Hackett for being a support and a sounding board.

A special thank you for Helen Warren for guiding me through this whole process and for always having a smile.

This research was funded by an Ontario Graduate Scholarship.

Table of Contents

List of Figures

Chapter 1

Introduction

The study of the foundations of quantum mechanics has existed in one for or another for more than 75 years[7]. In that time the goals of the subject have varied significantly, with some programs even working counter to others[4, 5, 22, 8, 3, 33, 27, 29]. Though these goals are mostly reflective of the philosophies of the researchers that work on these subjects, it is not always the case that the framework used in research is chosen through this bias. It is for this reason that it is prudent to examine whether there is the potential for consolidation of approaches or at the least the transfer of knowledge between them. For this reason we will examine different approaches to quantum foundations including their motivations, their applications and their descriptions of reality. We will then examine whether the opportunity to build bridges between these subjects exists and the implications of looking for such connections.

From the attempts to create a comprehensible realist interpretation of quantum me-

chanics we have gained insight into what distinguishes quantum mechanics from the classical world. One such attempt - due to Nelson [22] - proposed that the departure from classical physics was unnecessary, instead demonstrating that the Schrodinger equation (the principle equation of quantum mechanics) can be derived from the theory of Brownian motion. Though Nelson's Mechanics was an early candidate in a line of attempts to give a realist interpretation of quantum mechanics, it still remains one of the canonical examples in the study of quantum foundations. In section 1.1 we present the derivation of this theory. Then in section 2.4.2 we will compare this theory with another realist interpretation of quantum mechanics due to Bohm[4, 5].

One of the advances of the modern study of the foundations of quantum mechanics is the unification of different approaches to quantum mechanics which share commonalities. Combining approaches in this way allow us to determine which properties are invariant under the assumptions on which the theories differ, and in doing so develop tools with which to study them. One example of this method is the development of the ontic models framework (see for example [12]). In section 1.2 we will present this framework. Then in section 2.2 we will use this framework to illustrate the concept of contextuality and its generalizations, and in section 2.3 we will present the concept of deficiency within this framework and discuss its relations with contextuality.

Though fundamentally quantum foundations is concerned with trying to understand quantum mechanics, it can also take a broader vision which include looking at how quantum mechanics can interface with other things we know about reality. One such interface that is cause for concern is that with general relativity. General relativity possesses a very different structure from quantum mechanics mostly owing to differences in their treatment

of the causal structure of space-time, and this discord is reflected in the difficulty of constructing a theory of quantum gravity[30]. To this end the causaloid [10, 11] framework was constructed in order to study what would happen if we attempted to generalize how we treat quantum mechanics to allow for a more general causal structure. We will present the causaloid framework in section 1.3. We will also present a means of constructing a concept of entropy in such a framework in section 2.1.

The rest of this thesis will take the form of an evaluation of whether the lessons from each of these different approaches to quantum foundations can be applied to the others. To that end we will discuss the compatibilities and incompatibilities of these frameworks and from these discussions we'll further examine the idea of whether the concept of contextuality has a physical meaning (in section 3.1.1).

1.1 Nelson's Mechanics

In [22], Nelson attempted to construct a realist model of quantum mechanics. The attempt assumes that the trajectories of particles can be described by a modification of classical mechanics. This modification is based upon the assumption that whatever real processes lead to quantum mechanics averages out to give a noise-like modification to the classical dynamics proportional in magnitude to Planck's constant. This theory, referred to as Nelson's Mechanics is cast in the language of stochastic processes. The central result of [22] is in constructing the wave function in this language and demonstrating that it is a solution to the Schrodinger equation. Though foundational, the derivation in [22] is not fully rigorous and has some gaps. In the following sections we will tighten the derivation

in [22], while filling in the details missing from the calculations.

1.1.1 Stochastic Mechanics

We will first build the machinery to study classical mechanics with stochastic processes. Since we will be dealing with processes that are not necessarily differentiable, we must define a different type of derivative. Central to this is what we'll define as the mean forward derivative

$$Dx(t) = \lim_{\Delta t \to 0^+} \langle \frac{x(t+\Delta t) - x(t)}{\Delta t} \rangle, \tag{1.1}$$

and the mean backward derivative

$$D_\star x(t) = \lim_{\Delta t \to 0^+} \langle \frac{x(t) - x(t-\Delta t)}{\Delta t} \rangle. \tag{1.2}$$

If the mean forward derivative and the mean backward derivative are equal, then the process is differentiable. We will use these to study Brownian motion with friction in a potential and kinematics of Markoff processes, which we will later use to produce quantum mechanics from Brownian motion.

Brownian motion with friction in a potential

From Newton's second law, we know that the acceleration $\vec{K}$ of a particle of mass m due to a potential V is given by

$$m\vec{K} = -\nabla V \tag{1.3}$$

A frictional force is dependent not on the position of a particle, but instead on its velocity. We use $m\beta$ for our coefficient of friction, with β what is commonly termed the drag coefficient (to reflect that the frictional force is due to a particle moving through a space rather than along a surface).

We then get Langevin equations of the form

$$\mathrm{d}x(t) = \mathrm{d}v(t)\mathrm{d}t \tag{1.4}$$

$$\mathrm{d}v(t) = -\beta v(t)\mathrm{d}t + \vec{K}(x(t))\mathrm{d}t + \mathrm{d}\vec{B}(t). \tag{1.5}$$

We have also introduced a Wiener process $\vec{B}(t)$ - a purely diffusive term - where $\mathrm{d}\vec{B}$ is a Gaussian with

$$\langle \mathrm{d}\vec{B} \rangle = 0 \tag{1.6}$$

$$\langle \mathrm{d}B^2 \rangle = \frac{6\beta k_B T}{m}\mathrm{d}t \tag{1.7}$$

Here k_B is the Boltzmann constant, and the constant T arises from the assumption that for a distribution of such systems, we would assume the distribution to arise from a Maxwell-

Boltzman distribution characterized by a temperature T. In requiring causality and by the nature of randomness, we also get

$$\langle(\mathrm{d}\vec{B}(t))x(s)\rangle = 0 \quad \forall \quad s \leq t \tag{1.8}$$

$$\langle \mathrm{d}\vec{B}(t)v(s)\rangle = 0 \quad \forall \quad s \leq t. \tag{1.9}$$

Clearly, these equations are asymmetric in time. We impose time symmetry by defining $\mathrm{d}\vec{B}_\star(t)$ such that

$$\langle \mathrm{d}\vec{B}_\star(t)x(s)\rangle = 0 \quad \forall \quad s \geq t \tag{1.10}$$

$$\langle \mathrm{d}\vec{B}_\star(t)v(s)\rangle = 0 \quad \forall \quad s \geq t. \tag{1.11}$$

We then get the Langevin equation

$$\mathrm{d}\vec{v}(t) = \beta\vec{v}(t) + \vec{K}(\vec{x}(t)) + \mathrm{d}\vec{B}_\star(t). \tag{1.12}$$

Applying the definitions of the mean forward and backward derivative we have that - as $\vec{x}(t)$ is differentiable -

$$D\vec{x}(t) = D_\star\vec{x}(t) = \vec{v}(t) \tag{1.13}$$

where we've used the definition of velocity in terms of position

$$\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} = \vec{v}(t). \tag{1.14}$$

From equation 1.7, we know that

$$\mathrm{d}\vec{B} \propto \sqrt{\mathrm{d}t}. \tag{1.15}$$

As $\mathrm{d}\vec{B}$ appears in the Langevin equation for $\vec{v}(t)$ linearly, we then have that $\vec{v}(t)$ is not differentiable.

Examining $\vec{B}$ we see that as $\vec{B}(t+\Delta t) - \vec{B}(t)$ is independent of the pair $\vec{x}(t), \vec{v}(t)$ for $\Delta t > 0$, and has mean 0. This implies that

$$D\vec{B}(t) = 0. \tag{1.16}$$

As D is a linear operation, and so using the Langevin equation we can extract

$$D\vec{v}(t) = \lim_{\Delta t \to 0} \langle \frac{\vec{v}(t+\Delta t) - \vec{v}}{\Delta t} \rangle \tag{1.17}$$

$$= \lim_{\Delta t \to 0} \langle \frac{\Delta \vec{v}(t)}{\Delta t} \rangle \tag{1.18}$$

$$= \lim_{\Delta t \to 0} \langle \frac{-\beta \vec{v} \Delta t + \vec{K} \Delta t + \Delta \vec{B}}{\Delta t} \rangle \tag{1.19}$$

$$= -\beta \vec{v} + \vec{K} + \lim_{\Delta t \to 0} \frac{1}{\Delta t} \langle \mathrm{d}\vec{B} \rangle = -\beta \vec{v}(t) + \vec{K}(\vec{x}(t)), \tag{1.20}$$

where we've used the fact that the mean of the Wiener process is 0. In the same way we can extract from the Langevin equation obtained from the Wiener process $\vec{B}_\star$

$$D_\star \vec{v}(t) = \beta \vec{v}(t) + \vec{K}(\vec{x}(t)). \tag{1.21}$$

We now can consider a free particle - where $\vec{K} = 0$. In this situation we have that

$$D\vec{v}(t) = -D_\star\vec{v}(t) = -\beta\vec{v}(t). \tag{1.22}$$

Using our knowledge of the solutions to the differential equation

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -\beta a \tag{1.23}$$

we know that the velocity tends towards 0 for either choice of time direction.

Returning again to a system with a potential, we can combine equations 1.22 and 1.23 together with the definitions of the mean forward and backward velocities. Doing so we get

$$\begin{aligned} \frac{1}{2}DD_\star\vec{x} + \frac{1}{2}D_\star D\vec{x} &= \frac{1}{2}\left(D\vec{v} + D_\star\vec{v}\right) && (1.24)\\ &= \frac{1}{2}\left(-\beta\vec{v} + \vec{K} + \beta\vec{v} + \vec{K}\right) && (1.25)\\ &= \vec{K} && (1.26) \end{aligned}$$

From this, we define the *mean second derivative* of $\vec{x}(t)$ to be

$$\vec{a}(t) = \frac{1}{2}\left(DD_\star\vec{x}(t) + D_\star D\vec{x}(t)\right). \tag{1.27}$$

We then get the generalization of Newton's second law in Ornstein-Uhlenbeck theory: The

mean acceleration is equal to the external force divided by the mass of the particle.

$$\vec{F} = m\vec{a} \tag{1.28}$$

Kinematics of Markoff Processes

For a sufficiently large time scale compared to the relaxation time (characterized by β^{-1}), we can describe the Brownian motion of a free particle in a fluid by a Wiener process $\vec{w}(t)$. The Wiener infinitesimals are mutually independent Gaussians with

$$\langle \mathrm{d}\vec{w} \rangle = 0 \tag{1.29}$$

and

$$\langle \mathrm{d}\vec{w}_i(t)\vec{w}_j(t) \rangle = \frac{2k_B T}{m\beta}\mathrm{d}t \tag{1.30}$$

As k_B, T and m do not occur independently we combine them with β into the diffusion coefficient ν.

$$\nu = \frac{k_b T}{m\beta} \tag{1.31}$$

We can then further introduce an external force to the scenario, or equivalently, a current within the fluid by introducing a time dependent vector field $\vec{b}\,(\vec{x}(t), t)$. This would give us an equation for the position vector of our particle of the form

$$\mathrm{d}\vec{x}(t) = \vec{b}\,(\vec{x}(t), t)\,\mathrm{d}t + \mathrm{d}\vec{w}(t) \tag{1.32}$$

Like in section 1.1.1, we have that the $\mathrm{d}\vec{w}(t)$ are independent of the previous positions ($\vec{x}(s)$ with $s \leq t$):

$$\langle \mathrm{d}\vec{w}(t)\vec{x}(s)\rangle = 0 \quad \forall \quad s \leq t \tag{1.33}$$

This gives us that our mean forward velocity is solely due to the current

$$D\vec{x}(t) = \vec{b}(\vec{x}(t), t). \tag{1.34}$$

Again, in the same form as section 1.1.1 we define a backward Wiener process $\mathrm{d}\vec{w}_\star$, with

$$\langle \mathrm{d}\vec{w}_\star\rangle = 0. \tag{1.35}$$

We also have that the backwards Wiener process is independent of the future positions:

$$\langle \mathrm{d}\vec{w}_\star(t)\vec{x}(s)\rangle = 0, \quad \forall \quad s \geq t. \tag{1.36}$$

This allows us to have the Langevin equation from the backward Wiener process:

$$\mathrm{d}\vec{x}(t) = \vec{b}_\star\left(\vec{x}(t), t\right)\mathrm{d}t + \mathrm{d}\vec{w}_\star(t). \tag{1.37}$$

This gives us the mean backward velocity:

$$D_\star\vec{x}(t) = \vec{b}_\star\left(\vec{x}(t), t\right). \tag{1.38}$$

We now move to considering a probability density $\rho(\vec{x}, t)$. This density then satisfies

the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = -\sum_{i=1}^{3} \frac{\partial}{\partial x_i} \left(A_i \rho \right) + \frac{1}{2} \sum_{i,j=1}^{3} \frac{\partial^2}{\partial x_i \partial x_j} \left(B_{ij} \rho \right). \tag{1.39}$$

Here the A_i are defined by

$$\vec{A} = \lim_{\Delta t \to 0^+} \frac{\langle \Delta \vec{x} \rangle}{\Delta t} = \vec{b}, \tag{1.40}$$

and the B_{ij} are defined by

$$B_{ij} = \lim_{\Delta t \to 0^+} \frac{\langle \Delta x_i \Delta x_j \rangle}{\Delta t} = 2\nu \delta_{ij}. \tag{1.41}$$

From this we arrive at the *forward Fokker-Planck equation*:

$$\frac{\partial \rho}{\partial t} = -\vec{\nabla} \cdot \left(\vec{b} \rho \right) + \nabla^2 \left(\nu \rho \right). \tag{1.42}$$

Proceeding instead with the backward derivative instead of the forward changes our equation slightly. The A_i are then defined in terms of the backward current by

$$\vec{A} = \lim_{\Delta t \to 0^-} \frac{\langle \Delta \vec{x} \rangle}{\Delta t} = \vec{b}_\star, \tag{1.43}$$

and the B_{ij} are defined by

$$B_{ij} = \lim_{\Delta t \to 0^-} \frac{\Delta x_i \Delta x_j}{\Delta t} = -\lim_{\Delta t \to 0^+} \frac{\Delta x_i \Delta x_j}{\Delta t} = -2\nu \delta_{ij}. \tag{1.44}$$

This then gives us what we'll call the *backward Fokker-Planck equation*:

$$\frac{\partial \rho}{\partial t} = -\vec{\nabla} \cdot \left(\vec{b}_\star \rho \right) - \nabla^2 \left(\nu \rho \right). \tag{1.45}$$

We can take an evenhanded approach between the forward and backward equations by defining the *current velocity*:

$$\vec{v} = \frac{1}{2} \left(\vec{b} + \vec{b}_\star \right). \tag{1.46}$$

With this we can combine the forward and backward Focker-Planck equations to get

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \left(\vec{v} \rho \right). \tag{1.47}$$

We now consider the Taylor expansion of a function f of the position $\vec{x}$ and the time t:

$$f\left(\vec{x_4} + \mathrm{d}\vec{x_4}\right) = f\left(\vec{x_4}\right) + \left[\left(\mathrm{d}\vec{x_4} \cdot \nabla_4 \right) f \right]_{\vec{x_4}} + \frac{1}{2} \left[\left(\mathrm{d}\vec{x_4} \cdot \nabla_4 \right)^2 f \right]_{\vec{x_4}} + \ldots \tag{1.48}$$

where we've used the subscript 4 to denote that we're taking the infinitesimal both with respect to the position and the time, but without the relativistic implications of the d'Alembertian. Defining the infinitesimal of the function f to be

$$\mathrm{d}f = f(\vec{x_4} + \mathrm{d}\vec{x_4}) - f(\vec{x_4}). \tag{1.49}$$

With this definition, and returning to separate space and time co-ordinates we reach

$$\mathrm{d}f = \frac{\partial f}{\partial t}\mathrm{d}t + \mathrm{d}\vec{x}\cdot\nabla f\frac{1}{2}\sum_{i,j=1}^{3}\mathrm{d}x_i\mathrm{d}x_j\frac{\partial^2}{\partial x_i\partial x_j}f \tag{1.50}$$

where we've dropped higher order terms (including second order terms in time).

