Locating Oneself in a Quantum World

by

Charles Thorne Sebens

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in the University of Michigan
2015

Doctoral Committee:

Associate Professor David J. Baker, Co-Chair
Professor Laura Ruetsche, Co-Chair
Professor Ratindranath Akhoury
Professor Gordon Belot

http://henrygale.deviantart.com/art/Blasenkammer-52418546

2015

Preface

What is matter made of and what rules govern its motion? It appears that everything—boulders, sea water, human brains, distant stars—is made of the same building blocks obeying a single set of laws. However, as of now we know neither the building blocks nor the laws. We are at a stage in physics where we have no theory which fits all of the data, no theory which combines our best understanding of gravity with quantum mechanics. I believe a key reason why such a unification has been difficult to achieve is that there is wild disagreement about the foundations of quantum mechanics, disagreement about how it answers the core question above. There are a variety of mathematically and physically distinct theories that have been proposed to explain why the methods of quantum mechanics are so successful. These are often called "interpretations" of quantum mechanics, but I prefer to think of them as distinct physical theories. I focus on four that I believe to be particularly promising: the many-worlds interpretation, Bohmian mechanics, Newtonian quantum mechanics, and GRW theory. If these theories are all as successful as their proponents take them to be, then quantum mechanics presents us with a fascinating case of underdetermination of theory by data; there would be multiple mathematically and physically distinct theories that successfully reproduce the empirical predictions of quantum mechanics. As it turns out, enumerating the empirically successful theories is not a straightforward task. This dissertation uses the tools of formal epistemology—prominently, the concept of self-locating uncertainty—in addition to those of physics and philosophy of science to investigate which of the theories that have been put forward really are empirically adequate and what alternatives can be

devised.

I begin by proposing and defending a new theory, Newtonian QM, which combines aspects of two well-developed alternatives: Bohmian mechanics and the many-worlds interpretation. In the second chapter Sean Carroll and I examine how probability arises in the many-worlds interpretation and argues that the theory's probabilistic predictions are correct. In the third chapter I explore when exactly GRW theory is capable of recovering the predictions of textbook quantum mechanics. The final chapter is an original introduction to relativistic quantum field theory which lays groundwork for evaluating the prospects for extending interpretations of quantum mechanics to this more sophisticated theory. In Newton's physics, to determine how a body will move one simply needs to add up the various forces acting on it: gravitational, electric, magnetic, etc. This framework is generally taken to be inadequate for explaining the quantum behavior of subatomic particles like electrons and protons. We are told to revise our classical picture of the world in favor of a quantum one. In the first chapter of the dissertation I argue that if we can stomach the existence of parallel worlds, distinct from our own but no less real, we can account for quantum phenomena without overthrowing Newton's mechanics; without (at the fundamental level) introducing wave functions, Schrödinger's equation, superpositions, entanglement, etc. In addition to the normal forces, a quantum force is introduced which explains why particles don't follow classical trajectories (in fact, they end up following essentially the same trajectories as particles in Bohmian mechanics). This approach is arrived at by taking seriously the ways in which quantum mechanics resembles hydrodynamics. In quantum mechanics electrons and photons behave sometimes like particles and sometimes like waves. According to this new theory, Newtonian quantum mechanics, electrons and photons are fundamentally particles. The arcane waves of quantum mechanics are taken to be made of particles, like ordinary water or sound waves.

The universal wave function—a fundamental entity in both Bohmian mechanics and the many-worlds interpretation—can be represented as a field on configuration space, the space of possible ways particles can be arranged. According to Bohmian mechanics a single point in this space is special; it represents the way all of the particles actually are arranged. Newtonian quantum mechanics takes many of the points in configuration space to be special; for each, there is a world in which particles are arranged that way. There will always be multiple worlds so similar that an agent cannot tell directly from experience which they are in—self-locating uncertainty is unavoidable. In Newtonian quantum mechanics, assigning equal probability to each of these worlds immediately yields the desired quantum probabilities.

The next chapter—coauthored with Sean Carroll—defends the many-worlds interpretation. In this version of quantum mechanics, processes like quantum measurements cause agents to split into multiple copies. There are good reasons to reject the simple strategy of treating each copy as equiprobable, which is fortunate as in this case doing so would yield probabilities at odds with experimental data. We introduce an epistemic principle demarcating what facts about the world one's credences might reasonably depend on—only facts about what's happening around here—and argue from this principle that the many-worlds interpretation recovers the correct quantum probabilities. This principle explains why treating multiple copies of oneself as equiprobable is correct in some cases of classical self-locating uncertainty but not in cases of self-locating uncertainty arising from quantum measurements (and allows us to generate probabilities for cases involving both kinds of self-locating uncertainty).

In the third chapter I focus on a fourth option in precisifying quantum mechanics: Ghirardi-Rimini-Weber theory (GRW). According to GRW the normal time evolution—as described by the Schrödinger equation—is interrupted on occasion by a process known as "collapse of the wave function." It is well-understood why the

rate at which these collapses occur cannot be too high: the theory makes testable and incorrect predictions (for example, that macroscopic superpositions are unavoidably unstable). But could the theory ever be falsified or could one just make the collapse rate arbitrarily small? As the rate goes to zero, GRW becomes more and more like the many-worlds interpretation where collapse never occurs and the evolution of the wave function is always governed by the Schrödinger equation. I argue that when the rate is very small, parallel universes have time to form and one should be initially uncertain which they are in. Most of these universes are eventually destroyed by collapse events. Our continued survival provides empirical evidence against these variants of GRW, generating a lower bound on the collapse rate and thereby a way of potentially falsifying the theory.

In each of the first three chapters, there is a key point at which agents are forced to ask where in the quantum multiverse they are located. They must try to locate themselves in a quantum world. The papers that compose this dissertation are each distinct and can be read in isolation. Still, they are related to one another in complex ways. One theme that emerges from reading the first three together is that understanding the degree to an agent should expect to find themselves in a particular world—and how the agent's expectations should change when worlds are altered, created, or destroyed—is crucial to understanding which versions of quantum mechanics are supported by the data. Debates about the foundations of quantum mechanics have often centered on the easy case: non-relativistic quantum mechanics. Some proposals which are very promising in this context—like Bohmian mechanics and GRW theory—may not extend to the harder case of relativistic quantum field theory. I am interested in understanding and overcoming the barriers to such extensions. Uniting our best theory of particle physics, relativistic quantum field theory, with our best theory of gravity, general relativity, is one of the central goals of contemporary theoretical physics. I believe progress has been impeded by the

obscurity of relativistic quantum field theory. In my opinion, this is largely due to the fact that there are multiple very different proposals about how to understand non-relativistic quantum mechanics. In the future I would like to examine which "interpretations" of quantum mechanics can be extended to relativistic quantum field theory. Chapter four of the dissertation is the beginning of this endeavor.

In the final chapter of the dissertation, relativistic quantum field theory is presented in an atypical way which makes it especially clear what changes are involved in moving from non-relativistic quantum mechanics or classical field theory to relativistic quantum field theory. The move from non-relativistic quantum mechanics to relativistic quantum field theory is made in two steps: first, the state space is expanded so that the total number of particles can change with time; second, the non-relativistic expression for the kinetic energy in the Schrödinger equation is replaced with a relativistic expression. These two relatively simple changes suffice to move from non-relativistic quantum mechanics to relativistic quantum field theory. However, in this story it is unclear what quantum field theory has to do with fields. To clarify the connection, quantum field theory is next arrived at via a second path, from classical relativistic field theory: wave functionals over the space of possible field configurations are introduced to allow fields to enter superpositions. It is then shown that there is a sense in which both paths yield the same theory.

At times, one might get the feeling that the theories and scenarios under consideration in this dissertation are too wild to be relevant to the scientific question of what laws govern our universe. Unfortunately, I fear this is unavoidable. Quantum mechanics requires us to rethink some elements of our classical picture of the world. Exactly what elements must be rethought is a difficult question. There appear to be multiple options for jettisoning this and keeping that, none of which are, in the end, wholly intuitive and familiar. By being non-committal and imprecise, one may be able to give the impression that quantum mechanics can be understood without

much weirdness. As far as we know, it cannot. Here I try to tackle the strangeness head on, looking at physical theories which—although each odd in its own way—are at least sincere attempts to avoid vagueness, obfuscation, and imprecision; attempts to describe what electrons might be doing when no one is looking.

Table of Contents

List of Figures

Dissertation Abstract

There appear to be multiple mathematically and physically distinct theories that successfully reproduce the empirical predictions of quantum mechanics, so-called "interpretations" of quantum mechanics. This dissertation uses the tools of formal epistemology (prominently, the concept of self-locating uncertainty) to investigate which of the theories that have been put forward really are empirically adequate and what alternatives can be devised. The first chapter introduces a novel theory that incorporates aspects of two well-developed alternatives, Bohmian mechanics and the many-worlds interpretation. The quantum wave function can be represented as a field on configuration space, the space of possible ways particles can be arranged. According to Bohmian mechanics a single point in this space is special; it represents the way all of the particles actually are arranged. The newly introduced theory holds that many of the points in configuration space are special; for each, there is a world in which particles are arranged that way. In general, there will be multiple worlds that are so similar that an agent cannot tell directly from experience which they are in; self-locating uncertainty is unavoidable. The next chapter argues for the empirical adequacy of the many-worlds interpretation. In this version of quantum mechanics, processes like quantum measurements cause agents to split into multiple copies and enter periods of self-locating uncertainty. An epistemic principle demarcating which facts about the world one's credences might reasonably depend on (only facts about what's happening around here) is used to derive probabilistic predictions from the many-worlds interpretation. In the third chapter self-locating uncertainty is employed in evaluating another version of quantum mechanics, Ghirardi-Rimini-Weber theory.

In this theory, random collapse events prevent our world from splitting into many. If the collapse events are rare, this prevention fails. Other worlds have time to form but most are short-lived. Our survival provides evidence against that kind of theory. The final chapter is an original introduction to relativistic quantum field theory which lays groundwork for evaluating the prospects of extending various interpretations of quantum mechanics to this more sophisticated theory.

Chapter 1

Quantum Mechanics as Classical Physics

Abstract: Here I explore a novel no-collapse interpretation of quantum mechanics which combines aspects of two familiar and well-developed alternatives, Bohmian mechanics and the many-worlds interpretation. Despite reproducing the empirical predictions of quantum mechanics, the theory looks surprisingly classical. All there is at the fundamental level are particles interacting via Newtonian forces. There is no wave function. However, there are many worlds. [The published version of this chapter appears in *Philosophy of Science* 82 (2015).]

1.1 Introduction

On the face of it, quantum physics is nothing like classical physics. Despite its oddity, work in the foundations of quantum theory has provided some palatable ways of understanding this strange quantum realm. Most of our best theories take that story to include the existence of a very non-classical entity: the wave function. Here I offer an alternative which combines elements of Bohmian mechanics and the many-worlds interpretation to form a theory in which there is no wave function. According to this theory, all there is at the fundamental level are particles interacting via Newtonian forces. In this sense, the theory is classical. However, it is still undeniably strange as it posits the existence of a large but finite collection of worlds, each completely and utterly real. When an experiment is conducted, every result with appreciable Born

Rule probability does actually occur in one of these worlds. Unlike the many worlds of the many-worlds interpretation, these worlds are fundamental, not emergent; they are interacting, not causally isolated; and they never branch. In each of these worlds, particles follow well-defined trajectories and move as if they were being guided by a wave function in the familiar Bohmian way.

In this chapter I will not attempt to argue that this theory is unequivocally superior to its competitors. Instead, I would like to establish it as a surprisingly successful alternative which deserves attention and development, hopefully one day meriting inclusion among the list of promising realist responses to the measurement problem.

In §1.2, I briefly review why quantum mechanics is in need of a more precise formulation and discuss two no-collapse theories: the many-worlds interpretation and Bohmian mechanics. I then go on to offer a rather unlikable variant of Bohmian mechanics which adds to the standard story a multitude of worlds all guided by the same wave function. This theory is useful as a stepping stone on the way to Newtonian QM. Newtonian QM is then introduced. As soon as Newtonian QM is on the table, §1.5 & 1.6 present one of the most significant costs associated with the theory: the space of states must be restricted if the theory is to recover the experimental predictions of quantum mechanics. In §1.7, 1.8, & 1.9, I discuss the advantages of this new theory over Everettian and Bohmian quantum mechanics in explaining the connection between the squared amplitude of the wave function and probability. In §1.10, I consider the possibility of modifying the theory so that it describes a continuous infinity of worlds instead of a finite collection, concluding that such a modification would be inadvisable. In §1.11, I propose two options for the fundamental ontology of Newtonian QM. In §1.12, I use Newtonian QM to explain the way the wave function transforms under time reversal and Galilean boosts. Spin is then discussed in §1.13.

Some limitations of the theory presented here are worth stating up front. First, just as hydrodynamics relies on approximating a discrete collection of particles as a continuum, in its current form this theory must treat the discrete collection of worlds as a continuum. As this is merely an approximation, empirical equivalence with standard quantum mechanics is likely only approximate (§1.5). Second, one must impose a significant restriction on the space of states if the predictions of QM are to be reproduced (the Quantization Condition, §1.6). Third, I will not discuss extending the theory to handle multiple particles with spin or relativistic quantum physics.

Newtonian QM is a realist version of quantum mechanics based on the theory's hydrodynamic formulation (originally due to Madelung, 1927). For recent and relevant discussions of quantum hydrodynamics, see Wyatt (2005); Holland (2005). An approach much like Newtonian QM was independently arrived at by Hall *et al.* (2014). Newtonian QM is somewhat similar to Böstrom's (2012) metaworld theory[1] and the proposal in Tipler (2006). Related ideas about how to remove the wave function are explored in Poirier (2010); Schiff & Poirier (2012), including a suggestion of many worlds.

To avoid confusion, throughout the chapter I'll use "universe" to denote the entirety of reality, what philosophers call "the actual world" and what in these contexts is sometimes called the "multiverse," reserving "world" for the many worlds of quantum mechanics.

[1]The key difference with Newtonian QM being that Böstrom's theory does not as thoroughly excise the wave function (the dynamics being given by (1.1) not (1.16)).

1.2 The Measurement Problem

If the state of the universe is given by a wave function and that wave function always evolves in accordance with the Schrödinger equation, then quantum measurements will typically not have single definite outcomes. Actual measurements of quantum systems performed in physics laboratories do seem to yield just one result. This, in brief, is the measurement problem. There are various ways of responding.

According to Everettian quantum mechanics, a.k.a. the many-worlds interpretation, the wave function Ψ is all there is. The evolution of the wave function is always given by the Schrödinger equation,

$$i\hbar\frac{\partial}{\partial t}\Psi(\vec{x}_1, \vec{x}_2, ..., t) = \left(\sum_k \frac{-\hbar^2}{2m_k}\nabla_k^2 + V(\vec{x}_1, \vec{x}_2, ..., t)\right)\Psi(\vec{x}_1, \vec{x}_2, ..., t)\ , \qquad (1.1)$$

where Ψ is a function of particle configuration $(\vec{x}_1, \vec{x}_2, ...)$ and time t, m_k is the mass of particle k, ∇_k^2 is the Laplacian with respect to $\vec{x}_k$, and V is the classical potential energy of particle configuration $(\vec{x}_1, \vec{x}_2, ...)$ at t. When an observer performs a quantum measurement, the universal wave function enters a superposition of the observer seeing each possible outcome. This is not to be understood as one observer seeing many outcomes, but as many observers each seeing a single outcome. Thus, the theory is not obviously inconsistent with our experience of measurements appearing to have unique outcomes. According to Everettian quantum mechanics, there is nothing more than the wave function and therefore things like humans, measuring devices, and cats must be understood as being somehow composed of or arising out of wave function. (Wallace, 2003, 2012 takes these things to be patterns or structures in the universal wave function.) To summarize, here is what the Everettian QM says that

there is (the ontology) and how it evolves in time (the dynamical laws).

Ontology: (I) universal wave function $\Psi(\vec{x}_1, \vec{x}_2, ..., t)$

Law: (I) Schrödinger equation (1.1)

A second option in responding to the measurement problem is to expand the ontology so that the universe contains both a wave function evolving according to (1.1) *and* particles with definite locations. The time-dependent position of particle k can be written as $\vec{x}_k(t)$ and its velocity as $\vec{v}_k(t)$. The wave function pushes particles around by a specified law,

$$\vec{v}_k(t) = \frac{\hbar}{m_k} \mathrm{Im} \left[\frac{\vec{\nabla}_k \Psi(\vec{x}_1, \vec{x}_2, ..., t)}{\Psi(\vec{x}_1, \vec{x}_2, ..., t)} \right] . \tag{1.2}$$

Experiments are guaranteed to have unique outcomes because humans and their scientific instruments are made of particles (not wave function). These particles follow well-defined trajectories and are never in two places at once. This theory is Bohmian mechanics, a.k.a. de Broglie-Bohm pilot wave theory.

Ontology: (I) universal wave function $\Psi(\vec{x}_1, \vec{x}_2, ..., t)$

(II) particles with positions $\vec{x}_k(t)$ and velocities $\vec{v}_k(t)$

Laws: (I) Schrödinger equation (1.1)

(II) guidance equation (1.2)

From (1.1) and (1.2), one can derive an expression for the acceleration of each particle,

$$m_j \vec{a}_j(t) = -\vec{\nabla}_j \Big[Q(\vec{x}_1, \vec{x}_2, ..., t) + V(\vec{x}_1, \vec{x}_2, ..., t) \Big] , \tag{1.3}$$

where $Q(\vec{x}_1, \vec{x}_2, ..., t)$ is the quantum potential, defined by

$$Q(\vec{x}_1, \vec{x}_2, ..., t) = \sum_k \frac{-\hbar^2}{2m_k} \left(\frac{\nabla_k^2 |\Psi(\vec{x}_1, \vec{x}_2, ..., t)|}{|\Psi(\vec{x}_1, \vec{x}_2, ..., t)|} \right) . \tag{1.4}$$

Since the focus of this chapter is not on Everettian or Bohmian quantum mechanics, I've sought to present each as simply as possible. The best way to formulate each theory—ontology and laws—is a matter of current debate.

1.3 Prodigal QM

As a precursor to the theory I'll propose, consider the following interpretation of quantum mechanics which has both a many-worlds and a Bohmian flavor. The wave function always obeys the Schrödinger equation. There are many different worlds, although a finite number, each represented by a point in configuration space[2]. There are more worlds where $|\Psi|^2$ is large and less where it is small. Each world is guided by the single universal wave function in accordance with the Bohmian guidance equation and thus each world follows a Bohmian trajectory through configuration space. Let's call this ontologically extravagant theory *Prodigal QM*.[3] Why include a multitude of worlds when we only ever observe one, our own? We could simplify the theory by removing all of the worlds but one, arriving at Bohmian mechanics (Valentini, 2010, §7). But, less obviously, it turns out that there is another route to simplification: keep the multitude of worlds but remove the wave function. This option will be explored in the next section.

According to Prodigal QM, the universe contains a wave function $\Psi(\vec{x}_1, \vec{x}_2, ..., t)$ on configuration space and a large number of worlds which can be represented as points moving around in configuration space. The arrangement of the worlds in configuration space is described by a number density, $\rho(\vec{x}_1, \vec{x}_2, ..., t)$, normalized

[2]The location of a single particle is given by a point in space, $(\vec{x})$. The locations of all particles are given by a point in *configuration space*, $(\vec{x}_1, \vec{x}_2, ...)$, where $\vec{x}_i$ is the location of particle i.

[3]With a continuous infinity of worlds, Prodigal QM is mentioned in Valentini (2010, §7) and in Barrett (1999) (in Barrett's terminology, it is a Bohmian many-threads theory in which all of the threads are taken to be completely real); a closely related proposal is discussed in Dorr (2009).

so that integrating ρ over all of configuration space gives one, $\int d^3x_1 d^3x_2 ... \ \rho = 1$. Integrating $\rho(\vec{x}_1, \vec{x}_2, ..., t)$ over a not-too-small volume of configuration space gives the proportion of all of the worlds that happen to be in that volume at t. By hypothesis, worlds are initially distributed so that

$$\rho(\vec{x}_1, \vec{x}_2, ..., t) = |\Psi(\vec{x}_1, \vec{x}_2, ..., t)|^2 \ . \tag{1.5}$$

The velocities of the particles are described by a collection of velocity fields indexed by particle number, k,

$$\vec{v}_k(\vec{x}_1, \vec{x}_2, ..., t) = \frac{\hbar}{m_k} \mathrm{Im} \left[\frac{\vec{\nabla}_k \Psi(\vec{x}_1, \vec{x}_2, ..., t)}{\Psi(\vec{x}_1, \vec{x}_2, ..., t)} \right] \ , \tag{1.6}$$

In Prodigal QM, if there is a world at $(\vec{x}_1, \vec{x}_2, ...)$ at t the velocity of the kth particle in that world is $\vec{v}_k(\vec{x}_1, \vec{x}_2, ..., t)$.[4] With these velocity fields, the equivariance property of the Bohmian guidance equation (1.2) ensures that ρ is always equal to $|\Psi|^2$ if it ever is (see Dürr *et al.* , 1992, §3).

Ontology: (I) universal wave function $\Psi(\vec{x}_1, \vec{x}_2, ..., t)$

(II) particles in many worlds described by a world density $\rho(\vec{x}_1, \vec{x}_2, ..., t)$ and velocity fields $\vec{v}_k(\vec{x}_1, \vec{x}_2, ..., t)$

Laws: (I) Schrödinger equation (1.1)

(II) guidance equation (1.6)[5]

The use of densities and velocity fields is familiar from fluid dynamics. A quick review will be helpful. Consider a fluid composed of N point particles which each

[4] This is not true for Newtonian QM (see §1.5).

[5] Actually, the second dynamical law is more specific than (1.6) since it requires not just that the velocity *fields* obey (1.6) but that each world follows an *exact* Bohmian trajectory (see §1.5). The connection between ρ and Ψ in (1.5), though not a *dynamical* law, might best be thought of as a third law of Prodigal QM.

have mass m. The number density of these particles is $n(\vec{x},t)$, normalized so that $\int d^3x_1 d^3x_2 ...\ n = N$. The mass density is $m \times n(\vec{x},t)$. Integrating $n(\vec{x},t)$ over a not-too-small volume gives the number of particles in that volume at t. Whereas $n(\vec{x},t)$ gives the density of *particles* in *three*-dimensional space, ρ gives the density of *worlds* in *configuration* space. The velocity field for the fluid is $\vec{u}(\vec{x},t)$, defined as the mean velocity of particles near $\vec{x}$ at t.[6] For an inviscid compressible fluid with zero vorticity, the time evolution of n and $\vec{u}$ are determined by a continuity equation

$$\frac{\partial n(\vec{x},t)}{\partial t} = -\vec{\nabla} \cdot \Big(n\,(\vec{x},t)\ \vec{u}\,(\vec{x},t) \Big)\ , \tag{1.7}$$

and a Newtonian force law

$$m\,\vec{a}\,(\vec{x},t) = -\vec{\nabla} \left[\frac{p(\vec{x},t)}{n(\vec{x},t)} + V\,(\vec{x},t) \right]\ , \tag{1.8}$$

where V is the external potential, p is the pressure, and

$$\vec{a}\,(\vec{x},t) = \frac{D\,\vec{u}\,(\vec{x},t)}{Dt} = \Big(\vec{u}(\vec{x},t) \cdot \vec{\nabla} \Big)\ \vec{u}\,(\vec{x},t) + \frac{\partial\,\vec{u}\,(\vec{x},t)}{\partial t}\ . \tag{1.9}$$

The acceleration is given by the material derivative of $\vec{u}$ not the partial derivative because a particle's position in the fluid is time dependent.

The three quantum theories on the table thus far are applied to the double-slit experiment in figure 1.1. In the bottom-right diagram is Everettian QM where the

[6] More precisely, the number density and velocity field provide a good description of the particle trajectories if to a good approximation: $n(\vec{x},t)$ gives the average number of particles in a small-but-not-too-small region $\mathcal{R}$ centered about $\vec{x}$ over a short-but-not-too-short period of time $\mathcal{T}$ around t divided by the volume of $\mathcal{R}$, and $\vec{u}(\vec{x},t)$ gives the average velocities of the particles in $\mathcal{R}$ over $\mathcal{T}$. For more detail, see Chapman & Cowling (1970, §2.2). The connection between ρ and the $\vec{v}_k$s and the trajectories of individual worlds could be spelled out along similar lines, but full rigor in the context of Newtonian QM would require a better understanding of the dynamics (see §1.5 and Hall *et al.* , 2014).

universe is just a wave function. The particle's wave function is initially peaked at the two slits and then spreads out and interferes as time progresses. When the particle hits the detector, a multitude of worlds will separate via decoherence and in each the particle will be observed hitting at a particular point on the screen. In Bohmian mechanics, one adds to the wave function an actual particle which follows a definite trajectory in accordance with the guidance equation. In Prodigal QM, there is a wave function *and* a collection of worlds, each of which contains a particle following a Bohmian trajectory. In Newtonian QM, which will be introduced at the end of §1.4, one retains the multitude of worlds but removes the wave function.

1.4 Removing the Wave Function

One can derive an equation for the dynamics of particles in Prodigal QM that makes no reference to the wave function. Once this is done, we can formulate an alternate theory where the superfluous wave function has been removed. This new theory, Newtonian QM, will be the focus of the remainder of the chapter. The mathematical manipulations presented in this section are familiar from discussions of Bohmian mechanics, but take on a different meaning as derivations of particle dynamics in Prodigal QM. Those who wish to skip the derivation should simply note that (1.16) is derivable from (1.1), (1.5), and (1.6).

As $\rho = |\Psi|^2$ (1.5), the wave function can be written in terms of the world-density and a phase factor as

$$\Psi(\vec{x}_1, \vec{x}_2, ..., t) = \sqrt{\rho(\vec{x}_1, \vec{x}_2, ..., t)} e^{i\theta(\vec{x}_1, \vec{x}_2, ..., t)} \ . \tag{1.10}$$

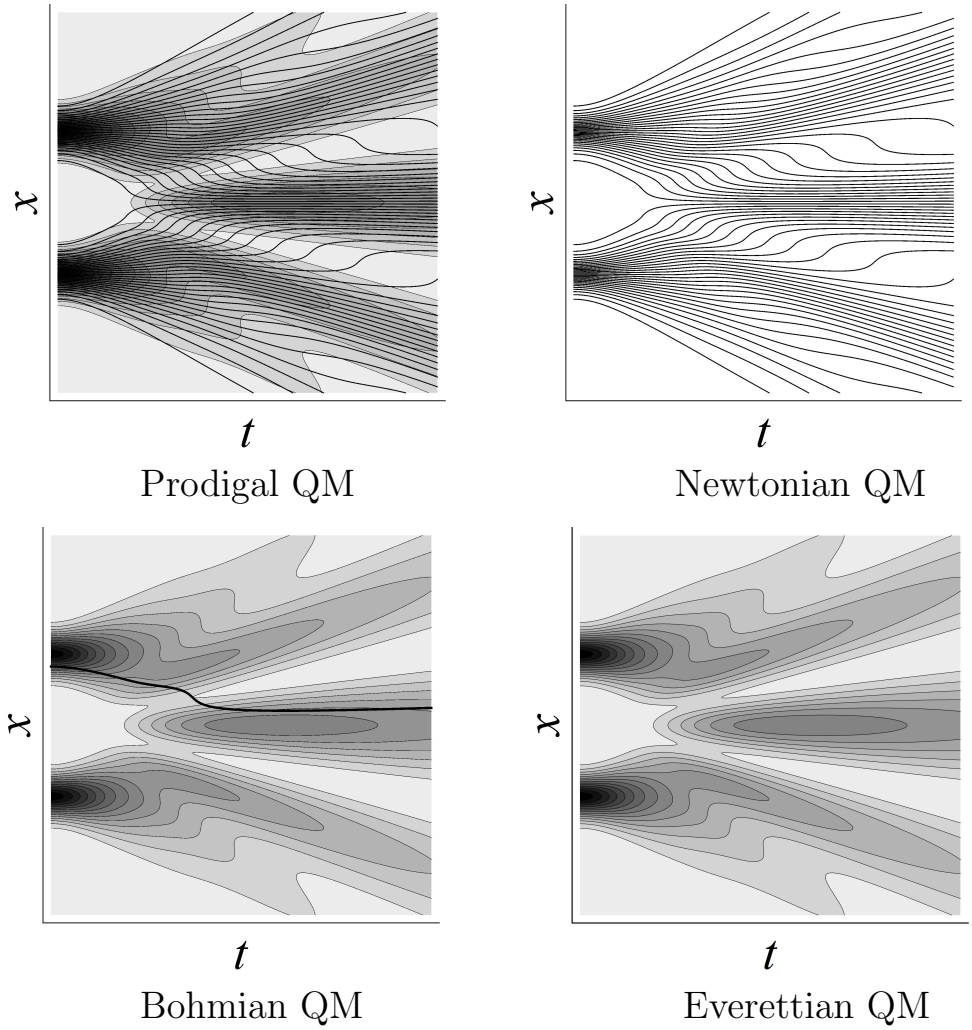

Figure 1.1: **Four Quantum Theories** Diagrams of the evolution of a single particle in the double-slit experiment according to four different no-collapse theories. The vertical axis gives the position of the single particle and the horizontal axis time. $|\Psi|^2$ is shown as a contour plot and particle trajectories as lines.

Plugging (1.10) into the guidance equation (1.6) generates

$$\vec{v}_k(\vec{x}_1, \vec{x}_2, ..., t) = \frac{\hbar}{m_k}\vec{\nabla}_k\theta(\vec{x}_1, \vec{x}_2, ..., t) \ , \tag{1.11}$$

relating $\vec{v}_k$ and θ. (At this point, I will stop repeating the arguments of Ψ, ρ, θ, and

$\vec{v}_k$; they *all* depend on the configuration of particles and the time.)

The evolution of the wave function Ψ is given by the Schrödinger equation (1.1). Dividing both sides of (1.1) by Ψ and using (1.10), one can derive that

$$\frac{i\hbar}{2\rho}\frac{\partial\rho}{\partial t} - \hbar\frac{\partial\theta}{\partial t} = \sum_k \frac{-\hbar^2}{2m_k}\left[\frac{\nabla_k^2\sqrt{\rho}}{\sqrt{\rho}} + \frac{2i}{\sqrt{\rho}}\left(\vec{\nabla}_k\sqrt{\rho}\right)\cdot\left(\vec{\nabla}_k\theta\right) + i\,\nabla_k^2\theta - \left|\vec{\nabla}_k\theta\right|^2\right] + V \ . \tag{1.12}$$

Equating the imaginary parts, using (1.11), yields

$$\frac{\partial\rho}{\partial t} = -\sum_k \vec{\nabla}_k\cdot(\rho\,\vec{v}_k) \ , \tag{1.13}$$

a continuity equation similar to (1.7). Equating the *real* parts of (1.12), using (1.11), yields

$$\frac{\partial\theta}{\partial t} = \sum_k\left\{\frac{\hbar}{2m_k}\frac{\nabla_k^2\sqrt{\rho}}{\sqrt{\rho}} - \frac{m_k}{2\hbar}|\vec{v}_k|^2\right\} - \frac{V}{\hbar} \ . \tag{1.14}$$

Acting with $\frac{\hbar}{m_j}\vec{\nabla}_j$ on both sides of (1.14) and rearranging, making use of (1.11) and the fact that

$$\vec{a}_j = \sum_k(\vec{v}_k\cdot\vec{\nabla}_k)\vec{v}_j + \frac{\partial\vec{v}_j}{\partial t} \tag{1.15}$$

gives

$$m_j\vec{a}_j = -\vec{\nabla}_j\left[\sum_k\frac{-\hbar^2}{2m_k}\left(\frac{\nabla_k^2\sqrt{\rho}}{\sqrt{\rho}}\right) + V\right] \ . \tag{1.16}$$

We have derived an equation of motion of the form $F = ma$, similar to both (1.3) and (1.8).[7] The last term in the brackets gives the *classical* potential energy of the configuration of particles and makes no reference to the other worlds. The other term looks like an *interaction* between the worlds. This term is the quantum potential Q familiar from Bohmian mechanics (1.4), with $|\Psi|$ replaced by $\sqrt{\rho}$.

[7] This is the multi-particle version of Wyatt (2005, eq. 1.7); Holland (2005, eq. 4.9).

Within Prodigal QM, we've seen that one can derive an equation which determines the dynamics for all of the particles in all of the worlds *without ever referencing the wave function*. (1.16) gives a way of calculating the acceleration of a particle that doesn't mention Ψ, as (1.6) does, but only depends on the density of worlds ρ and the potential V. In Prodigal QM, this equation is derived, not part of the statement of the theory in the previous section. But, what if we took it to be the primary equation of motion for the particles? One can remove the wave function from Prodigal QM leaving only the corresponding ρ and $\vec{v}_k$s. So long as one enforces (1.16), the dynamics for particles will be essentially as they were in Prodigal QM.

Now we can formulate a new theory: *Newtonian QM*. Reality consists of a large but finite number of worlds whose distribution in configuration space is described by $\rho(\vec{x}_1, \vec{x}_2, ..., t)$. The velocities of the particles in the worlds are described by the velocity fields $\vec{v}_k(\vec{x}_1, \vec{x}_2, ..., t)$. The dynamical law for the velocity fields is (1.16), a Newtonian force law. As the particles move, the resultant shift in the distribution ρ is determined by (1.13). According to Newtonian QM, quantum mechanics is nothing but the Newtonian mechanics of particles in many different worlds.

> **Ontology:** (I) particles in many worlds described by a world density $\rho(\vec{x}_1, \vec{x}_2, ..., t)$ and velocity fields $\vec{v}_k(\vec{x}_1, \vec{x}_2, ..., t)$
>
> **Law:** (I) Newtonian force law (1.16)[8]

Comparing this statement of Newtonian QM to the formulation of Bohmian mechanics in §1.2, Newtonian QM is arguably the simpler theory. The theory has a single dynamical law and the fundamental ontology consists only of particles. However, this quick verdict could certainly be contested, especially in light of the

[8]The continuity equation (1.13), although used alongside (1.16) to calculate the dynamics, is not here considered a dynamical *law* since it merely encodes the fact that worlds are neither created nor destroyed. As is mentioned in §1.6, the Quantization Condition might be considered a non-dynamical law.

discussion below: (1.16) is not a fundamental law (§1.5); an unnatural restriction must be put on the space of states (§1.6); there are multiple ways to precisify ontology of the theory (§1.11).

1.5 The Continuum Approximation

Since the number of worlds is taken to be finite, the actual distribution of worlds will be highly discontinuous; some locations in configuration space will contain worlds and others will not. Still, we can use a smooth density function ρ to describe the distribution of worlds well enough at a coarse-grained level (see footnote 6). The velocity field $\vec{v}_k(\vec{x}_1, \vec{x}_2, ...)$ gives the mean velocity of the k-th particle in worlds near $(\vec{x}_1, \vec{x}_2, ...)$, but the k-th particle in a world at $(\vec{x}_1, \vec{x}_2, ...)$ may have a somewhat different velocity from $\vec{v}_k(\vec{x}_1, \vec{x}_2, ...)$. So, in Newtonian QM worlds will typically only approximately follow Bohmian trajectories through configuration space just as fluid particles do not exactly follow pathlines.[9]

In fluid dynamics, the use of a description of the fluid in terms of n and $\vec{u}$ is justified by the fact that we can calculate the dynamics of these coarse-grained properties (and others) without needing to know exactly what all the particles are doing. Also, it is the coarse-grained properties that we measure (Batchelor, 1967, §1.2; Chapman & Cowling, 1970, §5). What justifies the use of ρ and the $\vec{v}_k$s to describe the collection of worlds? As it turns out, we can calculate the dynamics of these properties without worrying about the exact locations of worlds via (1.13) and (1.16). Once the evolution of ρ and the $\vec{v}_k$s are known, we can use $\rho(t)$ to get probabilities (§1.9) and the $\vec{v}_k(t)$s to determine pathlines (showing that particles follow Bohmian trajectories).

[9] A *pathline* gives the trajectory of a particle always traveling at the mean velocity $\vec{u}$.

The equation of motion for the theory (1.16) treats the collection of worlds as a continuum. It fails to be a fundamental law since it does not describe the precise evolution of each world and is not valid if there are too few worlds to be well-described as a continuum. Slight deviations from standard quantum mechanical behavior should be expected due to the fact that there are only a finite number of worlds; worse deviations the fewer worlds there are. Future experiments may observe such deviations and support Newtonian QM. As textbook quantum mechanics works well, we have reason to believe there are a very large number of worlds. (The situation here is similar to that of spontaneous collapse theories, which are in principle empirically testable.) Ultimately, the quantum contribution to the force in (1.16) should be derivable from a more fundamental inter-world interaction. One should be able to calculate the forces when there are only a handful of worlds. Hopefully future research will explain how the continuum approximation arises from a "micro-dynamics" of worlds just as fluid dynamics arises from the micro-dynamics of molecules. For some progress in this direction, see Hall *et al.* (2014).

1.6 Reintroducing the Wave Function

In §1.4 we saw that for any wave function $\Psi(t)$ obeying the Schrödinger equation, there exists a world-density $\rho(t)$ and a collection of velocity fields $\vec{v}_k(t)$ obeying (1.16) such that the relations between Ψ, ρ, and the $\vec{v}_k$s expressed in (1.5) and (1.6) are satisfied at all times. The converse does not hold. There are some combinations of ρ and the $\vec{v}_k$s, that is, some ways the universe might be according to Newtonian QM, that do not correspond to any wave function. In general, we'll restrict our attention to combinations of ρ and the $\vec{v}_k$s that can be derived from a wave function via (1.5) and (1.6) as it is these states which reproduce the predictions of quantum physics. For such states, it may be useful to *introduce* a wave function, Ψ, even

though it is not a fundamental entity and does not appear in the equation of motion of the theory (1.16). The wave function serves as a convenient way of *summarizing* information about the positions and velocities of particles in the various worlds; the magnitude encodes the density of worlds (1.5) and the phase encodes the velocities of particles (1.11). The wave function need not be mentioned in stating the theory or (in principle) for deriving empirical predictions, but introducing a wave function is useful for making contact with standard treatments of quantum mechanics.

As was just mentioned, there are some states of the universe in Newtonian QM that do not correspond to quantum wave functions.[10] That is, there are some combinations of ρ and the $\vec{v}_k$s for which one cannot find a wave function Ψ that satisfies (1.5) and (1.6). The amplitude of Ψ follows straightforwardly from ρ, but not every set of velocity fields $\vec{v}_k$ can be expressed as $\frac{\hbar}{m_k}$ times the gradient of a phase (1.11). For this to be the case, we must impose a constraint on the velocity fields.[11]

Quantization Condition Integrating the momenta of the particles along any closed loop in configuration space gives a multiple of Planck's constant, $h = 2\pi\hbar$.

$$\oint \left\{ \sum_k \left[m_k \vec{v}_k \cdot d\vec{\ell}_k \right] \right\} = nh \ . \tag{1.17}$$

If the Quantization Condition is satisfied initially, (1.16) ensures that it will be satisfied at all times.

To see one sort of constraint this requirement imposes, think about the following case: a single electron orbiting a hydrogen nucleus in the $n = 2$, $l = 1$, $m = 1$ energy

[10] This point was made concisely by Wallstrom (1994) in the context of quantum hydrodynamics; it was noted earlier by Takabayasi (1952); see also Holland (2005, eq. 4.14).

[11] This is loosely analogous to the constraint on the fluid velocity field $\vec{u}$ that it be irrotational (everywhere zero vorticity) which is required to introduce a velocity potential (and for the validity of (1.8)).

eigenstate. For simplicity, take the nucleus to provide an external potential and the universe to contain many worlds with a single electron in each. The electron's wave function is

$$\Psi_{2,1,1}(r,\theta,\phi) = \frac{-1}{8\sqrt{a^5\pi}} e^{\frac{-r}{2a}} e^{i\phi} r\sin\theta \ , \tag{1.18}$$

where a is the Bohr radius. The guidance equation tells us that the particle in each world executes a circle around the z-axis with velocity $v_\phi = \frac{\hbar}{mrsin\theta}$, entirely in the $\widehat{\phi}$ direction (here ϕ is the azimuthal angle). (1.17) is trivially satisfied since $\frac{m\hbar}{mrsin\theta} \times 2\pi rsin\theta = h$. But, if the electrons were circling the z-axis a bit faster or a bit slower the integral wouldn't turn out right and (1.17) wouldn't be satisfied; they could orbit twice as fast but not 1.5 times as fast.

Without the Quantization Condition, Newtonian QM has too large a space of states. There are ways the universe might be that are quantum mechanical and others that are not. It is easy to specify what universes should be excluded, those that violate (1.17), but hard to give a principled reason why those states should be counted as un-physical, improbable, or otherwise ignorable. For now, I think it is best to understand the Quantization Condition as an empirically discovered feature of the current state of the universe, or equivalently, of the initial conditions. However, one might prefer to think of it as a non-dynamical law. A better explanation of the Quantization Condition's satisfaction would help strengthen Newtonian QM as it might seem that the best possible explanation of the condition's satisfaction is the *existence* of a wave function (backtracking to Prodigal QM). In the remainder of the chapter I will assume that the Quantization Condition is satisfied.

Suppose the world density and the velocity fields at a time are given. Provided the Quantization Condition is satisfied, there exists a wave function satisfying (1.5) and (1.6). But, is it unique?[12] That is, can (1.5) and (1.6) be used to define Ψ in

[12]Here the question is considered at the level of the continuum description. Because there

terms of ρ and the $\vec{v}_k$s?[13] First consider the case where ρ is everywhere nonzero. The magnitude of Ψ can be derived from (1.5), and (1.11) gives the phase up to a global constant. The wave function can be determined up to a global phase. This would be insufficient if the overall phase mattered, but as the global phase is arbitrary this gives exactly what we need. Actually, it's even better this way. The fact that the dynamics don't care about the overall phase is explained in Newtonian QM by the fact that changes in the global phase of the wave function don't change the state of the universe; that is, they don't change ρ or the $\vec{v}_k$s.

If the region in which $\rho \neq 0$ is not connected, the wave function is not uniquely determined by ρ and the $\vec{v}_k$s—one can introduce arbitrary phase differences between the separate regions. As an example of the breakdown of uniqueness, consider the second energy eigenstate of a single particle in a one-dimensional infinite square well of length L. In this case the wave function is

$$\psi_a(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{2\pi x}{L}\right) \quad . \tag{1.19}$$

This describes a universe with ρ and $\vec{v}$ given by

$$\begin{aligned} \rho(x) &= \frac{2}{L} \sin^2\left(\frac{2\pi x}{L}\right) \\ \vec{v}(x) &= \begin{cases} 0 & \text{if } x \neq \frac{L}{2} \\ undefined & \text{if } x = \frac{L}{2} \end{cases} \quad . \end{aligned} \tag{1.20}$$

The velocity field $\vec{v}$ is undefined where there are no worlds. These expressions for ρ

are multiple ways of coarse-graining, there will be multiple not-too-different ρs and $\vec{v}_k$s that well-describe any *finite* collection of worlds and thus many wave functions. It may be that some ways of coarse-graining avoid the problems raised below by ensuring that the velocity fields are always well-defined. If they do, the derivability of Ψ from ρ and the $\vec{v}_k$s comes at the cost of limiting the wave functions one can recover, losing those in (1.18), (1.19), and (1.21).

[13] See also the discussion in Holland (2005, §4).

and $\vec{v}$ are also compatible with[14]

$$\psi_b(x) = \sqrt{\frac{2}{L}} \left| \sin\left(\frac{2\pi x}{L}\right) \right| \ . \tag{1.21}$$

This exposes an inconvenient indeterminism: The time evolution of ψ_a is trivial as it is an energy eigenstate. Since ψ_b is not differentiable at $L/2$, its time evolution cannot be calculated straightforwardly using the Schrödinger equation (1.1). As (1.5) and (1.6) do not determine which wave function is to be used to describe the state in (1.20), it is not clear how the state will evolve. The future evolution of the universe is not uniquely determined by the instantaneous state (1.20), the continuity equation (1.13), and the equation of motion (1.16). This indeterminacy arises because ρ is zero and the velocity field is undefined at $L/2$, so $\frac{\partial \rho}{\partial t}$ and $\vec{a}$ are undefined at $L/2$. There is reason to think this indeterminism is an artifact of the continuum approximation where (1.13) and (1.16) need the velocity fields to be well-defined at every point in configuration space—even where there are no worlds—to yield a unique time evolution. The fundamental dynamics should take as input a specification of the position of each world in configuration space and the velocities of the particles in those worlds, all of which will be well-defined (§1.5).

Consider a slightly different problem from that just considered: Suppose one would would like to find a wave function $\Psi(t)$ which describes a *history* of $\rho(t)$ and the $\vec{v}_k(t)$s,

[14]The wave function ψ_b has the disreputable property of not being smooth. It should be noted that there exist pairs of distinct smooth non-analytic wave functions which agree on ρ and $\vec{v}$ at a time. (Thanks to Gordon Belot for suggesting an example like this.) For example,

$$\psi_\alpha(x) = \begin{cases} Ce^{\frac{-1}{1-(x+2)^2}} & \text{if } -3 < x < -1 \\ -Ce^{\frac{-1}{1-(x-2)^2}} & \text{if } 1 < x < 3 \\ 0 & \text{else} \end{cases}$$
$$\psi_\beta(x) = |\psi_\alpha(x)| \ .$$

satisfying (1.13) and (1.16) over some time interval. There will be a collection of wave functions which satisfy (1.5) and (1.6) at each time. For any such wave function, one can multiply it by a spatially homogeneous time-dependent phase factor, $e^{if(t)}$, to get another wave function which always satisfies (1.5) and (1.6). (The global phase at each time is arbitrary and (1.5) and (1.6) do nothing to stop you from picking whatever global phase you'd like at each time.) In general, some of these wave functions will satisfy the Schrödinger equation (1.1) and others will not. To constrain the time-dependence of the phase when using a wave function to describe histories, (1.14) can be imposed as a third link between the wave function and the particles (in addition to (1.5) and (1.6)). Because (1.5), (1.6), (1.13), and (1.14) hold, the wave function must obey the Schrödinger equation.

This section began with the observation that there are states in Newtonian QM that cannot be described by a wave function. However, these can be excised by imposing the Quantization Condition. Given a state that *can* be described by a wave function, one might hope that this wave function would be unique. Sometimes it is not. A wave function aptly describes a state in Newtonian QM *at a time* if (1.5) and (1.6) are satisfied. But, if these are the only constraints, a *history* in Newtonian QM can always be described by many wave functions. So, there is freedom to add a third connection between the wave function and the particles. Imposing (1.14) proves a convenient choice as it guarantees that the wave function obeys the Schrödinger equation—a desirable feature since the point of introducing a wave function was to clarify the connection between Newtonian QM and standard treatments of quantum mechanics.

Because a wave function can be introduced to describe the world density and the velocity fields, one is free to use well-known techniques to calculate the time evolution of the wave function and use that to determine how the world density and velocity fields evolve. However, there is evidence that it is sometimes easier to use

the trajectories of worlds to calculate the time evolution (Wyatt, 2005; Hall *et al.* , 2014).

1.7 Probability: Versus Everettian Quantum Mechanics

The Born Rule is easier to justify in Newtonian QM than in the many-worlds interpretation. In Everettian QM, there is dispute over how one can even make sense of assigning probabilities to measurement outcomes when the way the universe will branch is deterministic and known (the incoherence problem). There is also the quantitative problem of why the Born Rule probabilities are the right ones to assign. Recent derivations tend to appeal to complex decision-theoretic arguments, which, although they may ultimately be successful, are not uncontroversially accepted (Saunders *et al.* , 2010). Things look worrisome because there are some *prima facie* plausible ways of counting agents which yield the result that *the vast majority of agents* see relative frequencies of experimental outcomes which deviate significantly from those predicted by the Born Rule (although the total amplitude-squared weight of the branches in which agents see anomalous statistics is small). Newtonian QM does not run into similar problems since the number of worlds in a particular region of configuration space is always proportional to $|\Psi|^2$. At any time, most agents are in high amplitude regions. So, in typical measurement scenarios, most agents will see long-run frequencies which agree with the predictions of the Born Rule.

Were a proponent of Prodigal QM to claim similar advantages over Everettian QM, one could reasonably object that the Born Rule is recovered only because it was put in by hand. In Prodigal QM, (1.5) is an additional postulate. In Newtonian QM, it is not. The density of worlds is given by $|\Psi|^2$ because Ψ is definitionally

related to the density of worlds by (1.5) (see §1.6). The wave function is, after all, not fundamental but a mere description of ρ and the $\vec{v}_k$s.

1.8 Probability: Versus Bohmian Mechanics

Although it is widely agreed that the Born Rule can be justified in Bohmian mechanics, there is disagreement about how exactly the story should go. In this section I will briefly discuss three ways of justifying the Born Rule in Bohmian mechanics and then argue that Newtonian QM can give a cleaner story. First, though, note an important similarity between the two theories. According to Newtonian QM each world follows an approximately Bohmian path through configuration space. So if you think that worlds in which particles follow Bohmian trajectories are able to reproduce the results of familiar quantum experiments, you should think worlds in Newtonian QM can too.

In Bohmian mechanics, not all initial conditions reproduce the statistical predictions of quantum mechanics. That is, not all specifications of the initial wave function $\Psi(0)$ and particle configuration $(\vec{x}_1(0), \vec{x}_2(0), ...)$ yield a universe in which experimenters would see long-run statistics of measurements on subsystems which agree with the predictions of the Born Rule. Why should we expect to be in one of the universes with Born Rule statistics? One way to respond to this problem is to add a *postulate* to the theory which ensures that ensembles of particles in the universe will (or almost certainly will) display Born Rule statistics upon measurement (e.g., Holland, 1993, §3.6.3). A second option is to argue that *typical* universes are such that Born Rule frequencies will be observed when measurements are made (Dürr *et al.* , 1992). To say that such results are "typically" observed is to say that: for any initial wave function $\Psi(0)$, the vast majority of initial particle configurations reproduce Born Rule statistics. Speaking of the "vast majority" of initial configurations only makes

sense relative to a way of measuring the size of regions of configuration space; here the measure used is given by $|\Psi|^2$. A third option: one could argue that many initial states will start to display Born Rule statistics sufficiently rapidly that, since we are not at the beginning of the universe, we should expect to see Born Rule frequencies now even if such frequencies were not displayed in the distant past (Valentini & Westman, 2005).

