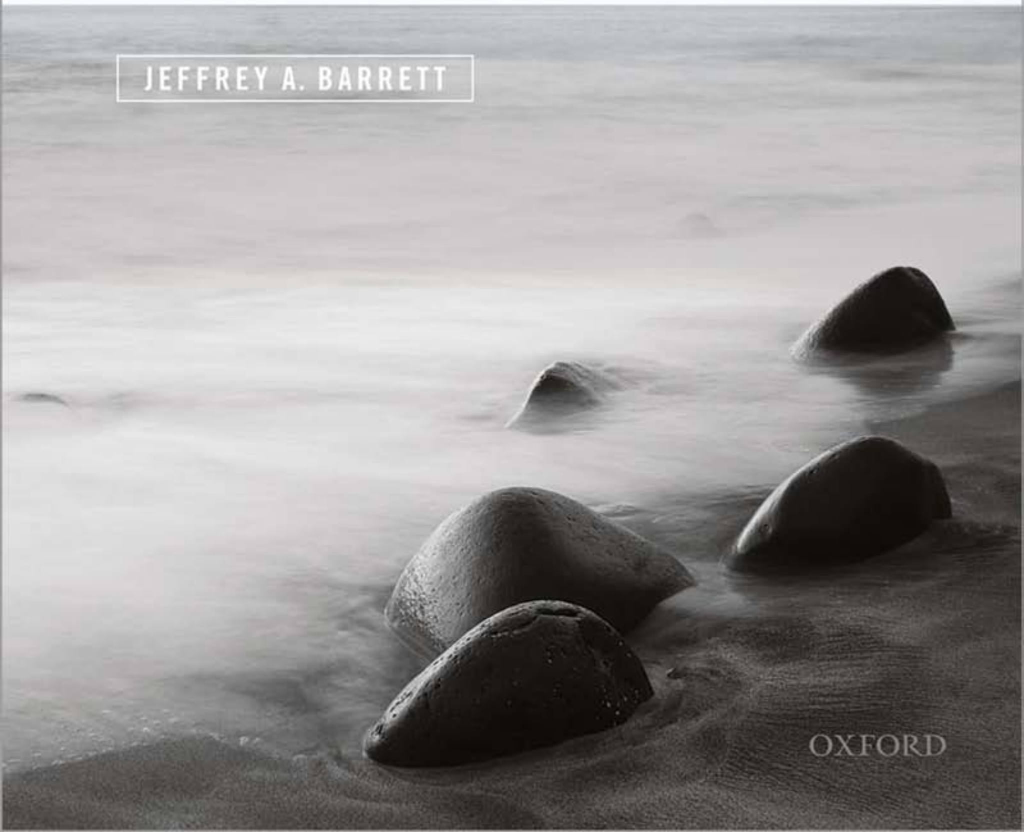
the conceptual foundations of quantum mechanics

JEFFREY A. BARRETT



OXFORD

The Conceptual Foundations of Quantum Mechanics

The Conceptual Foundations of Quantum Mechanics

JEFFREY A. BARRETT

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2019

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2019945690

ISBN 978–0–19–884468–6 (hbk.)
ISBN 978–0–19–884469–3 (pbk.)

Printed and bound in Great Britain by
Clays Ltd, Elcograf S.p.A.

For Martha

Contents

Preface

The aim of this book is to introduce the conceptual foundations of quantum mechanics. To this end, we will start with the empirical evidence that quantum mechanics is meant to explain, develop the standard formulation of quantum mechanics and discuss how it explains the empirical evidence, and then turn to the measurement problem and the various attempts to resolve it.

Since quantum mechanics is written in the language of linear algebra, we will need some basic mathematical notions in order to say precisely what the theory is. With few exceptions, almost everything should be accessible to a reader comfortable with high-school algebra.

This book owes much to the insights of David Bohm, Hugh Everett III, Eugene Wigner, John Bell, and David Albert. Everett and Wigner were the first to characterize the measurement problem in the form that we will discuss it. Bohm and Bell helped to set the stage for what a satisfactory resolution of the measurement problem might look like. And Albert's influence is evident throughout the book.

Theory selection involves a sort of cost–benefit analysis. We start with rough intuitions concerning the nature of the world and a sense of the sorts of predictions and explanations we want from a satisfactory physical theory, then we see how much we can get given the constraints of inquiry. From this perspective, this book might be thought of as a map of the conceptual options and explanatory trade-offs one faces in constructing a theory that explains and predicts quantum phenomena.

It is impossible to evaluate the conceptual foundations of quantum mechanics without reflecting on general philosophical questions concerning what we can know about the physical world and what it means for a physical theory to explain and predict. That the world is so deeply counterintuitive that one cannot trust even one's most cherished intuitions is one of the philosophical lessons of quantum mechanics. Part of the argument here is that there are nevertheless better and worse ways of understanding the quantum world.

The book is written the way I would explain quantum mechanics to a friend. While there is a natural order of sorts, the structure is not linear. This means that we will sometimes briefly discuss something before it is carefully explained. I have put forward-looking references in a number of places to mark this.

This book is the result of many discussions with friends and colleagues and contributions from students. Discussions with David Albert, Tim Maudlin, and David Wallace over the years were particularly salient. I would also like to thank Valia Allori, Thomas Barrett, Jeffrey Bub, Daniel Herrmann, J. B. Manchak, Guillaume Massas, and Lant Pritchett for helpful comments on the book as it was being written; Abel Peña for a close reading and excellent suggestions on an early version of the manuscript; the editors at OUP for their expert guidance; and Martha Tueller Barrett for her support at every turn. Finally, this project benefited from an FQXi mini-grant and from the kind hospitality of the colleagues at the University of Salzburg.

Jeff Barrett
Irvine, CA

6 July 2019

List of Figures

1
Classical Mechanics

1.1 Quantum Mechanics and Philosophical Intuition

Quantum mechanics is arguably the most accurate physical theory we have ever had, but it is deeply counterintuitive. It must be, in order to be empirically adequate. The physical world does not accord well with our commonsense and philosophical intuitions. And quantum phenomena are among the most recalcitrant aspects of our empirical experience.

That even our most cherished intuitions are unreliable as a guide to the physical world has an immediate philosophical consequence. There is no formulation of quantum mechanics that accords with all of our classical intuitions. That said, it is not at all clear which intuitions we should give up. While all formulations of quantum mechanics are counterintuitive, different formulations involve different conceptual sacrifices. We will start with the standard formulation of quantum mechanics, study its properties and the conceptual problems it encounters, and then consider some of the alternatives.

That there are different formulations of quantum mechanics that suggest quite different sets of metaphysical commitments also has philosophical implications. Alternative formulations of quantum mechanics concretely illustrate why it is difficult to judge our current physical theories to be true, probably true, or even probably approximately true. This is something we will return to after we have had a chance to see the radically different ways that alternative formulations of quantum mechanics describe the world.

The standard formulation of quantum mechanics has been used to illustrate the contingent nature of even the most certain philosophical intuitions. The philosopher and logician W. V. Quine famously appealed to the theory to argue that logical truth was empirical. The core of his argument was that one might have to modify even laws that had been taken by philosophers as necessary truths in order to account for the recalcitrant experience described and predicted by quantum mechanics. As Quine put the point, "[r]evision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle?" (1980, 43).

The standard formulation of quantum mechanics does not require one to give up the law of excluded middle at the level of *state attribution*. A given quantum-mechanical state either obtains or not for a particular physical system as usually understood. But on the standard interpretation of states, the law of the excluded middle does not hold for *property attribution*. As a quick example, my favorite bicycle might be in a *superposition* of being in the garage and not in the garage. If so, it is not in fact determinately in the garage and not in fact determinately not in the garage. And it is not partly in and partly out of the garage in any classical sense either. On the standard interpretation of quantum-mechanical states, this is a basic metaphysical issue, not an epistemic issue regarding what I may or may not know regarding the location of the bicycle.

One might say that property attribution in quantum mechanics is geometrical, not logical. Precisely what this means is something we will discuss in detail later, but there is a matter of philosophical methodology to discuss now. When one learns that the standard formulation of quantum mechanics treats properties geometrically rather than logically, one might suppose that we have learned something deep about the metaphysics of the physical world and that our pre-theoretic intuitions concerning the metaphysics of property attribution have been shown to be wrong. While this move of *naturalizing* metaphysics is tempting, to imagine that one can simply read necessary metaphysical truths from quantum mechanics would be to commit a mistake that is as bad as trusting the metaphysical intuitions one found intuitively compelling before reflecting on the nature of our quantum-mechanical experience. The standard formulation of quantum mechanics is certainly one of the best physical theories we have ever had, but any metaphysical conclusions one draws from it should be taken to be thoroughly contingent.

There are several reasons for this. To begin, while the standard formulation of quantum mechanics makes incredibly accurate predictions for sometimes wildly counterintuitive experimental results (on the order of fourteen significant figures for some measurements), there is good reason to believe that the theory is, at best, incomplete and, on a strict reading, logically inconsistent. This is the quantum measurement problem (something we will discuss in detail in Chapter 7). Insofar as the standard theory is at best incomplete and arguably inconsistent, it is a poor candidate for a reliable guide to metaphysical truth. Second, there are many mutually incompatible alternative formulations of quantum mechanics. These theories make empirical predictions that are similar to those made by the standard formulation of quantum mechanics for the experiments we have performed so far, but they explain the results of these experiments in often strikingly different ways. In some cases they disagree regarding the results of future experiments. Third, even if one had a particular favorite formulation of quantum mechanics, one would still have a number of options regarding precisely how to use it to provide explanations, and such alternative explanations often suggest quite different metaphysical commitments. Finally, the history of quantum mechanics provides ample evidence for the ways in which the particular formulations of even our most fundamental physical theories are contingent and hence unsuitable as a source of necessary truths regarding the basic nature of the world.[1]

It is important to be clear about this. The reason we know that the standard formulation of quantum mechanics is wrong is not pessimistic induction over the descriptive failures of our past theories. Rather, the standard theory is either incomplete in an empirically significant way on the most charitable reading or logically inconsistent on an uncharitable but natural reading. Further, its basic dynamical structure fails to mesh well with special relativity, the other cornerstone of modern physics.

The measurement problem is not that quantum mechanics is counterintuitive. It must be counterintuitive in order to make the right empirical predictions. The measurement problem, rather, is that one can describe an experiment in the language of quantum mechanics such that on plausible background assumptions the standard theory predicts logically inconsistent results. One can avoid the inconsistency by insisting that the standard theory is incomplete because it does not say which result in fact occurs. But then one has the task of fixing the theory.

There are a number of very different ways to address the quantum measurement problem. Each involves addressing the standard theory's potential inconsistency while retaining its successful empirical predictions.

A brief example may be helpful. Bohmian mechanics is one of the options for how one might revise the standard formulation of quantum mechanics to resolve the quantum measurement problem.[2] We will discuss this theory in detail in Chapter 11. Bohm's theory does not treat property attribution in the same way that the standard formulation of quantum mechanics does. Here, when the quantum state of my bike is a superposition of it being in the garage and it not being in the garage, the bike really is either determinately in the garage or determinately not in the garage. While the law of the excluded middle applies to position attribution on this formulation of quantum mechanics, this does not mean that one has rescued a classical set of philosophical intuitions from quantum mechanics. Position is treated in a special, privileged way in Bohmian mechanics. This means that most of what one classically takes as intrinsic observable physical properties (e.g. velocity, kinetic energy, etc.) are not really physical properties at all. Rather, one might think of them as the result of an elaborate illusion predicted by the theory. Hence while Bohmian mechanics saves some classical intuitions, it violates others.

Bohmian mechanics is not the only theoretical alternative to the standard formulation of quantum mechanics. But even a firm commitment to Bohmian mechanics would not provide a definitive metaphysics. This is because there

[1] See Cushing (1994) for an extended discussion of the historical contingency of quantum mechanics.

[2] This formulation of quantum mechanics was first proposed by David Bohm (1952) in response to the quantum measurement problem.

are always a number of quite different ways to understand the metaphysical commitments associated with any particular formulation of quantum mechanics.[3]

The upshot is that while quantum mechanics illustrates how our philosophical intuitions may mislead us, it does not deliver a ready-made description of the world to replace those intuitions. Rather, it provides a number of quite different, and not entirely satisfactory, metaphysical options.

Before we consider the conceptual problems that the standard formulation of quantum mechanics encounters and how these problems might be addressed, we need to understand how the theory works. A good place to start is by considering where our classical intuitions go wrong.

1.2 Newtonian Mechanics

Classical mechanics was proposed by Isaac Newton (1687) in his *Philosophiæ Naturalis Principia Mathematica* as a unified account of celestial and terrestrial motion. It describes the physical world in terms of the motions of its constituent objects. One might take particles to be the simplest sort of object.[4] A particle has a position $\boldsymbol{x}$, a velocity $\boldsymbol{v}$, and a mass m, and it is acted on by various forces $\boldsymbol{F}$.[5] The mass of a particle and its initial position and velocity and the forces acting on it determine how it moves.

Newton (1687, 83) presented the basic principles of his theory as three axioms, or laws, of motion.

Law I. Every body perseveres in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by forces impressed thereon. [That is, an object moves at a constant velocity unless it is acted on by a force.]

Law II. The alteration of motion is ever proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed. [That is

$$\boldsymbol{F} = m\boldsymbol{a} = m(d\boldsymbol{v}/dt) = d\boldsymbol{p}/dt, \tag{1.1}$$

[3] In the case of Bohmian mechanics, among one's options are how one understands configuration space, the dynamical laws of the theory, and which observables one chooses as determinate.

[4] For his part, Newton did not simply assume that the basic constituents of the world were particles. He also provided an extensive development and presentation of continuum mechanics in the *Principia*. In the continuum mechanics, the basic stuff of the physical world is modeled as smooth and uniform substance that need not be thought of as an approximation for a large number of particles. Continuum mechanics avoids some of the conceptual problems one encounters with particle mechanics. It does not, however, solve the problems that led to quantum mechanics. See Mark Wilson (2015) for a discussion of classical continuum mechanics. For now we will suppose that the basic constituents of the world are particles.

[5] We will use bold type to represent classical vector quantities like position, velocity, and momentum. We will have more to say about vectors in Chapter 4.

where $\boldsymbol{F}$ is the force impressed, m is the object's mass, $\boldsymbol{a}$ the object's acceleration, $\boldsymbol{v}$ the object's velocity, and $\boldsymbol{p}$ the object's momentum.]

Law III. To every action there is always opposed an equal reaction: or the mutual actions of two bodies upon each other are always equal, and directed to contrary parts. [That is

$$\boldsymbol{F}_{ab} = -\boldsymbol{F}_{ba}.] \tag{1.2}$$

Law II is the theory's dynamics. It tells us how the velocity of an object changes as forces are applied to it. The object's velocity, in turn, tells us how its position changes.

Given a plausible set of assumptions, classical mechanics is a deterministic theory where how each particle moves is fully determined by the forces imposed on it.[6] The current state of a physical system is given by the position, velocity, and mass of each of its constituent objects. Given the state at any time together with the forces acting between the objects, Law II fully determines the state at all future and past times.

In this sense, classical mechanics describes a clockwork universe where everything that happens is determined by the initial state of the fundamental physical objects and the forces that act between the objects. However, it remains to say what forces there are and how the forces themselves behave.

Newton's account of gravitational interactions provided an example of a force law. In modern notation, the gravitational force between two objects a and b is given by the law of universal gravitation

$$\boldsymbol{F}_{ab} = \frac{-Gm_a m_b}{\boldsymbol{r}_{ab}^2}, \tag{1.3}$$

where $\boldsymbol{F}_{ab}$ is the gravitational force on a, G is the gravitational constant, m_a and m_b are the masses of a and b respectively, and $\boldsymbol{r}_{ab}$ is the directed distance between the two objects.

If one supposes that gravitation is the only force, then the theory describes a physical world where objects move along continuous well-defined trajectories determined by the mutual attraction of the gravitational forces acting between them. If the other forces, whatever those may be, could be similarly described, it was thought that one would have a complete description and explanation of the behavior of the physical world.

[6] If the forces imposed on the objects are unbounded or not always well defined, then all bets are off, and the theory may not be deterministic. See Earman (1986) for a discussion of this and closely related issues. If a system does not satisfy the required conditions, then it might not evolve in a deterministic way, but, even then, the sort of indeterministic behavior that might be exhibited by a classical system is believed to be fundamentally different from the sort indeterministic behavior predicted by the standard formulation of quantum mechanics.

As an indicator of the reception of this theory, Edmund Halley, the English astronomer, mathematician, and physicist, was inspired to write an ode to Newton's achievement. The ode served as an introduction to Newton's (1687) *Principia*. Among other things, the ode gives Halley's assessment of the cognitive status of the Newton's theory:

> Lo, for your gaze, the pattern of the skies!
> What balance of the mass, what reckonings
> Divine! Here ponder too the Laws which God,
> Framing the Universe, set not aside
> But made the fixed foundations of his work.

These were the laws that God himself set not aside in the creation of the world. Not only did Newton have the physics right, but he had read the mind of God.

It was clear to everyone, including Newton himself, that the theory as described in the *Principia* was incomplete. Perhaps most saliently there were forces other than gravitation. To get a complete physical theory, one would need to characterize these with similar precision. But what remained to be done looked to many like just a bit of tidying up. Having the dynamics in hand that had been divinely established to govern the motion of all physical objects provided a framework for filling in the rest of the details.

Laws for other forces were subsequently added to the theory. Looking much like Newton's law of universal gravitation, Coulomb's law gives the electric force between charged objects:

$$\boldsymbol{F}_{ab} = \frac{kq_a q_b}{\boldsymbol{r}_{ab}^2}, \tag{1.4}$$

where $\boldsymbol{F}_{ab}$ is the electric force on a, k is Coulomb's constant, q_a and q_b are the charges of a and b respectively, and $\boldsymbol{r}_{ab}$ is the directed distance between the two objects.

But here the tidying up required some subtlety. It turned out that the electric force was dynamically coupled to another force—the magnetic force. What was needed was a dynamical account of how these forces were mediated and how their behavior was interrelated. To this end, the electric and magnetic forces were characterized as being mediated by the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$ respectively, and laws were proposed to describe how these fields interacted with charged material objects and how the fields themselves interacted and evolved. The result was four laws:

$$\nabla \cdot \boldsymbol{E} = 4\pi\rho$$
$$\nabla \cdot \boldsymbol{B} = 0$$

$$\nabla \times \boldsymbol{E} = -\frac{1}{c}\frac{\partial \boldsymbol{B}}{\partial t}$$

$$\nabla \times \boldsymbol{B} = \frac{1}{c}\left(4\pi \boldsymbol{J} + \frac{\partial \boldsymbol{E}}{\partial t}\right),$$

where ρ is the total electric charge density, c is the speed of light, and $\boldsymbol{J}$ is the total current density.[7] In order, these are Gauss's law, Gauss's law for magnetism, the Maxwell–Faraday equation, and Ampère's circuital law. These laws describe how electric and magnetic fields evolve and how they interact with matter and each other. They are known collectively as Maxwell's equations.

Two consequences of Maxwell's equations will prove particularly important. First, just as Newton's second law describes material objects moving in a deterministic way that depends on forces, Maxwell's equations describe electric and magnetic fields evolving in a deterministic way that depends on the motions of material objects and fields. This maintains the picture of a clockwork universe, just one that includes fields in addition to material objects. Second, Maxwell's equations predict that an accelerating charge will radiate electromagnetic waves that will travel at the speed of light c. An accelerating charge, then, loses energy. We will return to this in a moment.

With particles as fundamental material objects, we are left with the following classical picture. Particles always have determinate positions, velocities, and masses. The positions and velocities of the particles change in a deterministic way as forces are imposed on them, and the particles follow determinate trajectories over time. The motions of the particles and the fields determine the motions of the fields, which, in turn, affect the forces on, and hence the motions of, the particles. This picture is more complex than Newton's first pass, but it is still a thoroughly clockwork world under a plausible set of background assumptions.

There remained some tidying to do, but up to the late nineteenth century it looked like classical mechanics was on track to provide a complete and unified account of the physical world. And the overall picture was relatively intuitive. There are fundamental physical objects, those objects have a set of basic physical properties, and those properties evolve over time according to regular laws. Things move when they are pushed, how fast they move depends on how hard they are pushed, and when they move they follow determinate and well-defined trajectories.

Together, the laws of classical mechanics allowed for detailed predictions and explanations over a broad range of phenomena. Indeed, it was the predictive strength of the theory that ultimately showed that it could not be right.

[7] These equations involve a bit of vector analysis. This is one of the relatively few occasions in the book where we will not discuss the mathematics in detail. What matters to us most here is the two consequences of Maxwell's equations described below.

1.3 Problems with Classical Mechanics

Quine's belief that even the laws of logic were empirical followed from his understanding of how our commitments change as we learn from empirical experience.

> Any statement can be held true come what may, if we make drastic enough changes elsewhere in the system. . . . A recalcitrant experience can . . . be accommodated by any of various alternative reevaluations in various alternative quarters of the total system. (1980, 43–4)

In this he was certainly right, both logically and historically.

No particular empirical problem was by itself fatal to classical mechanics. Among the moves available to someone who wants to save classical mechanics from recalcitrant experience are the introduction of new forces. On the face of it, one just needs to postulate whatever forces it takes to get the right motions at the right times.

Even so, empirical problems began to pile up. By the early twentieth century it was clear that these problems were interrelated and pointed to something deeply counterintuitive about the behavior of the physical world. We will briefly consider two empirical problems that were historically important, then consider two closely related problems where the clash between experience and our pre-quantum intuitions was particularly clear.

One of the central problems with classical mechanics arose in trying to account for black-body radiation. Black-body radiation is the electromagnetic radiation (including light) that is emitted from a perfectly opaque and non-reflective object by dint of its temperature alone. To get an idea of the phenomenon, imagine a hollow sphere of non-reflective metal with a small hole punctured in one side. In this setup the hole acts as a black body. Light going into the hole from outside the sphere does not reflect back out. Hence the hole looks black at room temperature. But as the sphere is heated, the hole will begin to glow dull red then orange then bright yellow, perhaps eventually becoming bright blue if the material can withstand the heat. While there is typically a dominant color to the light emitted at a particular temperature, the full range of light emitted is always a mixture of colors. This mixture is the frequency distribution of light from the black body.

The problem is that classical mechanics makes entirely the wrong predictions for this distribution. Indeed, on plausible background assumptions, classical statistical mechanics predicts that a black body should emit a beam of light with infinite energy no matter what the positive temperature. If one were to build a metal sphere with a hole in the side, one might imagine a death-ray emanating from the hole destroying the universe. But nothing even approximately like that happens.

The physicist Max Planck got the right predictions for black-body radiation in 1900, but to do so he made the implausible assumption that energy is always exchanged in discrete packets, or quanta.[8] What makes this implausible is that classical mechanics places no constraints whatsoever on the quantity of energy that might be exchanged between objects in an interaction.

The same implausible assumption helped to explain another puzzle. When light strikes a metal surface, negatively charged particles, later identified as electrons, were sometimes found to be emitted from the surface. But, contrary to what one would expect from classical mechanics, the number of electrons emitted was not a simple function of the energy density of the light. Very bright red light might do nothing, while even the faintest blue light might produce a shower of electrons. This is the photoelectric effect. The effect is inexplicable if light is electromagnetic radiation whose interaction with matter is fully characterized by Maxwell's equations. To be sure, electromagnetic radiation should affect charged particles, but the magnitude of the effect should depend on the intensity of the radiation, not its frequency.

Albert Einstein explained the photoelectric effect in 1905 by assuming that energy can only be transmitted in discrete quanta and that the energy associated with each photon or quantum of light is given by hv, where h is Planck's constant and v is the frequency of the light.[9] This explanation, for which Einstein won his Nobel Prize in Physics, is flatly incompatible with the view that light consists in electromagnetic waves as described by Maxwell's equations.

There was ample evidence from the early nineteenth century on for the wave-like nature of light. As a form of electromagnetic radiation, it propagates, disperses, diffracts, and interferes, as one would expect of a wave. Light's wave-like behavior was well-described by Maxwell's equations. But there was also increasing evidence for the particle-like nature of light. When low-intensity light strikes phosphorescent or photographic material, it appears to strike it at very precise, discrete locations rather than hitting it over a broad smudged-out region as a wave would. One sees sharply localized flashes on a phosphorescent screen or single black dots on a photographic plate.

The classical two-slit experiment illustrates the dual nature of light. Consider a light source (a laser, for example), a screen with two parallel slits cut in it, and a plain screen beyond the slits (as in Figure 1.1A). If the slits are sufficiently close to each other, a pattern of bright and dark lines can be seen on the screen beyond the slits. The graph on the right of the figure indicates the relative brightness of the light hitting the screen in each region. This sort of interference pattern is characteristic of wave phenomena. It is as if the light waves incident on the barrier produce

[8] Planck (1900) and see Pais (1982, 364–72) for a discussion putting this into historical context.
[9] See Pais (1982, 378–86).

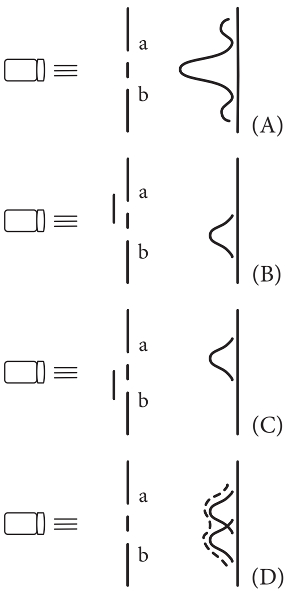


Figure 1.1. Two-slit interference patterns.

two light sources, one at each slit, that constructively and destructively interfere with each other before hitting the screen. Where there is constructive interference, the waves add and there is a bright line on the screen. Where there is destructive interference, the waves cancel and there is a dark line on the screen.

But other details of the experiment seem flatly incompatible with a wave model of light. At low intensities, one can observe individual flashes that suggest *particles of light* hitting the screen at precise, but largely random, locations. The bright lines in the interference pattern are just where more photons hit and the dark lines are where fewer hit. These particles came to be called *photons*.

That said, a particle model of light also appears untenable since particles passing through slits should not exhibit an interference pattern at all. By lowering the intensity of the light sufficiently, one can ensure that there is only one photon traveling through the device at a time. Suppose each such photon follows a trajectory that either passes through slit a or through slit b. If a photon passes through slit a, it should hit the screen directly behind slit a, and if it passes through slit b it should hit the screen directly behind slit slit b. And, indeed, when one blocks one or the other of the two slits, one observes the photons that still make it through simply hitting the screen behind the unblocked slit (as in Figures 1.1B and C).

Since we are sending the photons through the device one at a time, if they can be thought of as classical particles at all, one should see the sum of the two one-slit patterns when both slits are open (as in Figure 1.1D). But that is not what happens. When both slits are open the photons behave like waves and produce an interference pattern (as in Figure 1.1A) *even when the light is so dim that there* *is only one particle moving through the device at a time*. If one thinks of each photon as passing through a particular slit, then one must suppose that the particle somehow detected whether the other slit was blocked. It then moves straight back to the screen if the other slit is blocked but moves in a puzzling way that eventually produces the distinctive wave-like interference pattern if the other slit is unblocked.

All of this represents a fundamental explanatory failure of classical mechanics. If one supposes that light is wave-like, classical mechanics can explain the interference pattern, but cannot explain finding photons at particular locations. If one supposes that light is particle-like, it can explain finding photons at particular locations, but cannot explain how they got there.

In order to explain how light behaves we need to know whether to think of it as a wave, a particle, or something else. And, however we end up thinking of it, we then need consistent dynamical laws that characterize its behavior.

Since each of the problems we have considered so far involves the nature and behavior of light, one might imagine that classical mechanics does just fine predicting the behavior of ordinary material objects like books or tables or rocks. But this is not the case. Indeed, classical mechanics predicts that such ordinary material objects—things like baseballs, trees, and such—are impossible.

1.4 The Stability of Matter

By the end of the nineteenth century there was good reason to believe that matter was composed, in part, of relatively light, negatively charged particles. It was these particles that were emitted in cathode ray tubes and released from a metal surface by light in the photoelectric effect. These particles were called electrons.

But the very existence of electrons posed an immediate classical problem for the stability of matter. Since electrons are negatively charged, they will, by Coulomb's law, repel each other, so there is no stable configuration of particles consisting of just electrons. There must also be something that is positively charged to keep the electrons from flying off and to cancel the negative charge on the electrons and explain how an object might be electrically neutral.

The British physicist J. J. Thomson's solution was to suppose that the atoms that made up the fundamental elements like hydrogen, carbon, and potassium consisted of a number of negatively charged electrons "enclosed in a sphere of uniform positive electrification" (1904). There was just enough of this positive electrification in each atom to balance the negative charge of the electrons and render the atom electrically neutral (as in Figure 1.2C). Thomson's plum-pudding model had many virtues. Among these was that the electrons in an atom were simply embedded in a positive-charged pudding and did not move, a point we will return to in a moment.