Using equation 1.32 we can replace $\mathrm{d}\vec{x}_i$ with $\mathrm{d}\vec{w}_i$, which allows us to proceed to get:

$$Df = \lim_{\Delta t\to 0}\frac{\langle \mathrm{d}f\rangle}{\Delta t} \tag{1.51}$$

$$= \frac{\partial f}{\partial t} + \vec{b}\cdot\vec{\nabla}f + \frac{1}{2}\langle \mathrm{d}w^2\rangle\nabla^2 f. \tag{1.52}$$

Finally we can implement equation 1.30 to get

$$Df = \frac{\partial f}{\partial t} + \vec{b}\cdot\vec{\nabla}f + \nu\nabla^2 f. \tag{1.53}$$

We can reproduce this same process using the backward derivative:

$$D_\star f = \lim_{\Delta t\to 0^-}\frac{\langle \Delta f\rangle}{\Delta t} \tag{1.54}$$

$$= \frac{\partial f}{\partial t} + \vec{b}_\star\cdot\vec{\nabla}f - \nu\nabla^2 f. \tag{1.55}$$

We now observe the following: as ρ is a probability density its integral across all of space is 1 and so its distribution across space-time is invariant. From [21] we can then define adjoints with respect to $\rho\mathrm{d}^3x\mathrm{d}t$ as a measure, and find that the operators A and B

defined by

$$A = \frac{\partial}{\partial t} + \vec{b} \cdot \nabla + \nu \nabla^2 \tag{1.56}$$

$$B = -\frac{\partial}{\partial t} - \vec{b}_\star \cdot \nabla + \nu \nabla^2 \tag{1.57}$$

are adjoint to one another with respect to this measure. That is to say that

$$\int f(Ag)\rho \, \mathrm{d}^3x\mathrm{d}t = \int (Bf)g\rho \, \mathrm{d}^3x\mathrm{d}t. \tag{1.58}$$

We contrast this to the standard adjoint - which we'll denote by a † - with respect to the measure $\mathrm{d}^3x\mathrm{d}t$, defined by

$$\int (Af)g\mathrm{d}^3x\mathrm{d}t = \int f(A^\dagger g)\mathrm{d}^3x\mathrm{d}t. \tag{1.59}$$

From this we can form an equation relating $A^\dagger$ and B by conjugation by the probability distribution ρ:

$$\rho^{-1}\left(\frac{\partial}{\partial t} + \vec{b} \cdot \nabla + \nu \nabla^2\right)^\dagger \rho = -\frac{\partial}{\partial t} - \vec{b}_\star \cdot \nabla + \nu \nabla^2. \tag{1.60}$$

Multiplying from the right by ρ^{-1} and applying both sides to the trivial function, we can use the definition of the adjoint as being left acting to find

$$\frac{\partial \rho^{-1}}{\partial t} + \vec{b} \cdot \nabla \rho^{-1} + \nu \nabla^2 \rho^{-1} = -\frac{\partial \rho^{-1}}{\partial t} - \left(\vec{b}_\star \cdot \nabla\right)\rho^{-1} + \nu \nabla^2 \rho^{-1}. \tag{1.61}$$

With some manipulation we reach

$$2\frac{\partial \rho^{-1}}{\partial t} = -\frac{1}{\rho}\left(\vec{b} + \vec{b}_\star\right) \cdot \nabla \rho^{-1} \tag{1.62}$$

$$2\frac{-1}{\rho^2}\frac{\partial \rho}{\partial t} = \frac{1}{\rho^2}\left(\vec{b} + \vec{b}_\star\right) \cdot \nabla \rho \tag{1.63}$$

$$2\frac{\partial \rho}{\partial t} = -\left(\vec{b} + \vec{b}_\star\right) \cdot \nabla \rho \tag{1.64}$$

Applying the Fokker-Planck equation we can eliminate the time derivative and arrive at

$$2\left(-\nabla \cdot \left(\vec{b}\rho\right) + \nu \nabla^2 \rho\right) = -\left(\vec{b} + \vec{b}_\star\right) \cdot \nabla \rho. \tag{1.65}$$

Using the product rule we have that

$$\nabla \cdot \left(\rho \vec{b}\right) = (\nabla \rho) \cdot \vec{b} + \rho \nabla \cdot \vec{b}, \tag{1.66}$$

and likewise for $\vec{b}_\star$. We now recall that both of $\vec{b}$ and $\vec{b}_\star$ are currents of a fluid, and that they should thus be divergence free, allowing us to eliminate the second term in the right hand side of the product rule. This allows us to continue to simplify equation 1.65 through

$$-2\nabla \cdot \left(\vec{b}\rho\right) + 2\nu \nabla^2 \rho = -\nabla \cdot \left(\vec{b} + \vec{b}_\star\right) \rho. \tag{1.67}$$

Using left cancelation of the gradient we reach

$$-2\rho\vec{b} + 2\nu \nabla \rho = -\rho\vec{b} - \rho\vec{b}_\star, \tag{1.68}$$

which we rearrange to give us

$$\vec{b}_\star = \vec{b} - 2\nu\frac{\nabla\rho}{\rho}. \tag{1.69}$$

Additionally, if we call the difference between the forward and backward currents the *osmotic velocity*, $\vec{u}$:

$$\vec{u} = \frac{1}{2}\left(\vec{b} - \vec{b}_\star\right), \tag{1.70}$$

then equation 1.69 is equivalent to stating that

$$\vec{u} = \nu\frac{\nabla\rho}{\rho}. \tag{1.71}$$

We can compare this to the difference between the forward and backward Fokker-Planck equations, given by

$$0 = \frac{\partial\rho}{\partial t} + \nabla\cdot\left(\vec{b}_\star\rho\right) + \nu\nabla^2\rho - \frac{\partial\rho}{\partial t} - \nabla\cdot\left(\vec{b}\rho\right) + \nu\nabla^2\rho \tag{1.72}$$

$$= \nabla\cdot\left(\left(\vec{b}_\star - \vec{b}\right)\rho\right) + 2\nu\nabla^2\rho \tag{1.73}$$

$$= \nabla\cdot\left(\vec{u}\rho - \nu\nabla\rho\right) \tag{1.74}$$

Which under left cancelation would give us equation 1.71 again. Equation 1.71 can also be written

$$\vec{u} = \nu\nabla\left(\ln\rho\right). \tag{1.75}$$

We now proceed to calculate the time derivative of $\vec{u}$:

$$\frac{\partial \vec{u}}{\partial t} = \nu \nabla \frac{\partial \ln \rho}{\partial t} \tag{1.76}$$

$$= \nu \nabla \left(\frac{1}{\rho} \frac{\partial \rho}{\partial t} \right). \tag{1.77}$$

Using the continuity equation -

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\vec{v}\rho) \tag{1.78}$$

- we can study the time derivative:

$$\frac{\partial \vec{u}}{\partial t} = -\nu \nabla \left(\frac{1}{\rho} \nabla \cdot (\vec{v}\rho) \right) \tag{1.79}$$

$$= -\nu \left(\frac{-1}{\rho^2} \left(\nabla \rho \left(\nabla \cdot (\rho \vec{v}) \right) \right) - \frac{1}{\rho^2} \nabla \rho \left(\rho \nabla \cdot \vec{v} \right) \right. \tag{1.80}$$

$$\left. + \frac{1}{\rho} \nabla \left(\nabla \rho \cdot \vec{v} \right) + \frac{1}{\rho} \left(\nabla \rho \right) \left(\nabla \cdot \vec{v} \right) + \frac{1}{\rho} \rho \left(\nabla \left(\nabla \cdot \vec{v} \right) \right) \right)$$

$$= -\nu \left(\frac{-1}{\rho^2} \nabla \rho \left(\nabla \rho \cdot \vec{v} \right) + \frac{1}{\rho} \nabla \left(\nabla \rho \cdot \vec{v} \right) + \nabla \left(\nabla \cdot \vec{v} \right) \right) \tag{1.81}$$

$$= -\nu \nabla \left(\nabla \cdot \vec{v} \right) - \nabla \left(\vec{v} \cdot \vec{u} \right). \tag{1.82}$$

where in the last step we've used that

$$\nabla \left(\vec{v} \cdot \vec{u} \right) = \nabla \left(\vec{v} \cdot \frac{\nabla \rho}{\rho} \right) \tag{1.83}$$

$$= \left(\nabla \frac{1}{\rho} \right) \left(\vec{v} \cdot \nabla \rho \right) + \frac{1}{\rho} \left(\nabla \left(\vec{v} \cdot \nabla \rho \right) \right) \tag{1.84}$$

$$= \frac{-1}{\rho^2} \nabla \rho \left(\vec{v} \cdot \nabla \rho \right) + \frac{1}{\rho} \nabla \left(\nabla \rho \cdot \vec{v} \right). \tag{1.85}$$

With all of these results we can finally approach our goal in this section; to relate the $\vec{b}$ and $\vec{b}_\star$ of a Markoff process to an acceleration. To do this we begin with

$$\vec{a}(t) = \frac{1}{2}\left(D\vec{b}_\star + D_\star\vec{b}\right), \tag{1.86}$$

where

$$D\vec{b}_\star = \left(\frac{\partial}{\partial t} + \left(\vec{b}\cdot\nabla\right) + \nu\nabla^2\right)\vec{b}_\star \tag{1.87}$$

$$D_\star\vec{b} = \left(\frac{\partial}{\partial t} + \left(\vec{b}_\star\cdot\nabla\right) - \nu\nabla^2\right)\vec{b}. \tag{1.88}$$

We then get that

$$\vec{a}(t) = \frac{\partial}{\partial t}\left[\frac{1}{2}\left(\vec{b}+\vec{b}_\star\right)\right] + \frac{1}{2}\left(\vec{b}\cdot\nabla\right)\vec{b}_\star + \frac{1}{2}\left(\vec{b}_\star\dot{\nabla}\right)\vec{b} - \nu\nabla^2\left[\frac{1}{2}\left(\vec{b}-\vec{b}_\star\right)\right] \tag{1.89}$$

$$= \frac{\partial\vec{v}}{\partial t} + \frac{1}{2}\left[\vec{u}\dot{\nabla} + \vec{v}\cdot\nabla\right](\vec{v}-\vec{u}) + \frac{1}{2}\left[\vec{v}\cdot\nabla - \vec{u}\cdot\nabla\right](\vec{u}+\vec{v}) - \nu\nabla^2\vec{u} \tag{1.90}$$

$$= \frac{\partial\vec{v}}{\partial t} - (\vec{u}\cdot\nabla)\,\vec{u} + (\vec{v}\cdot\nabla)\,\vec{v} - \nu\nabla^2\vec{u}. \tag{1.91}$$

Where we remind the reader that

$$\vec{b} = \vec{u} + \vec{v} \tag{1.92}$$

$$\vec{b}_\star = \vec{v} - \vec{u}. \tag{1.93}$$

We can then rearrange to get

$$\frac{\partial \vec{v}}{\partial t} = \vec{a}(t) + (\vec{u} \cdot \nabla)\, \vec{u} - (\vec{v} \cdot \nabla)\, \vec{v} + \nu \nabla^2 \vec{u}, \tag{1.94}$$

which is our final result of this section.

1.1.2 Using Brownian Motion to Model Quantum Mechanics

To model quantum mechanics through Brownian motion we will make a set of assumptions based upon what we know from the study of quantum mechanics. First we choose our diffusion co-efficient in such a manner that we achieve classical mechanics in both the high mass limit and the limit as we take another constant (we'll use $\hbar$ for reasons that should be clear, and will be verified later) to zero. For this reason we set

$$\nu = \frac{\hbar}{2m}. \tag{1.95}$$

We assume that the particle moves in a frictionless fluid (either vacuum or a frictionless ether), so as not to contradict our inability to detect a preferred reference frame. This means that the Brownian motion will not be smooth and that velocities will not exist. We describe the motion with a Markoff process in co-ordinate space, subject to dynamics arising from from Newtonian dynamics

$$\vec{F} = m\vec{a}. \tag{1.96}$$

Specifying $\vec{b}$ and $\vec{b}_\star$ - or equivalently $\vec{u}$ and $\vec{v}$ - we fix the Markoff process

$$\frac{\partial \vec{u}}{\partial t} = -\frac{\hbar}{2m}\nabla\left(\nabla\cdot\vec{v}\right) - \nabla\left(\vec{v}\cdot\vec{u}\right) \tag{1.97}$$

$$\frac{\partial \vec{v}}{\partial t} = \frac{\vec{F}}{m} - \left(\vec{v}\cdot\nabla\right)\vec{v} + \left(\vec{u}\cdot\nabla\right)\vec{u} + \frac{\hbar}{2m}\nabla^2\vec{u}. \tag{1.98}$$

From this we can fully determine $\vec{u}$ and $\vec{v}$ given a set of initial distributions $\vec{u}(\vec{x}(t), t_0)$ and $\vec{v}(\vec{x}(t), t_0)$.

Real Time-Independent Schrodinger Equation

We consider a force arising from a potential

$$\vec{F} = -\nabla V. \tag{1.99}$$

Suppose first that $\vec{v} = 0$, then by equation 1.97 we have that

$$\frac{\partial \vec{u}}{\partial t} = 0 \tag{1.100}$$

and from equation 1.98 we get that

$$\frac{\nabla V}{m} = \left(\vec{u}\cdot\nabla\right)\vec{u} + \frac{\hbar}{2m}\nabla^2\vec{u}. \tag{1.101}$$

As $\vec{u}$ is a gradient field - and the curl of a gradient is zero - we get

$$\left(\vec{u}\cdot\nabla\right)\vec{u} = \frac{1}{2}\nabla\left(\vec{u}\cdot\vec{u}\right). \tag{1.102}$$

Likewise, we also get

$$\nabla^2 \vec{u} = \nabla \left(\nabla \cdot \vec{u} \right). \tag{1.103}$$

This allows us to simplify equation 1.101 to

$$\frac{\nabla V}{m} = \nabla \left(\frac{\vec{u}^2}{2} + \frac{\hbar}{2m} \left(\nabla \cdot \vec{u} \right) \right) \tag{1.104}$$

$$\frac{V}{m} - \frac{E}{m} = \frac{1}{2}\vec{u}^2 + \frac{\hbar}{2m} \left(\nabla \cdot \vec{u} \right) \tag{1.105}$$

where E is a constant with dimensions of energy. We can multiply across this by $m\rho$ and integrate across space to find

$$\int \frac{1}{2} m \vec{u}^2 \rho \mathrm{d}^3 x + \frac{\hbar}{2} \int \left(\nabla \cdot \vec{u} \right) \rho \mathrm{d}^3 x = \int V \rho \mathrm{d}^3 x - E. \tag{1.106}$$

We now perform a quick manipulation:

$$\left(\nabla \cdot \vec{u} \right) \rho = \left(\nabla \cdot \frac{\hbar}{2m} \frac{\nabla \rho}{\rho} \right) \rho \tag{1.107}$$

$$= \frac{\hbar}{2m} \left(\frac{-1}{\rho^2} \left(\nabla \rho \right)^2 + \frac{1}{\rho} \nabla^2 \rho \right) \rho \tag{1.108}$$

$$= \frac{\hbar}{2m} \left(\nabla^2 \rho - \frac{\left(\nabla \rho \right)^2}{\rho} \right). \tag{1.109}$$

Integrating both sides we can drop the first term by using Gauss's divergence theorem and the conservation of probability, giving us

$$\int \left(\nabla \cdot \vec{u} \right) \rho \mathrm{d}^3 x = - \int \vec{u} \cdot \nabla \rho \mathrm{d}^3 x. \tag{1.110}$$

Substituting back into equation 1.106 we get

$$\int \frac{1}{2} m\vec{u}^2 \rho \mathrm{d}^3x - \frac{\hbar}{2} \int (\vec{u} \cdot \nabla \rho)\, \mathrm{d}^3x = \int V \rho \mathrm{d}^3x - E. \tag{1.111}$$

Returning again to the same term we have

$$\frac{\hbar}{2} \vec{u} \cdot \nabla \rho = \frac{\hbar}{2} \vec{u} \cdot \left(\frac{2m\rho}{\hbar} \vec{u} \right) \tag{1.112}$$

$$= m\vec{u}^2 \rho, \tag{1.113}$$

which allows us to simplify equation 1.111 to

$$E = \int \frac{1}{2} m\vec{u}^2 \rho \mathrm{d}^3x + \int V \rho \mathrm{d}^3x. \tag{1.114}$$

As ρ is a probability density, E then has an alternate interpretation as the expectation value of $\frac{1}{2} m\vec{u}^2 + V$.