Each of these proposals faces challenges. The additional postulates which might be added to the theory look ad hoc. The measure used to determine typicality must be satisfactorily justified.[15] The desirable evolution of states described in the third option has only been demonstrated in relatively simple cases. Also, there will certainly exist initial conditions that do not come to display Born Rule statistics sufficiently rapidly and these must somehow be excluded. To the extent that one finds these objections to Bohmian strategies worrisome, it is an advantage of the new theory that it avoids them.

Although Newtonian QM, like Bohmian mechanics, permits a particular world to have a history of measurement results where the frequencies of outcomes do not match what one would expect from the Born Rule, it is *impossible* for the density of worlds to deviate from $|\Psi|^2$. So, in light of the results in Dürr *et al.* (1992), it will always be the case that Born Rule statistics are observed in the vast majority of *worlds* in any *universe* of Newtonian QM. Since we're not sure which world we are in, we should expect to be in one in which Born Rule statistics are observed.

[15] For a statement of the objection, see Dickson (1998, §5.4). For a variety of reasons to regard the measure as natural, see Goldstein & Struyve (2007).

1.9 Probability: Newtonian QM

If the universe's evolution is deterministic and the initial state is known, what is there left for an agent to assign probabilities to? There is no incoherence problem in Newtonian QM since, given the state of the universe, one is generally uncertain which of the many distinct worlds one is in. There will always be many possibilities consistent with one's immediate experiences. The uncertainty present here is *self-locating uncertainty* (see Lewis, 1979). Of course, there will generally also be uncertainty about the state of the universe.

On to the quantitative problem:[16] Given a particular distribution of worlds ρ and set of velocity fields $\vec{v}_k$, that is, given a specification of the state of the universe, one ought to assign *equal credence* to being in any of the worlds consistent with one's evidence.[17] Because there are only a finite number worlds, this advice is unambiguous. As it turns out, this basic indifference principle suffices to derive the correct quantum probabilities. Consider an idealized case in which the agent knows the world density and the velocity fields, and knows that there is an agent in each of these worlds having experiences indistinguishable from her own. In this case, the above indifference principle tells her to assign probabilities to being in different regions of configuration space in accordance with ρ. Since $\rho = |\Psi|^2$, she must assign credences in accordance with $|\Psi|^2$ and thus in agreement with the Born Rule. Next, suppose this agent learns the outcome of an experiment. Then she ought to assign zero credence to the worlds inconsistent with her evidence and reapportion that credence among those which remain (keeping the probability of each non-eliminated world equal). This updating is analogous to learning which branch you are on after a measurement in Everettian

[16] See also the discussion in Boström (2012, §2.4).

[17] This follows from a more general epistemic principle defended in Elga (2004).

QM.

In general, the probability agent S ought to assign to her own world having property A, conditional on a particular state of the universe at a certain time, is

$$\begin{aligned}\Pr\big(A\big|\rho, \vec{v}_1, \vec{v}_2, ...\big) &= \frac{\text{number of worlds with property } A \text{ and a copy of } S}{\text{number of worlds with a copy of } S} \\ &= \frac{\int dV_{AS}\ \rho(\vec{x}_1, \vec{x}_2, ...)}{\int dV_S\ \rho(\vec{x}_1, \vec{x}_2, ...)} = \frac{\int dV_{AS}\ |\Psi(\vec{x}_1, \vec{x}_2, ...)|^2}{\int dV_S\ |\Psi(\vec{x}_1, \vec{x}_2, ...)|^2}\ . \end{aligned} \tag{1.22}$$

Here A could be something like, "the pointer indicates 7" or "the particle just fired will hit in the third band of the interference pattern." The volume V_S delimits the set of worlds, specified by a region of configuration space, compatible with S's data. Worlds in this region are such that previous experiments had the outcomes S remembers them having, macroscopic arrangements of particles match what S currently observes, and some person is having the same conscious experiences as S.[18] The volume V_{AS} gives the set of worlds compatible with S's data in which A holds.[19] These conditional probabilities can be used to test hypotheses about ρ and the $\vec{v}_k$s and thus to learn about the state of the universe (not just one's own world) from experience.

1.10 Continuous Infinity or Mere Multitude of Worlds?

So far, we have taken ρ to describe the distribution of a large but finite number of worlds. But, one might be tempted to defend a variant of Newtonian QM in which there are a continuous infinity of worlds, one at every point at which ρ is

[18] For simplicity, I have neglected the possibility that S's memories or current observations are deceptive.

[19] Note that the boundaries of V_S and V_{AS} will often depend on ρ and the $\vec{v}_k$s.

non-zero. This causes trouble. The *meaning* of ρ becomes unclear if we move to a continuous infinity of worlds since we can no longer understand ρ as yielding the *proportion* of all worlds in a given volume of configuration space upon integration over that volume. There would be infinite worlds in any finite volume (where $\rho \neq 0$) and infinite total worlds. If ρ doesn't give the proportion of worlds in a region, it is unclear why epistemic agents should apportion credences as recommended in the previous section. So, the continuous variant, if sense can be made of it, faces the quantitative probability problem head on.

As discussed in §1.5, the dynamical law proposed for Newtonian QM (1.16) is not fundamental. If it somehow turns out that we cannot view the force caused by the quantum potential as arising from an interaction between individual worlds, this would provide a reason to accept a continuous infinity of quantum worlds over a mere multitude. It might appear to be a strength of the continuous variant that its laws can already be precisely stated, but I expect that this advantage will evaporate when possible fundamental interactions are formulated for the discrete variant. The continuous variant does have a serious advantage: the continuum approximation (§1.5) is no approximation. Particles will unerringly follow Bohmian trajectories.

1.11 Ontology

According to Newtonian QM, what the universe contains is a finite collection of worlds. There are at least two ways to precisify this idea. First, one might take *configuration space* to be the fundamental space, inhabited by point-particles (worlds). Second, one might take the fundamental space to be ordinary *three-dimensional space*, inhabited by particles in different worlds.

According to the first picture, on the fundamental level, the universe is $3N$-dimensional and contains a large number of point particles, each of which has

dynamics so complex that it merits the name of "world" or "world-particle." Forces between these world-particles are Newtonian and the dynamics are local. Here Newtonian QM is a theory of the Newtonian dynamics of a fluid of world-particles in $3N$-dimensional space. Albert (1996) has argued that the one world of Bohmian mechanics can be understood as a world-particle which moves around in configuration space guided by the wave function. He provides a way of explaining how the appearance of a three-dimensional world arises from the motion of this world-particle which applies *mutatis mutandis* to Newtonian QM in which there are more world-particles executing the same old Bohmian dances.

On the second picture there are particles interacting in three-dimensional space, nothing more.[20] Space is very densely packed with particles, but not all particles are members of the same world. Some particles are members of world #1, some of world #2, etc. What world a particle belongs to might be a primitive property, like its mass or charge. The equation of motion for a particle in world #827, (1.16), says that the force from the potential V depends only on the positions of the other particles in world #827. However, the quantum potential introduces an inter-world force whereby particles that are not members of world #827 can still impact the trajectory of a particle in this world. So, particles which happen to be members of the same world interact in one way, whereas particles which are members of different worlds interact another way.

In the many-worlds interpretation, one must tell a somewhat complicated story about how people and quantum worlds arise as emergent entities in the time-evolving quantum state (e.g., Wallace, 2003). This story may not be successful. It might be the case that wave functions evolving in accordance with the Schrödinger equation

[20] This second option resembles the novel ontology for the many-worlds interpretation proposed by Allori *et al.* (2011).

are incapable of supporting life or at least lives that feel like ours (Maudlin, 2010). If that's right, Newtonian QM has a potential advantage. On the second ontological picture, people are built from particles in the usual way. On the first ontological picture, there is a story about emergence that must be told but the details of the story are very different from the Everettian one and it succeeds or fails independently.

If, on the other hand, the Everettian story about emergence is successful, then Bohmian mechanics (as formulated here) faces the *Everett-in-denial objection* (Deutsch, 1996; Brown & Wallace, 2005; Valentini, 2010). Both Everettian QM and Bohmian mechanics contain in their fundamental ontology a wave function which always obeys the Schrödinger equation. If such a wave function is sufficient for there to be creatures experiencing what appears upon not-too-close inspection to be a classical world, then Bohmian mechanics, like Everettian QM, includes agents who see every possible outcome of a quantum measurement. If the Everettian story about emergence works and the Everett-in-denial objection against Bohmian mechanics is successful, then Newtonian QM has a serious advantage over Bohmian mechanics. Newtonian QM cannot be accused of being a many worlds theory in disguise since the theory embraces its many worlds ontology.

1.12 Symmetries: Time Reversal and Galilean Boosts

Newtonian QM can help us understand symmetry transformations in quantum mechanics. First, consider time reversal. Albert (2000) proposes an intuitive and general account of time reversal symmetry in physical theories which judges QM, in all of its familiar precisifications, to *fail* to be time-reversal invariant. A deterministic physical theory specifies which sequences of instantaneous states

are allowed and which are forbidden through dynamical laws. If the laws allow the time-reversed history of instantaneous states for any allowed history of instantaneous states, then the theory is deemed time-reversal invariant. In theories like Bohmian mechanics or Everettian QM, the instantaneous state includes the wave function at a time $\Psi(\vec{x}_1, \vec{x}_2, ..., t)$ and a complete history includes the wave function at all times. The time reverse of the history is $\Psi(\vec{x}_1, \vec{x}_2, ..., -t)$. $\Psi(\vec{x}_1, \vec{x}_2, ..., -t)$ will not necessarily satisfy the Schrödinger equation whenever $\Psi(\vec{x}_1, \vec{x}_2, ..., t)$ does—so quantum mechanics is judged not to be time-reversal invariant. However, $\Psi^*(\vec{x}_1, \vec{x}_2, ..., -t)$ will *always* satisfy the Schrödinger equation whenever $\Psi(\vec{x}_1, \vec{x}_2, ..., t)$ does (standard textbook accounts take this to be the time reversed history and thus judge the theory to be time-reversal invariant).

In Newtonian QM, it is straightforward to show that time reversing the history of particle trajectories amounts to changing the history of the wave function from $\Psi(\vec{x}_1, \vec{x}_2, ..., t)$ to $\Psi^*(\vec{x}_1, \vec{x}_2, ..., -t)$. The *instantaneous* state of the world is specified by giving the locations (but *not* the velocities) of all of the particles in all of the worlds. The time reversal operation thus takes the history $\rho(\vec{x}_1, \vec{x}_2, ..., t)$ and $\vec{v}_k(\vec{x}_1, \vec{x}_2, ..., t)$ to $\rho(\vec{x}_1, \vec{x}_2, ..., -t)$ and $-\vec{v}_k(\vec{x}_1, \vec{x}_2, ..., -t)$. By (1.11), flipping the phase generates a wave function which describes the flipped velocities of particles in the time-reversed history. The complex conjugation in the textbook time reversal operation for quantum mechanics can be explained as deriving from a reversal in the velocities of the particles.

Newtonian QM is time-reversal invariant according to Albert's account. Even if one doesn't agree with Albert's account of time-reversal invariance, it is a virtue of this theory over others that it can give a simple explanation of *why* the wave function transforms in the textbook way under time-reversal.

Next, consider Galilean boosts. In a similar spirit to Albert's criticism of the standard account of time-reversal, one could argue that quanrtum mechanics

is not invariant under Galilean boosts since the equations of motion are not generally obeyed when we take $\Psi(\vec{x}_1, \vec{x}_2, ..., t)$ to $\Psi(\vec{x}_1 - \vec{w}t, \vec{x}_2 - \vec{w}t, ..., t)$.[21] The invariance of quantum mechanics under Galilean boosts is sometimes demonstrated by showing that, for certain potentials, there *exists* a transformation of the state which appropriately shifts the probability density and guarantees satisfaction of the Schrödinger equation (e.g., Ballentine, 1998, §4.3). Under a boost by $\vec{w}$, the wave function is supposed to transform as

$$\Psi_0(\vec{x}_1, \vec{x}_2, ..., t) \xrightarrow{\vec{w}} \Psi(\vec{x}_1, \vec{x}_2, ..., t) = e^{\frac{i}{\hbar}\sum_k \left\{m_k \vec{w}\cdot\vec{x}_k - \frac{1}{2}m_k|\vec{w}|^2 t\right\}} \Psi_0(\vec{x}_1 - \vec{w}t, \vec{x}_2 - \vec{w}t, ..., t)\ . \tag{1.23}$$

It's interesting that there exists a transformation which moves probability densities in the right way and guarantees that the Schrödinger equation is invariant under boosts, but it is unclear why this particular transformation is the one that really represents Galilean boosts. In Newtonian QM this transformation of the wave function results from boosting the velocities of all of the particles in all of the worlds.

Adding $\vec{w}$ to the velocity of each particle transforms the original density $\rho_0(t)$ and the original velocity fields $\vec{v}_{0k}(t)$ to

$$\begin{aligned} \rho(\vec{x}_1, \vec{x}_2, ..., t) &= \rho_0(\vec{x}_1 - \vec{w}t, \vec{x}_2 - \vec{w}t, ..., t) \\ \vec{v}_k(\vec{x}_1, \vec{x}_2, ..., t) &= \vec{v}_{0k}(\vec{x}_1 - \vec{w}t, \vec{x}_2 - \vec{w}t, ..., t) + \vec{w}\ . \end{aligned} \tag{1.24}$$

Suppose $\Psi_0(t)$, $\rho_0(t)$, and the $\vec{v}_{0k}(t)$s satisfy (1.5), (1.6), and (1.14); that is, $\Psi_0(t)$ *describes* this density and these velocity fields. Then, the new wave function $\Psi(t)$ generated by the transformation in (1.23) will satisfy (1.5), (1.6), and (1.14) for the $\rho(t)$ and $\vec{v}_k(t)$s in (1.24), provided that the potential V is translation invariant (as

[21] A point made by Albert in presentations. See also Valentini (1997).

the reader can verify). Thus, (1.23) gives a general recipe for finding a wave function which correctly describes the boosted particles.

1.13 Spin-1/2 Particles

There appears to be serious trouble on the horizon for this new theory. In Bohmian mechanics spin is often treated as a property of the wave function, not the particles pushed along by it.[22] So, if we remove the wave function, it looks like we'll lose all of the information about the spin of the system! Actually, there is a very natural way to extend Newtonian QM to a single particle with spin. If we endow the particle with a *definite* spin in every world, we can recover the standard dynamics. Here I'll apply to Newtonian QM a strategy which has been used in quantum hydrodynamics and (a version of) Bohmian mechanics (see Holland, 1993, ch. 9 and references therein).

Consider the dynamics of a single spin-1/2 particle. To our basic ontology, consisting of a distribution of worlds $\rho(\vec{x},t)$ where the particle has velocity $\vec{v}(\vec{x},t)$ in each world, let us add a property to the particle in each world: spin magnetic moment. The spin magnetic moment $\vec{u}(\vec{x},t)$ of a particle can be specified by a polar angle $\alpha(\vec{x},t)$, an azimuthal angle $\beta(\vec{x},t)$, and a constant μ (for an electron, $\mu \approx \frac{-e\hbar}{2m}$, where e is the magnitude of the electron's charge).

$$\vec{\mu} = \mu \begin{pmatrix} \sin\alpha\cos\beta \\ \sin\alpha\sin\beta \\ \cos\alpha \end{pmatrix} \tag{1.25}$$

Alternatively, we can speak of the particle's internal angular momentum $\vec{S}$, which is

[22] e.g., Dürr & Teufel (2009, §8.4) and Albert (1994, ch. 7)

related to $\vec{\mu}$ by

$$\vec{S} = \frac{\hbar}{2\mu}\vec{\mu} \ . \tag{1.26}$$

With the magnetic moment in hand, we can partially define[23] the spinor wave function χ from ρ and $\vec{\mu}$ by

$$\chi = \begin{pmatrix} \chi_+ \\ \chi_- \end{pmatrix} = \begin{pmatrix} \sqrt{\rho}\cos\frac{\alpha}{2}\, e^{i\theta} \\ \sqrt{\rho}\sin\frac{\alpha}{2}\, e^{i\theta+i\beta} \end{pmatrix} \ , \tag{1.27}$$

similar to (1.10). Here the z-spin basis is used to represent the spinor.

The Bohmian guidance equation for a spin-1/2 particle is

$$\vec{v} = \frac{\hbar}{m}\mathrm{Im}\left[\frac{\chi^\dagger\vec{\nabla}\chi}{\chi^\dagger\chi}\right] \ . \tag{1.28}$$

Inserting the expression for χ in (1.27) yields

$$\vec{v} = \frac{\hbar}{m}\vec{\nabla}\theta + \frac{\hbar}{m}\sin^2\frac{\alpha}{2}\,\vec{\nabla}\beta \ , \tag{1.29}$$

similar to (1.11).

The Pauli equation for a spin-1/2 particle in the presence of an external magnetic field is

$$i\hbar\frac{\partial}{\partial t}\chi = \left\{\frac{-\hbar^2}{2m}\nabla^2 + V - \mu\vec{B}\cdot\sigma\right\}\chi \ , \tag{1.30}$$

where σ are the Pauli spin matrices. To focus on spin, the contributions to the Hamiltonian arising because the particle has a charge (not just a magnetic moment) have been omitted. From (1.27), (1.29), and (1.30) one can derive the time dependence

[23] This definition is only partial as θ is left unspecified.

of $\vec{\mu}$ and $\vec{v}$. The magnetic moment vector evolves as

$$\begin{aligned}\frac{\hbar}{2\mu}\frac{d\vec{\mu}}{dt} =& \frac{\hbar^2}{4m\mu^2\rho}\vec{\mu}\times\left[\partial_a\left(\rho\,\partial_a\vec{\mu}\right)\right]+\vec{\mu}\times\vec{B}\\ \frac{d\vec{S}}{dt} =& \vec{\mu}\times\vec{B}_{\text{Tot}}\ ,\end{aligned} \tag{1.31}$$

using the Einstein summation convention over spatial index a.[24] The right hand side gives the net torque on the particle, which arises from a quantum and a classical contribution. These torques can be combined by defining

$$\vec{B}_{\text{Tot}} \equiv \vec{B} + \frac{\hbar^2\left[\partial_a\left(\rho\,\partial_a\vec{\mu}\right)\right]}{4m\mu^2\rho}\ . \tag{1.32}$$

The net magnetic field $\vec{B}_{\text{Tot}}$ is the sum of a classical and a quantum contribution. (1.31) gives the classical dynamics for the angular momentum of a magnetic dipole in the presence of the magnetic field $\vec{B}_{\text{Tot}}$.

From (1.27), (1.29), and (1.30), it follows that the acceleration can be expressed as

$$m\vec{a} = -\vec{\nabla}\left[Q+Q_P+V\right]+\mu_a\vec{\nabla}B_{\text{Tot}a}\ . \tag{1.33}$$

This is simply the equation of motion for a particle without spin (1.16) with two new terms: the classical force on a particle with magnetic moment $\vec{\mu}$ from a magnetic field $\vec{B}_{\text{Tot}}$ and a spin-dependent contribution to the quantum potential,

$$Q_P = \frac{\hbar^2}{8m\mu^2}\vec{\mu}\cdot\left(\nabla^2\vec{\mu}\right) = \frac{1}{2m}\vec{S}\cdot\left(\nabla^2\vec{S}\right)\ . \tag{1.34}$$

[24] Result as in Holland (1993, eq. 9.3.15). Note that different conventions are adopted for the sign of μ. (1.33) is in agreement with Holland's eq. 9.3.19, although written in a more suggestive form.

As with the quantum potential Q discussed in §1.4, this new term represents an interaction between worlds (as does the quantum contribution to the net magnetic field $\vec{B}_{\text{Tot}}$). Together, the above equations of motion for $\vec{\mu}$ and $\vec{v}$, (1.31) and (1.33), serve to define Newtonian QM for a single spin-1/2 particle. We can omit any mention of the spinor wave function χ or the phase θ in the fundamental laws. The equations of motion for $\vec{\mu}$ and $\vec{v}$, which govern the evolution of ρ via (1.13), will guarantee that ρ, $\vec{\mu}$, and $\vec{v}$ will evolve as if they were governed by a spinor wave function satisfying the Pauli equation, provided that the velocity field obeys a constraint like the one imposed for spin-0 particles in §1.6,

$$\oint \left(m\vec{v} - \hbar \sin^2 \frac{\alpha}{2} \vec{\nabla}\beta \right) \cdot d\vec{\ell} = nh \ . \tag{1.35}$$

In Newtonian QM, particles have well-defined spin magnetic moments at all times. How can the theory recover the results of standard experiments involving spin if particles are never in superpositions of different spin states? Consider, for example, a z-spin "measuring" Stern-Gerlach apparatus. Suppose the wave function is in a superposition z-spin up and z-spin down: $\frac{1}{\sqrt{2}} |\uparrow_z\rangle + \frac{1}{\sqrt{2}} |\downarrow_z\rangle$. When passed through the inhomogeneous magnetic field, the wave function will split in two. On the standard account, the particle will be found in either the upper region (corresponding to z-spin up) or the lower region (corresponding to z-spin down) upon measurement with equal probability. In Newtonian QM, there is initially an ensemble of worlds, in each of which the particle has some initial position in the wave packet and in all of which the particle's spin magnetic moment points squarely in the x-direction. A particle in the top half of the initial wave packet has its spin rotated to point in the z-direction as it passes through the Stern-Gerlach apparatus (in accordance with (1.31)); a particle in the lower portion will end up with spin pointing in the negative z-direction. In this theory, the Stern-Gerlach apparatus does not *measure* z-spin, but

instead forces particles to align their magnetic moments along the z-axis. This is also how Stern-Gerlach measurements are interpreted in versions of Bohmian mechanics where particles have definite spins (see Dewdney *et al.* , 1986; Holland, 1993, ch. 9).

1.14 Conclusion

An optimistic synopsis: Once we realize that Newtonian QM is a viable way of understanding non-relativistic quantum mechanics, we see that we never needed to overthrow Newtonian mechanics with a quantum revolution. One can formulate quantum mechanics in terms of point particles interacting via Newtonian forces. The mysterious wave function is merely a way of summarizing the properties of particles, not a piece of fundamental reality.

There are a variety of reasons not to like this theory. First, there is arguably a cost associated with the abundance of other worlds which, although detectable via their interactions with our own world, are admittedly odd. Second, the space of states for the theory is larger than one might like in two distinct ways: There are possible combinations of ρ and the $\vec{v}_k$s that do not correspond to any wave function because the velocity fields cannot be expressed as the gradient of a phase (§1.6). There are also states of the universe where the number of worlds is not sufficiently large for the continuum description to be valid (§1.5). Even if there are a great many worlds, slight divergence from the predictions of standard quantum mechanics is to be expected. Third, it is a shortcoming of the current formulation of Newtonian QM that we must approximate the actual distribution of worlds as continuous and cannot yet formulate the fundamental equation of motion precisely for a discrete collection of worlds (§1.5). Finally, the theory is limited in that it is not here extended to systems of multiple particles with spin or to relativistic quantum physics.

In addition to its seductive conservatism, I view the following comparative

strengths as most compelling. Against the many-worlds interpretation, Newtonian QM has two main advantages. First, there is no incoherence problem or quantitative probability problem—the Born Rule can be justified quickly from self-locating uncertainty (§1.7). Second, the theory avoids the need to explain how worlds emerge from the wave function—worlds are taken to be fundamental (§1.11). Compared to Bohmian mechanics, the theory is arguably simpler—it replaces an ontology of wave functions and particles with one just containing particles (§1.4). Newtonian QM's explanation of why we should expect our world to reproduce Born Rule statistics is potentially more compelling than the Bohmian stories (§1.8). Also, Newtonian QM is forthright about its many worlds character, sidestepping the Everett-in-denial objection (§1.11).

Acknowledgements: Thanks to David Baker, Gordon Belot, Cian Dorr, Detlef Dürr, J. Dmitri Gallow, Michael Hall, Sheldon Goldstein, Daniel Peterson, Laura Ruetsche, Ward Struyve, Nicola Vona, and two anonymous referees for very useful feedback on drafts of this chapter. Thank you to Adam Becker, Sean Carroll, Dirk-André Deckert, Neil Dewar, Benjamin Feintzeig, Sophie Monahan, Cat Saint-Croix, Jonathan Shaheen, and Howard Wiseman for helpful discussions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 0718128.

References

Albert, David Z. 1994. *Quantum Mechanics and Experience*. Harvard University Press.

Albert, David Z. 1996. Elementary Quantum Metaphysics. *Boston Studies in the Philosophy of Science*, **184**, 277–284.

Albert, David Z. 2000. *Time and Chance*. Harvard University Press.

Allori, Valia, Goldstein, Sheldon, Tumulka, Roderich, & Zanghì, Nino. 2011. Many Worlds and Schrödinger's First Quantum Theory. *The British Journal for the Philosophy of Science*, **62**(1), 1–27.

Ballentine, Leslie E. 1998. *Quantum Mechanics: A Modern Development*. World Scientific Publishing Company Incorporated.

Barrett, Jeffrey A. 1999. *The Quantum Mechanics of Minds and Worlds*. Oxford University Press.

Batchelor, George K. 1967. *An Introduction to Fluid Mechanics*. Cambridge University Press.

Boström, Kim Joris. 2012. *Combining Bohm and Everett: Axiomatics for a Standalone Quantum Mechanics*. arXiv:1208.5632 [quant-ph].

Brown, Harvey R., & Wallace, David. 2005. Solving the Measurement Problem: De Broglie–Bohm Loses Out to Everett. *Foundations of Physics*, **35**(4), 517–540.

Chapman, Sydney, & Cowling, Thomas G. 1970. *The Mathematical Theory of Non-Uniform Gases*. Cambridge University Press.

Deutsch, David. 1996. Comment on Lockwood. *The British Journal for the Philosophy of Science*, **47**(2), 222–228.

Dewdney, Christopher, Holland, Peter R., & Kyprianidis, A. 1986. What Happens in a Spin Measurement? *Physics Letters A*, **119**(6), 259 – 267.

Dickson, William M. 1998. *Quantum Chance and Non-Locality: Probability and Non-Locality in the Interpretation of Quantum Mechanics*. Cambridge University Press.

Dorr, Cian. 2009. Finding Ordinary Objects in Some Quantum Worlds. Manuscript.

Dürr, Detlef, & Teufel, Stefan. 2009. *Bohmian Mechanics*. Springer.

Dürr, Detlef, Goldstein, Sheldon, & Zanghì, Nino. 1992. Quantum Equilibrium and the Origin of Absolute Uncertainty. *Journal of Statistical Physics*, **67**(5), 843–907.

Elga, Adam. 2004. Defeating Dr. Evil with Self-Locating Belief. *Philosophy and Phenomenological Research*, **69**(2), 383–396.

Goldstein, Sheldon, & Struyve, Ward. 2007. On the Uniqueness of Quantum Equilibrium in Bohmian Mechanics. *Journal of Statistical Physics*, **128**(5), 1197–1209.

Hall, Michael J. W., Deckert, Dirk-André, & Wiseman, Howard M. 2014. Quantum Phenomena Modeled by Interactions between Many Classical Worlds. *Phys. Rev. X*, **4**, 041013.

Holland, Peter. 2005. Computing the Wavefunction from Trajectories: Particle and Wave Pictures in Quantum Mechanics and Their Relation. *Annals of Physics*, **315**(2), 505–531.

Holland, Peter R. 1993. *The Quantum Theory of Motion: An Account of the de Broglie-Bohm Causal Interpretation of Quantum Mechanics*. Cambridge University Press.

Lewis, David K. 1979. Attitudes De Dicto and De Se. *The Philosophical Review*, **88**(4), 513–543.

Madelung, Erwin. 1927. Quantentheorie in Hydrodynamischer Form. *Zeitschrift für Physik A Hadrons and Nuclei*, **40**(3), 322–326.

Maudlin, Tim. 2010. Can the World be Only Wavefunction? In Saunders *et al.* (2010).

Poirier, Bill. 2010. Bohmian Mechanics without Pilot Waves. *Chemical Physics*, **370**(13), 4 – 14.

Saunders, Simon, Barrett, Jonathan, Kent, Adrian, & Wallace, David (eds). 2010. *Many Worlds?: Everett, Quantum Theory, & Reality*. Oxford University Press.

Schiff, Jeremy, & Poirier, Bill. 2012. Communication: Quantum Mechanics without Wavefunctions. *The Journal of Chemical Physics*, **136**, 031102.

Takabayasi, Takehiko. 1952. On the Formulation of Quantum Mechanics associated with Classical Pictures. *Progress of Theoretical Physics*, **8**(2), 143–182.

Tipler, Frank J. 2006. *What about Quantum Theory? Bayes and the Born*

Interpretation. arXiv:quant-ph/0611245.

Valentini, Antony. 1997. On Galilean and Lorentz Invariance in Pilot-wave Dynamics. *Physics Letters A*, **228**(4), 215–222.

Valentini, Antony. 2010. De Broglie-Bohm Pilot-Wave Theory: Many Worlds in Denial? in Saunders *et al.* (2010).

Valentini, Antony, & Westman, Hans. 2005. Dynamical Origin of Quantum Probabilities. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, **461**(2053), 253–272.

Wallace, David. 2003. Everett and Structure. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **34**(1), 87–105.

Wallace, David. 2012. *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. Oxford University Press.

Wallstrom, Timothy C. 1994. Inequivalence Between the Schrödinger Equation and the Madelung Hydrodynamic Equations. *Physical Review A*, **49**(3), 1613.

Wyatt, Robert E. 2005. *Quantum Dynamics with Trajectories: Introduction to Quantum Hydrodynamics*. Springer. Corey J. Trahan (contributor).

Chapter 2

Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics

Charles T. Sebens and Sean M. Carroll

Abstract: A longstanding issue in attempts to understand the Everett (Many-Worlds) approach to quantum mechanics is the origin of the Born rule: why is the probability given by the square of the amplitude? Following Vaidman, we note that observers are in a position of self-locating uncertainty during the period between the branches of the wave function splitting via decoherence and the observer registering the outcome of the measurement. In this period it is tempting to regard each branch as equiprobable, but we argue that the temptation should be resisted. Applying lessons from this analysis, we demonstrate (using methods similar to those of Zurek's envariance-based derivation) that the Born rule is the uniquely rational way of apportioning credence in Everettian quantum mechanics. In doing so, we rely on a single key principle: changes purely to the environment do not affect the probabilities one ought to assign to measurement outcomes in a local subsystem. We arrive at a method for assigning probabilities in cases that involve both classical and quantum self-locating uncertainty. This method provides unique answers to quantum Sleeping Beauty problems, as well as a well-defined procedure for calculating probabilities in quantum cosmological multiverses with multiple similar observers. [The published version of this chapter is forthcoming in the *British Journal for the Philosophy of*

Science.]

2.1 Introduction

The Everett (or Many-Worlds) approach to quantum mechanics is distinguished by its simplicity. The dynamics of the theory consists of a single deterministic evolution law, the Schrödinger equation. There are no separate rules for dealing with 'wave function collapse' and quantum measurement, nor are there additional hidden variables. A longstanding challenge for such an approach is to reproduce the Born rule: the probability of a measurement yielding eigenvalue a of the observable $\widehat{A}$, given that the system was prepared in state $|\psi\rangle$, is given by $|\langle a|\psi\rangle|^2$, where $|a\rangle$ is the[1] eigenstate with eigenvalue a. Indeed, it seems like quite a challenge to explain how Everettian quantum mechanics could provide a theory of probabilities at all, given that states can be specified with arbitrary precision and all evolution is perfectly deterministic.

In this paper we argue that probability arises in Everettian quantum mechanics because observers with perfect knowledge (of their immediate circumstances and the state of the universe as a whole) necessarily evolve into conditions of self-locating uncertainty, in which they do not know which approximately isolated semi-classical world (or 'branch') they inhabit.[2] We propose a general principle, the Epistemic Separability Principle (*ESP*), which captures the idea that predictions made by local agents should be independent of their environment. Given *ESP*, we show that there is a unique rational way for such an agent to assign a credence to each of the quantum worlds they might be in. These credences are precisely the ones recommended by the Born rule. The probabilities are fundamentally *subjective* in the sense that that

[1] Assuming there is a unique eigenstate with eigenvalue a.

[2] The increase in the number of branches over time is a consequence of unitary evolution, not an additional postulate (Wallace, 2003a, 2010a).

they are not written into the laws—as they are in spontaneous collapse theories—but instead capture the degrees of belief of a rational agent; however, they are *objective* in the sense that a rational agent *must* assign Born rule probabilities (if *ESP* is correct).[3]

We believe that recent proofs of the Born rule (Deutsch, 1999; Wallace, 2010c; Zurek, 2005) are on the right track, latching onto symmetries in the theory that serve to explain the validity of the Born rule in Everettian quantum mechanics. However, there are serious objections to the approaches already explored and many remain unconvinced, so we offer this derivation as a novel alternative. We seek to provide an *epistemic*—as opposed to a *decision-theoretic*—derivation of the Born rule which connects quantum uncertainty to the sort of self-locating uncertainty present in very large universes (discussed recently in Hartle & Srednicki, 2007; Page, 2007; Srednicki & Hartle, 2010). This approach shares formal features with Zurek's (2003a; 2003b; 2005) argument based on the idea of envariance, while offering a clearer explanation of the way in which probabilities arise in a deterministic setting.

At first glance, standard treatments of self-locating uncertainty seem to generate a serious problem for the many-worlds interpretation. Elga (2004) has put forward a compelling principle of indifference for cases of self-locating uncertainty, roughly: an observer should give equal credence to any one of a discrete set of locations in the universe that are consistent with the data she has. A simplistic extension to the quantum case would seem to favor treating each world as equiprobable (branch-counting) rather than the Born rule, since Everettian observers on different branches find themselves in situations of identical data. We believe that the reasoning behind Elga's principle, when properly applied to Everettian quantum mechanics, actually leads to the Born rule—not branch-counting.

[3]One who finds the probabilities discussed here insufficiently physical/objective may want to adopt the strategy for moving from unique rational subjective probabilities to objective physical probabilities discussed in (Wallace, 2006).

One common assumption appealed to in determining what probabilities one ought to assign is that certain features of the case are irrelevant, in particular, one can alter the environment in many ways without changing what probability assignments are rational. We formalize this idea by proposing a constraint on rational credences (*ESP*) and showing that, somewhat surprisingly, the principle suffices to derive the Born rule. A careful derivation will be provided after extensive stage-setting in §2.3.2 and appendix 2.C, but the spirit of the approach can be seen in this simplified version of the argument for two equal amplitude branches. Imagine that there are two quantum measurement devices which measure the z-spin of an x-spin up particle, the first displaying either ♡ or ♢, the second either ♣ or ♠ depending on the result of the measurement. The post-measurement state of the detectors can be written as

$$\frac{1}{\sqrt{2}}\left|\heartsuit\right\rangle\left|\clubsuit\right\rangle+\frac{1}{\sqrt{2}}\left|\diamondsuit\right\rangle\left|\spadesuit\right\rangle \quad . \tag{2.1}$$

If we assume there is an experimenter who has yet to observe the outcome, they ought to be uncertain about which of the two branches they are on. The probabilities assigned to the possible states of detector 1 (♡/♢) should be the same if the case is modified so that detector 2 is wired to show the opposite symbols from those in (2.1) as the state of the second detector is irrelevant to to the experimenter's question about their relation to the first detector,

$$\frac{1}{\sqrt{2}}\left|\heartsuit\right\rangle\left|\spadesuit\right\rangle+\frac{1}{\sqrt{2}}\left|\diamondsuit\right\rangle\left|\clubsuit\right\rangle \quad . \tag{2.2}$$

The probability of ♡ in (2.1)—which is the probability of being in the *first* branch of (2.1)—is the same as that of ♡ in (2.2)—which is the probability of being in the *first* branch of (2.2). If instead we focus on assigning probabilities to the possible states of detector 2 (♣/♠), they too should be the same in (2.1) and (2.2). Focusing on

♠ yields the requirement that the probability of being in the *second* branch of (2.1) be the same as the probability of being in the *first* branch of (2.2). And thus the probabilities assigned to each branch of (2.1) must be equal.

In §2.2, we begin by introducing the many-worlds interpretation and discussing how self-locating uncertainty arises during quantum measurements. If we extend standard methods of handling self-locating uncertainty to quantum cases in a simple way, we get the odd recommendation that branches should be weighted equally regardless of their quantum amplitudes. We argue that this cannot be the right way to handle the uncertainty that arises in cases of quantum branching. In §2.3 we provide an alternative principle for handling self-locating uncertainty across physical theories (*ESP*) and show that it yields the Born rule when applied to Everettian quantum mechanics. In §2.4, we show that this principle (when strengthened) yields the result that each copy of oneself should be judged equiprobable in cases of classical self-locating uncertainty. We then use the principle to generate probabilities in cases which involve both uncertainty arising from quantum measurement and from duplication of one's experiences elsewhere in the universe, such as 'quantum sleeping beauty' scenarios and quantum measurements in very large universes. Then, in §2.5 we argue that although in Everettian quantum mechanics there is nothing to be uncertain of before measurement (in idealized cases where the wave function and the dynamics are known), we can still generate rules for rational decision making pre-measurement and theory testing post-measurement. In §2.6 we compare our treatment with the approaches of Zurek, Deutsch, and Wallace.

Before diving in, a choice of terminology. In this article we'll use 'universe' to denote the collection of all branches as described by the universal wave function, reserving 'world' as a way of denoting a particular branch of the wave function, i.e., one of the many worlds of Everettian quantum mechanics. Using this terminology, 'quantum multiverse'='universe.'

2.2 Preliminaries: Many-worlds, Self-locating Uncertainty, and Branch-counting

2.2.1 The Many-worlds Interpretation

If the state of the universe is given by a wave function and that wave function's time evolution is at all times in accordance with the Schrödinger equation, strange things start to happen. Many of us have come to accept that tiny particles can be in superpositions of distinct states, e.g., passing through two slits at once. But, if the wave function always obeys the Schrödinger equation then these microscopic superpositions can be amplified, leading to cases where macroscopic measuring devices and even human beings are in superpositions of distinct states.

ONCE At t_1 Alice has prepared particle a in the x-spin up eigenstate,

$$|\uparrow_x\rangle_a = \frac{1}{\sqrt{2}}\Big(|\uparrow_z\rangle_a + |\downarrow_z\rangle_a \Big) \ . \tag{2.3}$$

She then measures the z-spin of the particle. At t_4, she has just observed the outcome of the experiment. If the particle were in a z-spin eigenstate, Alice and the measuring device would simply record that outcome. But, since the particle is in a superposition of z-spins, Alice and her device evolve into a superposition of different recorded outcomes. The state of Alice (A), her particle (a), the measuring device (D), and the rest of the universe evolve as below. Here R indicates the ready state and $\uparrow/\downarrow$ denote the possible outcomes.

$$\begin{aligned} |\Psi_{\text{ONCE}}(t_1)\rangle &= |R\rangle_A |R\rangle_D |\uparrow_x\rangle_a |E_0\rangle \\ |\Psi_{\text{ONCE}}(t_4)\rangle &= \frac{1}{\sqrt{2}} |\uparrow\rangle_A |\uparrow\rangle_D |\uparrow_z\rangle_a |E_1\rangle + \frac{1}{\sqrt{2}} |\downarrow\rangle_A |\downarrow\rangle_D |\downarrow_z\rangle_a |E_2\rangle \end{aligned} \tag{2.4}$$

In $|\Psi_{\text{ONCE}}(t_4)\rangle$ both Alice and the detector have entered superpositions of recording up and down as the result of the measurement. In practice, just one outcome is observed. One could modify the theory to ensure that only a single outcome actually occurs. This is the strategy taken by Bohmian mechanics, GRW theory, and other venerable responses to the measurement problem. However, such a modification may be unnecessary. According to the many-worlds interpretation, $|\Psi_{\text{ONCE}}(t_4)\rangle$ is a state where there are two (or more) copies of Alice, each of which has just observed the readout of a z-spin measurement. There's a version of Alice who saw up and a version who saw down, but no version that saw both up and down. Seeing a single definite outcome is not precluded by the state evolution given above, it is *guaranteed* since every one of Alice's successors sees a definite outcome.

Even if each copy of Alice has an ordinary experience at t_4, one still might be worried about the fact that there exist parts of the universe in which each possible outcome happens. Particularly, one might hope that after a measurement like this is made and Alice dutifully records up as the result, she can forget about the fact that the particle might have been down for the sake of all her future calculations. But, if the state of the universe is as in (2.4), then the possibility that the result might have been down is no mere possibility, that part of the wave function is still out there and could potentially interfere with Alice's part. If the many-worlds interpretation is to be viable, there must be some reason why Alice can ignore the other worlds in which the experiment turned out differently. Fortunately, there is: decoherence.

As the result of the measurement is recorded in the device and observed by Alice, many traces of this result will appear in the environment. Due to the large number of traces present in the environment, the states E_1 and E_2 are (at least approximately) orthogonal and can be expected to stay orthogonal as time progresses. Because these environment states are and will be orthogonal, the two components of $|\Psi_{\text{ONCE}}(t_4)\rangle$ will evolve in a non-interacting way, each component evolving in accordance with the

Schrödinger equation as if the other were not present. They can thus be treated as separate worlds (or collections of worlds) since they are effectively causally isolated and within each of them there are versions of Alice having clear and determinate experiences.[4]

In quantum mechanics, the state of a subsystem is given by a reduced density matrix, generated by taking a partial trace over the Hilbert space of the environment (Schlosshauer, 2007, §2.4; Nielsen & Chuang, 2010, §2.4). In ordinary cases of quantum measurement, the process of decoherence diagonalizes the reduced density matrix for the macroscopic degrees of freedom describing our observer and measuring apparatus in a very specific basis, the so-called pointer basis (Zurek, 1981, 1982, 1993, 2003a). The pointer basis is distinguished by being robust with respect to ordinary interactions with the environment.

The state of Alice, the detector, and particle a at t_1 can be written as a reduced density matrix by tracing out the environment,

$$\begin{aligned}\widehat{\rho}_{ADa}(t_1) &= \mathrm{Tr}_E\Big(\left|\Psi_{\textsc{Once}}(t_1)\right\rangle \left\langle\Psi_{\textsc{Once}}(t_1)\right| \Big) \\ &= \frac{1}{2}\left|R\right\rangle_A \left|R\right\rangle_D \left|\uparrow_z\right\rangle_a \left\langle\uparrow_z\right|_a \left\langle R\right|_D \left\langle R\right|_A + \frac{1}{2}\left|R\right\rangle_A \left|R\right\rangle_D \left|\uparrow_z\right\rangle_a \left\langle\downarrow_z\right|_a \left\langle R\right|_D \left\langle R\right|_A \\ &\quad + \frac{1}{2}\left|R\right\rangle_A \left|R\right\rangle_D \left|\downarrow_z\right\rangle_a \left\langle\uparrow_z\right|_a \left\langle R\right|_D \left\langle R\right|_A + \frac{1}{2}\left|R\right\rangle_A \left|R\right\rangle_D \left|\downarrow_z\right\rangle_a \left\langle\downarrow_z\right|_a \left\langle R\right|_D \left\langle R\right|_A \ . \end{aligned} \tag{2.5}$$

As the z-spin of the particle has yet to be entangled with the measuring device and the environment, the diagonal terms of the reduced density matrix (the middle two terms of the sum in the second line of (2.5)) are non-zero and potentially very relevant to the evolution of the system—e.g., if the x-spin of the particle were to be measured.

[4]For more on how Everettian branches arise from decoherence, see (Schlosshauer, 2005; Saunders, 2010a, §1.3) and references therein. One might object that this sort of story about the emergence of separate worlds cannot be assumed in a derivation of the Born rule. See appendix 2.B.

After the measurement is made and the environment gets entangled with the particle's spin, the reduced density matrix becomes

$$\begin{aligned}\widehat{\rho}_{ADa}(t_4) = & \tfrac{1}{2}\langle E_1|E_1\rangle |\uparrow\rangle_A |\uparrow\rangle_D |\uparrow_z\rangle_a \langle\uparrow_z|_a \langle\uparrow|_D \langle\uparrow|_A + \tfrac{1}{2}\langle E_1|E_2\rangle |\uparrow\rangle_A |\uparrow\rangle_D |\uparrow_z\rangle_a \langle\downarrow_z|_a \langle\downarrow|_D \langle\downarrow|_A \\ & + \tfrac{1}{2}\langle E_2|E_1\rangle |\downarrow\rangle_A |\downarrow\rangle_D |\downarrow_z\rangle_a \langle\uparrow_z|_a \langle\uparrow|_D \langle\uparrow|_A + \tfrac{1}{2}\langle E_2|E_2\rangle |\downarrow\rangle_A |\downarrow\rangle_D |\downarrow_z\rangle_a \langle\downarrow_z|_a \langle\downarrow|_D \langle\downarrow|_A \\ = & \tfrac{1}{2} |\uparrow\rangle_A |\uparrow\rangle_D |\uparrow_z\rangle_a \langle\uparrow_z|_a \langle\uparrow|_D \langle\uparrow|_A + \tfrac{1}{2} |\downarrow\rangle_A |\downarrow\rangle_D |\downarrow_z\rangle_a \langle\downarrow_z|_a \langle\downarrow|_D \langle\downarrow|_A \ . \end{aligned} \tag{2.6}$$

Here we've assumed for simplicity that E_1 and E_2 are perfectly orthogonal, $\langle E_1|E_2\rangle = 0$, and the diagonal terms vanish; in realistic cases $\langle E_1|E_2\rangle$ is very close to zero. The diagonal terms have dropped out and the state is now decomposable into two parts—branches—of the reduced density matrix, one in which the result was up and a second in which it was down. It is common to speak of the universal *wave function* splitting into several branches upon measurement, components that will no longer interact because of decoherence.[5] Similarly, one can just as easily speak of the *density matrix* for the universe as being decomposable into branches. *Reduced* density matrices can also evolve into pieces that because of decoherence will no longer interact. We call these pieces 'branches' too, although this is a non-standard use of the term 'branch.' Branches of a reduced density matrix should not be confused with branches of the universal wave function. A branch of a reduced density matrix may correspond to many branches of the universal wave function, as occurs in the possible elaboration of Once below.

Once-or-Twice Alice's particle (a) and Bob's particle (b) are both initially prepared in the x-spin up eigenstate. Alice's device measures the z-spin of her particle first. Then, Bob's device, which is connected to Alice's, measures z-

[5] Although we will use the language of 'branching' throughout the paper, we do not mean by this to prejudge the issue of whether these branches diverge or overlap (discussed in Saunders, 2010b; Wilson, 2012). Our strategy can be implemented on either picture (see footnote 29).

spin of particle b *only if* particle a was measured to have z-spin up. By t_1 the setup is prepared; by t_2 Alice's particle has been measured but Bob's has not; by t_3 both particles have been measured. Bob has been watching as the results of the experiments are recorded. Up through t_3, Alice has not looked at the measuring devices and is unaware of the results. By t_4, Alice has looked at her device and seen the result of the measurement of particle a, although she remains ignorant about the z-spin of particle b. The branching structure of this scenario is shown in figure 2.1. The evolution of the state is shown below. Here Bob (B) and his particle (b) have been pulled out of the environment, but otherwise the state is broken down as in (2.4) (Bob's device is treated as part of the environment; X indicates that the measurement was not made).

$$
\begin{aligned}
|\Psi_{\mathrm{OT}}(t_1)\rangle = {} & |R\rangle_A |R\rangle_D |{\uparrow_x}\rangle_a |R\rangle_B |{\uparrow_x}\rangle_b |E_R\rangle \\
|\Psi_{\mathrm{OT}}(t_2)\rangle = {} & \tfrac{1}{\sqrt{2}} |R\rangle_A |{\uparrow}\rangle_D |{\uparrow_z}\rangle_a |{\uparrow}\rangle_B |{\uparrow_x}\rangle_b |E_{\uparrow}\rangle + \tfrac{1}{\sqrt{2}} |R\rangle_A |{\downarrow}\rangle_D |{\downarrow_z}\rangle_a |{\downarrow}\rangle_B |{\uparrow_x}\rangle_b |E_{\downarrow}\rangle \\
|\Psi_{\mathrm{OT}}(t_3)\rangle = {} & \tfrac{1}{2} |R\rangle_A |{\uparrow}\rangle_D |{\uparrow_z}\rangle_a |{\uparrow},{\uparrow}\rangle_B |{\uparrow_z}\rangle_b |E_{\uparrow\uparrow}\rangle + \tfrac{1}{2} |R\rangle_A |{\uparrow}\rangle_D |{\uparrow_z}\rangle_a |{\uparrow},{\downarrow}\rangle_B |{\downarrow_z}\rangle_b |E_{\uparrow\downarrow}\rangle \\
& + \tfrac{1}{\sqrt{2}} |R\rangle_A |{\downarrow}\rangle_D |{\downarrow_z}\rangle_a |{\downarrow},\mathrm{X}\rangle_B |{\uparrow_x}\rangle_b |E_{\downarrow\mathrm{X}}\rangle \\
|\Psi_{\mathrm{OT}}(t_4)\rangle = {} & \tfrac{1}{2} |{\uparrow}\rangle_A |{\uparrow}\rangle_D |{\uparrow_z}\rangle_a |{\uparrow},{\uparrow}\rangle_B |{\uparrow_z}\rangle_b \left|E'_{\uparrow\uparrow}\right\rangle + \tfrac{1}{2} |{\uparrow}\rangle_A |{\uparrow}\rangle_D |{\uparrow_z}\rangle_a |{\uparrow},{\downarrow}\rangle_B |{\downarrow_z}\rangle_b \left|E'_{\uparrow\downarrow}\right\rangle \\
& + \tfrac{1}{\sqrt{2}} |{\downarrow}\rangle_A |{\downarrow}\rangle_D |{\downarrow_z}\rangle_a |{\downarrow},\mathrm{X}\rangle_B |{\uparrow_x}\rangle_b \left|E'_{\downarrow\mathrm{X}}\right\rangle
\end{aligned}
\tag{2.7}
$$

If we focus on the state of Alice, the detector, and particle a at t_4, the reduced density matrix for Once-Or-Twice will be exactly as it was in Once (2.6). However, this is clearly a case where there are at least three branches of the universal wave function corresponding to the three different possible outcomes of the series of experiments as recorded by Bob. In this case, the branch of $\widehat{\rho}_{ADa}$ in which Alice observes up corresponds to at least two branches of the universal wave function, as there are two possible outcomes Bob might have observed.