The problem with Thomson's model was that it did not account well for subsequent empirical data. In a series of experiments conducted between 1908 and 1913, Ernest Rutherford, Hans Geiger, and Ernest Marsden sent beams of positively charged alpha particles, each of which thousands of times heavier than an electron (and which were later found to be composed of two positively charged protons and two neutral neutrons), at thin metal foils. Most of the alpha particles passed straight through the foil. But they found that a few of the alpha particles were deflected by large angles. Indeed, sometimes they were bounced almost straight back.[10] If Thomson's model were correct, the positively charged alpha particles would not *bounce* off the much lighter electrons or the positively charged pudding. That they did bounce meant that there were small positively charged objects in the atom that were much heavier than the alpha particles themselves.

So Thomson's plum-pudding model of the atom was wrong. Rather than a positively charged pudding with negatively charged electrons embedded in it, there must be heavy, positively charged nuclei with much lighter negatively charged electrons flying around them. But this was a model that Thomson had already ruled out when he was thinking about the problem, and for good reason.

Consider how a negatively charged electron would move about a heavier, positively charged nucleus. The problem is not that classical mechanics fails to make predictions—it is that the predictions it makes are manifestly false.

Suppose that the electron is not moving (as in Figure 1.2A). Since the electron and the nucleus are oppositely charged and the electron is much lighter, the electron will accelerate and move toward the nucleus. As an accelerating charge, Maxwell's equations require the electron to emit electromagnetic radiation as it falls into the nucleus. So the atom will quickly fall in on itself in a burst of electromagnetic radiation.

So suppose that the electron is moving (as in Figure 1.2B). Since they are oppositely charged, the electron will feel the attractive force of the nucleus and accelerate toward it. Given that the electron is initially moving, it may begin to orbit the heavier nucleus, but to orbit it must change its direction of motion and hence accelerate, and as an accelerating charge, Maxwell's equations again require the electron to emit electromagnetic radiation. This radiation will carry energy away from the electron. If the radiant energy comes from the electron's potential energy, then the electron will move closer to the nucleus. If it comes from the electron's kinetic energy, then the electron will move closer to the nucleus. Either way, the electron spirals toward the nucleus emitting radiation as if falls. So the atom will quickly fall in on itself in a burst of electromagnetic radiation.

While one might imagine that there might be *some* stable configuration of positive and negative charges where the charges are precisely positioned so that

[10] See Geiger (1908) and Geiger and Marsden (1913) for descriptions of these experiments.

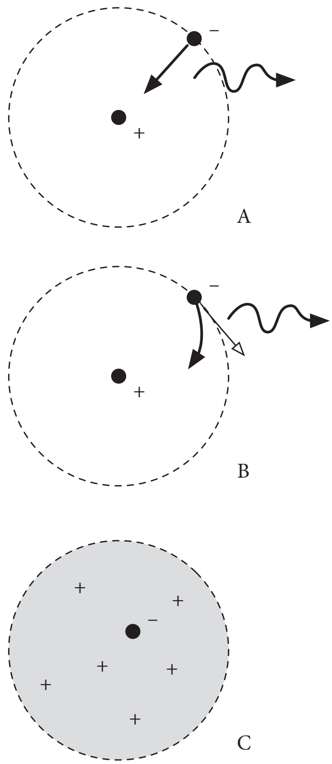


Figure 1.2. Moving electrons and Thomson's model.

their forces cancel, it was proven by the British mathematician Samuel Earnshaw (1842) that there is no stable stationary equilibrium for a collection of point charges. And if the point charges are moving, then they will either leave the region that is supposed to be occupied by the object or they will have to accelerate to stay in the region. If they accelerate, then they will radiate, so there can be no stable moving configuration either. Hence ordinary material objects like books or tables or rocks cannot be stable under the dynamical laws of classical mechanics if they are made of charged particles. And they are.

The upshot is that classical mechanics is flatly incompatible with the very existence of the books, chairs, rocks, and other objects that make up the bulk of our everyday experience. If there ever were such an object, it would immediately self-destruct in a blinding flash of radiation. This is not an issue of fine-tuning the theory. Classical mechanics makes perfectly clear predictions—it is just that its predictions are utterly wrong.

Since the electrons in ordinary material objects cannot be stationary and they cannot be moving, a natural place to start in constructing a new physical theory would be to say what these electrons are in fact *doing*. But this turned out to be so difficult that the Danish physicist Niels Bohr, one of the central figures in the early development of quantum mechanics, concluded that

> There is no quantum world. There is only an abstract quantum physical description. It is wrong to think that the task of physics is to find out how Nature *is*. Physics concerns what we can *say* about Nature.
> [As reported by his colleague Aage Petersen (1963, 12), emphasis in original.]

More specifically, in the case of the electrons orbiting the nucleus of an atom, Bohr believed that it was *conceptually impossible* to describe their motion.

The Irish physicist John Bell much later described the reaction of Bohr's generation this way:

> The physicists who first came upon such phenomena found them so bizarre that they despaired of describing them in terms of ordinary concepts like space and time, position and velocity. The founding fathers of quantum theory decided even that no concepts could possibly be found which could permit direct description of the quantum world. (1987, 170)

One can understand this reaction given the mounting empirical evidence that that matter behaves in a deeply nonclassical, counterintuitive way. But, given human nature as we find it, it is hard to resist the temptation of at least *trying* to describe the world and explain our experience.

By the early 1930s there was a physical theory that provided dynamical explanations for the behavior of quantum-mechanical systems like electrons. But before describing that theory, we should first consider more systematically the empirical evidence that it needs to explain. We will start by considering the behavior of electrons in more detail.[11]

[11] See Becker (2018) for a spirited introduction to the conceptual history of quantum mechanics.

2
Quantum Phenomena

2.1 Spin Properties

We will begin by considering the behavior of electrons in a number of simple, idealized experiments. This behavior illustrates the most basic sort of empirical evidence that quantum mechanics will have to cover in order to be empirically adequate.[1]

Electrons exhibit *spin* properties.[2] The first experiments to observe spin properties were performed by Otto Stern and Walther Gerlach (1922a, 1922b, 1922c) when they set out to measure the magnetic moment of silver atoms by passing the atoms through an inhomogeneous magnetic field. Classical mechanics predicts that a beam of silver atoms will simply spread out in the direction of the field gradient. But Stern and Gerlach found that the field split the initial beam of silver atoms into two sharply defined beams (as in the bottom of Figure 2.1). And it split the beam into precisely two beams regardless of the orientation of the field.

We will refer to the different directions that the inhomogeneous field might be oriented by as the coordinate axes x, y, z, and various angles relative to these axes. Silver atoms that are deflected one way when the field is oriented in the x-direction are identified as x-spin up atoms and those deflected the other way are identified as x-spin down atoms. If the field is oriented in the z-direction, one beam is identified as the z-spin up atoms and the other as the z-spin down atoms. And so on for the other directions the field might be oriented.

Individual electrons behave similarly. We will consider these properties in the context of idealized boxes like those described by David Albert (1992) (as in Figure 2.2). There are three types of *Albert box*. An x-spin box has three doors. One labeled "in," one labeled "$\uparrow_x$," and one labeled "$\downarrow_x$." A z-spin box also has three doors. One labeled "in," one labeled "$\uparrow_z$," and one labeled "$\downarrow_z$." The x-spin and z-spin boxes function much like idealized Stern–Gerlach devices. In particular, they correlate an electron's spin properties with its position. The third type of box

[1] As we will see later, quantum mechanics must do significantly more than cover these experiments in order to be empirically adequate. Among the quantum phenomena that a fully satisfactory theory must be able to explain are pair production and annihilation, scattering, EPR-type correlations, and the curious behavior of neutral K mesons. But this is a good place to start.

[2] Many particles exhibit intrinsic angular momentum and hence spin properties. Electrons are leptons and, like all leptons, are spin-1/2 particles. Photons, light quanta, are spin-1 particles.

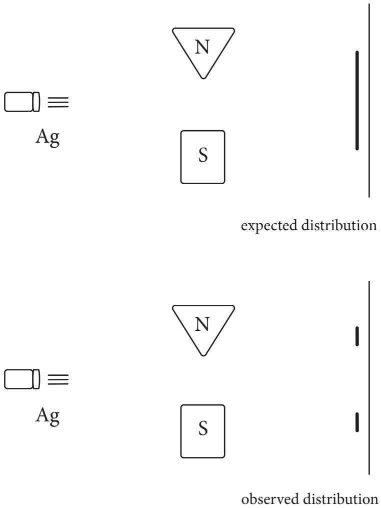


Figure 2.1. Expected vs observed Stern–Gerlach distributions.

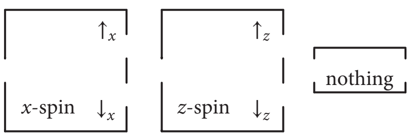


Figure 2.2. Albert boxes.

is what Albert calls a total-of-nothing box. It has two doors. One labeled "in" on the left and one labeled "out" on the right. We will start by considering the x-spin and z-spin boxes then discuss the total-of-nothing box later.[3]

In the following experiments we will suppose that we are supplied with a collection of electrons that have not been prepared in any special way and that we can send them through the boxes one at a time. We will keep track of what happens when we do so. The predictions of an empirically adequate formulation of quantum mechanics must agree with the results of these experiments. If one's formulation of quantum mechanics also allows one to follow the dynamical evolution of the states of the electrons as they move through the boxes, then one will be able to explain how the electrons are behaving to produce the observed phenomena.

[3] One might think of the Albert x-spin and z-spin boxes as idealized Stern–Gerlach devices that can handle individual charged particles like electrons. The behavior of an electron in a classical Stern–Gerlach device is complicated by the fact that, unlike neutral silver atoms, its trajectory will also be significantly affected by its charge. Albert boxes provide a helpful idealization. We will use similar idealizations when they allow us to characterize basic empirical or theoretical concepts more clearly. The thought is that one can add the complicating details back in once one understands the basic conceptual foundations.

2.2 Basic Properties of the Boxes

If one sends an electron that has not been prepared in any special way into an x-spin box, it exits from the $\uparrow_x$ door about half of the time and from the $\downarrow_x$ door about half of the time (as in Figure 2.3).

If one takes an electron that exited from the $\uparrow_x$ door and sends it into a second x-spin box, then it will exit from the $\uparrow_x$ door of that box as well, as long as it is not disturbed as it travels between the boxes. Similarly, an electron that exits from the $\downarrow_x$ door of the first box will exit from the $\downarrow_x$ door of the second box as long as it is not significantly disturbed as it travels between the boxes. If one is careful, the same behavior can be repeated indefinitely for a particular electron. But if the electron is knocked around between the boxes, then all bets are off—the door it exits from on the second x-spin box may well not agree with the door it exited from on the first. And the more it is knocked around between boxes, the lower the probability that the initial outcome will be repeated.

Similarly, if one sends an unprepared electron into a z-spin box, it exits from the $\uparrow_z$ door about half of the time and from the $\downarrow_z$ door about half of the time. And, if it is not disturbed, it will do the same thing it did with the first z-spin box if it subsequently encounters other z-spin boxes.

One might naturally conclude that electrons have the property of being either x-spin up or x-spin down and the property of being either z-spin up or z-spin down, and that the two types of box *sort* electrons accordingly. More specifically, the fact that careful repeated measurements yield the same result suggests that the

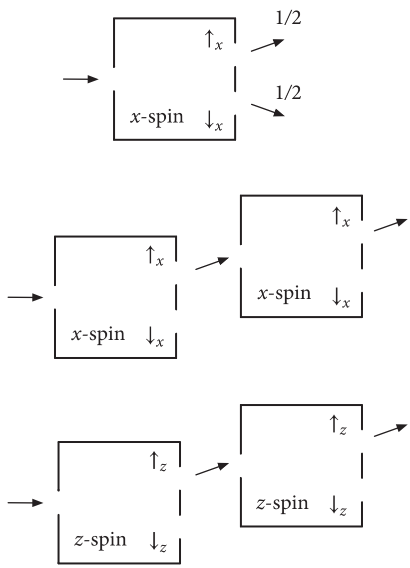


Figure 2.3. Basic behavior of spin boxes.

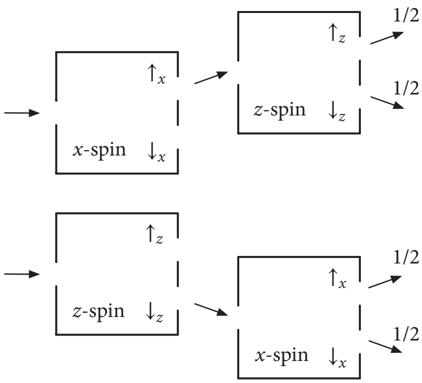


Figure 2.4. The empirical relationship between x-spin and z-spin.

electrons possess determinate spin properties. And the fact that the spin properties of an electron change if it is knocked around indicates that these properties are not immutable. But nothing so far has ruled out x-spin and z-spin being the *same* physical property. It could be that the two types of box are really the same, just differently labeled.

Consider the empirical relationship between x-spin and z-spin. Suppose one sends an electron into an x-spin box and it comes out the $\uparrow_x$ door. If the electron is then sent into a z-spin box, it exits from the $\uparrow_z$ door about half of the time and from the $\downarrow_z$ door about half of the time (as in Figure 2.4). Further, there is no predictable pattern whatsoever to which z-spin door it exits from. It appears then that the z-spin result is *randomly* determined. And the same is true for electrons exiting the first box from the $\downarrow_x$ door.

Similarly, if one knows the z-spin of an electron, this tells one nothing whatsoever about its x-spin. A z-spin up electron is just as likely to be found to be x-spin up as it is to be found x-spin down. And the same is true for a z-spin down electron.

Since the outcomes of x-spin and z-spin observations are completely uncorrelated, one might naturally conclude that the two spin quantities are independent. Further, one might imagine that it is possible to know the values of both quantities at the same time. But this is where one encounters another distinctively quantum-mechanical phenomenon.

Suppose one sends an electron into an x-spin box, then into a z-spin box, then into a second x-spin box (as in Figure 2.5). Suppose the electron exits from the $\uparrow_x$ door of the first x-spin box. Given the evidence that x-spin and z-spin are uncorrelated, the probability of it exiting from each door on the z-spin box is 1/2. Suppose it exits from the $\downarrow_z$ door then enters the second x-spin box. The electron exits the second x-spin box from the $\uparrow_x$ door about half of the time and from the $\downarrow_x$ door about half of the time. And the same thing happens if the electron exits from the first two boxes from other doors. It seems that the intervening z-spin box

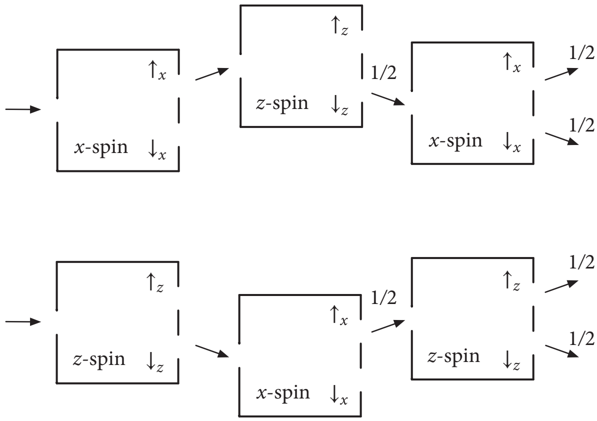


Figure 2.5. The randomizing behavior of alternating spin boxes.

completely *randomizes* the x-spin of an electron passing through it. Similarly, the z-spin of an electron seems to be completely randomized by an x-spin box.

This effect is different from knocking the electron around between x-spin boxes. There the likelihood of repeating the original x-spin result is a function of how much one disturbs the electron. Here it is not a matter of degree. An intervening z-spin box *completely randomizes* the observed x-spin of the electron, and an x-spin box completely randomizes the z-spin of an electron. And it is impossible to build an x-spin or z-spin box that exhibits the type of sorting behavior we have been considering but does not completely randomize the other spin property. The observables x-spin and z-spin stand in a special relationship to each other. Given the randomizing effect of an intervening observation of the complementary spin property, one might conclude that it is impossible to know both the x-spin and z-spin of an electron at the same time. While this is sometimes how quantum-mechanical uncertainty is characterized, the actual situation, as we will see, is significantly more subtle.[4]

There are three aspects of the experiments discussed so far that are notably non-classical. First, since measuring spin properties involves the interaction between smooth electromagnetic fields, one would expect a continuous range of possible outcomes. But for a particular orientation of the magnetic field, electrons are either deflected up or down.

[4] The Heisenberg uncertainty principle is sometimes taken to be that one does not know the position of a particle if one knows its velocity, and vice versa. But, on the standard interpretation of states, quantum-mechanical uncertainty is not about uncertainty regarding the values of *possessed* physical properties. The orthodox view is that if a particle has a determinate velocity, it simply has no determinate position. We will see why when we consider non-commuting spin properties in the context of the two-path experiment.

Second, while classical mechanics is deterministic, quantum phenomena are fundamentally random and probabilistic. A Stern–Gerlach device is just a smooth inhomogeneous magnetic field. Given its simplicity, one would not expect it to *randomize* the properties of systems sent through it. Further, the interaction between a particle and such a field is precisely the sort of thing that one would expect classical mechanics to get right. A satisfactory formulation of quantum mechanics, then, should tell us something about the source of the observed randomness in this simple setup.

Third, the phenomena so far suggest that there are fundamental quantum-mechanical constraints concerning one's epistemic access to physical properties. In classical mechanics there is no reason in principle why one cannot determine a particle's position and velocity, or any other set of properties, at the same time. And one's measurement results will be accurate to whatever degree of precision one's measuring device allows given the care with which it was designed and built. In contrast, if one knows the x-spin of an electron, one must be maximally uncertain concerning the result of a subsequent z-spin measurement, and if one knows the z-spin of and electron, one must be maximally uncertain concerning the result of a subsequent x-spin measurement. Unlike the situation in classical mechanics, there appears to be no physical information about the electron's spin properties or anything else that would allow one to make the future prediction. This epistemic limitation is closely related to the fundamentally stochastic nature of quantum mechanics. The full specification of the physical state of a particle with a determine x-spin simply fails to determine the result of a subsequent z-spin measurement.

2.3 Quantum-Mechanical Interference

The two-slit experiment discussed earlier illustrates quantum-mechanical interference involving photons. Material particles like electrons also exhibit interference effects. Here we will consider an experiment involving spin properties.

Interference effects are a characteristic variety of quantum-mechanical phenomena. Importantly, the effect here does not involve interference between *different* electrons. Rather, it occurs even when a *single* electron at a time is sent through the apparatus.

Consider the two-path experiment illustrated in Figure 2.6. A z-spin up electron e^- starts in region I and is sent into an x-spin box.[5] The device is set up so that if the electron exits from the $\uparrow_x$ door, then it will travel along path A to region II and if the electron exits from the $\downarrow_x$ door, then it will travel along path B to region II. In each case, we will suppose that the electron is carefully deflected in such a way that (1) the length of each of the two paths from I to II is two meters, and the speed

[5] One would know that the electron is z-spin up if it just came out the $\uparrow_z$ door of a z-spin box.

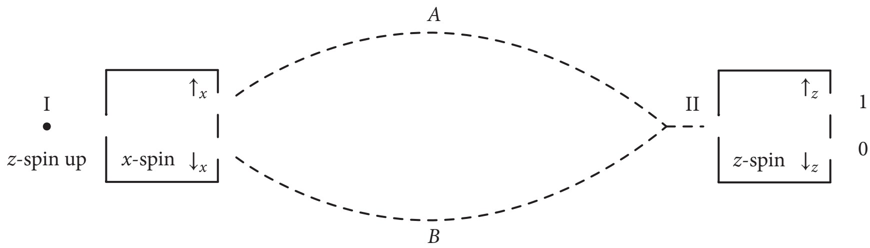


Figure 2.6. The two-path setup.

of the electron is a constant one meter per second, so it takes precisely two seconds to make the trip; (2) the electron's final velocity in region II is directly to the right just as it was in region I; and (3) the spin of the electron is not disturbed in any way on path *A* or *B*. We can test this last condition by sending electrons of known spin down each of the two paths and checking their spins with the appropriate box when they get to region II to make sure that the device has not disturbed the electron's initial state. Finally, we will also suppose that nothing in the environment becomes correlated to the electron's position or spin as it travels from I to II.

Suppose that the electron is sent through a *z*-spin box when it gets to region II. The question is which door one should expect it to exit from.

We know that if the *z*-spin up electron were just sent through the *z*-spin box directly, then it would come out the $\uparrow_z$ door. But here we are sending it though an *x*-spin box first, and our earlier experiments have given us very good reason to believe that an *x*-spin box randomizes *z*-spin. More specifically, we know that the probability of an *x*-spin up electron being found to be *z*-spin up is 1/2 and the probability of it being found to be *z*-spin down is 1/2. And we know that the same is true for an *x*-spin down electron. So if the electron exits the *x*-spin box from the $\uparrow_x$ door and hence takes path *A*, then it will have a 1/2 chance of being found *z*-spin up in region II. And if the electron exits the *x*-spin box from the $\downarrow_x$ door and takes path *B*, then it will have a 1/2 chance of being found *z*-spin up in region II. So whichever path the electron takes, the probability of it exiting from the $\uparrow_z$ door of the *z*-spin box at the end of the experiment is 1/2. Hence, if we send a series of electrons through the device one at a time, waiting for each to exit before we send the next in, about half of the electrons will exit from the $\uparrow_z$ door and about half will exit from the $\downarrow_z$ door.

But this is not what happens at all. Rather, if no one is looking at the paths, *all* of the electrons exit from the $\uparrow_z$ door.[6] Hence, electrons that start *z*-spin up in region I are all *z*-spin up again by the time they get to region II. In this setup, the intervening *x*-spin box *does nothing whatsoever* to the *z*-spin of the electrons

[6] And electrons that are initially *z*-spin down *always* exit from the the $\downarrow_z$ door in the two-path experiment.

passing through it. An x-spin box does not randomize z-spin after all. And there is good evidence here that it does not sort the electrons by x-spin either. If an electron is x-spin up, then it is always found to exhibit unbiased, random dispositions when it encounters a z-spin box. And similarly for x-spin down electrons. If the x-spin box were sorting the electrons by x-spin and then sending each along the path corresponding to its spin, then, depending on which path they took, the electrons in region II would be either x-spin up or x-spin down, so they would exhibit unbiased, random dispositions when they encountered a z-spin box. But they are all found to be z-spin up. Hence, we have good reason to believe that the x-spin box never sorted them by x-spin to begin with. That all of the electrons are found to be z-spin up in region II is an example of a quantum-mechanical interference effect.

This experiment also illustrates why it is difficult to say how electrons move in a way that satisfies our classical intuitions. Consider an electron that starts in region I. If it takes path A, then it is an x-spin up electron. But an x-spin up electron will exhibit the unbiased, random disposition to exit from either door when it encounters a z-spin box. So since the electron exhibits the sure-fire disposition to exit from the $\uparrow_z$ door, it did not take path A. But if the electron takes path B, then, one might reason classically, it is an x-spin down electron. But an x-spin down electron will also exhibit an unbiased, random disposition to exit from either door when it encounters a z-spin box. So since the electron exhibits the sure-fire disposition to exit from the $\uparrow_z$ door, it did not take path B either. And there is good empirical evidence that the electron does not take both paths or neither path. The electron is always found on precisely one of the two paths whenever one looks for it. Further, it is always found to be precisely where it should be given how fast it is moving. Since it takes two seconds to get from region I to region II given the setup, after one second, the electron will be found in the middle of path A or in the middle of path B.

Just as one would expect from the earlier experiments, the probability of finding the electron on either particular path if one looks is 1/2. Further, if one finds it on path A, it will also be found to be x-spin up. And if one finds it on path B, it will be x-spin down. So when one *looks* for the electron, it seems that the x-spin box *is* sorting electrons by their x-spin and that their x-spins are randomly distributed with equal probabilities. And if one finds the electron on one of the paths, it will be either x-spin up or x-spin down, so it will exhibit an unbiased, random disposition to exit from either door when it encounters the z-spin box in region II, or anywhere else one wants to put it beyond the place where one found the electron. So *finding* the electron on one of the two paths destroys the interference effect of its being guaranteed to be found z-spin up in region II.

Now we can see why it is not simply a matter of *not knowing* which path the electron takes from I to II. If the electron has the sure-fire disposition to exit by the $\uparrow_z$ door after it gets to region II, *it could not have determinately taken path A or determinately taken path B*. Had it taken one of these two paths, regardless of the issue of whether we knew which it took, then it was either x-spin up or it was x-spin down. And either way, we know that such an electron will have only a probability of 1/2 of exiting by the $\uparrow_z$ door after passing through region II. We have good empirical evidence then that the electron did not simply take one or the other of the two paths and we just don't know which.

Just as in the two-slit interference experiment with photons, this two-path experiment provides empirical evidence that the motions of fundamental particles fail to satisfy our classical intuitions. Further, this failure is of a radical sort. To put it in philosophical language, one might have thought that it was analytic to the notion of a material particle that it always had a determinate position and hence always followed a well-defined trajectory. But this now seems entirely wrong.

In addition to the issue of how the electrons are moving, the two-path experiment poses another puzzle. The earlier spin experiments we considered seemed to provide good empirical reason to suppose that an x-spin box completely randomizes z-spin, but now we have even better empirical reason to suppose that the box is doing nothing random whatsoever. But if that's right, one might naturally wonder what produces the random behavior we saw eariler of an electron that enters a z-spin box after exiting from an x-spin box.

It turns out that the question of how quantum randomness arises is closely related to the question of how electrons move. Indeed, one might distinguish between different strategies for formulating quantum mechanics by how one answers each of these two questions.[7]

2.4 How Electrons Move

There are a number of variants on the two-path experiment that one might consider in order to get a better idea of how a z-spin electron moves here. If one looks for the electron after one second, one always finds it to be either on path A or on path B, with an equal chance of each. If one looks on path B and does *not* find the electron, then one can infer that it is on path A. But if it is on path A, then it is x-spin up, and x-spin up electrons exhibit unbiased random dispositions when they encounter a z-spin box. And that is precisely what happens. If one looks at path B and does not find the electron, this destroys the interference effect and the electron may exit by either the $\uparrow_z$ or $\downarrow_z$ door of the second box, each with a probability of 1/2. More specifically, one might watch path B (as in Figure 2.7) while several z-spin up electrons are fed into the x-spin box, one by one, with enough time for the last to leave the entire apparatus before the next enters, and *even if one sees none of these*

[7] The thought is that one might distinguish between different formulations of quantum mechanics by considering how they explain determinate measurement records and the statistical distribution of such records. These two issues will arise repeatedly as we discuss explanatory options.

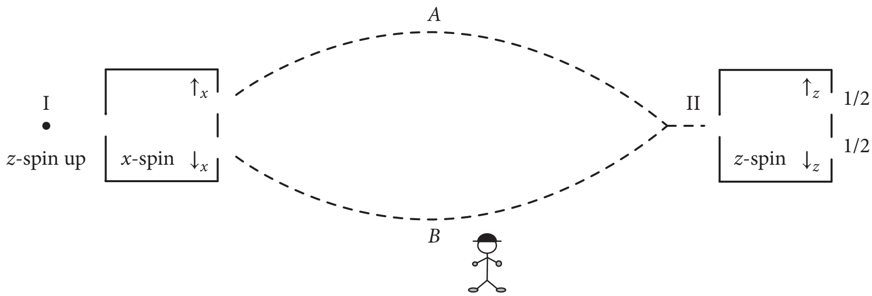


Figure 2.7. Two paths with an observer (who may see nothing on his path).

electrons on path B, and hence presumably never interacts with any of them because they were never on the observed path, about half will exit the z-spin box from the $\uparrow_z$ door and about half from the $\downarrow_z$ door. And as soon as one stops watching the path, all of the electrons will again exit from the $\uparrow_z$ door. What matters is whether one is watching the empty path when the electron would have been there had it taken that path. So simply looking and seeing nothing instantaneously affects the behavior of distant electrons.[8]

Given classical intuitions, it should not be possible to change the state of an electron without interacting with it. When an observer looks at an empty path and changes the statistical dispositions of a distant electron, the observer cannot be interacting with the distant electron directly. Further, classical mechanics provides no explanation for how the observer might be interacting with the electron indirectly either. There are simply no good candidates in classical mechanics for what might mediate such an interaction. The effect persists when the distance between the two paths is increased and when they are shielded from each other for electromagnetic interactions. Gravitational interactions are presumably too weak. And the strong and weak nuclear forces are poor candidates given their short range. More significantly, classical mechanics provides no dynamical account for how these fields, or any other, might affect the *statistical dispositions* of a distant electron when an observer looks at an empty path.

One can get a similar sort of quantum nonlocal behavior without an observer. Placing a barrier on path B that would gently stop an electron that takes that path (as in Figure 2.8) will also destroy the interference effect of getting a sure-fire z-spin up result at the second box. If an electron does not get to region II, then one will find it at the barrier. The electrons that get to region II took path A and hence never even got close to the barrier. Nevertheless, they will exit from the $\uparrow_z$ door about half of the time and from the $\downarrow_z$ door about half of the time. This happens regardless of

[8] The experimental setup described here is relatively small. Similar behavior is seen when the distances between the paths are much larger. Two-path experiments akin to this one have been performed with light over astronomical distances.