We now recall that

$$\vec{u} = \frac{\hbar}{m} \left(\frac{1}{2} \nabla \ln \rho \right) \tag{1.115}$$

$$\Rightarrow \frac{m\vec{u}}{\hbar} = \nabla \left(\frac{1}{2} \ln \rho \right) = R. \tag{1.116}$$

From this we say that R is the real potential associated with $\frac{m\vec{u}}{\hbar}$. We can make a change

of variables from ρ to ψ seemingly inspired by this fact and Born's rule to get

$$\psi^2 = \rho \tag{1.117}$$

$$\psi = e^R. \tag{1.118}$$

We then have that

$$\frac{m\vec{u}}{\hbar} = \nabla\left(\ln\psi\right) = \frac{\nabla\psi}{\psi}, \tag{1.119}$$

which we substitute into equation 1.114 -

$$\frac{\hbar^2}{2m^2}\frac{\left(\nabla\psi\right)^2}{\psi^2} + \frac{\hbar^2}{2m^2}\nabla\cdot\left(\frac{\nabla\psi}{\psi}\right) = \frac{1}{m}\left(V - E\right) \tag{1.120}$$

$$\frac{\hbar^2}{2m}\left[\frac{\left(\nabla\psi\right)^2}{\psi^2} + \left(\frac{-1}{\psi^2}\right)\left(\nabla\psi\right)^2 + \frac{1}{\psi}\nabla^2\psi\right] = V - E \tag{1.121}$$

$$\frac{-\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi \tag{1.122}$$

- to reach the time independent Schrodinger equation.

The Time-Dependent Schrodinger Equation

Now - in place of the assumptions of the previous section - we attempt to construct a more general solution to equations 1.97 and 1.98. To do this, we make the assumption that $\nabla\cdot\vec{v} = 0$ and additionally that v is a gradient of some S:

$$\nabla S = \frac{m}{\hbar}\vec{v}. \tag{1.123}$$

We now propose - as an ansatz - that

$$\psi = e^{R+iS} \tag{1.124}$$

is a solution to the time independent Schrodinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{-\hbar^2}{2m}\nabla^2\psi + V\psi. \tag{1.125}$$

Substituting in to the left hand side of the equation we get

$$i\hbar\left[\psi\frac{\partial}{\partial t}\left(R+iS\right)\right] \tag{1.126}$$

$$-\hbar\psi\frac{\partial S}{\partial t} + i\hbar\psi\frac{\partial R}{\partial t}. \tag{1.127}$$

Substituting in to the left hand side of the equation we get

$$\frac{-\hbar^2}{2m}\nabla\cdot\left(\psi\nabla\left(R+iS\right)\right) + V\psi \tag{1.128}$$

$$= \frac{-\hbar^2}{2m}\left[\psi\left[\nabla\left(R+iS\right)\right]^2 + \psi\nabla^2\left(R+iS\right)\right] + V\psi \tag{1.129}$$

$$= \frac{-\hbar^2}{2m}\psi\left[\left(\nabla R\right)^2 - \left(\nabla S\right)^2 + \nabla^2 R + i\left(2\nabla R\cdot\nabla S + \nabla^2 S\right)\right] + V\psi. \tag{1.130}$$

Dividing through both sides by ψ and taking the gradient of both sides we get that our right hand side is

$$-\hbar\frac{\partial}{\partial t}\left(\frac{m}{\hbar}\vec{v}\right)+i\hbar\frac{\partial}{\partial t}\left(\frac{m}{\hbar}\vec{u}\right) \tag{1.131}$$

$$=-\frac{\partial}{\partial t}\left(m\vec{v}\right)+i\frac{\partial}{\partial t}\left(m\vec{u}\right) \tag{1.132}$$

and the left hand side is

$$\frac{-\hbar^2}{2m}\nabla\left(\frac{m^2}{\hbar^2}\vec{u}^2-\frac{m^2}{\hbar^2}\vec{v}^2+\nabla\cdot\left(\frac{m}{\hbar}\vec{u}\right)+i\left(2\frac{m^2}{\hbar^2}\vec{u}\cdot\vec{v}+\nabla\cdot\frac{m}{\hbar}\vec{v}\right)\right)-\vec{F} \tag{1.133}$$

$$=-m\left(\vec{u}\cdot\nabla\right)\vec{u}+m\left(\vec{v}\cdot\nabla\right)\vec{v}-\frac{\hbar^2}{2}\nabla^2\vec{u}+i\left[-m\nabla\left(\vec{u}\cdot\vec{v}\right)-\frac{\hbar^2}{2}\nabla^2\vec{v}\right]-\vec{F}. \tag{1.134}$$

Comparing the real parts of the two sides gives us

$$\frac{\partial\vec{v}}{\partial t}=\frac{F}{m}+\left(\vec{u}\cdot\nabla\right)\vec{u}-\left(\vec{v}\cdot\nabla\right)\vec{v}+\frac{\hbar^2}{2m}\nabla^2\vec{u} \tag{1.135}$$

while comparing the imaginary parts gives us

$$\frac{\partial\vec{u}}{\partial t}=-\nabla\left(\vec{u}\cdot\vec{v}\right)-\frac{\hbar}{2m}\nabla^2\vec{v}. \tag{1.136}$$

These two equations are the same as equations 1.97 and 1.98, which arose from the study of a particle subject to a potential and Brownian motion. This completes the derivation: we have demonstrated the existence of a method for finding the solutions to the Schrodinger equation through solutions to a stochastic process. It is then possible that such a process could be the underlying physics behind quantum mechanics, or at the least an effective

description of some underlying theory.

1.2 Ontic Models

The study of realist theories of quantum mechanics has progressed significantly from its origins. Rather than considering specific models of underlying physical structures and laws, research has instead focused on constructing general frameworks which include all such models so that the common properties of these theories can be studied. The ontic models framework is such an attempt, assuming nothing about the underlying ontic state space other than its existence. In what follows we shall present this framework.

We will use the notation from [12]. We begin by introducing an ontic state space Λ, we will then introduce epistemic states μ which are probability distributions over Λ - which reflects our inability to know the true ontic state. We will also allow our epistemic state to depend on the method which we use to prepare our state, S_P. We thus write the epistemic state as $\mu(\lambda|\ S_P)$, and we impose the requirement that there must be at least one ontic state which describes the system (along with normalization of probability) by requiring that

$$\int_\Lambda \mathrm{d}\lambda \mu(\lambda|\ S_P) = 1. \tag{1.137}$$

We then introduce the measurement which we allow to depend on the procedure with which the measurement is taken - the exact form of this dependence within a model will reflect whether or not the model is contextual in the traditional sense. Additionally, we allow for the possibility that the ontic states still give probabilistic predictions for measurements.

For a given measurement procedure S_M, with outcomes indexed by a parameter k, we give a probability distribution for the measurement by $\xi\left(k|\ \lambda, S_M\right)$, where ξ is the probability that for a given ontic state λ we will get the kth measurement outcome for our procedure. We then require that this distribution is normalized by saying

$$\sum_k \xi\left(k|\ \lambda, S_M\right) = 1. \tag{1.138}$$

We can then extract probabilities of outcomes for experiments by combining these two quantities.

$$\int \mu\left(\lambda|\ S_P\right) \xi\left(k|\ \lambda, S_M\right) \mathrm{d}\lambda = \Pr\left(k|\ S_P, S_M\right) \tag{1.139}$$

This gives us probabilities of measurement outcomes only in terms of preparations and measurement procedures, given the assumption of an underlying ontic model.

In [12] this model was explored further and given context in terms of an operational understanding, its relation to quantum mechanics and how to have the ontic framework represent different concepts. Here we will present their perspectives on how to understand this framework.

The operational understanding of the ontic models framework comes with taking it to describe a system S. The preparation method S_P then corresponds to having S interact with a preparation device P (with macroscopic settings we still label with S_P). After this the system S then interacts with a measurement device M which records an outcome based on the state S and the measurement settings S_M. This gives us a picture of how to relate the ontic models framework to experiments and the physical world.

We can use these operational definitions to relate this framework to standard quantum mechanics. To do this we first observe that the result of a system interacting with the preparation device in quantum mechanics is a density matrix, which we'll label ρ_{S_P}. We then take the measurement process as giving a POVM effect E_k with k indexing the outcome, corresponding to S_M. (Note: A POVM (positive operator valued measurement) is a set of positive semidefinite, self-adjoint operators on a Hilbert space such that the sum over the elements of the set is the identity.) To have consistency between quantum mechanics and the ontic models framework we impose matching with the Born rule:

$$\Pr\left(k \mid S_P, S_M\right) = \mathrm{tr}\left(E_k \rho_{S_P}\right). \tag{1.140}$$

The form of the ontic state space Λ reflects the individual ontic theory that we are describing with the framework. An example of this is if we consider the possibility of the wave function being the complete description of reality then Λ would be the projective Hilbert space of the wave functions. In this case the epistemic state takes a simple form of

$$\mu\left(\lambda \mid \psi\right) = \delta\left(\lambda - \lambda_\psi\right), \tag{1.141}$$

where λ_ψ is just the assignment of ψ to the space Λ. If ψ were instead an incomplete description of reality then we could decompose Λ into the projective Hilbert space together with supplementary variables which complete the description. If the wave function only represents the state of our knowledge - and not anything real - then the Λ cannot even be decomposed into the projective Hilbert space of the wave function together with other

variables.

We can also discuss properties of a model that are encoded outside of the structure of Λ. One such property is outcome determinism. In section 1.2 we allowed for probabilistic predictions for measurements. If we instead require that each λ gives a completely determined outcome for a given preparation and measurement we have an outcome deterministic theory. In the language of the ontic models framework we do this by requiring that for any k,λ and S_M that

$$\xi\left(k \mid \lambda, S_M\right) \in \{0, 1\}. \tag{1.142}$$

The last point to be made from the observations in [12] is that though ontic models typically treat the preparation system P and measurement system and M as external to the theory it is possible to instead treat them as part of the system described by the ontological space. This description is equivalent to treating them externally provided we satisfy three assumptions:

- Separability: The global ontic space Γ is the Cartesian product of the ontic spaces of the three systems S,P and M:

$$\Gamma = \Lambda_P \times \Lambda_S \times \Lambda_M, \tag{1.143}$$

  where Λ_S is the ontic space that we'd previously called Λ.

- Statistical independence: The effect of the preparation device's ontic state is mediated through S. That is to say that M and P must be statistically independent of

one another.

- Measurement: The measurement outcome must depend on the measurement settings rather than the particular ontic state of the measurement device.

Provided these requirements are satisfied, an ontological description of the preparation and measurement systems is equivalent to the standard picture.

1.3 The Causaloid Framework

In Newtonian physics, physical processes are understood with respect to a fixed spatial coordinate system and a time parameter, which is absolute and ever increasing. Predictions are entirely deterministic. Quantum theory and general relativity depart from this classical picture in opposing manners. Quantum theory gives probabilistic predictions as to the outcomes of measurements, but retains fixed space and time coordinates. On the other hand, general relativity is deterministic, but shows that space and time form a dynamical structure. Reconciling these fundamental philosophical differences is one of the many challenges one is faced with in trying to construct a theory of quantum gravity. There have been many different approaches to this problem with many different results [28, 13, 31, 35, 33, 1]. One way of moving forward is to dismiss classical assumptions and create a probabilistic theory that has a dynamic causal structure. However, what results is indefinite casual structure. This is more radical than either probabilistic predictions or dynamical space-time structure. In general relativity, a separation between space-time locations is either space-like or time-like. An indefinite causal structure would allow for a separation

between space-time locations to be something like a quantum superposition of a space-like and a time-like separation. While we may be uncertain of the causal structure of the path between measurements, we know where in space-time we make measurements, what measurements we have made, and what outcomes we get. With this data, we can examine probabilistic correlations for information. The causaloid framework [9, 10, 11] which we outline below provides us with the necessary tools to examine an indefinite causal structure.

1.3.1 The Picture

Every experiment results in a set of data from making measurements on a system. Each piece of data could be thought of as a card with three indications on it; where the measurement is made in space-time, what is measured, and what the result of the measurement is. We will represent each card (or piece of data) as (x, f_x, y_x) where x denotes the space-time location, f_x represents the apparatus configuration, and y_x denotes the outcome of the measurement. The set of all possible cards (i.e. all possible measurements with all possible outcomes with every space-time configuration) is denoted V. We can imagine running an experiment an infinite number of times so as to be able to obtain relative frequencies. (Note that the concept of relative frequencies is utilized here only has a illustrative tool and is not meant to specify an ontology.) In order for the cards to tell us the relative frequencies, we must systematically sort them.

Each distinct x is defined as an *elementary region* of space-time. A *composite region*, denoted $\mathcal{O}_1$, is a set of elementary regions. (Note: These definitions of "elementary region" and "composite region" differ from those in [9, 10, 11].) Therefore, these cards can be sorted

according to their associated space-time region. The set of all possible cards with the same space-time location x on them is the *elementary information set*. We denote this set as R_x. The *composite information set*, R_1, is the union of all elementary information regions for which the associated elementary regions are contained within the composite region $\mathcal{O}_1$. More concisely,

$$R_1 \equiv \bigcup_{x \in \mathcal{O}_1} R_x \tag{1.144}$$

Note that a composite information set has no more or less structure than an elementary region of space-time. Therefore, without adding structure or losing generality, we can treat the sets R_x as elementary regions. From this point forward, the term *region* will be used interchangeably to refer to objects of type x or $\mathcal{O}_1$ and type R_x or R_1.

If we chose a particular measurement to perform in the region x (indexed by α_x), the set of cards consistent with $f_x^{\alpha_x}$ is called the *procedure information set* and is denoted $F_x^{\alpha_x}$.

$$F_x^{\alpha_x} \equiv \bigcup_{all\, y_x} \{(x, f_x^{\alpha_x}, y_x)\} \tag{1.145}$$

For composite regions, we can then define the following:

$$F_1^{\alpha_1} \equiv \bigcup_{x \in \mathcal{O}_1} F_x^{\alpha_x} \tag{1.146}$$

For the measurement $f_x^{\alpha_x}$, the set of cards consistent with the outcome $y_x^{\alpha_x}$ is called the *outcome information set* and is denoted $Y_x^{\alpha_x}$

$$Y_x^{\alpha_x} \equiv \{(x, f_x^{\alpha_x}, y_x^{\alpha_x})\} \tag{1.147}$$

For composite regions, we have

$$Y_1 \equiv \bigcup_{x \in \mathcal{O}_1} Y_x^{\alpha_x} \tag{1.148}$$

The set of all measurement and outcome pairs (i.e. all $\alpha = (\mathrm{Y}, \mathrm{F})$) is denoted Υ.

Notice that the set of all cards V can be viewed as all the cards from all (elementary) regions.

$$V = \bigcup_{all\, x} R_x \tag{1.149}$$

So V is the largest of all regions that can be considered.

These definitions provide a firm foundation on which the causaloid framework rests both mathematically and conceptually.