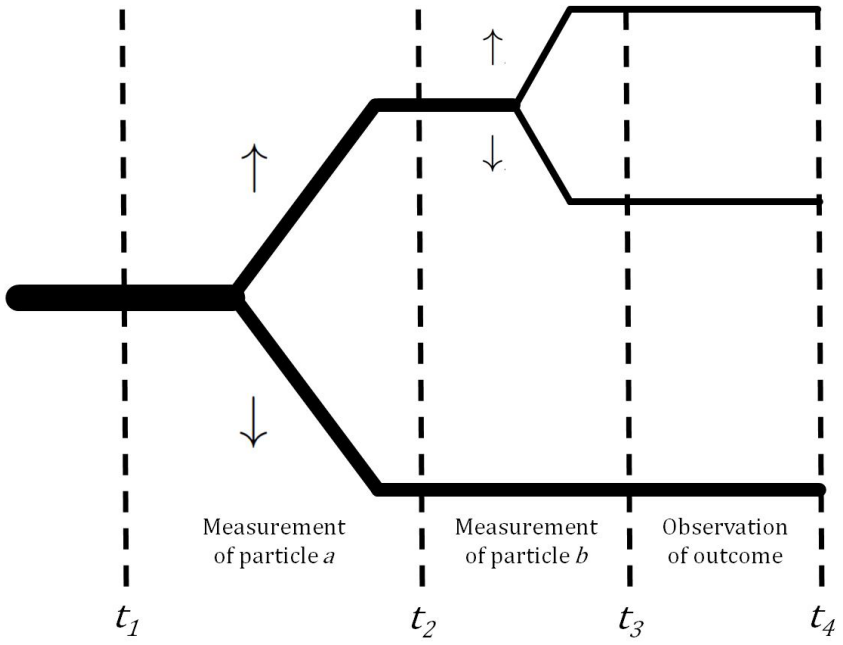

Figure 2.1: **Once-Or-Twice**

2.2.2 Self-locating Uncertainty and the Everettian Multiverse

On many interpretations of quantum mechanics, quantum measurements yield just one result. Before a measurement is made, the experimenter is typically uncertain about which outcome will actually occur, even if he knows as much about the universal wave function as could be helpful. This uncertainty can be quantified using the Born rule. In the many-worlds interpretation, if we assume that the experimenter knows the relevant information about the wave function, it is unclear what the agent might be uncertain of before a measurement is made. They know that every outcome will occur and that they will have a successor who sees each possible result. Before we can discuss the quantitative problem of what numerical probabilities an agent in an Everettian multiverse should assign to different outcomes, we must answer the

question: What can one assign probabilities to? Our answer will be that agents performing measurements pass through a period of self-locating uncertainty, in which they can assign probabilities to being one of several identical copies, each on a different branch of the wave function.

Consider Alice in Once-or-Twice. At time t_2 the state can be written in either of two equivalent forms,

$$|\Psi_{\mathrm{OT}}(t_2)\rangle = \frac{1}{\sqrt{2}} |R\rangle_A \Big(|\uparrow\rangle_D |\uparrow_z\rangle_a |\uparrow\rangle_B |\uparrow_x\rangle_b |E_\uparrow\rangle + |\downarrow\rangle_D |\downarrow_z\rangle_a |\downarrow\rangle_B |\uparrow_x\rangle_b |E_\downarrow\rangle \Big) \tag{2.8}$$

$$= \frac{1}{\sqrt{2}} |R\rangle_A |\uparrow\rangle_D |\uparrow_z\rangle_a |\uparrow\rangle_B |\uparrow_x\rangle_b |E_\uparrow\rangle + \frac{1}{\sqrt{2}} |R\rangle_A |\downarrow\rangle_D |\downarrow_z\rangle_a |\downarrow\rangle_B |\uparrow_x\rangle_b |E_\downarrow\rangle \,. \tag{2.9}$$

In the first way of writing the state, it appears as if there is only one copy of Alice (represented by $|R\rangle_A$), while the rest of the state has branched in two. But in the second expression, it appears as if there are two branches with two identical copies of Alice. (2.8) and (2.9) are manifestly mathematically identical, but suggest different physical pictures, corresponding to two plausible attitudes about wave function branching. One attitude would be that Alice herself doesn't branch until her own reduced density matrix splits into multiple distinct branches (in the pointer basis) sometime between t_3 and t_4; in that case (2.8) gives the right physical picture, and there is only one copy of Alice even after the first measurement has occurred. The other attitude is that branching happens throughout the whole wave function whenever it happens anywhere. When the universal wave function splits into multiple distinct and effectively non-interacting parts, the entire world splits—along with every object and agent in it. Then (2.9) is the more perspicuous way of writing the state. There are two identical post-measurement copies of Alice, one on each branch.[6]

[6] Wallace (2012, §8.6) advocates a version of the first picture and Vaidman (2014a, 2014b, forth-

We will work under the assumption that the second attitude (branching happens globally throughout the wave function whenever it happens anywhere) is the right way to think about the process.[7] The branching structure of the wave function in Everettian quantum mechanics is not uniquely defined by the theory, which is just a state space and a dynamical law; like coarse-graining in classical statistical mechanics, it is imposed by us with a degree of arbitrariness. We must decide how well-separated two putative branches must be to count as distinct. However, treating branching on a subsystem-by-subsystem basis—as advised by the first attitude—is needlessly arbitrary; it depends on how the universe is carved into subsystems. Some very small piece of a post-measurement detector might remain in the same state on every branch, but it seems wrong to think that whether the single piece has branched into many depends on whether it is considered as part of the detector or as a separate subsystem in its own right. For Alice's purposes, it makes sense to distinguish between the two copies of herself in (2.9); even though they are currently unentangled with the rest of the state, in the future they will evolve differently (and independently of one another) as Alice becomes aware of the measurement outcome.

The non-local nature of the globally-branching view might cause some discomfort. It implies that observers here on Earth could be (and almost surely are) branching all the time, without noticing it, due to quantum evolution of systems in the Andromeda Galaxy and elsewhere throughout the universe.[8] We take this to be one among many psychologically unintuitive but empirically benign consequences of Everettian

coming) the second.

[7] In our derivation of the Born rule, §2.3.2, this second attitude plays an important role in allowing us to examine cases like (2.17) where the agent is in a physically identical state on each branch.

[8] However, the reduced density matrix describing the state of the observer on Earth is unaffected by the events occurring in the Andromeda galaxy. In a relativistic context, the number of copies of an observer will be frame-relative. But, the probabilities assigned to measurement outcomes will not be (as the reduced density matrix is all that is relevant to the calculation of probabilities, see §2.3.1).

quantum mechanics.

There are two copies of Alice at t_2 in Once-or-Twice. Each copy can reasonably wonder which one she is. Thus even if she (incredibly) knows the universal wave function exactly, Alice still has something to be uncertain of. She isn't uncertain about the way the universe is; by supposition she knows the wave function and this gives a complete specification of the state of the universe. Alice is uncertain about where she is in the quantum multiverse (as has been emphasized by Vaidman 1998, 2011, 2014a, 2014b). She doesn't know if she's in the branch of the wave function in which the detector displays up or the one in which it shows down. Alice has *self-locating uncertainty* (see Lewis, 1979; Bostrom, 2002). We call this period in which self-locating uncertainty is present, after the measurement has been made and branching has occurred via decoherence but before the experimenter has registered the result, the 'post-measurement pre-observation' period.

During the post-measurement pre-observation there are multiple copies of Alice seeing different results. Are any or all of them the same person as Alice before the measurement? This is a tricky metaphysical question upon which we will not speculate. We will indiscriminately refer to a post-branching version of Alice as a 'copy,' or a 'successor,' or simply 'Alice.' In appealing to self-locating uncertainty during the post-measurement pre-observation period, all that is needed is that each copy of Alice can coherently wonder what sort of branch they inhabit.

At this point it may be objected that Once-or-Twice is highly atypical as Alice is for a time unaware of the measurement result. Actually, self-locating uncertainty is generic in quantum measurement. In Everettian quantum mechanics the wave function branches when the system becomes sufficiently entangled with the environment to produce decoherence. The normal case is one in which the quantum system interacts with an experimental apparatus (cloud chamber, Geiger counter, electron microscope, or what have you) and then the observer sees what the apparatus has

recorded. For any realistic room-temperature experimental apparatus, the decoherence time is extremely short: less than 10^{-20} seconds. Even if a human observer looks at the quantum system directly, the state of the observer's eyeballs will decohere in a comparable time. In contrast, the time it takes a brain to process a thought is measured in tens of milliseconds. No matter what we do, real observers will find themselves in a situation of self-locating uncertainty (after decoherence, before the measurement outcome has been registered).[9] The observer may not have enough time to think and reorganize their credences before learning the outcome (Wallace, 2006, §4.2), but, in trying to approximate ideal rationality, the agent can attempt to reconstruct the probabilities they ought to have assigned during the post-measurement pre-observation period. Despite often being short-lived, these probabilities are relevant for deciding what to believe after seeing the measurement outcome, §2.5.2, and deciding how to act in anticipation of the measurement, §2.5.1.

Self-locating uncertainty has been discussed in great depth by formal epistemologists. The general question of how to extend Bayesian updating to cases involving self-locating uncertainty is very difficult and a matter of contemporary philosophical debate, but many different proposals[10] agree on one minimal constraint on rational credences (a.k.a. subjective probabilities or degrees of belief):

Indifference Suppose that in a given possible universe U, agent S has a finite number N_U of copies in an internally qualitatively identical state to S. Further, suppose that the hypothesis H is true for N_{UH} of these copies. Then, S should assign

[9] Although we think that self-locating uncertainty is typical, it may not be necessary for our account that it is always present. See the end of §2.5.2.

[10] In particular, we're thinking of the proposals of (Bostrom, 2002; Meacham, 2008; Manley, unpublished) as formulated in (Manley, unpublished).

the following conditional credence to H being true given that S is in U:

$$P(H|U) = \frac{N_{UH}}{N_U} \tag{2.10}$$

Indifference was originally put forward in (Elga, 2004); the version here is from (Manley, unpublished) (paraphrased). Two agents are in the same *internal qualitative state* if they have identical current evidence: the patterns of colors in their visual fields are identical, they recall the same apparent memories, they both feel equally hungry, etc. This principle should not be confused with a Laplacean 'principle of indifference.'

To see *Indifference* in action, consider the following case from Elga (2004):

Duplicating Dr. Evil Dr. Evil is plotting the destruction of Earth from his lunar battle station when he receives an unwelcome message. Back on earth some pesky philosophers have duplicated the entirety of his battle station, perfectly replicating every piece of furniture, every weapon, and every piece of food, even replicating the stale moon air and somehow the weaker gravitational field. They went so far that at t they created a duplicate of Dr. Evil himself, Dup. Dup's immediate experiences are internally identical to Dr. Evil's. The two men pace, think, and scratch themselves in perfect synchronicity.

Upon learning of his duplication, what credence should Dr. Evil assign to being in his lunar base and not the terrestrial fake? According to *Indifference*, Dr. Evil should assign a credence of one half (and Dup should do the same). Dr. Evil knows that in any universe U he might inhabit there are two copies of himself ($N_U = 2$) and that one is on the moon ($N_{UH} = 1$ for $H =$ 'I am on the moon.').

2.2.3 Indifference and the Quantitative Probability Problem

In the previous section we showed that in the post-measurement pre-observation period it is possible to assign probabilities to the different possible outcomes of a completed experiment. But, what probabilities ought one to assign? If you take *Indifference* to be a constraint on rational probability assignments and count agents in the most naive way, there appears to be a simple rule for assigning probabilities: branch-counting.[11] The probability that H is true is given by the fraction of branches in which H holds. Since there is a single copy of the experimenter on each branch, the fraction of branches is the same as the fraction of agents and branch-counting is a direct consequence of *Indifference*.

There is a well-known problem for this kind of approach.[12] The number of branches in which a certain outcome occurs might not be well-defined. Branches are structures in the wave function which emerge via decoherence, and although this process will guarantee that on each branch a definite outcome is observed, it will generally not yield a good answer to the question of how many branches feature a certain outcome or even how many branches there are in total. In contrast, the total mod-squared amplitude of branches in which a certain outcome occurs is well-defined (although not to arbitrary precision). The difficulty of branch-counting provides good reason to think *Indifference* will often be unhelpful in Everettian scenarios. However, *Indifference* does give clear recommendations in idealized scenarios, so for the time being we'll restrict our attention to cases where the number of branches is well-defined.

Assume that Once-or-Twice really is a case in which one branch becomes three, as depicted in figure 2.1. At t_3, what probability should Alice assign to particle a having been measured to be z-spin down? By the methods of textbook quantum

[11] Lewis (2009a) considers a proposal along these lines.
[12] For more in-depth discussion, see Wallace (2007, §9, 2012, §3.11).

mechanics, we should use the Born rule to calculate the probability. That is, we should square the coefficient of the third branch of the wave function in (2.7),

$$P_{t_3}(\text{down}) = \left(\frac{1}{\sqrt{2}}\right)^2 = \frac{1}{2} \ . \tag{2.11}$$

To apply *Indifference* we simply need to determine the fraction of branches in which the result of the first measurement was down,

$$P_{t_3}(\text{down}) = \frac{\text{number of down branches}}{\text{number of branches}} = \frac{1}{3} \ , \tag{2.12}$$

the coefficients of $1/2$, $1/2$, and $1/\sqrt{2}$ of the branches in $|\Psi_{\text{OT}}(t_3)\rangle$ are irrelevant to the calculation. Supposing *Indifference* is true and people are to be counted in this straightforward way, textbook quantum mechanics and Everettian quantum mechanics make different predictions about the probability of seeing down. Thus, we can empirically test the many-worlds interpretation in any case where the fraction of branches in which some hypothesis H is true is unequal to the combined weights of branches in which H is true (the generic case).

We now have the strongest kind of argument against Everettian quantum mechanics: it is empirically inadequate. The theory fails to reproduce the empirical predictions of textbook quantum mechanics and thus has been conclusively falsified by the data typically taken as evidence for quantum mechanics.[13] However, this argument relies on the assumption that *Indifference* is a universally valid principle applicable to Everettian multiverses as well as cases of classical duplication like Duplicating Dr. Evil. As we will see, it is not.

[13] Admittedly, the force of this argument is weakened by the fact that we are working under an idealization. Branch number may not be well-defined in realistic cases and then it becomes unclear what probabilities the proponent of *Indifference* thinks one should assign.

2.2.4 Against Branch-counting

When *Indifference* is applied to Once-or-Twice, something odd happens. At t_2 the principle advises each copy of Alice to assign a probability of one half to particle a being down as half of the copies of Alice are in a branch where the particle was measured to be down. Then, at t_3 Alice is supposed to assign down a probability of one third (as in (2.12)). According to the Born rule, the recommended probability assignment is one half at both t_2 and t_3. If *Indifference* is right, there's a strange switch in the probabilities between t_2 and t_3. Is there any reason to think this undermines the branch-counting strategy advocated by *Indifference*? David Wallace has argued that such a switch violates a constraint he calls 'diachronic consistency.' In appendix 2.A, we argue that this is not the right diagnosis of the problem with the switch in credences. This kind of 'inconsistency' is a common result of *Indifference* and not something which should be taken to refute the principle. Still, we agree that there's something wrong with the probability switch.

Between t_2 and t_3 what happens? Particle b is measured and Bob takes note of the result. Nothing happens to Alice, particle a, or Alice's device. If nothing about Alice or her detector changes, why should her degree of belief that she bears a certain relation to the detector change? At t_2, the state of Alice and her detector is given by,

$$\widehat{\rho}_{AD}(t_2) = \frac{1}{2}\left|R\right\rangle_A\left|\uparrow\right\rangle_D\left\langle\uparrow\right|_D\left\langle R\right|_A + \frac{1}{2}\left|R\right\rangle_A\left|\downarrow\right\rangle_D\left\langle\downarrow\right|_D\left\langle R\right|_A \ . \tag{2.13}$$

At this time Alice is trying to locate herself in one of the two branches of this Alice+Detector system and, if she applies *Indifference*, decides that she should assign a probability of one half to each branch. At t_3, the state of the Alice+Detector system is unchanged, it is still exactly as in (2.13). Now she decides that the probability of being in the second branch is one third. Why should her probability for being in

different subsystems (branches of $\widehat{\rho}_{AD}$) of the Alice+Detector system change when nothing about that system changes and she knows that she is somewhere in that system? It shouldn't. In the next section we will use this insight to motivate a replacement for *Indifference*.

It is tempting to think that the number of copies of Alice cannot change without her physical state changing—this is the way things work in classical physics. But, in Everettian quantum mechanics, changes that purely affect her environment can change the number of copies of Alice in existence. For example, the change of state from t_2 to t_3 in Once-or-Twice, (2.7). Two intuitive constraints come into conflict: *Indifference*, and the belief that Alice's probabilities should be unaffected by changes in the state of her environment. We recommend rejecting the former in favor of the latter.

2.3 The Epistemic Irrelevance of the Environment

2.3.1 The Epistemic Separability Principle

In §2.2.4 we argued that branch-counting is unreasonable because it requires Alice to change her credences about the result of the measurement of particle a when things change elsewhere in the universe. In particular, when particle b is measured. In the Duplicating Dr. Evil case, whatever one thinks of *Indifference*, it seems clear that certain facts are irrelevant. The probability $P(\text{I'm Dr. Evil}|U)$ should not depend on what's happening deep inside the Earth's core or what's happening on the distant planet Neptune or any other remote occurrences. Unless, of course, the actions of the Earth's core cause earthquakes which the terrestrial Dup feels but the

lunar Dr. Evil does not. Or, if what's happening on Neptune includes *another* copy of the laboratory with another duplicate Dr. Evil in it. If there's a duplicate on Neptune, Dr.Evil can no longer be sure that Neptune is in fact a distant planet and not the one under his feet (and thus cannot treat it as irrelevant to his probability assignments). As long as the copies of Dr. Evil are unaffected and no new copies are created elsewhere, $P(\text{I'm Dr. Evil}|U)$ should be unaffected by changes to the environment. This thought, which was essential to our argument in §2.2.4, can be stated in imprecise slogan form as:

ESP-gist The credence one should assign to being any one of several observers having identical experiences is independent of the state of the environment.

In more precise but still theory-independent language, this epistemic principle becomes:

ESP Suppose that universe U contains within it a set of subsystems $\mathcal{S}$ such that every agent in an internally qualitatively identical state to agent A is located in some subsystem which is an element of $\mathcal{S}$. The probability that A ought to assign to being located in a particular subsystem $X \in \mathcal{S}$ given that they are in U is identical in any possible universe which also contains subsystems $\mathcal{S}$ in the same exact states (and does not contain any copies of the agent in an internally qualitatively identical state that are not located in $\mathcal{S}$).

$$P(X|U) = P(X|\mathcal{S}) \tag{2.14}$$

ESP stands for 'Epistemic Separability Principle,' as the principle allows one to separate the relevant parts of the universe from the rest; that is, to separate the set of subsystems where the agent might, for all they know, be located from everything else. X, U, and $\mathcal{S}$ are not propositions; (2.14) is shorthand that must be clarified.

$P(X|U)$ is the probability that A assigns to being *in* X given that they are *in* U. In $P(X|\mathcal{S})$, '$\mathcal{S}$' is shorthand for: there exist subsystems $\mathcal{S}$ and there are no internally qualitatively identical copies of A outside of these subsystems. To be precise, let's say that one is 'located in' subsystem X just in case one is a proper part of the subsystem (no limb or brain cell is omitted). There may be more than one copy of A in a given subsystem. In general, There will be many ways to carve out a set of subsystems from the universe. *ESP* applies to any such carving. The subsystems in $\mathcal{S}$ together need not cover the entire world, there may well be parts of U that are not in any of these subsystems, however these omitted parts cannot contain copies of A. The principle is restricted to cases where $\mathcal{S}$ has a finite number of members. The subsystems need not be located at the same time. One may, for example, be unsure if they are the person waking up in their bed on Monday or Tuesday (see §2.4.2).

We need to be somewhat careful about what exactly a 'subsystem' is supposed to be, although we will not attempt to give a rigorous characterization applicable to any conceivable theory. The essential idea is that a subsystem is a part of the larger system that can be considered as a physical system in its own right. Slightly more formally, we imagine that the overall state of a system can be decomposed into the states of various subsystems, so that two constraints are satisfied: (1) the state of each subsystem, perhaps with some additional information about how the subsystems are connected, can be used to uniquely reconstruct the original state; and (2) the information contained within each subsystem's state is enough to specify its immediate dynamical evolution, as long as the other subsystems are not influencing it.[14] So, in classical particle physics, a subsystem might be a collection of particles at a time and its state specified by giving the masses, positions, and velocities of the particles in

[14]Work would need to be done to formulate these constraints precisely, here they serve as rough guides.

the collection. By contrast, giving only the positions (or only the velocities) of a collection of particles would not specify a subsystem, since that wouldn't be sufficient to determine the collection's evolution. In quantum mechanics, systems can be divided into subsystems in two fundamentally different ways—quite unlike classical physics. The state of a system can be written as a density matrix $\widehat{\rho}_{sys}$ corresponding to some factor of the Hilbert space $\mathscr{H}_{sys}$ (a *reduced* density matrix unless the system is the whole universe). As $\mathscr{H}_{sys}$ can itself be decomposed into factors, $\mathscr{H}_{sys} = \mathscr{H}_1 \otimes \mathscr{H}_2$, one can treat the reduced density matrices $\widehat{\rho}_1$ and $\widehat{\rho}_2$ as subsystems of $\widehat{\rho}_{sys}$. Alternatively, if—as in (2.13)—branching has occurred, the different branches of $\widehat{\rho}_{sys}$ can be regarded as separate subsystems.[15]

Examination of the argument given by Elga (2004) for *Indifference* reveals that something like *ESP* is taken for granted. In his Toss&Duplication thought experiment, Elga assumes that the outcome of an additional coin toss should not affect the credence we assign to being either the original or a duplicated person with identical experiences; the justification for such an assumption would have to be something like *ESP*. *ESP* is compatible with *Indifference* in standard cases of classical self-locating uncertainty like Duplicating Dr. Evil. Requiring that all one care about in assigning credences between Dr. Evil and Dup is what's happening in the lunar laboratory, X, and the terrestrial replica, Y, (together, $\mathcal{S} = \{X, Y\}$) looks like it allows any assignment of credences to the two copies at all, provided one is consistent across universes that vary only in the state of the world outside the two laboratories. Actually, in §2.4.1 we'll see that although *ESP* is compatible with *Indifference*, the rule is not as permissive as it might initially seem. At least, not if we strengthen it as in §2.4.1. In classical cases, the strengthened principle requires one to consider each

[15] In specifying the state of each branch of a reduced density matrix like (2.13) one must retain the numerical prefactors before each term as they would be necessary to reconstruct the state of the system as a whole.

copy of oneself to be equiprobable—in agreement with *Indifference*.

Now that *ESP* is on the table, let's apply it to cases of measurement in Everettian quantum mechanics. Consider Once-or-Twice. At t_2, after particle a has been measured, Alice has branched into multiple copies. Alice knows that she is somewhere in the Alice+Detector system which is characterized by the reduced density matrix in (2.13). The density matrix $\widehat{\rho}_{AD}$ itself can be divided into branches X & Y corresponding to the particle spin being measured as either up or down. Any universal wave function which agrees on the density matrix $\widehat{\rho}_{AD}$ agrees on the state of X and Y. So, the probability of being in X, which is the probability of an up result, must be the same in any universe with the same reduced density matrix $\widehat{\rho}_{AD}$.

More generally, suppose that an experimenter A has just measured observable $\widehat{O}$ of system S and the measuring device has recorded some eigenvalue O_i on each branch of the wave function.[16] As discussed in §2.2.1, the reduced density matrix, $\widehat{\rho}_{AD}$, will be diagonalized in the pointer basis for the Agent+Detector subspace, $\mathscr{H}_A \otimes \mathscr{H}_D$, by the decoherence process. Each pointer state (with nonzero amplitude) defines a branch of the reduced density matrix, $\widehat{\rho}_{AD}$, on which the detector D registered a particular outcome O_i.[17] In assigning credences to the different outcomes, the agent is assigning probabilities to being located in these different branches. Specifying the state of $\widehat{\rho}_{AD}$ determines the state of all its branches and thus of all the subsystems in which the agent might find themselves ($\mathcal{S}$).

[16] *ESP* is not *only* applicable to Everettian quantum mechanics in cases of measurement. It can be applied whenever one is trying to locate oneself within a collection of decohered branches of the wave function. For example, this decoherence might have been caused by chaotic processes instead of quantum measurement (Wallace, 2012, ch. 3).

[17] After a measurement, the branching structure of $\widehat{\rho}_{AD}$—the set of pointer states—can be derived either by examining the interaction with the environment or by seeing in which basis the matrix is diagonal. One need not know anything about the state of the environment; the reduced density matrix alone is sufficient.

Thus in ordinary cases of quantum measurement, we can formulate a less general version of *ESP* which will be sufficient for our derivation of the Born rule.

ESP-QM Suppose that an experiment has just measured observable $\widehat{O}$ of system S and registered some eigenvalue O_i on each branch of the wave function. The probability that agent A ought to assign to the detector D having registered O_i when the universal wave function is Ψ, $P(O_i|\Psi)$, only depends on the reduced density matrix of A and D, $\widehat{\rho}_{AD}$:

$$P(O_i|\Psi) = P(O_i|\widehat{\rho}_{AD}) \tag{2.15}$$

This principle tells us that when observers assign probabilities to recorded outcomes of measurements that have already occurred, these probabilities should only depend on the Agent+Detector state, $\widehat{\rho}_{AD}$ (not on other features of the universal wave function). By applying this principle to quantum cases instead of *Indifference*, we are now able to shake the unrealistic assumption that the number of branches in which a certain outcome occurs is well-defined.

In formulating *ESP-QM*, we've relied on the fact that the state of a quantum subsystem is specified by a reduced density matrix. Although this is the standard way of representing subsystems in quantum mechanics, one might worry that its use here requires further justification. We discuss this concern in appendix 2.B.

2.3.2 Deriving the Born Rule

In this section we will derive the Born rule probabilities as the rational assignment of credences post-measurement pre-observation. We will first derive the rule in a case with two branches that have equal amplitudes, then use similar techniques to treat a case with two branches of unequal amplitude. It is straightforward to extend these

methods to more general cases (see appendix 2.C). Mathematically, our argument is most similar to that of Zurek (2005) and not far from those of the decision theoretic approach (Deutsch, 1999; Wallace, 2003b, 2010c, 2012). The interest of this proof is not its mathematical ingenuity but the facts that (a) it applies to cases where uncertainty is undeniably present and (b) it is based on a single well-motivated principle of rationality, *ESP*.

Proof of the Born rule for $\frac{1}{2}/\frac{1}{2}$ case: Alice measures the z-spin of a single particle in the x-spin up state. One display ($D1$) will show the result of the experiment. If the spin is up, a second display ($D2$) will show $\heartsuit$. If it is down, a $\diamondsuit$ will appear on the second display. Alice is not immediately affected by the result; in particular, she is for a time unaware of the experiment's outcome.[18] The wave function of Alice, the detectors, the particle, and the environment (the rest of the universe) evolves from

$$|\Psi_0\rangle = |R_0\rangle_A |R\rangle_{D1} |R\rangle_{D2} |\uparrow_x\rangle |E_R\rangle \tag{2.16}$$

to

$$|\Psi_1\rangle = \frac{1}{\sqrt{2}} |R\rangle_A |\uparrow\rangle_{D1} |\heartsuit\rangle_{D2} |\uparrow_z\rangle |E_{\uparrow\heartsuit}\rangle + \frac{1}{\sqrt{2}} |R\rangle_A |\downarrow\rangle_{D1} |\diamondsuit\rangle_{D2} |\downarrow_z\rangle |E_{\downarrow\diamondsuit}\rangle \ . \tag{2.17}$$

To use *ESP-QM* to demonstrate that $P(\uparrow|\Psi_1) = P(\downarrow|\Psi_1) = 1/2$, we will need to also consider an *alternate scenario* where the computer (part of the environment) is programmed differently so that $\heartsuit$ displays if down is measured and $\diamondsuit$ displays if up.

[18] That is, we assume that Alice is in the same physical state on both branches of the wave function ($|R\rangle_A$ in (2.17)). As time passes and the result of the measurement has effects on various parts of the universe, Alice's state will come to differ on the two branches (e.g., if she observes the result). But, immediately after measurement her state will be unaffected.

Then the post-measurement pre-observation wave function would be

$$|\Psi_2\rangle = \frac{1}{\sqrt{2}} |R\rangle_A |\uparrow\rangle_{D1} |\diamondsuit\rangle_{D2} |\uparrow_z\rangle |E_{\uparrow\diamondsuit}\rangle + \frac{1}{\sqrt{2}} |R\rangle_A |\downarrow\rangle_{D1} |\heartsuit\rangle_{D2} |\downarrow_z\rangle |E_{\downarrow\heartsuit}\rangle \ . \quad (2.18)$$

Step 1: Focus first on Alice and $D1$. The Alice+Detector 1 reduced density matrices for Ψ_1 and Ψ_2 are the same,[19]

$$\widehat{\rho}_{AD1}(\Psi_1) = \widehat{\rho}_{AD1}(\Psi_2) = \frac{1}{2} |R\rangle_A |\uparrow\rangle_{D1} \langle R|_A \langle\uparrow|_{D1} + \frac{1}{2} |R\rangle_A |\downarrow\rangle_{D1} \langle R|_A \langle\downarrow|_{D1} \ . \quad (2.19)$$

ESP-QM requires that the probabilities Alice assigns to the possible spin results be the same in these two universes as they have the same Observer+Detector reduced density matrix,

$$P(\downarrow|\Psi_1) = P(\downarrow|\Psi_2) \ . \quad (2.20)$$

Step 2: If we ask what probability Alice should assign to the display being $\heartsuit$, we need to consider the reduced density matrix generated by tracing over $D1$, the spin of the particle, and the environment. Ψ_1 and Ψ_2 agree on $\widehat{\rho}_{AD2}$. By *ESP-QM*, the probabilities assigned to $\heartsuit$ must be equal,

$$P(\heartsuit|\Psi_1) = P(\heartsuit|\Psi_2) \ . \quad (2.21)$$

Step 3: Next, note that the $\heartsuit$-branches *just are* the $\uparrow$-branches in Ψ_1 and the $\heartsuit$-branches *just are* the $\downarrow$-branches in Ψ_2. Thus Alice is in the $\heartsuit$-branch of $\widehat{\rho}_{AD2}(\Psi_1)$ if and only if she is in the $\uparrow$-branch of $\widehat{\rho}_{AD1}(\Psi_1)$. Similarly, she is in the $\heartsuit$-branch of $\widehat{\rho}_{AD2}(\Psi_2)$ if and only if she is in the $\downarrow$-branch of $\widehat{\rho}_{AD1}(\Psi_2)$. Therefore, Alice must

[19] Here we assume for simplicity that the environments are perfectly orthogonal. What is crucial is that we choose Ψ_2 such that $\widehat{\rho}_{AD1}(\Psi_1) = \widehat{\rho}_{AD1}(\Psi_2)$ and $\widehat{\rho}_{AD2}(\Psi_1) = \widehat{\rho}_{AD2}(\Psi_2)$, both of which can be easily satisfied even if the environment states are not perfectly orthogonal.

assign

$$P(\uparrow|\Psi_1) = P(\heartsuit|\Psi_1)$$
$$P(\downarrow|\Psi_2) = P(\heartsuit|\Psi_2) \ . \tag{2.22}$$

Step 4: Putting together the results in (2.20) - (2.22), we see that in Ψ_1 the probability of being on a $\uparrow/\heartsuit$-branch must be the same as that for being on a $\downarrow/\diamondsuit$-branch: $P(\uparrow|\Psi_1) = P(\downarrow|\Psi_1)$. So, the unique rational degrees of belief in the first scenario consider each branch to be equiprobable. Since these are the only two alternatives, the probability of each outcome is one half. In fact, using *ESP-QM*, we have shown that for *any* case in which the reduced density matrix is as in (2.19), the two spin states are equiprobable; there doesn't have to be a second display present. The result applies to a state in the general form:

$$\frac{1}{\sqrt{2}} |R\rangle_A |\uparrow\rangle_{D1} |\uparrow_z\rangle |E_\uparrow\rangle + \frac{1}{\sqrt{2}} |R\rangle_A |\downarrow\rangle_{D1} |\downarrow_z\rangle |E_\downarrow\rangle \tag{2.23}$$

Proof of the Born rule for $\frac{1}{3}/\frac{2}{3}$ case: Suppose Alice measures a particle in the state

$$\sqrt{\frac{2}{3}} |\uparrow_z\rangle + \sqrt{\frac{1}{3}} |\downarrow_z\rangle \ , \tag{2.24}$$

in which case upon measurement $\widehat{\rho}_{AD1}$ would be

$$\widehat{\rho}_{AD1}(\Psi_2) = \frac{2}{3} |R\rangle_A |\uparrow\rangle_{D1} \langle R|_A \langle\uparrow|_{D1} + \frac{1}{3} |R\rangle_A |\downarrow\rangle_{D1} \langle R|_A \langle\downarrow|_{D1} \ . \tag{2.25}$$

To determine the probabilities in this scenario, we will consider two different ways of having three displays linked to the measurement outcomes in the post-measurement

pre-observation state,

$$|\Psi_\alpha\rangle = \sqrt{\frac{1}{3}}\Big(|R\rangle_A |\uparrow\rangle_{D1} |\diamondsuit\rangle_{D2} |\clubsuit\rangle_{D3} |\uparrow_z\rangle |E_{\alpha 1}\rangle + |R\rangle_A |\uparrow\rangle_{D1} |\heartsuit\rangle_{D2} |\spadesuit\rangle_{D3} |\uparrow_z\rangle |E_{\alpha 2}\rangle$$
$$+ |R\rangle_A |\downarrow\rangle_{D1} |\heartsuit\rangle_{D2} |\clubsuit\rangle_{D3} |\downarrow_z\rangle |E_{\alpha 3}\rangle \Big)$$
$$|\Psi_\beta\rangle = \sqrt{\frac{1}{3}}\Big(|R\rangle_A |\uparrow\rangle_{D1} |\heartsuit\rangle_{D2} |\clubsuit\rangle_{D3} |\uparrow_z\rangle |E_{\beta 1}\rangle + |R\rangle_A |\uparrow\rangle_{D1} |\heartsuit\rangle_{D2} |\clubsuit\rangle_{D3} |\uparrow_z\rangle |E_{\beta 2}\rangle$$
$$+ |R\rangle_A |\downarrow\rangle_{D1} |\diamondsuit\rangle_{D2} |\spadesuit\rangle_{D3} |\downarrow_z\rangle |E_{\beta 3}\rangle \Big) \tag{2.26}$$

Step 1: Using *ESP-QM* to ignore $D2$, $D3$, and the environment, we can focus on the first display and compare the probabilities for $\downarrow$,

$$P(\downarrow|\Psi_\alpha) = P(\downarrow|\Psi_\beta) \ . \tag{2.27}$$

Step 2: Focusing on the second display gives

$$P(\diamondsuit|\Psi_\alpha) = P(\diamondsuit|\Psi_\beta) \ . \tag{2.28}$$

Since the $\diamondsuit$-branches of Ψ_β are the $\downarrow$-branches, we have

$$P(\diamondsuit|\Psi_\beta) = P(\downarrow|\Psi_\beta) \ . \tag{2.29}$$

Step 3: Similarly, focusing on $D3$ and noting that the $\downarrow$-branches of Ψ_β are the $\spadesuit$-branches yields

$$P(\spadesuit|\Psi_\alpha) = P(\downarrow|\Psi_\beta) \ . \tag{2.30}$$

Step 4: Combining (2.27) - (2.30) gives

$$P(\diamondsuit|\Psi_\alpha) = P(\spadesuit|\Psi_\alpha) = P(\downarrow|\Psi_\alpha) \ . \tag{2.31}$$

Since these are all of the possibilities, the probability of each is a third and

$$P(\downarrow|\Psi_\alpha) = \frac{1}{3}$$
$$P(\uparrow|\Psi_\alpha) = \frac{2}{3}\ . \tag{2.32}$$

This result holds whenever the reduced density matrix is as in (2.25).

The logic of this section suggests a way of thinking about the Born rule at an intuitive level. Our recipe amounts to the following prescription: write the state vector as a sum of orthogonal vectors with equal amplitudes by unitarily transforming the environment. Then (and only then), *ESP* justifies according equal credence to each such basis vector. In that sense, the Born rule is simply a matter of counting. However, we don't want to take this picture too literally. In particular, the counting occurs only after the original state is transformed and the branching structure altered. Also, the orthogonal basis vectors of the transformed state do not necessarily each correspond to a single quantum world. Thus the relevant counting is not simply a matter of counting agents in the original state.

Now that *ESP-QM* has revealed its power, one might reasonably suspect that we have given it too much. Let's take a moment to reflect on the principle's plausibility. As was discussed in §2.2.4 and 2.3.1, if *ESP-QM* is correct then Alice's credence that the result was up should not change between t_2 and t_3 in Once-Or-Twice *even though* the number of copies of Alice changes. The proof for the $\frac{1}{3}/\frac{2}{3}$ case above relies on the same trick: the reduced density matrix in (2.25) could describe a case of three detectors with three different combinations of outputs—as in (2.26)—or a single detector with two different outputs. By *ESP-QM* the probability of an up result must be the same in either case. When *ESP* was introduced it was immediately restricted to only apply in cases where the changes made to the environment do not involve the creation of additional copies of the agent elsewhere. Why? We cannot claim

that the problem is that such changes increase the number of copies in existence—the change from t_2 to t_3 in Once-Or-Twice does too. One problem with omitting the restriction is that it would be impossible to assign non-zero credences to the additional copies in the environment, since the agent would have to assign the same probabilities to all of the original copies whether or not additional copies are present. This problem does not arise when *ESP-QM* is applied to the quantum case. The change from t_2 to t_3 in Once-Or-Twice increases the number of copies of Alice in existence, but she is not forced to assign any of the copies probability zero. The fact that *ESP-QM* avoids this particular problem does not fully exonerate the principle. The change in the quantum state from t_2 to t_3 *could* conceivably turn out to be relevant. The motivation for *ESP-QM*, emphasized earlier, is that when Alice is wondering about her relation to some detector D, things happening elsewhere are *in fact not* relevant. We believe the principle to be well-motivated but not established beyond any doubt, and thus our derivation is provisional: if one can focus on the reduced density matrix in calculating probabilities as recommended by *ESP-QM*, the Born rule follows.

2.4 Varieties of Uncertainty

2.4.1 ESP and Indifference

In classical physics, you experience self-locating uncertainty when, for example, the universe is so large that you should expect there to exist a distant planet where someone is having the exact same immediate experiences that you are, or when the universe survives so long that you should expect short-lived Boltzmann brains to pop out of the vacuum in the exact subjective state you are in now, or when, as in Duplicating Dr. Evil, you have reason to think someone has purposefully created a duplicate of you. Such phenomena can also occur in quantum mechanical

contexts, leading to within-branch uncertainty. Once the principle is strengthened, *ESP* mandates that this sort of uncertainty should be treated with *Indifference*. This brings out an important virtue of the epistemic principle we've proposed: *ESP* explains why *Indifference* was a good heuristic valid in a wide variety of cases and also explains why the uncertainty arising from quantum measurements should be treated differently, using the Born rule.[20]

To apply *ESP* fruitfully to classical cases, it will help to strengthen the principle. The stronger version of *ESP* is not concise when stated precisely (below), but the basic idea is simple: Not only should it be irrelevant what's going on outside of the subsystems in $\mathcal{S}$, it also should be irrelevant where in spacetime each subsystem in $\mathcal{S}$ is located. Consider again Duplicating Dr. Evil. It seems irrelevant where the philosophers decide to build the replica of the lab. They could make it in America or Japan or on Mars. The choice shouldn't affect the probability that Dr. Evil assigns to being on the moon. Perhaps more controversially, we believe it is irrelevant *when* they build the replica. Suppose they tell Dr. Evil that they are scanning his lab now and will (unstoppably) make the replica next week. In such a case, Dr. Evil should start to wonder whether he's mistaken about the date. Further, we believe his doubts about being on the moon should not be mitigated by the temporal separation between himself and the replica.

Strong ESP Suppose that universe U contains within it a set of subsystems $\mathcal{S}$ such that every agent in an internally qualitatively identical state to agent A is located in some subsystem which is an element of $\mathcal{S} = \{X, Y, ...\}$. Let

[20] In a recent paper, (Wilson, forthcoming) has offered a clever alternative explanation as to why within-branch uncertainty should be treated differently from which-branch uncertainty: different Everettian worlds are in fact different possible worlds and thus *Indifference* does not constrain one's treatment of which-branch uncertainty. Our analysis goes further than just explaining the difference between the two kinds of uncertainty, we also explain the commonality (by giving a unified account of how both the Born rule, quantifying which-branch uncertainty, and *Indifference*, quantifying within-branch uncertainty, arise from a single core epistemic principle, *ESP*).

$\mathcal{S}' = \{M_X(X), M_Y(Y), ...\}$ where each $M(\cdot)$ is a transformation which rotates, spatially translates, and/or temporally shifts[21] the subsystem. The probability that A ought to assign to being located in a particular subsystem $X \in \mathcal{S}$ given that they are in U is identical to the probability that they ought to assign to being in $M_X(X) \in \mathcal{S}'$ given that they are in some universe U' (which contains within it a set of subsystems $\mathcal{S}'$ such that every agent in an internally qualitatively identical state to agent A is located in some subsystem which is an element of $\mathcal{S}'$).

$$P(X|U) = P(M_X(X)|U') \tag{2.33}$$

Here $M_X(X)$ is a relocated version of X, $M_Y(Y)$ is a relocated version of Y, etc. Since $M_Y(Y)$ is simply a relocated version of subsystem Y, being in $M_Y(Y)$ will feel just like being in Y. One might worry that the fact that $M_Y(Y)$ is surrounded by a different local environment than that around Y could make $M_Y(Y)$ and Y distinguishable. If the replica of the laboratory is built floating in outer space instead of sitting on Earth's surface, the instantaneous arrangement of furniture and fermions may be the same, but the copy of Dr. Evil would quickly notice the absence of a force keeping his feet on the floor. Still, if we focus on the *instant*[22] when the outer space replica really is a perfectly shifted version of the terrestrial replica, all of the particles in the each agent's brain are in the same arrangement and the outer space copy must be having exactly the same experiences as those had by the terrestrial copy (and thus

[21] *Strong ESP* loses some of its plausibility if these three transformations are not symmetries of the dynamical laws of U and U'. Here we assume that they all are.

[22] If one thinks (reasonably enough) that it takes time to have an experience, the subsystems X, Y, ... must be taken to be temporally extended if they are to contain agents having experiences. For example, let Y be the terrestrial laboratory persisting over the course of, say, one minute. Moving Y into outer space wouldn't change what's happening inside the laboratory during that minute at all. It would, however, involve a violation of the laws of nature (as the felt downward force would have no source) unless other appropriate changes were made (like the addition of an appropriately sized planet beneath the laboratory's floor).

the same experiences as those of the original Dr. Evil).

Using *Strong ESP*, it is straightforward to prove that one must follow the recommendations of *Indifference* in cases of classical duplication. In fact, the proof is so quick that it may cause you to doubt the principle. It shouldn't. *Indifference* was intuitively plausible.[23] *Strong ESP* retains those intuitive recommendations while avoiding the unacceptable recommendations in cases like Once-or-Twice (§2.2.4).

We'll first apply *Strong ESP* to Duplicating Dr. Evil and then move to the general case. Let X be the lunar laboratory and all of its contents (including Dr. Evil) and Y be the terrestrial replica and its contents. Let $M_X(X)$ move the lunar laboratory to where the terrestrial replica was and $M_Y(Y)$ move the lunar laboratory to where the terrestrial replica was ($M_X(X) = Y$ & $M_Y(Y) = X$, provided the two laboratories are in truly identical physical states). Take U to be the original world and U' to be the universe you get by switching the two laboratories (which, as it happens, doesn't change anything: $U' = U$). Then, (2.33) yields

$$P(X|U) = P(M_X(X)|U') = P(Y|U) \ . \tag{2.34}$$

Generalizing to arbitrary many copies of an agent is simple. By pairwise swaps like the one above it can be shown that any two copies are equiprobable. Thus when we restrict attention to cases where the agent's copies are *physically* identical and not merely having identical *experiences*, the recommendations of *Strong ESP* align with those of *Indifference*. The recommendations even agree in cases of quantum measurement when the agent's copies are truly identical, as in equal amplitude superpositions like (2.17).

The above justification for assigning equal credence to being in the lunar and

[23] In fact, it's not just intuitive, there are arguments for it. The argument given by Elga (2004) essentially relies on something like *ESP*, as we discussed in Section 2.3.1.

terrestrial laboratories in the Duplicating Dr. Evil case at no point relied on the laws being classical. One could equally well apply the reasoning to cases of within-branch uncertainty. For simplicity, focus on a single branch which contains a copy of Dr. Evil on the moon and another on Earth. The state might then be represented as $|\Psi\rangle = |L\rangle|T\rangle|E\rangle$, where $|L\rangle$ is the lunar lab, $|T\rangle$ is the terrestrial lab, and $|E\rangle$ is everything else. Take subsystem X to be the lunar laboratory, represented by the reduced density matrix $|L\rangle\langle L|$, and Y to be terrestrial one, $|T\rangle\langle T|$. We can define two unitary shift operators: $\widehat{S}_X$ which takes $|L\rangle$ to $|T\rangle$ and $\widehat{S}_Y$ which takes $|T\rangle$ to $|L\rangle$. With these operators we can express M_X and M_Y mathematically as $M_X(\cdot) = \widehat{S}_X(\cdot)\widehat{S}_X^\dagger$ and $M_Y(\cdot) = \widehat{S}_Y(\cdot)\widehat{S}_Y^\dagger$. The remainder of the argument proceeds as above. Note that the division into subsystems is different here than in §2.3.2. There we assumed you know what particles you are made of and are trying to determine what branch of a particular reduced density matrix you inhabit. Here we've focused on the question of what you're made of—that is, which reduced density matrix gives the state of the particles that compose your body.

2.4.2 Mixed Uncertainties

We have now seen that, depending on the scenario, probability in quantum mechanical contexts is sometimes handled by the Born rule and other times handled by *Indifference*. What about cases where the two types of uncertainty are mixed—when one knows neither where they are in spacetime nor which branch they are on? Here we must rely on the rule from which both individual prescriptions arise: *Strong ESP*. In this section, we'll discuss two illustrative 'quantum sleeping beauty' scenarios that combine both kinds of uncertainty (Lewis, 2007, 2009b; Papineau & Durà-Vilà, 2009a, 2009b; Peterson, 2011; Bradley, 2011, forthcoming; Wilson, 2014; Groisman *et al.* , 2013).

Two-Branch-Beauty The experimental subject, Beauty, will be put to sleep on Sunday night. While she is asleep, there will be a z-spin measurement of particle a which is initially x-spin up. If the result is up, she will be awoken on Monday and then her memory of Monday's events will be erased so that on Tuesday when she wakes up her last memories will be of going to sleep Sunday night. If the result is down, she will not have her memory erased. Beauty knows everything that might be relevant about the setup.

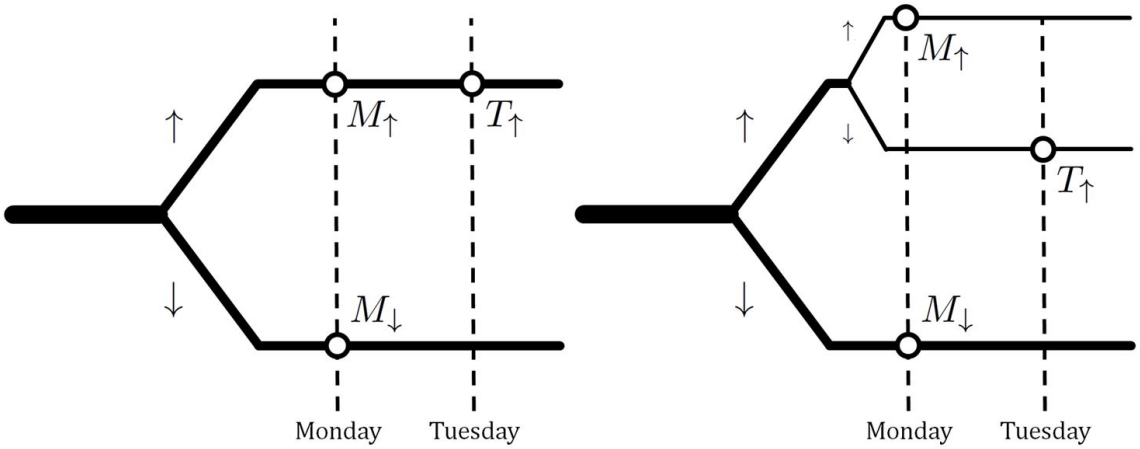

Figure 2.2: **Two Quantum Sleeping Beauty Scenarios**

Consider Beauty's situation upon waking with her most recent memories being those of going to bed Sunday evening. Her evidence doesn't discern between three possible locations in the multiverse where she might be: Monday morning on a branch where a was measured to be up, $M_\uparrow$, Tuesday morning on a branch where a was up, $T_\uparrow$, or Monday morning on a branch where a was down, $M_\downarrow$. Upon waking, what probability should Beauty assign to particle a being up? That is, what probability should she assign to $M_\uparrow \vee T_\uparrow$? By *ESP*-based arguments like those in §2.3.2, she

should assign equal probability to $M_\uparrow$ and $M_\downarrow$,

$$P\left(M_\uparrow\right) = P\left(M_\downarrow\right) \ . \tag{2.35}$$

However, since $M_\uparrow$ and $M_\downarrow$ are not the only alternatives, it does not follow that she should assign a probability of one half to each.

Given that a was spin up, we're dealing with a case of within-branch uncertainty about what day it is. Here *Indifference* and *Strong ESP* agree that Alice should assign equal probability to it being Monday or Tuesday morning (provided that the Monday and Tuesday Alices are in identical physical states, a useful albeit unrealistic assumption),

$$P\left(M_\uparrow\right) = P\left(T_\uparrow\right) \ . \tag{2.36}$$

Since $M_\uparrow$, $M_\downarrow$, and $T_\uparrow$ are the only three options, it follows from (2.35) and (2.36) that Alice should assign each alternative a probability of one third. The answer to our original question is that $P\left(\uparrow\right) = P\left(M_\uparrow \vee T_\uparrow\right)$ ought to be two thirds. We thus recover the popular 'thirder' result for this kind of sleeping beauty scenario.[24]

Two-Branch-Beauty may appear to be merely a fanciful philosopher's concoction, of little to no physical importance. In fact, the case is importantly similar to inflation scenarios where some quantum branches have no life, others have life, and some have so much life that one can reasonably expect that one's own experiences are duplicated somewhere else in spacetime. If we hope to develop a framework for testing theories like this, we need epistemic principles which can handle uncertainty about both which branch one is on and where one is within that branch.