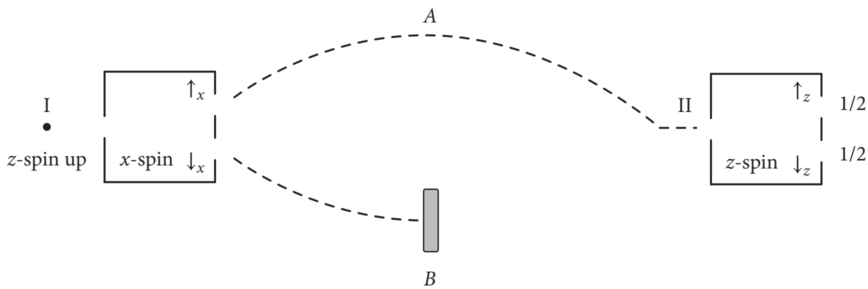


Figure 2.8. Two paths with a barrier on one.

the path in which one places the barrier, and it does not depend on whether one ever looks to see if the electron did in fact hit the barrier. Further, it does not matter when one decides which path to place the barrier on as long as the barrier is at a location at a time *when it would have blocked the electron* had the electron taken that path. Then the very presence of the barrier *in a location where the electron demonstrably did not go* will cause it to exhibit unbiased, random dispositions when it encounters the z-spin box in region II. Otherwise, without the barrier in place on the path not taken, the electron will always exit from the $\uparrow_z$ door. It is as if an electron that gets through the device "knows" whether or not there was a barrier on the path that it did not travel that would have blocked it had it traveled that path.

If one puts a barrier on both paths, then the electron does not get to region II and one always finds that it hit precisely one of the two barriers if one checks. But as we will see later, the standard formulation of quantum mechanics tells us that even this apparently straightforward situation is more subtle than one might have thought. The issue here turns on the phrase *if one checks*.

Suppose one wanted to determine which path the electron traveled without disturbing the interference effect of the sure-fire disposition of the electron to be found to be z-spin up in region II. One might try putting a camera on one of the paths to detect the electron. But no matter how careful one is, if the camera is sensitive enough to reliably record whether the electron took that path, then it will destroy the inference effect even if the electron does not take the path the camera is on and the resulting photograph is blank.

The simplest type of camera might involve a single particle. Suppose that one places a single negatively charged particle p near path B so that if the electron e^- takes path B, then particle p would be moved from region a to region b and so that if e^- takes path A, then particle p would stay in region a (as in Figure 2.9). In this setup, particle p *records* the path taken by the electron in its position. If p ends up in region a, then one might infer, given one's classical intuitions concerning physical records, that e^- took path A. And if p ends up in region b, then one might infer that e^- took path B. But the very presence of the recording particle p on path B, even if e^- does not take path B and particle p hence does not move at all,

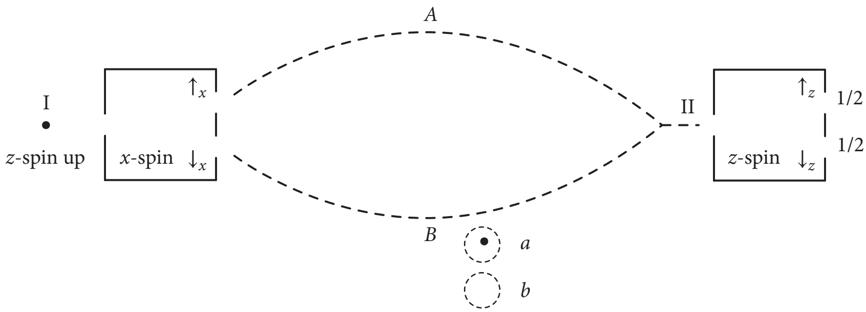


Figure 2.9. Two paths with a single recording particle.

again completely destroys the interference effect that the electron would otherwise exhibit in region II. More generally, if there is any record whatsoever of which path the electron followed or of its x-spin after passing through the x-spin box, or any correlation between the x-spin of the electron and its environment, intentional or not, then the interference effect will be destroyed and it will exit the z-spin box from each door with probability 1/2.

Another way to tell which path the electron followed would be to slow electrons on path B so that they take 2.1 seconds to get to region II instead of the usual 2 seconds. Then one knows that the electron took path A if it gets to region II in 2 seconds and that it took path B if it gets there in 2.1 seconds. But making electrons on one path take longer than those on the other again immediately destroys the interference effect in region II regardless of how long it takes each particular electron to get there.

In order to avoid interacting with the electron at all, one might try making one of the two paths slightly longer than the other, say 2.1 meters rather than 2 meters. Then the difference in the time it would take to travel each path would allow one to tell which path the electron took. But again, making one path longer than the other immediately destroys the interference effect in region II.

In short, the interference effect in region II is destroyed if one might know now or could have any way of reliably inferring later which path the electron took. This is related to the fact that one cannot simultaneously know both the x-spin and z-spin of an electron. If one could know which path the electron took without destroying the interference effect in region II, then one would know both its x-spin and z-spin at the same time. The standard formulation of quantum mechanics tells us that it is impossible to know the values of non-commuting physical quantities simultaneously, and we have never had any empirical reason to suppose that the theory is wrong in this.[9]

But again, there is good reason to believe that the real issue here is not one of what one *knows* about the electron. The electron presumably cannot *in fact*

[9] There is, however, a sense in which one might know the values of non-commuting quantities in a special context where one performs a self-measurement. See chapter 8 of Albert (1992).

be taking either of the two classically possible paths. If it were, it would not be guaranteed to be z-spin up at region II.

The standard formulation of quantum mechanics tells us that the problem is with our classical intuitions. Specifically, it says that when one is looking, the electron does not determinately travel either of the two classically possible paths, it does not determinately travel both, and it does not determinately travel neither. It also tells us that the reason an observer looking at an empty path destroys the interference effect in region II is *different* from the reason placing a barrier on the path not traveled or making one path longer than the other destroys the interference effect. Further, it tells us that the reason that the interference effect is destroyed when one records the path the electron takes in the position of a single particle depends on whether or not one observes that record *before* the electron gets to region II. We will see how all this works beginning in Chapter 4.

The results of the two-path experiments we have considered also suggest that the world exhibits a sort of nonlocality. We will consider a more subtle sort of quantum nonlocality when we discuss the EPR–Bell statistics in Chapter 6. A satisfactory formulation of quantum mechanics must predict and explain both the two-path phenomena and the EPR–Bell statistics.

2.5 Superpositions, Property Attribution, and the Total-of-Nothing Box

The standard formulation of quantum mechanics tells us that if a test electron is initially z-spin up in the sort of experiment we have been discussing, then it will follow a *superposition* of the two paths. Saying precisely what this means is one of the tasks of the theory, but we can get started on this here. An electron that follows a superposition of the two trajectories does not determinately travel path A or path B or both paths or neither path. Rather, the standard formulation of quantum mechanics tells us that such an electron with have no determinate classical trajectory whatsoever.

It will be important to make clear precisely what this means. In the two-path setup, the theory tells us that after one second, the electron is not in fact on path A, it is not in fact on path B, it is not in fact on both paths, and it is not in fact on neither path. That is, there is simply no fact of the matter concerning its classical location—it just does not have one. As David Albert has put the point, to ask where such an electron is, on the standard interpretation of states, is to commit a category mistake. It is like asking for the color of the number 7. Just as 7 is not the sort of thing that has a color, an electron in a superposition of different positions is not the sort of thing that has a position. But that the position of the electron is indeterminate does not mean that we do not know where it is, and nor does it mean that we cannot determine its position—it means that it simply fails to have a determinate position. And when the electron has a determinate x-spin, it simply fails to have a determinate z-spin, and the other way around.

Property attribution in quantum mechanics is *geometric* rather than *logical*.[10] Inasmuch as the electron does not *behave* like it is on either one of the two paths or on both paths or on neither path, it is at least in some sense unsurprising that quantum mechanics describes the electron as being in a state that is different from each of these four classical options. As we will see later, one might say that after one second the electron is at a 45° angle to being on path A and x-spin up and at a 45° angle to being on path B and x-spin down. But the theory will need to tell us how to understand such descriptions.

We will consider precisely how quantum mechanics uses geometric property attribution to describe the motion of the electron once we have the theory. In the meantime, let's consider one last bit of evidence against the reliability of our classical intuitions.

The third type of box is the one that Albert calls a total-of-nothing box. When an electron with a known spin is sent through such a box, it exhibits precisely the same spin after emerging from the other side as it did before—an x-spin up electron stays x-spin up, an x-spin down electron stays x-spin down, a z-spin up electron stays z-spin up, a z-spin down electron stays z-spin down, etc. Given that there are no fields in the box that impose forces on the electron, it *should do nothing to an electron* given our best *classical* understanding of the physical world. And on this the standard formulation of quantum mechanics agrees. On the standard interpretation of quantum-mechanical states, a total-of-nothing box indeed does nothing whatsoever to an electron that passes through it. Nor does it do anything to an electron that goes around it. But quantum mechanics predicts that such a box may do something striking to an electron that is in a *superposition of passing through it and around it*.

In the two-path setup a z-spin up electron that starts in region I, enters the x-spin box, and travels a superposition of the two paths without being disturbed or recorded in any way is always found to be z-spin up in region II. If under the same conditions a total-of-nothing box is placed on path B (as in Figure 2.10), a z-spin

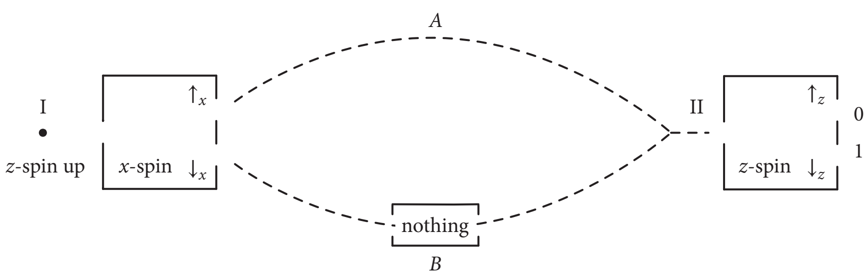


Figure 2.10. Two paths with a total-of-noting box on one.

[10] Jeff Bub (1997) describes this distinction well.

up electron that starts in region I and enters the x-spin box is always found to be z-spin *down* in region II. This is further empirical evidence that the electron does not determinately take either of the two classically possible paths. The total-of-nothing box demonstrably does nothing to an electron that goes through it and it demonstrably does nothing to an electron that goes around it. But it dramatically affects an electron that is in a *superposition* of going though it *and* around it.[11]

2.6 Random, Nonlocal, and Indeterminate

These idealized experiments illustrate some of the ways in which the physical world violates our basic classical intuitions. There is good evidence that physical systems exhibit random, in principle unpredictable behavior; that observations in one place may have nonlocal effects on systems somewhere else even when nothing is observed; that physical objects do not follow determinate trajectories; and that classical property attribution fails more generally. The intuitions violated here are sufficiently basic that one might reasonably, though entirely mistakenly, have thought that they were metaphysically necessary. One might have taken a situation where one observes nothing but affects the properties of a distant object without anything mediating the distant interaction to be *flatly unintelligible* and hence *impossible* or that physical property attribution *must necessarily* satisfy the laws of classical logic.

While it is unclear at this point in the story which of our intuitions might be preserved and at what cost, that the physical world is counterintuitive provides good reason up front not to trust our pre-theoretic philosophical intuitions. We will need to weigh such intuitions against each other in order to choose between alternative formulations of quantum mechanics later, but inasmuch as the empirical evidence itself is deeply counterintuitive, no empirically adequate formulation of quantum mechanics can preserve all of our cherished philosophical intuitions.

The task is to find a physical theory that gets the counterintuitive phenomena right and instructs our pre-theoretic intuitions. We will begin with the standard formulation of quantum mechanics, but first we will need some mathematical tools to say what the theory is.

[11] More generally, the electron will be found to be z-spin up in region II if there are an even number of total-of-nothing boxes anywhere on either of the two paths, and it will be found to be z-spin down if there are an odd number of total-of-nothing boxes on the paths. The total-of-nothing box illustrates a version of the Aharonov–Bohm effect (1959).

3
The Mathematics of Quantum Mechanics

3.1 Hilbert Space

We need a few basic mathematical notions to say what quantum mechanics is. In this chapter we will characterize the geometric and algebraic properties of a number of mathematical objects and begin to discuss their physical interpretation.[1]

Quantum mechanics is written in the language of linear algebra. The state of a physical system is represented by a unit-length vector $|\psi\rangle$ in Hilbert space $\mathcal{H}$.[2] A Hilbert space is a collection of vectors that can be added to each other and multiplied by real or complex numbers. It also has an inner product that takes any pair of vectors and returns a number.[3]

Geometrically, a vector might be thought of as an arrow with a magnitude and direction. If one defines an appropriate set of operations, the arrows in Figure 3.1(a) can be taken to form a two-dimensional, real-valued vector space. If one introduces a Cartesian coordinate system, one might represent a vector in this space algebraically by translating the base of the arrow to the origin of the coordinate system without changing its direction then taking the ordered pair consisting of the x and y coordinates of the tip of the arrow (x, y) as the algebraic representation. Figure 3.1(b) illustrates this for the vectors in Figure 3.1(a).[4]

For an n-dimensional vector space, one will in general need an n-tuple of complex numbers $(v_1, v_2, \ldots, v_n)$ to represent its vectors. There are two kinds of vectors we will consider. Both satisfy the axioms for being a vector space. A ket-vector is written as a column of complex numbers

$$|v\rangle = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}. \tag{3.1}$$

[1] While the physical significance of the mathematics is discussed here, readers with a background in linear algebra may wish to proceed directly to the specification of the standard formulation of quantum mechanics in the next chapter. One can find there a more efficient, though less user-friendly, presentation of the relationship between the mathematics and physics.

[2] This notation for representing state vectors was introduced by P. A. M. Dirac. It is widely used as it makes the sort of mathematical expressions that show up in quantum mechanics particularly easy to read. We will see how the notion works as we go.

[3] A formal characterization of a Hilbert space, complex numbers, and related notions can be found in Appendix A.

[4] As we will see, introducing a Cartesian coordinate system is equivalent to choosing an orthonormal basis. Ultimately, we will talk in terms of a choice of basis in the context of a Hilbert space.

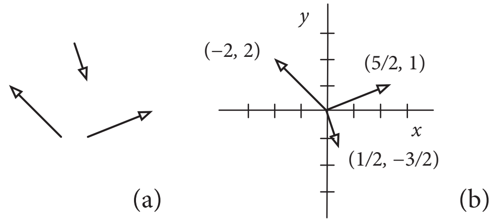


Figure 3.1. Arrows as vectors.

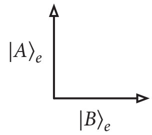


Figure 3.2. Orthogonal quantum states.

Corresponding to each ket-vector is a bra-vector that can be written as a row of complex numbers

$$\langle v| = \begin{pmatrix} v_1^* & v_2^* & \ldots & {v_n}^* \end{pmatrix}. \tag{3.2}$$

Each number in the bra-vector is the complex conjugate of the corresponding number in the ket-vector. The bra-vector corresponding to $|\phi\rangle + |\chi\rangle$ is $\langle\phi| + \langle\chi|$ and the bra vector corresponding to $\alpha|\phi\rangle$ is $\alpha^*\langle\phi|$. If all the numbers in a ket-vector are real, then the corresponding bra-vector will contain precisely the same real numbers.[5]

One might also construct a Hilbert space using a set of functions as vectors. One just needs to make sure that the functions and one's operations on them satisfy the axioms for a Hilbert space (as described in Appendix A). We will start by supposing that vectors are represented by n-tuples of real numbers. More specifically, we will be able to represent the interactions between electrons and spin boxes with real-valued two-dimensional vectors—vectors that can be represented as ordered pairs of real numbers. We will, however, need more general representations later.

The first connection between the mathematics and physics concerns how physical states are represented. The *quantum-mechanical state* of a physical system is represented by a vector of length one. The state of an electron e^- that is determinately on path A is represented by a unit-length vector $|A\rangle_e$, and the state where it is determinately on path B is represented by a unit-length vector $|B\rangle_e$ that is at a right angle to $|A\rangle_e$ (as in Figure 3.2). Note that the physical system that a state vector describes is indicated by a subscript on the vector.

[5] A bit more detail may be helpful here. The bra-vectors form a co-space that is isomorphic to the vector space formed by the ket-vectors. Each vector in the ket-space corresponds to a canonically dual bra-vector in the co-space. Because of the canonical relation between the two spaces, a vector and its dual can be thought of as the same vector, just written down differently. Physically, a state vector and its dual represent precisely the same state.

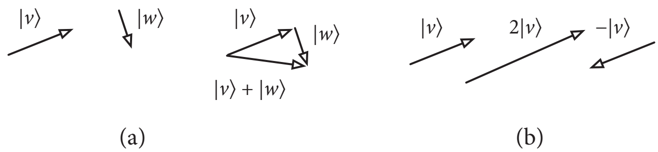


Figure 3.3. Vector addition and scalar multiplication.

Geometrically, one might think of the vector $|v\rangle + |w\rangle$ as what one gets by putting the tail of $|w\rangle$ at the tip of $|v\rangle$ and drawing an arrow from the tail of $|v\rangle$ to the tip of $|w\rangle$ (as in Figure 3.3(a)). One gets the same result regardless of the order in which one carries out this construction. If $|v\rangle$ and $|w\rangle$ are two-dimensional vectors, one might represent their sum algebraically by matrix addition

$$|v\rangle + |w\rangle = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix}. \tag{3.3}$$

Higher-dimensional vectors add the same way.

One can think of *scalar multiplication* as an operation that stretches or shortens a vector without moving it off the ray that it is on (as in Figure 3.3(b)). Multiplying a vector by a negative number flips the direction of the vector by 180°, but scalar multiplication always keeps the vector on the same ray (as in Figure 3.3(b)). Algebraically, multiplying a vector by a scalar just multiplies each number in the matrix by the scalar.

$$\alpha|v\rangle = \alpha \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \end{pmatrix}. \tag{3.4}$$

Vector addition together with scalar multiplication is used to represent *superpositions* of quantum-mechanical states. If one is only interested in its position, the state of an electron e^- that starts z-spin up is

$$\frac{1}{\sqrt{2}}|A\rangle_e + \frac{1}{\sqrt{2}}|B\rangle_e$$

after it passes through the x-spin box in the two-path experiment. This state describes the electron as being on a superposition of path A and path B. This is a unit length vector, so it can represent a state. And the state it represents is at a 45° angle to the vector representing e being on path A and at a 45° angle to the vector representing e being on path B (as in Figure 3.4). Unit-length vectors at different angles to $|A\rangle_e$ and $|B\rangle_e$ like

$$\frac{\sqrt{3}}{2}|A\rangle_e + \frac{1}{2}|B\rangle_e$$

represent different physical superpositions.

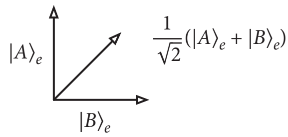


Figure 3.4. An electron on a superposition of paths.

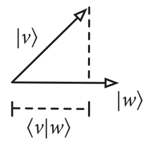


Figure 3.5. The inner product $\langle v|w\rangle$.

The superposition represented by the vector at a 45° angle in Figure 3.4 is not a state where the electron is on one or the other or both paths. If e were on path A, its state would be $|A\rangle_e$, and if it were on path B, its state would be $|B\rangle_e$. But paths A and B are not irrelevant to the electron's state. While it is not determinately on either path, it is associated with a quantum-mechanical *amplitude* of $1/\sqrt{2}$ of being on each path. The amplitudes are given by the scalar coefficients in the superposition. Amplitudes determine quantum-mechanical probabilities.

Bra- and ket-vectors are used to characterize the inner product $\langle v|w\rangle$. The inner product is related to both the length of vectors and the angle between them. The length of a real-valued vector $|v\rangle$ is equal to $\sqrt{\langle v|v\rangle}$. Geometrically, the inner product of two real-valued vectors $\langle v|w\rangle$ is the length of $|v\rangle$ times the length of $|w\rangle$ times the cosine of the angle between the two vectors. So, if each of the vectors is unit-length, which is always true when the vectors represent quantum-mechanical states, $\langle v|w\rangle$ is just the cosine of the angle between the two vectors or the orthogonal projection of one vector onto the other (as in Figure 3.5). Consequently, if the inner product of two real-valued, unit-length vectors is 1, then they are parallel, and hence the same vector representing the same physical state. And if it is 0, then they are orthogonal representing different physical states. States where physical properties differ *classically* are orthogonal.

Algebraically, the inner product of $\langle v|$ and $|w\rangle$ is

$$\langle v|w\rangle = \begin{pmatrix} v_1^* & v_2^* & \dots & v_n{}^* \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = v_1^* w_1 + v_2^* w_2 + \cdots + v_n^* w_n. \tag{3.5}$$

This can be thought of as matrix multiplication. If the vectors are two-dimensional and real valued, as they will be for the next few chapters, then this expression simplifies to

$$\langle v|w\rangle = \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = v_1 w_1 + v_2 w_2. \tag{3.6}$$

Probabilities in quantum mechanics are closely related to the inner product. Specifically, if a system S is in initial state $|\phi\rangle_S$, the probability it will be found in state $|\chi\rangle_S$ if one observes it is

$$|\langle\phi|\chi\rangle|^2 = \langle\phi|\chi\rangle\langle\chi|\phi\rangle,$$

or the norm-squared of the inner product of the two states.

There are always many ways to express a given vector as a linear combination (or sum with various coefficients on each of the terms) of other vectors in the vector space. A basis for the space $\mathcal{H}$ is a linearly independent set of vectors such that any vector $|\phi\rangle$ in the space can be represented as a linear combination of vectors in the set

$$|\phi\rangle = \sum_i \alpha_i |b_i\rangle.$$

where α_i are complex numbers and the index i goes from 1 to n picking up each of the elements in the basis. If this condition is satisfied, we say that the set *spans* the space $\mathcal{H}$. A basis is minimal in the sense that it will no longer span the space if any vector is removed. Every basis for $\mathcal{H}$ has the same cardinality or number of vectors. The *dimension* of the space $\mathcal{H}$ is the number of vectors in one of its bases.

A basis is *orthonormal* if every element in the basis is unit-length and orthogonal to every other element in the basis—that is, for different elements in the basis $|b_i\rangle$ and $|b_k\rangle$, $\langle b_i|b_k\rangle = 0$. While one can always represent a vector as a linear combination of basis vectors, an orthonormal basis has properties that make it especially easy to work with. If the basis is orthonormal, the coefficient on each term is just the inner product of the vector represented by the expansion and the corresponding element of the basis

$$|\phi\rangle = \sum_i |b_i\rangle\langle b_i|\phi\rangle. \tag{3.7}$$

In quantum mechanics, every observable physical property or quantity corresponds to an orthonormal basis. The state where electron e is determinately x-spin up is represented by a unit-length vector $|\uparrow_x\rangle_e$ and the state where it is determinately x-spin down is represented by a unit-length vector $|\downarrow_x\rangle_e$ that is orthogonal to the first. The two determinate x-spin states, then, form an orthonormal basis for a two-dimensional real-valued spin space (as in Figure 3.6(a)). And, as such, any state $|\phi\rangle_e$ in the space can be written as a linear combination of these two vectors

$$|\phi\rangle_e = \alpha|\uparrow_x\rangle_e + \beta|\downarrow_x\rangle_e,$$

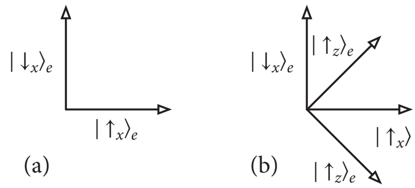


Figure 3.6. Bases representing x-spin and z-spin.

where the coefficients α and β are such that the length of $|\phi\rangle_e$ is 1.[6] This orthonormal basis, then, represents the physical observable x-spin.

The physical observable z-spin is represented by a different orthonormal basis. The z-spin basis is at an angle of 45° to the x-spin basis (as in Figure 3.6(b)). The vector $|{\uparrow_z}\rangle_e$ represents the state where e is determinately z-spin up, and $|{\downarrow_z}\rangle_e$ represents the state where e is determinately z-spin down. Note that when the electron has a determinate x-spin it does not have a determinate z-spin and when it has a determinate z-spin it does not have a determinate x-spin. *This* is the precise sense in which quantum-mechanical property attribution is geometric rather than logical. It is the *direction* of the state vector that represents the properties of the physical system.

The vectors of a Hilbert space can be acted on by operators. One might think of an operator $\hat{O}$ as a machine that takes an input vector $|\phi\rangle$ and produces an output vector $\hat{O}|\phi\rangle$. The vector resulting from the operation might be stretched, rotated, reflected, or projected or some combination of such actions, or one might just get the same vector out that one put in. Quantum mechanics uses *linear operators* to represent physical properties and to specify how physical states evolve over time. An operator $\hat{L}$ is *linear* if and only if

$$\hat{L}(\alpha|\phi\rangle + \beta|\chi\rangle) = \alpha\hat{L}|\phi\rangle + \beta\hat{L}|\chi\rangle.$$

That is, a linear operator yields the same result acting on the sum of two vectors as it does acting on each vector then taking their sum and carrying over the original coefficients.

Algebraically, a linear operator on an n-dimensional space can be represented as an $n \times n$ matrix, and the action of the operator on the vector as matrix multiplication. So in the two-dimensional case

$$\hat{A}|v\rangle = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 \\ a_{21}v_1 + a_{22}v_2 \end{pmatrix}. \tag{3.8}$$

[6] The state space must be complex-valued to represent all possible spin properties of an electron. That said, we can explain how electrons interact with x-spin and z-spin boxes restricting the coefficients to real numbers. This also makes it easier to picture the states geometrically.

For every linear operator on an n-dimensional Hilbert space $\mathcal{H}$ there is an $n \times n$ matrix that represents the operator, and for every $n \times n$ matrix there is a linear operator. Linear operators on an infinite-dimensional function space might take the form of differential operators. But in the case of electron spin properties, one needs only 2×2 matrices to represent the relevant linear operators.

Sometimes acting on a vector $|\psi\rangle$ with an operator $\hat{A}$ gives the original vector times back a scalar constant $\lambda|\psi\rangle$. When

$$\hat{A}|\psi\rangle = \lambda|\psi\rangle$$

we say that $|\psi\rangle$ is an *eigenvector* of operator $\hat{A}$ with *eigenvalue* λ. In this case, the action of the operator may shrink or stretch the vector, but it does not rotate it off the ray it is on.

Hermitian operators are a special type of linear operator that quantum mechanics uses to represent observable physical properties. An operator is *Hermitian* if one gets back the matrix representing the operator by reflecting the matrix along its diagonal and taking the complex conjugate of every element in the matrix. That is, a Hermitian operator is equal to its conjugate transpose $\hat{H} = \hat{H}^*$. In terms of the elements of a 3×3 matrix, this means that

$$\hat{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11}^* & a_{21}^* & a_{31}^* \\ a_{12}^* & a_{22}^* & a_{32}^* \\ a_{13}^* & a_{23}^* & a_{33}^* \end{pmatrix}. \tag{3.9}$$

A Hermitian operator has two properties that are particularly important for its physical interpretation: its eigenvectors form an orthonormal basis and its eigenvalues are real numbers.

As we saw earlier, there is a correspondence between physical observables and orthonormal bases. Since the eigenvectors of a Hermitian operator form an orthonormal basis, this means that there is also a correspondence between physical observables and Hermitian operators. Given an appropriate state space $\mathcal{H}$, for every physical observable there is a Hermitian operator, and for every Hermitian operator there is a physical observable. The real eigenvalue associated with an eigenvector is the value of the observable when the physical system is in the eigenstate represented by that eigenvector.

The representation of spin states that we have been using, written in the x-spin basis, is

$$|\uparrow_x\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, |\downarrow_x\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{3.10}$$

and

$$|\uparrow_z\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, |\downarrow_z\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}. \tag{3.11}$$

One has a significant degree of freedom in choosing such a representation. Given the precise way in which quantum-mechanical property attribution is geometric, all that matters is the angles or inner products between the state vectors.

The states where x-spin is determinate and the states where z-spin is determinate each form an orthonormal basis for the two-dimensional spin space and fully represent the corresponding physical observable.[7] These two observables might equivalently be represented by two Hermitian operators $\hat{\sigma}_x$ and $\hat{\sigma}_z$. The first has the determinate x-spin states as eigenvectors and the second has the determinate z-spin states as eigenvectors. Specifically, one can write the x-spin operator in the x-spin basis as

$$\hat{\sigma}_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

It is easy to check that

$$\hat{\sigma}_x|\uparrow_x\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = +1\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

So $|\uparrow_x\rangle$ is an eigenvector of $\hat{\sigma}_x$ with eigenvalue $\lambda = +1$. And since

$$\hat{\sigma}_x|\downarrow_x\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix} = -1\begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$|\downarrow_x\rangle$ is an eigenvector of $\hat{\sigma}_x$ with eigenvalue $\lambda = -1$. The physical interpretation is that if the quantum-mechanical state of an electron is an eigenstate of $\hat{\sigma}_x$ with eigenvalue $\lambda = +1$, then the electron determinately has the property of being x-spin up, and if it is in an eigenstate of $\hat{\sigma}_x$ with eigenvalue $\lambda = -1$, then it determinately has the property of being x-spin down.

Similarly, one can write the z-spin operator in the x-spin basis as

$$\hat{\sigma}_z = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

[7] Once one has a mathematical representation of the physics, it is sometimes convenient to use mathematical and physical language interchangeably as in this sentence.