1.3.2 First level physical compression

The most basic quantity that we would want to be able to calculate is the probability that a certain (set of) outcome(s) is observed given that a certain (set of) measurement(s) has been performed at a (set of) location(s) in space and time. Suppose that the set of locations we are interested in is $\mathcal{O}_1$. Then the information set of interest is R_1. The set comprised of all the cards not in R_1 is $V - R_1$. We call (Y_{V-R_1}, F_{V-R_1}) the *generalized preparation* because it is the information that surrounds R_1 not only from the immediate past, but from the future and the rest of space-time as well. By the choices we make in setting up the experiment, we can put conditions on the generalized preparation such that

$\text{Prob}(Y_V|F_V)$ is well-defined. (See Ref.[9] for details.) Then we can write

$$\text{Prob}(Y_V|F_V) = \text{Prob}(Y_1, Y_{V-R_1}|F_1, F_{V-R_1}) \tag{1.150}$$

For a specific pair $\alpha_1 \Leftrightarrow (Y_1^{\alpha_1}, F_1^{\alpha_1})$, we can write this probability as

$$\text{Prob}(Y_1^{\alpha_1}, Y_{V-R_1}|F_1^{\alpha_1}, F_{V-R_1}) \tag{1.151}$$

We use the short-hand p_{α_1} to denote the probability defined in Eq.(1.151). One way to specify the state of a system is to list all the possible p_{α_1} for elements of R_1.

$$\begin{pmatrix} \vdots \\ p_{\alpha_1} \\ \vdots \end{pmatrix} \quad \alpha_1 \in \Upsilon_1 \tag{1.152}$$

However, this over-specifies the state. We do not usually need to know the probability of every outcome of every measurement in order to determine what the complete state of the system is. Physical theories tell us what relationships exist between variables and what constraints those relationships place on the variables of the system. These relationships and constraints can be used to determine a reduced set of probabilities from which all other probabilities can be represented. The reduced set of probabilities is defined such that any probability can be written as a linear combination of the probabilities in the reduced set. Let us denote the reduced or fiducial set in R_1 as $\Omega_1 \subseteq \Upsilon_1$. This process of going from the set of all the probabilities to the smallest essential set we call *first level*

physical compression. This can be expressed as

$$\mathbf{p} = \begin{pmatrix} \vdots \\ p_{l_1} \\ \vdots \end{pmatrix} \quad l_1 \in \Omega_1 \subseteq \Upsilon_1 \tag{1.153}$$

such that

$$p_{\alpha_1} = \mathbf{r}_{\alpha_1} \cdot \mathbf{p} \tag{1.154}$$

where $\mathbf{r}_{\alpha_1}$ encodes the physical compression and therefore, is determined by the details of the physical theory. Of course, the compression is not unique. We can define a *decompression matrix*, $\Lambda_{\alpha_1}^{l_1}$ such that

$$\Lambda_{\alpha_1}^{l_1} \equiv \mathbf{r}_{\alpha_1}\big|_{l_1} \tag{1.155}$$

where $\mathbf{r}_{\alpha_1}\big|_{l_1}$ means the l_1 component of $\mathbf{r}_{\alpha_1}$. While the decompression matrix may seem superfluous at the first level, it becomes a useful notation for higher level compression.

1.3.3 Second level physical compression and the causaloid product

Let us consider two distinct regions $R_1, R_2 \subset V$. In a similar fashion to the single region case,

$$p_{\alpha_1\alpha_2} = \text{Prob}(Y_1^{\alpha_1}, Y_2^{\alpha_2}, Y_{V-R_1-R_2} | F_1^{\alpha_1}, F_2^{\alpha_2}, F_{V-R_1-R_2}) \tag{1.156}$$

We specify the state of the system by listing all $p_{\alpha_1\alpha_2}$.

$$\begin{pmatrix} \vdots \\ p_{\alpha_1\alpha_2} \\ \vdots \end{pmatrix} \quad \alpha_1\alpha_2 \in \Upsilon_1 \times \Upsilon_2 \tag{1.157}$$

where $\times$ is the cartesian product. It can be shown that

$$p_{\alpha_1\alpha_2} = \sum_{l_1 l_2 \in \Omega_1 \times \Omega_2} \Lambda_{\alpha_1}^{l_1} \Lambda_{\alpha_2}^{l_2} p_{l_1 l_2} \tag{1.158}$$

which implies that the following list of probabilities is sufficient.

$$\begin{pmatrix} \vdots \\ p_{l_1 l_2} \\ \vdots \end{pmatrix} \quad l_1 l_2 \in \Omega_1 \times \Omega_2 \subseteq \Upsilon_1 \times \Upsilon_2 \tag{1.159}$$

This is effectively first level compression on each index. But if a physical theory has some connection between the two regions, $\Omega_1 \times \Omega_2$ may no longer be the smallest set that is sufficient to represent all possible states. Then *second level physical compression* is possible. It is defined to be

$$\mathbf{p} = \begin{pmatrix} \vdots \\ p_{k_1 k_2} \\ \vdots \end{pmatrix} \quad k_1 k_2 \in \Omega_{12} \subseteq \Omega_1 \times \Omega_2 \tag{1.160}$$

such that

$$p_{\alpha_1\alpha_2} = \mathbf{r}_{\alpha_1\alpha_2} \cdot \mathbf{p} = \sum_{k_1k_2 \in \Omega_{12}} \mathbf{r}_{\alpha_1\alpha_2}\big|_{k_1k_2} p_{k_1k_2} \tag{1.161}$$

When $\Omega_{12} = \Omega_1 \times \Omega_2$, second level compression is trivial. But it is proven in [10] that it is possible that $\Omega_{12} \subset \Omega_1 \times \Omega_2$.

Now we can define a second level decompression matrix. By comparing Eq.(1.158) and Eq.(1.161), we infer that

$$\mathbf{r}_{\alpha_1\alpha_2}\big|_{k_1k_2} = \sum_{l_1l_2 \in \Omega_1 \times \Omega_2} \Lambda^{l_1}_{\alpha_1} \Lambda^{l_2}_{\alpha_2} \Lambda^{k_1k_2}_{l_1l_2} \tag{1.162}$$

where

$$\Lambda^{k_1k_2}_{l_1l_2} = \mathbf{r}_{l_1l_2}\big|_{k_1k_2} \tag{1.163}$$

which is the desired second level decompression matrix. This matrix encodes how we move from $p_{l_1l_2}$'s to $p_{k_1k_2}$'s. Using the definition of the first level decompression matrix, Eq.(1.162) becomes

$$\mathbf{r}_{\alpha_1\alpha_2}\big|_{k_1k_2} = \sum_{l_1l_2 \in \Omega_1 \times \Omega_2} \Lambda^{k_1k_2}_{l_1l_2} \mathbf{r}_{\alpha_1}\big|_{l_1} \mathbf{r}_{\alpha_2}\big|_{l_2} \tag{1.164}$$

This defines the *causaloid product*, denoted $\mathbf{r}_{\alpha_1} \otimes^{\Lambda} \mathbf{r}_{\alpha_2}$ which unifies the different causal structure-specific products. Explicitly,

$$\mathbf{r}_{\alpha_1} \otimes^{\Lambda} \mathbf{r}_{\alpha_2} = \mathbf{r}_{\alpha_1\alpha_2} \tag{1.165}$$

It is this product that allows us to look at the probabilistic correlations between arbitrary locations in space-time without specifying the causal relationship.

We have shown second level compression for the case where we have two regions. This is easily generalized for any number of regions. The object that would encode the compression for three regions would be $\Lambda^{k_1k_2k_3}_{l_1l_2l_3}$, for four regions would be $\Lambda^{k_1k_2k_3k_4}_{l_1l_2l_3l_4}$, etc. After second level compression over multiple regions, we have

$$\begin{pmatrix} \Lambda^{k_1}_{\alpha_1} \\ \Lambda^{k_1k_2}_{l_1l_2} \\ \Lambda^{k_1k_2k_3}_{l_1l_2l_3} \\ \vdots \end{pmatrix} \tag{1.166}$$

There is a third level of physical compression that compresses these multi-region Λ-matrices to give the *Causaloid*, $\mathbf{\Lambda}$, which is defined as

$$\mathbf{\Lambda} \equiv \left(\{\Lambda\} \mid \{\Lambda\} \subseteq \{\Lambda^{k_1}_{\alpha_1}, \Lambda^{k_1k_2}_{l_1l_2}, \ldots\}\right) \tag{1.167}$$

where $\{\Lambda\}$ is determined by the rules of the physical theory (for detailed discussion of how this works see [2]). By decompressing the set $\{\Lambda\}$, we can obtain the Λ-matrix for any set of regions. This means that the Causaloid gives us the ability to perform any calculation that the physical theory allows for.

1.3.4 Well-defined probabilities

Up to this point we have exclusively dealt with probabilities conditioned on procedures. It is more useful to also be able to condition on outcomes. Specifically, we'd like an expression

for the following:

$$\text{Prob}(Y_2^{\alpha_2}|Y_1^{\alpha_1}, F_1^{\alpha_1}, F_2^{\alpha_2}) \tag{1.168}$$

Using Bayes' Theorem, this becomes

$$\text{Prob}(Y_2^{\alpha_2}|Y_1^{\alpha_1}, F_1^{\alpha_1}, F_2^{\alpha_2}) = \frac{\text{Prob}(Y_1^{\alpha_1}, Y_2^{\alpha_2}|F_1^{\alpha_1}, F_2^{\alpha_2})}{\sum_{Y_2^{\beta_2} \sim F_2^{\alpha_2}} \text{Prob}(Y_1^{\alpha_1}, Y_2^{\beta_2}|F_1^{\alpha_1}, F_2^{\alpha_2})} \tag{1.169}$$

where $Y_2^{\beta_2} \sim F_2^{\alpha_2}$ denotes that the sum is over all possible outcomes corresponding to the measurement $F_2^{\alpha_2}$ (in R_2). (For simplicity, we have suppressed the part of the notation denoting the generalized preparation.) In the causaloid framework, this becomes

$$\text{Prob}(Y_2^{\alpha_2}|Y_1^{\alpha_1}, F_1^{\alpha_1}, F_2^{\alpha_2}) = \frac{\mathbf{r}_{\alpha_1\alpha_2} \cdot \mathbf{p}}{\mathbf{r}_{\alpha_1 \frown_2} \cdot \mathbf{p}} \tag{1.170}$$

where $\mathbf{r}_{\alpha_1 \frown_2} = \sum_{\beta_2} \mathbf{r}_{\alpha_1\beta_2}$. (The sum being over β_2 in this notation has the same meaning as the sum being over all outcomes consistent with F_2.) In order for this probability to be considered well-defined, the right hand side cannot depend on $V - R_1 - R_2$. Since $\mathbf{r}_{\alpha_1\alpha_2}$ and $\mathbf{r}_{\alpha_1 \frown_2}$ are determined exclusively by the physical theory, neither has any dependence on $V - R_1 - R_2$. However, $\mathbf{p}$ does depend on $V - R_1 - R_2$. This implies that in order for the probability Eq.(1.170) to be well defined (i.e. not depend on $V - R_1 - R_2$), it must not vary with $\mathbf{p}$. The dependence on $\mathbf{p}$ can be removed altogether by requiring that $\mathbf{r}_{\alpha_1\alpha_2}$ be parallel to $\mathbf{r}_{\alpha_1 \frown_2}$. Therefore, the above probability is well defined if and only if

$$\mathbf{r}_{\alpha_1\alpha_2} \parallel \mathbf{r}_{\alpha_1 \frown_2} \tag{1.171}$$

With this condition, we get

$$\text{Prob}(Y_2^{\alpha_2}|Y_1^{\alpha_1}, F_1^{\alpha_1}, F_2^{\alpha_2}) = \frac{|\mathbf{r}_{\alpha_1\alpha_2}|}{|\mathbf{r}_{\alpha_1\frown_2}|} \tag{1.172}$$

1.3.5 $\odot^\Gamma$ product

Consider two distinct regions; R_A and R_P. By definition

$$\mathbf{r}_{\alpha_A\alpha_P} = \mathbf{r}_{\alpha_A} \otimes^\Lambda \mathbf{r}_{\alpha_P}$$
$$\mathbf{r}_{\beta_A\alpha_P} = \mathbf{r}_{\beta_A} \otimes^\Lambda \mathbf{r}_{\alpha_P}$$

Suppose we wanted to take the dot product between two vectors of the above form. Using decompression matrices, we can write

$$\begin{aligned} \mathbf{r}_{\alpha_A\alpha_P} \cdot \mathbf{r}_{\beta_A\alpha_P} = & \left(\mathbf{r}_{\alpha_A} \otimes^\Lambda \mathbf{r}_{\alpha_P}\right) \cdot \left(\mathbf{r}_{\beta_A} \otimes^\Lambda \mathbf{r}_{\alpha_P}\right) \\ = & \sum_{k_A k_P} \left(\sum_{l_A l_P} \Lambda_{l_A l_P}^{k_A k_P} \mathbf{r}_{\alpha_A}\big|_{l_A} \mathbf{r}_{\alpha_P}\big|_{l_P} \right) \left(\sum_{l'_A l'_P} \Lambda_{l'_A l'_P}^{k_A k_P} \mathbf{r}_{\beta_A}\big|_{l'_A} \mathbf{r}_{\alpha_P}\big|_{l'_P} \right) \end{aligned} \tag{1.173}$$

where $k_A k_P \in \Omega_{AP}$, $l_A l_P \in \Omega_A \times \Omega_P$, and $l'_A l'_P \in \Omega_A \times \Omega_P$. Notice that we can write

$$\sum_{l_A l_P} \Lambda_{l_A l_P}^{k_A k_P} \mathbf{r}_{\alpha_A}\big|_{l_A} \mathbf{r}_{\alpha_P}\big|_{l_P}$$

as

$$\sum_{l_A \in \Omega_A} \left[\mathbf{r}_{\alpha_A}\big|_{l_A} \left(\sum_{l_P \in \Omega_P} \Lambda_{l_A l_P}^{k_A k_P} \mathbf{r}_{\alpha_P}\big|_{l_P} \right) \right]$$

Similarly,

$$\sum_{l'_A l'_P} \Lambda^{k_A k_P}_{l'_A l'_P} \mathbf{r}_{\beta_A}\big|_{l'_A} \mathbf{r}_{\alpha_P}\big|_{l'_P} = \sum_{l'_A \in \Omega_A} \left[\mathbf{r}_{\beta_A}\big|_{l'_A} \left(\sum_{l'_P \in \Omega_P} \Lambda^{k_A k_P}_{l'_A l'_P} \mathbf{r}_{\alpha_P}\big|_{l'_P} \right) \right]$$

Define

$$\Gamma^{k_A k_P}_{l_A}(\mathbf{r}_{\alpha_P}) \equiv \sum_{l_P \in \Omega_P} \Lambda^{k_A k_P}_{l_A l_P} \mathbf{r}_{\alpha_P}\big|_{l_P}$$

and, similarly,

$$\Gamma^{k_A k_P}_{l'_A}(\mathbf{r}_{\alpha_P}) \equiv \sum_{l'_P \in \Omega_P} \Lambda^{k_A k_P}_{l'_A l'_P} \mathbf{r}_{\alpha_P}\big|_{l'_P}$$

Using this, Eq.(1.173) becomes

$$\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\beta_A \alpha_P} = \sum_{k_A k_P} \left(\sum_{l_A l'_A} \Gamma^{k_A k_P}_{l_A}(\mathbf{r}_{\alpha_P}) \Gamma^{k_A k_P}_{l'_A}(\mathbf{r}_{\alpha_P}) \mathbf{r}_{\alpha_A}\big|_{l_A} \mathbf{r}_{\beta_A}\big|_{l'_A} \right) \tag{1.174}$$

where $k_A k_P \in \Omega_{AP}$ and $l_A l'_A \in \Omega_A \times \Omega_A$. This suggests that the essence of $\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\beta_A \alpha_P}$ is a relationship between $\mathbf{r}_{\alpha_A}$ and $\mathbf{r}_{\beta_A}$ mediated by matrices that depend on $\mathbf{r}_{\alpha_P}$. Therefore, we can view Eq.(1.173) as kind of product of $\mathbf{r}_{\alpha_A}$ and $\mathbf{r}_{\beta_A}$. Dot products of this form come up frequently enough that we will define this as the Γ-dot product and denote it as

$$\mathbf{r}_{\alpha_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\beta_A} \equiv \mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\beta_A \alpha_P} = \left(\mathbf{r}_{\alpha_A} \otimes^{\Lambda} \mathbf{r}_{\alpha_P} \right) \cdot \left(\mathbf{r}_{\beta_A} \otimes^{\Lambda} \mathbf{r}_{\alpha_P} \right) \tag{1.175}$$

Chapter 2

Discussions of the Applicability of the Models

Equipped with the three frameworks from chapter 1 we now set about to demonstrate that each of these frameworks can be applied to make progress. Each of these frameworks are very different in their structure and their intent, and so each is suitable for studying different problems. With this in mind we will focus here on the applications to which each framework is well suited rather than pushing them beyond their natural capabilities.

The causaloid framework is a probabilistic framework without a fixed causal structure. The result of creating such a framework is that we lose contact with many of the quantities that allow us to have a physical intuition with which to interpret results. One such quantity that we lose contact with is entropy: the second law of thermodynamics is in particular intertwined with the notions of causality and time, and so it will require effort to reconnect

with it. In section 2.1 I introduce the concept of causally unbiased entropy in an effort to produce a quantity which will allow us to gain a physical intuition for the framework.