[24] Vaidman (2011, §6) gets the same result in this case through his consideration of self-locating uncertainty.

There's something odd about the recommendations of *Strong ESP* in Two-Branch-Beauty. The probability that the principle tells Beauty to assign to up is 2/3 whereas the Born rule recommends a probability of 1/2. Is the Born rule invalid?[25] We think that the right lesson to draw is that there are two sources of uncertainty here and the Born rule is perfectly correct for quantifying the uncertainty brought about by quantum measurement. However, there is also uncertainty arising from the duplication of Beauty's experiences on the up branch. When both types of uncertainty are present, we need a rule for aggregating them. *Strong ESP* is capable of doing this, as discussed here and in §2.4.3.

To understand what's special about the Two-Branch-Beauty case discussed above, it is worth considering a variant of the case.

Three-Branch-Beauty Beauty will be put to sleep on Sunday night. While she is asleep, there will be a z-spin measurement of particle a which is initially x-spin up. If the result is up, particle b (identically prepared) will also be measured: If b is up, Beauty will awake on Monday morning; if down, Beauty will be kept asleep until Tuesday morning. If particle a is measured to be down, particle b will not be measured and Beauty will awake on Monday. Beauty knows the setup.

This is simply a case of repeated quantum measurements, essentially the same setup as Once-or-Twice. Here we can run arguments like those in §2.3.2 to show that Beauty should assign Born rule probabilities: $P(M_\uparrow) = 1/4$, $P(T_\uparrow) = 1/4$, and $P(M_\downarrow) = 1/2$.[26] Comparing Three-Branch-Beauty and Two-Branch-

[25] See the discussion of cosmological cases where it appears that the Born rule is untenable in (Page, 2009a); see also §2.4.3.

[26] A wrinkle: Before we can apply the methods of §2.3.2, one must first shift attention to a universe where the $T_\uparrow$ copy of Beauty is temporally translated back to Monday. *Strong ESP* says that the probabilities will be the same in such a universe. However, this new case is easier to handle as it is exactly analogous to Once-or-Twice.

Beauty, we see that *Strong ESP* explains how the popular thirder solution can be correct for Two-Branch-Beauty whereas the halfer solution is correct in the quantum Three-Branch-Beauty case (solving the mystery of Lewis, 2007).

2.4.3 Large Universe Cosmology and the Quantum Multiverse

The tricks of §2.4.2 can be used to give recommendations in the general case. Consider a universe U—recall that in our terminology 'universe' means 'entire quantum multiverse'—with a set of observers $\{\mathcal{O}_i\}$ who find themselves in indistinguishable circumstances.[27] They may be on the same branch of the wave function, on different branches, or even at different times. For each observer $\mathcal{O}_i$ existing at some time t_i, the overall state describing the universe can be written in the form

$$|\Psi(t_i)\rangle = \alpha_i|\phi_i\rangle + \beta_i|\phi_i^{\perp}\rangle \ , \tag{2.37}$$

where $|\phi_i\rangle$ is the branch on which the observer lives and $|\phi_i^{\perp}\rangle$ is the remainder of the quantum state, including all other branches. $|\phi_i^{\perp}\rangle$ is (at least approximately) orthogonal to $|\phi_i\rangle$.

In this case, by the considerations above, *Strong ESP* provides an unambiguous procedure for assigning credences. To each observer we assign a weight

$$w_i = |\alpha_i|^2 \ . \tag{2.38}$$

[27] Here we work under the fiction that the observers can be easily distinguished, labeled, and counted. In actuality there may be multiple ways of carving the wave function into branches (see §2.2.2). Corresponding to different ways of carving the wave function into branches, there will be different collections of observers $\mathcal{O}_i$. However, predictions derived using (2.39) will not depend significantly on the particular carving. In applying *Strong ESP* we also assume that the observers are physically identical (modulo the weights of their branches).

Then the probability for being observer $\mathcal{O}_i$ is simply[28]

$$P(\mathcal{O}_i|U) = \frac{w_i}{\sum_j w_j} \ . \tag{2.39}$$

Note that the weights $\{w_i\}$ will not in general sum to unity for a variety of reasons: the weights may be calculated at different times for different observers; there may be multiple observers on a single branch; there may be branches with no observers. This rule reduces to *Indifference* in the case of multiple observers on a single branch and to the Born rule when there is a single observer on each branch.

This recipe provides a resolution of the ambiguity in applying the Born rule in large universes that was identified by Page (2009a, 2009b, 2010); see also (Aguirre & Tegmark, 2011; Albrecht & Phillips, 2014). Consider a universe that is large enough to contain multiple observers with identical experiences on the same branch of the wave function. Imagine that each observer plans to measure the z-spin of their particle, each particle being in a potentially different pre-measurement state:

$$|\sigma_i\rangle = \gamma_i|\uparrow_z\rangle + \delta_i|\downarrow_z\rangle \ . \tag{2.40}$$

In simple cases of quantum measurement, the Born rule can be expressed as the statement that the probability of an observational outcome is given by the expectation value of a projection operator. For example, in an individual state of the form (2.40), the probability of observing spin-up is $P(\uparrow_z|\sigma_i) = \langle\sigma_i|\hat{\Pi}_\uparrow|\sigma_i\rangle = |\gamma_i|^2$, where $\hat{\Pi}_\uparrow = |\uparrow_z\rangle\langle\uparrow_z|$. Page shows that there is no projection operator that gives the probability that an observer, not knowing which observer they are, will measure $\uparrow_z$ or $\downarrow_z$; in that sense the Born rule is insufficient to fix the probabilities of measurement outcomes. *Strong ESP* resolves this ambiguity; using (2.39), the probability of z-spin up (post-

[28] Groisman *et al.* (2013) have proposed essentially the same rule.

measurement, pre-observation) is simply given by

$$P(\uparrow_z|U) = \sum_i P(\mathcal{O}_i|U) \times |\gamma_i|^2 \ , \tag{2.41}$$

and analogously for z-spin down. Here $P(\mathcal{O}_i|U)$ are the probabilities assigned to being different observers *before* the measurement is conducted.

This result has consequences for the measure problem in cosmology (Freivogel, 2011; Salem, 2012; Vilenkin, 2012), although we will not explore them in detail here. We will only note that the probability one should assign to being a particular observer in the multiverse clearly depends on the amplitude of the branch on which that observer finds themselves. For example, consider one evolving branch on which the temporal density of observers grows exponentially, $n(t) \propto e^{\omega t}$, such as might happen in inflationary cosmology. Let $\mathcal{B}$ be the subset of observers who live on that branch. According to our prescription, the probability one should assign to being on that branch is not proportional to the integral of $n(t)$ over time, which is obviously infinite. Rather, it should be weighted by the amplitude $\alpha(t)$ of the corresponding branch:

$$P(\mathcal{B}|U) \propto \int dt\, |\alpha(t)|^2 n(t). \tag{2.42}$$

If the the amplitude is decaying exponentially in time, $\alpha(t) \propto e^{-\gamma t}$, we will have a well-defined finite probability for being a member of $\mathcal{B}$ as long as $\gamma > \omega/2$. If $\gamma < \omega/2$, the numerator and denominator of (2.39) both go to infinity and the rule fails to give a well-defined probability for being in $\mathcal{B}$.

2.5 Probability in Practice

In §2.3.2 we showed that, post-measurement pre-observation, agents should assign probabilities to measurement outcomes in accordance with the Born rule. This alone is not a sufficiently strong result to show that the many-worlds interpretation recovers all of the important aspects of quantum probability. There remain two key problems, identified in Papineau (1996); Greaves (2007a). First, *the practical problem*. When faced with decisions whose repercussions depend upon the outcomes of future quantum measurements, why should agents act as they would if only one outcome were going to occur with probability determined by the Born rule? Second, *the epistemic problem*. Why can we infer facts about the wave function from observed long-run frequencies? Also, why do the data usually taken as evidence for quantum mechanics provide evidence for *Everettian* quantum mechanics? As we are primarily concerned with our reasons for believing in the many-worlds interpretation, we focus on the epistemic problem. The practical problem is less urgent. If Everettian quantum mechanics is well-confirmed by the evidence but turns out to recommend that we act differently, then we would do well to adjust the way we make decisions. In §2.5.1 we suggest that perhaps there is no adjustment, that an Everettian can continue to act as if a single outcome will occur with probability given by the Born rule.

A common objection to Everettian quantum mechanics is that there is no way for probabilities to arise in a deterministic theory when the entire physical state and the laws are known. Consider Once-or-Twice. At t_1 before the measurements have been made, Alice knows what will happen. She will have a successor who sees that particle a was measured to be down and two who see up. If she knows the universal wave function, there is nothing for her to be uncertain of. Of course, as was discussed in §2.2.2, she will experience self-locating uncertainty at t_2 and t_3.

But, at t_1 there is no uncertainty. Still, it may be that Alice should act *as if* she was uncertain (perhaps there is, as Vaidman 2011, 2014a would say, an 'illusion of probability'). Regardless of how Alice is to act at t_1, when the data is collected and the theories are tested, between t_3 and t_4, the requisite uncertainty *is* present and confirmation can proceed as usual. Here we follow Greaves (2004) in admitting that Everettian quantum mechanics lacks pre-measurement subjective uncertainty but arguing that it is not thereby refuted.[29] The core problem of this section, whether post-measurement pre-observation probabilities are sufficient for solving the practical and epistemic problems, has been discussed by Tappenden (2011) and we are largely in agreement with his conclusions.

2.5.1 Betting and Branching

Imagine that, in Once-or-Twice, at t_1 Alice is offered a bet which costs \$20 and pays \$50 if the first measurement yields down, nothing if up. The net reward or cost is assessed at t_4. Should Alice accept the bet? At t_2, all of her successors will wish that she had taken the bet. They each assign a probability of 0.5 to up and 0.5 to down, so the expected value of the bet is \$5. At t_3, the expected value of the bet is the same and again all of Alice's successors will wish she had taken it. Alice knows ahead of time that although she'll gain no new information about the

[29] Although we will not explore the option here, one could try to revive some notion of pre-measurement uncertainty which would be present even in cases where the wave function is known; strategies include conceiving of persons as four-dimensional worms, taking Everettian worlds to be diverging as opposed to overlapping, and/or trying to most charitably interpret words like 'uncertainty' as uttered by agents who unknowingly reside in an Everettian multiverse (see Saunders & Wallace, 2008; Saunders, 2010b; Wilson, 2012; Wallace, 2012, ch. 7). Our proof of the Born rule post-measurement could potentially be combined with such an account to justify setting the pre-measurement probabilities in alignment with the Born rule. However, it is not clear that true pre-measurement uncertainty is needed to solve the practical or epistemic problems. Although there may be some way of making sense of pre-measurement probabilities, we hope to show that our approach does not rely on the success of such a program (here we adopt the strategy of Greaves, 2007a, §1.3).

outcome as time progresses, once the measurement has occurred she will wish she'd taken the bet. So, it seems reasonable that at t_1 she should gamble in the way her future selves will wish she had and accept the bet. Generalizing this reasoning, Alice should always bet as if one outcome were going to occur with probability given by the Born rule since during the inevitable post-measurement pre-observation period all of her successors will approve of choices made under this supposition. (This type of argument is presented and assessed in Tappenden, 2011.)

When Alice decides, pre-measurement, to accept the bet, she is not doing so because she is trying to make a decision under uncertainty. Instead, she is trying to make a decision about how to distribute goods among her successors. If she distributes goods as recommended above, each successor will think the decision reasonable before they come to know which successor they are. Once they come to realize which branch they're on, they may well not endorse the bet. This should not be surprising. Even the most careful gambler can lose and in such cases would prefer that the bet was never made.

The strategy for arriving at effective pre-measurement uncertainty outlined above has three distinct shortcomings. First, the claim that one's pre-measurement decisions must align with the post-measurement preferences of one's successors must be justified. This alignment might be enforced by some sort of decision-theoretic reflec-

tion principle, but such a principle would need to be precisely stated and defended.[30] Second, this strategy only works if the preferences of agents are narrow: each successor only cares about what's happening in their own branch.[31] If each of Alice's successors only cares about their own wealth, then the bet described above seems lucrative. There is a 50% chance of winning \$30 and a 50% chance of losing \$20. The probability of being on a certain branch acts, for all practical purposes, like the probability that a certain thing is happening. But, if the successors care about, say, the average wealth of all successors, then at t_2 it seems like a good bet but at t_3 it does not. (Here we've assumed an idealized case, as in §2.2.3, where the number of successors is well-defined. Removing the idealization makes decision making harder, but doesn't fix the problem that one's preferences may extend beyond the goings on in one particular branch. For example, one might conceivably desire that no copy of oneself elsewhere in the multiverse be experiencing a truly miserable life (Price, 2010, §7).) Third, the strategy outlined above does not directly address cases where there is no period of uncertainty at all because observation happens immediately upon measurement. We believe such cases are atypical, but perhaps possible (§2.2.2). In such a case, one can use the strategy above to argue that *had there been* a period of uncertainty of any length at all, it would have been rational to treat the psi-squared

[30] Some work has already been done in this direction. Wallace (2002, §8.1) introduces a decision-theoretic reflection principle which is discussed in Greaves (2004, §4.2) and Greaves (2007a, §5.2.1). The requirement is roughly that one's betting behavior ought not change over a period of time in which no new evidence is gained (and this is coupled with the idea that the occurrence of an expected branching event provides no new evidence). Wallace (2010c,2012) defends a requirement of diachronic consistency which does similar work. Unfortunately, both of these principles condemn agents we take to be acting rationally in various cases where self-locating uncertainty is important (such cases often violate *epistemic* reflection principles, see Lewis, 2009a; Arntzenius, 2003). For example, the principles condemn an agent who accepts the Duplicating Dr. Evil Dutch book in appendix 2.A as the agent's betting behavior changes once the duplicate is created. We thus believe that these two principles cannot be correct as stated.

[31] Tappenden (2011) is aware of this problem (which was raised in the context of anticipated branching by Price, 2010).

branch weights as probabilities for the sake of decision making—to care about one's successors in proportion to the weights of the branches they occupy. As the limit of the amount of time after measurement before observation is taken to zero, the same decisions remain rational. It would be very odd if they were not the right actions to take when there is no period of uncertainty at all.[32]

2.5.2 Theory Confirmation

For the purposes of empirically testing competing theories—about the state of the system or the laws that govern its evolution—it is necessary that when the data come in we can judge whether the data are considered probable or improbable by the various theories under consideration. That is, immediately before the agent *observes* the result, probabilities of different outcomes must be well-defined. However, it is inessential that there be well-defined probabilities before the *measurement* is made and the outcome is recorded by the measuring device (the point at which the wave function branches). Post-measurement pre-observation probabilities are sufficient for theory confirmation. On our account, immediately before looking at the outcome of an experiment an agent in an Everettian multiverse is uncertain of what the observed outcome will be and is perfectly capable of quantifying that uncertainty. Theories can be tested in familiar ways.

Although theory testing in Everettian QM proceeds essentially as usual, we discuss two kinds of learning scenarios here for the sake of illustration. First, consider testing theories about the wave function of the system by gathering data about which eigenvalues are measured. We will treat the problem from a Bayesian perspective, but one could apply alternative methods (see Wallace, 2012, §6.2 & 6.3). For simplicity,

[32] A similar limiting argument is made by Greaves (2007a, §5.2.1), and another appealing to a hypothetical period of uncertainty by Tappenden (2011, §4)

consider a case where there are just two theories under consideration:

WHAT WAVE FUNCTION? Alice will measure the z-spin of a single particle which she knows to be prepared in either state '$P\uparrow$' ('probably up') or state '$P\downarrow$,'

$$|\Psi_{P\uparrow}\rangle = \sqrt{\frac{9}{10}}\ |\uparrow_z\rangle + \sqrt{\frac{1}{10}}\ |\downarrow_z\rangle$$
$$|\Psi_{P\downarrow}\rangle = \sqrt{\frac{1}{10}}\ |\uparrow_z\rangle + \sqrt{\frac{9}{10}}\ |\downarrow_z\rangle\ . \tag{2.43}$$

Before the experiment, she considers either wave function equally likely. Let t_1 be a time after preparation before measurement, t_2 be post-measurement pre-observation, and t_3 be immediately post-observation.

On our account it is not hard to see that observing an up result confirms $\Psi_{P\uparrow}$. At t_2, Alice assigns conditional probabilities $P_{t_2}(\uparrow_z|\Psi_{P\uparrow}) = 0.9$ and $P_{t_2}(\uparrow_z|\Psi_{P\downarrow}) = 0.1$. As she knows branching has occurred, these are the probabilities she assigns to being on an up branch conditional on particular initial wave functions. If upon observation she sees that the spin was up, she should update her credence in $\Psi_{P\uparrow}$ by conditionalizing on her new evidence,

$$\begin{aligned} P_{t_3}(\Psi_{P\uparrow}) &= \frac{P_{t_2}(\uparrow_z|\Psi_{P\uparrow})\, P_{t_2}(\Psi_{P\uparrow})}{P_{t_2}(\uparrow_z|\Psi_{P\uparrow})\, P_{t_2}(\Psi_{P\uparrow}) + P_{t_2}(\uparrow_z|\Psi_{P\downarrow})\, P_{t_2}(\Psi_{P\downarrow})} \\ &= \frac{0.9 \times 0.5}{0.9 \times 0.5 + 0.1 \times 0.5} = 0.9\ . \end{aligned} \tag{2.44}$$

The credence Alice assigns to $\Psi_{P\uparrow}$ jumps from 0.5 to 0.9, as it should. Continuing to observe more up results than down in identically prepared systems would further confirm the $\Psi_{P\uparrow}$ hypothesis over $\Psi_{P\downarrow}$. In analyzing this case we've assumed that Alice should update as usual, by (2.44), even though some of the probabilities involved

quantify her self-locating uncertainty, like $P_{t_2}(\uparrow_z|\Psi_{P\uparrow})$.[33]

Next, consider a second type of learning scenario. We can empirically confirm Everettian quantum mechanics over competing empirically inequivalent hypotheses about the physical laws by observing Born rule statistics. Suppose we run a large number of experiments without observing the outcomes (one could equally well look after each experiment, but this way things will be a bit simpler). Some sequences of results would be deemed likely by the Born rule and others would be judged very improbable. Suppose that the actual sequence of outcomes, $\mathscr{S}$, is one that the Born rule judges likely. For example, imagine that of a large number of measurements of $\widehat{A}$ on systems prepared in the same state $|\Psi\rangle$, the fraction in which eigenvalue a is observed is approximately $|\langle a|\Psi\rangle|^2$. Then, the probability of the evidence given the theory, $P(\mathscr{S}|\text{EQM}\&\Psi)$, is high for Everettian quantum mechanics, EQM, because Everettian probabilities match the Born rule probabilities (§2.3.2). If another theory considered such sequences to be less probable, then updating on $\mathscr{S}$ would support Everettian quantum mechanics over this competing theory. If the alternative theory assigned the same probability to $\mathscr{S}$, the data would not discern between the two theories.

In general, confirmation will work as usual in cases where Alice has a period of uncertainty before the outcome is revealed to her. What happens if in What Wave Function? Alice sees that the particle is $\uparrow_z$ *exactly when* the branching occurs so that there is no post-measurement pre-observation period? This is a tricky

[33]There is a way in which the case of theory testing here is different from normal cases. On either of the two hypotheses about the wave function, it was guaranteed that a version of Alice would see spin up. This leads to a worry: What Alice learns upon seeing up is that 'in my universe, one of the copies saw up'—which doesn't provide evidence for either theory over the other as Alice already knew with certainty that it would happen—and that 'I'm the one who saw up'—which is purely self-locating information about where she is in the universe and thus not informative about the way the universe is. We side with Titelbaum (2008) and Greaves & Myrvold (2010, pp. 294-295) in rejecting the final step, 'purely self-locating evidence' *can be* informative about the way the world is. For a theory about learning from evidence that disagrees on this point, see Meacham (2008).

question. Alice might try to update her credences immediately after branching by conditionalization as in (2.44) (now moving directly from the pre-measurement time t_1 to the post-observation time t_3 as there is no post-measurement pre-observation period),

$$P_{t_3}\left(\Psi_{P\uparrow}\right) = \frac{P_{t_1}\left(\uparrow_z|\Psi_{P\uparrow}\right) P_{t_1}\left(\Psi_{P\uparrow}\right)}{P_{t_1}\left(\uparrow_z|\Psi_{P\uparrow}\right) P_{t_1}\left(\Psi_{P\uparrow}\right) + P_{t_1}\left(\uparrow_z|\Psi_{P\downarrow}\right) P_{t_1}\left(\Psi_{P\downarrow}\right)} \,. \tag{2.45}$$

In this equation, the probability $P_{t_1}\left(\uparrow_z|\Psi_{P\uparrow}\right)$ is hard to interpret. It's definitely not the probability that the particle was up given $\Psi_{P\uparrow}$ *at* t_1 since the particle was in a superposition of up and down before the measurement was made. It cannot be Alice's pre-measurement probability that '*I* will see up' given that the particle is currently in state $\Psi_{P\uparrow}$ since the claim is ill-defined; some successors will see up and others will see down, all are Alice's descendants. It should not be interpreted as the probability that *some* successor of Alice sees up. In cases where stochastic theories are under consideration this proposal would make Everettian QM too easy to confirm—all outcomes that one might observe are given probability one (see Greaves, 2007a, p. 140). Standard Bayesian methods break down. The question of how to revise Bayesian confirmation theory is difficult and numerous proposals have been made—many of which rely on Elga's indifference principle and thus are incompatible with *ESP* (see footnote 2.2.2). To thoroughly address the question of how to update when the outcome is immediately observed upon branching, we must await philosophical

progress.[34] There is reason to be optimistic. Plausibly, in the case of immediate observation one should adjust their beliefs in the way they would *if there were* a short period of self-locating uncertainty as the adjustment is the same no matter how short the period of self-locating uncertainty is (Tappenden, 2011, p. 107).

2.6 Comparison to Other Approaches

The main purpose of this article is to present our derivation of the Born rule. It is consistent with the success of our derivation that other attempts to derive the Born rule are also satisfactory. However, there would have been little motivation to embark on this project if we did not have concerns about existing approaches. In this section we will highlight the differences between our derivation and two existing programs: Zurek's envariance-based derivation and the decision-theoretic approach.

2.6.1 Zurek's Envariance-based Derivation of the Born Rule

Here we will briefly present Zurek's (2005) argument in the simplest case and mention some of the limitations of his approach.

According to Zurek, the probabilities for a system S to manifest various properties upon measurement depend only on the state of the system ('Fact 2'). Further, the state of a system is not affected by unitary transformations on the environment ('Fact 1'). From these assumptions, one can prove that the probabilities for different

[34] One appealing recent suggestion is to replace $P_{t_1}(\uparrow_z|\Psi_{P\uparrow})$ with $\tau(\Psi_{P\uparrow} \to \uparrow_z)$ where $\tau(\Psi_{P\uparrow} \to \uparrow_z)$ is a 'transition probability,' capturing something like the degree to which the copies of Alice on $\uparrow_z$ branches are Alice's successors if the initial wave function is $\Psi_{P\uparrow}$; or, more operationally, specifying how much of Alice's credence in $\Psi_{P\uparrow}$ should shift to the $\uparrow_z$ branches at t_3 before she takes account of any new evidence (Schwarz, forthcoming). If Schwarz's suggestion is adopted, one could argue that the transition probabilities should be given by the amplitude-squared of the wave function by requiring that belief update when there is no period self-locating uncertainty agree with the case of an arbitrarily small period of self-locating uncertainty.

measurement outcomes are in agreement with the Born rule in ordinary scenarios. For example, suppose the state of the universe is

$$\frac{1}{\sqrt{2}}\Big(|\uparrow_z\rangle_S |E_1\rangle + |\downarrow_z\rangle_S |E_2\rangle \Big) , \tag{2.46}$$

where E_1 and E_2 are orthogonal. It is taken to follow from the perfect entanglement between system and environment that $P(\uparrow_z) = P(E_1)$. The following universal state can be reached by a unitary swap of the environment states (or a unitary swap of the system states),

$$\frac{1}{\sqrt{2}}\Big(|\downarrow_z\rangle_S |E_1\rangle + |\uparrow_z\rangle_S |E_2\rangle \Big) . \tag{2.47}$$

This swap does not change the state of the system or the environment, it only affects the entanglement between the two. The perfect entanglement of (2.47) yields $P(\downarrow_z) = P(E_1)$. Combining these two results gives $P(\uparrow_z) = P(\downarrow_z)$, in agreement with the Born rule.[35]

Zurek has provided a compelling argument that the Born Rule gives the only *sensible* probability measure in Everettian quantum mechanics. However, that has arguably already been established by Gleason's Theorem (Gleason, 1957). What Zurek fails to explain is how probabilities arise at all in this deterministic theory—self-locating uncertainty is not discussed. Thus in his derivation the nature of the probabilities involved is obscure. Zurek claims that he is calculating probabilities for *future* measurement outcomes, but he does not explain clearly what these are probabilities *of*, saying that the 'observer can be ignorant of his future state, of the outcome of the measurement he has decided to carry out' (Zurek, 2005, §VII.C).[36] This would be reasonable if only one outcome were expected to occur. However,

[35] In our exposition of (Zurek, 2005, §II.C) we take option 3 since it seems strongest and is most similar to the reasoning in our derivation.

[36] See also FAQ #5 of (Zurek, 2010).

as explained in §2.2.2, it is not obvious that there is anything for one to assign probabilities to in the many-worlds interpretation when all outcomes will occur and the future evolution of the wave function is known.[37] One could reasonably argue that Facts 1 and 2 are insufficiently motivated in Zurek's treatment, especially given that the probabilities involved are not the usual sort.[38] Although our derivation relies on an assumption as well, *ESP-QM* (which is similar to Facts 1 and 2 taken together), we have attempted to provide a more thorough justification of our assumption and a more philosophically careful treatment of the probabilities involved.

2.6.2 The Decision-theoretic Program

Starting with (Deutsch, 1999), there have been a variety of attempts to justify Born rule probabilities using decision theory (Greaves, 2004, 2007a; Greaves & Myrvold, 2010; Wallace, 2003b; 2010c, 2012; Wilson, 2013). The basic strategy is to argue that plausible constraints on rational preferences ensure that when we make bets about the outcomes of future quantum measurements we will act as if we assign Born rule probabilities to the various branches. The success of their program is compatible with the success of our purely epistemic arguments. In fact, the mathematical methods used to derive the Born rule are quite similar.[39] At this point, many remain un-

[37] Zurek (2005, §III.C) recognizes this challenge and responds by departing from the many-worlds interpretation and discussing his Existential interpretation instead. This raises doubts about the applicability of his proof to Everettian quantum mechanics.

Wallace (2012, ch. 7, esp. §7.6) argues that we should expand our concept of subjective probability so that one can not only assign probabilities to ways the world might be and locations they might have in it, but also to different futures within a branching Everettian multiverse. Doing this would help preserve the correctness of language used by agents in Everettian worlds. Wallace may be right that such an extension is possible and advisable, but to the extent that it is, it requires a significant change in the way we understand probability. In this new context, fact 2 is far from obvious. Wallace's derivation of the Born rule is not susceptible to similar concerns as his assumptions are about the *preferences* of ideal agents, not their *probability assignments*.

[38] See also (Schlosshauer & Fine, 2005, esp. §III.F2; Albert, 2010, §3.1).

[39] Compare Wallace's, 2012, §5.5, use of erasure to ignore irrelevant details and our use of *ESP-QM* in §2.3.2 to do the same.

convinced by the decision-theoretic program (Baker, 2007; Albert, 2010; Price, 2010; Dizadji-Bahmani, forthcoming; Maudlin, 2014). We focus on one particular type of concern here to highlight an advantage of our approach.

The constraints proposed in the decision-theoretic approaches may reasonably be doubted as they manifestly conflict with a *prima facie* reasonable epistemic principle, *Indifference*. Deutsch's original formulation of the decision-theoretic argument for the Born rule implicitly appeals to the following assumption (Wallace, 2003b):

> **Measurement Neutrality.** A rational agent is indifferent between any two quantum bets that agree on the state $|\Psi\rangle$ on which the measurement is to be performed, the observable $\widehat{X}$ to be measured, and the 'payoff function' P from the spectrum of $\widehat{X}$ to the set of consequences. (Greaves, 2007b, p. 119)

Wallace (2007) has proved that *Measurement Neutrality* is equivalent to:

> **Equivalence.** A rational agent is indifferent between any two quantum bets that agree, for each possible reward, on the mod-squared measure of branches on which that reward is given. (Greaves, 2007b, p. 119)

The arguments in (Deutsch, 1999; Greaves, 2004, 2007b; Wallace, 2003b) appeal to this assumption in one of its manifestations. *Measurement Neutrality* may have some intuitive plausibility but, looking at *Equivalence*, it is obvious that this assumption will be inconsistent with a branch-counting rule for assigning probabilities. The principle seems to beg the question against the defender of *Indifference* who takes branch-counting to be the proper way to assign probabilities in Everettian quantum mechanics. Defenders of the decision-theoretic program typically point out that the conflict with branch-counting should not cause doubt about *Equivalence* since there are independent reasons to think branch-counting is wrong. Two main reasons are given. First, branches cannot generally be counted so branch-counting is not a rule one could actually apply. As discussed in §2.2.3, there are cases, albeit somewhat

contrived, where branch number is well-defined and in these cases branch-counting gives definite recommendations which are in conflict with the Born rule. The fact that it is unclear how to apply a rule in some cases is not necessarily a reason to think it incorrect when it yields a clear judgment. Second, branch-counting is diachronically inconsistent. We do not take this to be a strong reason to reject branch-counting (see appendix 2.A).

In more recent decision-theoretic derivations of the Born rule, Wallace (2010c, 2012) introduces a somewhat different collection of rationality axioms, multiple of which might be contested by the defender of branch-counting. Wallace argues in favor of these axioms and does not take their conflict with branch-counting (Wallace, 2012, §5.8.1) as a reason to think them incorrect, since branch-counting can be shown to be irrational. For one who believes his axioms, this gives reason to doubt branch-counting. But for the proponent of *Indifference*, it gives reason to doubt Wallace's axioms. We also reject *Indifference* because it conflicts with an assumption about rationality (*ESP*). However, we take our argument to be potentially more persuasive to the proponent of *Indifference* for two main reasons: *ESP* is a single, simple, general epistemic principle which we believe has similar (if not more) initial plausibility than *Indifference*; *ESP* is consistent with *Indifference* (and *Strong ESP* entails it) in cases of classical self-locating uncertainty and thus it is capable of explaining why (and how) quantum cases are to be handled differently (though see footnote 20). A strength of our approach, over both the decision-theoretic program and Zurek's derivation, is that it can be applied in both cases of classical duplication and in cases of quantum branching.

2.7 Conclusion

In this paper we have presented a justification for the Born rule in Everettian quantum mechanics in which self-locating uncertainty played a fundamental role. The policy of reacting to self-locating uncertainty by treating each observer as equiprobable, *Indifference*, is reasonable in classical scenarios but yields strange recommendations in quantum contexts. Instead we proposed the *Strong Epistemic Separability Principle* (*Strong ESP*), which extends smoothly from *Indifference* in the classical regime to the Born rule in cases of quantum measurement.

Following the consequences of *Strong ESP* leads us to a simple and physically transparent derivation of the Born rule in the many-worlds interpretation. The appearance of probabilities in a deterministic theory is explained by evolution from perfect knowledge to unavoidable self-locating uncertainty. Our approach provides a unified perspective on uncertainties in both the classical and quantum contexts, with implications for large-universe cosmology as well as for the foundations of quantum mechanics.

There are a few ways in which future work could strengthen the approach to probability in Everettian quantum mechanics developed in this article. First, although *Strong ESP* has some intuitive appeal, one could seek further philosophical justification for the principle and a more precise formulation of it. Second, more work could be done in justifying the appeal to reduced density matrices as the correct way of representing subsystems in Everettian quantum mechanics. Third, one might hope to reach more definitive conclusions about the connection between post-measurement probabilities and pre-measurement decision making.

Funding: National Science Foundation (DGE 0718128 to C.S.); Department of

Energy (DE-FG02-92ER40701 to S.C.); Gordon and Betty Moore Foundation through Caltech Moore Center for Theoretical Cosmology and Physics (776 to S.C. and C.S.).

Acknowledgments: Charles Sebens would like to thank John-Mark Allen, David Baker, Gordon Belot, Adrian Kent (see Kent, 2014), David Manley, Daniel Peterson, Barry Loewer, and Laura Ruetsche for very useful comments on drafts of this article. He would also like to thank Simon Saunders and David Wallace for a wonderful introduction to the Everett interpretation. Sean Carroll would like to thank Scott Aaronson, David Albert, Andreas Albrecht, Hirosi Ooguri, Simon Saunders, and David Wallace for helpful conversations.

2.A What's Not Wrong with Branch-counting

Wallace (2012, §4.3) gives the following diagnosis of why the switch from $P_{t_2}(\text{up}) = \frac{1}{2}$ to $P_{t_3}(\text{up}) = \frac{2}{3}$ in Once-or-Twice is irrational: it is diachronically inconsistent.[40] According to 'the standard rules of the probability calculus,' the probability of particle a being up at t_3 ought to be

$$P_{t_3}(\text{up at } t_3) = P_{t_2}(\text{up at } t_3|\text{up at } t_2)P_{t_2}(\text{up at } t_2) + P_{t_2}(\text{up at } t_3|\text{down at } t_2)P_{t_2}(\text{down at } t_2) \ . \tag{2.48}$$

But, it clearly is not if

$$\begin{aligned} P_{t_3}(\text{up at } t_3) &= \frac{2}{3} \\ P_{t_2}(\text{up at } t_3|\text{up at } t_2) &= 1 \\ P_{t_2}(\text{up at } t_3|\text{down at } t_2) &= 0 \\ P_{t_2}(\text{up at } t_2) = P_{t_2}(\text{down at } t_2) &= \frac{1}{2} \ , \end{aligned} \tag{2.49}$$

as recommended by *Indifference*. Without the subscripts, the point is certainly valid. At t_2 or t_3 (or any other time), it must be the case that

$$P(\text{up at } t_3) = P(\text{up at } t_3|\text{up at } t_2)P(\text{up at } t_2) + P(\text{up at } t_3|\text{down at } t_2)P(\text{down at } t_2) \ , \tag{2.50}$$

[40] He credits Deutsch with giving a version of the argument in conversation (Wallace, 2013, footnote 15).

provided 'up at t_2' and 'down at t_2' are mutually exclusive and exhaustive alternatives. This forbids certain unsophisticated ways of filling in the branch-counting story, but is of no concern to the defender of *Indifference*. At t_2, $P(\text{up at } t_3)$ is only $1/2$. At t_3, $P(\text{up at } t_2)$ has risen to $2/3$.

(2.48) is not a requirement of *consistency* for one's credences at a time, it is a constraint on the way one *ought to* adjust their credences over time and not one that a proponent of *Indifference* should accept. At t_2 Alice knows that there's a 50/50 chance that she's on an up branch and that if she is on one she'll definitely still be on one later. However, once the second measurement is made she realizes there are three branches she might be on and that none of her evidence discerns between the three possibilities. So, by the logic of *Indifference*, she should think it more likely that she's on an up branch since there are twice as many of them.

Wallace (2012, §5.4) rightly notes that because of the prevalence of branching in Everettian quantum mechanics, such credence shifts will be so common that deliberative action will be near impossible. Such are the dangers of living in a wildly branching multiverse.

The violation of (2.48) can be made more worrisome as it leads to an Everettian Dutch book.[41] If Alice reasons by *Indifference*, there are a series of bets which can be given to her and her successors such that each bet will be judged fair but the combination of bets guarantees a loss on every branch. Here is a way of constructing the Dutch book (from Peterson, 2011, §2): At t_2 a bet is offered which pays \$15 if particle a is down and $-$\$15 if up. At t_3 a bet is offered which pays \$10 if particle a is up and $-$\$20 if down. All costs/rewards will be collected/paid at t_4. If Alice assigns probabilities as in (2.49), these bets will seem fair. However, if she accepts

[41] Wallace would likely see this as a friendly strengthening of his quick argument (see Wallace, 2012, §4.8 together with Wallace, 2010b, p. 247). For an alternate strategy, see (Wallace, 2010b, §II).

both bets then she will lose \$5 on each branch. There is an apparent inconsistency here. Looking forward from t_1 Alice can see that this series of bets will guarantee her a loss and thus would presumably choose not to take them if offered as a package before the branching starts. However, as she goes through the experiment and her credences shift (in a way which she could have anticipated), she finds each bet fair when offered.

The defender of *Indifference* should not be perturbed.[42] Consider this potential Dutch book for Dr. Evil in the Duplicating Dr. Evil case: Before t (that is, before the duplication), Dr. Evil is offered a bet that some time long after t will pay out \$100 to Dr. Evil and −\$300 to Dup. Not long after t, he is offered a second bet which will pay out \$200 to Dup and −\$200 to Dr. Evil. If both bets are accepted, Dr. Evil and Dup will each lose \$100. The first bet seems lucrative. Since Dup has yet to come into existence, Dr. Evil can be sure he's not the duplicate. The second bet is fair by *Indifference*. Thus it turns out that accepting *Indifference* in the original case where the principle was most plausible already leads to a diachronic Dutch book. The defender of *Indifference* finds the recommendations of the principle reasonable in cases like Duplicating Dr. Evil and thus must reject the idea that this kind of diachronic Dutch book demonstrates irrationality. The Everettian Dutch book discussed above raises no new concerns for *Indifference*.

The same point can be expressed in another way. If the fission in Once-or-Twice, depicted in figure 2.1, was classical fission instead of quantum fission then assigning credences in accord with *Indifference* and betting accordingly would be reasonable (imagine cases of amoeba-like division as discussed in, e.g., Parfit, 1971). If these beliefs and betting behaviors are acceptable in the classical case, why not in the quantum case as well? The charge of diachronic inconsistency does not raise

[42]Thanks to David Manley for essential discussion here. See (Lewis, 2009a) for a related point.

distinctively quantum problems for *Indifference*.

2.B Circularity

There are reasons to be concerned that an appeal to reduced density matrices in deriving the Born rule involves some sort of illegitimate circularity (Zeh, 1997; Baker, 2007; Schlosshauer, 2007, §8.2.2; Zurek, 2003b, 2005, 2010; Kent, 2010, §5.4). This objection comes in a variety of forms. We'll consider three.

First, one might be concerned that the very use of inner products, partial trace operations, and reduced density matrices is forbidden until one has derived a connection between squared inner product and probability. This cannot be right. How could one derive such a thing without ever writing an inner product? Partial trace operations and reduced density matrices are perfectly well-defined mathematically within the framework of quantum theory. Their definition does not *require* understanding any number as a probability. The concern about circularity must then be a concern about the *physical interpretation* of these mathematical objects (see also Zurek, 2010, FAQ #6).

Second, the idea that the reduced density matrix describes the state of a subsystem—which we appeal to in our derivation of the Born rule—could be doubted.

> It is not obvious that the reduced density operator for system A is in any sense a description for the state of system A. The physical justification for making this identification is that the reduced density operator provides the correct measurement statistics for measurements made on system A. (Nielsen & Chuang, 2010, §2.4.3)

It is true that the use of reduced density matrices to describe subsystems is often mo-

tivated by the fact that the reduced density matrix retains enough information about the state to deduce Born rule probabilities for outcomes of measurements performed solely on that system (Schlosshauer, 2007, §2.4.6; Nielsen & Chuang, 2010, box 2.6). However, we believe an alternative justification is possible which has nothing to do with probabilities or measurements (although we only gesture at such a justification here). In §2.3.1 we proposed that the mathematical representation of the state of a subsystem should (1) together with the states of all other subsystems and facts about the connections between the subsystems yield the total state, and (2) be sufficient to determine its own evolution when the subsystem is isolated. Condition (1), interpreted appropriately, seems to be satisfied for reduced density matrices. Specifying the reduced density matrices for A and B, along with facts about the entanglement between the two systems, should be sufficient to specify the density matrix for the composite system AB. (We are currently thinking about the precise sense in which this is so and hope to include a more careful discussion in the published version of this chapter.) To see that (2) is met, suppose that at least for a time subsystems A and B are isolated. Let $\widehat{U}_t$ be the unitary operator that gives the time evolution of the total state. Since A and B are non-interacting, $\widehat{U}_t = \widehat{U}_A \otimes \widehat{U}_B$. The time evolution of $\widehat{\rho}_A$ is then given by $\widehat{U}_A \widehat{\rho}_A \widehat{U}_A^\dagger$ ($\widehat{U}_B$ is irrelevant).

Third, there is a worry that we cannot assume a structure of branching quantum worlds without assuming squared-amplitudes give probabilities. Suppose that the reduced density matrix for a macroscopic system is approximately diagonal. Why are we justified in treating the diagonal terms as branches and the off-diagonal terms as somehow unimportant and certainly not branches themselves? (See Baker, 2007; brief reply in Wallace, 2012, pp. 243-254.) Our derivation treats the probabilities involved as purely *epistemic*, quantifying an agent's uncertainty about which world they are in. The question of how patterns in the wave function give rise to distinct worlds—and people who can wonder about which world they are in—is primarily *metaphysical*. In

this article we've offered no additional insights on how this important project is to be completed, choosing to work under the assumption that it can be.

It should be noted that the use of density matrices, while convenient, is not strictly necessary for our derivation of the Born rule. The proof of the Born rule could equally well be derived using purely the language of state vectors. Instead of starting from the idea that the state of the environment is irrelevant, we could begin with the closely related thought that changes in the environment—represented by unitary transformations—ought not change the probabilities one assigns. In Carroll & Sebens (2014) we gave a derivation along these lines.

2.C Generalization of the Born Rule Derivation

In the main text we showed how one can derive the Born rule in two simple cases. Here we extend the arguments of §2.3.2 to the more general case of N orthogonal branches with arbitrary rational squared-amplitudes and arbitrary phases. As the proof here is restricted to *rational* squared-amplitudes, one might be concerned about whether the account gives any advice at all when the squared-amplitudes are irrational. This is especially worrisome as there are more irrational numbers than rational numbers. However, if we assume (plausibly) that the probabilities vary continuously with small changes in the amplitudes, the restricted proof is sufficient. For any wave function with irrational squared-amplitudes there exist arbitrarily similar wave functions with rational squared-amplitudes (as the rationals are a dense subset of the reals).

A general state of the above form can be written as

$$\begin{aligned} |\Psi_0\rangle &= \frac{c_1}{T}e^{i\theta_1}\,|R\rangle_A\,|d_1\rangle_D\,|E_1\rangle + \frac{c_2}{T}e^{i\theta_2}\,|R\rangle_A\,|d_2\rangle_D\,|E_2\rangle + ... + \frac{c_N}{T}e^{i\theta_N}\,|R\rangle_A\,|d_N\rangle_D\,|E_N\rangle \\ &= \sum_{k=1}^{N}\frac{c_k}{T}e^{i\theta_k}\,|R\rangle_A\,|d_k\rangle_D\,|e_k\rangle_E\ , \end{aligned} \tag{2.51}$$

where each $c_k > 0$, $c_k^2 \in \mathbb{Z}^+$, $\theta_k \in \mathbb{R}$, and $T = \sqrt{\sum_k c_k^2}$. $|R\rangle_A$ is Alice's state which is taken to be the same on every branch. $|d_k\rangle_D$ are possible states of the detector or system of interest, D, the entity whose states Alice is assigning probabilities to. $|E_k\rangle$ are the possible states of the environment. Here Alice would like to know what probabilities to assign to each possible state d_k. We will show that the appropriate probability of d_k is $\frac{c_k^2}{T^2}$ (unless $d_k = d_j$ for some j, a situation where the particle is in

the same state for two or more different environments, in which case the probability is given by the sum of the amplitude-squared for each j such that $d_k = d_j$ including $j = k$, $\sum_j \frac{c_j^2}{T^2}$). The probability that Alice ought to assign to each p_k will be unchanged if we make the following transformation on each environment state (this follows from *ESP-QM*, since the transformation leaves the Alice+Detector reduced density matrix unaffected),

$$|E_k\rangle \longrightarrow \frac{1}{c_k}\left\{|E'_{k,1}\rangle + ... + |E'_{k,c_k^2}\rangle\right\} \ . \tag{2.52}$$

The environment state E_k is taken to a superposition of c_k^2 different environment states. These transformations take Ψ_0 to

$$|\Psi_{eqamp}\rangle = \sum_{k=1}^{N}\sum_{j=1}^{c_k^2}\frac{1}{T}e^{i\theta_k}\,|R\rangle_A\,|d_k\rangle_D\,\left|E'_{k,j}\right\rangle \ . \tag{2.53}$$

In this state, all of the components have equal amplitude. Further, there are c_k^2 terms where D is in state d_k (unless $d_k = d_j$ for some j, in which case there are $\sum_j c_j^2$ terms). Presently we will prove that each component of Ψ_{eqamp} has equal probability. Looking at the number of terms with each d_k, it is clear that this is all that is needed to show that the probabilities in Ψ_0 for each d_k are given by the Born rule. The above method for moving from a state of unequal amplitudes to one of equal amplitudes is also used in (Zurek, 2005, §II.D).

We've reduced the problem to showing that terms in equal amplitude superpositions are equiprobable. A general state with equal amplitude terms, such as Ψ_{eqamp}, can be written as

$$|\Psi\rangle = \sum_{k=1}^{N}\frac{1}{\sqrt{N}}e^{i\theta_k}\,|R\rangle_A\,|d_k\rangle_D\,|E_k\rangle \ . \tag{2.54}$$

Here some states d_k may be identical. Now consider two possible transformations which leave the Alice+Detector reduced density matrix unchanged. First, consider a

transformation which entangles N display screens with each environment state E_k. The symbol on the first screen, symbol 1, may be either S_1 or S_1', maybe either $\heartsuit$ or $\diamondsuit$, symbol 2 might be S_2 or S_2', etc. The first transformation entangles the kth state with N displays, most of which show the unprimed symbols but the kth display shows the primed symbol:

$$|E_k\rangle \longrightarrow |S_1\rangle\,|S_2\rangle\,...\,|S_k'\rangle\,...\,|S_N\rangle\,|E_k^*\rangle \ . \tag{2.55}$$

This takes the state Ψ to

$$|\Psi_\alpha\rangle = \sum_{k=1}^{N} \frac{1}{\sqrt{N}} e^{i\theta_k}\,|R\rangle_A\,|d_k\rangle_D\,|S_1\rangle\,...\,|S_k'\rangle\,...\,|S_N\rangle\,|E_k^*\rangle \ . \tag{2.56}$$

The second transformation gives most environment states the same set of symbols, except the Nth state which gets all of the primed symbols,

$$|E_k\rangle \longrightarrow \begin{cases} |S_1\rangle\,|S_2\rangle\,...\,|S_N\rangle\,|E_k^{**}\rangle & \text{if } k \neq N \\ \\ |S_1'\rangle\,|S_2'\rangle\,...\,|S_N'\rangle\,|E_k^{**}\rangle & \text{if } k = N \end{cases} \ . \tag{2.57}$$

This takes Ψ to

$$\begin{aligned} |\Psi_\beta\rangle = &\sum_{k=1}^{N-1} \frac{1}{\sqrt{N}} e^{i\theta_k}\,|R\rangle_A\,|d_k\rangle_D\,|S_1\rangle\,|S_2\rangle\,...\,|S_N\rangle\,|E_k^{**}\rangle \\ &+ \frac{1}{\sqrt{N}} e^{i\theta_N}\,|R\rangle_A\,|d_N\rangle_D\,|S_1'\rangle\,|S_2'\rangle\,...\,|S_N'\rangle\,|E_N^{**}\rangle \ . \end{aligned} \tag{2.58}$$

Now we use similar techniques as in §2.3.2 to show that the probabilities of each component of Ψ_α are the same as the probability of the N-th component of Ψ_β.

Focusing on the reduced density matrix of Alice+Display k, we see that

$$P\left(d_k|\Psi_\alpha\right) = P\left(d_N|\Psi_\beta\right) \ . \tag{2.59}$$

Combining these results for all k, we see that the probability of each term in Ψ_α must be equal and thus that the probability of each d_k in state Ψ is given by the Born rule (because Ψ_α, (2.56), and Ψ, (2.54), agree on the Alice+Detector reduced density matrix). In combination with the discussion in the previous paragraph, this completes the proof that the probabilities for different states d_k in the general state Ψ_0 are given by the Born rule.

References

Aguirre, A., & Tegmark, M. 2011. Born in an Infinite Universe: A Cosmological Interpretation of Quantum Mechanics. *Physical Review D*, **84**, 105002.

Albert, D. Z. 2010. Probability in the Everett Picture. pp. 355–68. in Saunders *et al.* (2010).

Albrecht, A., & Phillips, D. 2014. Origin of Probabilities and their Application to the Multiverse. *Physical Review D*, **90**, 123514.

Arntzenius, F. 2003. Some Problems for Conditionalization and Reflection. *The Journal of Philosophy*, **100**(7), pp. 356–70.

Baker, D. J. 2007. Measurement Outcomes and Probability in Everettian Quantum Mechanics. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **38**(1), pp. 153–69.

Bostrom, N. 2002. *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge.

Bradley, D. J. 2011. Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty. *The British Journal for the Philosophy of Science*, **62**(2), pp. 323–42.

Bradley, D. J. forthcoming. Everettian Confirmation and Sleeping Beauty: Reply to Wilson. *The British Journal for the Philosophy of Science*.

Carroll, S. M., & Sebens, C. T. 2014. Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics. *Pages 157–69 of:* Struppa, D., & Tollaksen, J. (eds), *Quantum Theory: A Two-Time Success Story, Yakir Aharonov Festschrift*. Springer.

Deutsch, D. 1999. Quantum Theory of Probability and Decisions. *Proceedings of the Royal Society of London*, **A458**, pp. 3129–37.

Dizadji-Bahmani, F. forthcoming. The Probability Problem in Everettian Quantum Mechanics Persists. *The British Journal for the Philosophy of Science*.