Here one can check that the vector $|\uparrow_z\rangle$ is an eigenvector of $\hat{\sigma}_z$ with eigenvalue $\lambda = +1$ and the vector $|\downarrow_z\rangle$ is an eigenvector of $\hat{\sigma}_z$ with eigenvalue $\lambda = -1$. The physical interpretation is that if the state of an electron is an eigenstate of $\hat{\sigma}_z$ with eigenvalue $\lambda = +1$, then the electron determinately has the property of being z-spin up, and if it is in an eigenstate of $\hat{\sigma}_z$ with eigenvalue $\lambda = -1$, then it determinately has the property of being z-spin down.

Since the eigenstates of z-spin form an orthonormal basis and since the eigenstates of x-spin form an orthonormal basis, determinate x-spin states can be written in terms of z-spin states using expression 3.7:

$$|\uparrow_x\rangle = \langle\uparrow_z|\uparrow_x\rangle|\uparrow_z\rangle + \langle\downarrow_z|\uparrow_x\rangle|\downarrow_z\rangle = \frac{1}{\sqrt{2}}|\uparrow_z\rangle + \frac{1}{\sqrt{2}}|\downarrow_z\rangle$$
$$|\downarrow_x\rangle = \langle\uparrow_z|\downarrow_x\rangle|\uparrow_z\rangle + \langle\downarrow_z|\downarrow_x\rangle|\downarrow_z\rangle = \frac{1}{\sqrt{2}}|\uparrow_z\rangle - \frac{1}{\sqrt{2}}|\downarrow_z\rangle,$$

and the other way around

$$|\uparrow_z\rangle = \langle\uparrow_x|\uparrow_z\rangle|\uparrow_x\rangle + \langle\downarrow_x|\uparrow_z\rangle|\downarrow_x\rangle = \frac{1}{\sqrt{2}}|\uparrow_x\rangle + \frac{1}{\sqrt{2}}|\downarrow_x\rangle$$
$$|\downarrow_z\rangle = \langle\uparrow_x|\downarrow_z\rangle|\uparrow_x\rangle + \langle\downarrow_x|\downarrow_z\rangle|\downarrow_x\rangle = \frac{1}{\sqrt{2}}|\uparrow_x\rangle - \frac{1}{\sqrt{2}}|\downarrow_x\rangle.$$

So an electron with a determinate x-spin is in a *superposition* of z-spins and an electron with a determinate z-spin is in a *superposition* of x-spins.

Unitary operators are another special type of linear operator. An operator $\hat{U}$ is unitary if and only if

$$\hat{U}^*\hat{U} = \hat{U}\hat{U}^* = \hat{I},$$

where the operation is standard matrix multiplication and $\hat{I}$ is the identity operator. Geometrically, a unitary operator rotates a vector without changing its length. Since the quantum dynamics that usually obtains is linear and the time-evolution of a quantum-mechanical state is another quantum-mechanical state, and since all states are unit-length, this is precisely the type of operator one needs to represent the usual dynamics. We will say more precisely how states evolve once we have the full theory.

The last mathematical tool that we will need is another type of multiplication. The *tensor product* of an n-dimensional Hilbert space $\mathcal{H}_1$ and an m-dimensional Hilbert space $\mathcal{H}_2$ is an $n \times m$-dimensional Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. If the n vectors $|\phi_i\rangle$ form an orthonormal basis for $\mathcal{H}_1$ and the m vectors $|\chi_j\rangle$ form an orthonormal basis for $\mathcal{H}_2$, then the $n \times m$ vectors $|\phi_i\rangle \otimes |\chi_j\rangle$ form an orthonormal basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$. The tensor product of the two-dimensional vectors $|v\rangle$ and $w\rangle$

can be represented algebraically as

$$|v\rangle \otimes |w\rangle = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \otimes \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} v_1 w_1 \\ v_1 w_2 \\ v_2 w_1 \\ v_2 w_2 \end{pmatrix}.$$

The tensor product of higher-dimensional vectors is represented analogously. The tensor product associates across vector addition and both associates and commutes with scalar multiplication.

Quantum mechanics uses tensor products to combine spaces that represent different properties of the same system or properties of different quantum-mechanical systems. Physically, the tensor product is read as logical *and*. The state $|\uparrow_x\rangle_{e_1} \otimes |\downarrow_x\rangle_{e_2}$ is one where electron e_1 is x-spin up *and* electron e_2 is x-spin down. The state $|\uparrow_x\rangle_e \otimes |A\rangle_e$ is one where electron e is both x-spin up *and* on path A.[8]

3.2 Spin Space

While a real-valued two-dimensional Hilbert space is all we will need for the next few chapters, it is useful to see how complex-valued representations work. To this end, we will briefly consider a full spin space representation for an electron.

It is possible to consider the spin of an electron along *any* ray in three-dimensional space, not just those in the xz-plane. One might hence build a y-spin box that is associated with the electron spin along the y-axis in physical space, an axis that is orthogonal to both the x-axis associated with an x-spin box and the z-axis associated with a z-spin box. But to represent y-spin we need a richer mathematical structure.

A two-dimensional *complex-valued* Hilbert space allows one to represent *any* spin state of an electron. Spin states, in general, are represented by vectors of the form

$$|\psi\rangle_S = \begin{pmatrix} a_1 + b_1 i \\ a_2 + b_2 i \end{pmatrix}$$

where a_1, b_1, a_2, and b_2 are real numbers; $i = \sqrt{-1}$; and $\sqrt{\langle\psi|\psi\rangle} = 1$.

The y-spin operator written in the x-spin basis is

$$\hat{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

[8] Given its physical interpretation as logical *and*, one might take the tensor product between *physical states* to commute.

The operator $\hat{\sigma}_y$ is Hermitian because taking the complex conjugate of each element then taking the transpose of the matrix gives one back $\hat{\sigma}_y$. The two eigenstates of y-spin are

$$|\uparrow_y\rangle = \langle\uparrow_x|\uparrow_y\rangle|\uparrow_x\rangle + \langle\downarrow_x|\uparrow_y\rangle|\downarrow_x\rangle = \frac{1}{\sqrt{2}}|\uparrow_x\rangle + \frac{i}{\sqrt{2}}|\downarrow_x\rangle$$

and

$$|\downarrow_y\rangle = \langle\uparrow_x|\downarrow_y\rangle|\uparrow_x\rangle + \langle\downarrow_x|\downarrow_y\rangle|\downarrow_x\rangle = \frac{1}{\sqrt{2}}|\uparrow_x\rangle - \frac{i}{\sqrt{2}}|\downarrow_x\rangle.$$

The state $|\uparrow_y\rangle_e$ is an eigenstate of $\hat{\sigma}_y$ with eigenvalue $\lambda = +1$, and the state $|\downarrow_y\rangle_e$ is an eigenstate of $\hat{\sigma}_y$ with eigenvalue $\lambda = -1$. These two states represent e having the determinate property y-spin up and y-spin down respectively.

While it is important to the physics to have a full spin representation for electrons, we can get by with a two-dimensional, real-valued Hilbert space as long as we are only concerned with the x-spin and z-spin boxes. Later we will need higher-dimensional, complex-valued representations.[9]

Table 3.1 summarizes the mathematical notions we have discussed and their physical interpretation in the standard formulation of quantum mechanics.

[9] See Susskind and Friedman (2014) for further details regarding the representation of spin.

Table 3.1. Mathematical notions and physical interpretation on the standard formulation of quantum mechanics (on the Schrödinger picture).

	Mathematical notion	Algebraic representation	Physical interpretation
1.	Unit-length vector	$\vert\psi\rangle_S$	Physical state of S
2.	Vector sum	$\vert\psi\rangle_S + \vert\chi\rangle_S$	Superposition of $\vert\psi\rangle_S$ and $\vert\chi\rangle_S$
3.	Scalar multiplication	$\alpha\vert\psi\rangle_S + \beta\vert\chi\rangle_S$	Quantum-mechanical amplitudes
4.	Orthonormal basis	$\{\vert\phi_i\rangle_S\}$	Determinate property states
5.	Inner product	$\vert\psi\rangle_S = \sum_i \langle\phi_i\vert\psi\rangle\vert\psi\rangle_S$	State decomposition
6.	Linear operator	$\hat{L}(\alpha\vert\phi\rangle + \beta\vert\chi\rangle) = \alpha\hat{L}\vert\phi\rangle + \beta\hat{L}\vert\chi\rangle$	Observables and linear dynamics
7.	Hermitian operator	$\hat{H} = \hat{H}^*$	Physical observable
8.	Eigenvector and eigenvalue	$\hat{H}\vert\phi\rangle_S = \lambda\vert\phi\rangle_S$	Value of determinate property state
9.	Unitary operator	$\hat{U}^*\hat{U} = \hat{U}\hat{U}^* = \hat{I}$	Linear dynamics
10.	Norm-squared	$\vert\langle\phi_k\vert\psi\rangle\vert^2$	Probability of collapse from $\vert\psi\rangle_S$ to $\vert\phi_k\rangle_S$
11.	Tensor product	$\vert\phi\rangle_S \otimes \vert\psi\rangle_S$, $\vert\phi\rangle_R \otimes \vert\psi\rangle_S$	Composition of properties and systems

4
The Standard Formulation of Quantum Mechanics

4.1 The von Neumann–Dirac Theory

The standard formulation of quantum mechanics was described by P. A. M. Dirac (1930) and revised by John von Neumann (1932) (1955).[1] The theory can be expressed in five rules using the mathematical notions from the last chapter. The rules provide a systematic treatment of the physical ideas introduced there.[2]

1. Representation of states: The state of a physical system S is represented by a vector $|\psi\rangle_S$ of unit length in a Hilbert space $\mathcal{H}$.
2. Representation of observables: A physical observable O is represented by a Hermitian operator $\hat{O}$ on the state space $\mathcal{H}$, and each Hermitian operator on $\mathcal{H}$ corresponds to some observable.
3. Interpretation of states: A system S has a determinate value for observable O if and only if it is in an eigenstate of O: that is, S has a determinate value for O if and only if $\hat{O}|\psi\rangle_S = \lambda|\psi\rangle_S$, where $\hat{O}$ is the Hermitian operator corresponding to O, $|\psi\rangle_S$ is the vector representing the state of S, and the eigenvalue λ is a real number. In this case, one would with certainty get the result λ if one measured O of S.
4. Laws of motion:

   I. Linear dynamics: if *no measurement* is made of a physical system S, it will evolve in a deterministic, linear way: if the state of S is given by $|\psi(t_0)\rangle_S$ at time t_0, then its state at a time t will be given by $\hat{U}(t_0, t_1)|\psi(t_0)\rangle_S$, where $\hat{U}$ is a unitary operator on $\mathcal{H}$ that depends on the energy properties of S.

   II. Nonlinear collapse dynamics: if a *measurement* is made of a system S, it will instantaneously and nonlinearly jump to an eigenstate of the observable

[1] See Albert (1992) for an conceptual introduction to the theory with as little mathematics as possible. Maudlin (2019) is also an excellent companion to the arguments we will discuss here.

[2] This is the Schrödinger picture of the standard theory. The theory can also be formulated in the Heisenberg picture with observables rather than states evolving. In the Heisenberg picture, a *density matrix* may represent an epistemic mixture over pure states. The Schrödinger picture is conceptually clearer inasmuch as one is not tempted to muddle straightforward epistemic probabilities concerning what state obtains with the probabilities that result from the collapse of the quantum-mechanical state. One of the main achievements of von Neumann (1932) was to show the precise sense in which these two formulations are equivalent (an earlier study of the equivalence of the theories was given by Schrödinger (1926) himself). Their equivalence allows us to discuss the standard theory in the Schrödinger picture without loss of generality.

being measured (a state where the system has a determinate value of the quantity being measured). If the initial state is given by $|\psi\rangle_S$ and $|\chi\rangle_S$ is an eigenstate of O, then the probability of S collapsing to $|\chi\rangle_S$ is equal to $|\langle\chi|\psi\rangle|^2$ (the square of the magnitude of the projection of the premeasurement state onto the eigenstate). If a measurement is made, then the system instantaneously and randomly jumps from the initial superposition to an eigenstate of the observable being measured

$$|\psi\rangle = \sum_k c_k|\psi_k\rangle \longrightarrow |\psi_j\rangle,$$

where $|c_j|^2$ is the probability of ending up in the eigenstate $|\psi_j\rangle$.

5. Composition rule: If system S_1 is represented by an element $|\phi\rangle$ of $\mathcal{H}_1$ and S_2 by an element $|\psi\rangle$ of $\mathcal{H}_2$, then the composite system $S_1 + S_2$ is represented by an element $|\phi\rangle \otimes |\psi\rangle$ of $\mathcal{H}_1 \otimes \mathcal{H}_2$. Similarly, if property P_1 of system S is represented by an element $|\phi\rangle$ of $\mathcal{H}_1$ and an independent (quantum-mechanically compatible) property P_2 of S by an element $|\psi\rangle$ of $\mathcal{H}_2$, then both properties can be simultaneously represented by an element $|\phi\rangle \otimes |\psi\rangle$ of $\mathcal{H}_1 \otimes \mathcal{H}_2$.

Rule 1 says that the quantum-mechanical state of a system is represented as a unit-length vector in a Hilbert space. The dimension of the space depends on the system and the physical properties one wants to represent. While one has considerable flexibility in one's representation, empirical results fix the angles between states.

If one is interested just in the x-spin and z-spin properties of an electron e, one can represent the electron's state in a two-dimensional, real-valued Hilbert space $\mathcal{H}_1$ that is spanned by the orthonormal basis $|\uparrow_x\rangle_e$ and $|\downarrow_x\rangle_e$ and also by the orthonormal basis $|\uparrow_z\rangle_e$ and $|\downarrow_z\rangle_e$ (as in Figure 3.6).

If one is interested just in the position of the electron and one only distinguishes between it being in region I or on path A or on path B or in region II (as in Figure 4.1), then its state can be represented in a four-dimensional Hilbert space $\mathcal{H}_2$ spanned by the orthonormal basis $|\mathrm{I}\rangle_e$, $|A\rangle_e$, $|B\rangle_e$, and $|\mathrm{II}\rangle_e$ (as in Figure 4.2). Each

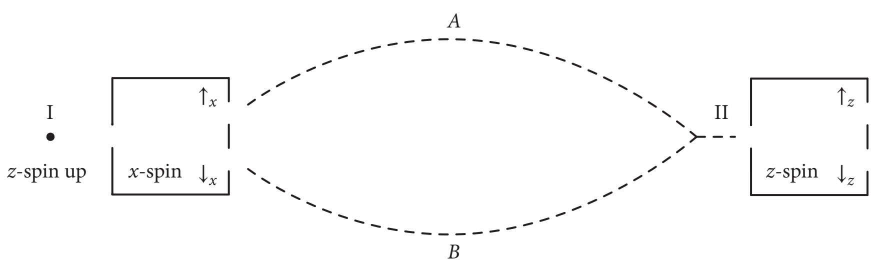


Figure 4.1. Two-path experiment with four regions.

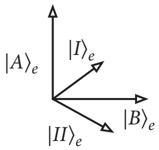


Figure 4.2. Orthogonal states for each of the four regions.

of these vectors represents a state where e is determinately in the corresponding region, and linear combinations of these vectors

$$|\psi\rangle_e = a_1|\mathrm{I}\rangle_e + a_2|A\rangle_e + a_3|B\rangle_e + a_4|\mathrm{II}\rangle_e$$

represent e in various superpositions of being in each of the regions.

Rule 2 says that all physical observables are represented by Hermitian operators and all Hermitian operators represent physical observables. Since the eigenvectors of a Hermitian operator form an orthonormal basis, one can also think of this as a correspondence between physical observables and orthonormal bases. A physical observable is a physical quantity like x-spin, position, or velocity. A physical property might be thought of as a system either having or not having a particular value for a specified observable.

Rule 3 is sometimes called the eigenvalue–eigenstate link. It provides the standard physical interpretation of quantum-mechanical states by describing how property attribution works.[3] In doing so, it connects the representation of states with the representation of observables. The rule stipulates that a system S has a value for some observable *if and only if* it is in one of the eigenstates of that observable. If so, the value of the observable is given by the eigenvalue associated with that eigenstate. The eigenvalues of an observable, then, are the possible results of a measurement of that observable. In terms of properties, a system S determinately has a particular physical property if and only it is in an eigenstate of having that property. A typical state will hence describe a system as neither determinately having nor determinately not having a particular specified property.

The Hermitian operator representing the x-spin of an electron e is

$$\hat{\sigma}_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

written in the x-spin basis. The state $|\uparrow_x\rangle_e$ is associated with eigenvalue $+1$, which, given the choice of representation and the eigenvalue–eigenstate link, corresponds to having the determinate property of being x-spin up. And the state $|\downarrow_x\rangle_e$ is

[3] This rule has been a part of the standard textbook presentation of quantum mechanics from Dirac and von Neumann to the present. See Gilton (2016) for a brief history.

associated with eigenvalue −1, which corresponds to having the determinate property of being x-spin down.

The operator representing z-spin is

$$\hat{\sigma}_z = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

in the x-spin basis. Here the state $|\uparrow_z\rangle_e$ is associated with eigenvalue $\lambda = +1$, and the state $|\downarrow_z\rangle_e$ is associated with eigenvalue $\lambda = -1$. These two states then correspond to e having the determinate property z-spin up and z-spin down respectively.

On the standard interpretation of states, the two vectors $|\phi\rangle_e$ and $-|\phi\rangle_e$ represent precisely the same physical state. What matters physically is the *ray* determined by the state vector.

The other direction of the eigenvalue–eigenstate link says that if a system S is not in an eigenstate of a particular observable, then it simply fails to have a value for the observable at all. A system that is in an eigenstate of z-spin, for example, has no x-spin. The operators representing these two observables do not commute.[4] On the standard interpretation of states, this indicates that the properties are incompatible and cannot be simultaneously determinate. It is not that the system in an eigenstate of z-spin has an x-spin but we just do not know what it is. This is not an epistemic issue. Rather, it is a part of the metaphysics of property atrribution on the standard interpretation of states.

Since the eigenvectors of a Hermitian operator form an orthonormal basis, there is also a physical observable corresponding to each orthonormal basis. This means that there are a continuous infinity of different physical properties that one can represent just in the two-dimensional space we are using to represent x-spin and z-spin. Start with the x-spin basis and rotate it by some angle, keeping the two basis vectors orthogonal. Each rotation, up to a right angle, corresponds to a different physical observable. In this case, these observables correspond to different orientations of the magnets in the Stern–Gerlach apparatus. Note that *every state* is an eigenstate of *some* physical observable. This means that while most of the physical properties of a given system will fail to be determinate, it will always have at least one determinate physical property.

The bases corresponding to the observables x-spin and z-spin are geometrically as different as possible. Position and momentum are another example of physical observables that are maximally different. If the value of one such observable is known, then the *result of a measurement* of the other is maximally uncertain. This is the structure of the Heisenberg uncertainty relations.

[4] That is, when one multiplies the first Hermitian operator by the second (using the standard rule for matrix multiplication), one gets a different result than when one multiplies the second by the first.

It is sometimes said that if one knows the x-spin of an electron, then one is maximally uncertain regarding its z-spin, and if one knows its z-spin, then one is maximally uncertain regarding its x-spin. But this way of putting it is *fundamentally wrong* given the standard interpretation of states. It is not an issue of one's epistemic uncertainty regarding possessed properties. If an electron has an x-spin, we are not uncertain regarding its x-spin. Rather, by the eigenvalue–eigenstate link, *it simply fails to have any determinate z-spin whatsoever*. What is uncertain is the result of a subsequent measurement of the non-commuting observable. And rule 4II tells us why.[5]

This has always been an essential feature of the standard formulation of quantum mechanics. Dirac contrasted the way that properties work in classical mechanics and quantum mechanics in the first edition of his *Principles of Quantum Mechanics* as follows:

> In classical mechanics an observable always has a particular value for any state. This is not so in quantum mechanics, where a special condition $[\hat{O}|\psi\rangle_S = \lambda|\psi\rangle_S]$ is necessary for an observable to have a particular value for a certain state.
>
> (1930, 31)

After all, when the initially z-spin up electron in the two-path experiment of Figure 4.1 ends up in a superposition of being on path A and being on path B, there is very good empirical reason to believe that it cannot in fact be on one of the two paths and we just do not know which. Specifically, if it were determinately on one or the other of the two paths, it would be x-spin up or x-spin down and hence could not have the sure-fire disposition to behave in the z-spin up way when it gets to region II.[6]

Dirac explained that the classical way of speaking *is* permissible when a system is in an eigenstate of having a particular property. In that case, the system simply *has* the property (1930, 30). Putting the two directions together, a physical system has a definite value for an observable *if and only if* it is in an eigenstate of that observable. Dirac then held the standard interpretation of quantum-mechanical states from his first description of the theory.[7]

Rule 3 is not a dynamical law. The act of measuring an observable of a system *does not* multiply the vector representing the system's state by the Hermitian

[5] Another way to put the point is that the Heisenberg relations do not represent uncertainty concerning the values of possessed properties but rather the geometric relationship between the bases representing those quantities. In the case of x-spin and z-spin, the uncertainty relations correspond to the fact that the basis for one observable is a 45° rotation of the basis for the other observable.

[6] We will see a way to avoid this conclusion when we discuss Bohmian mechanics in Chapter 11. But historically, people took such empirical results as direct evidence for the standard eigenvalue–eigenstate link.

[7] As discussed in the present section, von Neumann's understanding of the eigenvalue–eigenstate link was tied to his understanding of state completeness and of the role played by the two dynamical laws. See also his interpretational principle β (1955, 253). See Gilton (2016) for a general discussion of the history of the eigenvalue–eigenstate link.

operator representing the observable. The rule just says how Hermitian operators represent physical observables and their values. What *happens* to a system when it is not being measured is described by rule 4I. Rule 4II describes what happens when a measurement is made.

Rule 4 gives the dynamical laws of the theory, and in doing so it treats *measurement* as a primitive, undefined concept. Rule 4I says that when *no measurement* is made, the system evolves in a deterministic, linear way that depends only on its energy properties. This is the linear dynamics. Von Neumann called it process 2 (1955, 351–2 and 417–18). It describes how an electron *e* (or any other physical system) moves when no measurements are being made of it. In contrast, rule 4II says that when an observable of a system is *measured*, the system instantaneously and randomly jumps to a state where it determinately has or determinately does not have the property being observed with probabilities determined by the norm-squared of the magnitude of the projection of the state onto each possible outcome. This rule for assigning probabilities is sometimes called the Born rule, after Max Born (1926) who first described it.[8] Rule 4II is sometimes called the *collapse* of the quantum-mechanical state. Von Neumann called it process 1 (1955, 351–2 and 417–18). It explains the standard quantum statistics, why one gets a determinate result when one performs a measurement, and why a repeated measurement yields the same result if the system has not been disturbed between measurements.[9]

Dirac described the collapse dynamics in his textbook on quantum mechanics as follows:

> When we measure a real dynamical variable . . . the disturbance involved in the act of measurement causes a jump in the state of the dynamical system. From physical continuity, if we make a second measurement of the same dynamical variable . . . immediately after the first, the result of the second measurement must be the same as that of the first. Thus after the first measurement has been made, there is no indeterminacy in the result of the second. Hence, after the first measurement has been made, the system is in an eigenstate of the dynamical variable . . . . This conclusion must still hold if the second measurement is not actually made. In this way we see that a measurement always causes the system to jump to an eigenstate of the dynamical variable that is being measured . . . .
> (1957, 36)

What Dirac referred to as *the disturbance involved in the act of measurement* here is not something that can be understood classically nor is it something that can be understood in terms of the linear dynamics (rule 4I). Rather, on the formulation

[8] See Pais (1986, 255–61) for a discussion of the history of Born's proposal.

[9] The time-dependent Schrödinger dynamics is just a special case of the unitary dynamics described by rule 4I. It is used when one wants to describe the evolution of a wave function that assigns amplitude densities to a continuous range of possible positions. In that case, the dynamics is written as a differential equation to represent the continuous evolution of the wave function.

of quantum mechanics Dirac presented, the collapse dynamics (rule 4II) simply stipulates a fundamentally random evolution of the state of an object system on measurement.

In his sharpening of Dirac's initial presentation of the theory, von Neumann put it as follows

> We therefore have two fundamentally different types of interventions which can occur in a system S.... First, the arbitrary changes by measurements which are given by [rule 4II, von Neumann's process 1]. Second, the automatic changes which occur with the passage of time. These are given by [rule 4I, von Neumann's process 2]. (1955, 351)

He repeatedly indicated that the random behavior of a system under measurement was not the result of any lack of knowledge of "hidden parameters" that in fact determine the measurement result. Rather, von Neumann believed that one should "admit as a fact that the laws which govern the elementary process (i.e. the laws of quantum mechanics) are of a statistical nature." And he believed that a statistical interpretation of quantum mechanics involving fundamentally non-causal processes was "the only consistently enforceable interpretation of quantum mechanics" (1955, 209–10).[10]

According to rule 4II, the probability of a measurement outcome is given by the norm-squared of the inner product of the object system's state and the state corresponding to the outcome in question. The probability that an electron in state $|\uparrow_z\rangle_e$ would yield x-spin up as the result of an x-spin measurement is

$$|\langle\uparrow_z|\uparrow_x\rangle|^2 = \left|\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\right|^2 = \left|\left(\frac{1}{\sqrt{2}}\right)(1) + \left(\frac{1}{\sqrt{2}}\right)(0)\right|^2 = \left|\frac{1}{\sqrt{2}}\right|^2 = \frac{1}{2}.$$

And, if it did, the electron would be left in the state $|\uparrow_x\rangle_e$. Since

$$|\uparrow_z\rangle = \langle\uparrow_x|\uparrow_z\rangle|\uparrow_x\rangle + \langle\downarrow_x|\uparrow_z\rangle|\downarrow_x\rangle = \frac{1}{\sqrt{2}}|\uparrow_x\rangle + \frac{1}{\sqrt{2}}|\downarrow_x\rangle$$

[10] There is a puzzle here concerning what one means when one says that the state collapses described by rule 4II are *random*, or as von Neumann put it *arbitrary* or *capricious*. The uncontentious idea is that one expects that a sequence of quantum measurement results will not exhibit any specifiable pattern. Closely related, one expects that there will be no specifiable betting strategy that will allow one to do better than chance betting on these outcomes. One also expects the results of a random process to exhibit a sort of statistical independence. There are, however, a number of ways that one might understand these three conditions. Another puzzle concerns how one might have empirical evidence that the process that generates one's measurement results is in fact fundamentally random. While there is much to say here, there are higher-priority puzzles at hand, like saying precisely when and why each of the two specified dynamical laws obtains and why. See Barrett and Huttegger (2019) for a discussion of quantum randomness.

the probability of getting the result x-spin up is just the norm-squared of the coefficient on the $|\uparrow_x\rangle$ term. Similarly, the probability of getting x-spin down is the norm-squared of the coefficient on the $|\downarrow_x\rangle$ term, which is also 1/2.

Similarly, if the state were

$$\frac{\sqrt{3}}{2}|A\rangle_e - \frac{1}{2}|B\rangle_e$$

the probability of finding the electron on path A, if one were to look, would be 3/4 and the probability of finding it on path B would be 1/4, and in each case the collapse would leave it in the corresponding eigenstate of position. The probabilities, again, are given by the norm-squared of the coefficients on each term when the state is written in the basis associated with the observable one is considering. We were interested in x-spin in the last case; here we are interested in the electron's position.

The collapse dynamics (rule 4II) works closely with the eigenvalue–eigenstate link (rule 3). If a system in fact either determinately had or determinately did not have a property, then a measurement of the property would simply reveal whether it had it or not. But the eigenvalue–eigenstate link tells us that the system only determinately has or determinately does not have a particular property if it is either in an eigenstate of having the property or in an eigenstate of not having the property. Since most states do not satisfy this special condition, a system will typically fail to have or not have a particular specified property. So for the outcome of a measurement to reveal a determinate possessed property, the system must somehow first get into an eigenstate of the observable being measured. And to get the right quantum statistics, the eigenstate it ends up in must be randomly selected with the standard quantum probabilities, with the norm-squared of the quantum-mechanical amplitude associated with that possible outcome.

The eigenvalue–eigenstate link (rule 3) works closely with the *dynamics* to explain the standard quantum statistics. The eigenvalue–eigenstate link is hence an essential component of the standard formulation of quantum mechanics, not an *optional* interpretational principle. This is why it has played a central conceptual role from the earliest presentations of the standard theory.

While appealing to a distinction between the physical theory of quantum mechanics and its interpretation is commonplace, it is often in practice unclear how to distinguish between the two. How one understands quantum-mechanical states, observables, and probabilities must fit with the dynamics of one's particular formulation of the theory. This is perhaps easiest to see when the standard theory is contrasted with the alternative formulations of quantum mechanics like GRW (which we will discuss in Chapter 8) and Bohmian mechanics (which we will discuss in Chapter 11). The dynamical laws of GRW will require us to change how quantum-mechanical states are interpreted, and the interpretational strategy of Bohmian mechanics will require different dynamical laws.

There are other reasons to be skeptical regarding the idea that one can sharply distinguish between one's physical theory and its interpretation. Inasmuch as one takes one's physical theory to provide explanations of physical processes, phenomena, events, and so on, one's theory must incorporate at least some interpretational assumptions. One should want to include in one's specification of a theory those assumptions that play the most significant roles in its explanations and predictions. While it is not always easy to decide precisely what assumptions should count as part of a particular formulation of quantum mechanics, the eigenvalue–eigenstate link clearly works together with the dynamical laws to explain determinate measurement records and their statistical properties in the standard collapse theory.