The ontic models framework similarly has difficulties connecting its internal statements to reality due to its high level of abstraction. Despite this, the ontic models framework is very good at making statements about general properties of measurements. Due to this capability of the model it is the ideal framework to address the concept of contextuality - a property of the results of measurements to depend on the choice of what other measurements have been performed at the same time. In section 2.2 we will reproduce both the original idea of contextuality due to [16] and its extensions from [34]. Additionally in section 2.3 we will discuss the concept of deficiency from [12] and present a way of relating it to contextuality.

Nelson's mechanics has a very different character than the ontic models framework or the causaloid framework, instead being very definite in its form: dealing with trajectories rather than abstract states. For this reason, Nelson's mechanics can be used to examine an alternate set of questions from the other frameworks. In section 2.4.1 we will introduce Bohmian mechanics - another realist theory of quantum mechanics - and then in section 2.4.2 we will compare the construction of Bohmian mechanics to that of Nelson's mechanics.

2.1 Entropy in the Causaloid

The content of this section is a result of a collaboration with Lucien Hardy [18].

It is natural in discussions of causal structure to raise the question of entropy. The

second law of thermodynamics tells us that in an isolated system, entropy can increase or remain the same, but it can never decrease [17]. In information theory, entropy is viewed as being a measure of uncertainty before we measure a state or equivalently, the amount of information gained by upon learning the state of a system [23]. As is the basis for other work [19, 14, 15] we shall take an equivalence of these two types of entropy. Inherent in this concept of entropy is an assumed causal structure, specifically that there exists a background time[30]. The standard definition of entropy is in the context of a definite causal structure with reference to absolute time. In order to make sense of entropy in an indefinite causal structure, a clear definition must be established. To do so requires consideration of the following questions:

> *What are the concepts from the usual picture of entropy in a definite causal structure that are necessary to define entropy? What are the analogues to these concepts in a picture with indefinite causal structure?*

Using the formalism introduced in section 1.3, we are able to provide answers to these questions and then, define a causally-unbiased entropy.

Standard definitions of entropy assume fixed causal structure. Here we develop a causally-unbiased definition of entropy in the causaloid formalism.

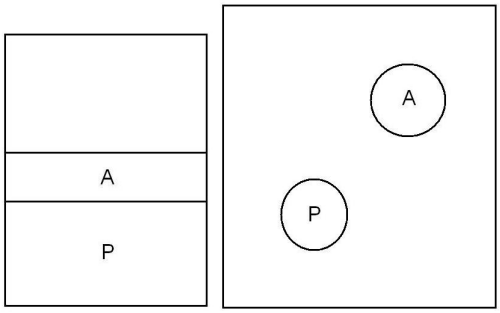

Figure 2.1: (a) Fixed causal structure (b) Indefinite causal structure

2.1.1 The picture

The Shannon entropy [23] for a classical state is defined as

$$S = -\sum_i p_i log_2 p_i \tag{2.1}$$

The definition of p_i used in this equation requires that the structure of space-time be organized with the following features:

- a region of interest, A
- an immediate past space-time region, P
- sufficient data about what happened in P
- a measurement F_A
- a set of outcomes, $\{Y_A^i\}$, corresponding to F_A

This allows us to write

$$p_i = \text{Prob}(Y_A^i | F_A, \text{data}_P) \tag{2.2}$$

Removing all time bias from these features of space-time structure, we get

- a region of interest, A
- a reference region P
- an outcome/measurement pair in P, $\{\alpha_P\} = \{(F_P, Y_P)\}$
- a measurement F_A
- a set of outcomes, $\{Y_A^i\}$, corresponding to F_A

The reference region can be thought of as a kind of preparation region that is not limited to being in the causal past. In fact, the choice of reference region is arbitrary as illustrated in Fig. 2.1b.

The definition of p_i in a causally-unbiased structure is

$$p_i = \text{Prob}(Y_A^i | Y_P, F_P, F_A) \tag{2.3}$$

(Since P is arbitrary, we should technically say 'p_i with respect to the reference region P'. However, for the sake of brevity, we will assume that 'with respect to P' is implied much as 'with respect to the past' is taken as implied in the causally-biased situation.)

Using the above definition of p_i, we define the entropy relative to the reference data (F_P, Y_P) as

$$S = -\sum_i \text{Prob}(Y_A^i|Y_P, F_P, F_A) \log_2 \left(\text{Prob}(Y_A^i|Y_P, F_P, F_A)\right) \tag{2.4}$$

Notice that this reduces to the causally-biased definition of entropy when P is the past; F_A measures the microstate in the classical case or measures in the basis where $\hat{\rho}$ is diagonal in the quantum case.

2.1.2 In the causaloid framework

Taking the probability to be well-defined, Eq.(1.172) and Eq.(2.4) give the following definition of entropy:

$$S = -\sum_{\alpha_A} \frac{|\mathbf{r}_{\alpha_A\alpha_P}|}{|\mathbf{r}_{\frown_A\alpha_P}|} \log_2 \left(\frac{|\mathbf{r}_{\alpha_A\alpha_P}|}{|\mathbf{r}_{\frown_A\alpha_P}|}\right) \tag{2.5}$$

Of course, this equation requires that $\mathbf{r}_{\alpha_A\alpha_P} \parallel \mathbf{r}_{\frown_A\alpha_P}$. We can also consider what happens when $\mathbf{r}_{\alpha_A\alpha_P}$ is nearly parallel to $\mathbf{r}_{\frown_A\alpha_P}$, using the definition of the probability from Eq.(1.170). The entropy associated with this is

$$S^\Lambda = -\sum_{\alpha_A} \left(\frac{\mathbf{r}_{\alpha_A\alpha_P} \cdot \mathbf{p}}{\mathbf{r}_{\frown_A\alpha_P} \cdot \mathbf{p}}\right) \log_2 \left(\frac{\mathbf{r}_{\alpha_A\alpha_P} \cdot \mathbf{p}}{\mathbf{r}_{\frown_A\alpha_P} \cdot \mathbf{p}}\right) \tag{2.6}$$

It becomes necessary to shorten the notation for the following work so $\mathbf{r}_{\alpha_A\alpha_P}$ will be denoted as $\mathbf{v}_i$ (where the index α_A is represented by i) and $\mathbf{r}_{\frown_A\alpha_P}$ will be denoted as $\mathbf{u}$. As with any vector, $\mathbf{v}_i$ can be decomposed into a component parallel to $\mathbf{u}$ and a component

perpendicular to $\mathbf{u}$ (i.e. components in $\hat{\mathbf{u}}^{\parallel}$ and $\hat{\mathbf{u}}^{\perp}$, respectively). That is,

$$\mathbf{v}_i = v_i^{\parallel}\hat{\mathbf{u}}^{\parallel} + v_i^{\perp}\hat{\mathbf{u}}^{\perp} \tag{2.7}$$

Using the unit vectors as defined, $\mathbf{p}$ can be decomposed as

$$\mathbf{p} = p_x\hat{\mathbf{u}}^{\parallel} + p_y\hat{\mathbf{u}}^{\perp} + \mathbf{p}^{\perp} \tag{2.8}$$

where $\mathbf{p}^{\perp}$ is the component of $\mathbf{p}$ that is perpendicular to the plane defined by $\mathbf{u}$ and $\mathbf{v}_i$. The probability of interest, p_i, then becomes

$$\begin{aligned} p_i = & \frac{\mathbf{v}_i \cdot \mathbf{p}}{\mathbf{u} \cdot \mathbf{p}} \\ = & \frac{v_i^{\parallel}}{u} + k\frac{v_i^{\perp}}{u} \end{aligned} \tag{2.9}$$

where $k = \frac{p_y}{p_x}$. Notice that the first term is equivalent to a well-defined probability (Eq. 1.172). We require the second term to be small since the deviation from well-defined should be small. Since we have already required that $\mathbf{v}_i^{\perp}$ be small, we need only place restrictions on k.

2.1.3 Bounds on k

For the purposes of this subsection, we will work in the plane defined by $\mathbf{u}$ and $\mathbf{v}_i$. Define the angle between $\mathbf{u}$ and the projection of $\mathbf{p}$ into the plane to be θ. Define the length of

the projection of $\mathbf{p}$ into the plane to be p_{xy}. Using basic trigonometry, we get

$$p_y = p_{xy} sin\theta \tag{2.10}$$

$$p_x = p_{xy} cos\theta \tag{2.11}$$

Therefore, k can be written in a form that is dependent on only one variable, as follows:

$$k = tan\theta \tag{2.12}$$

As θ tends towards $\pm\frac{\pi}{2}$, k tends to infinity. Therefore, to ensure that the second term of (2.9) is small, we require that k be finite. Assume it to be a property of the state space for $\mathbf{p}$ that there exists some $0 < \theta_{max}$. Clearly, $|\theta_{max}| < \frac{\pi}{2}$ in order for k to be finite. So θ is bounded as follows:

$$-\frac{\pi}{2} < -\theta_{max} \leq \theta \leq \theta_{max} < \frac{\pi}{2} \tag{2.13}$$

The k corresponding to θ_{max} will be denoted as k_{max}. Further bounds can be placed on k by the state space of the physical theory. For our purposes, it is sufficient that k is finite.

2.1.4 Q factor

In light of (2.9), entropy, as defined in (2.6), becomes

$$\begin{aligned} S^{\Lambda} = & -\sum_i \left(\frac{v_i^{\parallel}}{u} + k\frac{v_i^{\perp}}{u}\right) \log_2 \left(\frac{v_i^{\parallel}}{u} + k\frac{v_i^{\perp}}{u}\right) \\ = & -\sum_i \left(\frac{v_i^{\parallel}}{u} + k\frac{v_i^{\perp}}{u}\right) \left[\log_2 \left(1 + k\frac{v_i^{\perp}}{v_i^{\parallel}}\right) + \log_2 \left(\frac{v_i^{\parallel}}{u}\right)\right] \end{aligned} \tag{2.14}$$

Since $v_i^{\perp}$ is very small (as is implied by the fact that $\mathbf{v_i}$ and $\mathbf{u}$ are nearly parallel) and k is finite, we can take a Taylor expansion (to leading order) of the first $\log_2$ term. Doing this gives

$$\begin{aligned} S^{\Lambda} = & -\sum_i \left(\frac{v_i^{\parallel}}{u} + k\left(\frac{v_i^{\perp}}{u}\right)\right) \left[\frac{k}{\ln 2}\left(\frac{v_i^{\perp}}{v_i^{\parallel}}\right) + \mathcal{O}\left({v_i^{\perp}}^2\right) + \log_2 \left(\frac{v_i^{\parallel}}{u}\right)\right] \\ = & -\sum_i \left(\frac{v_i^{\parallel}}{u}\right) \log_2 \left(\frac{v_i^{\parallel}}{u}\right) + k\left(\frac{v_i^{\perp}}{u}\right) \log_2 \left(e\frac{v_i^{\parallel}}{u}\right) + \mathcal{O}\left({v_i^{\perp}}^2\right) \end{aligned} \tag{2.15}$$

Notice that the first term is equivalent to the definition of entropy where $\mathbf{u} \parallel \mathbf{v_i}$ and that S^{Λ} reduces to this definition when $v_i^{\perp} = 0$. That is, when $\mathbf{u} \parallel \mathbf{v_i}$ (or equivalently, $v_i^{\perp} = 0$)

$$S^{\Lambda} = S \equiv -\sum_i \left(\frac{v_i^{\parallel}}{u}\right) \log_2 \left(\frac{v_i^{\parallel}}{u}\right) \tag{2.16}$$

For $v_i^{\perp} \neq 0$, we will define

$$Q = -\sum_i \left(\frac{v_i^{\perp}}{u}\right) \log_2 \left(e\frac{v_i^{\parallel}}{u}\right) \tag{2.17}$$

Using k_{max} as defined in the previous section, we can regard $k_{max}Q$ as a kind of correction to the causally-biased entropy. Then, to leading order

$$S - k_{max}Q \leq S^{\Lambda} \leq S + k_{max}Q \tag{2.18}$$

2.1.5 Understanding Q

Q is an entirely new quantity with no direct classical analogue so understanding its physical interpretation is a non-trivial matter. If we consider entropy as a measure of uncertainty, then S is the measure of our uncertainty that the measurement F_A in region A will yield the specific outcome Y_A^i, given the data we have from the reference region P. Since our reference region P is arbitrary, one way to view Q is that it measures how completely the region P influences region A. In a definite causal structure, an immediate past region would be the exclusive influence on our region of interest and Q would be zero. However, in the causally-indefinite picture, we cannot require a priori that the reference region we have chosen will be the exclusive influence on our region of interest. If there are no influences on region A from outside region P, then the probability will be well-defined and Q will be zero. But if there are influences on region A from outside region P, then the magnitude of Q will reflect that.

2.1.6 Using the $\odot^{\Gamma}$ product

For the sake of completeness the $\mathbf{u}$'s and $\mathbf{v}_i$'s must be translated into $\mathbf{r}_{\frown_A \alpha_P}$'s and $\mathbf{r}_{\alpha_A \alpha_P}$'s. Notice that

$$\frac{v_i^{\parallel}}{u} = \frac{\mathbf{v_i} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \tag{2.19}$$

$$\frac{v_i^{\perp}}{u} = \sqrt{\frac{v_i^2 - {v_i^{\parallel}}^2}{u^2}} = \sqrt{\frac{\mathbf{v_i} \cdot \mathbf{v_i}}{\mathbf{u} \cdot \mathbf{u}} - \frac{(\mathbf{v_i} \cdot \mathbf{u})^2}{(\mathbf{u} \cdot \mathbf{u})^2}} \tag{2.20}$$

Substituting $\mathbf{r}_{\alpha_A \alpha_P}$ for $\mathbf{v}_i$ and $\mathbf{r}_{\frown_A \alpha_P}$ for $\mathbf{u}$ gives

$$\frac{v_i^{\parallel}}{u} = \frac{\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\frown_A \alpha_P}}{\mathbf{r}_{\frown_A \alpha_P} \cdot \mathbf{r}_{\frown_A \alpha_P}} \tag{2.21}$$

$$\frac{v_i^{\perp}}{u} = \sqrt{\frac{\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\alpha_A \alpha_P}}{\mathbf{r}_{\frown_A \alpha_P} \cdot \mathbf{r}_{\frown_A \alpha_P}} - \frac{(\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\frown_A \alpha_P})^2}{(\mathbf{r}_{\frown_A \alpha_P} \cdot \mathbf{r}_{\frown_A \alpha_P})^2}} \tag{2.22}$$

Using the Γ-dot product the above equations simplify to

$$\frac{v_i^{\parallel}}{u} = \frac{\mathbf{r}_{\alpha_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\frown_A}}{\mathbf{r}_{\frown_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\frown_A}} \tag{2.23}$$

$$\frac{v_i^{\perp}}{u} = \sqrt{\frac{\mathbf{r}_{\alpha_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\alpha_A}}{\mathbf{r}_{\frown_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\frown_A}} - \left(\frac{\mathbf{r}_{\alpha_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\frown_A}}{\mathbf{r}_{\frown_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\frown_A}} \right)^2} \tag{2.24}$$

This allows us to completely specify the entropy of R_A relative to a preparation R_P in the causaloid framework. It is straightforward to generalize this to define the joint entropy of

R_A and R_B with reference to a "preparation" R_P. Simply redefine $\mathbf{u}$ and $\mathbf{v_{ij}}$ as

$$\mathbf{v}_{ij} = \mathbf{r}_{\alpha_A \alpha_B \alpha_P} \tag{2.25}$$

$$\mathbf{u} = \mathbf{r}_{\frown_A \frown_B \alpha_P} \tag{2.26}$$

where

$$\mathbf{r}_{\alpha_A \alpha_B \alpha_P} = \mathbf{r}_{\alpha_A} \otimes^{\Lambda} \mathbf{r}_{\alpha_B} \otimes^{\Lambda} \mathbf{r}_{\alpha_P}$$

and

$$\mathbf{r}_{\frown_A \frown_B \alpha_P} = \sum_{\beta_A} \mathbf{r}_{\beta_A} \otimes^{\Lambda} \sum_{\beta_B} \mathbf{r}_{\beta_B} \otimes^{\Lambda} \mathbf{r}_{\alpha_P}$$

Using the same procedure as for one region, we get

$$\frac{v_{ij}^{\parallel}}{u} = \frac{\mathbf{r}_{\alpha_A \alpha_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\frown_A \frown_B}}{\mathbf{r}_{\frown_A \frown_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\frown_A \frown_B}} \tag{2.27}$$

$$\frac{v_{ij}^{\perp}}{u} = \sqrt{\frac{\mathbf{r}_{\alpha_A \alpha_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\alpha_A \alpha_B}}{\mathbf{r}_{\frown_A \frown_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\frown_A \frown_B}} - \left(\frac{\mathbf{r}_{\alpha_A \alpha_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\frown_A \frown_B}}{\mathbf{r}_{\frown_A \frown_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\frown_A \frown_B}} \right)^2} \tag{2.28}$$

In this manner, we can define causally-unbiased entropy in the causaloid framework for any number of regions.