Elga, A. 2004. Defeating Dr. Evil with Self-Locating Belief. *Philosophy and Phenomenological Research*, **69**(2), pp. 383–96.

Freivogel, B. 2011. Making Predictions in the Multiverse. *Classical and Quantum Gravity*, **28**, 204007.

Gleason, A. M. 1957. Measures on the Closed Subspaces of a Hilbert Space. *Journal of Mathematics and Mechanics*, **6**, pp. 885–94.

Greaves, H. 2004. Understanding Deutsch's Probability in a Deterministic Multiverse. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **35**(3), pp. 423–56.

Greaves, H. 2007a. On the Everettian Epistemic Problem. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **38**(1), pp. 120–52.

Greaves, H. 2007b. Probability in the Everett Interpretation. *Philosophy Compass*, **2**(1), pp. 109–28.

Greaves, H., & Myrvold, W. 2010. Everett and Evidence. pp. 264–304. in Saunders *et al.* (2010).

Groisman, B., Hallakoun, N., & Vaidman, L. 2013. The Measure of Existence of a Quantum World and the Sleeping Beauty Problem. *Analysis*, **73**(4), pp. 695–706.

Hartle, J. B., & Srednicki, M. 2007. Are We Typical? *Physical Review D*, **75**(12), 123523.

Kent, A. 2010. One World Versus Many: The Inadequacy of Everettian Accounts of Evolution, Probability, and Scientific Confirmation. pp. 307–54. in Saunders *et al.* (2010).

Kent, A. 2014. Does it Make Sense to Speak of Self-Locating Uncertainty in the Universal Wave Function? Remarks on Sebens and Carroll. *Foundations of Physics*, 1–7.

Lewis, D. K. 1979. Attitudes *De Dicto* and *De Se*. *The Philosophical Review*, **88**(4), pp. 513–43.

Lewis, P. J. 2007. Quantum Sleeping Beauty. *Analysis*, **67**(293), pp. 59–65.

Lewis, P. J. 2009a. Probability, Self-Location, and Quantum Branching. *Philosophy of Science*, **76**(5), pp. 1009–19.

Lewis, P. J. 2009b. Reply to Papineau and Durà-Vilà. *Analysis*, **69**(1), pp. 86–9.

Manley, D. unpublished. On Being a Random Sample.

Maudlin, T. 2014. *Critical Study* David Wallace, *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. *Noûs*, **48**(4), pp. 794–808.

Meacham, C. 2008. Sleeping Beauty and the Dynamics of De Se Beliefs. *Philosophical Studies*, **138**(2), pp. 245–69.

Nielsen, M. A., & Chuang, I. L. 2010. *Quantum Computation and Quantum Information*. Cambridge University Press.

Page, D. N. 2007. Typicality Defended. ⟨arxiv.org/abs/0707.4169⟩.

Page, D. N. 2009a. The Born Rule Fails in Cosmology. *Journal of Cosmology and Astroparticle Physics*, **0907**, 008.

Page, D. N. 2009b. Born Again. ⟨arxiv.org/abs/0907.4152⟩.

Page, D. N. 2010. Born's Rule Is Insufficient in a Large Universe. ⟨arxiv.org/abs/1003.2419⟩.

Papineau, D. 1996. Many Minds are No Worse than One. *The British Journal for the Philosophy of Science*, **47**(2), pp. 233–41.

Papineau, D., & Durà-Vilà, V. 2009a. A Thirder and an Everettian: A Reply to Lewis's "Quantum Sleeping Beauty". *Analysis*, **69**(1), pp. 78–86.

Papineau, D., & Durà-Vilà, V. 2009b. Reply to Lewis: Metaphysics Versus Epistemology. *Analysis*, **69**(1), pp. 89–91.

Parfit, D. 1971. Personal Identity. *The Philosophical Review*, **80**(1), pp. 3–27.

Peterson, D. 2011. Qeauty and the Books: A Response to Lewis's Quantum Sleeping Beauty Problem. *Synthese*, **181**(3), pp. 367–74.

Price, H. 2010. Decisions, Decisions, Decisions: Can Savage Salvage Everettian Probability? pp. 369–90. in Saunders *et al.* (2010).

Salem, M. P. 2012. Bubble Collisions and Measures of the Multiverse. *Journal of Cosmology and Astroparticle Physics*, **1201**, 021.

Saunders, S. 2010a. Many Worlds? An Introduction. pp. 1–49. in Saunders *et al.* (2010).

Saunders, S. 2010b. Chance in the Everett Interpretation. pp. 181–205. in Saunders *et al.* (2010).

Saunders, S., & Wallace, D. 2008. Branching and Uncertainty. *The British Journal for the Philosophy of Science*, **59**(3), pp. 293–305.

Saunders, S., Barrett, J., Kent, A., & Wallace, D. (eds). 2010. *Many Worlds?: Everett, Quantum Theory, & Reality*. Oxford University Press.

Schlosshauer, M. A. 2005. Decoherence, the Measurement Problem, and Interpretations of Quantum Mechanics. *Reviews of Modern Physics*, **76**(4), 1267.

Schlosshauer, M. A. 2007. *Decoherence and the Quantum-to-Classical Transition*. Springer Verlag.

Schlosshauer, M. A., & Fine, A. 2005. On Zurek's Derivation of the Born Rule. *Foundations of Physics*, **35**(2), pp. 197–213.

Schwarz, W. forthcoming. Belief Update across Fission. *The British Journal for the Philosophy of Science*.

Srednicki, M., & Hartle, J. 2010. Science in a Very Large Universe. *Physical Review D*, **81**(12), 123524.

Tappenden, P. 2011. Evidence and Uncertainty in Everett's Multiverse. *The British Journal for the Philosophy of Science*, **62**(1), pp. 99–123.

Titelbaum, M. G. 2008. The Relevance of Self-Locating Beliefs. *Philosophical Review*, **117**(4), pp. 555–606.

Vaidman, L. 1998. On Schizophrenic Experiences of the Neutron or Why We Should Believe in the Many-Worlds Interpretation of Quantum Theory. *International Studies in the Philosophy of Science*, **12**(3), pp. 245–61.

Vaidman, L. 2011. Probability in the Many-Worlds Interpretation of Quantum Mechanics. *In:* Ben-Menahem, Y., & Hemmo, M. (eds), *The Probable and the Improbable: Understanding Probability in Physics, Essays in Memory of Itamar Pitowsky*. Springer.

Vaidman, L. 2014a. Quantum Theory and Determinism. *Quantum Studies: Mathematics and Foundations*, **1**, 5–38.

Vaidman, L. 2014b. Many-Worlds Interpretation of Quantum Mechanics. *In:* Zalta, E. N. (ed), *The Stanford Encyclopedia of Philosophy*. ⟨plato.stanford.edu/archives/win2014/entries/qm-manyworlds/⟩.

Vaidman, L. forthcoming. Bell Inequality and Many-Worlds Interpretation. *In:* Bell, M., & Gao, S. (eds), *Quantum Non-locality and Reality: 50 Years of Bell's Theorem*. Cambridge University Press.

Vilenkin, A. 2012. Global Structure of the Multiverse and the Measure Problem. *AIP Conference Proceedings*, **1514**, pp. 7–13.

Wallace, D. 2002. Quantum Probability and Decision Theory, Revisited. ⟨arxiv.org/abs/quant-ph/0211104⟩.

Wallace, D. 2003a. Everett and Structure. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **34**(1), pp. 87–105.

Wallace, D. 2003b. Everettian Rationality: Defending Deutsch's Approach to Probability in the Everett Interpretation. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **34**(3), pp. 415–39.

Wallace, D. 2006. Epistemology Quantized: circumstances in which we should come to believe in the Everett interpretation. *The British Journal for the Philosophy of Science*, **57**(4), pp. 655–89.

Wallace, D. 2007. Quantum Probability from Subjective Likelihood: Improving on Deutsch's Proof of the Probability Rule. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **38**(2), pp. 311–32.

Wallace, D. 2010a. Decoherence and Ontology. pp. 53–72. in Saunders *et al.* (2010).

Wallace, D. 2010b. Diachronic Rationality and Prediction-Based Games. **CX**(3), pp. 243–66.

Wallace, D. 2010c. How to Prove the Born Rule. pp. 227–63. in Saunders *et al.* (2010).

Wallace, D. 2012. *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. Oxford University Press.

Wallace, D. 2013. The Everett Interpretation. *Pages 460–88 of:* Batterman, R. (ed), *The Oxford Handbook of Philosophy of Physics*.

Wilson, A. 2012. Everettian Quantum Mechanics without Branching Time. *Synthese*, **188**(1), pp. 67–84.

Wilson, A. 2013. Objective Probability in Everettian Quantum Mechanics. *The British Journal for the Philosophy of Science*, **64**(4), pp. 709–37.

Wilson, A. 2014. Everettian Confirmation and Sleeping Beauty. *The British Journal for the Philosophy of Science*, **65**(3), pp. 573–98.

Wilson, A. forthcoming. The Quantum Doomsday Argument. *The British Journal for the Philosophy of Science*.

Zeh, H. 1997. What is Achieved by Decoherence? *Pages 441–51 of:* Ferrero, M., & van der Merwe, A. (eds), *New Developments on Fundamental Problems in Quantum Physics*. Springer.

Zurek, W. H. 1981. Pointer Basis of Quantum Apparatus: Into What Mixture Does the Wave Packet Collapse? *Physical Review D*, **24**, pp. 1516–25.

Zurek, W. H. 1982. Environment Induced Superselection Rules. *Physical Review D*, **26**, pp. 1862–80.

Zurek, W. H. 1993. Preferred States, Predictability, Classicality and the Environment-Induced Decoherence. *Progress of Theoretical Physics*, **89**(2), pp. 281–312.

Zurek, W. H. 2003a. Decoherence, Einselection, and the Quantum Origins of the Classical. *Reviews of Modern Physics*, **75**(3), pp. 715–75.

Zurek, W. H. 2003b. Environment-Assisted Invariance, Entanglement, and Probabilities in Quantum Physics. *Physical Review Letters*, **90**(12), 120404.

Zurek, W. H. 2005. Probabilities from Entanglement, Born's Rule $p_k = |\psi_k|^2$ from Envariance. *Physical Review A*, **71**(5), 052105.

Zurek, W. H. 2010. Quantum Jumps, Born's Rule, and Objective Reality. pp. 409–432. in Saunders *et al.* (2010).

Chapter 3

Killer Collapse: Empirically Probing the Philosophically Unsatisfactory Region of GRW

Abstract: GRW theory offers precise laws for the collapse of the wave function. These collapses are characterized by two new constants, λ and σ. Recent work has put experimental upper bounds on the collapse rate, λ. Lower bounds on λ have been more controversial since GRW begins to take on a many-worlds character for small values of λ. Here I examine GRW in this odd region of parameter space where collapse events act as natural disasters that destroy branches of the wave function along with their occupants. Our continued survival provides evidence that we don't live in a universe like that. I offer a quantitative analysis of how such evidence can be used to assess versions of GRW with small collapse rates in an effort to move towards more principled and experimentally-informed lower bounds for λ. [The published version of this chapter is forthcoming in *Synthese*.]

3.1 Introduction

One central point of disagreement in the foundations of quantum mechanics is whether the collapse of the wave function is a genuine physical process. If collapse is to be taken seriously, we should seek to determine physical laws that might govern this process. Ghirardi-Rimini-Weber theory (GRW) offers possible precise laws which

guarantee that the wave function collapses during familiar quantum measurements. However, observers and measurements have no special status in the theory; collapses happen whether or not scientists are watching.

The laws of GRW include two new fundamental constants not present in textbook discussions of quantum mechanics. One parameter, σ, characterizes the precision of the collapse events and the other, λ, the rate at which collapses occur. If these parameters are chosen properly, the theory appears to succeed in generating the correct probabilistic predictions for experiments taken to be within the purview of non-relativistic quantum mechanics. However, as more experiments are conducted we continue to shrink the space of possible values for σ and λ. Potentially, the allowed region could shrink so much it disappears and GRW could be ruled out. Alternatively, new experiments might confirm GRW over its competitors. As of now, there seems to be a fair amount of leeway as to what values we may assign to the parameters (figure 3.1). Focus on the collapse rate λ. It is fairly well-understood how we can put experimental *upper* bounds on the collapse rate. If collapse events were too frequent, interference patterns would be destroyed by particles collapsing mid-experiment, isolated systems would heat up, undisturbed atoms would spontaneously emit photons, and in other varied ways the experimental predictions of the theory would be corrupted (these constraints have been reviewed recently in Adler, 2007; Feldmann & Tumulka, 2012; Bassi *et al.* , 2013).

In this chapter, I would like to explore how we might put experimental *lower* bounds on the collapse rate λ. The trend in the literature has been to dismiss low values of λ for non-empirical reasons or for reasons that presuppose the failure of the many-worlds interpretation. When λ is very small GRW becomes an odd theory. Macroscopic objects are not prevented from entering superpositions and the theory takes on a many-worlds character (§3.3). Such versions of GRW have been rejected as philosophically unsatisfactory. Surely they are. But, there has been

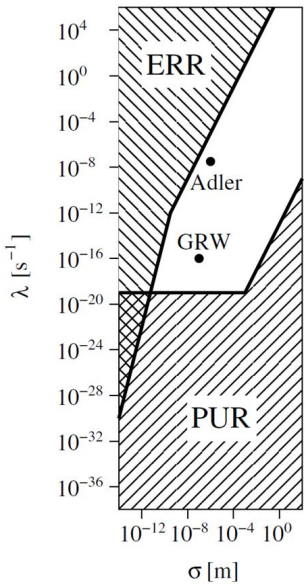

Figure 3.1: **Parameter Diagram of GRW Theory** Figure from Feldmann & Tumulka (2012). ERR is the "empirically refuted region." PUR is the "philosophically unsatisfactory region." The points labeled "GRW" and "Adler" indicate the values suggested in Ghirardi *et al.* (1986) and Adler (2007) respectively. It should be noted that Adler's proposal was made in the context of CSL, not GRW.

disagreement about exactly where the problems arise. Feldmann & Tumulka (2012) give the criterion, "We regard a parameter choice (σ, λ) as philosophically satisfactory if and only if the PO [primitive ontology] agrees on the macroscopic scale with what humans normally think macroscopic reality is like." Bassi *et al.* (2010) impose the requirement that "any superposition reaching the eye must be reduced before it is transformed into a perception in the brain.", building on a suggestion in Aicardi *et al.* (1991). Adler (2007) and Gisin & Percival (1993) argue that the formation of a microscopic latent image in a detector counts as a measurement even before this image is amplified to macroscopic scale. They believe that the collapse rate must be high enough that even these latent images do not enter superpositions.

I will argue that very small values of λ are not just *philosophically* problematic, they are *empirically* unacceptable *even if* the many-worlds interpretation is viable. In doing so, I hope to begin shifting the burden from philosophical considerations to empirical ones and to lay the foundation for a principled and experimentally informed approach to determining lower bounds on λ. Although the chapter will focus on GRW throughout, many of the lessons could be applied *mutatis mutandis* to other collapse

theories.

3.2 GRW Theory

In GRW theory, the evolution of the wave function is typically governed by the familiar Schrödinger equation,

$$i\hbar\frac{d}{dt}\left|\Psi(t)\right\rangle = \widehat{H}\left|\Psi(t)\right\rangle \ . \tag{3.1}$$

At some instants, the evolution of the wave function is discontinuous and not in accord with the Schrödinger equation. The wave function collapses. According GRW, collapse is a real physical process governed by well-defined laws and occurring frequently, not just during measurements. Humans and other observers play no spooky role, they are just particularly intelligent and perceptive collections of particles.

When a collapse occurs a randomly chosen particle has its position become extremely well-localized. Collapses occur randomly at a rate of $N\lambda$ where N is the total number of particles. That is, once a collapse occurs at T_1 the probability that the next collapse, at T_2, will happen within time interval Δt is given by

$$P(T_2 - T_1 < \Delta t) = 1 - e^{-N\lambda\Delta t} \ . \tag{3.2}$$

The collapse rate λ is one of two new constants of the theory, originally suggested to be on the order of $10^{-16}\mathrm{s}^{-1}$ (Ghirardi *et al.* , 1986).[1] The collapse localizes particle I (randomly chosen) around location $\boldsymbol{X}$, where $\boldsymbol{X}$ is chosen randomly with probability

[1] It has been suggested that different particles might collapse at different rates depending on their masses (Pearle & Squires, 1994). The analysis presented here could be applied to such a formulation.

density

$$\rho_I(\boldsymbol{x}) = \lim_{t \nearrow T} \langle \Psi(t) | \, \Lambda_I(\boldsymbol{x}) \, | \Psi(t) \rangle \ . \tag{3.3}$$

"$\lim_{t \nearrow T}$" denotes the limit as t approaches the time of collapse, T, from below. $\Lambda_i(\boldsymbol{x})$ is the collapse operator defined by

$$\Lambda_i(\boldsymbol{x}) = \frac{1}{(2\pi\sigma^2)^{3/2}} e^{-\frac{(\widehat{\boldsymbol{x}}_i - \boldsymbol{x})^2}{2\sigma^2}} \ , \tag{3.4}$$

where $\widehat{\boldsymbol{x}}_i$ is the position operator for particle i. The wave function after the collapse is given by the pre-collapse wave function multiplied by a tightly peaked three-dimensional Gaussian centered about $\boldsymbol{X}$ and normalized,

$$\lim_{t \searrow T} |\Psi(t)\rangle = \lim_{t \nearrow T} \frac{\Lambda_I(\boldsymbol{X})^{1/2} \, |\Psi(t)\rangle}{\langle \Psi(t) | \, \Lambda_I(\boldsymbol{X}) \, | \Psi(t) \rangle^{1/2}} \ . \tag{3.5}$$

The second new constant in GRW, σ, appears in (3.4) and characterizes the width of the Gaussian that localizes the particle. It was originally proposed to be on the order of 10^{-7}m (Ghirardi *et al.* , 1986). In the remainder of the chapter different values of λ will be considered, but σ will be kept fixed at about 10^{-7}m.

In the simplest version of GRW, GRW$\mathbf{0}$, the wave function is all there is and its evolution is determined by the Schrödinger equation (3.1) and the collapse process (3.2, 3.3, 3.5). In the limit where λ is taken to zero, collapse never occurs and GRW$\mathbf{0}$ becomes Everettian quantum mechanics (a.k.a. the many-worlds interpretation or S$\mathbf{0}$). All there is is the wave function and it always evolves in accordance with the Schrödinger equation. Defenders of Everettian quantum mechanics tend to view GRW$\mathbf{0}$ as the right way to think about GRW theory since they think that our experiences of reality can emerge from patterns in wave functions. For Everettians and others who prefer GRW$\mathbf{0}$ to the alternatives below, this chapter can be read as a discussion of GRW$\mathbf{0}$ in the strange regime where it approaches Everettian quantum

mechanics.

For some, GRW$\emptyset$ is unsatisfactory (e.g., Allori *et al.* , 2008, §4.3; Maudlin, 2010). According to GRW$\emptyset$ there are no objects in familiar three-dimensional space, there is only a wave function in an abstract space: a vector in Hilbert space, a complex-valued function on configuration space, or some other exotic beast. In GRWm, the universe contains a wave function which obeys the above dynamics, but that's not all there is, and, in some sense, that's not the important stuff. In particular, it's not the stuff we're made of. In addition to the wave function, there also exists a distribution of matter in three-dimensional space specified by a density,

$$m(\boldsymbol{x}, t) = \langle \Psi(t)| \, \widehat{M}(\boldsymbol{x}) \, |\Psi(t)\rangle \ . \tag{3.6}$$

Here $\widehat{M}(\boldsymbol{x})$ is the mass density operator defined by

$$\widehat{M}(\boldsymbol{x}) = \sum_{i=1}^{N} m_i \, \delta^3(\widehat{\boldsymbol{x}}_i - \boldsymbol{x}) \ . \tag{3.7}$$

In the limit as λ goes to zero, there is no collapse and GRWm becomes Sm, Schrödinger evolution with a mass density (discussed in Allori *et al.* , 2011). Sm is a many-worlds theory much like Everettian quantum mechanics, but where the universe contains a distribution of mass in three-dimensional space in addition to the unitarily evolving wave function. Some think that GRW$\emptyset$ and S$\emptyset$ are unsatisfactory because such laws would not give rise to creatures with conscious experiences like ours, perceiving an apparently three-dimensional world. Readers who think GRW$\emptyset$ is unsatisfactory can understand this chapter as a discussion of GRWm in the awkward bit of parameter space where it approaches Sm. In the following sections, I will not differentiate between GRW$\emptyset$ and GRWm. Read GRW in whichever way you think makes it the stronger theory. Read MWI as S$\emptyset$ if you're reading GRW as GRW$\emptyset$, as

Sm if you're reading GRW as GRWm.

Perhaps neither S$\emptyset$ nor Sm really are many-worlds theories. Without collapse, one might argue, it's not that every outcome of a quantum experiment is observed by a separate copy of the experimenter but that the single experimenter somehow experiences all outcomes at once or otherwise ceases to have a normal mental state. If this is the problem with these theories, GRW will become empirically inadequate as it approaches S$\emptyset$ or Sm. However, determining when and how it becomes inadequate would require a specific account of the abnormality that should be expected in the absence of collapse. For the purposes of this chapter, I will assume that the theory GRW limits to as the collapse rate is taken to zero really is a many-worlds theory.

There is a third version of GRW, GRWf. Here one supplements the wave function with a primitive ontology of flashes. Taking λ to be small in this version of the theory raises entirely different concerns from those faced by GRW$\emptyset$ and GRWm. The problem for GRWf when λ is small is not that human lives are constantly ending, but that such life may be absent altogether. Understanding the empirical adequacy of GRWf in this region of parameter space would require a very different kind of analysis and for that reason GRWf will not be discussed in the remainder of the chapter. A brief discussion of GRWf in this regime can be found in Feldmann & Tumulka (2012, §4).

3.3 Branches and Stumps

GRW was originally formulated with the rate of collapse $\lambda \approx 10^{-16}\text{s}^{-1}$. With this rate, when a measurement occurs the wave function just starts to branch into a superposition of outcomes when, with very high probability, the wave function

collapses to a single definite outcome.[2] This is how GRW solves the measurement problem: a single definite outcome is guaranteed by the rapid collapse of the wave function and the fact that probabilities for collapsing to different outcomes are (approximately) given by the Born rule is a non-trivial consequence of the collapse process (3.2, 3.3, 3.5). If the rate of collapse is taken to zero, then collapses never occur and GRW becomes MWI. In MWI, every possible outcome of a quantum measurement actually occurs.

What if λ is chosen so that it is not quite zero, but is very small ($\lambda \ll 10^{-16}\text{s}^{-1}$, keeping $\sigma \approx 10^{-7}\text{m}$[3])? In this regime collapses occur, but only very rarely. When a collapse occurs, the results are catastrophic. After a spin measurement, the laboratory enters into a superposition of a world in which the scientists record an up result and another in which they record down. Later, if any of the particles that compose the scientists or the measurement readout collapse, one of the worlds will be destroyed. Imagine 15 minutes pass between the moment when the measurement occurred and the time when collapse chooses a world to eliminate.[4] In this time, the scientists in both worlds can walk, think, and talk. After collapse, only one world remains. When a collapse like this occurs, all of the inhabitants of the other world are instantaneously and painlessly killed. Or, maybe the collapse doesn't cause the other world to go out of existence, but instead the tail of the Gaussian distorts the world and alters its

[2]There has been some debate over whether the destruction of other branches is successful; see the literature on the problem of tails. Here I assume that the problem can be solved. If it cannot, GRW is not a viable solution to the measurement problem. In particular, I will assume that if collapse chooses one part of the state and massively shrinks the rest, it is not merely improbable to find oneself in a part of the state that was not fortunate enough to be the center of the collapse, it is impossible. There is no life in those other parts soon after collapse.

[3]This ensures that, in general, a single collapse will be sufficient to destroy branches in which the measurement turned out differently.

[4]This would be typical if we choose λ to be on the order of 10^{-33}s^{-1} and assume that there are about 10^{30} fundamental particles brought into an entangled superposition by the experiment (using (3.2)).

evolution so that it is inhospitable to human life.[5] In this case, death is fairly quick but perhaps not instantaneous. Either way, in this region of parameter space collapses are not helpful shifts which prevent macroscopic superpositions from forming, they're colossal natural disasters.

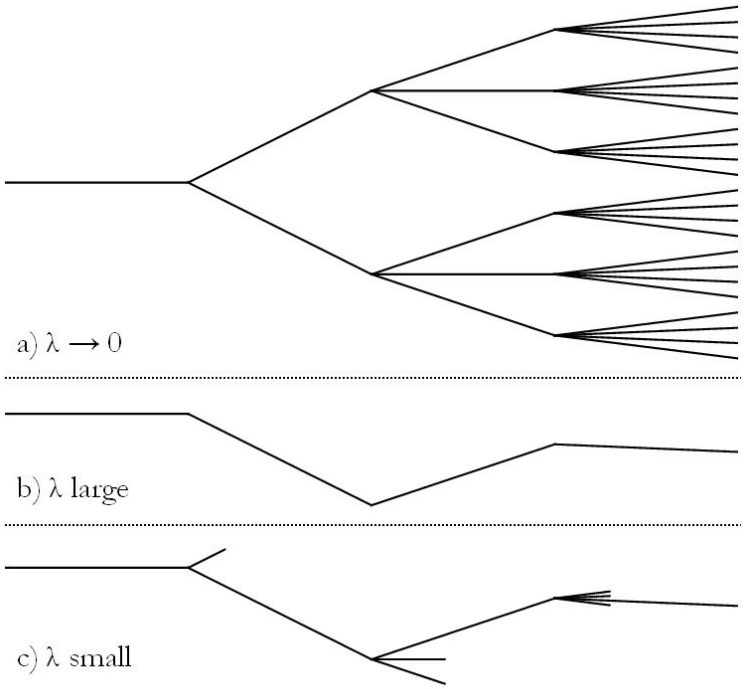

Figure 3.2: **GRW with Different Collapse Rates** Plot of GRW evolution for a sequence of three measurements for different values of λ.

The way the universe—the totality of quantum worlds—evolves in each of these three regions of parameter space is depicted in figure 3.2. With λ at or near zero, worlds branch every time a measurement occurs and each outcome happens on some branch. For standard values of λ, branching is prevented by the collapse of the wave function and each measurement has a definite outcome. For small values of

[5] See Wallace (2014); Vaidman (2014b, §8) for "solutions" to the tails problem along these lines (also briefly discussed in Allori *et al.* (2011, §4)).

λ branching occurs before collapse is able to prevent it; collapse events occur after branching. Living in such a universe is extremely dangerous as entire worlds are constantly being obliterated. If you are lucky enough to find yourself living a long life, you should be shocked. Repeated improbable occurrences often indicate failure of a theory. This is no exception. The data you receive from your survival provides strong *empirical evidence* against the theory.

3.4 The Rarity of Longevity

To judge the empirical adequacy of a given theory, I will focus on the likelihood of the evidence given the theory, $P(\mathcal{E}|\mathcal{T})$. If, for some evidence $\mathcal{E}$ and theories $\mathcal{T}_1$ and $\mathcal{T}_2$, $P(\mathcal{E}|\mathcal{T}_1) > P(\mathcal{E}|\mathcal{T}_2)$, then the evidence $\mathcal{E}$ confirms $\mathcal{T}_1$ over $\mathcal{T}_2$. If one updates on $\mathcal{E}$ by Bayesian conditionalization, then for any theory $\mathcal{T}$, the credence assigned to $\mathcal{T}$ after gaining the evidence can be expressed in terms of the prior probabilities as $P_{post}(\mathcal{T}) = P(\mathcal{T}|\mathcal{E})$.[6] It follows from $P(\mathcal{E}|\mathcal{T}_1)$ being greater than $P(\mathcal{E}|\mathcal{T}_2)$ that, if one changes their credences in response to $\mathcal{E}$ by Bayesian updating, the ratio of one's credence in $\mathcal{T}_1$ to their credence in $\mathcal{T}_2$ will rise,

$$\frac{P_{post}(\mathcal{T}_1)}{P_{post}(\mathcal{T}_2)} = \frac{P\left(\mathcal{E}|\mathcal{T}_1\right)}{P\left(\mathcal{E}|\mathcal{T}_2\right)} \frac{P(\mathcal{T}_1)}{P(\mathcal{T}_2)} > \frac{P(\mathcal{T}_1)}{P(\mathcal{T}_2)} \tag{3.8}$$

[6] Although I expect that this straightforward account of theory confirmation applies to the cases under discussion, one might reasonably be concerned. The situations considered involve *self-locating uncertainty* (see Sebens & Carroll, 2014; Vaidman, 2014a, §4.2) and Bayesian conditionalization must be somehow modified to handle such cases (see Arntzenius, 2003). Some modifications will vindicate the use of conditionalization here, others will not. To avoid controversy, I focus primarily on the probability of the evidence given the theory and not the posterior probabilities that result from updating on the evidence. I approach the problem from the familiar diachronic perspective, taking one's previous beliefs and evidence to together determine what one's current beliefs should be. Alternatively the problem could be approached synchronically, taking one's evidence together with what Meacham (2010) calls an "epistemic kernel" to determine what one's current beliefs should be (there are several competing ways of implementing this approach; see Manley, 2014).

Theories that are empirically equivalent will assign the evidence equal probability and the data that comes in will not discern between them.

The theories which will be compared are: versions of GRW with different parameter values, e.g., $\text{GRW}_{\lambda=10^{-16}\text{s}^{-1}}$; the many-worlds interpretation, MWI; and some unspecified theory which gives the correct Born rule probabilities and guarantees survival, QM.[7] The constraint that QM gives the Born rule probabilities is the constraint that: the probability of seeing the outcome corresponding to eigenvalue O_i of the observable operator $\widehat{O}$ is given by

$$P(O_i|\text{QM}) = |\langle O_i|\Psi\rangle|^2 \ . \tag{3.9}$$

Throughout I'll assume that the agent knows whatever is useful to know about the universal wave function, Ψ, including $|\langle O_i|\Psi\rangle|^2$ for all i. This allows us to focus on the confirmation of alternate dynamical theories without concerning ourselves with the way agents learn about the universe's wave function.

I will initially suppose that MWI is capable of recovering the Born rule probabilities.[8]

Convenient Conjecture In MWI, after a measurement of the observable $\widehat{O}$ has been made and before the outcome is observed, the probability one ought to assign to seeing the outcome corresponding to eigenvalue O_i is given by

[7] What wonderful theory succeeds in recovering the Born rule, as is demanded of the theory I've called "QM"? This will be a matter of disagreement. Let QM stand in for your favorite theory, whichever you think recovers the right probabilities, be it MWI, $\text{GRW}_{\lambda=10^{-16}\text{s}^{-1}}$, Bohmian mechanics, or something else. $\text{GRW}_{\lambda=10^{-16}\text{s}^{-1}}$ predicts deviations from the Born rule for certain yet-to-be-conducted experiments involving, e.g., macroscopic superpositions (which, even if perfectly isolated from the environment, would be predicted to be unstable). However, for the already-conducted experiments typically taken to provide support for quantum mechanics the predictions should (approximately) match those of the Born rule (setting aside the concerns raised in §3.6).

[8] For an extended defense of this conjecture, see Wallace (2012). See also Carroll & Sebens (2014); Sebens & Carroll (2014).

$$P(O_i|\text{MWI}) = |\langle O_i|\Psi\rangle|^2.$$

This is a highly controversial assumption, so let me clarify the spirit in which I am introducing it. In order to put empirical lower bounds on λ we need to consider cases where GRW becomes more and more like MWI. If we don't have quantitative predictions from MWI, it will not be possible to quantify the success of GRW in these bits of parameter space. Later I'll discuss how things change if the conjecture is false (§3.5).

In the notation used here, $\text{GRW}_{\lambda=0}$ is MWI. So, when a measurement is made, $P(O_i|\text{MWI}) = P(O_i|\text{GRW}_{\lambda=0})$. Thus if we are assuming that the Convenient Conjecture is true and thereby that MWI is empirically adequate, it follows that $\text{GRW}_{\lambda=0}$ is empirically adequate as well.

The question, then, is for what values of λ is GRW approximately empirically equivalent to QM and for what values do the predictions of GRW and QM diverge? If the predictions diverge significantly, GRW becomes empirically inadequate—the data we actually have fits the predictions of QM.[9] For the remainder of this section, take the rate of collapse λ to be sufficiently small that whenever a measurement occurs we can expect there to be copies of the experimenter who record each outcome. From the Convenient Conjecture and the fact that the dynamics are the same in GRW and MWI before collapse, it is reasonable to suppose that for these small values of λ the probability of seeing each result is given by

$$P\left(O_i|\text{GRW}_\lambda\right) = |\langle O_i|\Psi\rangle|^2 \ . \tag{3.10}$$

But, the observed experimental outcome is *not* the only data one has to update on. The experimenter should also take into account the fact that she has survived for a

[9]It's fine if the predictions for certain *future* experiments diverge (see footnote 7) since the data might (for all we know) support GRW over alternative formulations of quantum mechanics.

time Δt beyond the moment when the measurement was performed. The probability for surviving to Δt can be calculated as

$$\begin{aligned} P\left(\Delta t|\mathrm{GRW}_\lambda \& O_i\right) &= 1 - P\left(\text{fatal collapse by } \Delta t|\mathrm{GRW}_\lambda \& O_i\right) \\ &= 1 - P\left(\text{death}|\text{collapse by } \Delta t \& \mathrm{GRW}_\lambda \& O_i\right) \times P\left(\text{collapse by } \Delta t|\mathrm{GRW}_\lambda \& O_i\right) \ . \end{aligned} \tag{3.11}$$

The probability of a collapse occurring by Δt can be approximated using (3.2) along with the simplifying assumption that there are N_S particles whose collapse would cause a jump to a single outcome: $P\left(\text{collapse by } \Delta t|\mathrm{GRW}_\lambda \& O_i\right) = 1 - e^{-N_S \lambda \Delta t}$.[10] The probability of dying in the event of such a collapse is just the probability that the collapse is centered around some branch other than one's own: $1 - |\left\langle O_i|\Psi\right\rangle|^2$.[11,12] Inserting these two expressions into (3.11) yields

$$P\left(\Delta t|\mathrm{GRW}_\lambda \& O_i\right) = |\left\langle O_i|\Psi\right\rangle|^2 + e^{-N_S \lambda \Delta t} - |\left\langle O_i|\Psi\right\rangle|^2 e^{-N_S \lambda \Delta t} \ . \tag{3.12}$$

[10] More realistically, N_S would increase as a function of time.

[11] This is an optimistic estimate. In fact there will usually be many worlds corresponding to each outcome and thus even when a collapse is centered on the right outcome O_i, one's world might well be destroyed.

[12] Here, to keep things simple, it is assumed that life on the branches not selected by collapse ends immediately (setting aside the possibility of delayed death mentioned in the previous section and footnote 5).

The probability of the total evidence can be assessed by combining (3.10) and (3.12),[13]

$$\begin{aligned} P\left(O_i\&\Delta t|\mathrm{GRW}_\lambda\right) &= P\left(\Delta t|\mathrm{GRW}_\lambda\&O_i\right) \times P\left(O_i|\mathrm{GRW}_\lambda\right) \\ &= \left(|\left\langle O_i|\Psi\right\rangle|^2 + e^{-N_S\lambda\Delta t} - |\left\langle O_i|\Psi\right\rangle|^2 e^{-N_S\lambda\Delta t}\right)|\left\langle O_i|\Psi\right\rangle|^2 \ . \end{aligned} \tag{3.13}$$

We can better understand this formula by considering a simple case. Imagine $\lambda \approx 10^{-33}\mathrm{s}^{-1}$ and $N_S \approx 10^{30}$ so that the experimenter can expect to have approximately 15 minutes between measurement and collapse (as in footnote 4). In this time, she can form expectations about what will happen and look around. Suppose she sees an outcome, O_A, with low Born rule probability, $|\left\langle O_A|\Psi\right\rangle|^2 = \frac{1}{10}$. She should be somewhat surprised and also afraid. Now she knows that she only has a one in ten chance of survival. If she makes it through the day, she should be surprised again. The probability assigned to the total evidence (surviving and seeing that outcome) is $\frac{1}{10} \times \frac{1}{10} = \frac{1}{100}$, which follows from (3.13) with $\Delta t \gg \frac{1}{N_S\lambda}$.

Consider a variation of the previous case in which the experimenter does not observe the outcome until long after the measurement. Assume for simplicity that there are just two possible outcomes, O_A and O_B, and one branch corresponding to

[13]Two clarifications: First, the proposition signified by "$O_i\&\Delta t$" in (3.13) should be understood as the indexical claim "*I* am alive Δt after the experiment and in *my* world the result of the experiment is O_i." not the weaker claim that "There exists a copy of me who is alive Δt after the experiment and in a world where the result of the experiment was O_i." Long after the experiment, the probability of the second claim is given by $|\left\langle O_i|\Psi\right\rangle|^2$ since it is just the probability that the GRW collapse will select outcome O_i. Thus the weaker claim might appear friendlier to GRW with small λ. Why focus on the stronger claim? The weaker claim does not take into account one's full (indexical) evidence and using it to update probabilities in GRW and MWI leads to unacceptable results (see footnote 24).

Second, the probability of "$O_i\&\Delta t$" (stronger version) is difficult (perhaps impossible) to assess before the measurement since it is unclear whether one, all, or none of the post-branching copies are identical to the original experimenter. Fortunately, we can focus on the probability assigned to "$O_i\&\Delta t$" *immediately after* branching. Since the experimenter doesn't yet know which branch they are on or whether they will survive, it makes sense to assign probabilities at this point. Further, these later probabilities are what matter for theory confirmation as these are the probabilities assigned to the evidence right before the evidence is acquired.

each (see footnote 11). In this case her survival should not be much of a surprise; the probability is 82%. The probability of O_A is 10% and the chance of survival given O_A is 10%. The probability of the other outcome, O_B, is 90% and the chance of survival given O_B is 90%. Thus the total chance of survival is $\frac{1}{10} \times \frac{1}{10} + \frac{9}{10} \times \frac{9}{10} = \frac{82}{100}$. The probability she should assign to O_A given that she survived can be calculated by Bayes' theorem as the probability of survival conditional on O_A, $\frac{1}{10}$, times the probability of O_A, $\frac{1}{10}$, divided by the probability of survival, $\frac{82}{100}$. This yields $\frac{1}{82}$. The probability assigned to her total evidence is the probability of surviving times the probability of seeing O_A upon surviving, $\frac{82}{100} \times \frac{1}{82} = \frac{1}{100}$ (the same result as was obtained in the first case).

In the first case—observation before collapse—the probability assigned to seeing O_A was correct but the subsequent probability of survival was in disagreement with QM. In the second case—collapse before observation—the probability of survival was in disagreement with QM *and* the subsequent probability of O_A was incorrect. The fact that the probability of survival was less than one in the second case shows that GRW_λ (with λ very small) could be disconfirmed by repeated experiments even if no one bothers to look at the results of the experiments. In the second case, unlike the first, O_A is assigned a probability in disagreement with the Born rule. (The reason for this disagreement is that outcomes which were already improbable get further penalized for poorly predicting the agent's survival.) Thus the problem for small values of λ is not *merely* that the probability of survival is low but also that supposing one has survived these dangerous collapse events leads to poor predictions about the outcomes of measurements that have already been made. Note that by focusing on the probability of the total evidence, (3.13), we need not worry about whether collapse happens before or after observation of the outcome.

If λ is so small that no collapses are expected to occur within any reasonable length of time Δt and the Convenient Conjecture holds, the predictions of

GRW_λ approximately match those of QM. However, as has been noted (Feldmann & Tumulka, 2012, §4), there would be little motivation for such a theory. It would be simpler to just set λ to zero and remove the collapses all together, yielding MWI. As λ grows it becomes more likely that a collapse will have occurred within Δt and the disagreement between GRW_λ and QM gets worse. QM predicts that you will be alive whereas GRW_λ assigns a certain probability to your death. For fixed λ, the larger Δt is the larger the disagreement between QM and GRW_λ; see (3.13). However, once λ is sufficiently large the assumption that branching precedes collapse becomes invalid. In the next section I'll consider cases in which branching is prevented by collapse.

The fact that one's own continued survival is used as evidence for assessing theories is undeniably odd. Experimenters don't typically keep track of the time elapsed since the experiment was performed. But, epistemologists have contemplated cases much like this where survival *is* relevant data. Consider the following much-discussed example (Leslie, 1989; Swinburne, 1990):

Firing Squad Suppose that a dozen well-trained shooters are ordered to execute you by firing 12 shots each. While blindfolded you hear 144 shots ring out but you survive unscathed.

In such a scenario, your own survival provides evidence that the shooters intentionally let you live over the alternative hypothesis that you got lucky because each of the 144 shots missed its intended target.

The situation here is similar to *Firing Squad*. The hypothesis that the squad intentionally misses is like the hypothesis that QM is true and there are no cataclysmic collapse events. The hypothesis that the shooters were attempting to kill you is like the hypothesis that GRW_λ is true for some troublesome small-but-not-too-small choice of λ where worlds are constantly snuffed out quickly and without warning. However, there is an important difference: In *Firing Squad*, the target will either

survive or be killed. In GRW_λ with troublesome λ, there will be many versions of the experimenter that are killed and always at least one that survives. A closer non-quantum analogy is:

Prison Poisoning On New Year's Day you wake up in a nondescript prison cell, #27. A coin was flipped. On New Year's Eve, you were blindfolded and shipped either to *Alcatraz*, if heads, or *Arkham*, if tails. Each prison contains 100 numbered cells and you were randomly assigned to #27.[14] While you slept in your cell the new year began with a randomly chosen 99 of the 100 cells in Arkham being filled with deadly poison gas. Those in Alcatraz were safe. You knew the plan all along.

In this case, you should initially think it equally likely that you ended up in either prison. After surviving the night you should come to believe that you were probably shipped to Alcatraz since being shipped to Arkham would have likely resulted in your death. It was guaranteed that one of the prisoners in Arkham would survive, but it was not likely to be the one in cell #27. Alcatraz is like MWI and Arkham is like GRW with troublesome λ. The numbered cells represent 100 possible results of a measurement and the gas plays the role of collapse. For an analogue of GRW with a normal collapse rate, one could introduce the possibility of being sent to a third prison with a single cell, randomly numbered and free of poison.[15]

Cases like *Firing Squad* and *Prison Poisoning* have a curious feature: in both scenarios, one hypothesis cannot be confirmed by the subject. If the bullets and poison kill instantly, no course of experience would support the hypothesis that the squad was trying to kill you or that you were sent to Arkham. Similarly, if collapse

[14] For the closest analogy, imagine that each cell of your prison is occupied by a copy of you that resulted from a 1-to-100 fission midday on New Year's Eve.

[15] In this case, the fission in footnote 14 should not be supposed.

kills instantly there are no experiences one could have that would provide evidence for GRW with troublesome λ over QM (*if* the Convenient Conjecture holds). In *Firing Squad*, the problem with the hypothesis that the squad is trying to kill you is not that it predicts odd experiences but that it predicts your experiences will end. In GRW with small λ too the problem is not odd experiences. In fact, in a theory like $\mathrm{GRW}_{\lambda=10^{-33}\mathrm{s}^{-1}}$ if the outcomes of many repeated experiments are recorded one can expect[16] with high probability that at any given time the record will show a sequence of results that looks randomly generated with each outcome's probability weighted by $|\langle O_i|\Psi\rangle|^2$.[17] If a branching has just occurred, there may be multiple distinct versions of the record but each will show the same long-run frequencies for the various outcomes as the records will only differ in the last few entries. At any time, a typical observer will remember, and have records of, measurement results that fit the predictions of QM. This feature of the theory might cause one to doubt whether we could have empirical evidence against $\mathrm{GRW}_{\lambda=10^{-33}\mathrm{s}^{-1}}$, but it shouldn't.[18] As in *Prison Poisoning* and *Firing Squad*, the fact that one has survived *is relevant*

[16] That is, looking at the probabilities derived from the collapse process this is what one should expect. As discussed earlier in this section, if one performs experiments, survives for a long time, and doesn't look at the outcome, the probabilities that should be assigned to the different possible outcomes are not the standard Born rule probabilities. Seeing a sequence that fits *these* expectations should shift one's credence towards GRW with small lambda—but, only after the theory has been significantly disconfirmed by one's survival. Thus unlike *Firing Squad* and *Prison Poisoning*, there is in fact a way to get a piece of evidence that points towards the dangerous hypothesis. Still, one's total course of experience will never favor GRW with small λ over QM; as can be seen by noting that the expression in parentheses in (3.13) is at most one.

[17] If the collapse rate is much smaller, there will be many records only some of which show sequences deemed probable by the Born rule (as in MWI).

[18] Wallace (2014) in considering a similar situation seems to find this—"strictly speaking"—sufficient empirical success as the theory does manage to "explain why the scientific community has so far observed statistical results in accord with quantum mechanics (via the anthropic fact that worlds in which violations were observed are now radioactive deserts [the fate he believes befalls worlds in the tails]). And it explains why it is rational to act as if the predictions of quantum mechanics were true (because in those worlds where they turn out false, we're all doomed anyway)." Vaidman (2014b, §8) also seems untroubled by the possibility of death in the tail branches.

evidence in determining which hypothesis to believe. In contrast to GRW with small λ, MWI predicts that there will at any time be many observers whose memories and records don't fit the predictions of QM—every sequence of quantum measurement results is observed by the inhabitants of some quantum world. The theory then faces the challenge of explaining why we should not expect to be one of these observers (the challenge of establishing the truth of the Convenient Conjecture).

Those who are attracted to the idea of quantum immortality may object to the conclusions reached in this section. Consider a dangerous branching event from the perspective of the many-worlds interpretation (a "quantum suicide" scenario). Suppose you will survive on one branch and die immediately, or quickly, on all others. It is tempting to think you should expect survival with certainty. As Lewis (2004) put it, "The experience of being dead should never be expected to any degree at all, because there is no such experience." If death is indeed immediate on all branches but one, the thought has some plausibility. But if there is any delay it should be rejected. In such a case, there is a short period of time when there are multiple copies of you, each (effectively) causally isolated from the others and able to assign a credence to being the one who will live.[19] Only one will survive. Surely rationality does not compel you to be maximally optimistic in such a scenario.[20] The situation in GRW with a troublesome collapse rate is just like the delayed-death version of the above quantum suicide scenario and, as in that case, survival should not receive probability one.[21] If the collapse rate is raised so that the agent never splits into multiple copies,

[19] Do the copies need to last long enough to have thoughts to cause trouble? I think not. If you survive, you can consider what credences you should have assigned during the short period after splitting when you coexisted with the other copies.

[20] The situation here is like that of the prisoner in Arkham if the period between the splitting event (see footnote 14) and the deaths were made much shorter.

[21] Anticipating an upcoming branching and subsequent collapse, can one assign a probability to survival *before* splitting? If so, does the fact that some successor will survive the collapse mean that before splitting survival is certain? These questions need not be answered as the relevant probabilities are those assigned immediately after branching (footnote 13).

there is no danger of death and survival can be expected with certainty.

3.5 Averting Branching

If collapse occurs sufficiently soon after a measurement, branching can be averted. As the other branches of the universe where the outcome was different are just beginning to form the collapse event occurs, ensuring that the macroscopic readout gives a definite result and the experimenter sees a single outcome. The simplest way to incorporate this feature of the theory is by introducing a cutoff characterizing the amount of time that passes before branching occurs if there is no collapse. If a collapse happens within τ, branching is averted and a single outcome occurs. If collapse does not occur until after τ, then there is a branching of worlds before the collapse, as in the previous section.[22] Let $C_{<\tau}$ indicate that collapse occurs before the cutoff, $C_{>\tau}$ indicate after. Including both of these possibilities, the probability of the data given the theory can be expressed as

$$P\left(O_i\&\Delta t|\mathrm{GRW}_\lambda\right) = \overbrace{P\left(O_i\&\Delta t|\mathrm{GRW}_\lambda\&C_{>\tau}\right)}^{①} \times \overbrace{P\left(C_{>\tau}|\mathrm{GRW}_\lambda\right)}^{②} + \underbrace{P\left(O_i\&\Delta t|\mathrm{GRW}_\lambda\&C_{<\tau}\right)}_{③} \times \underbrace{P\left(C_{<\tau}|\mathrm{GRW}_\lambda\right)}_{④} . \tag{3.14}$$

The first piece, ①, is just as in (3.13) where it was assumed that branching preceded collapse. The fourth piece, ④, is the probability that a collapse happens by τ. This follows directly from (3.2), ④ $= 1-e^{-N_S\lambda\tau}$. The second piece is simply the probability that a collapse does *not* occur, ② $= 1-$ ④. The third piece, ③, is the probability that a given outcome resulted from the GRW collapse process in a case where branching

[22]The cutoff τ is not a free parameter and not derived from the collapse process. It could be calculated by determining when branching occurs in the absence of collapse (see §3.6).

does not occur. Here we have GRW working as intended and the probability should be in approximate agreement with the Born rule provided λ is not so large as to push us into the empirically refuted region of parameter space (figure 3.1), ③ $\approx |\langle O_i|\Psi\rangle|^2$.[23] Inserting these expressions in (3.14) and rearranging gives,

$$P\left(O_i\&\Delta t|\mathrm{GRW}_\lambda\right) = |\langle O_i|\Psi\rangle|^2 - \Big(1 - |\langle O_i|\Psi\rangle|^2\Big)\Big(1 - e^{-N_S\lambda\Delta t}\Big)|\langle O_i|\Psi\rangle|^2 e^{-N_S\lambda\tau}\ , \tag{3.15}$$

which limits to the Born rule probabilities as λ goes to zero or infinity. (3.15) is not valid if λ is large enough that the probabilities in ③ deviate significantly from those given by the Born rule. It cannot be extended in a simple and general manner as the way in which ③ deviates from $|\langle O_i|\Psi\rangle|^2$ will be depend on the particular experiment under consideration.

In this simplified story, the probability of surviving to Δt and seeing a certain outcome O_i depends dramatically and discontinuously on whether collapse happens before or after branching. The expressions for ① and ③ are quite different. A more careful analysis would ideally give a smooth transition or justify a precise cutoff, but this would require wading into the murky territory of collapses that occur *during* branching and settling questions of personal identity there (in particular, when exactly personal fission occurs and whether it can, in any relevant sense, partially occur). It might be seen either as intriguing or disconcerting that we must answer questions of personal identity in the context of MWI to put precise lower bounds on λ in GRW.