The eigenvalue–eigenstate link is closely related to the principle of state completeness. It has been widely held that the standard quantum-mechanical state provides a complete and accurate representation of the state of a physical system S; that there is simply nothing else that one could add. Suppose that this is true, and consider a physical system S in a symmetric superposition of eigenstates of having a classical property A and not having that property $\neg A$:

$$\frac{1}{\sqrt{2}}|A\rangle_S + \frac{1}{\sqrt{2}}|\neg A\rangle_S.$$

If one believes that this vector provides a *complete* physical description of system S, then whether or not it has property A cannot possibly be determinate as there is, by the symmetry of the state, nothing whatsoever that could decide the issue. Consider a non-symmetric state:

$$|\psi\rangle_S = \alpha|A\rangle_S + \beta|\neg A\rangle_S.$$

Insofar as one believes that this state must be compatible with a measurement yielding either the result A *or* the result $\neg A$, then if one believes that this vector provides a complete physical description of the current state of the system, whether or not S has property A *cannot be determinate* since there is nothing that could decide which of the two conditions in fact obtains. So if S is not in an eigenstate of having or an eigenstate of not having property A, then given the principle of state completeness property A cannot be determinate.

The other direction of the eigenvalue–eigenstate link is relatively straightforward. Insofar as S would have the sure-fire disposition of exhibiting property A on measurement when it is in state $|A\rangle_S$ and since, by completeness, there is nothing else that might determine this result, S has property A in this state.

As von Neumann put the point, “it is evident that everything that can be said about the state of a system must be derived from its wave function [the vector representing its state]” (1955, 196). Insofar as one cannot derive whether S has property A from the vector $|\psi\rangle_S$ above, there is nothing whatsoever that one can say regarding whether S has the property. The issue here concerns what physical facts there *are*, not just what one can know about those facts. As von Neumann explained later, “we believe that after having specified [the quantum state] we know the state completely” (1955, 207). If $|\psi\rangle_S$ is in fact *a complete specification of the state*, then whether it has property A must be *physically* indeterminate. A commitment to state completeness for von Neumann entails a commitment to the eigenvalue–eigenstate link.

One last thing about rule 4 itself. There is a sense in which the standard theory's dynamics is essentially incomplete. The theory says that the state collapses to a randomly determined eigenstate of the observable being measured when a *measurement* is made, but it does not tell us what constitutes a measurement. That *measurement* occurs in the standard theory as an undefined primitive term leads to the quantum measurement problem. We will return to consider this problem in detail after we see how the standard theory works (in Chapter 7). In the meantime, we will suppose that what physical interactions should count as measurements is clear enough to meaningfully apply the theory.

Rule 5 tells us how to combine the state representations for different subsystems to get a state representation for a composite system and how to combine the state representations involving quantum-mechanically independent properties of the same system to get a state representation for both. We will briefly consider each of these tasks in turn.

Consider two electrons, e_1 and e_2. The spin of e_1 can be represented by a vector in two-dimensional space $\mathcal{H}_{e_1}$. The spin of e_2 can also be represented by a vector in two-dimensional space $\mathcal{H}_{e_2}$. By rule 5, the spin state of the composite system e_1 *and* e_2 can be represented by a vector in the four-dimensional space $\mathcal{H}_{e_1} \otimes \mathcal{H}_{e_2}$ (as in Figure 4.3). Since $|\uparrow_x\rangle_{e_1}$ and $|\downarrow_x\rangle_{e_1}$ form an orthonormal basis for $\mathcal{H}_{e_1}$ and $|\uparrow_x\rangle_{e_2}$ and $|\downarrow_x\rangle_{e_2}$ form an orthonormal basis for $\mathcal{H}_{e_2}$, the four vectors $|\uparrow_x\rangle_{e_1} \otimes |\uparrow_x\rangle_{e_2}$,

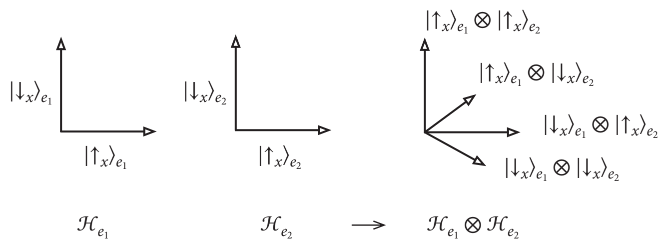


Figure 4.3. Tensor product representing different systems.

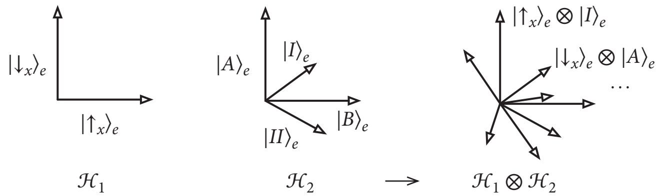


Figure 4.4. Tensor product representing different properties.

$|\uparrow_x\rangle_{e_1} \otimes |\downarrow_x\rangle_{e_2}$, $|\downarrow_x\rangle_{e_1} \otimes |\uparrow_x\rangle_{e_2}$, and $|\downarrow_x\rangle_{e_1} \otimes |\downarrow_x\rangle_{e_2}$ form an orthonormal basis for $\mathcal{H}_{e_1} \otimes \mathcal{H}_{e_2}$. Linear combinations of these vectors allows one to get the z-spin eigenstates for the two particles as well.

Consider the spin and position of an electron e. The spin of e can be represented by a vector in two-dimensional space $\mathcal{H}_1$ spanned by the orthonormal basis $|\uparrow_x\rangle_e$ and $|\downarrow_x\rangle_e$ (as in Figure 4.4). Similarly, the four possible locations in the two-path experiment can be represented by the vector in the four-dimensional space $\mathcal{H}_2$ spanned by the orthonormal basis $|\text{I}\rangle_e$, $|A\rangle_e$, $|B\rangle_e$, and $|\text{II}\rangle_e$. By rule 5, the spin and position of e can be represented by a vector in the eight-dimensional space $\mathcal{H}_1 \otimes \mathcal{H}_2$ spanned by the orthonormal basis $|\uparrow_x\rangle_e \otimes |\text{I}\rangle_e$, $|\uparrow_x\rangle_e \otimes |A\rangle_e$, ..., $|\downarrow_x\rangle_e \otimes |\text{I}\rangle_e$, $|\downarrow_x\rangle_e \otimes |A\rangle_e$, ..., and $|\downarrow_x\rangle_e \otimes |\text{II}\rangle_e$.

Physically, the tensor product is read as logical *and*. So $|\uparrow_x\rangle_{e_1} \otimes |\downarrow_z\rangle_{e_2}$ means that e_1 is x-spin up and e_2 is z-spin down, and $|\downarrow_x\rangle_e \otimes |A\rangle_e$ means that e is x-spin up and on path A. Since we will be writing the tensor product often, we will hereafter adopt the shorthand of writing $|\phi\rangle \otimes |\chi\rangle$ as $|\phi\rangle|\chi\rangle$. So $|\downarrow_x\rangle_e|A\rangle_e$ represents an x-spin up electron on path A.

Note that while e might have a determinate x-spin and a determinate position at the same time, it cannot have a determinate x-spin and a determinate z-spin at the same time. As a consequence, one *cannot* use rule 5 to represent the state of e as $|\uparrow_x\rangle_e \otimes |\downarrow_z\rangle_e$. Indeed, it feels wrong even typing this expression. The reason is that x-spin and z-spin are quantum-mechanically interrelated quantities. Formally, this is indicated by the fact that they correspond to different bases in the *same* Hilbert space. Quantum mechanics does not tell us which physical quantities are quantum-mechanically independent and which are not. This is something that must be determined empirically.

A closely related empirical constraint on the representation of states is the observed relative frequencies of one's measurements results. If e has a determinate x-spin and its z-spin is measured, it is found to be z-spin up about half the time and z-spin down about half of the time. And if it has a determinate z-spin and its x-spin is measured, it is found to be x-spin up about half the time and x-spin down about half of the time. This is what fixes the angle between the x-spin and z-spin bases at 45°.

There is a formal constraint on the representation of states that has to do with the *topology* of Hilbert space. Since every possible value that an observable might have corresponds to a different orthogonal vector in the state space, the dimension of the space must be at least the cardinality of the spectrum (the set of possible values) of the observable one wants to represent. Since the theory requires the state space to be separable, its dimension is at most countably infinite.[11] This places a significant constraint on what observables one might represent. In particular, one cannot represent quantities with continuous spectra like precise position, velocity, or energy. A standard way around this is to use a Hilbert space with an orthonormal basis corresponding to a countable discrete spectrum that provides the degree of precision one needs in order to represent the behavior of the particular physical system at hand.

Empirical considerations also constrain when each of the two dynamical laws is used. The basic methodological rule of thumb that one learns as a physics student is to use rule 4I unless one needs to explain a determinate measurement outcome. But empirical constraints often *require* one to use rule 4I. As we shall see presently, one *must* use the *linear dynamics* to describe the interactions between the electron and the x-spin, z-spin, and total-of-nothing boxes in order to get the right empirical results for the two-path experiments. We will return later (in Chapter 7) to the more general question of when each law should be used as this is central to the measurement problem.

The way that we have described the standard formulation of quantum mechanics here is sometimes called the Schrödinger picture or *wave mechanics*. This is in contrast to matrix mechanics or the Heisenberg picture. On the latter, one fixes a state vector then gives the time-evolution of the physical system at hand in terms of the evolution of the Hermitian operators representing one's observables. Among other things, von Neumann (1932) showed that these two ways of picturing the standard formulation of quantum mechanics are equivalent in a very strong mathematical sense and hence can be thought of as precisely the same physical theory. Given its relative conceptual clarity, we will use the Schrödinger picture throughout the book.

4.2 Spin Boxes and the Linear Dynamics

Consider an x-spin box (as in Figure 4.5(a)). The box is designed so that an x-spin up electron in region I would end up an x-spin up electron in region II

$$|\uparrow_x\rangle_e|\mathrm{I}\rangle_e \rightarrow |\uparrow_x\rangle_e|\mathrm{II}\rangle_e$$

[11] A vector space is separable if and only if it has countable dimension. If this condition is satisfied, then, given a specified basis, every vector in the space has a unique representation as a countable sum of the elements of the basis with appropriate coefficients. This notion of the separability of a vector space has nothing to do with quantum-mechanical notion of separability that we will discuss later in this chapter.

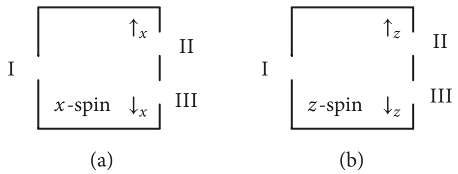


Figure 4.5. *x*-spin and *z*-spin boxes with labeled positions.

and an *x*-spin down electron in region I would end up an *x*-spin down electron in region III

$$|\downarrow_x\rangle_e|\mathrm{I}\rangle_e \rightarrow |\downarrow_x\rangle_e|\mathrm{III}\rangle_e.$$

These two dispositions fully characterize what an *x*-spin box does to the spin and position of an electron that encounters it.

Similarly, a *z*-spin box with similarly labeled positions (as in Figure 4.5(b)) is characterized by the following two dispositions:

$$|\uparrow_z\rangle_e|\mathrm{I}\rangle_e \rightarrow |\uparrow_z\rangle_e|\mathrm{II}\rangle_e$$

and

$$|\downarrow_z\rangle_e|\mathrm{I}\rangle_e \rightarrow |\downarrow_z\rangle_e|\mathrm{III}\rangle_e.$$

That is, it sends a *z*-spin up electron to region II without disturbing its spin and a *z*-spin down electron to region III without disturbing its spin. This is precisely how one would build boxes to *sort* electrons without disturbing them if spin properties behaved classically. But they don't.

Consider the *x*-spin box again. Given how it is designed, we know what it does to an *x*-spin up electron and to an *x*-spin down electron. But we need guidance from the theory to figure out what it does to a *z*-spin up electron.

Suppose that we start with a *z*-spin up electron *e* in region I, then send it through an *x*-spin box. Suppose further that no one looks at the electron, and suppose that the electron, consequently, does not count as being measured. The theory does not tell us what counts as a measurement, but, if anything does, presumably an observer looking at the electron for the purpose of determining its physical properties would. That the interaction with *x*-spin box *does not* count as a measurement is just a guess at this point. But as we will see when we get to the the empirical evidence from the two-path experiments, it is a good guess.

On the assumption that the interaction with the box does not count as a measurement, the time-evolution of *e* is given by the deterministic linear dynamics (rule 4I). In calculating how the state of the electron evolves, we will make essential use of the fact that this dynamics is *linear*. Recall that the action of a linear operator on a superposition of two states is the same as the action of the operator on each of the states individually, carrying over the coefficients from the superposition. That is, the time-evolution operator $\hat{U}$ has the property that

$$\hat{U}(\alpha|\phi\rangle + \beta|\chi\rangle) = \alpha\hat{U}|\phi\rangle + \beta\hat{U}|\chi\rangle.$$

The precise action of $\hat{U}$ is determined by the energy properties of the system, which are determined in this case by how the x-spin box is designed. Since the box sends an x-spin up electron to region II without disturbing its spin, we know that

$$\hat{U}|\uparrow_x\rangle_e|\text{I}\rangle_e = |\uparrow_x\rangle_e|\text{II}\rangle_e,$$

and since it sends an x-spin down electron to region III without disturbing its spin, we also know that

$$\hat{U}|\downarrow_x\rangle_e|\text{I}\rangle_e = |\downarrow_x\rangle_e|\text{III}\rangle_e.$$

What it does to a z-spin up electron follows from this, the linearity of the dynamics, and the relationship between x-spin and z-spin.

Suppose that it takes one second for the electron to go through the box. The initial state of e at time t_0 is

$$|\psi(t_0)\rangle_e = |\uparrow_z\rangle_e|\text{I}\rangle_e.$$

By rule 4I, its state after one second is

$$\begin{aligned}
|\psi(t_1)\rangle_e &= \hat{U}|\psi(t_0)\rangle_e \\
&= \hat{U}|\uparrow_z\rangle_e|\text{I}\rangle_e \\
&= \hat{U}\left[\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e\right]|\text{I}\rangle_e \\
&= \hat{U}\left[\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\text{I}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\text{I}\rangle_e\right] \\
&= \frac{1}{\sqrt{2}}\hat{U}|\uparrow_x\rangle_e|\text{I}\rangle_e + \frac{1}{\sqrt{2}}\hat{U}|\downarrow_x\rangle_e|\text{I}\rangle_e \\
&= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\text{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\text{III}\rangle_e.
\end{aligned}$$

Here we have written each step in the calculation so that the basic algebra is as clear as possible and to show how the linear operator $\hat{U}$ works. We will gradually adopt a more efficient notation.

The first thing to notice is that after the electron has passed through the box, it is no longer determinately z-spin up. But it is not that anything has happened to randomize its spin. So far the entire evolution of the state has been completely deterministic. Rather, it is that one can no longer represent the resultant state $|\psi(t_1)\rangle_e$ as the tensor product of z-spin up and something else. Hence, by rules 3 and 5, the electron is no longer determinately z-spin up.

The x-spin box puts the electron in a *superposition* of being x-spin up and in region II *and* being x-spin down and in region III. Such an electron is not determinately in region II, it is not determinately in region III, it is not determinately in both, and it is not determinately in neither. Indeed, since the x-spin box has correlated e's position with its spin and has left it in a state that cannot be factored into a tensor product of a spin state and a position state, it no longer even has a determinate spin or position *state* to call its own. When this happens, we say that the spin and the position of the electron are *entangled*. Entanglement is a decidedly nonclassical notion. It plays an essential role in how quantum mechanics predicts and explains a particularly subtle class of quantum phenomena.

The x-spin box does something similar to a z-spin down electron. Here the state at time t_0 is

$$|\psi(t_0)\rangle_e = |\downarrow_z\rangle_e|\mathrm{I}\rangle_e.$$

So by the linear dynamics, its state after one second is

$$\begin{aligned}
|\psi(t_1)\rangle_e &= \hat{U}|\psi(t_0)\rangle_e \\
&= \hat{U}|\downarrow_z\rangle_e|\mathrm{I}\rangle_e \\
&= \hat{U}\left[\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e - \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e\right]|\mathrm{I}\rangle_e \\
&= \hat{U}\left[\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{I}\rangle_e - \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{I}\rangle_e\right] \\
&= \frac{1}{\sqrt{2}}\hat{U}|\uparrow_x\rangle_e|\mathrm{I}\rangle_e - \frac{1}{\sqrt{2}}\hat{U}|\downarrow_x\rangle_e|\mathrm{I}\rangle_e \\
&= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{II}\rangle_e - \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{III}\rangle_e.
\end{aligned}$$

So the interaction leaves e in a *different* entangled superposition of being x-spin up and in region II and being x-spin down and in region III.

The linear operator $\hat{U}$ that represents the physical action of a z-spin box evolves states in an analogous way. In contrast to an x-spin box, a z-spin box does not do anything fancy to electrons that start in an eigenstate of z-spin—it just sends z-spin up electrons to region II and z-spin down electrons to region III. But, given this and the linearity of the dynamics, an electron that starts x-spin up

$$|\psi(t_0)\rangle_e = |\uparrow_x\rangle_e |\mathrm{I}\rangle_e$$

evolves as follows:

$$\begin{aligned}
|\psi(t_1)\rangle_e &= \hat{U}|\psi(t_0)\rangle_e \\
&= \hat{U}|\uparrow_x\rangle_e|\mathrm{I}\rangle_e \\
&= \hat{U}\left[\frac{1}{\sqrt{2}}|\uparrow_z\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e\right]|\mathrm{I}\rangle_e \\
&= \hat{U}\left[\frac{1}{\sqrt{2}}|\uparrow_z\rangle_e|\mathrm{I}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e|\mathrm{I}\rangle_e\right] \\
&= \frac{1}{\sqrt{2}}\hat{U}|\uparrow_z\rangle_e|\mathrm{I}\rangle_e + \frac{1}{\sqrt{2}}\hat{U}|\downarrow_z\rangle_e|\mathrm{I}\rangle_e \\
&= \frac{1}{\sqrt{2}}|\uparrow_z\rangle_e|\mathrm{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e|\mathrm{III}\rangle_e.
\end{aligned}$$

Because this state does not factor, the electron is no longer determinately x-spin up. Again, it does not even have determinate spin *state* to call its own. Rather, e is in an entangled superposition of being z-spin up and in region II *and* being z-spin down and in region III. And the z-spin box leaves an x-spin down electron in a similar entangled superposition.

We have been using the fact that if one knows what the unitary time-evolution operator $\hat{U}$ does to each element in a set of vectors that spans the possible initial states of a system S, then one knows what it does to every possible initial state. Since $\hat{U}$ is linear, one just writes the state of the system in terms of states with known evolutions, then time-evolves each term and keeps the coefficients with their corresponding terms. So if $\hat{U}|\phi(t_0)_k\rangle_S = |\phi(t_1)_k\rangle_S$, one carries out the transformation term by term to get the new time-evolved state

$$\begin{array}{llll}
|\psi(t_0)\rangle_S = a_1|\phi(t_0)_1\rangle_S & \cdots & + a_k|\phi(t_0)_k\rangle_S & \cdots \\
\qquad\qquad\quad \downarrow \hat{U} & & \quad \downarrow \hat{U} & \\
|\psi(t_1)\rangle_S = a_1|\phi(t_1)_1\rangle_S & \cdots & + a_k|\phi(t_1)_k\rangle_S & \cdots
\end{array}$$

One often knows how each element of a particular basis would evolve in a particular physical situation. When one does, one can simply rewrite the system's state in that basis, then calculate how the full state evolves by evolving each term of the state written in the new basis and keeping the coefficients. Here we knew what the spin boxes did to each of the elements of the corresponding spin basis because we knew how the boxes were designed. We will use this trick repeatedly to determine how physical systems will evolve under the linear dynamics.

4.3 Quantum Statistics

Consider again sending a z-spin up electron e through an x-spin box. The initial state is

$$|\psi(t_0)\rangle_e = |\uparrow_z\rangle_e|\mathrm{I}\rangle_e$$

and, on the linear dynamics, the final state is

$$|\psi(t_1)\rangle_e = \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{III}\rangle_e.$$

This is a superposition of being x-spin up and in region II *and* being x-spin down and in region III. Hence, given rules 3 and 5, one can no longer talk about e's spin at all without also talking about its position and one cannot talk about its position without talking about its spin. But when we do this experiment and check, we always find the electron with both a definite position and a definite spin after it passes through the box. About half of the time we find it in region II and x-spin up and about half the time we find it in region III and x-spin down. So, by the time we find the electron with determinate position and spin properties, it can no longer be in this entangled state. The challenge is to say *how* it ends up in an eigenstate of both position and x-spin and *when* this happens.

Whatever a *measurement* is, presumably an observer looking for an electron and either finding it or not counts. If so, then this means that if the electron is in the state $|\psi(t_1)\rangle_e$ and one observes its position, then the state of the electron must evolve according to the collapse dynamics (rule 4II). That is, the electron presumably must nonlinearly, randomly, and instantly collapse to an eigenstate of position the moment one looks for it. The probabilities associated with each possible measurement outcome are determined by the electron's state just before it was measured.

Let's start with a slightly simpler case, one where we are just considering the spin of the electron. Suppose one had a device that measured the x-spin of an electron directly. Suppose the electron is initially in the state

$$
\begin{aligned}
|\phi(t_1)\rangle_e &= \alpha|\uparrow_x\rangle_e + \beta|\downarrow_x\rangle_e \\
&= \alpha\begin{pmatrix}1\\0\end{pmatrix} + \beta\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}
\end{aligned}
$$

written in the x-spin basis, immediately before the measurement. Then rule 4II predicts that the probability that the post-measurement state $|\psi(t_2)\rangle_e$ will be $|\uparrow_x\rangle_e$ is

$$
|\langle\uparrow_x|\phi(t_1)\rangle|^2 = \left|\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right|^2 = |(1)(\alpha) + (0)(\beta)|^2 = |\alpha|^2
$$

and the probability that it will be $|\downarrow_x\rangle_e$ is

$$
|\langle\downarrow_x|\phi(t_1)\rangle|^2 = \left|\begin{pmatrix}0 & 1\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right|^2 = |(0)(\alpha) + (1)(\beta)|^2 = |\beta|^2.
$$

Again, the probability of each possible outcome is given by the norm-squared of the coefficient on the term in the initial state that corresponds to that outcome. The coefficient represents the quantum-mechanical *amplitude* associated with the outcome. And the probability of a particular result is equal to the norm-squared of the amplitude associated with that result.

Now consider the more complicated entangled state. If an electron is in the state

$$
|\psi(t_1)\rangle_e = \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{III}\rangle_e \tag{4.1}
$$

and we measure its position, there are two possible outcomes—it will be found to be in either region II or in region III. And each of these possibilities is associated with an amplitude of $1/\sqrt{2}$. Consequently, the collapse dynamics assigns each possibility a probability of $|1/\sqrt{2}|^2 = 1/2$ of being realized on measurement.

If the electron is found in region II, then it is determinately not in region III. So the collapse of the state multiplies the second term in $|\psi(t_1)\rangle_e$ by zero, and e ends up in the state

$$
|\psi(t_2)\rangle_e = |\uparrow_x\rangle_e|\mathrm{II}\rangle_e.
$$

Similarly, if e is found in region III, then it is determinately not in region II. So the collapse multiplies the first term in $|\psi(t_1)\rangle_e$ by zero, and e ends up in the state

$$
|\psi(t_2)\rangle_e = |\downarrow_x\rangle_e|\mathrm{III}\rangle_e.
$$

In the first case, the electron ends up x-spin up and in region II, and, in the second, x-spin down and in region III. In both cases, it now has both a determinate position *and* a determinate x-spin. The fact that the two quantities are perfectly correlated in the initial state ensures that if a collapse gives one of the quantities a determinate value it will also give the other the corresponding determinate value.

Now we know how the spin boxes work. On the assumption that they interact linearly with the electrons, an x-spin box just *correlates* the x-spin of the electron with its position. When one looks for the electron, one *measures* its position. It is *this* that gives the electron a determinate position which, in turn, gives it a determinate x-spin. Because its x-spin and position are correlated, one can read its x-spin from its position. On the assumption that boxes do not cause collapses, the random behavior of the electrons arises from the subsequent position measurement, not from the boxes. And the probabilities involved in the collapse process explain the standard quantum statistics.

Electrons that start z-spin down behave similarly. If the electron starts in state

$$|\psi(t_0)\rangle_e = |\downarrow_z\rangle_e|\mathrm{I}\rangle_e$$

and is sent through the x-spin box, it will end up in the state

$$|\psi(t_1)\rangle_e = \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{II}\rangle_e - \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{III}\rangle_e$$

by the linear dynamics. When its position is subsequently measured, it will end up in the state

$$|\psi(t_2)\rangle_e = |\uparrow_x\rangle_e|\mathrm{II}\rangle_e \tag{4.2}$$

with probability $|1/\sqrt{2}|^2 = 1/2$, and in state

$$|\psi(t_2)\rangle_e = |\downarrow_x\rangle_e|\mathrm{III}\rangle_e \tag{4.3}$$

with probability $|-1/\sqrt{2}|^2 = 1/2$. And z-spin boxes behave similarly on observation.

When the electron is in a physical state like $|\psi(t_1)\rangle_e$ that is simply represented by a vector in Hilbert space, we say that it is in a *pure state*. If the electron were already determinately in the state $|\uparrow_x\rangle_e|\mathrm{II}\rangle_e$ or determinately in the state $|\downarrow_x\rangle_e|\mathrm{III}\rangle_e$ (with equal probability) before it was observed, and we just didn't know which, one would say that it was in a *statistical mixture* of these two (pure) states. On the assumption that the boxes do not cause collapses, the linear dynamics predicts that, immediately before we observe the electron, the electron is in the pure state

$|\psi(t_1)\rangle_e$, *not* a statistical mixture of $|\uparrow_x\rangle_e|\text{II}\rangle_e$ and $|\downarrow_x\rangle_e|\text{III}\rangle_e$. In contrast, the collapse dynamics, if we believed that it applied to the interaction between the electron and the x-spin box, which we don't, would predict the statistical mixture of $|\uparrow_x\rangle_e|\text{II}\rangle_e$ and $|\downarrow_x\rangle_e|\text{III}\rangle_e$ with equal probability immediately before the electron is observed.

So, if we are right to believe that an electron follows the linear dynamics when it encounters an x-spin box (an assumption that we will return to in the next chapter), an x-spin box does not typically *sort* electrons by their x-spin. And it does not *randomize* their z-spin properties or anything else either. Rather, it just *correlates* an electron's spin with its position, which nearly always leads to an entangled superposition of spin and position properties. And it does so in a way that is fully deterministic. The electron's state only evolves randomly when a measurement is made. But, if this is right, then why does it *seem* that an intervening x-spin box randomizes z-spin *even when no one observes the electron*?

4.4 Combining Boxes

Suppose one knows that an electron e is z-spin up because one just observed it exiting from the $\uparrow_z$ door of a z-spin box. Consider what happens under the linear dynamics if one sends e through an x-spin box with a z-spin box positioned at the $\uparrow_x$ door (as in Figure 4.6). We need to keep track of five possible eigenstates of position, one for each of the five regions indicated by Roman numerals; so we need a five-dimensional vector space for the position part of the state. The spin part of the state is again represented by a vector in a two-dimensional vector space. We will suppose that e starts z-spin up in region I and takes one second to move through each box.

The state of the electron at time t_0 is

$$|\psi(t_0)\rangle_e = |\uparrow_z\rangle_e|\text{I}\rangle_e.$$

By the linear dynamics and how the x-spin box is designed, its state after one second is

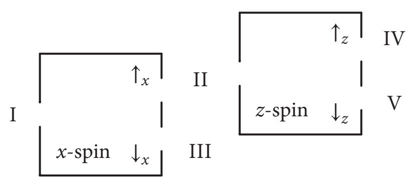


Figure 4.6. x-spin followed by z-spin with labeled positions.

$$
\begin{aligned}
|\psi(t_1)\rangle_e &= \hat{U}(t_0, t_1)|\psi(t_0)\rangle_e \\
&= \hat{U}(t_0, t_1)|\uparrow_z\rangle_e|\mathrm{I}\rangle_e \\
&= \hat{U}(t_0, t_1)\left[\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{I}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{I}\rangle_e\right] \\
&= \frac{1}{\sqrt{2}}\hat{U}(t_0, t_1)|\uparrow_x\rangle_e|\mathrm{I}\rangle_e + \frac{1}{\sqrt{2}}\hat{U}(t_0, t_1)|\downarrow_x\rangle_e|\mathrm{I}\rangle_e \\
&= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{III}\rangle_e.
\end{aligned}
$$

Again, we are going slowly in order to see how the theory works. Since the dynamics is linear, to calculate what happens in the next second we get the final state by considering what a z-spin box would do to each term of the state of the electron described by $|\psi(t_1)\rangle_e$ in the last line above. The z-spin box will not do anything to the second term as it describes an electron that does not even enter the box. So we just need to figure out what it will do to the first term.