2.1.7 Conclusions

In a definite causal structure, the only thing required for a definition of entropy that is not in an indefinite causal structure is an immediate past region. Since there is no reason in an indefinite causal structure to choose any reference region over any other, we simply choose

an arbitrary region. This ensures that we do not hold on to any pre-conceived notions of space-time and its connection to causality. The definition of the causally-unbiased entropy resulted in a correction to the causally-biased definition of entropy. In a sense, the Q factor gives us an emergent idea of causality. It is a measure of the extent to which our region of interest is causally connected to our reference (or "preparation") region. If it is zero, the traditional ideas of causality are recovered. The next step would be determining how the Q factor could potentially be physically observed. To do so may require us to know more of the theoretical and mathematical properties of Q. Which mathematical properties of Shannon entropy hold for causally-unbiased entropy? What is the status of the Second Law of Thermodynamics in an indefinite causal structure? To go about answering this, we could consider how S^{Λ} "evolves" along tubes through indefinite space-times.

2.2 Contextuality's Generalization in Ontic Models

The study of ontological models has a peculiar history. A portion of the seminal papers in the subject actually take the form of attempted 'no-go' theorems, intended to prove that the subject was a dead-end. Beginning with the result of Von Neumann in [36] which instead of the desired result (of proving what were at the time referred to as hidden variables theories were incorrect) led to a rebuttal by Bohm with a counterexample[4], and then the Kochen-Specker paper [16] which again purported to be a no-go result, but then instead became an introduction to the concept (and necessity within ontological models) of contextuality.

Contextuality in its basic sense is used to describe a theory in which the relationship

of the observable results depend not solely on the states of the theory, but also depend upon the set of what quantities are being observed. Classical mechanics is an example of a non-contextual theory, best evidenced by the fact that the observables in classical mechanics are also the states of the theory.

A picture of how such a property could exist in terms of the ontic models framework emerges from considering three measurements of a quantum system. The first measurement A commutes with the other two measurements we consider B and C (all of which take values of either 0 or 1). Contextuality takes the form of the portions of the ontic space corresponding to $0_A0_B \cup 0_A1_B$ and $0_A0_C \cup 0_A1_C$ not coinciding exactly. In terms of the ontic models framework this can be expressed as

$$\begin{aligned} &\text{Supp}\left[\xi\left(0_A0_B \mid \lambda, S_{AB}\right)\right] \cup \text{Supp}\left[\xi\left(0_A1_B \mid \lambda, S_{AB}\right)\right] \\ &\neq \text{Supp}\left[\xi\left(0_A0_C \mid \lambda, S_{AC}\right)\right] \cup \text{Supp}\left[\xi\left(0_A1_C \mid \lambda, S_{AC}\right)\right], \end{aligned} \tag{2.29}$$

for a deterministic ontological model. This means that we can't use such a union to devise a notion of the portion of the ontic space corresponding to 0_A (or likewise for 1_A) by using either of these measurement configurations. We illustrate this in figure 2.2 where the regions of support are labeled by the measurement outcomes.

We shall represent the central results from [16] and the extensions of the concept of contextuality from [34] here, in the hopes of providing an introduction to the subject. We will - where possible - reframe the results in the modern context of ontological models, and the modern understanding of the impact of the results.

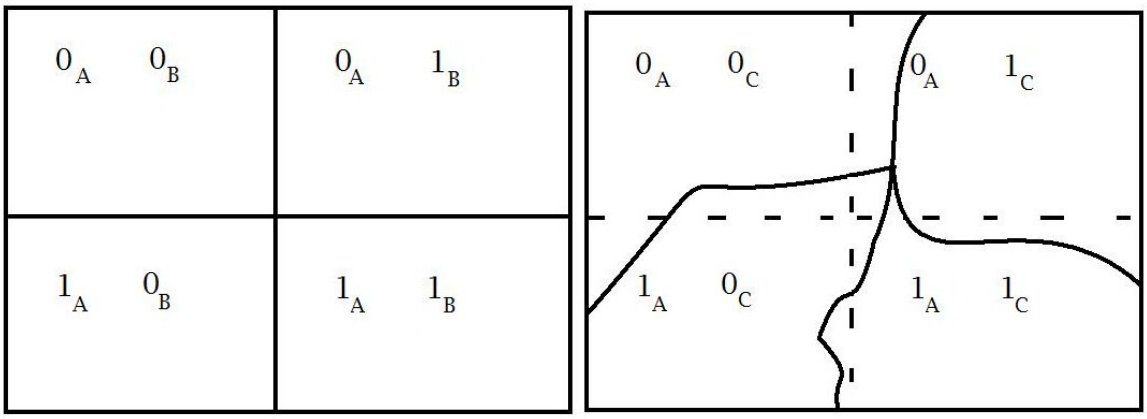

Figure 2.2: The ontic space from measuring A with two different contexts

2.2.1 Kochen and Specker

The Kochen-Specker theorem was originally introduced in [16] as a no-go theorem for hidden variables approaches to quantum mechanics. As with Von Neumann's no-go theorem, there is an 'out' to the theorem, which then became a restriction on viable hidden variables theories. As such the modern statement of the Kochen-Specker theorem is instead that hidden variables theories must be contextual. We will outline the argument that underlies the Kochen-Specker theorem, clarifying where the assumption of contextuality entered and thus demonstrating the more modern form of the theorem.

The central objects within the construction of the Kochen-Specker theorem are the so called 'commeasurable observables' which are a generalization of the commuting observables of standard quantum mechanics. Formally we give the definition that

A set of observables A_i (with i being elements of an indexing set) is said to

be ***comeasurable*** if there exists an observable B and some set of measurable functions f_i (strictly speaking the functions must be Borel, but this is a requirement on the source and target spaces and not on the functions themselves) such that for each i

$$A_i = f_i(B) \tag{2.30}$$

This definition then requires that commeasurability means what we would naively expect: by measuring B we can apply the different functions to our result to have measurements for all of the A_i simultaneously. As measurability of functions is preserved under addition, scalar multiplication and taking products of the functions, we can form what is called a partial algebra.

A ***Partial Algebra*** is a set A over a field $\mathbf{F}$, together with: addition, scalar multiplication, a product from $A \times A$ to A, an identity element $1 \in A$, and a reflexive symmetric binary relation $\bowtie$, subject to the following restrictions:

- $a \bowtie 1$ for all $a \in A$
- $\bowtie$ is closed under all the operations (addition, scalar multiplication and the product)
- Given three elements that each satisfy the binary relation with each other, the values of the polynomials in the three elements form a commutative algebra over the field $\mathbf{F}$

We can moreover define homomorphisms between partial algebras:

Given two partial algebras A and B over a field $\mathbf{F}$, a map $h : A \to B$ is a ***partial algebra homomorphism*** if it is compatible with each of the operations and with the binary relation. Namely, for $a, b \in A$, $\alpha, \beta \in \mathbf{F}$ and $a \bowtie b$:

$$h(a) \bowtie h(b) \tag{2.31}$$

$$h(\alpha a + \beta b) = \alpha h(a) + \beta h(b) \tag{2.32}$$

$$h(ab) = h(a)h(b) \tag{2.33}$$

$$h(1_A) = 1_B \tag{2.34}$$

Where 1_A and 1_B are the identity elements in A and B respectively.

Looking at a general commutative algebra K, we can see that we can construct from it a partial algebra by taking the set of elements following the binary relation to be the full space $K \times K$.

Working from these definitions we have a way of discussing the underlying ideas of the hidden variables formulation of quantum mechanics. Kochen and Specker put forward the premise that the underlying notion of a hidden variables theory is that the partial algebra of observables of quantum mechanics can be embedded into a commutative algebra.

We can then present information necessary for the Kochen-Specker theorem. Beginning with a hidden state space Λ we consider the set $\mathbf{R}^\Lambda$ of all functions from the hidden states to the real numbers, this space of functions forms a commutative algebra. It is this commutative algebra that we embed the partial algebra of observables into. Each hidden state $\lambda \in \Lambda$ then defines a homomorphism from the partial algebra of quantum

observables to the real numbers through an element of $\mathbf{R}^{\Lambda}$. We express this as follows: for any observable A in the algebra of partial observables, we define the homomorphism h to the real numbers through the element $f_A \in \mathbf{R}^{\Lambda}$ by $h(A) = f_A(\lambda)$.

This has within it an assumption of non-contextuality: consider three observables a, b, c with $a \bowtie b$, $a \bowtie c$, but b not being comeasurable with c. As a and b are comeasurable, there exists a single observable M such that both are determined from a measurement of M, likewise as a and c are comeasurable, there exists as observable N from which they can be determined. That the value of $h(a)$ depends on $f_a(\omega)$ and not on M or N, means that the value is independent of the context in which it is measured.

The primary result of [16] then takes the following form

> There exist physical systems for which such a homomorphism cannot exist. As a result, the only valid hidden variable theories are contextual.

We will construct the counter example as follows: consider the problem of measuring the square of the spin of a spin-1 object. From standard quantum mechanics we know that J_x^2, J_y^2 and J_z^2 are co-measurable as they commute. We likewise know that for a spin-1 system the sum of the values of these three observables must be 2 (in natural units). Extending from the fact that these objects only can have values of 0 or 1 we can reach the conclusion that for any assignment of values to the three observables only one can have a value of 0. We can also extend this to any triple of squared spin observables which are mutually orthogonal (and also therefore co-measurable).

We can then consider the following problem: for a non-contextual hidden variables theory there must exist a homomorphism which assigns values to all possible sets of these

observables simultaneously. The issue is that there exists a set of observables of this form so that there is no assignment of values to the directions that can be made without contradicting the requirement that only one of the three values of a set of mutually orthogonal can be 0.

The particular proof of the existence of such a set of directions is done by demonstrating that there exists a graph which is realizable on a sphere in the following sense: the points of the graph correspond to points on a sphere, and edges of the graph correspond to the points they connect being orthogonal. A trio of mutually orthogonal points then takes the form of a triangle within such a graph. If there exists a homomorphism then we should be able to label the vertices of any realizable graph with values 0 or 1 such that:

- No edge connects two vertices labeled with 0
- No triangle contains three 1s.

In particular, there exists a (rather complicated) graph which is proven in [16] to be realizable which does not admit a homomorphism. The existence of this graph gives us the desired contradiction: we therefore cannot construct a hidden variable theory which is non-contextual and gives a consistent value assignment to the observables for this scenario.

Examining the form of this proof gives a natural way of seeing how contextuality provides the 'out' for the no-go theorem as it was originally presented. Allowing for contextuality reduces our problem to only needing to construct a homomorphism which simultaneously assigns values for a set of commeasurable observables - in this case, a triangle. Assigning such a valuation in this case while respecting the requirements defined above is trivial.

It is interesting to note that the form of the Kochen Specker theorem does not prove that non-contextual hidden variable theories are always inconsistent. The statement of the result is not broad: it only states that there are physical systems which cannot consistently be described by non-contextual hidden variable theories. It is therefore possible that we can consistently characterize many physical system by using non-contextual hidden variable theories.

It was pointed out in [12] that there are three implicit restrictions within the Kochen-Specker theorem. From the perspective of ontic models we can see that the theorem only applies to outcome deterministic theories. Looking at the structure of the proof we can also see that it only considers projective-valued measures (PVMs) and not POVMs in general. Lastly it relies on the fact that the Hilbert space is 3 or more dimensional through the fact that the projector is not uniquely defined for more than 2 dimensions. To see this we observe that in 2 dimensions there is a unique perpendicular to any given direction, whereas in 3 or more dimensions we have an infinite number of ways to choose the perpendiculars. In constructing a PVM we construct the context of the measurement, and thus in 2 dimensions there is no ability to form a contextual measurement.

2.2.2 Spekkens

The work done in [34] extends the notions of contextuality from those introduced in [16] in three major ways. The first is to extend the idea to non-deterministic ontological models, the second is to extend the idea of contextuality to a similar property regarding preparations and lastly to extend the idea of contextuality to measurements associated with

positive-operator valued measures (POVMs). These extensions are evident in the form by which we have constructed our formalism for ontological models, but we will nonetheless go through each in kind.

The extension to non-deterministic ontological models is encapsulated in how we construct the functions associated with measurement in our ontological theory. In an ontological model which is deterministic the function ξ gives a probability of 1 to the measurement result corresponding to the ontological state and a probability of 0 to all others. By allowing ξ to take the form of a general probability distribution over the measurement outcomes, we achieve the extension of the notion of contextuality to non-deterministic ontological models simply through the allowance of ξ to depend on the measurement procedure.

Preparation contextuality is the extension of the idea of contextuality to the possibility that the means by which a state is prepared could influence the probability distribution μ over the ontic space. This possibility is reflected in our general framework through allowing μ to depend on S_P the preparation procedure for the state.

The last extension - to POVMs - follows from the extension to non-deterministic ontological models, along with our framework having the allowance for general measurement procedures. Thus all of the information for this extension is contained in ξ taking the form of $\xi\left(k|\ \lambda, S_M\right)$.

No Go Theorems

The results in [34] include no-go theorems in two dimensions for both preparation contextuality and also one for measurement contextuality for POVMs. We will replicate both of

these here.