To recap: If λ is so extremely small that you should not expect (relevant) collapses to have occurred in your lifetime (figure 3.2.a), then GRW_λ is empirically adequate if the Convenient Conjecture holds. If λ is large enough that collapses must be considered but small enough that branching typically precedes collapse (figure 3.2.c),

[23] Here it is also assumed that we're considering familiar experiments, not future ones that probe smaller values of λ (see footnote 7).

then early death is the norm and one's continued survival provides strong evidence against the theory. If λ is increased to around the initially proposed value of 10^{-16}s^{-1} (figure 3.2.b), the theory may again be empirically adequate as branching is prevented by collapse and the collapse process ensures that the probabilities of various outcomes are given by the Born rule. If λ is increased even further, so that $\lambda > 10^{-8}$, the theory is again empirically inadequate as collapses occur too frequently. Superpositions are destroyed mid-experiment and other maladies ensue (see Feldmann & Tumulka, 2012; Bassi *et al.* , 2013).

What happens if the Convenient Conjecture is false and MWI gives different probabilities from QM? Then, $\text{GRW}_{\lambda=0}$ is empirically inadequate as $\text{GRW}_{\lambda=0}$ is MWI. This failure also rules out GRW_{λ} for very small λ where collapses can be neglected. For larger values of λ where collapse is rare but relevant, there are now two ways in which the theory fails: the probabilities of the various outcomes are incorrect and there is, in general, some probability that one would not have survived to Δt.[24] For still larger values of λ that successfully avert branching, the theory again has a chance of being empirically adequate since the probabilities of outcomes are now determined by the collapse process and the MWI probabilities are irrelevant.

3.6 The Race: Decoherence versus Collapse

For GRW to be tenable, there must be values of λ for which the theory is empirically adequate. On the one hand, λ must be large enough that collapse practically never

[24] When Δt is large, the probability of the total evidence given by (3.13) is $|\langle O_i|\Psi\rangle|^4$ because the probability of seeing O_i is $|\langle O_i|\Psi\rangle|^2$ and the probability of subsequently surviving is also $|\langle O_i|\Psi\rangle|^2$. This suggests a crafty maneuver to save GRW with small λ. What if the probability of each outcome were one—after all, each outcome will be recorded by someone—so that the probability of total evidence is $|\langle O_i|\Psi\rangle|^2$ as in QM? This proposal faces a fatal problem: practically every experiment would confirm MWI over single-world theories (see Greaves, 2007, §4).

occurs after the experimenter has branched into multiple copies. Otherwise, one's continued survival empirically refutes GRW_λ, (3.15). On the other hand, λ must be small enough that collapses do not spoil the predictions for experiments that have already been performed. That is, λ must lie below the experimentally refuted region of figure 3.1. But, are there any values in this range? To answer this, we need to determine whether decoherence-induced branching tends to occur before or after collapse.[25]

We know that for values of λ near the originally suggested value, 10^{-16}s^{-1}, the experiment readout and the experimenter are in a well-defined state corresponding to a single outcome very soon after the measurement occurs. But, what is not clear is which of two possibilities occurred immediately after the measurement (figure 3.3): (a) the world briefly branched and then a collapse event destroyed some of the copies of the experimenter, or (b) there was never a branching event because collapse prevented the microscopic superposition from causing the experimenter to enter into a superposition.

A proper analysis is warranted, but beyond the scope of this chapter. Here is a *very* rough calculation of how quickly collapse would have to occur to prevent decoherence-induced branching: Decoherence is fast. A slow estimate might be 10^{-23}s for 1 gram of matter at room temperature in a superposition of two locations separated by one centimeter (Zurek, 2003). To ensure a 95% probability of collapse by 10^{-23}s, λ would have to be at least 3 s^{-1} (from (3.2), assuming the number of particles is on the scale of moles, $N = 10^{23}$). But, experiments restrict λ to being at most 10^{-8}s^{-1} (figure 3.1). This calculation suggests trouble. There may not be a safe region of parameter space.[26]

[25] See Schlosshauer (2005, §IV.E); Bacciagaluppi (2012, §3.1.2) for discussion of decoherence in GRW.

[26] This concern is corroborated by the calculations in Tegmark (1993); Benatti *et al.* (1995).

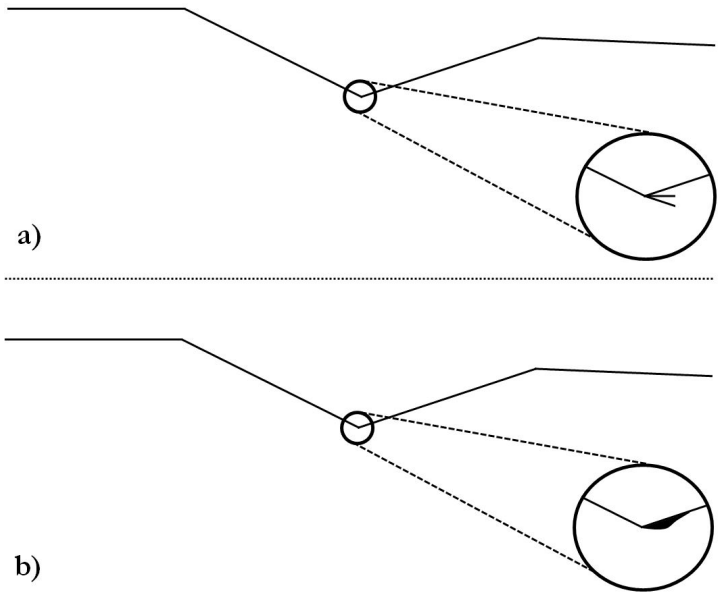

Figure 3.3: **Two Potential Close-ups of Figure 3.2b**

Let me highlight one particularly pernicious simplification in this rough calculation: It is assumed that the bit of matter starts in a superposition. In actuality, it would take time for the matter to enter a superposition and a collapse event could occur in this interval, preventing the macroscopic superposition from forming. It would (normally) take even longer for a large object like a human to enter a superposition of spatially disjoint states (although this is certainly asking too much; the presence of a far less dramatic superposition should be sufficient for the human to branch).

I'll close by summarizing the key lessons of the analysis. First, to determine precise experimental bounds on the parameters λ and σ in GRW, we must answer metaphysical and epistemological questions about MWI: When/how does branching occur in the absence of collapse (§3.5 & 3.6)? What probabilities should be assigned to different outcomes in MWI (§3.4)? This provides additional motivation for that ongoing research program. Second, even if the Convenient Conjecture

holds and MWI is empirically adequate, some of the philosophically unsatisfactory region of parameter space is also empirically refuted (§3.3, 3.4, & 3.5). Surprisingly, it is not refuted by the outcomes we observe, but by the fact that we live long enough to observe so many of them. Third, it is not clear how to draw a principled border for the philosophically unsatisfactory region if our dissatisfaction is purely "philosophical" (§3.1). But, with the realization that small values of the collapse rate λ are empirically refuted, we now have a method to begin drawing principled lower bounds on λ: determine whether the experimenter branches before or after collapse (§3.5 & 3.6). Simple calculations suggest that the lower bound generated from empirical considerations will be stronger than the bound generated from a distaste for long lasting macroscopic superpositions, perhaps strong enough to rule out GRW entirely (§3.6). This merits further study.

Acknowledgments: Thanks to David Albert, David Baker, Gordon Belot, Cian Dorr, J. Dmitri Gallow, Jeremy Lent, David Manley, Laura Ruetsche, Roderich Tumulka, and an anonymous referee. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 0718128.

References

Adler, Stephen L. 2007. Lower and Upper Bounds on CSL Parameters from Latent Image Formation and IGM Heating. *Journal of Physics A: Mathematical and Theoretical*, **40**(12), 2935.

Aicardi, Franca, Borsellino, Antonio, Ghirardi, Gian Carlo, & Grassi, Renata. 1991. Dynamical Models for State-Vector Reduction: Do they ensure that measurements have outcomes? *Foundations of Physics Letters*, **4**(2), 109–128.

Allori, Valia, Goldstein, Sheldon, Tumulka, Roderich, & Zanghì, Nino. 2008. On the Common Structure of Bohmian Mechanics and the Ghirardi–Rimini–Weber Theory. *The British Journal for the Philosophy of Science*, **59**(3), 353–389.

Allori, Valia, Goldstein, Sheldon, Tumulka, Roderich, & Zanghì, Nino. 2011. Many Worlds and Schrödinger's First Quantum Theory. *The British Journal for the Philosophy of Science*, **62**(1), 1–27.

Arntzenius, Frank. 2003. Some Problems for Conditionalization and Reflection. *The Journal of Philosophy*, **100**(7), 356–370.

Bacciagaluppi, Guido. 2012. The Role of Decoherence in Quantum Mechanics. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*, winter 2012 edn.

Bassi, Angelo, Deckert, Dirk-André, & Ferialdi, Luca. 2010. Breaking Quantum Linearity: Constraints from Human Perception and Cosmological Implications. *EPL (Europhysics Letters)*, **92**(5), 50006.

Bassi, Angelo, Lochan, Kinjalk, Satin, Seema, Singh, Tejinder P, & Ulbricht, Hendrik. 2013. Models of Wave-Function Collapse, Underlying Theories, and Experimental Tests. *Reviews of Modern Physics*, **85**(2), 471.

Benatti, F, Ghirardi, Gian Carlo, & Grassi, R. 1995. Quantum Mechanics with Spontaneous Localization and Experiments. *Pages 263–279 of:* Beltrametti, Enrico, & Lévy-Leblond, Jean-Marc (eds), *Advances in quantum phenomena*. Springer.

Carroll, Sean M., & Sebens, Charles T. 2014. Many Worlds, the Born Rule, and Self-locating Uncertainty. *In:* Struppa, Daniele, & Tollaksen, Jeff (eds), *Quantum Theory: A Two-Time Success Story: Yakir Aharonov Festschrift*. Springer.

Feldmann, William, & Tumulka, Roderich. 2012. Parameter Diagrams of the GRW and CSL Theories of Wavefunction Collapse. *Journal of Physics A: Mathematical and Theoretical*, **45**(6), 065304.

Ghirardi, GianCarlo, Rimini, Alberto, & Weber, Tullio. 1986. Unified Dynamics for Microscopic and Macroscopic Systems. *Physical Review D*, **34**(2), 470.

Gisin, Nicolas, & Percival, Ian C. 1993. The Quantum State Diffusion Picture of Physical Processes. *Journal of Physics A: Mathematical and General*, **26**(9), 2245.

Greaves, Hilary. 2007. On the Everettian Epistemic Problem. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **38**(1), 120–152.

Leslie, John. 1989. *Universes*. Routledge.

Lewis, David. 2004. How Many Lives has Schrödinger's Cat? *Australasian Journal of Philosophy*, **82**(1), 3–22.

Manley, David. 2014. On Being a Random Sample. Manuscript.

Maudlin, Tim. 2010. Can the World be Only Wavefunction? *Pages 121–143 of:* Saunders, Simon, Barrett, Jonathan, Kent, Adrian, & Wallace, David (eds), *Many Worlds?: Everett, Quantum Theory, & Reality*. Oxford University Press.

Meacham, Christopher J. G. 2010. Unravelling the Tangled Web: Continuity, Internalism, Uniqueness and Self-Locating Belief. *Pages 86–125 of:* Szabó Gendler, Tamar, & Hawthorne, John (eds), *Oxford Studies in Epistemology*, vol. 3.

Pearle, Philip, & Squires, Euan. 1994. Bound State Excitation, Nucleon Decay Experiments and Models of Wave Function Collapse. *Phys. Rev. Lett.*, **73**, 1–5.

Schlosshauer, M. 2005. Decoherence, the Measurement Problem, and Interpretations of Quantum Mechanics. *Reviews of Modern Physics*, **76**(4).

Sebens, Charles T., & Carroll, Sean M. 2014. Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics. arXiv:1405.7577 [quant-ph].

Swinburne, Richard. 1990. The Argument from the Fine Tuning of the Universe. *In:* Leslie, John (ed), *Physical Cosmology and Philosophy*. MacMillan, New York.

Tegmark, Max. 1993. Apparent Wave Function Collapse Caused by Scattering. *Foundations of Physics Letters*, **6**(6), 571–590.

Vaidman, Lev. 2014a. Many-Worlds Interpretation of Quantum Mechanics. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*, spring 2014 edn.

Vaidman, Lev. 2014b. Quantum Theory and Determinism. *Quantum Studies: Mathematics and Foundations*, **1**(1-2), 5–38.

Wallace, David. 2012. *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. Oxford University Press.

Wallace, David. 2014. Life and Death in the Tails of the GRW Wave Function. arXiv:1407.4746 [quant-ph].

Zurek, Wojciech H. 2003. Decoherence and the Transition from Quantum to Classical–REVISITED. quant-ph/0306072.

Chapter 4

A Laws-First Introduction to Quantum Field Theory

Abstract: Here I present an atypical introduction to the foundations of relativistic quantum field theory (QFT). I seek to be especially clear about the space of physical states and the laws of the theory, as well as the connection between QFT and two of its predecessors: quantum mechanics and classical field theory. Part I introduces QFT as an extension of non-relativistic quantum mechanics with two important modifications (introduced one at a time): the number of particles is allowed to be indeterminate and the energy of a state is given by a relativistic expression. In part II, I present QFT as a quantum version of the classical theory of fields where the the field is permitted to be in a superposition of distinct states, described by a wave functional. The limiting case of classical field theory is then derived using path integrals. Throughout, I use the Schrödinger picture. I hope to prepare readers for derivations of Feynman rules and experimental predictions, but I do not cover such machinery here. I further limit my treatment by not discussing (much) spin, fermions, or renormalization. I will instead focus on theories of interacting spin-0 bosons (or real scalar fields, depending on how you look at it).

4.1 Introduction: Schrödinger-picture QFT in the Particle and Field Bases

When we first meet relativistic quantum field theory (QFT) we are presented with a very strange theory that seems quite different from special relativity, electromagnetism, quantum mechanics, and other physical theories we have come to know and love. In moving from classical particle or field theories to QFT we are faced with a variety of changes: the theory is (in some sense) quantized, we move from talk of states to speaking almost exclusively of operators, the number of particles is allowed to vary with time, particles are treated as (in some sense) excitations of a field, and the theory is made relativistic. To provide a clear and intuitive entry into the theory, these modifications should be made carefully and in the right order. If this is done properly, students should be able to apply the intuitive grasp they had of other theories to better understand QFT and visualize possible phenomena. In this chapter, I offer two ways to develop QFT that I find intuitive and thus hope will be pedagogically useful and helpful for foundational and philosophical work. My treatment will ignore the history of rejected theories leading to QFT and instead focus on the logical structure of the successful theory we have arrived at.[1]

It is too easy to learn QFT without answering the following central questions

[1]The treatment here is largely a synthesis of bits and pieces from various sources. Part I develops QFT along the lines of Schweber (1961) and Teller (1997). The variable-particle QM theory I introduce in §4.3 & 4.4 is similar to the one briefly discussed in (Srednicki, 2007, pg. 11-13). The version of QFT presented in part I is similar to that utilized in Bell-type Bohmian QFTs (Tumulka, 2007; Dürr *et al.* , 2004, 2005). Part II presents a similar formulation to Hatfield's (1992) "Schrödinger representation." The treatment of path integrals in appendix 4.A is significantly different from, but closest to, that in (Hatfield, 1992). The classical field theory limit comes from Zee's (2010) textbook. The version of QFT in part II provides the starting point for Bohm and Hiley's (1993) version of Bohmian QFT. Part III is largely original, but incorporates insights from (Hatfield, 1992, ch. 10 & 11).

(which will be emphasized in this introduction): What are the physical states of systems in QFT? What are the dynamical law(s) that these states obey? How could QFT possibly limit to both QM in the low energy regime and classical field theory when quantum effects are negligible? What is the quantum field operator $\widehat{\phi}(\vec{x})$? I will provide answers to these questions in this chapter.

The order in which one introduces the complexities associated with QFT is key to the intelligibility of the presentation. I will begin part I by briefly reviewing non-relativistic quantum mechanics (QM). In section 4.3 I will move to a new version of QM in which a system can be in a superposition of states with different total particle number, allowing for the possibility that particles may be created or destroyed. In section 4.5 I will introduce a relativistic Hamiltonian in order to complete the move from non-relativistic quantum mechanics to relativistic quantum field theory. In section 4.6, I will introduce some more sophisticated technical machinery including the Lagrangian and the quantum field operator $\widehat{\phi}$. In the last section of part I, I will discuss a simple interaction term.

In part II, I present QFT as a quantum version of classical field theory. In quantum mechanics, the instantaneous state of the world is representable as a wave function over the space of possible particle configurations (instead of a point in configuration space, as it is in classical particle mechanics). In quantum field theory, we can represent the state of the world as a wave *functional* over the space of possible *field* configurations. The evolution of this wave functional is governed by a version of the Schrödinger equation. Using this Schrödinger equation, we can derive an expression for the path integral in QFT (appendix 4.A) and use the path integral to show that QFT limits to classical field theory in the appropriate regime.

In part III I will show that the presentations of QFT in part I and II, although quite different in appearances, are in fact equivalent. Part I presents QFT in a particular *basis*, the particle basis. Part II presents QFT in the field basis. To demonstrate that

the theories differ only in a choice of basis, we will show that the definitions and laws introduced in part I can be derived from the formulation in part II (and vice versa). At the end of part III, we will very briefly discuss the measurement problem as it manifests itself in QFT and use the development of QFT presented here to introduce some proposed solutions.

The ideal audience for this piece is someone who has struggled through the first few chapters of a QFT textbook and/or the first few weeks of an introductory graduate course in quantum field theory (or much more), but feels like they still don't have an intuitive grip on the theory (who does?). Srednicki's (2007) textbook would be an ideal accompaniment to the presentation here as it is best aligned with the order of development and choice of notation, but Peskin and Schroeder's (1995) would be fine too. I hope that this introduction proves useful to both physicists and philosophers of physics.

There are many places where the story told here may be found lacking in sufficient rigor for the precise philosopher or mathematician. There are interesting questions to be asked about defining localized particle states, rigorously computing path integrals, carefully explicating the ways in which certain theories are reduced to others, etc. Here I don the hat of the working physicist and attempt to give as intuitive a picture of textbook QFT as I can, leaving the task of further precisification to others.

In this chapter I will focus quantum field theories for spin-0 bosons, ignore the complications introduced by spin, and, as much as possible, not worry about renormalization. Throughout the chapter I will take $\hbar = c = 1$.

Part I

From Quantum Mechanics to QFT

4.2 Non-relativistic Quantum Mechanics in Configuration Space

In this section we will very briefly review quantum mechanics in the position basis. In QM, the state of a single isolated spin-0 particle at time t is specified by giving a complex valued function $\Psi(\vec{x}, t)$ over position space. The Schrödinger equation determines how $\Psi(\vec{x}, t)$ evolves over time: $i\frac{\partial}{\partial t}\Psi(\vec{x}, t) = \widehat{H}\Psi(\vec{x}, t)$. For a single particle in a time-independent potential, the Hamiltonian $\widehat{H}$ would be: $\left(-\frac{\nabla^2}{2m} + V(\vec{x})\right)$. The probability of finding the particle in a volume V at time t is given by $\int_V d^3x \, |\Psi(\vec{x}, t)|^2$.

When we move to a system which contains n particles, the wave function is defined over *configuration space* (the $3n$-dimensional space of all possible positions for each of the n particles): $\Psi_n(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n)$. The Schrödinger equation for a general state vector is:

$$i\frac{\partial}{\partial t}\Psi_n(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n, t) = \widehat{H}\Psi_n(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n, t)\,. \tag{4.1}$$

The (configuration space) probability density for a certain configuration of particle positions is given by $|\Psi(\vec{x}_1, ..., \vec{x}_n)|^2$. This can be integrated over a volume for each particle to get the probability of finding the particles in those regions. States must

be normalized as follows:

$$\int d^3x_1 \int d^3x_2 ... \int d^3x_n \Psi^*(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n)\Psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n) = 1\,. \tag{4.2}$$

As mentioned earlier, we will be focusing on bosons in this chapter. For identical bosons the wave function must be symmetric under particle permutations:

$$\forall i, j \quad \Psi(\vec{x}_1, ..., \vec{x}_i, ..., \vec{x}_j, ..., \vec{x}_n) = \Psi(\vec{x}_1, ..., \vec{x}_j, ..., \vec{x}_i, ..., \vec{x}_n)\,. \tag{4.3}$$

One could limit the space on which Ψ is defined to the space of *unique* configurations of particles (that is, to the space of configurations that do not differ by simply swapping identical particles). As it makes the calculations easier, we will simply require that the Ψs we consider obey eq. 4.3. In the box below, QM is summarized so we can easily compare it to theories introduced later.[2]

[2]Here the theory is presented in a rough way so that the details could be spelled out by the various interpretations of QM. I present it in this simplified way so we don't (yet) have to address questions like: Is Schrödinger's equation the only equation of motion or does collapse occur? What counts as an "appropriate experiment"? See §4.13 for a brief discussion of questions like these.

Theory 1: n-particle QM (in the position basis)

States The states of the theory are wave functions on n-particle configuration space: $\Psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n)$

Schrödinger Equation The Hamiltonian as it appears in Schrödinger's equation can be decomposed into a free part and a term representing interactions and external potentials:

$$i\frac{\partial}{\partial t}\Psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n, t) = \left(-\frac{\nabla^2_{x_1}}{2m} - ... - \frac{\nabla^2_{x_n}}{2m} + \widehat{H}_{int}\right)\Psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n, t)\,. \tag{4.4}$$

Probabilities The probability of observing the system in state $|\Phi\rangle$ (or probability *density*, depending on the state $|\Phi\rangle$) in an appropriate experiment is given by the square of the inner product of $|\Phi\rangle$ and the system's state $|\Psi\rangle$:

$$|\langle\Phi|\Psi\rangle|^2 = \left|\int d^3x_1 d^3x_2 ... d^3x_n \Phi^*(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n)\Psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n)\right|^2 . \tag{4.5}$$

A note on **Probabilities**: If $|\Phi\rangle$ is properly normalized, like the ground state of the harmonic oscillator, then $|\langle\Phi|\Psi\rangle|^2$ gives a probability. If $|\Phi\rangle$ is delta-function normalized, like a position or momentum eigenstate, then $|\langle\Phi|\Psi\rangle|^2$ gives a probability density.

4.3 Superpositions of Particle Number

In QM as discussed above, the wave function is defined over a space where every state it gives a probability to has the same number of particles. If the wave function is a function of the positions of 7 particles, then the states it assigns probability to are all different configurations of those 7 particles. However, in the real world we know that particles can be created and destroyed. So, it would be nice to have a theory that assigns non-zero probability to changes in particle number.

It is simple enough to extend QM to allow for superpositions of states with different numbers of particles. In the last section we noted that for every number of particles n you might have, there is a different $3n$-dimensional space over which the wave function is defined. Why can't we just define the wave function as assigning a complex number to every point in each one of these configuration spaces? In other words, why not have the wave function assign a complex number to every possible arrangement of any number of particles? To pick out this wave function $|\Psi\rangle$, we will have to specify a function $\Psi_1(\vec{x}_1, t)$ which gives an amplitude to each possible position that a single particle might be in; and we will have to specify a function $\Psi_2(\vec{x}_1, \vec{x}_2, t)$ which gives an amplitude to each possible set of positions that two particles might have; and we will have to give Ψ_3 and Ψ_4 and Ψ_5 and etc. all the way out to infinity (if the probability of a large number of particles falls off sufficiently quickly, the state can still be normalized). Further, there will be a complex number $\Psi_0(t)$ which gives an amplitude to the vacuum (a state of zero particles). The wave function is now a map from elements of the disjoint union of all n-particle configuration spaces to complex numbers.[3] This constitutes a move from defining the state $|\Psi\rangle$ as an element of a

[3]See (Dürr *et al.* , 2004) for a different choice of notation for representing variable-particle wave functions.

Hilbert space to defining $|\Psi\rangle$ as an element of a *Fock* space. Figure 4.1 provides a visual representation of the wave function (squared) for a variable-particle-number system.

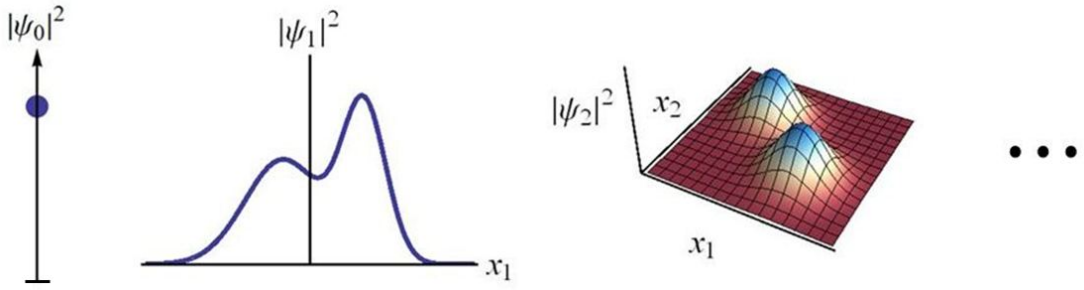

Figure 4.1: **Wave Function in Variable-particle QM** When we allow the number of particles to vary, we define Ψ on the disjoint union of n-particle configuration spaces. These graphs depict the amplitude-squared of a *single* wave function $|\Psi\rangle$ in *one* spatial dimension. The first plot shows the amplitude-squared assigned to the possibility of zero particles, the next shows the possible locations of a single particle and their probability densities, etc.. Note that in the two particle space, $|\Psi_2|^2$ is invariant under a permutation of the two particles.

The equation of motion for this state is simply the Schrödinger equation (eq. 4.1) imposed on *each* configuration space. For the case of non-interacting particles where each particle experiences the same time-independent potential $V(\vec{x})$, the equation becomes:

$$\forall n, \quad i\tfrac{\partial}{\partial t}\Psi_n(\vec{x}_1, ..., \vec{x}_n, t) = \left(-\tfrac{\nabla^2_{x_1}}{2m} - ... - \tfrac{\nabla^2_{x_n}}{2m} + V(\vec{x}_1) + ... + V(\vec{x}_n)\right)\Psi_n(\vec{x}_1, ..., \vec{x}_n, t)\,. \tag{4.6}$$

With this Hamiltonian, there is no probability for changes in particle number. If the wave function is initially only non-zero on the configuration space for 64 particles, there will be zero probability for any state in the future with more or less than 64 particles. This is *not* true for a general Hamiltonian. Probability is conserved so

long as the Hamiltonian is Hermitian[4], but many Hermitian Hamiltonians in which particles can interact will allow for changes in total particle number (see end of §4.4 for an example).

We can easily define the inner product on this space. We can take states of different definite particle number to be orthogonal and states of identical particle number to have the standard inner product on a configuration space of that dimension. For a pair of arbitrary states, this leads to the expression:

$$\langle\Phi|\Psi\rangle = \Phi_0^*\Psi_0 + \int d^3x_1 \Phi_1^*(\vec{x}_1)\Psi_1(\vec{x}_1) + \int d^3x_1 d^3x_2 \Phi_2^*(\vec{x}_1, \vec{x}_2)\Psi_2(\vec{x}_1, \vec{x}_2) + \dots . \tag{4.7}$$

To get a feel for this theory, let's begin by calculating the probability density for finding exactly one particle at $\vec{x}$ and no particles anywhere else. This is simply given by the square of the matrix element of the state Ψ with a one particle state. Since any multi-particle state will be orthogonal to a one particle state, the only contribution comes from Ψ_1:

$$|\langle\vec{x}|\Psi\rangle|^2 = |\Psi_1(\vec{x})|^2 . \tag{4.8}$$

This expression follows from eq. 4.7 with the wave function for $|\vec{x}\rangle$ given by $\Psi_1(\vec{x}_1) = \delta(\vec{x}_1 - \vec{x})$ and $\Psi_0 = \Psi_2(\vec{x}_1, \vec{x}_2) = \dots = 0$. In general, the probability density for n particles in a certain configuration is simply $|\Psi_n(\vec{x}_1, \dots, \vec{x}_n, t)|^2$. This leads to the following normalization condition for a state $|\Psi\rangle$:

$$\langle\Psi|\Psi\rangle = |\Psi_0|^2 + \int d^3x_1 |\Psi_1(\vec{x}_1)|^2 + \int d^3x_1 d^3x_2 \, |\Psi_2(\vec{x}_1, \vec{x}_2)|^2 + \dots = 1 . \tag{4.9}$$

[4]Proof that a Hermitian Hamiltonian conserves probability:
$\frac{\partial}{\partial t}(\langle\Psi|\Psi\rangle) = \frac{\partial}{\partial t}(\langle\Psi|)|\Psi\rangle + \langle\Psi|\frac{\partial}{\partial t}(|\Psi\rangle)$
$= \langle\Psi|i\widehat{H}^\dagger|\Psi\rangle + \langle\Psi| - i\widehat{H}|\Psi\rangle$ by eq. 4.1
$= 0$ if $\widehat{H} = \widehat{H}^\dagger$

Next, let us consider the number density of particles at $\vec{x}$ (integrating this over a region gives the expected number of particles in that region):

$$\begin{aligned}\langle \text{number of particles at } \vec{x}\rangle &= |\langle \vec{x}|\Psi\rangle|^2 + 2\int d^3x_2 |\langle \vec{x}, \vec{x}_2|\Psi\rangle|^2 + ... \\ &= |\Psi_1(\vec{x})|^2 + 2\int d^3x_2 \, |\Psi_2(\vec{x}, \vec{x}_2)|^2 + \dots . \end{aligned} \tag{4.10}$$

The contribution from the possibility of a single particle is intuitive: we simply add the probability density for that particle being at $\vec{x}$. For 2-particles, we need to sum the probability that particle 1 is at $\vec{x}$ (regardless of where particle 2 is) with the probability that particle 2 is at $\vec{x}$.

$$\int d^3x_2 \, |\Psi_2(\vec{x}, \vec{x}_2)|^2 + \int d^3x_1 \, |\Psi_2(\vec{x}_1, \vec{x})|^2 \ . \tag{4.11}$$

Since the wave function over two dimensional configuration space is symmetric (as we are considering identical particles), these two terms can be combined to give the second contribution in eq. 4.10.

4.4 Creation and Annihilation Operators

It turns out that it will be convenient to be able to talk about states (wave functions) as operators acting on the vacuum. In the new theory introduced in the previous section, a state is an assignment of an amplitude to each point in each n-particle configuration space, an amplitude to any way any number of particles might be arranged. An operator is simply a mapping from states in our theory to other states in the theory (possibly un-normalized states, possibly zero which isn't really a state). Let us *define* $a_0^\dagger(\vec{x})$ as the operator which adds a particle definitely located at $\vec{x}$ to

any state.[5,6] Acting on the vacuum, $a_0^\dagger$ gives: $a_0^\dagger(\vec{x})|0\rangle = |\vec{x}\rangle$. Acting on a state of n particles, $a_0^\dagger$ brings it to a new—un-normalized—state with $n+1$ particles: $a_0^\dagger(\vec{x})|\Psi_n\rangle = \sqrt{n+1}|\Psi^{new}_{n+1@\vec{x}}\rangle$ (see fig. 4.2).[7,8] Here $|\Psi^{new}_{n+1@\vec{x}}\rangle$ is simply the original state $|\Psi_n\rangle$ with an additional particle localized at $\vec{x}$ (and is itself normalized). $a_0^\dagger$ is called a "creation" operator since it adds a particle to the state. From this definition it follows that $a_0^\dagger(\vec{x})$'s conjugate is an annihilation operator which decreases the total particle number by 1 by removing a particle localized at $\vec{x}$ if there is one to be removed (and takes $|0\rangle$ to 0).

$$\begin{aligned} \langle\Psi^{new}_{n+1@\vec{x}}|a_0^\dagger(\vec{x})|\Psi_n\rangle &= \sqrt{n+1} \\ \therefore \langle\Psi_n|a_0(\vec{x})|\Psi^{new}_{n+1@\vec{x}}\rangle &= \sqrt{n+1}\,. \end{aligned} \tag{4.12}$$

Here we start with the known (stipulated) action of $a_0^\dagger(\vec{x})$ on an arbitrary state and show that $a_0(\vec{x})$ takes a state with a particle localized at $\vec{x}$ to a state with one less particle located at $\vec{x}$ (and pulls out a factor of $\sqrt{n}$). We can now *derive* the

[5]Here I follow Teller (1997, pg. 41) in defining the creation operator by its action on states, not its canonical commutation relations.

[6]The subscript 0 is used to indicate that this creation operator has a non-relativistic normalization; $a^\dagger$ proper will be introduced in eq. 4.43.

[7]To be precise, the wave function of $\Psi^{new}_{n+1@\vec{x}}$ can be given in terms of Ψ_n by: $\Psi^{new}_{n+1@\vec{x}}(\vec{x}_1,...,\vec{x}_{n+1}) = \frac{1}{\sqrt{n+1}}(\delta(\vec{x}-\vec{x}_1)\Psi_n(\vec{x}_2,...,\vec{x}_{n+1}) + \delta(\vec{x}-\vec{x}_2)\Psi_n(\vec{x}_1,\vec{x}_3,...,\vec{x}_{n+1}) + ... + \delta(\vec{x}-\vec{x}_{n+1})\Psi_n(\vec{x}_2,...,\vec{x}_n))$ (as in Weinberg, 1995, eq. 4.2.3)

[8]One could, of course, introduce the creation operation without the $\sqrt{n+1}$ factor which makes the new state un-normalized, but I choose to include the factor since it allows us to use $a_0^\dagger(\vec{x})a_0(\vec{x})$ as the number operator (see eq. 4.17).

commutation relations for creation and annihilation operators[9]:

$$\left[a_0^\dagger(\vec{x}), a_0^\dagger(\vec{x}')\right] = [a_0(\vec{x}), a_0(\vec{x}')] = 0$$
$$\left[a_0(\vec{x}), a_0^\dagger(\vec{x}')\right] = \delta^3(\vec{x} - \vec{x}')\,. \tag{4.14}$$

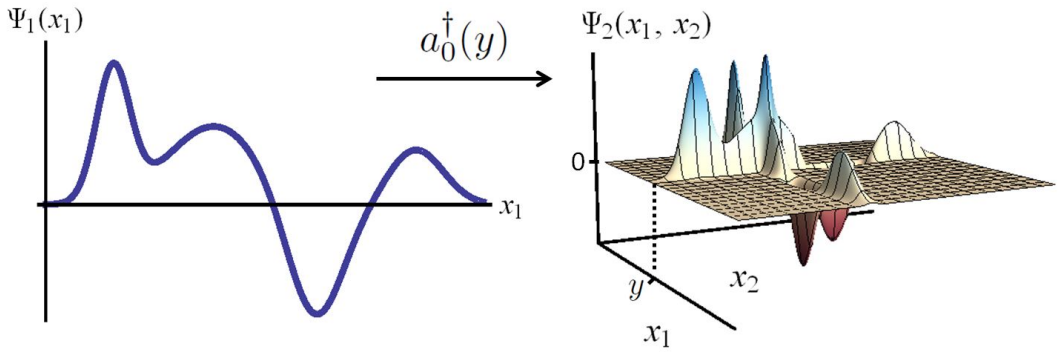

Figure 4.2: **Particle Creation** The creation operator $a_0^\dagger(y)$ takes a one-particle state (in one spatial dimension) to a two particle state where the second particle is localized at position y. Here we examine a special case where the wave function is real (so it can be plotted easily). The delta function is represented as a very steep Gaussian.

It should be clear that two of these creation operators will create two particles (the order doesn't matter because boson states are symmetric under particle permutations; alternatively, because $a_0^\dagger$'s commute): $a_0^\dagger(\vec{x}_1)a_0^\dagger(\vec{x}_2)|0\rangle = a_0^\dagger(\vec{x}_2)a_0^\dagger(\vec{x}_1)|0\rangle =$

[9] *Derivation*: Let $|\Psi\rangle$ be an arbitrary state that can be decomposed into particle number eigenstates as follows $|\Psi\rangle = \sum_n \alpha_n |\Psi_n\rangle$

$$\begin{aligned}\langle\Psi|\left[a_0(\vec{x}), a_0^\dagger(\vec{x}')\right]|\Psi\rangle &= \langle\Psi|a_0(\vec{x})a_0^\dagger(\vec{x}') - a_0^\dagger(\vec{x}')a_0(\vec{x})|\Psi\rangle \\ &= \sum_n |\alpha_n|^2 \left(\langle\Psi^{new}_{n+1@\vec{x}}|(n+1)|\Psi^{new}_{n+1@\vec{x}'}\rangle - \langle\Psi^{new}_{n-1@\vec{x}'}|n|\Psi^{new}_{n-1@\vec{x}}\rangle\right) \\ &= \sum_n |\alpha_n|^2 \left(\delta^3(\vec{x} - \vec{x}') \times (n+1-n)\right) = \delta^3(\vec{x} - \vec{x}')\,. \end{aligned} \tag{4.13}$$

See (Teller, 1997, pg. 42-45) for a discussion of why the creation/annihilation operators for *fermions* differ in such a way that $\delta^3(\vec{x} - \vec{x}')$ is the *anti*-commutator of a and $a^\dagger$.

$\sqrt{2}|\vec{x}_1, \vec{x}_2\rangle$. A general state will be a superposition of such n-particle, localized states:

$$\begin{aligned}|\Psi\rangle &= \left(\Psi_0 + \int d^3x_1 \Psi_1(\vec{x}_1) a_0^\dagger(\vec{x}_1) + \frac{1}{\sqrt{2}} \int d^3x_1 d^3x_2 \Psi_2(\vec{x}_1, \vec{x}_2) a_0^\dagger(\vec{x}_1) a_0^\dagger(\vec{x}_2) + ... \right. \\ &\left. + \frac{1}{\sqrt{n!}} \int d^3x_1 ... d^3x_n \Psi_n(\vec{x}_1, ..., \vec{x}_n) a_0^\dagger(\vec{x}_1) ... a_0^\dagger(\vec{x}_n) \right) |0\rangle \\ &= \widehat{\Psi}|0\rangle \,. \end{aligned} \tag{4.15}$$

In the second equality we define $\widehat{\Psi}$ as shorthand for the operator given in the first equality which acts on the vacuum creating the state $|\Psi\rangle$. In this way we can associate a unique operator with each state in our theory.

To get some practice with these operators, let's calculate the probability for finding one particle at $\vec{x}$ and no particles anywhere else. As in eq. 4.8, we just take the square of the inner product of the state with a position eigenstate:

$$\begin{aligned}|\langle \vec{x}|\Psi\rangle|^2 &= \left|\langle 0|a_0(\vec{x}) \left(\Psi_0 + \textstyle\int d^3x_1 \Psi_1(\vec{x}_1) a_0^\dagger(\vec{x}_1) + \frac{1}{\sqrt{2}} \int d^3x_1 d^3x_2 \Psi_2(\vec{x}_1, \vec{x}_2) a_0^\dagger(\vec{x}_1) a_0^\dagger(\vec{x}_2) + ...\right) |0\rangle\right|^2 \\ &= |\Psi_1(\vec{x})|^2 \,. \end{aligned} \tag{4.16}$$

The vacuum amplitude commutes with the annihilation operator and thus allows it to annihilate the vacuum on the right. The components of the wave function corresponding to multi-particle states annihilate the vacuum on the left when commuted through the single annihilation operator. Only the single-particle component is relevant. Here we retrieve the same result as eq. 4.8. To calculate the number density, we simply take the expectation value of the number operator

$a_0^\dagger(\vec{x})a_0(\vec{x})$:

$$
\begin{aligned}
&\langle\Psi|a_0^\dagger(\vec{x})a_0(\vec{x})|\Psi\rangle \\
&= \langle\Psi|a_0^\dagger(\vec{x})a_0(\vec{x})\left(\Psi_0 + \int d^3x_1\Psi_1(\vec{x}_1)a_0^\dagger(\vec{x}_1) + \tfrac{1}{\sqrt{2}}\int d^3x_1 d^3x_2\Psi_2(\vec{x}_1,\vec{x}_2)a_0^\dagger(\vec{x}_1)a_0^\dagger(\vec{x}_2) + ...\right)|0\rangle \\
&= \langle\Psi|\left(\int d^3x_1\Psi_1(\vec{x}_1)a_0^\dagger(\vec{x})\delta^3\left(\vec{x}-\vec{x}_1\right)\right. \\
&\left. + \tfrac{1}{\sqrt{2}}\int d^3x_1 d^3x_2\Psi_2(\vec{x}_1,\vec{x}_2)a_0^\dagger(\vec{x})\left[a_0^\dagger(\vec{x}_2)\delta^3\left(\vec{x}-\vec{x}_1\right) + a_0^\dagger(\vec{x}_1)\delta^3\left(\vec{x}-\vec{x}_2\right)\right] + ...\right)|0\rangle \\
&= |\Psi_1(\vec{x})|^2 + 2\int d^3x_2\left|\Psi_2(\vec{x},\vec{x_2})\right|^2 + ... \,. \qquad (4.17)
\end{aligned}
$$

Here we have shown that the operator $a_0^\dagger(\vec{x})a_0(\vec{x})$ is the number operator by showing that its expectation value gives the expression previously derived in eq. 4.10.

It is straightforward to do everything we've done so far in the *momentum basis*. $a_0^\dagger(\vec{k})$ is defined as the operator which adds a particle with definite momentum $\vec{k}$ (just as $a_0^\dagger(\vec{x})$ was defined as the operator which adds a particle in a position eigenstate). It will be convenient to choose the $a_0^\dagger(\vec{k})$ operators to have the following normalization (as we did with $a_0^\dagger(\vec{x})$): $a_0^\dagger(\vec{k})|\Psi_n\rangle = \sqrt{n+1}|\Psi^{new}_{n+1@\vec{k}}\rangle$. This leads to similar commutation relations for $a_0(\vec{k})$: $\left[a_0(\vec{k}), a_0^\dagger(\vec{k}\,')\right] = \delta^3(\vec{k}-\vec{k}\,')$. Equations 4.9, 4.15, and 4.17 all hold in momentum space when we replace $\vec{x}$ by $\vec{k}$ and the complex functions Ψ_n with their Fourier transforms, defined (symmetrically) by:

$$
\begin{aligned}
\widetilde{\Psi}_n(\vec{k}_1,...,\vec{k}_n) &\equiv \int \frac{d^3x_1...d^3x_n}{(2\pi)^{3n/2}} e^{-i\vec{k}_1\cdot\vec{x}_1}...e^{-i\vec{k}_n\cdot\vec{x}_n}\Psi_n(\vec{x}_1,..,\vec{x}_n) \\
\Psi_n(\vec{x}_1,..,\vec{x}_n) &= \int \frac{d^3k_1...d^3k_n}{(2\pi)^{3n/2}} e^{i\vec{k}_1\cdot\vec{x}_1}...e^{i\vec{k}_n\cdot\vec{x}_n}\widetilde{\Psi}_n(\vec{k}_1,...,\vec{k}_n)\,. \qquad (4.18)
\end{aligned}
$$

The rules for Fourier-transforming the wave functions in eq. 4.18 induce rules for transforming between $a_0^\dagger(\vec{x})$ and $a_0^\dagger(\vec{k})$. Since it should not matter whether we express a single particle wave function in the position or the momentum basis, we

must have,

$$|\Psi\rangle = \int d^3x \Psi(\vec{x}) a_0^\dagger(\vec{x})|0\rangle = \int d^3k \widetilde{\Psi}(\vec{k}) a_0^\dagger(\vec{k})|0\rangle \,. \tag{4.19}$$

It follows from eq. 4.18 and 4.19 that,

$$a_0^\dagger(\vec{x}) = \int \frac{d^3k}{(2\pi)^{3/2}} e^{-i\vec{k}\cdot\vec{x}} a_0^\dagger(\vec{k}) \qquad a_0^\dagger(\vec{k}) = \int \frac{d^3x}{(2\pi)^{3/2}} e^{i\vec{k}\cdot\vec{x}} a_0^\dagger(\vec{x}) \,. \tag{4.20}$$

Using the number operator in momentum space, we can rewrite the free Schrödinger equation, eq. 4.6 with $V = 0$. Our strategy will be to count the number of particles expected to be in each momentum eigenstate and weight this counting by the energy associated with momentum $\vec{k}$.

$$i\frac{\partial}{\partial t}|\Psi(t)\rangle = \int d^3k \frac{|\vec{k}|^2}{2m} a_0^\dagger(\vec{k}) a_0(\vec{k})|\Psi(t)\rangle \,. \tag{4.21}$$

It is important to remember that at this point we have introduced some new notation but fundamentally have done very little to alter the original theory (QM). We have simply extended the theory to *allow* states that are in superpositions of different particle numbers and permitted terms in the Hamiltonian that cause the total number of particles to change. In this section, we have introduced a slight generalization to QM that will make the shift to QFT easier. In the next section we will move to relativistic quantum field theory. The only difference between (real scalar) QFT and the variable-particle QM we have been discussing in this section (and the last) is the move to a relativistic Hamiltonian.

We can summarize the theory we have come up with as we did for n-particle QM. The main difference between theory 1 and theory 2 is a wider range of possible states. The version of the Schrödinger equation is just the natural extension of eq. 4.1 to this larger space (rewritten in terms of creation and annihilation operators). **Probabilities** is essentially unchanged.

Theory 2: Variable-particle QM

States The states of the theory are wave functions on the disjoint union of n-particle configuration spaces: $|\Psi\rangle$ is specified by $\langle \Psi_0, \Psi_1(\vec{x}_1), \Psi_2(\vec{x}_1, \vec{x}_2), ... \rangle$

Schrödinger Equation Basically eq. 4.6, but, a bit more generally, eq. 4.21 with the possibility of interactions:

$$i\frac{\partial}{\partial t}|\Psi(t)\rangle = \left(\int d^3k \frac{|\vec{k}|^2}{2m} a_0^\dagger(\vec{k}) a_0(\vec{k}) + \widehat{H}_{int} \right) |\Psi(t)\rangle \,. \tag{4.22}$$

Probabilities The probability (or probability density) of observing the system in state $|\Phi\rangle$ in an appropriate experiment is given by the square of the inner product of $|\Phi\rangle$ and the system's state $|\Psi\rangle$: $|\langle\Phi|\Psi\rangle|^2$ (where the inner product is defined in equation 4.7)

Before we move on, let's briefly consider the sort of interaction term you might want to add to eq. 4.22. An example:

$$\widehat{H}_{int} = \lambda \int d^3x \left(a_0^\dagger(\vec{x}) a_0^\dagger(\vec{x}) a_0(\vec{x}) + a_0^\dagger(\vec{x}) a_0(\vec{x}) a_0(\vec{x}) \right) . \tag{4.23}$$

Note that $\widehat{H}_{int}$, like the free Hamiltonian, is hermitian, so states will stay normalized.

This term would allow for interactions in which two particles are destroyed and one is created and interactions where one particle is destroyed and two are created (all at the same point $\vec{x}$). Fourier transforming the creation/annihilation operators in eq. 4.23 (using eq. 4.20) makes it clear that momentum is conserved in these interactions:

$$\widehat{H}_{int} = \frac{\lambda}{(2\pi)^{3/2}} \int d^3k_1 d^3k_2 \left(a_0^\dagger(\vec{k}_1) a_0^\dagger(\vec{k}_2) a_0(\vec{k}_1 + \vec{k}_2) + a_0^\dagger(\vec{k}_1 + \vec{k}_2) a_0(\vec{k}_1) a_0(\vec{k}_2) \right). \tag{4.24}$$

To see how this interacting Hamiltonian can be used to time evolve state, we can calculate how a momentum eigenstate $|\vec{k}, t=0\rangle$ evolves (using eq. 4.22):

$$\begin{aligned} \frac{\partial}{\partial t}|\vec{k}\rangle &= -i\left(\widehat{H}_{free} + \widehat{H}_{int}\right)|\vec{k}\rangle \\ &= -i\frac{|\vec{k}|^2}{2m}|\vec{k}\rangle - i\frac{\sqrt{2}\lambda}{(2\pi)^{3/2}} \int d^3k_1 |\vec{k}_1, \vec{k} - \vec{k}_1\rangle . \end{aligned} \tag{4.25}$$

Although momentum is conserved explicitly by eq. 4.24, there is no guarantee that *kinetic energy* will be the same in the state observed after a period of Schrödinger equation time evolution. However, the probability of a large difference in kinetic energy is small. Treating $\widehat{H}_{int}$ as a small perturbation in the Hamiltonian, we can calculate the time evolution of a momentum eigenstate: $|\vec{k}, t=0\rangle$. Let's look at how the two-particle part of wave function, which is initially uniformly zero, evolves with time (to first order):

$$\widetilde{\Psi}_2(\vec{k}_1, \vec{k}_2, t) = -\frac{\lambda\sqrt{2}}{(2\pi)^{3/2}} \left(\frac{\text{Exp}\left[-i\frac{|\vec{k}|^2}{2m}t\right] - \text{Exp}\left[-i\frac{|\vec{k}_1|^2+|\vec{k}_2|^2}{2m}t\right]}{\frac{|\vec{k}_1|^2+|\vec{k}_2|^2}{2m} - \frac{|\vec{k}|^2}{2m}} \right) \delta^3(\vec{k}_1 + \vec{k}_2 - \vec{k}) . \tag{4.26}$$

Here we can see that the denominator of the middle factor suppresses large differences in energy. Similarly, the interactions that arise in *relativistic QFT* do not enforce relativistic energy, $\sqrt{|\vec{k}|^2 + m^2}$, conservation, but the amplitude for large differences

in energy between the initial and final states is small (an energy conserving delta function appears when the initial and final states are taken to $\pm\infty$, as they are in standard derivations of the Feynman rules).

An important feature of the theory described by eq. 4.24: the vacuum *is* the ground state of the theory (this is not true in general), but the one particle states $|\vec{k}\rangle$ and $|\vec{x}\rangle$ are *not* eigenstates of the Hamiltonian. Here I make an important distinction between the vacuum state and the ground state. The *vacuum* is the zero particle state with wave function: $\Psi_0 = 1$, $\Psi_1 = \Psi_2 = ... = 0$. The *ground state* is the lowest energy eigenstate of the Hamiltonian.