While we do not know what a z-spin box will do to an x-spin up electron just from the construction of the box, we do know what it will do to a z-spin up or z-spin down electron. The trick is to rewrite this term in the z-spin basis, then time-evolve each term in the resulting expression. Since

$$
|\uparrow_x\rangle_e = \frac{1}{\sqrt{2}}|\uparrow_z\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e
$$

we can substitute this expression into the last line above to get

$$
\begin{aligned}
|\psi(t_1)\rangle_e &= \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|\uparrow_z\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e\right)|\mathrm{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{III}\rangle_e \\
&= \left(\frac{1}{2}|\uparrow_z\rangle_e + \frac{1}{2}|\downarrow_z\rangle_e\right)|\mathrm{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{III}\rangle_e \\
&= \frac{1}{2}|\uparrow_z\rangle_e|\mathrm{II}\rangle_e + \frac{1}{2}|\downarrow_z\rangle_e|\mathrm{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{III}\rangle_e.
\end{aligned}
$$

Now we know what the z-spin box would do to each of the terms in this expression. It would evolve the state represented by the first term to $|\uparrow_z\rangle_e|\mathrm{IV}\rangle_e$ and the state represented by the second term to $|\downarrow_z\rangle_e|\mathrm{V}\rangle_e$, and it would do nothing at all to the state represented by the third term. So

$$
\begin{aligned}
|\psi(t_2)\rangle_e &= \hat{U}(t_1, t_2)|\psi(t_1)\rangle_e \\
&= \hat{U}(t_1, t_2)\left(\frac{1}{2}|\uparrow_z\rangle_e|\mathrm{II}\rangle_e + \frac{1}{2}|\downarrow_z\rangle_e|\mathrm{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{III}\rangle_e\right)
\end{aligned}
$$

$$= \frac{1}{2}\hat{U}(t_1,t_2)|\uparrow_z\rangle_e|\mathrm{II}\rangle_e + \frac{1}{2}\hat{U}(t_1,t_2)|\downarrow_z\rangle_e|\mathrm{II}\rangle_e + \frac{1}{\sqrt{2}}\hat{U}(t_1,t_2)|\downarrow_x\rangle_e|\mathrm{III}\rangle_e$$

$$= \frac{1}{2}|\uparrow_z\rangle_e|\mathrm{IV}\rangle_e + \frac{1}{2}|\downarrow_z\rangle_e|\mathrm{V}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{III}\rangle_e.$$

And the linear dynamics leaves the electron in an entangled superposition of being z-spin up and in region IV *and* being z-spin down and in region V *and* being x-spin down and in region III.

If one looks for the electron in this state, its state will instantly and randomly collapse with probabilities determined by the amplitudes associated with each possible outcome. The probability of finding it in region IV and hence z-spin up is $|1/2|^2 = 1/4$. The probability of finding it in region V and hence z-spin down is $|1/2|^2 = 1/4$. And the probability of finding it in region III and hence x-spin down is $|1/\sqrt{2}|^2 = 1/2$.

And *these probabilities* are precisely what one would expect if the x-spin box completely randomized z-spin so that each box sent about half of the electrons out of each of its doors. But nothing random occurred when the electron encountered the x-spin box and interacted with it linearly. It just *appears* that something random occurred. That is how one would explain the statistics if one believed that each electron followed a determinate classical trajectory. But here the only random event is the collapse that occurs when one observes the electron's final position at the end of the experiment.

At this point, it is natural to wonder again why we are so sure that the electrons obey the linear dynamics and hence typically do not follow classical trajectories when they interact with the boxes. The empirical evidence for this is given by the results we get in the two-path experiments. We will discuss this in the next chapter but first we will end this chapter by considering two other quick examples of quantum-mechanical systems.

4.5 Physical Properties more Generally: *K* Mesons and Qubits

So far, we have been talking about the spin properties and positions of electrons, but quantum mechanics treats all physical properties in the same basic way. One represents a physical observable by finding a Hilbert space spanned by an orthonormal basis corresponding to the observable one seeks to represent. Each element of the basis will correspond to a different value that the observable might take. Different rigid rotations of one orthonormal basis will correspond to the various other physical properties that can be represented in the space. One's empirical evidence determines how one identifies these properties physically—the statistics of one's measurement results determines the angles between the properties.

The same vector space may represent different sets of related physical properties. The following examples illustrate how this works in the context of very different sets of physical properties.

Because of the way the standard theory treats properties, there is sometimes no simple matter of fact about what sort of particle a particular fundamental particle *is*. Neutral K mesons, like many other particles, have antiparticles with very different physical properties. A K^0 meson has a strangeness of $+1$, for example, while its antiparticle, a $\bar{K}^0$ meson, has a strangeness of -1.[12] Neutral K mesons are typically found to be in *superpositions* of being K^0 mesons and being $\bar{K}^0$ mesons at the same time. This means that on the standard interpretation of quantum-mechanical states, a neutral K meson is typically not a K^0 meson, not a $\bar{K}^0$ meson, not both, and not neither; rather, it is in a superposition of being both types of particle simultaneously.

The quantum-mechanical representation of neutral K meson states works very much like the representation of electron spin.[13] The state where a neutral K meson is determinately a K^0 particle is $|K^0\rangle$ and the state where it is determinately a $\bar{K}^0$ particle is $|\bar{K}^0\rangle$. These two states are mutually orthogonal and provide a basis for the space of possible neutral K meson particle states.

The K-short meson K_S has a mean time to decay of about 8.95×10^{-11} seconds and the K-long meson K_L has a mean time to decay of about 5.11×10^{-8} seconds. Each of these particles decays in a different way producing a different cascade of particles. K_S and K_L mesons can be represented as linear superpositions of determinate K^0 and $\bar{K}^0$ states

$$|K_S\rangle = \frac{1}{\sqrt{2}}(|K^0\rangle + |\bar{K}^0\rangle)$$

and

$$|K_L\rangle = \frac{1}{\sqrt{2}}(|K^0\rangle - |\bar{K}^0\rangle).$$

The geometric relationship between each of these states can be seen in Figure 4.7(a). The state of a particular neutral K meson will wobble about in this state space when the particle is unobserved. As it does, the probabilities of finding it as each type of particle if it is observed wobble with its state.[14]

[12] Strangeness, like spin, is a basic property of fundamental particles. It is conserved in strong and electromagnetic interactions but not in weak interactions.

[13] The present description follows Baym's (1969, 38–45) more detailed account.

[14] The behavior of neutral K mesons and particle decay and creation more generally is among the phenomena that a satisfactory formulation of quantum mechanics must be able to explain. As we will see, some of the formulations of quantum mechanics devised to address the quantum measurement

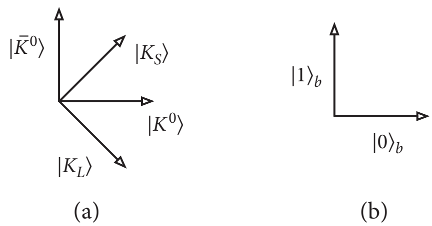


Figure 4.7. State representations for *K* mesons and qubits.

A qubit in a quantum computer might also be represented by vectors in such a space (as in Figure 4.7(b)). A bit in a classical computer can take the value 0 or 1. In contrast, a qubit b might take any superposition of the two orthogonal states $|0\rangle_b$ and $|1\rangle_b$

$$\alpha|0\rangle_b + \beta|1\rangle_b$$

where α and β are complex numbers.

The qubits in a quantum computer evolve according to linear dynamics rule 4I when unobserved. This allows one to execute a superposition of different classical operations on the qubits. Since the dynamics is linear, classical operations on a superposition of inputs will produce a superposition of classical results. If one added the number n to a superposition of the two numbers j and k, for example, one would get a superposition of $j + n$ and $k + n$ as the result. In this case, there is a sense in which one has performed two calculations in a single computational step. One might, at least in principle, carry out an *infinite* superposition of different calculations similarly. The aspirational idea behind a quantum computer is to exploit such massively parallel processing. The problem is that if one simply reads the result of a parallel calculation, one collapses it to a particular randomly selected classical calculation. And performing a single, randomly determined calculation is clearly worse than just doing the single classical calculation one cares about most.

There are, however, a few clever quantum algorithms that get around this problem. While no quantum algorithm can solve a problem beyond what a classical computer could solve if given enough time (that is, on the standard way of setting them up, quantum computers are Turing-equivalent machines), there are quantum algorithms that allow for faster computations than any known classical algorithm. The best example of a useful quantum algorithm is Peter Shor's (1996) algorithm for factoring a number into its composite prime factors in polynomial time. No known classical algorithm can factor a number this quickly.

problem (Chapter 7) account for the spin behavior of electrons more easily than they do field phenomena like particle annihilation and creation.

5

Quantum Interference

5.1 The Simple Two-Path Experiment

We are now in a position to see how the standard formulation of quantum mechanics explains the results of the two-path interference experiments. Consider the two-path experiment where we start with a z-spin up electron e in region I at time t_0 (as in Figure 5.1). We will suppose that the electron is halfway through the apparatus in one second t_1 and at region II in two seconds t_2 and that the changes in direction as it travels do not affect its spin properties. We will also suppose that no one looks at either of the two paths during the experiment and that e is kept well-isolated from its environment. Finally, we will continue to assume that the interaction between the electron and the spin boxes is correctly characterized by the deterministic linear dynamics. That is, we will continue to suppose that this interaction does not count as a measurement that collapses the state of the electron.

Before we calculate what happens with a z-spin up electron, let's consider what the two-path apparatus does to x-spin up and x-spin down electrons. An x-spin up electron would travel through the device like this

$$
\begin{gathered}
|\psi(t_0)\rangle_e = |\uparrow_x\rangle_e |\mathrm{I}\rangle_e \\
\downarrow \\
|\psi(t_1)\rangle_e = |\uparrow_x\rangle_e |A\rangle_e \\
\downarrow \\
|\psi(t_2)\rangle_e = |\uparrow_x\rangle_e |\mathrm{II}\rangle_e.
\end{gathered}
$$

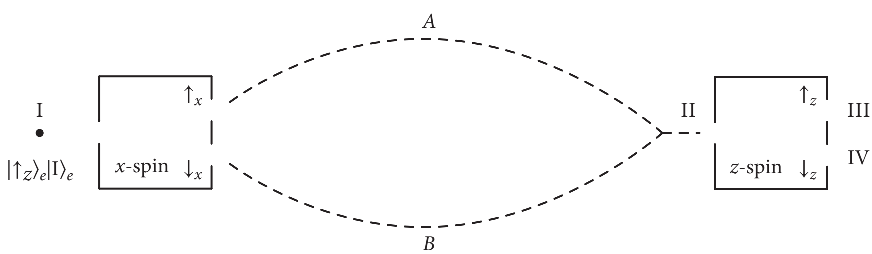


Figure 5.1. Two-path experiment with labeled positions.

And an x-spin down electron would travel through the device like this

$$\begin{gathered}|\psi(t_0)\rangle_e = |\downarrow_x\rangle_e|\mathrm{I}\rangle_e \\ \downarrow \\ |\psi(t_1)\rangle_e = |\downarrow_x\rangle_e|B\rangle_e \\ \downarrow \\ |\psi(t_2)\rangle_e = |\downarrow_x\rangle_e|\mathrm{II}\rangle_e.\end{gathered}$$

Since any spin state of an electron in region I can be represented as a linear combination of $|\uparrow_x\rangle_e|\mathrm{I}\rangle_e$ and $|\downarrow_x\rangle_e|\mathrm{I}\rangle_e$, the dispositions of an electron in each of these two x-spin eigenstates fully determines the unitary time-evolution of any electron that starts in region I. Since the evolution is linear, we just write the initial state in the x-spin basis, then time-evolve each term in the superposition, keeping the original coefficients.

The state of the z-spin up electron at time t_0 is

$$|\psi(t_0)\rangle_e = |\uparrow_z\rangle_e|\mathrm{I}\rangle_e.$$

So, given the linearity of the dynamics and how the x-spin box is designed, its state after one second is

$$\begin{aligned}|\psi(t_1)\rangle_e &= \hat{U}(t_0,t_1)|\psi(t_0)\rangle_e \\ &= \hat{U}(t_0,t_1)|\uparrow_z\rangle_e|\mathrm{I}\rangle_e \\ &= \hat{U}(t_0,t_1)\left[\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{I}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{I}\rangle_e\right] \\ &= \frac{1}{\sqrt{2}}\hat{U}(t_0,t_1)|\uparrow_x\rangle_e|\mathrm{I}\rangle_e + \frac{1}{\sqrt{2}}\hat{U}(t_0,t_1)|\downarrow_x\rangle_e|\mathrm{I}\rangle_e \\ &= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e.\end{aligned}$$

Here the electron is in an entangled superposition of being x-spin up and on path A *and* being x-spin down and on path B, and hence does not have any determinate spin or position properties.

Given the linearity of the dynamics and the construction of the two-path apparatus, the electron's state after two seconds is

$$\begin{aligned}|\psi(t_2)\rangle_e &= \hat{U}(t_1,t_2)|\psi(t_1)\rangle_e \\ &= \hat{U}(t_1,t_2)\left(\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e\right)\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2}}\hat{U}(t_1, t_2)|\uparrow_x\rangle_e|A\rangle_e + \frac{1}{\sqrt{2}}\hat{U}(t_1, t_2)|\downarrow_x\rangle_e|B\rangle_e \\
&= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\text{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\text{II}\rangle_e \\
&= \left(\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e\right)|\text{II}\rangle_e \\
&= |\uparrow_z\rangle_e|\text{II}\rangle_e.
\end{aligned}
$$

So, putting the pieces together, the full time-evolution of the electron under the linear dynamics is

$$
\begin{aligned}
|\psi(t_0)\rangle_e &= |\uparrow_z\rangle_e|\text{I}\rangle_e \\
&\downarrow \\
|\psi(t_1)\rangle_e &= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e \\
&\downarrow \\
|\psi(t_2)\rangle_e &= |\uparrow_z\rangle_e|\text{II}\rangle_e.
\end{aligned}
$$

And after two seconds the electron is simply z-spin up and in region II. And that's why, when no one looks, it exhibits the sure-fire disposition to exit from the $\uparrow_z$ door of the z-spin box. It *is* z-spin up.[1]

The *interference effect* of the electron being determinately z-spin up at the end of the two-path apparatus is also how we know that it evolves by the linear dynamics (rule 4I) and not by the collapse dynamics (rule 4II) when it interacts with an x-spin box. For e to be z-spin up in region II on the standard theory it *must* have traveled a *superposition* of the two paths, as predicted by the linear dynamics. While it is in a superposition of being on path A and being on path B, it does not have any determinate spin or position properties because its spin and position properties are entangled. This is represented by the fact that the electron's state after one second cannot be factored into a tensor product of a spin state and a position state. But after two seconds, its state *can* be factored into a product of spin and position. And it is just a z-spin up electron in region II.

5.2 Measurement

Suppose that we run the same two-path experiment but we put an observer on path B to see whether the electron follows that path (as in Figure 5.2). Again, the electron e starts in state

[1] And if the electron is initially z-spin down, a similar narrative explains why it will certainly end up z-spin down at the end of the two-path apparatus.

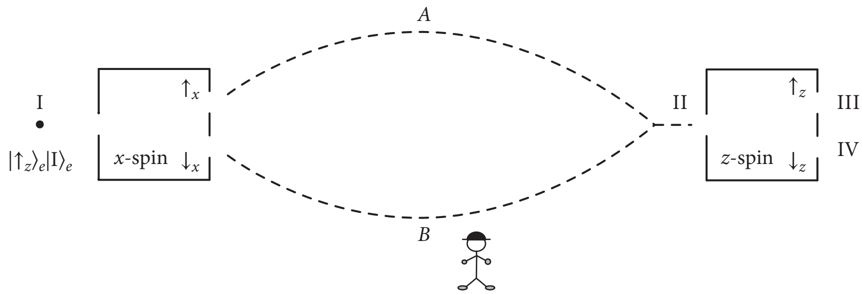


Figure 5.2. Two paths with an observer.

$$|\psi(t_0)\rangle_e = |\uparrow_z\rangle_e|\mathrm{I}\rangle_e$$

and by the linear dynamics ends up in the state

$$|\psi(t_1)\rangle_e = \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e$$

after it has passed through the x-spin box.

While this is not a state where the electron has a determinate position, it is a state where an observer will find it on one or the other of the two paths *if she looks for it*. And an observer intentionally looking for the electron to determine its position presumably counts as a *measurement* of the electron's position. In which case, the collapse dynamics (rule 4II) tells us that the state of the electron will randomly and nonlinearly collapse to an eigenstate of position with probabilities determined by the norm-squared of the projection of the initial state $|\psi(t_1)\rangle_e$ onto each possible state where it has a determinate position.

Suppose that the observer looks for the electron on path B. The probability that she will find it is the norm-squared of the amplitude corresponding to it being on path B, which in this case is $|1/\sqrt{2}|^2 = 1/2$. If she finds it, then the probability of the electron being on path A is zero and hence the amplitude associated with it being on path A must also be zero. So if the electron is found on path B, then it is also x-spin down, and its quantum state is $|\downarrow_x\rangle_e|B\rangle_e$. That is, the effect of finding the electron on path B is to multiply the first term of $|\psi(t_1)\rangle_e$ by zero. One then renormalizes the vector to get the resultant state.

The probability that the observer will *not* find the electron on path B is one minus the probability that she will: $1 - 1/2 = 1/2$. If she does not find it on path B, then the probability of it being on path B is zero and hence the amplitude associated with it being on path B must also be zero. This means that if it is not found on path B, then the state $|\psi(t_1)\rangle_e$ must collapse to $|\uparrow_x\rangle_e|A\rangle_e$ when the observer looks for the electron and does not find it. So the effect of not finding the electron on path B is to multiply the second term of $|\psi(t_1)\rangle_e$ by zero.

In short, the collapse dynamics together with the standard interpretation of states requires that finding the electron on path *B* collapses the state to $|\downarrow_x\rangle_e|\mathrm{B}\rangle_e$, and not finding it on path *B* collapses the state to $|\uparrow_x\rangle_e|A\rangle_e$. This is what explains why if one finds the electron on path *B*, then it is *x*-spin down. And it explains why if one does not find it on path *B*, then one will find it on path *A* if one looks for it there and it will be *x*-spin up.

If the observer finds the electron on path *B* then allows it to continue moving through the apparatus, the linear dynamics predicts that the *x*-spin down electron will end up in the symmetric superposition of exiting from each of the *z*-spin box doors

$$\frac{1}{\sqrt{2}}|\uparrow_z\rangle_e|\mathrm{III}\rangle_e - \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e|\mathrm{IV}\rangle_e.$$

Then if one looks for it again, one will collapse the electron to an eigenstate of position with probabilities given by the norm-squared of the amplitudes associated with each possible outcome and hence, in this case, find it in region III and *z*-spin up with probability $|1/\sqrt{2}|^2 = 1/2$ and find it in region IV and *z*-spin down with probability $|-1/\sqrt{2}|^2 = 1/2$.

Similarly, if the observer does *not* find the electron on path *B*, then the collapse has put it on path *A* and *x*-spin up, so the linear dynamics predicts that its state will be

$$\frac{1}{\sqrt{2}}|\uparrow_z\rangle_e|\mathrm{III}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e|\mathrm{IV}\rangle_e$$

after passing through the *z*-spin box. Consequently, one will also find it in region III and *z*-spin up with probability 1/2 and in region IV and *z*-spin down with probability 1/2.

Again, if the observer does not look at path *B*, then one will *always* find the electron in region III and *z*-spin up. The observer changes the state of the electron by looking for it *even if she does not find it*. If she does not look for it, the state of the electron just before it enters the *z*-spin box will be

$$|\uparrow_z\rangle_e|\mathrm{II}\rangle_e.$$

But if she does look for it, she causes it to collapse to an eigenstate of the position, even if she does not see it, and the electron ends up in a statistical mixture of being either determinately *x*-spin up

$$|\uparrow_x\rangle_e|\mathrm{II}\rangle_e$$

with probability 1/2 or determinately x-spin down

$$|\downarrow_x\rangle_e|\text{II}\rangle_e$$

also with probability 1/2 just before it enters the z-spin box. And in each of these two cases, the electron will end up in a superposition of exiting from each of the z-spin doors.

There is something nonlocal about the observer looking in one place, not finding the electron there, but affecting its dispositions somewhere else. But the situation here is subtle. As the theory tells it, it is not that the state of a distant electron is affected by the observer looking at the empty path. Rather, it is that the observer's looking at path B and not finding the electron there instantaneously causes the electron to in fact be on path A when there was no such determinate matter of fact before the observer looked. To be sure, looking for it affects the state of the electron, but, since it is in an entangled state without any determinate position (or even a determinate position *state*) before the observer looks, the action is not one that affects the state of an object that is *determinately* distant. Rather, it gives an object with no determinate position at all a determinate distant position.

The measurement does not simply reveal that the electron is not on the path that the observer is watching. The electron was decidedly *not* determinately on path A before she looked. If it had been, then it would also have had a determinate x-spin and, hence, would not exhibit the sure-fire disposition to exit from the $\uparrow_z$ door of the z-spin box had the observer not looked. The measurement does not reveal a preexisting determinate fact, it *makes* the fact determinate.

For looking at a path and not seeing anything to count as a measurement, the observer must be looking at a time when the theory predicts that she might find the electron. That is, there must be a nonzero amplitude of the electron being in the region that she is checking when she is checking it. Then, even when she does not find it, the effect of the measurement is to collapse the state to one where the electron is determinately not where she looked.

Such a collapse gives the same posterior probabilities for the result of subsequent measurements as conditioning on new evidence. Bayes' theorem tells us that the probability of the electron being found in region R given that it is not found in the disjoint region S is

$$P(R|\neg S) = \frac{P(R)}{1 - P(S)},$$

where $P(R)$ and $P(S)$ are the prior probabilities of finding e in R and S as given by rule 4II. So, by the principle of strict conditionalization, the posterior probability of finding e in R after not finding it in S is

$$P'(R) = \frac{P(R)}{1 - P(S)}.$$

5.3 Barriers

The theory tells us that something subtly different happens if one puts a physical barrier on path B in the place of an observer (as in Figure 5.3). Suppose that the electron starts z-spin up in region I at time t_0. Suppose also that the barrier is gentle so that if the electron hits it, it just slows the electron to a stop leaving it on path B and not disturbing its spin properties. Suppose that at one second t_1 the electron is on the paths, at two seconds it has had time to get to region II, and at three seconds t_3 it has had time to get through the z-spin box. Finally, suppose that the electron evolves in the standard linear, deterministic manner all the way through the apparatus.

With this setup, an electron that started x-spin up would make it all the way through the apparatus evolving as follows:

$$\begin{aligned}
|\psi(t_0)\rangle_e &= |\uparrow_x\rangle_e|\text{I}\rangle_e \\
&\downarrow \\
|\psi(t_1)\rangle_e &= |\uparrow_x\rangle_e|A\rangle_e \\
&\downarrow \\
|\psi(t_2)\rangle_e &= |\uparrow_x\rangle_e|\text{II}\rangle_e \\
&\downarrow \\
|\psi(t_3)\rangle_e &= \frac{1}{\sqrt{2}}|\uparrow_z\rangle_e|\text{III}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e|\text{IV}\rangle_e.
\end{aligned}$$

And an electron that started x-spin down would be stopped by the barrier on path B and, hence, evolve as follows:

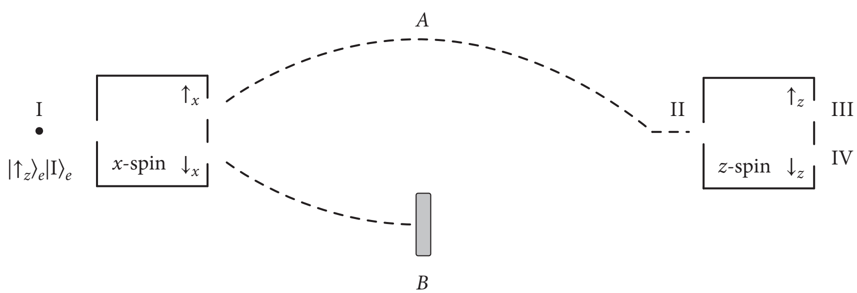


Figure 5.3. Two paths with a barrier.

$$|\psi(t_0)\rangle_e = |\downarrow_x\rangle_e|\mathrm{I}\rangle_e$$
$$\downarrow$$
$$|\psi(t_1)\rangle_e = |\downarrow_x\rangle_e|B\rangle_e$$
$$\downarrow$$
$$|\psi(t_2)\rangle_e = |\downarrow_x\rangle_e|B\rangle_e$$
$$\downarrow$$
$$|\psi(t_3)\rangle_e = |\downarrow_x\rangle_e|B\rangle_e.$$

Again, these two dispositions fully characterize the interaction between an electron and two-path apparatus.

It follows by the linear dynamics and the design of the apparatus that if the state of the electron at time t_0 is

$$|\psi(t_0)\rangle_e = |\uparrow_z\rangle_e|\mathrm{I}\rangle_e,$$

then its state after one second is

$$\begin{aligned}
|\psi(t_1)\rangle_e &= \hat{U}(t_0, t_1)|\psi(t_0)\rangle_e \\
&= \hat{U}(t_0, t_1)|\uparrow_z\rangle_e|\mathrm{I}\rangle_e \\
&= \hat{U}(t_0, t_1)\left[\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{I}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{I}\rangle_e\right] \\
&= \frac{1}{\sqrt{2}}\hat{U}(t_0, t_1)|\uparrow_x\rangle_e|\mathrm{I}\rangle_e + \frac{1}{\sqrt{2}}\hat{U}(t_0, t_1)|\downarrow_x\rangle_e|\mathrm{I}\rangle_e \\
&= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e.
\end{aligned}$$

Here the electron is in an entangled superposition of being on path A and x-spin up and being on path B and x-spin down. After two seconds, its state is

$$\begin{aligned}
|\psi(t_2)\rangle_e &= \hat{U}(t_1, t_2)|\psi(t_1)\rangle_e \\
&= \hat{U}(t_1, t_2)\left(\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e\right) \\
&= \frac{1}{\sqrt{2}}\hat{U}(t_1, t_2)|\uparrow_x\rangle_e|A\rangle_e + \frac{1}{\sqrt{2}}\hat{U}(t_1, t_2)|\downarrow_x\rangle_e|B\rangle_e \\
&= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e.
\end{aligned}$$

Which means that the electron is now in a superposition of being in region II and x-spin up (because it took path A) and still being on path B and x-spin down (because the barrier stopped it).

Note that, unlike the situation after two seconds in the original two-path experiment, this state does not factor into a tensor product of a state where the electron has a determinate spin and a state where it has a determinate position. The effect of the barrier is to keep the electron's spin and position properties entangled. And this changes entirely the way the electron behaves when it encounters the z-spin box.

By the linear dynamics, the state of the electron after three seconds is

$$\begin{aligned}|\psi(t_3)\rangle_e &= \hat{U}(t_2, t_3)|\psi(t_2)\rangle_e \\ &= \hat{U}(t_2, t_3)\left(\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\text{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e\right).\end{aligned}$$

The z-spin box will clearly do nothing to the second term that describes the electron as still being on path B. But it will affect the first term.

We can calculate what the apparatus does to the first term by rewriting just that term in the z-spin basis then time-evolving each term in the expression:

$$\begin{aligned}|\psi(t_3)\rangle_e &= \hat{U}(t_2, t_3)\left[\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|\uparrow_z\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e\right)|\text{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e\right] \\ &= \hat{U}(t_2, t_3)\left[\frac{1}{2}|\uparrow_z\rangle_e|\text{II}\rangle_e + \frac{1}{2}|\downarrow_z\rangle_e|\text{II}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e\right] \\ &= \frac{1}{2}|\uparrow_z\rangle_e|\text{III}\rangle_e + \frac{1}{2}|\downarrow_z\rangle_e|\text{IV}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e.\end{aligned}$$

So with a barrier on path B the electron ends up in an entangled superposition of being in region III and z-spin up, being in region IV and z-spin down, and being on path B and x-spin down. Given the amplitudes associated with each term, when one looks for the electron, one will cause it to collapse to an eigenstate of position, and one will find it in region III and z-spin up with probability $|1/2|^2 = 1/4$, in region IV and z-spin down with probability $|1/2|^2 = 1/4$, and on path B and x-spin down with probability $|1/\sqrt{2}|^2 = 1/2$.

The presence of an *observer* on path B would prevent an electron that starts z-spin up in region I from ending up determinately z-spin up in region II by collapsing its state to one where it is either determinately x-spin up and on path A or determinately x-spin down and on path B. In contrast, the presence of a *barrier* on path B prevents the electron from ending up determinately z-spin up in region II by keeping the electron's spin and position properties from disentangling. The interference effect of finding a z-spin up electron in region II is destroyed in each case but for completely different reasons.[2]

It is important to be clear about this. The two ways of destroying the interference effect here involve entirely different sequences of physical states exhibiting radically different causal stories. The effect of the observer is to cause a nonlinear, random collapse of the quantum-mechanical state on measurement. This produces a determinate measurement result and leads to the statistical mixture of either

$$|\text{up}\rangle_e = |\uparrow_x\rangle_e |\text{II}\rangle_e \tag{5.1}$$

or

$$|\text{down}\rangle_e = |\downarrow_x\rangle_e |\text{II}\rangle_e \tag{5.2}$$

with probability 1/2 for each at time t_2. The effect of the barrier on the electron is more subtle and involves nothing nonlinear or random.