Both proofs make use of the same set of six two dimensional vectors, and in many ways are reminiscent of the proof from Kochen and Specker in their use of pairs of orthogonal vectors to construct contradictory statements. We first give the set of six vectors:

$$v_a = (1, 0) \tag{2.35}$$

$$v_A = (0, 1)$$

$$v_b = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$$

$$v_B = \left(\frac{\sqrt{3}}{2}, \frac{-1}{2}\right)$$

$$v_c = \left(\frac{1}{2}, \frac{-\sqrt{3}}{2}\right)$$

$$v_C = \left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$$

We can see that for any letter pairing the inner product between them is 0. We can then construct the rank 1 density operators and the projective measurement operators associated to each of these vectors (these are the same), we will label them σ and D respectively. We

find for these

$$\sigma_a = D_a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{2.36}$$

$$\sigma_A = D_A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\sigma_b = D_b = \begin{pmatrix} \frac{1}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{3}{4} \end{pmatrix}$$

$$\sigma_B = D_B = \begin{pmatrix} \frac{3}{4} & \frac{-\sqrt{3}}{4} \\ \frac{-\sqrt{3}}{4} & \frac{1}{4} \end{pmatrix}$$

$$\sigma_c = D_c = \begin{pmatrix} \frac{1}{4} & \frac{-\sqrt{3}}{4} \\ \frac{-\sqrt{3}}{4} & \frac{3}{4} \end{pmatrix}$$

$$\sigma_C = D_C = \begin{pmatrix} \frac{3}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{1}{4} \end{pmatrix}$$

The orthogonality of the letter pairings of these matrices follows from the orthogonality of the corresponding vectors.

$$\sigma_a \sigma_A = D_a D_A = 0 \tag{2.37}$$

$$\sigma_b \sigma_B = D_b D_B = 0$$

$$\sigma_c \sigma_C = D_c D_C = 0$$

Lastly we can construct the sum of the density or projective measurement operators labeled by the same letter in upper and lower case, along with the sums of the triple of lower case and triple of upper class letters, and find them all to have the same value.

$$\begin{aligned}
\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} &= \frac{1}{2}\sigma_a + \frac{1}{2}\sigma_A = \frac{1}{2}D_a + \frac{1}{2}D_A \\
&= \frac{1}{2}\sigma_b + \frac{1}{2}\sigma_B = \frac{1}{2}D_b + \frac{1}{2}D_B \\
&= \frac{1}{2}\sigma_c + \frac{1}{2}\sigma_b = \frac{1}{2}D_c + \frac{1}{2}D_C \\
= \frac{1}{3}\sigma_a + \frac{1}{3}\sigma_b + \frac{1}{3}\sigma_c &= \frac{1}{3}D_a + \frac{1}{3}D_b + \frac{1}{3}D_c \\
= \frac{1}{3}\sigma_A + \frac{1}{3}\sigma_B + \frac{1}{3}\sigma_C &= \frac{1}{3}D_A + \frac{1}{3}D_B + \frac{1}{3}D_C
\end{aligned} \tag{2.38}$$

We can now consider the question of preparation contextuality. The proof of preparation contextuality (similarly to the one in [16]) is by contradiction, so we will assume that our probability distributions are not functions of the preparation method. Next we observe that for two preparation procedures to be distinguishable their two probability distributions should have no common support on the ontic space, that is

$$\mu(\lambda)\mu'(\lambda) = 0 \quad \forall\lambda. \tag{2.39}$$

Now, consider six preparation procedures corresponding to the pure states defined by the vectors from equation 2.35, which we'll call P_a through P_C. These preparations would give us the density operators from equation 2.36. We can also consider new preparations P_{aA}, P_{bB}, P_{cC}, P_{abc} and P_{ABC}, by allowing us to create a preparation corresponding to each of

the subscripts with even probability. These mixed state procedures would all correspond to the same density matrix $\frac{\mathbf{1}}{2}$. In the ontic model framework each of the procedures would correspond to a probability distribution in the ontic state, which we label μ_a through μ_{ABC}. As each of the pairs of states for each letter correspond to completely distinguishable results, we know that the probability distributions must obey

$$\begin{aligned} \mu_a(\lambda)\mu_A(\lambda) &= 0 \\ \mu_b(\lambda)\mu_B(\lambda) &= 0 \\ \mu_c(\lambda)\mu_C(\lambda) &= 0. \end{aligned} \tag{2.40}$$

Additionally, in order for the ontic model framework to be consistent the mixed state procedures must correspond to a convex combination of the constituent preparation procedures's probability distributions, so we have:

$$\begin{aligned} \mu_{aA}(\lambda) &= \frac{1}{2}\mu_a(\lambda) + \frac{1}{2}\mu_A(\lambda) \\ \mu_{bB}(\lambda) &= \frac{1}{2}\mu_b(\lambda) + \frac{1}{2}\mu_B(\lambda) \\ \mu_{cC}(\lambda) &= \frac{1}{2}\mu_c(\lambda) + \frac{1}{2}\mu_C(\lambda) \\ \mu_{abc}(\lambda) &= \frac{1}{3}\mu_a(\lambda) + \frac{1}{3}\mu_b(\lambda) + \frac{1}{3}\mu_c(\lambda) \\ \mu_{ABC}(\lambda) &= \frac{1}{3}\mu_A(\lambda) + \frac{1}{3}\mu_B(\lambda) + \frac{1}{3}\mu_C(\lambda) \end{aligned} \tag{2.41}$$

We now add the implication of our theory being non-contextual. A non-contextual theory would mean that different preparation procedures corresponding to the same observable state should correspond to the same probability distribution over the ontic state space,

this gives us that there should exist a single distribution $\mu\lambda$ equal to all of these composite distributions

$$\begin{aligned}
\mu\lambda &= \frac{1}{2}\mu_a(\lambda) + \frac{1}{2}\mu_A(\lambda) \\
&= \frac{1}{2}\mu_b(\lambda) + \frac{1}{2}\mu_B(\lambda) \\
&= \frac{1}{2}\mu_c(\lambda) + \frac{1}{2}\mu_C(\lambda) \\
&= \frac{1}{3}\mu_a(\lambda) + \frac{1}{3}\mu_b(\lambda) + \frac{1}{3}\mu_c(\lambda) \\
&= \frac{1}{3}\mu_A(\lambda) + \frac{1}{3}\mu_B(\lambda) + \frac{1}{3}\mu_C(\lambda).
\end{aligned} \tag{2.42}$$

We now must ask the question of if we can satisfy equations 2.40 and 2.42. The three equations of products constrain us the choosing three of our distributions to be 0, and attempting to implement those in equations 2.42 gives us all of the other distributions to be zero. This means that the only solution is the trivial one, which isn't a permissible distribution in our framework, yielding a contradiction. We thus get a no-go result for two dimensional non-contextual ontic theories.

We can now use much of this proof in constructing the proof of measurement contextuality for POVMs for a two dimensional system. We can consider three measurements $M_a = \{D_a, D_A\}$, $M_b = \{D_b, D_B\}$ and $M_c = \{D_c, D_C\}$, with the D's as defined above. As each of the pairs of the D's span the space and are orthogonal, we get that the sum of the pairs is the identity (as mentioned in equation 2.38) and we have that the products of the elements of the pairs are zero (much as we had for the density operators in the previous proof). Putting this into our ontological framework, we get that each of the measurements

has two corresponding probability distributions over the ontic space, for example M_a has two functions $\xi_a(\lambda)$ and $\xi_A(\lambda)$ providing the probability of receiving results corresponding to v_a or v_A for each value of λ. As these are mutually exclusive, and exhaustive options, we get that

$$\begin{aligned} \xi_a(\lambda) + \xi_A(\lambda) &= 1 \\ \xi_b(\lambda) + \xi_B(\lambda) &= 1 \\ \xi_c(\lambda) + \xi_C(\lambda) &= 1. \end{aligned} \tag{2.43}$$

We also make one further assumption: that for PVMs (like those that we have considered so far) we have outcome determinism. This means that for each λ we have a unique assignment of the result of a measurement. In our binary situation, this means that one of the pair will be 0 and the other 1 for each letter pairing. We can represent this as

$$\begin{aligned} \xi_a(\lambda)\,\xi_A(\lambda) &= 0 \\ \xi_b(\lambda)\,\xi_B(\lambda) &= 0 \\ \xi_c(\lambda)\,\xi_C(\lambda) &= 0. \end{aligned} \tag{2.44}$$

We can now construct a POVM from these PVMs by considering a random process where we perform one of M_a, M_b and M_c, each with equal probability, additionally we don't record the details of which process was performed (only recording whether a lower case result or capital result is obtained). We then have a POVM which we'll call $M = \left\{\frac{1}{3}D_a + \frac{1}{3}D_b + \frac{1}{3}D_c, \frac{1}{3}D_A + \frac{1}{3}D_B + \frac{1}{3}D_C\right\}$. In our framework these measurement proce-

dures must represented by convex sums of the probability distributions of the measurements it is composed of. We then get that M is represented by

$$\left\{\frac{1}{3}\xi_a(\lambda)+\frac{1}{3}\xi_b(\lambda)+\frac{1}{3}\xi_c(\lambda),\frac{1}{3}\xi_A(\lambda)+\frac{1}{3}\xi_B(\lambda)+\frac{1}{3}\xi_C(\lambda)\right\} \tag{2.45}$$

However, we have forms for the matrices D_x in the first definition of M, and so we know that the POVM should have the form $\left\{\frac{1}{2}\mathbf{1},\frac{1}{2}\mathbf{1}\right\}$. As we assume non-contextuality, our probability distributions over the ontic space for the POVM should not depend on our preparation method. We should then be able to achieve the same probability distribution by any procedure which assigns equal probabilities to two results independent of the ontic state. This gives us that our distributions must each be equal to $\frac{1}{2}$ and so we get that

$$\begin{aligned}\frac{1}{3}\xi_a(\lambda)+\frac{1}{3}\xi_b(\lambda)+\frac{1}{3}\xi_c(\lambda)&=\frac{1}{2}\\ \frac{1}{3}\xi_A(\lambda)+\frac{1}{3}\xi_B(\lambda)+\frac{1}{3}\xi_C(\lambda)&=\frac{1}{2}\end{aligned} \tag{2.46}$$

We now can get our contradiction: the equations labeled 2.46 and the equations labeled 2.43 and 2.44 do not have a consistent solution, and so we have that for two dimensions we must have measurement contextuality for POVMs.

It should be noted here that these results only apply to ontological models constructed in such a framework. This should be contrasted to the Kochen-Specker theorem which applies to any realist construction of quantum mechanics.

2.2.3 Harrigan & Rudolph

In [12] contextuality was examined further from the perspective of ontic models. They refer to the contextuality of section 2.2.1 as traditional contextuality and define it as follows:

> A model is traditionally contextual if it is both out outcome deterministic and if there exists at least one projection operator such that measurement outcome is dependent on the specific PVM used.

This definition is intended to contrast with a generalized definition of context which they give as:

> The context of an outcome is all of the measurement settings that do not effect the statistics of the outcome.

From this view point we can give an alternate definition of measurement contextuality:

> A model is measurement contextual if the indicator function (ξ) is not unique to a POVM element. Rather it depends on some additional information as well (i.e. context).

Likewise we then express preparation contextuality

> A model is preparation contextual if the epistemic state μ is not unique to the density matrix ρ, rather depending on other information.

These two types of contextuality are independent, in that one can have one, the other, both or neither.

The last observation that we put forward from [12] is that within the ontic models framework there are two ways to achieve traditional contextuality. The first - ξ-contextuality - gives rise to traditional contextuality by varying the support for the indicator function as measurement settings change. The second - λ-contextuality - gives rise to traditional contextuality by instead varying the ontic state of the system as the measurement setting changes. The key is that in either case it forces that how λ is in the support of ξ is dependent on the settings (i.e. the context).

2.3 Deficiency in Ontic Models

The strength of the ontic models framework lies in its ability to discover properties of models of quantum mechanics which would not be clear when looking at these models in the standard formalism. Contextuality is one such property, but in [12] another property, termed deficiency, was put forward.

If $\exists$ a pure quantum state $|\psi\rangle$ such that

$$\text{Supp}\left[\mu\left(\lambda \mid \psi, S_p\right)\right] \subset \text{Supp}\left[\xi\left(\psi \mid \lambda, S_M\right)\right] \tag{2.47}$$

for some choice of S_p and S_M, then the ontological model is deficient.

Just from its obscure nature this is an interesting property to consider, and one that might not arise naturally in the language of a particular model (rather than the general ontic model framework). We will examine this concept in two steps: First, we will attempt to provide an understanding of deficiency as a property and secondly, we will present a theorem from [12] demonstrating that traditional contextuality implies deficiency for ontic models.

Deficiency, as laid out formally, is rather abstract and so the meaning of it can easily be missed. The first thing to note is that it is a statement of existence, and so need only be true for a single ψ along with a single pairing of measurement setup and preparation method. With this in mind, the left hand side of the subset statement then speaks to the set of ontic states which the system could be in if prepared in the state ψ with method S_p, whereas the right hand side speaks to the set of ontic states the system could be in and still be measured to be the state ψ with method S_M. That one of these could be a proper subset of the other implies that our measurement method and preparation methods are essentially giving different definitions of the state ψ in terms of the ontic states. We will discuss the implications of this in section 2.3.1.

Theorem 1. *Any ontic model that is traditionally contextual must be deficient.*

Proof

We divide the proof into two parts. First we address outcome indeterministic models. Suppose for a contradiction that our model is not deficient. Then for all ψ and choices of S_M

and S_P we have

$$Supp\left[\mu\left(\lambda \mid \psi, S_p\right)\right] = Supp\left[\xi\left(\psi \mid \lambda, S_M\right)\right]. \tag{2.48}$$

As ψ should always be measured to be itself we require

$$\int \mathrm{d}\lambda \ \mu\left(\lambda \mid \psi\right)\xi\left(\psi \mid \lambda\right) = 1. \tag{2.49}$$

However

$$\int \mu\left(\lambda \mid \psi\right)\mathrm{d}\lambda = 1, \tag{2.50}$$

and so equation 2.49 can only be satisfied if

$$\xi\left(\psi \mid \lambda\right) = 1 \forall \lambda \in Supp\left[\mu\left(\lambda \mid \psi\right)\right]. \tag{2.51}$$

This means though that all indicator functions for all ψ are deterministic indicator functions, which gives us a contradiction with our assumption of outcome indeterminism.

We now assume our model is outcome deterministic, traditionally contextual and assume for a contradiction that our model is also not deficient. We thus have that for a given S_P

$$Supp\left[\mu\left(\lambda \mid \psi, S_p\right)\right] = Supp\left[\xi\left(\psi \mid \lambda, S_M\right)\right] \forall S_M. \tag{2.52}$$

We can conclude from this that for any choices of ψ, S_M and S'_M that we have

$$Supp\left[\xi\left(\psi \mid \lambda, S_M\right)\right] = Supp\left[\xi\left(\psi \mid \lambda, S'_M\right)\right]. \tag{2.53}$$

This means that our measurements will not be contextual for any choice of ψ giving us a contradiction and completing our proof.

We note that our proof differs slightly from [12] since they do not note that the negation of deficiency requires that equality hold for all ψ rather than for a single ψ. The structure of the proof is not impacted, rather this just closes a logical gap in their presentation.

2.3.1 Measurement, Preparation and Deficiency

The examination of deficiency as a property of ontic models lends itself to discussing the division between measurement methods and preparation methods in the ontic models framework. Even after dissecting the definition of deficiency step by step it can be difficult to fully understand, however this can be alleviated by treating measurement and preparation in a more even handed fashion.

Preparation methods which result in a state ψ in many ways must be considered a measurement as well: that we are preparing a system in a given state means that in some fashion we have measured the system to be in that particular state. If we only consider models which are blind to our intentions or labeling of processes, we can then consider all preparations to in fact be measurements. We can then assume that the measurement settings and preparation methods are part of the same set, and treat them equivalently.

Returning to the notion of deficiency, we note that we can use this consistent treatment of measurement and preparation to clarify it. As deficiency is only the requirement of the existence of some choice ψ, S_M, S_p which obeys the set relation, we can exclude from it all such choices where S_M and S_p are identical. The argument for this exclusion is one of consistency - appealing again to the notion that our model is blind to our intentions. Once we've excluded this case we can see that deficiency relates very much to the concepts of measurement and preparation contextuality. Deficiency becomes the statement that the ontic states associated with a state ψ are dependent upon the method used to prepare the state (or equivalently to measure it). In particular we get that there exists some ψ, S_M and S'_M

$$\text{Supp}\left[\xi\left(\psi \mid \lambda, S_M\right)\right] \subset \text{Supp}\left[\xi\left(\psi \mid \lambda, S'_M\right)\right], \tag{2.54}$$

or equivalently that

$$\text{Supp}\left[\mu\left(\lambda \mid \psi, S_p\right)\right] \subset \text{Supp}\left[\mu\left(\lambda \mid \psi, S'_p\right)\right]. \tag{2.55}$$

These statements then imply the requirement that the model be both preparation and measurement contextual if it is deficient and blind to our intentions with respect to whether processes are measurements or preparations.

2.4 Nelsons Mechanics vs. Bohmian Mechanics

Nelsons mechanics and Bohmian mechanics are often compared to one another due to their similar structure: both attempt to formulate quantum mechanics in terms of a modified classical dynamics. Despite this similarity, the two theories produce very different physical pictures: Bohmian mechanics has a quantum potential which is determined by the solutions of the Schrodinger equation, whereas Nelson's mechanics instead assumes only Brownian motion with a diffusion constant proportional to $\hbar$. In section 2.4.1 we will review the derivation of Bohmian mechanics, and then in section 2.4.2 we will compare the applicability and meaning of these theories.

2.4.1 Bohmian Mechanics

Bohm's goals for his mechanics [4, 5] were to obtain a theory which gave a realist description of individual ensembles (instead of for distributions) and to counter two assumptions of the orthodox interpretation of quantum mechanics that he objected to. The objections were to the assumption that the wave function is a complete description of reality and the assumption that measurement is an inherently unpredictable process. These objections were on the basis of the fact that the assumptions did not give rise to a set of predictions which were uniquely determined by having chosen these assumptions. Here we reproduce Bohm's derivation.