4.5 A Relativistic Hamiltonian

Now we can handle theories in which the number of particles in a system varies as a function of time and we have a way of writing states in terms of operators acting on the vacuum state. Let us attempt to create a *relativistic* theory. One of the simplest things we can try is modifying the free non-relativistic Schrödinger equation (eq. 4.21) to have a Hamiltonian operator which corresponds to the relativistic energy. Since, $E_{rel} = \sqrt{|\vec{k}|^2 + m^2}$, you might try $\widehat{H} = \int d^3x \sqrt{-\nabla^2 + m^2}$. Unfortunately, the square root of a differential operator is a bit hard to make sense of. Instead, we can work directly from the idea that the energy of a *momentum* eigenstate is $\sqrt{|\vec{k}|^2 + m^2}$. By decomposing the state into momentum eigenstates and summing their relativistic energies (as in eq. 4.21 & 4.22) we get an expression for the (free) Hamiltonian which we can plug into the Schrödinger equation to get a law of temporal evolution for the state $|\Psi\rangle$:

$$i\frac{\partial}{\partial t}|\Psi(t)\rangle = \int d^3k \sqrt{|\vec{k}|^2 + m^2}\; a_0^\dagger(\vec{k}) a_0(\vec{k}) |\Psi(t)\rangle \,. \tag{4.27}$$

As you can see, we needed creation and annihilation operators to be able to express this Hamiltonian and that is why we needed to develop theory 2 before we could move to the relativistic extension of QM.

Taking the theory developed in §4.3 & 4.4 and using a relativistic Hamiltonian like eq. 4.27, we're done. That is all you need to have a relativistic quantum field theory. To recap: the state of the system is an assignment of complex numbers to points in each n-particle configuration space. The law of temporal evolution is still Schrödinger's equation, just with a relativistic Hamiltonian (eq. 4.27 + interactions). States are normalized according to eq. 4.9 and number densities can be calculated by eq. 4.10, or equivalently, eq. 4.17. The creation operator $a^\dagger(\vec{x})$ is defined by its action on states (as in first paragraph of §4.4) and the momentum creation operator is defined similarly and related by a Fourier transformation, eq. 4.20.[10]

Using our free relativistic Schrödinger equation (eq. 4.27), we can calculate the way some simple states evolve in time. We will begin by deriving the time dependence of single particle momentum eigenstates. Here $|\Psi\rangle = |\vec{k}, t\rangle$ where $|\vec{k}, 0\rangle = a_0^\dagger(\vec{k})|0\rangle$. $|\vec{k}, 0\rangle$ is an eigenstate of the Hamiltonian with energy $\sqrt{|\vec{k}|^2 + m^2}$, so the time evolution is just a constant phase rotation:

$$|\vec{k}, t\rangle = e^{-i\sqrt{|\vec{k}|^2+m^2}t} a_0^\dagger(\vec{k})|0\rangle\,. \tag{4.28}$$

The time evolution of a position eigenstate can be calculated by decomposing the state into its Fourier components (using eq. 4.20):

$$|\vec{x}, 0\rangle = \frac{1}{(2\pi)^{3/2}} \int d^3k e^{-i\vec{k}\cdot\vec{x}} a_0^\dagger(\vec{k})|0\rangle\,. \tag{4.29}$$

[10]The most similar account of QFT to this that I've seen is the formulation of QFT used in Bell-type approaches to Bohmian QFT Dürr *et al.* , 2004, 2005.

Then, using the known time evolution of momentum eigenstates from eq. 4.28:

$$|\vec{x},t\rangle = \frac{1}{(2\pi)^{3/2}}\int d^3k e^{-i\vec{k}\cdot\vec{x}}e^{-i\sqrt{|\vec{k}|^2+m^2}t}a_0^\dagger(\vec{k})|0\rangle\,. \tag{4.30}$$

Be careful. This equation represents how a state localized at $\vec{x}$ at $t=0$ evolves, it does not represent the evolution of a particle localized at $\vec{x}$ at time t. In general, a particle localized at $\vec{x}$ at time x_0 will evolve in time t according to (it is apparent that the particle is localized at $\vec{x}$ at x_0 since this equation become $|\vec{x}\rangle$ for $t=x_0$):

$$|x,t\rangle = \frac{1}{(2\pi)^{3/2}}\int d^3k e^{-i\vec{k}\cdot\vec{x}}e^{+i\sqrt{|\vec{k}|^2+m^2}\,x_0}e^{-i\sqrt{|\vec{k}|^2+m^2}\,t}a_0^\dagger(\vec{k})|0\rangle\,. \tag{4.31}$$

Here we've dropped the vector hat on $\vec{x}$ to denote a four-vector. Defining $\omega(\vec{k})$ as the energy of a particle with momentum $\vec{k}$, $k_0=\omega(\vec{k})\equiv\sqrt{|\vec{k}|^2+m^2}$, we can rewrite eq. 4.31 in terms of the four dimensional dot product (using the Minkowski metric with $(-,+,+,+)$ signature):

$$|x,t\rangle = \frac{1}{(2\pi)^{3/2}}\int d^3k e^{-ik\cdot x}e^{-i\sqrt{|\vec{k}|^2+m^2}\,t}a_0^\dagger(\vec{k})|0\rangle \tag{4.32}$$

We can calculate the time dependence of an arbitrary state. This will be just like eq. 4.15 but now all of the position eigenstates evolve with t.

$$\begin{aligned}|\Psi(t)\rangle = \Big(&\Psi_0(0)+\textstyle\int d^3x_1 d^3k_1\Psi_1(\vec{x}_1,0)e^{-i\vec{k_1}\cdot\vec{x_1}-ik_{10}x_{10}}a_0^\dagger(\vec{k}_1)+\ldots\\ &+\tfrac{1}{\sqrt{n!}}\textstyle\int d^3x_1\ldots d^3x_n d^3k_1\ldots d^3k_n\Psi_n(\vec{x}_1,\ldots,\vec{x}_n,0)e^{-i\vec{k}_1\cdot\vec{x}_1-ik_{10}x_{10}}a_0^\dagger(\vec{k}_1)\ldots e^{-i\vec{k}_n\cdot\vec{x}_n-ik_{n0}x_{n0}}a_0^\dagger(\vec{k}_n)\Big)|0\rangle\,.\end{aligned} \tag{4.33}$$

The probability density for a certain configuration of n particles is (as in theories 1 & 2):

$$|\langle\vec{x}_1,\ldots,\vec{x}_n,t|\Psi\rangle|^2 = |\Psi_n(\vec{x}_1,\ldots,\vec{x}_n,t)|^2\,. \tag{4.34}$$

We have deftly avoided problems with probability that plagued early attempts at developing relativistic versions of QM. States stay normalized as they evolve because the Hamiltonian is Hermitian, but, as an exercise, one can show that it follows from eq. 4.33.[11]

Again, we summarize the theory we have arrived at. Note that the *only* difference between theory 2 and theory 3 is the expression for the energy of free particles that appears in the Hamiltonian of the free theory.

---

[11] *Proof* (using momentum basis to expand $|\Psi(t)\rangle$):

$$\begin{aligned}
&\langle\Psi(t)|\Psi(t)\rangle = \langle\Psi(t)|0\rangle\langle 0|\Psi(t)\rangle + \int d^3k_1 \langle\Psi(t)|\vec{k}_1, t\rangle\langle\vec{k}_1, t|\Psi(t)\rangle + \int d^3k_1 d^3k_2 \langle\Psi(t)|\vec{k}_1, \vec{k}_2, t\rangle\langle\vec{k}_1, \vec{k}_2, t|\Psi(t)\rangle + ... \\
&= |\Psi_0|^2 + \int d^3k_1 \left|\langle 0|e^{i\omega(\vec{k}_1)t} a_0(\vec{k}_1) \int d^3k \widetilde{\Psi}_1(\vec{k}) e^{-i\omega(\vec{k})t} a_0^\dagger(\vec{k})|0\rangle\right|^2 \\
&+ \int d^3k_1 d^3k_2 \left|\langle 0|\tfrac{1}{\sqrt{2}} e^{i(\omega(\vec{k}_1)+\omega(\vec{k}_2))t} a_0(\vec{k}_1) a_0(\vec{k}_2) \tfrac{1}{\sqrt{2}} \int d^3k d^3k' \widetilde{\Psi}_2(\vec{k}, \vec{k}') e^{-i(\omega(\vec{k})+\omega(\vec{k}'))t} a_0^\dagger(\vec{k}) a_0^\dagger(\vec{k}')|0\rangle\right|^2 + ... \\
&= |\Psi_0|^2 + \int d^3k_1 \left|\langle 0| \int d^3k \widetilde{\Psi}_1(\vec{k}) e^{i\omega(\vec{k}_1)t} e^{-i\omega(\vec{k})t} \delta^3(\vec{k} - \vec{k}_1)|0\rangle\right|^2 \\
&+ \int d^3k_1 d^3k_2 \left|\langle 0|\tfrac{1}{2} \int d^3k d^3k' \widetilde{\Psi}_2(\vec{k}, \vec{k}') e^{i(\omega(\vec{k}_1)+\omega(\vec{k}_2))t} e^{-i(\omega(\vec{k})+\omega(\vec{k}'))t} \left(\delta^3(\vec{k} - \vec{k}_1)\delta^3(\vec{k}' - \vec{k}_2) + \delta^3(\vec{k} - \vec{k}_2)\delta^3(\vec{k}' - \vec{k}_1)\right)|0\rangle\right|^2 + ... \\
&= |\Psi_0|^2 + \int d^3k_1 |\widetilde{\Psi}_1(\vec{k}_1)|^2 + \int d^3k_1 d^3k_2 \left|\widetilde{\Psi}_2(\vec{k}_1, \vec{k}_2)\right|^2 + ... = 1\,. \qquad (4.35)
\end{aligned}$$

The first line of eq. 4.35 gives the identity element in Fock space. For single particle QM we had:

$$\int d^3k |\vec{k}, t\rangle\langle\vec{k}, t| = \hat{1}\,. \qquad (4.36)$$

In Fock space we must allow any number of particles to have a full basis of states:

$$|0\rangle\langle 0| + \int d^3k_1 |\vec{k}_1, t\rangle\langle\vec{k}_1, t| + \int d^3k_1 d^3k_2 |\vec{k}_1, \vec{k}_2, t\rangle\langle\vec{k}_1, \vec{k}_2, t| + ... = \hat{1}\,. \qquad (4.37)$$

To get the second line of eq. 4.35, we use eq. 4.19 to expand $|\Psi\rangle$ and rewrite the momentum eigenstates using eq. 4.28. Then we use eq. 4.14 to eliminate the creation and annihilation operators. Finally, we integrate out the delta functions to get a form which matches our requirement for the initial wave function's normalization (eq. 4.9, in momentum space).

Theory 3: Real Scalar QFT (in the *particle* basis)

States The states of the theory are wave functions on the disjoint union of n-particle configuration spaces: $|\Psi\rangle$ is specified by $\langle \Psi_0, \Psi_1(\vec{x}_1), \Psi_2(\vec{x}_1, \vec{x}_2), ... \rangle$

Schrödinger Equation As presented below or, equivalently, writing the free Hamiltonian as in eq. 4.46.

$$i\frac{\partial}{\partial t}|\Psi(t)\rangle = \left(\int d^3k \sqrt{|\vec{k}|^2 + m^2}\; a_0^\dagger(\vec{k}) a_0(\vec{k}) + \widehat{H}_{int} \right) |\Psi(t)\rangle \,. \tag{4.38}$$

Probabilities The probability (or probability density) of observing the system in state $|\Phi\rangle$ in an appropriate experiment is given by the square of the inner product of $|\Phi\rangle$ and the system's state $|\Psi\rangle$: $|\langle\Phi|\Psi\rangle|^2$ (where the inner product is defined in equation 4.7)

It is easy to see how this theory limits to QM in the free case or when there is a fixed external potential. If the momenta involved are sufficiently small relative to m (the particles are traveling slowly relative to the speed of light), then $\sqrt{|\vec{k}|^2 + m^2}$ is, to first order in $|\vec{k}|^2$, $m + \frac{|\vec{k}|^2}{2m}$. So, the free part of the Hamiltonian that appears in eq. 4.38 is the same as theory 2's (eq. 4.22) up to an invisible constant[12] shift in the energy: $\int d^3k\, m\, a_0^\dagger(\vec{k}) a_0(\vec{k})$. It is more complicated to show that interactions like those in the Standard Model can limit to the sort of simple inter-particle interactions discussed in QM (How do photon exchanges in QFT manifest themselves as $1/r^2$

[12] Actually, the shift is dependent on the total number of particles (which can change when interactions are present). But, it is constant if we confine the wave function to a particular n-particle configuration space (and don't allow interactions which take it out of there).

attractive forces in a Hydrogen atom?). Examination of equation 4.38 suggests that theory 3 should limit to special relativity in the same way that QM limits to classical particle mechanics.[13]

4.6 The Lagrangian and the Quantum Field

We have now seen the basic laws of relativistic quantum field theory and can introduce some more sophisticated notation. I would like to define and discuss an operator that is of particular interest in the theory: the field operator $\widehat{\phi}$ and its conjugate momenta $\widehat{\Pi}$. In this section, we will define these operators in terms of our creation and annihilation operators, show that the Hamiltonian for free bosons can be rewritten in terms of $\widehat{\phi}$ and $\widehat{\Pi}$, and discuss some properties of these new operators. By reformulating our version of QFT in terms of these new operators, we connect the treatment thus far to more standard textbook treatments and prepare for part II by presenting the theory in a way that makes it appear similar to the version discussed there (in part III we will see that the similarity goes beyond mere appearances, it's the same theory). So, this new notation will look like an unnecessary complication of a clean and simple theory. But don't worry, the motivation will become clear in part III when we compare theory 3 and theory 4 (introduced in part II).

[13] Here's how one might try to prove it: One should be able to develop a path integral approach to QFT starting with eq. 4.38 and make an argument similar to that in §4.11 that when variations in the action are large relative to $\hbar$, we can use a steepest decent approximation to derive the classical equations of motion for special relativity.

Define the operators $\widehat{\phi}(\vec{k})$ and $\widehat{\Pi}(\vec{k})$ as:

$$\widehat{\phi}(\vec{k}) \equiv \frac{1}{\sqrt{2\omega(\vec{k})}} \left(a_0(\vec{k}) + a_0^\dagger(-\vec{k}) \right)$$
$$\widehat{\Pi}(\vec{k}) \equiv -i\sqrt{\frac{\omega(\vec{k})}{2}} \left(a_0(\vec{k}) - a_0^\dagger(-\vec{k}) \right) . \tag{4.39}$$

In terms of these operators, we can rewrite the Hamiltonian in eq. 4.27 as:

$$\begin{aligned} \widehat{H} &= \int d^3k \sqrt{\omega(\vec{k})}\; a_0^\dagger(\vec{k}) a_0(\vec{k}) \\ &= \int d^3k \left\{ \frac{1}{2}\widehat{\Pi}(\vec{k})\widehat{\Pi}(-\vec{k}) + \frac{|\vec{k}|^2}{2}\widehat{\phi}(\vec{k})\widehat{\phi}(-\vec{k}) + \frac{m^2}{2}\widehat{\phi}(\vec{k})\widehat{\phi}(-\vec{k}) - \frac{\omega(\vec{k})}{2}\delta(0) \right\} . \end{aligned} \tag{4.40}$$

We can rewrite the Hamiltonian as an integral over space by Fourier transforming $\widehat{\phi}$ and $\widehat{\Pi}$. Omitting the spatial integral, we can write the Hamiltonian density as:

$$\widehat{\mathscr{H}} = \frac{1}{2}\widehat{\Pi}^2(\vec{x}) + \frac{1}{2}(\nabla\widehat{\phi}(\vec{x}))^2 + \frac{1}{2}m^2\widehat{\phi}^2(\vec{x}) - \Omega_0 . \tag{4.41}$$

Here $\Omega_0 \equiv \int d^3k\; \frac{\omega(\vec{k})}{2(2\pi)^3}$ (yes this is infinite, we'll discuss it soon). This looks like the Hamiltonian for a classical Klein-Gordon field (except that ϕ and Π are operators not classical fields). In part II we will see how "quantizing" the classical Klein-Gordon field leads to a (relativistic) quantum field theory where eq. 4.41 is the Hamiltonian that appears in the theory's Schrödinger equation (teaser from part II: $\langle\Psi|\widehat{\phi}(\vec{x},t)|\Psi\rangle$ gives the expectation value of a measurement of a classical field at $\vec{x},t$). Although the operator $\widehat{\phi}$ is introduced via a different definition in part II (eq. 4.66), we'll see in part III that they are indeed the same operator. In part III we will examine how the equivalence of the Hamiltonians in eq. 4.27 and eq. 4.41 (just proved) can be used to

show that theory 3 (developed here) is equivalent to the theory introduced in part II.

Fourier transforming the definitions of $\widehat{\phi}$ and $\widehat{\Pi}$ in eq. 4.39 gives the following expressions:

$$\begin{aligned}\widehat{\phi}(\vec{x}) &= \int \widetilde{d^3k} \left(e^{-i\vec{k}\cdot\vec{x}} a^\dagger(\vec{k}) + e^{i\vec{k}\cdot\vec{x}} a(\vec{k})\right) \\ \widehat{\Pi}(\vec{x}) &= \int \widetilde{d^3k}\; i\omega(\vec{k}) \left(e^{-i\vec{k}\cdot\vec{x}} a^\dagger(\vec{k}) - e^{i\vec{k}\cdot\vec{x}} a(\vec{k})\right) .\end{aligned} \tag{4.42}$$

Here we have replaced the familiar creation and annihilation operators with renormalized versions defined by:[14]

$$a^\dagger(\vec{k}) \equiv (2\pi)^{3/2}(2\omega(\vec{k}))^{1/2} a_0^\dagger(\vec{k}) . \tag{4.43}$$

The commutation relation (eq. 4.14) then becomes: $\left[a(\vec{k}), a^\dagger(\vec{k}')\right] = (2\pi)^3 2\omega(\vec{k})\delta^3(\vec{k} - \vec{k}')$. $a^\dagger(\vec{k})$ is a convenient operator to use for relativistic theories since it transforms simply under Lorentz transformations. In eq. 4.42 we have also introduced the notation $\widetilde{d^3k}$ for the following relativistically invariant integration measure which will appear often:

$$\widetilde{d^3k} \equiv \frac{d^3k}{(2\pi)^3 2\omega(\vec{k})} . \tag{4.44}$$

An expression for $\widehat{\phi}(\vec{x}, t)$, which denotes acting with $\widehat{\phi}(\vec{x})$ on the state at some time t later than now, can be found by time evolving a state forward, acting with the field operator, and then time evolving the state back (similarly for $\widehat{\Pi}$).

$$\widehat{\phi}(\vec{x}, t) = e^{i\widehat{H}t} \widehat{\phi}(\vec{x}) e^{-i\widehat{H}t} . \tag{4.45}$$

[14] These operators now match what you see throughout many QFT textbooks, for example (Srednicki, 2007, pg. 28).

With our new notation, the Hamiltonian in eq. 4.27 can be written as:

$$\widehat{H} = \int \widetilde{d^3k}\ \omega(\vec{k}) a^\dagger(\vec{k}) a(\vec{k})\,. \tag{4.46}$$

Using this Hamiltonian and eq. 4.45, we see that (in the free case), the time dependence of ϕ is given by:[15]

$$\widehat{\phi}(\vec{x},t) \equiv \int \widetilde{d^3k}\ \left(e^{-ik\cdot x} a^\dagger(\vec{k}) + e^{ik\cdot x} a(\vec{k})\right)\,. \tag{4.47}$$

Here we have moved from three-dimensional to four-dimensional dot products in the exponentials (with $x_0 = t$).[16] The corresponding expression for $\widehat{\Pi}$ can be calculated in the same manner:

$$\widehat{\Pi}(\vec{x},t) = \int \widetilde{d^3k}\ i\omega(\vec{k}) \left(e^{-ik\cdot x} a^\dagger(\vec{k}) - e^{ik\cdot x} a(\vec{k})\right)\,. \tag{4.49}$$

From eq. 4.47 & 4.49 and the commutation relations for a and $a^\dagger$, we can show:

$$[\widehat{\phi}(\vec{x},t), \widehat{\phi}(\vec{x}',t)] = [\widehat{\Pi}(\vec{x},t), \widehat{\Pi}(\vec{x}',t)] = 0$$

$$[\widehat{\phi}(\vec{x},t), \widehat{\Pi}(\vec{x}',t)] = i\delta^3(\vec{x} - \vec{x}')\,. \tag{4.50}$$

[15] In deriving this, the following operator identity is useful (the Hadamard lemma of the Campbell-Baker-Hausdorff formula): $e^{\hat{A}}\hat{B}e^{-\hat{A}} = \hat{B} + [\hat{A},\hat{B}] + \frac{1}{2!}[\hat{A},[\hat{A},\hat{B}]] + ...$

[16] We can show that $\widehat{\phi}(\vec{x},t)$ (eq. 4.47) must satisfy the Klein-Gordon equation ($\partial^2 f(\vec{x},t) = m^2 f(\vec{x},t)$).

$$\begin{aligned}\left(-\partial^2 + m^2\right)\widehat{\phi}(\vec{x},t) &= \left(\partial_0^2 - \partial_{\vec{x}}^2 + m^2\right)\int \widetilde{d^3k}\left(e^{-ik\cdot x} a^\dagger(\vec{k}) + e^{ik\cdot x} a(\vec{k})\right) \\ &= \int \widetilde{d^3k}\left(-\omega(\vec{k})^2 + \vec{k}^2 + m^2\right)\left(e^{-ik\cdot x} a^\dagger(\vec{k}) + e^{ik\cdot x} a(\vec{k})\right) = 0\,.\end{aligned} \tag{4.48}$$

We can derive the Hamiltonian in the free relativistic Schrödinger equation (eq. 4.27, or equivalently, as in eq. 4.41) from a specified Lagrangian density (operator) defined in terms of $\widehat{\phi}$.

$$\widehat{\mathscr{L}} = -\frac{1}{2}\partial^{\mu}\widehat{\phi}\partial_{\mu}\widehat{\phi} - \frac{1}{2}m^2\widehat{\phi}^2 + \Omega_0\,. \tag{4.51}$$

If we take this as the Lagrangian, we can derive the Hamiltonian by defining the conjugate momentum operator to $\widehat{\phi}(\vec{x},t)$ as $\widehat{\Pi}(\vec{x},t) = \frac{\partial\widehat{\mathscr{L}}}{\partial\left\{\partial_0\widehat{\phi}(\vec{x},t)\right\}} = \partial_0\widehat{\phi}(\vec{x},t)$[17] and defining the Hamiltonian density in terms of the Lagrangian by: $\widehat{\mathscr{H}} = \widehat{\Pi}\left(\partial_0\widehat{\phi}\right) - \widehat{\mathscr{L}}$ (replacing time derivatives of the field operator by their expression in terms of the conjugate momenta).[18]

The Ω_0 term that appears first in eq. 4.41 has the effect of shifting the energy by a constant and thus does not change the physics. The factor is included in eq. 4.41 so that the Hamiltonian is exactly equal to that in eq. 4.27 and so that the energy of the ground state of the free theory is zero. We could avoid the funny Ω_0 term by *normal ordering* the terms that appear in the Hamiltonian. The normal ordering $:\widehat{O}:$ of an operator $\widehat{O}$, is given by expanding $\widehat{O}$ in terms of a's and $a^\dagger$'s and then ordering all products of creation and annihilation operators so that the annihilation operators appear to the right of the creation operators.[19] For example, the normal ordering of $\int d^3x\, a(\vec{x})a^\dagger(\vec{x})a(\vec{x})$ is $:\int d^3x\, a(\vec{x})a^\dagger(\vec{x})a(\vec{x}):\, = \int d^3x\, a^\dagger(\vec{x})a(\vec{x})a(\vec{x})$. We can thus rewrite eq. 4.41 without Ω_0 as:

$$\widehat{\mathscr{H}} = \frac{1}{2}:\widehat{\Pi}^2: + \frac{1}{2}:(\nabla\widehat{\Pi})^2: + \frac{1}{2}m^2:\widehat{\phi}^2:\,. \tag{4.52}$$

[17] This equation is consistent with the expressions for $\widehat{\phi}$ and $\widehat{\Pi}$ in eq. 4.47 & 4.49.

[18] Forgive me the abuse of notation used in defining the derivative with respect to an operator in $\widehat{\Pi}(\vec{x},t) = \frac{\partial\widehat{\mathscr{L}}}{\partial\left\{\partial_0\widehat{\phi}(\vec{x},t)\right\}}$.

[19] Weinberg (1995, §4.2) proves that every operator can be written in terms of creation and annihilation operators.

4.7 An Interacting Quantum Field Theory

In this section we will briefly consider what happens when you add an interaction term to the Lagrangian in eq. 4.51, examining how the Hamiltonian is affected by the new term and explicitly considering the infinitesimal time evolution of a single particle momentum eigenstate. Let's consider adding the term $-\lambda : \widehat{\phi}^4 :$ to our free Lagrangian. This will have the effect of changing the Hamiltonian to:

$$\begin{aligned}
\widehat{H}_{new} &= \widehat{H}_{Free} + \int d^3x \ : \lambda \widehat{\phi}^4(\vec{x}) : \\
&= \int \widetilde{d^3k}\ \omega(\vec{k}) a^\dagger(\vec{k}) a(\vec{k}) \\
&+ \lambda(2\pi)^3 \Bigg\{ \int \widetilde{d^3k_1} ... \widetilde{d^3k_4}\ a(\vec{k}_1) a(\vec{k}_2) a(\vec{k}_3) a(\vec{k}_4) \delta^3(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 + \vec{k}_4) \\
&\quad + 4 \int \widetilde{d^3k_1} ... \widetilde{d^3k_4}\ a^\dagger(\vec{k}_1) a(\vec{k}_2) a(\vec{k}_3) a(\vec{k}_4) \delta^3(\vec{k}_1 - \vec{k}_2 - \vec{k}_3 - \vec{k}_4) \\
&\quad + 6 \int \widetilde{d^3k_1} ... \widetilde{d^3k_4}\ a^\dagger(\vec{k}_1) a^\dagger(\vec{k}_2) a(\vec{k}_3) a(\vec{k}_4) \delta^3(\vec{k}_1 + \vec{k}_2 - \vec{k}_3 - \vec{k}_4) \\
&\quad + 4 \int \widetilde{d^3k_1} ... \widetilde{d^3k_4}\ a^\dagger(\vec{k}_1) a^\dagger(\vec{k}_2) a^\dagger(\vec{k}_3) a(\vec{k}_4) \delta^3(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 - \vec{k}_4) \\
&\quad + \int \widetilde{d^3k_1} ... \widetilde{d^3k_4}\ a^\dagger(\vec{k}_1) a^\dagger(\vec{k}_2) a^\dagger(\vec{k}_3) a^\dagger(\vec{k}_4) \delta^3(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 + \vec{k}_4) \Bigg\} .
\end{aligned} \tag{4.53}$$

In interacting theories, $a^\dagger(\vec{k})$ and $\widehat{\phi}(\vec{x})$ are defined exactly as they were in the free case. What is the effect of all these new terms on an arbitrary state? The first term would annihilate four particles that have a net momentum of zero, the second would annihilate three and create one with the sum of the momenta of the original particles, etc. The terms in 4.53 should look a bit familiar since the interaction term we experimented with at the end of §4.4 (in eq. 4.24) was quite similar. Unlike the

toy theory discussed there, here the vacuum is *not* the ground state of the theory (it should be clear that the vacuum is not even an energy eigenstate).

To see explicitly how a state evolves in this theory, let's consider acting with the Hamiltonian on a one particle momentum eigenstate (as we did in eq. 4.25). After an infinitesimal period of time, the single particle state has some probability of staying a single particle state, some probability of evolving into a three particle state, and some probability of being joined by four particles created from the vacuum:

$$\begin{aligned}\frac{\partial}{\partial t}|\vec{k}\rangle = &-i\omega(\vec{k})|\vec{k}\rangle - i\frac{\sqrt{3!}\lambda}{(2\pi)^3}\int d^3k_1 d^3k_2 \left[\frac{1}{\omega(\vec{k})\omega(\vec{k}_1)\omega(\vec{k}_2)\omega(\vec{k}-\vec{k}_1-\vec{k}_2)}\right]^{\frac{1}{2}} |\vec{k}_1, \vec{k}_2, \vec{k}-\vec{k}_1-\vec{k}_2\rangle \\ &-i\frac{\sqrt{5!}\lambda}{(2\pi)^3}\int d^3k_1 d^3k_2 d^3k_3 \left[\frac{1}{\omega(\vec{k})\omega(\vec{k}_1)\omega(\vec{k}_2)\omega(\vec{k}_3)\omega(-\vec{k}_1-\vec{k}_2-\vec{k}_3)}\right]^{\frac{1}{2}} |\vec{k}_1, \vec{k}_2, \vec{k}_3, -\vec{k}_1-\vec{k}_2-\vec{k}_3, \vec{k}\rangle\,.\end{aligned} \tag{4.54}$$

Notice that in the second and third terms the probability of producing terms with high momentum decreases because of the $\omega^{-\frac{1}{2}}$ factors. The larger λ is, the greater the chance of a change in particle number.

In this chapter we will not go on to derive testable predictions from interacting QFTs. But, you are now prepared to derive the Feynman rules for ϕ^4-theory and calculate amplitudes for various transitions to occur in particle collisions (from which you can derive cross-sections and decay rates). You can start with Peskin and Schroeder's perturbative calculation of transition amplitudes in (Peskin & Schroeder, 1995, §4.2) (then come back and read parts II and III). Or, if you prefer to study QFT from textbooks that start with path integrals (such as Srednicki, 2007 or Zee, 2010), I recommend reading through parts II and III before/while studying their calculations of transition amplitudes.

Part II

From Classical Field Theory to QFT

In part I we moved from QM to QFT by allowing for the creation and annihilation of particles and introducing a relativistic expression for the Hamiltonian in the Schrödinger equation. In part II, we will re-introduce QFT as a quantum version of classical field theory (CFT). We will begin by moving from classical states of definite field configuration to quantum states with indeterminate field configuration (wave functionals over the space of possible field configuration). The time evolution of these states is determined by a Schrödinger equation which uses the CFT expression for the energy of a state. Since the particular field theory we will start from is already relativistic (Klein-Gordon field theory), the quantum theory we end up with will be relativistic as well. In order to show that this formulation of QFT is empirically equivalent to the textbook variety, we will use path-integrals to derive the generating functional for interacting field theories from which the Feynman rules can be derived by standard perturbative methods. Once the path-integral analysis has been completed, it will be straightforward to show that the classical field theory emerges as a limiting case. The version of QFT that we will arrive at in part II, by starting with CFT, will look different from the form examined in Part I. In part III we will demonstrate that the two formulations are just manifestations of a different choice of basis for the abstract, Dirac-notation, formulation of the theory.

4.8 The Space of States: Wave Functionals over Field Configurations

In classical particle dynamics, one specifies the *instantaneous* state[20] of a system of identical particles by picking out a point in configuration space which determines the locations of all the particles: $(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n)$. When we move to a quantum description of the same system, the state becomes a wave function which assigns a complex number to every possible classical configuration for the system: $\Psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n)$. In CFT, the instantaneous state of a region is given by a classical field configuration, a function which assigns a real number to each point in space: $\phi(\vec{x})$. In a quantum description, the state of an isolated region will be an assignment of a complex number to each possible field configuration. That means, the state will be a *wave functional* which takes a *function*, the classical field configuration, as its argument and returns a complex number: $\Psi[\phi(\vec{x})]$ (Ψ depends on the function ϕ but not on a specific value of $\vec{x}$, so I will omit the argument of ϕ in the future but continue to use square brackets to remind ourselves that Ψ is a *functional*, not a function).[21] The role of the wave function is depicted in figure 4.3.

Working with wave functionals over continuous spaces is difficult, less intuitive, and harder to do rigorously, so we will take the discrete case to be central. When

[20] The use of "instantaneous" here is taken from Albert's *Time and Chance* (Albert, 2000, ch. 1, §3). A point in phase space gives more than just the instantaneous state of the system, it says something (independent of the dynamical laws) about how the state of the system is changing.

[21] Wave functional methods are often discussed in textbook treatments of QFT (for example, Huang, 2008, pg. 29-33) but I have rarely seen them bestowed with the fundamentality I give them here. In (Weinberg, 1999), Weinberg suggests their foundational importance, but doesn't provide much explanation about how the story is supposed to go. Wallace (2006, §2.3) provides a wave functional account of states, but does not give a detailed discussion of the dynamics of such states. Hatfield's (1992, ch. 9-14) textbook gives a nice, in-depth treatment of the wave functional approach similar to that developed here. My derivation of the generating functional from the path integral is significantly different from Hatfield's and fits well with the treatment in (Srednicki, 2007, ch. 6-9).

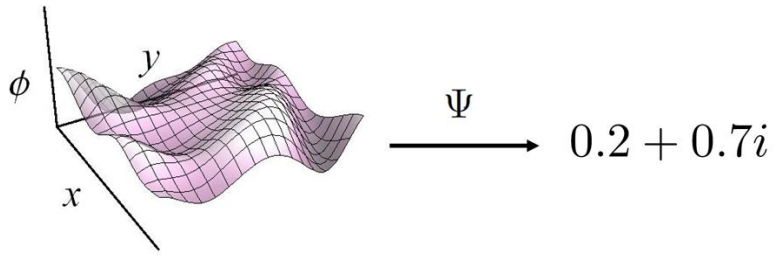

Figure 4.3: **Wave Functional** A 2-D classical field configuration $\phi(x, y)$ is mapped by the wave functional $\Psi[\phi]$ to a complex number.

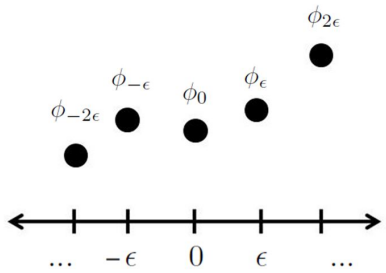

Figure 4.4: **Discrete Field Configuration** A field configuration in 1-D discrete space.

in doubt, any calculation of a transition probability or other prediction of the theory can be done in discrete space and then upon completion the lattice spacing can be taken to zero.

Consider the simple case where space is 1-dimensional and discrete with lattice spacing ϵ, as in figure 4.4. Classically, at each point there is a number ϕ_i that gives the value of the field at that point. Specifying ϕ_i for all i specifies the classical field configuration: $\phi(x) \to \{..., \phi_{-\epsilon}, \phi_0, \phi_\epsilon, ...\}$. So, the quantum wave functional becomes a *function* of as many arguments as there are points in space: $\Psi[\phi(x)] \to \Psi(..., \phi_{-\epsilon}, \phi_0, \phi_\epsilon, ...)$. In this case, it is easy to define the inner product of states by integrating over all possible field configurations:

$$\langle \Psi | \Psi' \rangle = \int ...d\phi_{-\epsilon} d\phi_0 d\phi_\epsilon ... \Psi^*(..., \phi_{-\epsilon}, \phi_0, \phi_\epsilon, ...) \Psi'(..., \phi_{-\epsilon}, \phi_0, \phi_\epsilon, ...) \,. \tag{4.55}$$

Moving to three dimensions, we can define $\prod_{\vec{x},\epsilon}$ as the product over all points on our lattice (with lattice spacing ϵ) so that:

$$\prod_{\vec{x},\epsilon} d\phi_{\vec{x}} = ...d\phi_{-\hat{z}}, d\phi_0, d\phi_{\hat{x}}, d\phi_{\hat{y}}, d\phi_{\hat{z}}, d\phi_{2\hat{x}}... \,. \tag{4.56}$$

Here $\hat{x}/\hat{y}/\hat{z}$ is the vector of length ϵ pointing in the $x/y/z$-direction. In three dimensions, the inner product becomes:

$$\langle \Psi | \Psi' \rangle = \int \prod_{\vec{x},\epsilon} d\phi_{\vec{x}} \Psi^*(...\phi_{-\hat{z}}, \phi_0, \phi_{\hat{x}}, ...) \Psi'(...\phi_{-\hat{z}}, \phi_0, \phi_{\hat{x}}, ...) \,. \tag{4.57}$$

The probability that a certain state Ψ' is observed when the wave functional of the system is Ψ is, as in QM, given by $|\langle \Psi' | \Psi \rangle|^2$. States are normalized by requiring that: $|\langle \Psi | \Psi \rangle|^2 = 1$. In the case of a discrete lattice, mathematically difficult-to-understand integrations over the space of functions become more straightforward integrations over field values at points:

$$\int D\phi F[\phi] \quad \longrightarrow \quad \int \prod_{\vec{x},\epsilon} d\phi_{\vec{x}} F(...\phi_{-\hat{z}}, \phi_0, \phi_{\hat{x}}, ...) \,. \tag{4.58}$$

Here $\int D\phi F[\phi]$ is the integral of the functional $F[\phi]$ over all possible field configurations ϕ.

We can define a derivative of a functional with respect to a function at a specified point $\vec{x}$ as follows:

$$\frac{\delta F[\phi]}{\delta \phi(\vec{x})} \equiv \lim_{h \to 0} \frac{F[\phi^*] - F[\phi]}{h} \,. \tag{4.59}$$

ϕ^* is equal to the function ϕ everywhere except at the point $\vec{x}$ where its value is $\phi^*(\vec{x}) = \phi(\vec{x}) + h$. We are considering the effect of changing the function by an arbitrarily small amount at $\vec{x}$ and seeing how the value of the functional depends on

this shift. In the discrete case, this becomes:

$$\frac{\delta F[\phi]}{\delta \phi(\vec{x})} \quad \longrightarrow \quad \lim_{h \to 0} \quad \frac{F(..., \phi_{\vec{x}} + h, ...) - F(..., \phi_{\vec{x}}, ...)}{h} \,. \tag{4.60}$$

Here we consider the difference between the value of the functional at the field configuration ϕ and the configuration obtained by shifting the value of the field at $\vec{x}$ by an arbitrarily small amount (leaving the field value at all other points unchanged, as depicted in figure 4.5). It is useful to note that functional derivatives obey the product rule and the chain rule.[22]

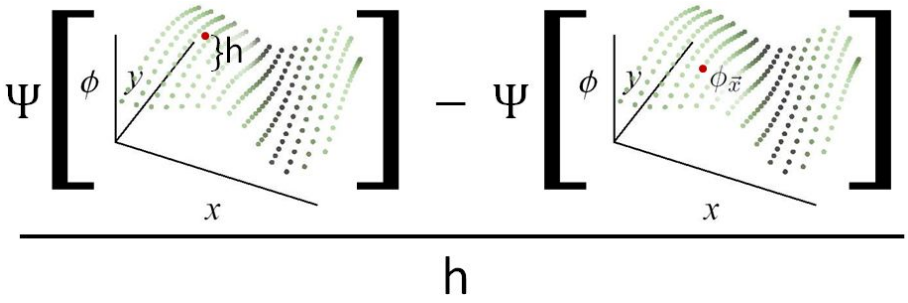

Figure 4.5: **The Functional Derivative** Visual depiction of the definition of a functional derivative in the discrete case (with two spatial dimensions). Here we consider the amplitude assigned to to the function that differs from ϕ by an arbitrarily small displacement in the field value at a particular point $\vec{x}$.

We can move from a wave functional representation to a "ket" by integrating over all possible states of definite field configuration:

$$|\Psi\rangle = \int D\phi \Psi[\phi] |\phi\rangle \,. \tag{4.61}$$

We require that $|\phi\rangle$ is normalized so that the inner product of two states is a delta *functional*. In the discrete case, the delta functional becomes an infinite product of

[22] See (Hatfield, 1992, ch. 9) for a detailed discussion of functional differentiation and methods for solving functional differential equations.

delta functions of the field value at each point in space:

$$\langle\phi|\phi'\rangle = ...\delta\left(\phi_{-\hat{z}} - \phi'_{-\hat{z}}\right)\delta\left(\phi_0 - \phi'_0\right)\delta\left(\phi_{\hat{x}} - \phi'_{\hat{x}}\right)... = \prod_{\vec{x},\epsilon}\left[\delta\left(\phi_{\vec{x}} - \phi'_{\vec{x}}\right)\right] . \tag{4.62}$$

With this requirement, eq. 4.57 follows from eq. 4.61. Note that the state $|\phi\rangle$ is *not* properly normalized, $\langle\phi|\phi\rangle = \infty$ (just as $\langle\vec{x}|\vec{x}'\rangle = \delta^3\left(\vec{x} - \vec{x}'\right)$ makes $|x\rangle$ un-normalized in QM[23]). The wave functional for the state $|\phi\rangle$ is given by: $\Psi_\phi[\phi'] = \prod_{\vec{x},\epsilon}\left[\delta\left(\phi_{\vec{x}} - \phi'_{\vec{x}}\right)\right]$ (this follows from eq. 4.62 & 4.57). From eq. 4.57 and our expression for $\Psi_\phi[\phi']$, we see that $\langle\phi|\Psi\rangle = \Psi[\phi]$, so eq. 4.61 can be written as:

$$|\Psi\rangle = \int D\phi\,\langle\phi|\Psi\rangle|\phi\rangle\,. \tag{4.63}$$

Thus we have the following representation of the identity:

$$1 = \int D\phi\,|\phi\rangle\langle\phi|\,. \tag{4.64}$$

[23] See the note on probability at the end of §4.2.

4.9 The Schrödinger Equation for Wave Functionals

The fundamental law of temporal evolution in this wave functional understanding of QFT is[24]:

$$\begin{aligned} i\tfrac{\partial}{\partial t}\Psi[\phi,t] &= \widehat{H}\Psi[\phi,t] \\ &= \int d^3x \left[\tfrac{1}{2}\left(\widehat{\Pi}(\vec{x})\right)^2 + \tfrac{1}{2}\left(\nabla\widehat{\phi}(\vec{x})\right)^2 + \tfrac{1}{2}m^2\left(\widehat{\phi}(\vec{x})\right)^2 - \Omega_0 + V_0\left(\widehat{\phi}(\vec{x})\right)\right]\Psi[\phi,t]\,. \end{aligned} \tag{4.65}$$

This is simply the QM Schrödinger equation where states are now mappings from field configurations and times to complex numbers, wave functionals, instead of functions on configuration space and the Hamiltonian gives the energy of a Klein-Gordon field (with interactions if $V_0 \neq 0$). In the free case, we need to shift the energy by Ω_0 if we would like the Hamiltonian that appears in the (free) Schrödinger equation (eq. 4.65) to agree with the one introduced in part I (eq. 4.27), as we've already shown that the Hamiltonian in eq. 4.46 is the same as that in eq. 4.41, which is the Hamiltonian in the Schrödinger equation above, eq. 4.65 (except that eq. 4.41 has $V_0 = 0$). In appendix 4.A I argue that Ω_0 is also needed in interacting theories so that we can ensure that the ground state of the theory has zero energy (which is an experimentally unobservable but calculationally useful feature). To understand eq. 4.65, we must understand the operators involved. In part I, the operators denoted by $\widehat{\phi}$ and $\widehat{\Pi}$ were introduced in eq. 4.39. For now, pretend that you don't know about those definitions and you've just encountered two new operators (in §4.12 we'll show

[24] This equation is given in (Hatfield, 1992, eq. 10.12) and is used as the starting point for a Bohmian version of bosonic QFT in (Bohm & Hiley, 1993, eq. 11.26).

that we are discussing the same operators here as those in eq. 4.42, but that shouldn't be obvious yet). $\widehat{\phi}(\vec{x})$, when acting on a field configuration eigenstate, pulls out the value of the field at $\vec{x}$ according to that field configuration: $\widehat{\phi}(\vec{x})|\phi\rangle = \phi(\vec{x})|\phi\rangle$. $\widehat{\Pi}$ acts on a state by taking the functional derivative of the wave functional with respect to the field value at $\vec{x}$.[25] The operators $\widehat{\phi}$ and $\widehat{\Pi}$ are defined by:

$$\begin{aligned}\widehat{\phi}(\vec{x})|\Psi\rangle &\equiv \int D\phi\, \phi(\vec{x})|\phi\rangle\langle\phi|\Psi\rangle \\ &= \int D\phi\, \phi(\vec{x})\Psi[\phi]|\phi\rangle \\ \widehat{\Pi}(\vec{x})|\Psi\rangle &= \int D\phi\ -i\frac{\delta\Psi[\phi]}{\delta\phi(\vec{x})}|\phi\rangle\,. \end{aligned} \tag{4.66}$$

Be careful! Here $\widehat{\phi}$ is an operator, ϕ is a specific field configuration, and $|\phi\rangle$ is a field configuration eigenstate with definite field configuration ϕ. From these definitions, we can derive the commutation relations for $\widehat{\phi}$ and $\widehat{\Pi}$ (and they agree with those of eq. 4.50):

$$[\widehat{\phi}(\vec{x}), \widehat{\phi}(\vec{x}')] = [\widehat{\Pi}(\vec{x}), \widehat{\Pi}(\vec{x}')] = 0$$

$$[\widehat{\phi}(\vec{x}), \widehat{\Pi}(\vec{x}')] = i\delta^3\left(\vec{x} - \vec{x}'\right)\,. \tag{4.67}$$

Here we use the fact that $\frac{\delta\phi(\vec{x})}{\delta\phi(\vec{x}')} = \delta^3\left(\vec{x} - \vec{x}'\right)$. Note that (so long as $\langle\Psi|\widehat{H}|\Psi\rangle$ is finite) the hermiticity of $\widehat{H}$ guarantees conservation of probability.[26]

[25] One might reasonably wonder whether the definition of the conjugate field momenta operator $\widehat{\Pi}(\vec{x})$ in eq. 4.66 is consistent with the definition of conjugate momenta in eq. 4.49 as: $\widehat{\Pi}(\vec{x},t) \equiv \frac{\partial\mathscr{L}}{\partial\{\partial_0\widehat{\phi}(\vec{x},t)\}}$ which is $\partial_0\widehat{\phi}(\vec{x},t)$ for our Lagrangian from eq. 4.51. In fact it is. Taking the time derivative of $\widehat{\phi}(\vec{x},t) = e^{i\widehat{H}t}\widehat{\phi}(\vec{x})e^{-i\widehat{H}t}$ (see eq. 4.45) with $\widehat{H}$ defined in terms of $\widehat{\phi}$ and $\widehat{\Pi}$ by eq. 4.65 (using the commutator from eq. 4.67) gives $\partial_0\widehat{\phi}(\vec{x},t) = \widehat{\Pi}(\vec{x},t)$.

[26] *Proof*:

$$\frac{\partial}{\partial t}\langle\Psi|\Psi\rangle = \frac{\partial}{\partial t}\left(\int D\phi\Psi^*[\phi]\Psi[\phi]\right) = \int D\phi\left(\frac{\partial}{\partial t}\Psi^*[\phi]\right)\Psi[\phi] + \int D\phi\Psi^*[\phi]\left(\frac{\partial}{\partial t}\Psi[\phi]\right) = i\langle\widehat{H}^\dagger\rangle - i\langle\widehat{H}\rangle = 0\,. \tag{4.68}$$

As we did in Part I, let us briefly summarize the theory we've developed:

Theory 4: Real Scalar QFT (in the *field* basis)

States The states of the theory are wave functionals on the space of possible field configurations: $\Psi[\phi]$

Schrödinger Equation

$$i\tfrac{\partial}{\partial t}\Psi[\phi,t] = \left(\int d^3x \left[\tfrac{1}{2}\left(\widehat{\Pi}(\vec{x})\right)^2 + \tfrac{1}{2}\left(\nabla\widehat{\phi}(\vec{x})\right)^2 + \tfrac{1}{2}m^2\left(\widehat{\phi}(\vec{x})\right)^2 - \Omega_0\right] + \widehat{H}_{int}\right)\Psi[\phi,t]\,. \tag{4.69}$$

Probabilities The probability (or probability density) of observing the system in state $|\Phi\rangle$ in an appropriate experiment is given by the square of the inner product of $|\Phi\rangle$ and the system's state $|\Psi\rangle$: $|\langle\Phi|\Psi\rangle|^2$. The inner product is defined in equation 4.57 which in the continuous case becomes:

$$\langle\Phi|\Psi\rangle = \int D\phi\ \Phi^*[\phi]\Psi[\phi]\,. \tag{4.70}$$

4.10 To Predictions via Path Integrals

Now, if this were a QFT textbook we would be obligated to go on to derive the Feynman rules and calculate cross-sections. These skills are important, but as we are only trying to clarify the theory's foundations, we will not go on to calculate predictions of the theory here. However, it is important that you are convinced that

the formulation of QFT discussed here is capable of entailing the standard Feynman rules for interacting Klein-Gordon fields. To show this capability, appendix 4.A uses the Schrödinger equation for wave functionals (eq. 4.65) to derive the following equation for ground state expectation values of time-ordered field operators (as in Srednicki, 2007, eq. 9.5)[27]:

$$\langle\Omega|T\widehat{\phi}(\vec{x},t)\widehat{\phi}(\vec{x}^{*},t^{*})...|\Omega\rangle = \left[\frac{1}{i}\frac{\delta}{\delta J(\vec{x},t)}\frac{1}{i}\frac{\delta}{\delta J(\vec{x}',t')}...\int \mathbf{D}\phi\,\mathrm{Exp}\left[i\int d^{4}x(\mathscr{L}+J\phi)\right]\right]_{J=0}. \tag{4.71}$$

Here $\int \mathbf{D}\phi$ is an integral over all paths through the space of possible field configurations, $|\Omega\rangle$ is the time-independent[28] ground state (which is in general not equal to the zero-particle vacuum state), and $\widehat{\phi}(\vec{x},t)$ pulls out the value of the field at t (eq. 4.45 still holds as the definition of $\widehat{\phi}(\vec{x},t)$ in terms of $\widehat{\phi}(\vec{x})$). T in eq. 4.71 denotes the time-ordered product of the field operators, enforcing a re-ordering of the field operators so that those with later times appear first when reading from left to right. From eq. 4.71, one can use the LSZ reduction formula to calculate the Feynman rules for the theory (see, for example, Srednicki, 2007, ch. 5, 9, 10).[29] Empirical predictions of the theory, like decay rates and cross-sections, can then be calculated from the Feynman rules. In order to keep in mind the similarities between

[27] One is not forced to use path integrals and reduction formulas to calculate measurable predictions from interacting scalar field theories in the field basis. Hatfield (1992, ch. 11) shows that you can calculate cross sections by perturbatively calculating the time evolution of wave functionals using standard (non-path-integral) methods from QM (like those used to derive eq. 4.26).

[28] The state is time-independent because Ω_0 is chosen so that the ground state has zero energy.