One might think of the barrier as producing a *decoherence effect* involving the *internal* degrees of freedom of the electron. To say what this means, it will be useful to first understand *environmental decoherence*, something we will turn to in the next section. For now, note that, on the assumption that the barrier just stops the electron, the standard theory predicts the following *pure state* for time t_2 when there is a barrier instead of an observer:

$$|\text{super}\rangle_e = \frac{1}{\sqrt{2}} |\uparrow_x\rangle_e |\text{II}\rangle_e + \frac{1}{\sqrt{2}} |\downarrow_x\rangle_e |\text{B}\rangle_e. \tag{5.3}$$

This state is completely different in character from each of the two states that might evolve by the collapse dynamics when an observation is made. While all three states exhibit the same unbiased statistics for electrons that make it through the z-spin box, they are different physical states with different empirical properties. An electron in state $|\text{up}\rangle_e$ will be found to be x-spin up, an electron in state $|\text{down}\rangle_e$ will be found to be x-spin down, and an electron in state $|\text{super}\rangle_e$ will not exhibit a determinate x-spin but will have sure-fire interference properties that are not shared by an electron in state $|\text{up}\rangle_e$ or $|\text{down}\rangle_e$.

There are other things that would act like the barrier here to prevent the interference effect of the electron being z-spin up in region II. Suppose that we do not stop the electron, but just slow it down on path B so that if it were to take that path determinately, it would get to region II later than if it had taken path A. In that case, an electron that took path A would be in region II at a time when

[2] The final state with the barrier here is very much like the state one gets when one sends an electron through an x-spin followed directly by a z-spin box with no observation between the two boxes. Both of these cases might be thought of as examples of *internal decoherence*.

an electron that took path *B* would still be on path *B*, so the state predicted by the linear dynamics would take precisely the same form as $|\text{super}\rangle_e$. It is just that the second term might represent the electron as being a bit further along path *B*. One would also destroy the interference effect if one made one of the paths longer than the other in order to determine which path the electron traveled to get to region II. To get the interference effect both components of the state must describe the electron as being in region II at the same time; otherwise, the state does not factor as the electron being *z*-spin up and in region II.

One could randomly choose which path to put the barrier in *after* the electron has passed the *x*-spin box. This is called a *delayed-choice experiment*. Given the discussion so far, it should be clear that all that matters for destroying the interference effect is that a barrier be in one or the other of the two paths at a time such that it being there then would prevent a state that can be factored into spin and position from evolving.

In the case of a delayed-choice experiment one might worry that the electrons that get to the *z*-spin box are affected by the presence of the barrier on *the path that they did not take* when they must have "decided which path to take" before the barrier was even in place and "never even got close" to the barrier. But, as the standard theory has it, the situation is significantly less intuitive than such worries suggest. The electron that one finds when one looks for it at the *z*-spin box did not get there by determinately following the path without the barrier. Indeed, according to the theory, the electron was not determinately *there* before one looked for it. Rather, it was in an entangled superposition of being at the *z*-spin box and at the barrier.

When one imagines that an electron that gets through the apparatus must have determinately followed a barrier-free trajectory, one is appealing to intuitions that the theory tells us are simply mistaken. And the presence of interference effects when there is no barrier provides good empirical reason to believe that the theory is getting things right.

5.4 Decoherence

Suppose that, rather than directly looking to see which path the electron travels, we set something up to record the trajectory it took. This might be done by placing a camera on one of the paths. Or, more simply, we might just put a single particle *p* on path *B* and arrange it so that it moves from region *a* to region *b*, without disturbing *e*'s motion, if and only if the electron takes path *B* (as in Figure 5.4). Here the record of the electron's trajectory is made in terms of particle *p*'s position.

While it is perhaps less clear whether taking a photograph with a macroscopic camera should count as a measurement in the theory, we have very good empirical evidence that interactions between fundamental particles are accurately described

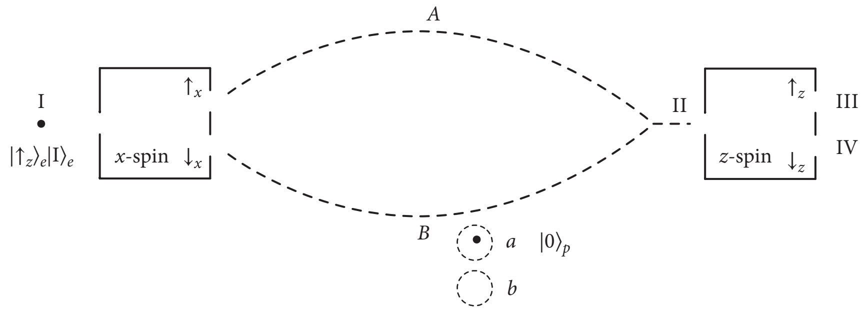


Figure 5.4. Two paths with a recording particle.

by the linear dynamics. Hence, we will suppose that the evolution of the state of the electron and recording particle are described by the linear dynamics.

This is the first time we have used the theory to describe a *composite* quantum-mechanical system. We will continue to use the space we have been using to represent the state of the electron. In addition, we need to keep track of the position of the recording particle. For the purposes at hand, the position of the recording particle can be represented by a vector in a two-dimensional space spanned by the state $|a\rangle_p$ where p is determinately in region a and the state $|b\rangle_p$ where it is determinately in region b. The state of the composite two-particle system then is represented by a vector in the tensor product of these two spaces.

We will suppose that the electron starts in region I at time t_0, is through the x-spin box in a half-second $t_{1/2}$, has had time to interact with particle p by three halves of a second $t_{3/2}$, is at region II in two seconds t_2, and is through the z-spin box by three seconds t_3. We will start by considering what the apparatus does over the first two seconds for electrons that start in an eigenstate of x-spin.

If the electron starts x-spin up and in region I and the recording particle p starts in region a, then the state of the composite system evolves as follows

$$
\begin{aligned}
|\psi(t_0)\rangle_{ep} &= |\uparrow_x\rangle_e|\text{I}\rangle_e|a\rangle_p \\
&\downarrow \\
|\psi(t_{1/2})\rangle_{ep} &= |\uparrow_x\rangle_e|A\rangle_e|a\rangle_p \\
&\downarrow \\
|\psi(t_{3/2})\rangle_{ep} &= |\uparrow_x\rangle_e|A\rangle_e|a\rangle_p \\
&\downarrow \\
|\psi(t_2)\rangle_{ep} &= |\uparrow_x\rangle_e|\text{II}\rangle_e|a\rangle_p.
\end{aligned}
$$

In this case, particle p stays in region a without moving, indicating that the electron took path A and is hence x-spin up.

If the electron starts x-spin down and in region I and the recording particle p starts in region a, then the state of the composite system would evolve as follows

$$
\begin{aligned}
|\psi(t_0)\rangle_{ep} &= |\downarrow_x\rangle_e |\mathrm{I}\rangle_e |a\rangle_p \\
&\downarrow \\
|\psi(t_{1/2})\rangle_{ep} &= |\downarrow_x\rangle_e |B\rangle_e |a\rangle_p \\
&\downarrow \\
|\psi(t_{3/2})\rangle_{ep} &= |\downarrow_x\rangle_e |B\rangle_e |b\rangle_p \\
&\downarrow \\
|\psi(t_2)\rangle_{ep} &= |\downarrow_x\rangle_e |\mathrm{II}\rangle_e |b\rangle_p .
\end{aligned}
$$

Here particle p moves from a to b, indicating that the electron took path B and is x-spin down.

So if the composite system begins in the state

$$
\begin{aligned}
|\psi(t_0)\rangle_{ep} &= |\uparrow_z\rangle_e |\mathrm{I}\rangle_e |a\rangle_p \\
&= \left(\frac{1}{\sqrt{2}} |\uparrow_x\rangle_e + \frac{1}{\sqrt{2}} |\downarrow_x\rangle_e \right) |\mathrm{I}\rangle_e |a\rangle_p \\
&= \frac{1}{\sqrt{2}} |\uparrow_x\rangle_e |\mathrm{I}\rangle_e |a\rangle_p + \frac{1}{\sqrt{2}} |\downarrow_x\rangle_e |\mathrm{I}\rangle_e |a\rangle_p
\end{aligned}
$$

it will evolve as follows on the linear dynamics

$$
\begin{aligned}
|\psi(t_0)\rangle_{ep} &= \frac{1}{\sqrt{2}} |\uparrow_x\rangle_e |\mathrm{I}\rangle_e |a\rangle_p + \frac{1}{\sqrt{2}} |\downarrow_x\rangle_e |\mathrm{I}\rangle_e |a\rangle_p \\
&\downarrow \\
|\psi(t_{1/2})\rangle_{ep} &= \frac{1}{\sqrt{2}} |\uparrow_x\rangle_e |A\rangle_e |a\rangle_p + \frac{1}{\sqrt{2}} |\downarrow_x\rangle_e |B\rangle_e |a\rangle_p \\
&\downarrow \\
|\psi(t_{3/2})\rangle_{ep} &= \frac{1}{\sqrt{2}} |\uparrow_x\rangle_e |A\rangle_e |a\rangle_p + \frac{1}{\sqrt{2}} |\downarrow_x\rangle_e |B\rangle_e |b\rangle_p \\
&\downarrow \\
|\psi(t_2)\rangle_{ep} &= \frac{1}{\sqrt{2}} |\uparrow_x\rangle_e |\mathrm{II}\rangle_e |a\rangle_p + \frac{1}{\sqrt{2}} |\downarrow_x\rangle_e |\mathrm{II}\rangle_e |b\rangle_p .
\end{aligned}
$$

The final state here is a superposition of the electron being x-spin up and in region II and p being in region a *and* the electron being x-spin down and in region II and p being in region b. In this state the electron is in fact determinately in region II, and, keeping in mind the physical interpretation of the tensor product as logical *and*, we could factor out its position if we wanted. But the interaction between e and p has left the electron's x-spin *entangled* with particle p's position. Here the electron no longer even has a spin *state* to call its own, nor does particle p have a position state of its own—while we initially had the choice of representing e's spin properties or p's position properties independently or together, now they can only be represented together.

To see what happens, we will write the state of the composite system at time t_2 in terms of e's z-spin basis as

$$
\begin{aligned}
|\psi(t_2)\rangle_{ep} &= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\text{II}\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\text{II}\rangle_e|b\rangle_p \\
&= \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|\uparrow_z\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e\right)|\text{II}\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|\uparrow_z\rangle_e - \frac{1}{\sqrt{2}}|\downarrow_z\rangle_e\right)|\text{II}\rangle_e|b\rangle_p \\
&= \frac{1}{2}|\uparrow_z\rangle_e|\text{II}\rangle_e|a\rangle_p + \frac{1}{2}|\downarrow_z\rangle_e|\text{II}\rangle_e|a\rangle_p + \frac{1}{2}|\uparrow_z\rangle_e|\text{II}\rangle_e|b\rangle_p - \frac{1}{2}|\downarrow_z\rangle_e|\text{II}\rangle_e|b\rangle_p.
\end{aligned}
$$

We know how each term in this expression will evolve when the electron encounters the z-spin box. By the linearity of the dynamics

$$
\begin{aligned}
|\psi(t_2)\rangle_{ep} &= \frac{1}{2}|\uparrow_z\rangle_e|\text{II}\rangle_e|a\rangle_p + \frac{1}{2}|\downarrow_z\rangle_e|\text{II}\rangle_e|a\rangle_p + \frac{1}{2}|\uparrow_z\rangle_e|\text{II}\rangle_e|b\rangle_p - \frac{1}{2}|\downarrow_z\rangle_e|\text{II}\rangle_e|b\rangle_p \\
&\downarrow \\
|\psi(t_3)\rangle_{ep} &= \frac{1}{2}|\uparrow_z\rangle_e|\text{III}\rangle_e|a\rangle_p + \frac{1}{2}|\downarrow_z\rangle_e|\text{IV}\rangle_e|a\rangle_p + \frac{1}{2}|\uparrow_z\rangle_e|\text{III}\rangle_e|b\rangle_p - \frac{1}{2}|\downarrow_z\rangle_e|\text{IV}\rangle_e|b\rangle_p.
\end{aligned}
$$

Here the probability of the electron being found in region III on measurement is the sum of the probabilities of the state represented by the first and third terms in this expression being realized: $|1/2|^2 + |1/2|^2 = 1/4 + 1/4 = 1/2$. And the probability of it being found in region IV on measurement is the sum of the probabilities of the state represented by the second and fourth terms being realized: $|1/2|^2 + |-1/2|^2 = 1/4 + 1/4 = 1/2$.[3]

So *the very presence* of a recording particle p whose position becomes correlated with the position of the electron destroys the interference effect in region II (of e exhibiting the sure-fire disposition of ending up in region III when it starts z-spin up) regardless of whether one ultimately finds p in region a or b. And it does this without causing a collapse of the quantum-mechanical state.

[3] Note that observing the position of p in this state does not give the electron a determinate position. Rather, it leaves the electron in a symmetric superposition of being in region II and being in region III. More generally, a collapse to an eigenstate of one observable only makes another observable determinate if the two observables were perfectly correlated in the initial state. See the discussion of quantum records in the next section for further details regarding what one can infer regarding the state of e from an observation of p. It turns out that much depends on *when* one observes p.

The situation here is similar to the two-path experiment with the barrier on path B. Both are examples of decoherence. While the electron does determinately get to region II in two seconds in the environmental decoherence experiment we just considered, the interaction between the electron and the recording particle p leaves the electron's spin properties entangled with p's position so one cannot factor out a determinate spin state for the electron. This happens if the state of anything whatsoever in the electron's environment or anywhere else becomes either intentionally or unintentionally correlated to its x-spin as it moves through the apparatus. In the two-path experiment with a barrier on path B, the electron's *own position* becomes correlated with its x-spin leaving its spin properties entangled with its position. This *internal* correlation destroys the interference effect in precisely the same way as an environmental correlation.

In general, if one does anything at all to record which path the electron travels in the two-path experiment, one will necessarily be correlating the value of one's physical record with the electron's position, and hence with its x-spin, and consequently destroying the interference effect of finding it z-spin up in region II. Such decoherence effects destroy interference effects but without causing a collapse of the state. The interactions that produce the decoherence effect are perfectly linear. The standard theory predicts the same statistics for a z-spin measurement of an electron in region II if an observer previously collapses its state, putting it on one or the other of the two paths, as it does if the state of a particle (or anything else) becomes correlated with the position of the electron in such a way as to record which path it traveled. But, again, the causal stories are entirely different.

This difference is critically important to how the theory works. Both the interaction with the recording particle p and the presence of a barrier on one of the paths destroy the interference effect by producing a physical correlation that prevents one from being able to factor the electron's state into a spin part and a position part after two seconds. In the two-particle experiment involving e and p, the correlation that destroys the interference effect is between the electron's x-spin and p's position. In the one-particle barrier experiment, it is the correlation between the electron's x-spin and its own position. But if one observes which path the electron is on, it is the random, nonlinear collapse of the quantum state that destroys the interference effect.

The collapse of the state on measurement cannot be modeled as a decoherence effect. If the position of the recording particle p is simply correlated to the electron's position by means of a linear interaction, the resultant state of the composite system after two seconds will be

$$|\psi(t_2)\rangle_{ep} = \left(\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|b\rangle_p\right)|\text{II}\rangle_e.$$

But if one supposed that the interaction with recording particle *p* somehow *measured* the electron's state, the state of the composite system would be

$$|\uparrow_x\rangle_e |a\rangle_p |\mathrm{II}\rangle_e$$

or

$$|\downarrow_x\rangle_e |b\rangle_p |\mathrm{II}\rangle_e$$

with probability 1/2 for each.

As in the experiment with the barrier on one path, the three states here exhibit different physical properties. The first describes the electron and the recording particle *p* as being in an entangled superposition of the electron being *x*-spin up and *p* being in region *a and* the electron being *x*-spin down and *p* being in region *b*. Here *p*'s position is quantum-mechanically correlated to the electron's *x*-spin, but neither *p*'s position nor the electron's *x*-spin are determinate on the standard interpretation of states. While the electron will exhibit no single-particle interference effect in region II in this state, since the recording particle *p*'s position is not even determinate, this state could not possibly represent a measurement where its position provided a determinate record of the measurement result. Particle *p*'s position is, however, both correlated to the electron's *x*-spin *and* perfectly determinate in the second and third states. In contrast with the first state, the recording particle *p*'s position in either of these states might represent the determinate record of an *x*-spin measurement of the electron.

Given the standard interpretation of states, there is a good argument that the state of a system *must* collapse on measurement. The entangled state produced by internal or environmental decoherence is simply incompatible with there even being a determinate measurement record.

5.5 Quantum Records

A quantum record, just like a classical record, involves correlating one's recording variable with the value of the physical quantity that one wants recorded. But the sort of historical inferences one might make from a classical record typically don't hold for quantum records. To be sure, if the electron starts in an eigenstate of *x*-spin and hence determinately travels one or the other of the two classically possible trajectories, then looking at the position of the recording particle *p* will tell one which path the electron took. If *p* ends up in region *a*, then the electron took path *A*; and if *p* ends up in region *b*, then the electron took path *B*. But if the electron starts in anything other than an eigenstate of *x*-spin, then the value of one's record will not allow one to infer which determinate trajectory the electron traveled, for the simple reason that it will not have traveled a determinate trajectory at all.

If the electron's initial state is z-spin up, then the state of the composite system will be

$$|\psi(t_{3/2})\rangle_{ep} = \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e|b\rangle_p$$

after 3/2 seconds. If one observes the position of p in this state, with probability 1/2 one will collapse the state to

$$|\text{up record}\rangle_{ep} = |\uparrow_x\rangle_e|A\rangle_e|a\rangle_p \tag{5.4}$$

and find p in region a and with probability 1/2 one will collapse the state to

$$|\text{down record}\rangle_{ep} = |\downarrow_x\rangle_e|B\rangle_e|b\rangle_p \tag{5.5}$$

and find p in region b. So if one finds p in region a, then one will certainly also find the electron x-spin up and on path A if one looks for it. One might naturally infer that it got there by traveling along path A and that the fact that the electron never got close enough to push it explains why particle p did not get pushed to region b. And if one finds p in region b, one will also certainly find the electron x-spin down and on path B if one looks. One might similarly infer that it got there by traveling along path B and that particle p was pushed from a to b when the electron passed by it. The upshot is that wherever one finds the recording particle p when one looks at the record, nothing that one might *subsequently* observe regarding the position of e or p will contradict the classical assumption that the position of p accurately records the path that the electron in fact followed. But the classical history of the electron's motion that one intuitively infers from the record is utterly false according to the theory. There is an important sense in which what one supposes to be a *record of the electron's trajectory* is not a *record* at all.

Consider the same experiment but where one decides *not* to observe the position of p. As we have seen, the correlation between the electron's x-spin and p's position would destroy the interference effect of finding a sure-fire z-spin up electron in region II. But one would be able to show empirically that the state was $|\psi(t_{3/2})\rangle_{ep}$ and hence that the electron did not in fact take either path even here by measuring a two-particle observable $\hat{O}$ that had $|\psi(t_{3/2})\rangle_{ep}$ as an eigenstate with eigenvalue $+1$ and the orthogonal state

$$\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e|a\rangle_p - \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e|b\rangle_p$$

as an eigenstate with eigenvalue −1. In state $|\psi(t_{3/2})\rangle_{ep}$, measuring $\hat{O}$ would certainly yield the result +1, but in state $|\text{up record}\rangle_{ep}$ or state $|\text{down record}\rangle_{ep}$ it would yield the result +1 half the time and −1 the other half of the time. The sure-fire disposition to get +1 for a measurement of $\hat{O}$ represents a new interference effect predicted by the linear dynamics.

The crucial point here is that when decoherence destroys one interference effect, it always produces another interference effect that one might at least in principle observe. While there is no longer a single-particle interference effect involving just *e* that that would show that the linear dynamics obtains, the theory predicts that the composite system of $e + p$ will exhibit a two-particle interference effect. Specifically, if the linear dynamics alone correctly describes the behavior of the composite system, then measuring $\hat{O}$ will yield a sure-fire result of +1. Further, we have very good empirical evidence that simple two-particle systems like this always obey the linear dynamics when unobserved. We will have more to say about this later.

There is also something *nonlocal* about what happens to the electron when one observes the position of the recording particle *p*. The situation here is a little different from our earlier reflections regarding potential quantum nonlocality. Suppose that one waits two seconds before observing *p*'s position. The state of the composite system, with *e*'s position factored out, will be

$$|\psi(t_2)\rangle_{ep} = \left(\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|b\rangle_p\right)|\text{II}\rangle_e.$$

Here the electron is determinately in region II, so nothing that one does to *p* will affect the electron's position. But the electron's *x*-spin remains entangled with *p*'s position. As a consequence, if one measures *p*'s position and finds it in region *a*, the electron *instantaneously and nonlocally* collapses to a state where it is determinately *x*-spin up; and if one finds *p* in region *b*, then the electron *instantaneously and nonlocally* collapses to a state where it is determinately *x*-spin down. So measuring the position of *p* here *instantaneously and nonlocally* gives the electron a determinate *x*-spin when it did not initially have one *no matter how far apart the two particles may be*. Inasmuch as the possessed properties of the distant electron immediately change when *p* is observed, this is a straightforward example of action at a distance.

If the states of two systems are entangled, then one cannot represent the state of either alone. Rather, the two systems must be understood as *a single system* in an entangled state. Observing the state of either of the subsystems may instantaneously and nonlocally change the state of the other subsystem by changing the state of the composite system, no matter what the distance between its parts is. If the electron and recording particle are entangled as in state $|\psi(t_{3/2})\rangle_{ep}$, the electron might be 4.37 light years away in orbit about α Centauri and a position measurement of p on Earth would both instantaneously *disentangle* the two systems and randomly give the electron a determinate x-spin that it did not determinately have before p's position was observed.

5.6 Total-of-Nothing

Albert's total-of-nothing box provides additional empirical evidence that the standard formulation of quantum mechanics is getting things right. Given the construction of the box, the theory predicts that if the spin state of an electron e is $|\phi\rangle_e$ when it enters the box, then it will be $-|\phi\rangle_e$ when it exits from the other side. This vector is on the same ray as $|\phi\rangle_e$ and hence represents precisely the same physical state according to the theory.

Taking these vectors to represent precisely the same state makes sense given how one calculates probabilities. The probability of a measurement result is given by the norm-squared of an inner product. The minus sign on a state will affect the sign of the inner product, but it will have no effect on the *norm-squared* of that inner product. Hence an electron in state $|\phi\rangle_e$ exhibits precisely the same dispositions as an electron in state $-|\phi\rangle_e$. To insist that there is a difference would be to insist on a difference that has no empirical consequences. And, indeed, the spin properties of an electron that passes through a total-of-nothing box are empirically the same as its spin properties before. An x-spin up electron stays x-spin up, a z-spin down electron stays z-spin down, and so on.

In order to calculate what happens to an electron that was initially z-spin up if a total-of-nothing box is placed on path B of a two-path apparatus (as in Figure 5.5), we will first consider what it would do to an x-spin up and an x-spin down electron. An x-spin up electron would determinately take path A, miss the box, and hence evolve as follows:

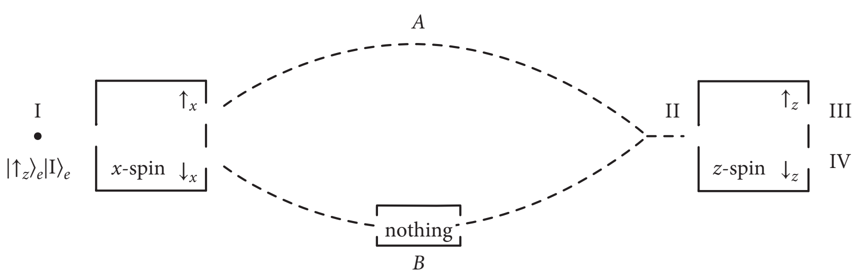


Figure 5.5. Two paths with a total-of-nothing box.

$$
\begin{gathered}
|\psi(t_0)\rangle_{ep} = |\uparrow_x\rangle_e|\mathrm{I}\rangle_e \\
\downarrow \\
|\psi(t_{1/2})\rangle_{ep} = |\uparrow_x\rangle_e|A\rangle_e \\
\downarrow \\
|\psi(t_{3/2})\rangle_{ep} = |\uparrow_x\rangle_e|A\rangle_e \\
\downarrow \\
|\psi(t_2)\rangle_{ep} = |\uparrow_x\rangle_e|\mathrm{II}\rangle_e.
\end{gathered}
$$

An x-spin down electron would determinately take path B and, hence, go through the total-of-nothing box:

$$
\begin{gathered}
|\psi(t_0)\rangle_{ep} = |\downarrow_x\rangle_e|\mathrm{I}\rangle_e \\
\downarrow \\
|\psi(t_{1/2})\rangle_{ep} = |\downarrow_x\rangle_e|B\rangle_e \\
\downarrow \\
|\psi(t_{3/2})\rangle_{ep} = -|\downarrow_x\rangle_e|B\rangle_e \\
\downarrow \\
|\psi(t_2)\rangle_{ep} = -|\downarrow_x\rangle_e|\mathrm{II}\rangle_e.
\end{gathered}
$$

In this case, the box multiplies the state of the electron by -1, but this does nothing whatsoever to its physical dispositions. It is still just an x-spin down electron when it gets to region II.

Given this and the linearity of the dynamics, an electron that begins in the state z-spin up would evolve as follows:

$$
\begin{aligned}
|\psi(t_0)\rangle_{ep} &= |\uparrow_z\rangle_e|\mathrm{I}\rangle_e \\
&= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{I}\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{I}\rangle_e \\
&\downarrow \\
|\psi(t_{1/2})\rangle_{ep} &= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e \\
&\downarrow \\
|\psi(t_{3/2})\rangle_{ep} &= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e - \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e \\
&\downarrow \\
|\psi(t_2)\rangle_{ep} &= \frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\mathrm{II}\rangle_e - \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\mathrm{II}\rangle_e \\
&= |\downarrow_z\rangle_e|\mathrm{II}\rangle_e.
\end{aligned}
$$

So a z-spin up electron that enters the apparatus will be z-spin *down* when it gets to region II, hence it will exhibit the sure-fire disposition to end up in region IV rather than region III when it exits from the z-spin box.

The total-of-nothing box, then, does nothing to an electron that goes through it and it does nothing to an electron that goes around it, but it dramatically affects the state, and hence the dispositions, of an electron that is in a *superposition of going through it and around it*. Multiplying a state by -1 does nothing since it leaves the state on its original ray, but multiplying just one term of a superposition by -1 moves the state off its original ray. In the present case, it rotates the state by 90°, which, in the context of the two-path apparatus, ultimately produces an electron where the z-spin is flipped.

Note that the total-of-nothing box does not accomplish this by directly affecting the spin of the electron in any way at any time. For this reason, it would presumably be wrong to say that any action of the total-of-nothing box *caused* the z-spin of the electron to flip. Nevertheless, it would not have flipped if the total-of-nothing box had not been on path B.

5.7 The Wave Function

Among the historical motivations for quantum mechanics we started with were the behavior of light and the stability of matter. We are now in a position to say something regarding how the standard formulation of quantum mechanics explains such phenomena.

In the two-slit experiment, the standard theory describes each photon as being in a *superposition* of going through slit A and through slit B. The *linear dynamics* predicts that such a photon will end up in a superposition of hitting the screen in different regions. When one observes the screen, this *measurement* causes it to collapse to a state where it has a determinate approximate position. The norm-squared of the quantum-mechanical amplitude associated with each region gives the probability of the photon being found in that region when one looks. The distinctive interference distribution of photons on the screen is given by these probabilities. This is a two-path interference effect akin to finding a z-spin up electron at the end of the two-path apparatus discussed above. The final state of the photon at the screen here is similarly the result of applying the linear dynamics to the vector sum of the state of the photon going through slit A and through slit B. And the standard formulation of quantum mechanics gets this state, and hence the statistical distribution, just right.

With regard to the stability of matter, consider a hydrogen atom where the electron is in the lowest energy state possible. Here the *linear dynamics* predicts that the electron is in a spherically symmetric *superposition* of different locations centered on the positively charged proton that constitutes the nucleus of the hydrogen atom. It also tells us that, while it does not have a determinate position, there is a sense in which the electron is determinately not moving. This stationary superposition, known as the 1*s* orbital, is stable under the unitary dynamics. Consequently, as long as no one looks for the electron or kicks it around by more conventional means, it will just sit there in a superposition of different *positions* and a stable eigenstate of *energy*.

As with the photon in the two-slit experiment, the electron here is not determinately in any particular region before it is observed. Rather, its state is represented as a complex-valued *wave function* $|\psi(r)\rangle_e$ over possible positions r. The norm-squared of the wave function gives the probability density for finding the electron at a particular position. The wave function can be understood as a unit-length vector in a Hilbert space $\mathcal{H}$ whose elements are functions. What it means for the function $|\psi(r)\rangle$ to be unit-length is that

$$\int_R |\langle\psi(r)|\psi(r)\rangle|^2 dr = 1,$$

taking the integral here over all possible positions R. The inner product of $|\phi(r)\rangle$ and $|\chi(r)\rangle$ is

$$\int_R \langle\psi(r)|\chi(r)\rangle dr$$

over all possible positions R.