The derivation begins with the Schrodinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = -\left(\frac{\hbar^2}{2m}\right)\nabla^2\psi + V(x)\psi, \tag{2.56}$$

where ψ is still considered a complex function over space. From here we break the function ψ down using a polar decomposition in terms of real functions S and R:

$$\psi = Re^{\frac{iS}{\hbar}}. \tag{2.57}$$

With this decomposition the Schrodinger equation then becomes

$$\frac{\partial R}{\partial t} = \frac{-1}{2m}R\nabla^2 S + 2\nabla R\cdot\nabla S \tag{2.58}$$

$$\frac{\partial S}{\partial t} = -\left[\frac{(\nabla S)^2}{2m} + V(x) - \frac{\hbar^2}{2m}\frac{\nabla^2 R}{R}\right]. \tag{2.59}$$

We now introduce P which in standard quantum mechanics would be interpreted as the probability density from the quality

$$P(x) = R^2(x). \tag{2.60}$$

With P we can substitute into equations 2.58 and 2.59 to reach

$$\frac{\partial P}{\partial t} + \nabla\cdot\left(P\frac{\nabla S}{m}\right) = 0 \tag{2.61}$$

$$\frac{\partial S}{\partial t} + \frac{(\nabla S)^2}{2m} + V(x) - \frac{\hbar^2}{4m}\left(\frac{\nabla^2 P}{P} - \frac{1}{2}\frac{(\nabla P)^2}{P^2}\right) = 0. \tag{2.62}$$

In the limit that $\hbar$ goes to zero, we recover the Hamilton-Jacobi equation from equation 2.62 with

$$\frac{\nabla S}{m} = \vec{v}(x) \tag{2.63}$$

the velocity at each point. We also then have that equation 2.61 is conservation of probability

$$\frac{\partial P}{\partial t} + \nabla \cdot (P\vec{v}(x)) = 0. \tag{2.64}$$

Bohm then observed that the term in $\hbar^2$ could be interpreted as an addition to the classical potential. We call this the quantum potential:

$$U(x) = \frac{-\hbar^2}{4m}\left(\frac{\nabla^2 P}{P} - \frac{1}{2}\frac{(\nabla P)^2}{P^2}\right) \tag{2.65}$$

$$= \frac{-\hbar^2}{2m}\frac{\nabla^2 R}{R}. \tag{2.66}$$

We then have that S is the solution to the Hamilton-Jacobi equation for a probability distribution subject to the potential $V + U$, along with conservation of probability.

These equations can also be used to give equations of motion for individual particles possessing precise positions and momenta. From this we get that a particle is subject to a modification of Newton's laws which includes a quantum potential:

$$m\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2} = -\nabla\left(V(x) - \frac{\hbar^2}{2m}\frac{\nabla^2 R}{R}\right). \tag{2.67}$$

where the initial particle momentum is related to the wave function of quantum mechanics through

$$\vec{p} = \nabla S(x), \tag{2.68}$$

though this fact is not necessary for the use of the modified Newton equation.

The resulting mechanics is then deterministic which presents us with a question as to how it could reproduce the results of quantum mechanics. The solution to this issue is through uncontrollable disturbances from measurement devices: the inaccuracies inherent in measurement restrict our knowledge to probability distributions. Though we have exact trajectories of particles, because we can never measure those trajectories exactly without disturbance this is then characterized as a hidden variable model.

The picture of Bohmian mechanics that comes from this derivation is that of classical mechanics with the addition of a quantum potential that is determined by the solutions of the Schrodinger equation. The corollary of this is that we're left with a modern equivalent of the question of why gravitational mass and inertial mass are the same: why is it that we observe the probability density to be equal to the square of the magnitude of the wave function? In Bohmian mechanics there is no a priori reason that these two must be equal.

2.4.2 Comparing the two mechanics

Equipped with the two derivations from sections 1.1 and 2.4.1 we now contrast the form of the two theories. Before we proceed in contrasting their implications it is interesting to

note that the motivations behind the two theories are very different: Bohm's mechanics is motivated by a desire for a realist description of the universe, whereas Nelson's was motivated by a desire to apply a recent mathematical advance. Despite Nelson's differing motivations he still takes a similar view that there should exist an underlying reality that leads to quantum mechanics. It is to keep these realist structures that both theories share a common starting point in newtonian mechanics, and both attempt not to stray from these roots.

The most striking difference between them is the role of the Schrodinger equation in their derivations. Bohmian mechanics not only assumes the Schrodinger equation, but it then requires that the solution to the equation and its derivative are combined to makeup a potential. Nelsons mechanics instead assumes only diffusion at the level of Planck's constant, and arrives at the conclusion that a probability density under these conditions gives rise to the Schrodinger equation.

The substantive difference in these two theories comes from a subtle difference: Bohmian mechanics assigns to the trajectories the role of a hidden variable, whereas Nelson's mechanics considers them to be accessible but inherently random. The hidden variables in Nelson's mechanics are not made manifest in the theory: the excitations of the hidden variables relax on a time scale which is small compared to the time scale of the trajectories and so they are in effect averaged out to noise-like fluctuations.

Due to these differences in structure the two theories may have very different applications within quantum foundations. In particular Bohmian mechanics acts as a viable candidate for a realist theory of quantum mechanics, whereas Nelson's mechanics instead

tells us that a theory whose hidden variables have dynamics on sufficiently small time scales can lead to probability distributions evolving according to the schrodinger equation.

It is interesting to look at the criticisms that have been made of Nelson's mechanics in [37, 38] and see whether they apply to Bohmian mechanics as well. The criticisms of Nelson's mechanics relate to the need for the S in the breakdown of ψ to be multivalued in certain situations. For instance where the wavefunction is a solution to a problem with angular momentum and S can therefore take on values differing by integer multiples of the azimuthal angle, we then have that the wavefunction is only single valued if these differences are multiples of 2π (which results in multiplication by a factor of 1). What we then have is a constraint on what values of S will give viable solutions to the Schrodinger equation. In these situations we have a problem in Nelson's mechanics: there is no reason that we should assume our solutions also observe this constraint, and so in general we get behaviour which does not correspond to a solution of the Schrodinger equation. On the other hand, as the direction of implication is reversed in Bohmian mechanics - we begin with solutions to the Schrodinger equation and then derive from them the trajectories - the difficulty never arises and this criticism is avoided.

Chapter 3

Discussion and Conclusions

3.1 Compatibilities and Incompatibilities

As we've shown in chapter 2, each of these frameworks can be applied in studying different concepts. What we now want to question is whether the advances within each of these frameworks can be transferred to the others, or whether the choices that went into constructing each framework mean that their results are inapplicable elsewhere.

Both the causaloid and ontic models share a common obstacle to progress: having sacrificed a description of reality in terms of variables based upon our experience and observations for a description in terms of mathematical states. This obstacle is manifest in different ways for each of the two theories. As the causaloid doesn't have an assumption of realism its obstacle is solely computation complexity: that to produce any physical calculation could take an arbitrary amount of effort. The ontic models framework is more

constrained than this. It is troubled both by the issue of complexity, but is also constrained by its assumption of realism. This means that it cannot be used to progress to a deeper description of reality: an assumption of the model is that it is impossible to create a description of its variables.

Given these considerations as a background we will look at the notion of contextuality in section 3.1.1, which we discussed originally in the context of ontic models. Using the other subjects as a reference we examine whether the notion of contextuality can have physical meaning.

3.1.1 Physical Contextuality

The concept of contextuality introduced in section 2.2 is used to describe a property of a theory, not a property of reality itself. We will examine the consequences of relaxing this fact while attempting to construct a consistent description of reality.

For reality to be contextual, there would have to exist situations where any consistent description would exhibit contextuality. We use this requirement to distinguish physical contextuality from a contextual description of non-contextual reality. We also require such situations to not be those which are contextual simply due to a poor choice of the division between observer and observed.

The second requirement of physical contextuality is not a trivial escape which would render the definition worthless. It is possible that such a better choice of division may not exist. Consider for an illustration the problem of the observer in quantum general relativity[6] here the object being observed is the entirety of the universe, and so there is

no way to separate the observer from the observed. There is a contradiction here though: in this picture there is no notion of repeatability or guarantee of relationship between different measurements. We then have a complex question: could we observe something which is actually physically contextual? For the sake of illustration this would mean that there is some system which we could take multiple choices of simultaneous measurements (say A, B, C or A, B', C') and the choice of which measurements are taken influences the results of the measurements. The complication here is that we've presupposed that we know that we can simultaneously take these measurements. In quantum mechanics we have a definition of what it means for observables to be simultaneously measurable: they must commute. How do we know that two measurements should in fact be comeasurable? The reality is that we cannot. Given any scenario where we believe two sets of measurements with a common element should be co-measurable and find that the result is contextual, we have to accept the alternative possibility that there is in fact some physical phenomena which prevents the measurement sets from being comeasurable. This gives us another possibility: any observed physical contextuality could instead be replaced by a contextual theory with different physical phenomena.

We should then ask ourself what types of phenomena could lead to apparent physical contextuality. To answer this we must ask ourselves what conditions would make us believe that two things might not have influence upon one another. The most obvious answer to this question is distance: if two objects are sufficiently distant we would think that all forces between them should fall below levels at which they could influence one another. Any phenomenon then which acted irrespective of distance would then potentially be confused with physical contextuality.

In a situation where we have non-local phenomena giving rise to apparent physical contextuality, we are faced with a greater hurdle: we cannot repair such an issue by enlarging our observed system as there is no guarantee that any such enlargement would dilute the influence of the non-local effects. Here instead we are left with two options.

The first option is to move to a framework which describes reality not in the terms which we are used to, but instead moves to a description in terms of what parts of the universe can have influence on others. The causaloid framework (see section 1.3) is an example of a direct attempt at such a description, and taken in a broader context so is the ontic models framework (see section 1.2) which never makes reference to a space (or space-time) description, replacing it instead with an ontic space of be-ables.

The alternative option is to accept that reality is making it clear that one cannot probe to deeper levels. We must then accept either an inherent connectedness between all things - which gives rise to a hidden variable theory like Bohmian mechanics (see section 2.4.1) with its all-knowing quantum potential - or an inherent randomness - which gives rise to a stochastic theory like Nelson's mechanics (see section 1.1). It is interesting here to recall that Nelson's mechanics and Bohmian mechanics share a common equilibrium description of reality, meaning that these two options may have more commonality than assumed.

These conclusions leave us with interesting questions concerning the value of research into contextuality. If any physical meaning of contextuality can be replaced with another theory, is there any possible result that can emerge from its study?

3.2 Conclusions

In this thesis I have presented a broad perspective of quantum foundations in the attempt to show both the benefits and the pitfalls of the existence of a large variety of approaches to the subject. The breadth of the approaches to quantum foundations certainly has allowed for many successes, but the vast difference in these approaches means that many of these results cannot be applied to other approaches.

In chapter 1 I outlined three different frameworks that fall under the general heading of quantum foundations: ontic models, Nelson's mechanics, and the causaloid. Then in chapter 2 I discussed applications of these frameworks in studying entropy, contextuality and in understanding other approaches to quantum foundations. Lastly, in section 3.1.1 I presented an argument - rooted in the broader view of having examined multiple approaches to quantum foundations - that the property of contextuality may not have any physical meaning.

The study of quantum foundations has advanced significantly from its early origins, but I believe that until the field begins to learn what results are transferrable from different frameworks and which are simply artifacts of a particular construction, it will fail to live up to its promises.

References

[1] J. Ambjorn, J. Jurkiewicz, and R. Loll. Quantum Gravity, or The Art of Building Spacetime. 2006. Contribution to the book 'Approaches to Quantum Gravity', edited by D. Oriti, to appear at Cambridge Univ. Press.

[2] Guido Bacciagaluppi. Nelsonian mechanics revisited. *Found.Phys.Lett.*, 12:1–16, 1999.

[3] J.S. Bell. SPEAKABLE AND UNSPEAKABLE IN QUANTUM MECHANICS. COLLECTED PAPERS ON QUANTUM PHILOSOPHY. 1987.

[4] D. Bohm. A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I. *Physical Review*, 85:166–179, January 1952.

[5] D. Bohm. A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. II. *Physical Review*, 85:180–193, January 1952.

[6] Bryce S. DeWitt. Quantum theory of gravity. i. the canonical theory. *Phys. Rev.*, 160(5):1113–1148, Aug 1967.

[7] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum mechanical description of physical reality be considered complete? *Phys.Rev.*, 47:777–780, 1935.

[8] Hugh Everett. Relative state formulation of quantum mechanics. *Rev.Mod.Phys.*, 29:454–462, 1957.

[9] Lucien Hardy. Probability theories with dynamic causal structure: A New framework for quantum gravity. 2005.

[10] Lucien Hardy. Towards quantum gravity: A Framework for probabilistic theories with non-fixed causal structure. *J.Phys.A*, A40:3081, 2007. In honor of Giancarlo Ghirardi.

[11] Lucien Hardy. Formalism Locality in Quantum Theory and Quantum Gravity. 2008.

[12] N. Harrigan and T. Rudolph. Ontological models and the interpretation of contextuality. *ArXiv e-prints*, September 2007.

[13] S.W. Hawking. Quantum Gravity and Path Integrals. *Phys.Rev.*, D18:1747–1753, 1978.

[14] Ted Jacobson. Black hole entropy and induced gravity. 1994.

[15] E. T. Jaynes. Information Theory and Statistical Mechanics. *Phys. Rev.*, 106:620–630, 1957.

[16] Simon Kochen and E. Specker. The Problem of Hidden Variables in Quantum Mechanics. *J.Math.Mech.*, 17:59–87, 1968.

[17] L. D. Landau and E. M. Lifshitz. *Statistical Physics, Third Edition, Part 1: Volume 5 (Course of Theoretical Physics, Volume 5).* Butterworth-Heinemann, 3 edition, 1980.

[18] Sonia Markes and Lucien Hardy. Entropy for theories with indefinite causal structure. 2009.

[19] O. J. E. Maroney. Information and Entropy in Quantum Theory. *ArXiv Quantum Physics e-prints*, November 2004.

[20] Ryan Morris. Topics in Quantum Foundations: Ontological Models, and Distinguishability as a Resource. Master's thesis, University of Waterloo, 2009.

[21] Edward Nelson. The adjoint markoff process. *Duke Mathematical Journal*, 25:671–690, 1958.

[22] Edward Nelson. Derivation of the Schrodinger equation from Newtonian mechanics. *Phys.Rev.*, 150:1079–1085, 1966.

[23] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[24] G. N. Ord and R. B. Mann. Entwined pairs and Schrödinger's equation. *Annals of Physics*, 308:478–492, December 2003.

[25] G. N. Ord and R. B. Mann. Entwined paths, difference equations, and the Dirac equation. *pra*, 67(2), February 2003.

[26] Philip M. Pearle and Antony Valentini. Generalizations of quantum mechanics. 2005. To be published in: Encyclopedia of Mathematical Physics, eds. J.-P. Francoise, G. Naber and T.S. Tsun (Elsevier, 2006).

[27] R. Penrose. On gravity's role in quantum state reduction. *General Relativity and Gravitation*, 28(5):581–600, 1996.

[28] R. Penrose and Malcolm A.H. MacCallum. Twistor theory: An Approach to the quantization of fields and space-time. *Phys.Rept.*, 6:241–316, 1972.

[29] Carlo Rovelli. 'Incerto tempore, incertisque loci': Can we compute the exact time at which the quantum measurement happens? *Found.Phys.*, 28:1031–1043, 1998.

[30] Carlo Rovelli. Quantum gravity. 2004. Published in Cambridge Monographs on Mathematical Physics, pages 1-480, year 2004.

[31] Carlo Rovelli and Lee Smolin. Loop Space Representation of Quantum General Relativity. *Nucl.Phys.*, B331:80, 1990.

[32] Lee Smolin. Could quantum mechanics be an approximation to another theory? 2006.

[33] Rafael D. Sorkin. ON THE ROLE OF TIME IN THE SUM OVER HISTORIES FRAMEWORK FOR GRAVITY. *Int.J.Theor.Phys.*, 33:523–534, 1994.

[34] R. W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *pra*, 71(5):052108–+, May 2005.

[35] Thomas Thiemann. Modern canonical quantum general relativity. 2001.

[36] J. von Neumann. *Mathematical foundations of quantum mechanics*. 1955.

[37] T. C. Wallstrom. On the derivation of the Schrödinger equation from stochastic mechanics. *Foundations of Physics Letters*, 2:113–126, March 1989.

[38] T. C. Wallstrom. Inequivalence between the Schrödinger equation and the Madelung hydrodynamic equations. *pra*, 49:1613–1617, March 1994.