[29] To derive the LSZ formula, we need a Hamiltonian which satisfies various constraints listed in (Srednicki, 2007, ch. 5) (typically we start with one that doesn't and then change it through renormalization). A note on following Srednicki's derivation in light of part I: First, it is $a_0^{\dagger}$ not $a^{\dagger}$ which creates a normalized one particle state, so the wave function in eq. 5.6 must be chosen to counteract the un-normalized nature of $a^{\dagger}(\vec{k})|0\rangle$. Second, It is not obvious that the ground state of a renormalized interacting theory is the same as the vacuum, so it is not obvious that $a(\vec{k})$ will annihilate the ground state. However, if we have a Hamiltonian for which Srednicki's assumptions are satisfied (eq. 5.25 & 5.19), it follows that $a(\vec{k})$ (defined in terms of $\widehat{\phi}$ & $\widehat{\Pi}$ in eq. 4.75) *will* annihilate the ground state. So, *for renormalized theories*, the ground state is the vacuum.

QFT and QM, here is the single-particle QM analogue of eq. 4.71:

$$\langle \Omega, t = t_f | T\hat{x}(t)\hat{x}(t')...|\Omega, t = t_i\rangle = \left[\tfrac{1}{i}\tfrac{\delta}{\delta f(t)}\tfrac{1}{i}\tfrac{\delta}{\delta f(t')}... \int Dx \,\mathrm{Exp}\left[i\int_{t_i}^{t_f} dt(L + fx)\right]\right]_{f=0}. \tag{4.72}$$

Note that $\int Dx$ is an integral over the space of paths through *space*, the field operators are replaced by position operators, and the state $|\Omega, t = t_i\rangle$ is the ground state at $t = t_i$ (not the vacuum).

4.11 The Limiting case of Classical Field Theory

QM limits to classical mechanics when the variation in the action ($\int d^3x\mathscr{L}$) is large relative to $\hbar$. Similarly, QFT limits to classical field theory when the variation in the action is large relative to $\hbar$.[30] Let's begin with eq. 4.102 (a step in the derivation of eq. 4.71 from eq. 4.65 in appendix 4.A) with the factor of $\hbar$ restored.

$$\langle \phi'', t''|\phi', t'\rangle = \int \mathscr{D}\phi \mathrm{Exp}\left[\frac{i}{\hbar}\int_{t'}^{t''} dt \int d^3x \mathscr{L}(\dot{\phi}(\vec{x}, t), \phi(\vec{x}, t))\right]. \tag{4.73}$$

Here $\int \mathscr{D}\phi$ is an integral over all paths through the space of field configurations keeping the endpoints fixed at $\phi_0 = \phi'$ and $\phi_{N+1} = \phi''$. We are summing a contribution for each possible classical evolution of the field (see figure 4.6 in appendix 4.A for a visual depiction of the path integral in eq. 4.73). When the variation in the action is large relative to $\hbar$, we can use a steepest descent approximation to calculate the transition amplitude. The integral will be dominated by the region in the space of paths through field configuration space where $\int_{t'}^{t''} dt \int d^3x \mathscr{L}(\dot{\phi}(\vec{x}, t), \phi(\vec{x}, t))$ is minimized. The transition amplitude then (approximately) only gets a contribution

[30] This limit is appropriately introduced early in (Zee, 2010, §1.3).

from a single path (that is, a single time evolution for a definite-valued classical field):

$$\langle \phi'', t'' | \phi', t' \rangle \propto \text{Exp}\left[\frac{i}{\hbar} \int_{t'}^{t''} dt \int d^3x \mathscr{L}(\dot{\phi}^*(\vec{x}, t), \phi^*(\vec{x}, t)) \right] . \tag{4.74}$$

So, the system evolves along a determinate trajectory through the space of field configurations. What path $\phi^*(\vec{x}, t)$ minimizes $\int_{t'}^{t''} dt \int d^3x \mathscr{L}(\dot{\phi}^*(\vec{x}, t), \phi^*(\vec{x}, t))$? The path of least action! So, the CFT principal of least action (from which one can derive classical equations of motion for fields) arises from a limiting case of the QFT dynamics.

We have now shown how QFT limits to QM when the momenta are small (relative to mc, see §4.5) and to relativistic CFT when the variation in the action is large (relative to $\hbar$). However, we used different formulations of the theory to get each limiting case. In part III, we will show that the formulations in parts I and II (theories 3 & 4) are in fact equivalent.

Part III

Relating the Particle and Field Bases

In part I we developed theory 3, a formulation of QFT where the physical states are wave functions on the space of all possible positions for all possible configurations of any number of particles and time evolution is determined by eq. 4.38. In part II we developed theory 4, a formulation where the possible states are wave functionals on the space of possible field configurations and time evolution is determined by eq. 4.69. As it turns out, remarkably, theories 3 & 4 are actually equivalent! At the level of QFT, the world seems to be representable in terms of fields *or* particles, and we can freely go back and forth between representations at our convenience. Just as QM can be developed in the position or momentum bases, QFT can be developed in the particle or field bases (actually, it can be developed in the field basis, theory 4, or the field-momenta basis, using eq. 4.96, or the particle-position basis, theory 3, or the particle-momentum basis, §4.4 & 4.5, or the particle-energy basis (if the conditions are right), see Teller, 1997, ch. 3 with $\hat{A}^{(1)} = \widehat{H}$, or ...).[31]

Part III is short. In the first section, we will show that you can derive theory 3 from theory 4 if we posit a certain mapping between states. In the second and final section, I will very briefly introduce the measurement problem as it manifests itself in QFT.

[31] Thank you to David Wallace for suggesting to me in conversation the idea that the particle and field formulations of QFT are merely different bases in which we can represent states (I do not mean to suggest that he would endorse the presentation here, although he might).

4.12 Making Particles out of Fields

In this section we will show that theory 3 can be derived from theory 4. To-do list: The creation and annihilation operators $a^\dagger$ and a do not appear in part II so we must find a way to define them in the context of theory 4. In part I, we defined $\widehat{\phi}$, $\widehat{\Pi}$, and an inner product. Since these things were defined again in part II, we must show that the "definitions" in part I can be derived as *consequences* of the definitions in part II. Further, we introduced a Schrödinger equation (eq. 4.38) for theory 3, but if theory 4 is to underlie theory 3 we must show that eq. 4.38 can be derived from eq. 4.65 (actually, we've already basically done this in eq. 4.39-4.41). In positing the equivalence of the two formulations, we are free to choose a mapping between wave functions and wave functionals so as to preserve the machinery of theory 3.

$\widehat{\phi}$ and $\widehat{\Pi}$ were defined in part II by eq. 4.66, but a and $a^\dagger$ were not introduced. There are no wave functions on the union of configuration spaces in theory 4, so a and $a^\dagger$ cannot be defined by their actions on such wave functions (as in §4.4). However, we can freely define $a(\vec{k})$ and $a^\dagger(\vec{k})$ by the inversions of our expressions for $\widehat{\phi}$ and $\widehat{\Pi}$ (eq. 4.42[32]) from §4.6 in part I:

$$\begin{aligned} a(\vec{k}) &= \int d^3x \; e^{i\vec{k}\cdot\vec{x}} \left[i\widehat{\Pi}(\vec{x}) + \omega(\vec{k})\widehat{\phi}(\vec{x}) \right] \\ a^\dagger(\vec{k}) &= \int d^3x \; e^{-i\vec{k}\cdot\vec{x}} \left[-i\widehat{\Pi}(\vec{x}) + \omega(\vec{k})\widehat{\phi}(\vec{x}) \right]. \end{aligned} \tag{4.75}$$

From these, we can define a_0 and $a_0^\dagger$ by scaling a and $a^\dagger$ as in eq. 4.43.

We can now use the creation and annihilation operators to determine the mapping

[32] For the relation between $\widehat{\Pi}$ (as in eq. 4.49) and $\widehat{\mathscr{L}}$ to be $\widehat{\Pi}(\vec{x},t) = \frac{\partial \mathscr{L}}{\partial\{\partial_0 \widehat{\phi}(\vec{x},t)\}}$, the dependence of $\mathscr{L}$ on $\partial_0\widehat{\phi}$ must be as in eq. 4.51, which it would be if we are only considering interaction terms like those allowed in 4.65.

between wave functions and wave functionals. We can define the vacuum wave functional as the one which is annihilated by $a(\vec{k})$:

$$a(\vec{k})|0\rangle = \int D\phi\ a(\vec{k})|\phi\rangle\langle\phi|0\rangle = \int D\phi\ a(\vec{k})\Psi_0[\phi]|\phi\rangle\,. \tag{4.76}$$

This yields (derivation in Huang, 2008, §2.12[33]):

$$\Psi_0[\phi] = \eta\ \text{Exp}\left[-\int d^3x d^3x'\ \phi(\vec{x})\phi(\vec{x}')\int \widetilde{d^3k}\ \omega^2(\vec{k})e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}\right]\,. \tag{4.77}$$

η is the normalization of the state (given in Hatfield, 1992, pg. 203). Then, all other states are defined by acting on the vacuum with the appropriate creation operators. We will now explicitly construct one and two particle states to see what particle states look like in the field basis. First, the wave functional for a single particle with momentum $\vec{k}$:

$$\begin{aligned}
\Psi_{\vec{k}}[\phi] = \langle\phi|\,\vec{k}\,\rangle = \langle\phi|a_0^\dagger(\vec{k})|0\rangle &= \frac{1}{\sqrt{(2\pi)^3 2\omega(\vec{k})}}\langle\phi|a^\dagger(\vec{k})|0\rangle \\
&= \tfrac{1}{\sqrt{(2\pi)^3 2\omega(\vec{k})}}\int D\phi'\left\{\left(\int d^3x\ e^{-i\vec{k}\cdot\vec{x}}\left[-\tfrac{\delta}{\delta\phi'(\vec{x})}+\omega(\vec{k})\phi'(\vec{x})\right]\right)\langle\phi|\phi'\rangle\langle\phi'|0\rangle\right\} \\
&= \frac{2\sqrt{\omega(\vec{k})}}{\sqrt{2(2\pi)^3}}\int d^3x\ \phi(\vec{x})e^{i\vec{k}\cdot\vec{x}}\ \Psi_0[\phi]\,.
\end{aligned} \tag{4.78}$$

In moving to the second line, we insert a complete set of field eigenstates and use eq. 4.66 to replace the $\widehat{\phi}$ and $\widehat{\Pi}$ operators with the field ϕ' and the functional derivative with respect to it, $\frac{\delta}{\delta\phi'}$. In the next step we use the fact that the inner product of field eigenstates is the delta functional to integrate over ϕ' and act with the functional derivative on the vacuum state from eq. 4.77. Acting with two creation operators

[33] Alternate derivations in (Bohm & Hiley, 1993, §11.4) and (Hatfield, 1992, §10.1).

gives a similar result:[34]

$$\Psi_{\vec{k}_1,\vec{k}_2}[\phi] = \langle\phi|a_0^\dagger(\vec{k}_1)a_0^\dagger(\vec{k}_2)|0\rangle = \frac{4\sqrt{\omega(\vec{k}_1)\omega(\vec{k}_2)}}{2(2\pi)^3}\int d^3x d^3y\ \phi(\vec{x})\phi(\vec{y})e^{i\vec{k}_1\cdot\vec{x}}e^{i\vec{k}_2\cdot\vec{y}}\ \Psi_0[\phi]\,. \tag{4.79}$$

Now we can generate a mapping between mapping between wave functions and wave functionals. In part I, we could write an arbitrary state in terms of n-particle wave functions (see eq. 4.15):

$$|\Psi\rangle = \Psi_0|0\rangle + \int d^3k_1\ \widetilde{\Psi}_1(\vec{k}_1)|\vec{k}_1\rangle + \int d^3k_1 d^3k_2\ \widetilde{\Psi}_2(\vec{k}_1,\vec{k}_2)|\vec{k}_1,\vec{k}_2\rangle + \ldots\,. \tag{4.80}$$

In part II, we used the wave functional to specify a state (eq. 4.61):

$$|\Psi\rangle = \int D\phi\Psi[\phi]|\phi\rangle\,.$$

If we insert a complete set of n-particle states in eq. 4.61 (as in eq. 4.37), we can get it in the form of eq. 4.80.

$$\begin{aligned}|\Psi\rangle &= \int D\phi\ \left\{\Psi[\phi]\left(|0\rangle\langle 0| + \int d^3k_1|\vec{k}_1\rangle\langle\vec{k}_1| + \int d^3k_1d^3k_2|\vec{k}_1,\vec{k}_2\rangle\langle\vec{k}_1,\vec{k}_2| + \ldots\right)|\phi\rangle\right\}\\ &= \int D\phi\ \Psi_0^*[\phi]\Psi[\phi]|0\rangle + \int d^3k_1\int D\phi\ \Psi^*_{\vec{k}_1}[\phi]\Psi[\phi]|\vec{k}_1\rangle + \int d^3k_1d^3k_2\int D\phi\ \Psi^*_{\vec{k}_1,\vec{k}_2}[\phi]\Psi[\phi]|\vec{k}_1,\vec{k}_2\rangle + \ldots\end{aligned} \tag{4.81}$$

Here $\Psi_0[\phi]$, $\Psi_{\vec{k}_1}[\phi]$, and $\Psi_{\vec{k}_1,\vec{k}_2}[\phi]$ are specified by eq. 4.77, 4.78, and 4.79 respectively (wave functionals for more particles would be generated similarly). Thus, comparing eq. 4.80 and 4.81, the following relations specify the wave function in terms of the

[34] One must be very careful with commutators here. $a_0^\dagger(\vec{k}_1)a_0^\dagger(\vec{k}_2)\Psi_0[\phi]$ does not give the same result.

wave functional for a particular state $|\Psi\rangle$:

$$\begin{aligned}
\Psi_0 &= \int D\phi\ \Psi_0^*[\phi]\Psi[\phi] \\
\widetilde{\Psi}_1(\vec{k}_1) &= \int D\phi\ \Psi^*_{\vec{k}_1}[\phi]\Psi[\phi] \\
\widetilde{\Psi}_2(\vec{k}_1, \vec{k}_2) &= \int D\phi\ \Psi^*_{\vec{k}_1,\vec{k}_2}[\phi]\Psi[\phi] \\
&\vdots
\end{aligned} \tag{4.82}$$

The wave functional can be determined from the wave function by inserting a complete set of field eigenstates in eq. 4.80 and comparing to 4.61, yielding:

$$\Psi[\phi] = \Psi_0\Psi_0[\phi] + \int d^3k_1\ \widetilde{\Psi}_1(\vec{k}_1)\Psi_{\vec{k}_1}[\phi] + \int d^3k_1 d^3k_2\ \widetilde{\Psi}_2(\vec{k}_1, \vec{k}_2)\Psi_{\vec{k}_1,\vec{k}_2}[\phi] + \dots . \tag{4.83}$$

Next, we need to show that the inner product defined in part I, eq. 4.7, is equivalent to the inner product on the space of wave functionals (eq. 4.70).

$$\begin{aligned}
\langle\Phi|\Psi\rangle &= \Phi_0^*\Psi_0 + \textstyle\int d^3k_1\ \widetilde{\Phi}_1^*(\vec{k}_1)\widetilde{\Psi}_1(\vec{k}_1) + \int d^3k_1 d^3k_2\ \widetilde{\Phi}_2^*(\vec{k}_1, \vec{k}_2)\widetilde{\Psi}_2(\vec{k}_1, \vec{k}_2) + \dots \\
&= \textstyle\int D\phi D\phi'\ \Phi^*[\phi]\Psi[\phi']\Psi_0^*[\phi']\Psi_0[\phi] + \int d^3k_1\ \int D\phi D\phi'\ \Phi^*[\phi]\Psi[\phi']\Psi^*_{\vec{k}_1}[\phi']\Psi_{\vec{k}_1}[\phi] + \dots \\
&= \textstyle\int D\phi D\phi'\ \Phi^*[\phi]\Psi[\phi']\langle\phi'|\left\{|0\rangle\langle 0| + \int d^3k_1|\vec{k}_1\rangle\langle\vec{k}_1| + \int d^3k_1 d^3k_2|\vec{k}_1, \vec{k}_2\rangle\langle\vec{k}_1, \vec{k}_2| + \dots\right\}|\phi\rangle \\
&= \textstyle\int D\phi\ \Phi^*[\phi]\Psi[\phi]\,.
\end{aligned} \tag{4.84}$$

From line 1 to 2 we use eq. 4.82. In moving to the next line we rewrite the $\Psi_0[\phi]$, $\Psi_{\vec{k}_1}[\phi]$, etc. as inner products and rearrange terms. Then, we note that the complete set of states is the identity and get a delta functional from the inner product of field configuration eigenstates.

Finally we need to show that the particle basis Schrödinger equation (eq. 4.38) can be derived from eq. 4.65. In §4.6, eq. 4.39-4.41, we showed that *if* the relations

between $\widehat{\phi}$, $\widehat{\Pi}$, $a^\dagger$, and a posited in eq. 4.39 hold, then the Hamiltonian in eq. 4.38 is equivalent to 4.41, which is the Hamiltonian that appears in the field basis Schrödinger equation, eq. 4.65. By *defining* a and $a^\dagger$ by eq. 4.75, we have ensured that eq. 4.39 is valid.

Success! We have shown that theory 3 can be reduced to theory 4. It should be clear that you could use similar reasoning to go the other way and ground theory 4 in theory 3. So, the two theories are equivalent.

4.13 Setting up the Measurement Problem in QFT

The presentation of QFT provided in this chapter allows us to briefly introduce some possible solutions to the measurement problem in relativistic quantum field theory. The problem, simply stated, is to find out what makes **Probabilities** (in the statement of theories 3 & 4) true. In basis-independent language, the condition is:

Probabilities The probability of observing the system in state $|\Phi\rangle$ (or probability density) in an appropriate experiment is given by the square of the inner product of $|\Phi\rangle$ and the system's state $|\Psi\rangle$: $|\langle\Phi|\Psi\rangle|^2$.

One simple answer is to refuse to precisify condition or derive it. Some processes count as "appropriate experiments" and when you do one of these it is an axiom of the theory that probabilities for various results are given by $|\langle\Phi|\Psi\rangle|^2$. Upon being observed to be in a certain subset of all possible states, the wave function (or wave functional) collapses so that it assign non-zero amplitude only to states in this subset. This is the QFT-analogue of a simple version of the Orthodox or Copenhagen interpretation of QM (in QM, see von Neumann, 1955 for von Neumann's version and Albert, 1994, ch.5 for philosophical discussion of it).

Another option is to deny that collapse occurs and deny that **Probabilities** is a postulate of the theory, arguing that the entire content of the theory is captured by the space of states and the universally valid Schrödinger equation. This is the strategy advised by modern Everettians who defend the *many-worlds interpretation* of quantum theories.[35] David Wallace has argued briefly (Wallace, 2008, §7.3) that the Many-worlds interpretation extends unproblematically and smoothly to QFT and this seems plausible. In the framework presented here, the Everettian argues that the abstract representation of the theory introduced in part III is correct and **Probabilities** can somehow be derived from Schrödinger equation evolution (recent arguments pursue a decision theoretic line, see (Deutsch, 1999; Greaves, 2004, 2007; Wallace, 2010, 2012); for an alternative see (Carroll & Sebens, 2014; Sebens & Carroll, forthcoming)). Which representation of the theory (3 or 4 or something else) is most useful for representing the emergent branching of worlds would be settled by decoherence-based arguments (see Wallace, 2003; Schlosshauer, 2005).

Extending Bohmian mechanics (a.k.a. de Broglie-Bohm Pilot Wave theory) to QFT is certainly less straightforward. According to Bohmian mechanics, the world contains more than just a wave function: there are particles with definite positions.[36] These particles get "pushed around" by the wave function in a way that ensures that probabilities are typically given by the amplitude-squared. This is a "hidden variables" interpretation of QM, and has often been criticized for including some sort of relativistically-unacceptable action at a distance. But, this is a bit unfair. For the theory to be compatible with special relativity, we must surely move to a Bohmian version of *relativistic* quantum field theory. In extending the Bohmian approach to

[35]For a brief introduction to the theory in its modern form, see (Wallace, 2008, §4), or for a book-length treatment, see (Wallace, 2012). For the original formulation, see (Everett III, 1957).

[36]For a brief introduction to Bohmian mechanics in QM, see (Berndl *et al.* , 1995). For a more comprehensive treatment of the theory, see (Dürr & Teufel, 2009).

QFT, it is not obvious what the definite states should be: particles with definite positions or a field in a definite configuration? That is, should we work from theory 3 and add particles with definite position or work from theory 4 stipulating that the world is in fact in a definite field configuration at each time? Both strategies have been pursued and it seems that, at least in the case of spin-0 bosons / real scalar fields, both can be made to work. Bell-type quantum field theories take the world to be in a definite state in the disjoint union of n-particle configuration spaces. There really are a certain number of particles in a particular arrangement. One recent version argues that particles typically move in a deterministic manner, guided by the wave function, but there are stochastic jumps from one n-particle configuration space to another when particles are created or destroyed (see Dürr *et al.* , 2004, 2005). This theory suggests taking Feynman diagrams completely seriously as possible trajectories for real particles. Another option is to take the world to contain no particles, only fields and wave functionals. However, the fields have definite values at each spacetime point. This version of Bohmian mechanics takes the formulation of QFT in theory 4 as more fundamental. This kind of Bohmian QFT was proposed by David Bohm himself (together with Basil Hiley) in (Bohm & Hiley, 1993), so we will call them "Bohm-type" approaches to Bohmian QFT. A recent discussion of both types of Bohmian QFTs can be found in (Tumulka, 2007); a survey of Bohm-type approaches is given in (Struyve, 2010).

GRW's spontaneous collapse version of QM has been extended to non-interacting relativistic quantum mechanics (Tumulka, 2006c,a) and to non-relativistic quantum field theory (Tumulka, 2006b), but it has not yet been extended to the realm of relativistic quantum field theory.

Acknowledgments: I would like to especially thank Ratindranath Akhoury, David Baker, and Laura Ruetsche for comments on drafts. In addition, thanks to Daniel

Peterson, Emanuele Rossanese, and Mark Wise for helpful discussion. I would also like to that the audience at the So. Cal. Philosophy of Physics reading group on 8/8/12 at UC Irvine. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 0718128.

4.A Computing Path Integrals with Wave Functionals

In the following derivation I will try to highlight a couple (minor?) assumptions that I am somewhat unsatisfied with having to make in the hopes that some astute reader will show me why they are legitimate or unnecessary. We would like to derive the following expression (eq. 4.71)[37]:

$$\langle\Omega|T\widehat{\phi}(\vec{x},t)\widehat{\phi}(\vec{x}^*,t^*)...|\Omega\rangle = \left[\tfrac{1}{i}\tfrac{\delta}{\delta J(\vec{x},t)}\tfrac{1}{i}\tfrac{\delta}{\delta J(\vec{x}',t')}... \int \mathbf{D}\phi \,\text{Exp}\left[i\int d^4x(\mathscr{L}+J\phi(\vec{x},t))\right]\right]_{J=0}. \tag{4.85}$$

First, we will write $\langle\Omega|T\widehat{\phi}(\vec{x},t)\widehat{\phi}(\vec{x}^*,t^*)...|\Omega\rangle$ in terms of the inner product of states of definite field configuration: $\langle\phi'',t''|T\widehat{\phi}(\vec{x},t)\widehat{\phi}(\vec{x}^*,t^*)...|\phi',t'\rangle$. Then, we will evaluate these inner products using path integrals (ignoring the time-ordered product of field operators). Finally, we will reintroduce the field operators.

Step 1: Field Configuration Eigenstates

The most natural way to write $\langle\Omega|T\widehat{\phi}(\vec{x},t)\widehat{\phi}(\vec{x}^*,t^*)...|\Omega\rangle$ in terms of the inner product of states of definite field configuration would be to expand the ground state in terms of field configuration eigenstates:

$$\langle\Omega|T\widehat{\phi}(\vec{x},t)\widehat{\phi}(\vec{x}^*,t^*)...|\Omega\rangle = \int D\phi''D\phi'\Psi^*_\Omega[\phi'']\langle\phi'',t''|T\widehat{\phi}(\vec{x},t)\widehat{\phi}(\vec{x}^*,t^*)...|\phi',t'\rangle\Psi_\Omega[\phi']\,. \tag{4.86}$$

[37] For a refresher on path integrals in QM with a similar structure to the derivation presented here, see (Srednicki, 2007, ch. 6).

However, the ground state wave functional is complicated (see eq. 4.77) and, in interacting theories, often unknown. So, instead we will use a clever trick to isolate the ground state. We will *assume* that there is a unique minimum energy state $|\Omega\rangle$ and choose the constant Ω_0 in $\widehat{H}$ so that the energy of this state is zero: $\widehat{H}|\Omega\rangle = 0$ (in the free theory, $V_0 = 0$, the ground state is the vacuum state, but in general it won't be). If we replace the Hamiltonian that appears in Schrödinger's equation by $(1 - i\epsilon)\widehat{H}$, then a field configuration eigenstate at t' is given by:

$$|\phi', t'\rangle = e^{i(1-i\epsilon)\widehat{H}t'}|\phi'\rangle\,. \tag{4.87}$$

Here we use the Schrödinger equation to evolve the eigenstate at t' backwards to whatever state it would have been in at $t = 0$. Be careful: $|\phi', t'\rangle$ denotes a particular state at $t = 0$ which, if evolved forward by Schrödinger equation evolution for a time t', would become the field configuration eigenstate $|\phi'\rangle$. We can now introduce a complete set of energy eigenstates[38]:

$$|\phi', t'\rangle = \sum_{k=0}^{\infty} e^{iE_k t'} e^{E_k t' \epsilon} \Psi_k^*[\phi']|k\rangle\,. \tag{4.88}$$

If we then take the limit as $t' \to -\infty$ at constant ϵ, all contributions are annihilated except for that from the ground state.

$$\lim_{t'\to-\infty} |\phi', t'\rangle = \Psi_\Omega^*[\phi']|\Omega\rangle\,. \tag{4.89}$$

Similarly, in the limit as t'' goes to infinity, $\langle\phi'', t''| \to \Psi_\Omega[\phi'']\langle\Omega|$. So, in the limit as

[38]The summation is symbolic. It indicates integration where there is an infinite degeneracy of eigenstates or a continuous spectra of eigenvalues; what is important is that there is a single minimum energy state with energy zero.

the time interval goes to infinity:

$$\lim_{t'\to-\infty,\ t''\to+\infty} \int D\phi' D\phi'' \langle \phi'', t''|T\widehat{\phi}(\vec{x},t)\widehat{\phi}(\vec{x}^*,t^*)...|\phi',t'\rangle = Z\langle\Omega|T\widehat{\phi}(\vec{x},t)\widehat{\phi}(\vec{x}^*,t^*)...|\Omega\rangle\,. \tag{4.90}$$

Where $Z \equiv \int D\phi' D\phi'' \Psi^*_\Omega[\phi']\Psi_\Omega[\phi'']$ is a constant which I *assume* is non-zero and the time evolution of the states on the left-hand-side is determined by $\widehat{H} \to (1-i\epsilon)\widehat{H}$. The right-hand-side is what we want in eq. 4.71 (up to a constant), so we just need to find $\langle \phi'', t''|T\widehat{\phi}(\vec{x},t)\widehat{\phi}(\vec{x}^*,t^*)...|\phi',t'\rangle$.

Before moving on, I should note that there is an *even trickier* way to implement the $1-i\epsilon$ maneuver. If we are working with a Hamiltonian for which the non-zero energy states scale positively with m (which we would expect if they were particle-like excitations), then we can ensure that they are removed when the times are taken to infinity if we change m^2 to $m^2-i\epsilon$ in the Hamiltonian (this is how the trick is typically implemented and explains the origin of $m^2-i\epsilon$ in the Feynman rules). Although the Hamiltonian in eq. 4.65 with, for example, $V_0 = -\lambda\widehat{\phi}^4$ may not have the required spectrum to allow implementing this maneuver, if we were to renormalize the theory by adding counter terms, it would (see Srednicki, 2007, ch. 5).

Step 2: Evaluating the Path Integral

To make things a bit simpler, let's begin by ignoring the time-ordered product of fields and just try to calculate $\langle \phi'', t''|\phi', t'\rangle$ in terms of a path integral. In QM we considered paths through the space of possible trajectories of the system through configuration space. In QFT, we will also consider the possible trajectories of the classical system, but this time we will be considering paths through the space of possible *field* configurations.

Using the unitary time evolution operator that follows from eq. 4.65 (as in eq.

4.87), we can write the inner product as:

$$\langle \phi'', t''|\phi', t'\rangle = \left[\langle \phi''|e^{-i\widehat{H}_J(t''-t')}|\phi'\rangle\right]_{J=0} . \tag{4.91}$$

The Hamiltonian operator is as it appears in the Schrödinger equation except for the infinitesimal imaginary shift in m^2 and the addition of a J-dependent term which does not contribute since J is set to zero (the motivation for the addition of this term will be explained in step 3, but it should be clear that eq. 4.91 & 4.92 are legitimate):

$$\begin{aligned} \widehat{H}_J &= \int d^3x \left[\frac{1}{2}\left(\widehat{\Pi}(\vec{x})\right)^2 + V(\widehat{\phi}(\vec{x}))\right] \\ V(\widehat{\phi}(\vec{x})) &\equiv \frac{1}{2}\left(\nabla\widehat{\phi}(\vec{x})\right)^2 + \frac{1}{2}(m^2 - i\epsilon)\left(\widehat{\phi}(\vec{x})\right)^2 - \Omega_0 + V_0\left(\widehat{\phi}(\vec{x})\right) - J\widehat{\phi}(\vec{x}) . \end{aligned} \tag{4.92}$$

We can break eq. 4.91 into $N+1$ time steps of length δt, inserting a complete set of field-configuration eigenstates after each time step:

$$\langle \phi'', t''|\phi', t'\rangle = \left[\int \prod_{j=1}^{N} D\phi_j \langle \phi''|e^{-i\widehat{H}_J\delta t}|\phi_N\rangle\langle\phi_N|e^{-i\widehat{H}_J\delta t}|\phi_{N-1}\rangle ... \langle\phi_1|e^{-i\widehat{H}_J\delta t}|\phi'\rangle\right]_{J=0} . \tag{4.93}$$

Here we are calculating the amplitude for the field to transition from definite state ϕ' to definite state ϕ'' by looking at contributions from each path through the space of possible field configurations. The integrals in eq. 4.93 are depicted in figure 4.6. Using the Campbell-Baker-Hausdorff formula ($e^{\hat{A}+\hat{B}} = e^{\hat{A}}e^{\hat{B}}e^{-\frac{1}{2}[\hat{A},\hat{B}]+\cdots}$) and throwing out terms of order δt^2, we can write:

$$e^{-i\widehat{H}_J\delta t} = e^{-i\delta t\int d^3x \frac{1}{2}\left(\widehat{\Pi}(\vec{x})\right)^2} e^{-i\delta t\int d^3x V(\widehat{\phi}(\vec{x}))} . \tag{4.94}$$

We can expand a typical term in eq. 4.93 by inserting a complete set of field momenta

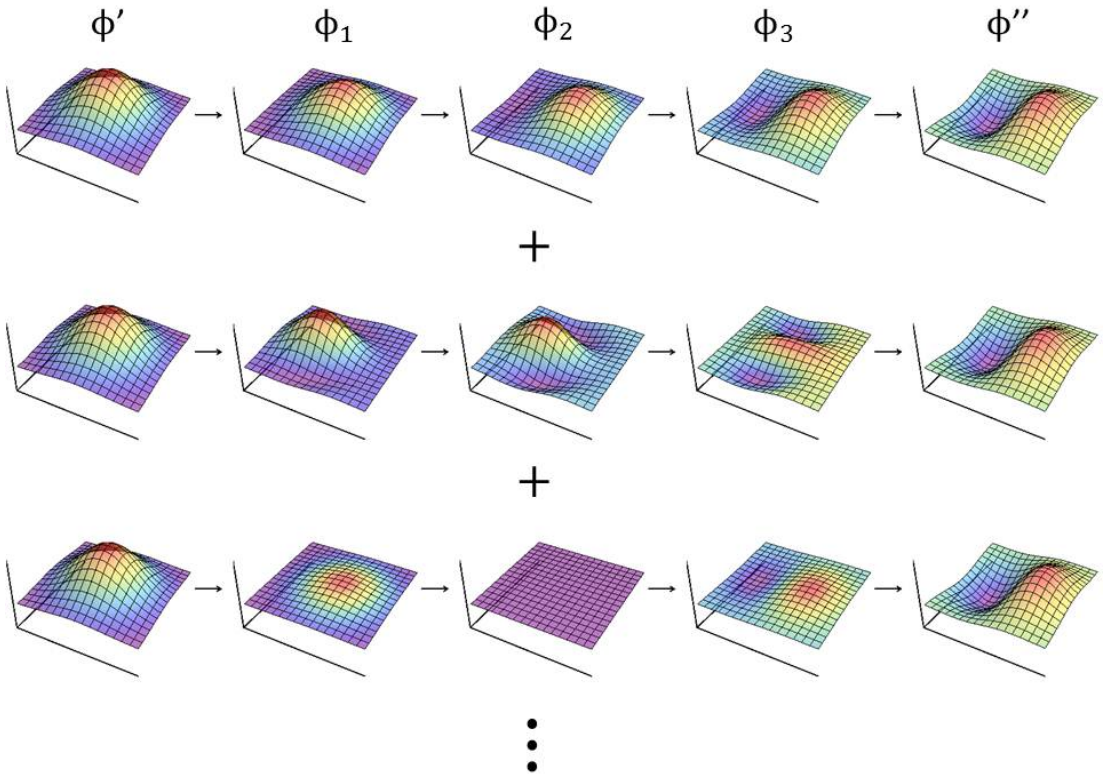

Figure 4.6: **The Path Integral for Fields** Here the summation of amplitudes in eq. 4.93 is depicted for two-dimensional fields for N=3 (4 time steps). The amplitudes for different possible evolutions of the field are added to give the amplitude for a field starting at ϕ' at t' evolving to ϕ'' at t''. Note that the initial and final field configuration is held fixed.

eigenstates:

$$\begin{aligned}\langle\phi_2|e^{-i\widehat{H}_J\delta t}|\phi_1\rangle &= \int D\Pi_1 \langle\phi_2|e^{-i\delta t\int d^3x\frac{1}{2}\left(\widehat{\Pi}(\vec{x})\right)^2}|\Pi_1\rangle\langle\Pi_1|e^{-i\delta t\int d^3x V(\widehat{\phi}(\vec{x}))}|\phi_1\rangle \\ &= \int D\Pi_1 e^{-iH_J(\Pi_1,\phi_1)\delta t}\langle\phi_2|\Pi_1\rangle\langle\Pi_1|\phi_1\rangle\,. \end{aligned} \tag{4.95}$$

To calculate $\langle \Pi_1 | \phi_1 \rangle$, we need an expression for a field-momentum eigenstate:

$$\Psi_{\Pi_1}[\phi] = \frac{1}{(2\pi)^{\mathbf{N}/2}} e^{i \int d^3x \Pi_1(\vec{x})\phi(\vec{x})} \,. \tag{4.96}$$

Here $\mathbf{N}$ is the number of points in space. The choice of normalization comes from the requirement that: $\langle \Pi | \Pi' \rangle = \prod_{\vec{x},\epsilon} [\delta(\phi_{\vec{x}} - \phi'_{\vec{x}})]$ (as in eq. 4.62). So, $\langle \phi_2 | e^{-i\hat{H}\delta t} | \phi_1 \rangle$ becomes:

$$\langle \phi_2 | e^{-i\widehat{H}_J \delta t} | \phi_1 \rangle = \frac{1}{(2\pi)^{\mathbf{N}}} \int D\Pi_1 e^{-iH_J(\Pi_1,\phi_1)\delta t} e^{i \int d^3x \Pi_1(\vec{x})(\phi_2(\vec{x}) - \phi_1(\vec{x}))} \,. \tag{4.97}$$

Now we repeat the analysis in 4.97 for each expectation value in 4.93 and we see that the formula becomes:

$$\langle \phi'', t'' | \phi', t' \rangle = \left[\frac{1}{(2\pi)^{\mathbf{N}(N+1)}} \int \prod_{j=1}^{N} D\phi_j \prod_{i=0}^{N} D\Pi_i e^{-iH_J(\Pi_i,\phi_i)\delta t} e^{i \int d^3x \Pi_i(\phi_{i+1} - \phi_i)} \right]_{J=0} \,. \tag{4.98}$$

Note that here ϕ_0 is ϕ' and ϕ_{N+1} is ϕ'' and that these fields are not integrated over but fixed by the incoming/outgoing field eigenstates (they will be integrated when we plug the expression for $\langle \phi'', t'' | \phi', t' \rangle$ into eq. 4.90). We can then complete the integrals over Π which are all similar to:

$$\begin{aligned} &\int D\Pi_1 e^{-i\delta t \int d^3x \frac{1}{2}(\Pi_1(\vec{x}))^2} e^{i \int d^3x \Pi_1(\vec{x})(\phi_2(\vec{x}) - \phi_1(\vec{x}))} \\ &= \int \prod_{\vec{x},\epsilon} \left[d\Pi_{1\,\vec{x}} e^{-i\delta t \frac{1}{2}\left(\Pi_{1\,\vec{x}}\right)^2} e^{i\Pi_{1\,\vec{x}}(\phi_{2\,\vec{x}} - \phi_{1\,\vec{x}})} \right] . \end{aligned} \tag{4.99}$$

Here we shift to the discrete case to compute the integral. The integral over the field momenta at each point in the space can then be calculated by completing the square,

yielding a contribution of:

$$\left(\frac{2\pi}{i\delta t}\right)^{\mathbf{N}/2} e^{\frac{i}{2\delta t}\int d^3x(\phi_2-\phi_1)^2}\,. \tag{4.100}$$

Incorporating such contributions from each integral over $D\Pi$, eq. 4.98 becomes:

$$\langle\phi'',t''|\phi',t'\rangle = \left[\int\prod_{j=1}^{N}\left[\frac{D\phi_j}{(2\pi i\delta t)^{\mathbf{N}/2}}\right] e^{\frac{i\delta t}{2}\int d^3x\frac{(\phi_{j+1}-\phi_j)^2}{\delta t^2}} e^{-i\delta t\int d^3x V(\phi_j)}\right]_{J=0}\,. \tag{4.101}$$

We can then take the limit as δt approaches zero where the equation becomes:

$$\begin{aligned}\langle\phi'',t''|\phi',t'\rangle &= \left[\int\mathscr{D}\phi\mathrm{Exp}\left[i\int_{t'}^{t''}dt\int d^3x\left(\frac{1}{2}\dot{\phi}^2(\vec{x},t)-V\left(\phi\left(\vec{x},t\right)\right)\right)\right]\right]_{J=0}\\ &= \int\left[\mathscr{D}\phi\mathrm{Exp}\left[i\int_{t'}^{t''}dt\int d^3x\mathscr{L}(\dot{\phi}(\vec{x},t),\phi(\vec{x},t))+J\widehat{\phi}(\vec{x},t)\right]\right]_{J=0}\,.\end{aligned} \tag{4.102}$$

Here we have defined $\mathscr{D}\phi = \prod_{j=1}^{N}\left[\prod_{\vec{x},\epsilon}\left[\frac{d\phi_{j,\vec{x}}}{(2\pi i\delta t)^{1/2}}\right]\right]$ (for discrete spacetime).[39] This equation is now a sum over all possible paths through the space of field configurations.

Step 3: The Time-Ordered Product of Field Operators

Now that we have derived an expression for $\langle\phi'',t''|\phi',t'\rangle$, we would like to generalize to the case where we include the time-ordered product of field operators as in eq. 4.90 (so we can derive eq. 4.71). I will examine the case of a single field operator and leave it to the reader to extend the treatment to multiple operators. Let us calculate (using eq. 4.45):

$$\langle\phi'',t''|\widehat{\phi}(\vec{x},t)|\phi',t'\rangle = \langle\phi''|e^{-i\widehat{H}(t''-t)}\widehat{\phi}(\vec{x})e^{i\widehat{H}(t'-t)}|\phi'\rangle\,. \tag{4.103}$$

[39] This measure is in agreement with that derived in (Hatfield, 1992, eq. 13.13).

Now we have broken the time evolution into two steps. First, we evolve the state from t' to t, then we act with $\widehat{\phi}(\vec{x})$ and evolve the rest of the way to t''. So, eq. 4.93 is modified to:

$$\langle \phi'', t''|\widehat{\phi}(\vec{x},t)|\phi',t'\rangle = \left[\int \prod_{j=1}^{N} D\phi_j \langle \phi''|e^{-i\widehat{H}_J\delta t}|\phi_N\rangle\langle ...\phi_{t+1}|e^{-i\widehat{H}_J\delta t}\widehat{\phi}(\vec{x})|\phi_t\rangle ...\langle \phi_1|e^{-i\widehat{H}_J\delta t}|\phi'\rangle\right]_{J=0}. \tag{4.104}$$

The field operator simply pulls out the value of the field at t and eq. 4.102 becomes:

$$\langle \phi'', t''|\widehat{\phi}(\vec{x},t)|\phi',t'\rangle = \left[\int \mathscr{D}\phi\phi(\vec{x},t)\text{Exp}\left[i\int_{t'}^{t''} ds \int d^3y \mathscr{L}(\dot{\phi}(\vec{y},s),\phi(\vec{y},s))\right]\right]_{J=0}. \tag{4.105}$$

When there are multiple field operators, the time ordering will ensure that we can use the same trick. Next, we will show that taking a functional derivative with respect to J gives the same result. We start with eq. 4.102, isolating the J dependent term:

$$\tfrac{1}{i}\tfrac{\delta}{\delta J(\vec{x},t)}\langle \phi'', t''|\phi',t'\rangle = \left[\int \mathscr{D}\phi\text{Exp}\left[i\int_{t'}^{t''} ds \int d^3y \mathscr{L}(\dot{\phi}(\vec{y},s),\phi(\vec{y},s))\right]\left(\tfrac{1}{i}\tfrac{\delta}{\delta J(\vec{x},t)}e^{J\widehat{\phi}(\vec{y},s)}\right)\right]_{J=0}. \tag{4.106}$$

The functional derivative can be calculated by the chain rule, using the fact that $\frac{\delta f(\vec{y},s)}{\delta f(\vec{x},t)} = \delta^3(\vec{x}-\vec{y})\delta(t-s)$. This gives:

$$\tfrac{1}{i}\tfrac{\delta}{\delta J(\vec{x},t)}\langle \phi'', t''|\phi',t'\rangle = \left[\int \mathscr{D}\phi\phi(\vec{x},t)\text{Exp}\left[i\int_{t'}^{t''} ds \int d^3y \mathscr{L}(\dot{\phi}(\vec{y},s),\phi(\vec{y},s))\right]\right]_{J=0}. \tag{4.107}$$

Since eq. 4.105 and 4.107 are equivalent, we can replace the time-ordered product of field operators with functional derivatives with respect to J. Now we can see why it was useful to add J to the Hamiltonian in eq. 4.92. Allowing for arbitrarily many

operators, the expression becomes:

$$\langle \phi'', t'' | T\widehat{\phi}(\vec{x}, t)\widehat{\phi}(\vec{x}^*, t^*)...|\phi', t'\rangle = \frac{1}{i}\frac{\delta}{\delta J(\vec{x}, t)}\frac{1}{i}\frac{\delta}{\delta J(\vec{x}^*, t^*)}...\langle \phi'', t''|\phi', t'\rangle \,. \quad (4.108)$$

Plugging eq. 4.108 into eq. 4.90 derived at the end of step 1 gives the following expression for the ground state expectation value of the time-ordered product of field operators:

$$\langle \Omega | T\widehat{\phi}(\vec{x}, t)\widehat{\phi}(\vec{x}^*, t^*)...|\Omega\rangle = \tfrac{1}{Z}\lim_{t'\to-\infty,\ t''\to+\infty}\int D\phi' D\phi'' \tfrac{1}{i}\tfrac{\delta}{\delta J(\vec{x},t)}\tfrac{1}{i}\tfrac{\delta}{\delta J(\vec{x}^*,t^*)}...\langle \phi'', t''|\phi', t'\rangle \,. \quad (4.109)$$

Now, we can insert the expression for $\langle \phi'', t''|\phi', t'\rangle$ derived in step 2, eq. 4.102, and taking the limit as the time interval goes to infinity:

$$\langle \Omega | T\widehat{\phi}(\vec{x}, t)\widehat{\phi}(\vec{x}^*, t^*)...|\Omega\rangle = \left[\tfrac{1}{i}\tfrac{\delta}{\delta J(\vec{x},t)}\tfrac{1}{i}\tfrac{\delta}{\delta J(\vec{x}',t')}... \int \mathbf{D}\phi \,\mathrm{Exp}\left[i\int d^4x(\mathscr{L} + J\phi)\right]\right]_{J=0} . \quad (4.110)$$

Here we define $\int \mathbf{D}\phi \equiv \frac{1}{Z}\int D\phi' D\phi'' \mathscr{D}\phi$, which is equivalent to $\frac{(2\pi i\delta t)^{\mathbf{N}}}{Z}\prod_{j=0}^{N+1}\left[\prod_{\vec{x},\epsilon}\left[\frac{d\phi_{j,\vec{x}}}{(2\pi i\delta t)^{1/2}}\right]\right]$ in the discrete case. The $\int \mathscr{D}\phi$ integrated over all paths through the space of field configurations holding the initial and final configurations fixed. $\int \mathbf{D}\phi$ also integrates over the field configuration at the end points.

At this point the derivation is complete. Eq. 4.110 was our goal from section 4.10.

References

Albert, David Z. 1994. *Quantum Mechanics and Experience*. Harvard University Press.

Albert, David Z. 2000. *Time and Chance*. Harvard University Press.

Berndl, Karin, Daumer, Martin, Dürr, Detlef, Goldstein, Sheldon, & Zanghì, Nino. 1995. A Survey on Bohmian Mechanics. *Il Nuovo Cimento B*, **110**(5), 737–750.

Bohm, David, & Hiley, Basil J. 1993. *The Undivided Universe: An Ontological Interpretation of Quantum Theory*. Routledge.

Carroll, Sean M., & Sebens, Charles T. 2014. Many Worlds, the Born Rule, and Self-Locating Uncertainty. *In:* Struppa, Daniele, & Tollaksen, Jeff (eds), *Quantum Theory: A Two-Time Success Story: Yakir Aharonov Festschrift*. Springer.

Deutsch, David. 1999. Quantum Theory of Probability and Decisions. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **455**(1988), 3129–3137.

Dürr, Detlef, & Teufel, Stefan. 2009. *Bohmian Mechanics*. Springer.

Dürr, Detlef, Goldstein, Sheldon, Tumulka, Roderich, & Zanghì, Nino. 2004. Bohmian Mechanics and Quantum Field Theory. *Physical review letters*, **93**(9).

Dürr, Detlef, Goldstein, Sheldon, Tumulka, Roderich, & Zanghì, Nino. 2005. Bell-type Quantum Field Theories. *Journal of Physics A: Mathematical and General*, **38**.

Everett III, Hugh. 1957. "Relative State" Formulation of Quantum Mechanics. *Reviews of Modern Physics*, **29**(3).

Greaves, Hilary. 2004. Understanding Deutsch's Probability in a Deterministic Multiverse. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **35**(3), 423–456.

Greaves, Hilary. 2007. On the Everettian Epistemic Problem. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **38**(1), 120–152.

Hatfield, Brian. 1992. *Quantum Field Theory of Point Particles and Strings*. Vol. 75.

Frontiers in Physics, Addison-Wesley.

Huang, Kerson. 2008. *Quantum Field Theory: From Operators to Path Integrals*. Wiley-VCH.

Peskin, Michael E., & Schroeder, Dan V. 1995. *An Introduction to Quantum Field Theory*. Westview press.

Saunders, Simon, Barrett, Jonathan, Kent, Adrian, & Wallace, David (eds). 2010. *Many Worlds?: Everett, Quantum Theory, & Reality*. Oxford University Press.

Schlosshauer, Maximillian A. 2005. Decoherence, the Measurement Problem, and Interpretations of Quantum Mechanics. *Reviews of Modern Physics*, **76**(4).

Schweber, Silvan S. 1961. *An Introduction to Relativistic Quantum Field Theory*. Row, Peterson, and Company.

Sebens, Charles T., & Carroll, Sean M. forthcoming. Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics. *British Journal for the Philosophy of Science*.

Srednicki, Mark A. 2007. *Quantum Field Theory*. Cambridge University Press.

Struyve, Ward. 2010. Pilot-wave theory and quantum fields. *Reports on Progress in Physics*, **73**, 106001.

Teller, Paul. 1997. *An Interpretive Introduction to Quantum Field Theory*. Princeton University Press.

Tumulka, Roderich. 2006a. Collapse and Relativity. *In:* A. Bassi, D. Dürr, T. Weber, & Zanghì, N. (eds), *Quantum Mechanics: Are there Quantum Jumps? and On the Present Status of Quantum Mechanics, AIP Conference Proceedings 844*. American Institute of Physics.

Tumulka, Roderich. 2006b. On spontaneous wave function collapse and quantum field theory. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, **462**(2070), 1897–1908.

Tumulka, Roderich. 2006c. A relativistic version of the Ghirardi–Rimini–Weber model. *Journal of Statistical Physics*, **125**(4), 821–840.

Tumulka, Roderich. 2007. The 'Unromantic Pictures' of Quantum Theory. *Journal of Physics A: Mathematical and Theoretical*, **40**.

von Neumann, John. 1955. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press.

Wallace, David. 2003. Everett and Structure. *Studies In History and Philosophy*

of Science Part B: Studies In History and Philosophy of Modern Physics, **34**(1), 87–105.

Wallace, David. 2006. In Defence of Naiveté: The Conceptual Status of Lagrangian Quantum Field Theory. *Synthese*, **151**(1), 33–80.

Wallace, David. 2008. The Interpretation of Quantum Mechanics. *In:* Rickles, D. (ed), *The Ashgate Companion to Contemporary Philosophy of Physics*. London: Ashgate.

Wallace, David. 2010. How to Prove the Born Rule. in Saunders *et al.* (2010).

Wallace, David. 2012. *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. Oxford University Press.

Weinberg, Steven. 1995. *The Quantum Theory of Fields, Vol. 1: Foundations*. Cambridge University Press.

Weinberg, Steven. 1999. What is Quantum Field Theory, and What Did We Think It Is? *Pages 241–251 of:* Yu Cao, T. (ed), *Conceptual Foundations of Quantum Field Theory*. Cambridge University Press.

Zee, Anthony. 2010. *Quantum Field Theory in a Nutshell*. Princeton University Press.