An orthonormal basis for $\mathcal{H}$ consists in a set of unit-length functions $\{\phi_1, \phi_2, \ldots\}$ such that any function in the space can be written as a linear combination of the elements of the space. This sum typically has an infinite number of terms

$$|\psi(r)\rangle_e = \sum_k a_k|\phi_k(r)\rangle_e = \sum_k \langle\psi(r)|\phi_k(r)\rangle|\phi_k(r)\rangle_e.$$

The probability of finding e in state $|\phi_k(r)\rangle_e$ if one measures the observable corresponding to the orthonormal basis is $|a_k|^2 = |\langle\psi(r)|\phi_k(r)\rangle|^2$. And the probability $P(R)$ of finding the electron in any particular region R is equal to

$$P(R) = \int_R |\langle\psi(r)|\psi(r)\rangle|^2 dr,$$

and $|\langle\psi(r)|\psi(r)\rangle|^2$ is the electron's probability density.

For systems that contain more than one particle, the wave function is a complex-valued function over *configuration space*. Configuration space has three dimensions for each particle in the system one wants to describe. One might think of a single point in the $3N$-dimensional configuration space associated with a system S

as giving the x_k, y_k, and z_k coordinate for each particle k in S. The coordinates of the point $Q(t)$ might then be written as

$$(x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_k, y_k, z_k, \ldots, x_N, y_N, z_N).$$

Configuration space for one particle is ordinary 3-space; for two particles, it is a six-dimensional space where each point represents the positions of each of the two particles in 3-space; and so on. An immediate consequence is that, except for the special case of a single particle, the wave function of a system cannot be thought of as a simple field living in ordinary three-dimensional space. While the wave function is important to the standard theory, a clear understanding of configuration space and the wave function over configuration space will be absolutely essential when we discuss GRW and Bohmian mechanics later.

One can often represent a system's quantum-mechanical state by considering the amplitude of its configuration being in various of a finite set of disjoint regions of configuration space. This has been our strategy up to now. When it works, this allows one to get by with a finite-dimensional Hilbert space. We will continue to do so when we can.

We have seen how the standard collapse formulation of quantum mechanics predicts and explains all of the phenomena we have considered so far. The theory is counterintuitive. But given that our experience of quantum-mechanical systems is counterintuitive, any theory that makes the right empirical predictions must be.

There are other sorts of phenomena that any empirically adequate formulation of quantum mechanics would need to be able to predict. Among the most salient of these are the EPR–Bell statistics we consider in the next chapter. These statistics illustrate a subtle sort of non-locality that is predicted by the standard theory and exhibited by the physical world. Inasmuch as the EPR–Bell statistics are also counterintuitive, any theory that predicts them must be as well.

6
Real and/or Local

6.1 The EPR Argument

While the standard formulation of quantum mechanics has been enormously successful in making precise empirical predictions concerning counterintuitive phenomena, there have always been discontents. Einstein was awarded the Nobel prize in physics in particular for his work in quantum mechanics, but he was a persistent critic of the theory.[1] His most influential argument was in the paper "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" (1935), which he coauthored with Boris Podolsky and Nathan Rosen. Einstein, Podolsky, and Rosen (EPR) answered the question posed by the title of their paper by arguing that quantum mechanics in fact could not be considered to provide a complete description of physical reality.

The EPR argument was based on two conditions they took to be uncontentious. For a physical theory to be *complete* every element of physical reality must have a counterpart in the theory. And, if one can predict the value of a quantity with certainty without in any way disturbing a system, then there is an element of *physical reality* corresponding to that quantity.

Both of these conditions make at least some intuitive sense. If there is a real physical quantity that is not represented by one's theory, then one's theory is indeed incomplete in the precise sense that there is a real physical quantity that it does not represent. And, while this is a bit more subtle, if one can predict the result of a future measurement *with certainty* without in any way disturbing a system, then one might imagine that there must be some real fact about the system before it was measured that explains one's successful prediction. More specifically, one might believe that the quantity that one is measuring must in fact already have had the value that one predicts getting before it was measured and that it is *this* that explains the successful prediction. And, if the quantity already has that value, then it has it whether or not it is ever in fact measured. Hence, one might conclude, there is an element of physical reality corresponding to the value of the quantity.

Putting the two conditions together, a physical theory is only *complete* if it has a counterpart for every quantity of a system that can be predicted with

[1] The citation for Einstein's 1921 Nobel prize (which he was awarded in November 1922) read "To Albert Einstein for his services to theoretical physics and especially for his discovery of the photoelectric effect" (Pais 1982, 386 and 503). Here *discovery* means *quantum-mechanical explanation*.

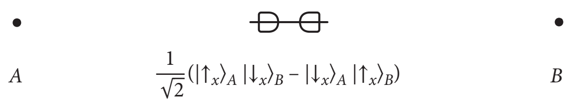


Figure 6.1. Two electrons in the EPR state.

certainty without disturbing the system in any way. EPR argue that quantum mechanics is *incomplete* by arguing that there are values of physical quantities that are not represented in the theory but that one can predict with certainty without disturbing the relevant physical system. While they used position and momentum for their argument, one can give the same argument in terms of x-spin and z-spin.

Consider two electrons A and B in something we will call the EPR state

$$\frac{1}{\sqrt{2}}\left(|\uparrow_x\rangle_A|\downarrow_x\rangle_B - |\downarrow_x\rangle_A|\uparrow_x\rangle_B\right). \tag{6.1}$$

Suppose that electron A is near Earth and electron B is near α Centauri some 4.37 light years away (as in Figure 6.1). Since the two particles do not move much through the EPR story, we will just consider the spin part of their state here.

The EPR state describes the x-spins of the two particles as being anti-correlated. Specifically, whatever x-spin one observes for one electron, one will observe the opposite x-spin for the other. But the EPR state has a further, more general property. It describes the spins of the two particles as anti-correlated in *every spin basis*.

Consider z-spin as an example. To calculate the EPR state in the z-spin basis, we start with the EPR state in the x-spin basis, substitute the expressions for z-spin, then simplify.

$$\begin{aligned}
&\frac{1}{\sqrt{2}}(|\uparrow_x\rangle_A|\downarrow_x\rangle_B - |\downarrow_x\rangle_A|\uparrow_x\rangle_B)\\
&\quad= \frac{1}{\sqrt{2}}\Bigg[\frac{1}{\sqrt{2}}(|\uparrow_z\rangle_A + |\downarrow_z\rangle_A)\frac{1}{\sqrt{2}}(|\uparrow_z\rangle_B - |\downarrow_z\rangle_B) - \frac{1}{\sqrt{2}}(|\uparrow_z\rangle_A - |\downarrow_z\rangle_A)\\
&\qquad\frac{1}{\sqrt{2}}(|\uparrow_z\rangle_B + |\downarrow_z\rangle_B)\Bigg]\\
&\quad= \frac{1}{\sqrt{2}}\Bigg[\frac{1}{2}|\uparrow_z\rangle_A|\uparrow_z\rangle_B - \frac{1}{2}|\uparrow_z\rangle_A|\downarrow_z\rangle_B + \frac{1}{2}|\downarrow_z\rangle_A|\uparrow_z\rangle_B - \frac{1}{2}|\downarrow_z\rangle_A|\downarrow_z\rangle_B\\
&\qquad-\frac{1}{2}|\uparrow_z\rangle_A|\uparrow_z\rangle_B - \frac{1}{2}|\uparrow_z\rangle_A|\downarrow_z\rangle_B + \frac{1}{2}|\downarrow_z\rangle_A|\uparrow_z\rangle_B + \frac{1}{2}|\downarrow_z\rangle_A|\downarrow_z\rangle_B\Bigg]\\
&\quad= -\frac{1}{\sqrt{2}}(|\uparrow_z\rangle_A|\downarrow_z\rangle_B - |\downarrow_z\rangle_A|\uparrow_z\rangle_B).
\end{aligned}$$

So whatever z-spin one observes for one particle, one will observe the opposite z-spin for the other.

The EPR argument goes as follows. By measuring the x-spin of particle A, one can predict the x-spin of particle B with certainty without disturbing it in any way. One just predicts the opposite for particle B of whatever one gets for the x-spin of particle A. And, given their classical and relativistic intuitions, EPR believed that measuring the x-spin of particle A could not possibly (instantaneously) disturb the distant particle B. So particle B must already have a determinate x-spin. It is that already-posessed x-spin that explains the result one knows one would certainly get when one measures particle B given that one by direct observation knows the x-spin of A. Similarly, by measuring the z-spin of particle A, one can predict the z-spin of particle B with certainty without disturbing it in any way. Again, one just predicts the opposite z-spin. And since measuring A cannot affect distant particle B, B must already have a determinate z-spin as well. Hence, the possessed values of both x-spin and z-spin must be fully determinate, real physical properties of particle B before we measure them of B. And since measuring particle A does not disturb the properties of particle B in any way, the possessed values of *both* x-spin and z-spin must be fully determinate, real physical properties of particle B even before we measure anything of A and even if we change our mind and decide not to measure A at all. And since this argument could also be run from the perspective of first measuring particle B then predicting the results of measurements on particle A, particle A must itself already have a determinate x-spin and z-spin as well. Since the full quantum-mechanical description of the EPR state does not tell us the values of the x-spin and z-spin of either particle, *quantum mechanics is incomplete*. Further, inasmuch as it denies that any particle could possibly have both a determinate x-spin and a determinate z-spin at the same time, one might take the very conceptual structure of quantum mechanics to *require* the descriptive incompleteness of the theory.

Because of its symmetry, the EPR state cannot possibly represent possessed values for the x-spin and z-spin of the two particles under any interpretation. More specifically, with regard to the standard theory, the eigenvalue–eigenstate link tells us that it is simply nonsense to suppose that an electron has both an x-spin and a z-spin at the same time. If an electron has a determinate x-spin, it cannot have a determinate z-spin; and if it has a determinate z-spin, it cannot have a determinate x-spin. So if one accepts EPR's argument against state completeness, one must also reject the standard eigenvalue–eigenstate link. If EPR are right, then the problem is not just that quantum mechanics is incomplete—it is that the standard quantum account of physical properties is fundamentally mistaken.

A key step in the argument turns on EPR's implicit assumption that measuring particle A could not possibly disturb the state of particle B in any way. Here they are appealing to an intuitively plausible notion of *locality* where no action on one system can instantaneously affect the state of a distant system. Of course, there are good physical reasons for assuming something like this. Perhaps most significantly, special relativity does not allow for causal influences that propagate faster than light.

For its part, the standard formulation of quantum mechanics predicts that measuring the x-spin of particle A instantaneously gives particle B an x-spin that it did not already have. If one were to measure the x-spin of particle A, the collapse dynamics predicts that one would get the result x-spin up with probability 1/2, and the final state of the composite system would be $|\uparrow_x\rangle_A|\downarrow_x\rangle_B$. In this case, one would indeed get the result x-spin down for a subsequent measurement of B, but not because B was x-spin down before one measured A. Rather, it is because the measurement of the x-spin of A instantaneously gave B a determinate x-spin when it did not have one before. Similarly, the collapse dynamics predicts that one would get x-spin down for particle A with probability 1/2 leaving the composite system in the state $|\downarrow_x\rangle_A|\uparrow_x\rangle_B$. Here one would get the result x-spin up for a subsequent measurement of B, but not because B was already x-spin up. Rather, it is again because the measurement of the x-spin of A instantaneously gave B a determinate x-spin when it did not have one before. In each case, the collapse of the state instantaneously changes the states of *both* particles by disentangling their spins. On this description of measurement and account of property attribution, one might argue that the standard collapse theory is complete in the sense that whenever one can predict a result with certainty, there is a determinate quantity that explains the reliability of the prediction. But it is not compatible with relativity.[2]

The collapse dynamics is closely tied to the eigenvalue–eigenstate link in the standard formulation of quantum mechanics. Inasmuch as (1) particle B has no x-spin on the standard interpretation of states before A is measured and (2) it has the opposite x-spin as particle A immediately after A is measured, measuring particle A's x-spin *must* instantaneously change the state of B. That is, the collapse dynamics *must be nonlocal* in the strong sense of being incompatible with relativity to give *entangled systems* the determinate properties they need when and where they need them on the standard interpretation of states.

The upshot is that when EPR conclude that by measuring the x-spin of particle A, one can predict the x-spin of particle B with certainty *without disturbing it in any way*, they are not so much arguing that the standard theory is incomplete as they are contradicting it. Their belief that measuring A cannot possibly affect the state of B is incompatible with both (1) the collapse dynamics and (2) how the eigenvalue–eigenstate link characterizes the properties of the entangled particles.

[2] Einstein had already described his worries about the locality and completeness of quantum mechanics at the 1927 Solvay Congress, at which he also described his concern with the mutual compatibility of quantum mechanics and relativity. See Bacciagaluppi and Valentini (2009) for a detailed discussion.

EPR wanted a formulation of quantum mechanics that was *complete* and *local* in a sense that is compatible with special relativity. A central question, then, is whether there is any empirically adequate formulation of quantum mechanics that is both complete and local in this sense.

6.2 Quantum Mechanics and Relativity

Quantum mechanics and special relativity are arguably the two best physical theories we have ever had. Together, they are the cornerstones of modern physics. But the standard collapse formulation of quantum mechanics and special relativity are also logically incompatible with each other. To see why we will start by briefly considering how special relativity works.[3]

Special relativity describes how physical events are spatially and temporally organized. In the theory, three-dimensional space and time are represented by a single four-dimensional structure—*spacetime*. The theory provides each unaccelerated observer with a way of identifying which events are simultaneous in her inertial frame. The set of events an inertial (unaccelerated) observer judges to be simultaneous with each tick of her clock forms a surface in spacetime. And the family of surfaces formed by her clock-ticks determines a foliation of spacetime.

Consider time t and one of the three spatial coordinates, call it x (as in Figure 6.2(a)). An observer in inertial frame $\mathcal{L}$ at rest relative to the laboratory would judge the spacetime events on each of the dashed lines as simultaneous with each of her clock-ticks while an observer moving to the right at a constant relative velocity v in inertial frame $\mathcal{L}'$ would judge the events on the dotted lines in this figure as simultaneous with each of her clock-ticks.

Now suppose instead that observer $\mathcal{L}$ is moving to the left relative to the laboratory frame (as in Figure 6.2(b)). In this case, she would judge the spacetime

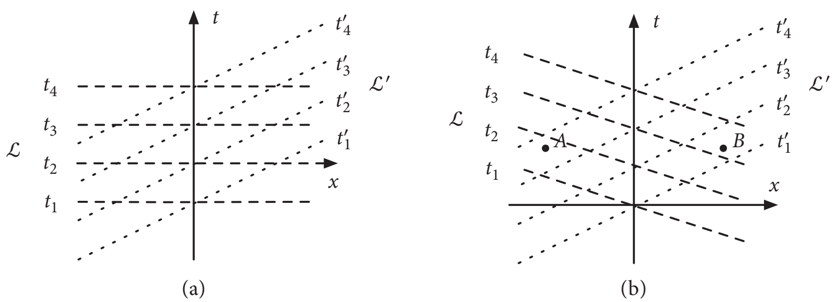


Figure 6.2. Order of measurement events in different inertial frames.

[3] See Susskind and Friedman (2017) for an introduction to special relativity.

events on each of the dashed lines in that figure as simultaneous with each of her clock-ticks. Consider events A and B. In frame $\mathcal{L}$ event A occurs between time t_1 and time t_2 and event B occurs between time t_3 and time t_4. So observer $\mathcal{L}$ would judge event A to occur *before* event B. But in $\mathcal{L}'$'s inertial frame, whose planes of simultaneity are represented by the dotted lines, event A occurs between time t'_3 and time t'_4 and event B occurs between time t'_1 and time t'_2. So observer $\mathcal{L}'$ would judge event A to occur *after* event B. The principle of relativity, on which special relativity is based, requires inertial observers to agree with each other regarding all matters of physical fact. Since observers in the two inertial frames *disagree* about the temporal order of the events, there is simply no physical matter of fact about whether event A occurs before or after event B. This, however, is logically incompatible with how the standard collapse formulation of quantum mechanics accounts for determinate measurement outcomes.

Suppose that particles A and B are in the EPR state and that particle A is in region A near Earth and particle B is in orbit about α Centauri. Suppose that an observer on Earth measures particle A's z-spin at noon 1 January 2050 (in the laboratory inertial frame) and that an observer on α Centauri measures the x-spin of particle B at noon 1 January 2050 plus one second (in the laboratory inertial frame). Suppose, finally, that Figure 6.2(b) represents the geometric relationship between the observers' inertial frames and the two measurement events A and B.

Now consider whether particle B has a determinate x-spin *just before* the α Centauri observer measures it. In the $\mathcal{L}$ frame, as in the laboratory frame, the Earth observer's measurement occurs first and the α Centauri observer's second. So by the time the α Centauri observer measures the x-spin of her particle, the Earth observer has caused a collapse of the state and the two particles are now either in the state $|\uparrow_x\rangle_A|\downarrow_x\rangle_B$ or the state $|\downarrow_x\rangle_A|\uparrow_x\rangle_B$. Consequently, particle B already has a perfectly determinate x-spin. In this inertial frame, the α Centauri observer's x-spin measurement will just reveal the local and fully determinate physical matter of fact.

But in the $\mathcal{L}'$ frame the α Centauri observer's measurement occurs first and the Earth observer's measurement occurs second. So here when the α Centauri observer measures the x-spin of her particle, the two particles are still in the entangled EPR state and there is, hence, simply no determinate matter of fact concerning the x-spin of particle B just before her measurement. Here the α Centauri observer's measurement is not simply *revealing* a determinate local property of her particle. It is instantaneously giving particle B a randomly determined x-spin, nonlocally giving particle A the opposite x-spin, and, in the process, disentangling the states of the two particles. But in the $\mathcal{L}$ frame, it is the Earth observer's measurement that gives each of the particles determinate x-spins.

Back to the question of whether particle B has a determinate x-spin just before the α Centauri observer measures it. Since particle B has a determinate x-spin just before the α Centauri observer measures it in the $\mathcal{L}$ frame and does not have a determinate x-spin in the $\mathcal{L}'$ frame, there can be no physical matter of fact concerning whether particle B has the *local property* of having a determinate x-spin just before it is measured.

This leaves us in a position where we cannot consistently assign a state to the two-particle system nor can we even consistently say whether or when one might assign determinate *local* properties to particle B. The sort of nonlocality exhibited by the standard theory is very strong—its account of determinate measurement outcomes is logically incompatible with relativistic constraints.

Now consider the closely related question of whether the state of particle B is *entangled* with the state of particle A just before the α Centauri observer measures it. Since it is in one inertial frame and isn't in another, *entanglement* cannot be a real physical relation between the two EPR particles according to relativity, at least not as entanglement is understood in the context of the standard theory. The problem with that, as we will see shortly, is that we use the notion of *entangled* spacelike-separated systems to explain quantum phenomena like the results of EPR–Bell-type experiments. The upshot is that there are basic features of standard quantum mechanics that are essential to how the theory explains quantum phenomena but that are flatly incompatible with the dynamical constraints of relativity.

Rather than argue that the standard theory was incompatible with relativity, EPR argued that the quantum-mechanical description of physical systems was incomplete. The historical reason for this is that the immediate target of their argument was Bohr's Copenhagen interpretation of quantum mechanics.[4] From the perspective of the standard theory, the problem is more one of compatibility with relativity than incompleteness. To get a formulation of quantum mechanics that is compatible with relativity, one might consider replacing the instantaneous collapse of the quantum-mechanical state with a dynamical law that does not require the choice of a preferred inertial frame.[5] But if one did so, one would also need to revise how one understands quantum-mechanical entanglement and property attribution more generally.

[4] Since their target was Bohr's Copenhagen interpretation of quantum mechanics, they could not argue that the theory was incompatible with relativity since it was entirely unclear whether the collapse dynamics should count as a *real* dynamical process given Bohr's particular brand of idealism. But that does not mean that Bohr's interpretation of quantum mechanics was a good one—just that it was, and still is, extremely difficult to pin down. See Becker (2018) for a compelling critical account of Bohr and his companions' views on quantum mechanics. See Pais (1982 and 1986) for more traditional sentiments.

[5] As we will see when we consider the idea of a hypersurface-collapse dynamics in connection with GRWf (Chapter 8), the new dynamical law might involve collapses, but it will have to be significantly more subtle than rule 4II. How one understands entanglement in a theory like GRWf is also very different from in the standard theory.

6.3 Bell's Theorem

John von Neumann was convinced that the standard quantum-mechanical state was complete and accurate, and hence that a "hidden variable" interpretation of quantum mechanics that added extra parameters to the standard state description in the way that EPR seem to have wanted was untenable. Indeed, many physicists and philosophers believed that von Neumann had *proven* that there could be no hidden-variable formulation of quantum mechanics. But this is simply wrong. Indeed, there has long been a concrete counterexample.

Bohmian mechanics (which we will discuss in Chapter 11) is a formulation of quantum mechanics developed by David Bohm in 1952 and is closely related to an idea that Louis de Broglie described at the 1927 Solvay Congress.[6] It is a *hidden-variable theory* that arguably makes the same statistical predictions as the standard formulation of quantum mechanics for the experiments we have discussed so far. Further, there is a sense in which it satisfies EPR's intuitions concerning the reality of physical quantities. In particular, it predicts that all of the *intrinsic* physical properties of a system are fully determinate, and it both represents and describes the evolution of these properties. But Bohmian mechanics is, like the standard collapse theory, nonlocal in the strong sense that its dynamical laws are flatly incompatible with the constraints of special relativity. The question then is whether one can get a local version of something like Bohmian mechanics that might satisfy the implicit locality constraint that EPR appeal to in their argument.

In his famous 1964 paper, "On the Einstein Podolsky Rosen Paradox," John Bell described the situation as follows:

> There have been attempts to show that . . . no "hidden variable" interpretation of quantum mechanics is possible [here Bell cites von Neumann (1932)]. These attempts have been examined elsewhere and found wanting [here Bell cites a paper that he would later publish (1966)]. Moreover, a hidden variable interpretation of elementary quantum theory has been explicitly constructed [here Bell cites Bohm's description of Bohmian mechanics (1952)]. That particular interpretation has indeed a grossly nonlocal structure. This is characteristic, according to the result to be proved here, of any such theory which reproduces exactly the quantum mechanical predictions. (1964, 195)

Bell then proceeded to prove his famous theorem. Unlike many who later commented on the significance of Bell's theorem, Bell himself did not take the theorem to count in any way against hidden-variable theories generally nor Bohmian mechanics in particular. Indeed, since the problem with Bohmian mechanics was that it was nonlocal, Bell just wanted to know whether one could somehow

[6] See Bacciagaluppi and Valentini (2009) for a discussion of de Broglie's proposal.

keep the sort of complete realism it provided while making the theory local. He took his theorem to show that this was impossible. Bell, consequently, concluded that Bohmian mechanics was the *best way* he could see to formulate quantum mechanics.

Bell considered an experiment where one starts with a pair of particles *A* and *B* in the EPR state and one can choose to measure various spin properties of each of the two particles. In this setup, EPR would believe that each particle must have an already-possessed spin property in every direction in which one might measure its spin. It is these preexisting properties that explain why we can predict with certainty the result of a measurement of any of these spin properties for a particle without disturbing it by measuring the same property of the other particle. So, if EPR are right, then there must be a more complete specification of the EPR state given by that set of physical parameters that would in fact determine the result of each individual spin measurement. It is this set of parameters that one would have to add to the standard quantum-mechanical state in order for the theory to be complete.

EPR also implicitly assumed that the measurement of one of the particles does not in any way affect the other particle. Bell took this to mean that the measurement of one particle does not affect the result of the spin measurement on the other particle. In particular, the result one gets for the measurement of particle *B* cannot depend on which spin observable one chooses to measure of particle *A*—that the value of the physical parameter that determines the result of the measurement on *B* at α Centauri cannot depend on which spin property one choses to measure of *A* on Earth.

Bell showed, however, that if a theory satisfies these assumptions, then it cannot make the same statistical predictions as the standard collapse formulation of quantum mechanics. And it was later shown that the standard theory makes the right *empirical* predictions for these measurements. Hence, no theory that satisfies both EPR's reality and locality assumptions can be empirically adequate.[7]

Again, Bell did not take this to mean that Bohm's theory was unsatisfactory. Rather, he concluded that *one cannot do better than Bohmian mechanics*. Bell was a champion of Bohmian mechanics throughout his career. He explained how the theory worked, defended it, extended it, and enumerated its virtues. We will examine both Bohmian mechanics and how Bell understood it in detail in Chapter 11. Here we will consider why he was not overly worried that it exhibited a strong variety of nonlocality.

[7] This does not mean that no formulation of quantum mechanics can be compatible with relativistic constraints. As we will see when we discuss GRWf later (in section 8.4 in particular), one can get a formulation of quantum mechanics that might be said to satisfy relativistic constraints. The theory, however, does not satisfy EPR's reality assumptions. One way to see this is that neither EPR particle has *any* spin properties on GRWf.

6.4 Bell-Type Theorems

We now have a family of Bell-type no-go theorems. These theorems are proven by starting with a set of metaphysical assumptions, then using the axioms of classical probability theory to prove that a particular statistical inequality must hold if those assumptions are satisfied. A particular Bell-type theorem is interesting if the assumptions used to derive it are intuitively plausible and the associated inequality is violated by nature. In that case, one knows that at least one of the intuitively plausible assumptions is in fact false. In order to see better how such Bell-type arguments work in general, consider the following Bell-type no-go theorem.[8]

Consider two electrons in the EPR state

$$\frac{1}{\sqrt{2}}(|\uparrow_x\rangle_A|\downarrow_x\rangle_B - |\downarrow_x\rangle_A|\uparrow_x\rangle_B).$$

And consider three physical directions along which one might measure their spin: +60°, 0°, and −60° (as in Figure 6.3). Each of these directions corresponds to a *different* spin property. The angles give the direction along which one would orient the magnets on a Stern–Gerlach device to measure the particular spin property of each electron. Given how the standard theory represents physical properties, each of these spin properties corresponds to a different orthonormal basis in the Hilbert space $\mathcal{H}$ we use to represent electron spin. More specifically, the 0°-spin basis is at a 30° angle to the +60°-spin basis in one direction and the −60°-spin basis in the other.

Now consider three intuitively plausible assumptions. Given their commitments and the observed spin properties of electrons, EPR would almost certainly have endorsed each of these.

Assumption 1 (value definiteness): Particles A and B in fact have determinate spin values in each of the three directions.

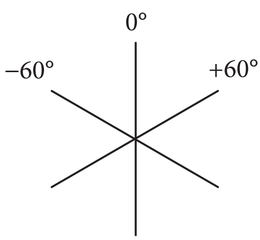


Figure 6.3. Three spin properties labeled by magnet angle.

[8] This theorem follows Wigner (1970) and (1976).

Since this mirrors one of EPR's central conclusions, they would certainly have accepted this assumption. On this assumption, the three spin values for each of the two particles

$$(A@+60°, A@0°, A@-60°; B@+60°, B@0°, B@-60°)$$

can be represented by six signs. The first sign is the value of particle *A*'s +60°-spin, the second is the value of particle *A*'s 0°-spin, etc. These are the real physical parameters that EPR would take to represent the possessed spin properties in each direction. In this notion, then, $(+-+;-+-)$ represents *A*'s +60°-spin as up, its 0°-spin as down, and its −60°-spin as up; and it represents *B*'s +60°-spin as down, its 0°-spin as up, and its −60°-spin as down. Let $P(+-+;-+-)$ be the epistemic probability of those particular parameter values in fact obtaining.

Assumption 2 (anti-correlated spins): The possessed spins of the two particles are opposite in each direction.

Since EPR took the possessed properties of the particles to be what explained the result of measuring that property, and since the results of spin measurements of the same spin property of the two particles in the EPR state are always found to be opposite, they certainly would have accepted this assumption. This reduces the number of possible determinate spin states for the two particles to eight:

$$(+++;---), (++-;--+), (+-+;-+-), (+--;-++)$$
$$(-++;+--), (-+-;+-+), (--+;++-), (---;+++).$$

On each trial of the experiment, each of the devices, one at location *A* and the other at location *B*, may be used to measure one of two possible spin properties. Suppose that each spin-measuring device begins ready to measure spin in the 0° direction and that one can *turn* the device at *A* to measure spin in the +60° direction and/or *turn* the device at *B* to measure spin in the −60° direction. If a device is *not turned*, then its measurement automatically defaults to measure spin in the 0° direction.

Assumption 3 (setting independence): The outcome of a spin measurement of a particle is given by the value of its possessed spin in the direction being measured.

One might think of this as a sort of locality assumption that also involves what it means to measure a physical property. In the present context, it captures both EPR's sense that the result of a measurement must be explained by the preexisting properties of the particle being measured and their sense that what one chooses to measure of one particle cannot affect the possessed properties of the other particle. Given Assumption 2, if one's measurement of a particle changed a spin property of that particle, then it would also have to change the corresponding spin property of the distant particle instantaneously in order to keep the spins opposite. So while no single assumption involves locality, a locality condition is entailed by the three assumptions together.[9]

Given that one's choice of what to measure does not change the value of what is being measured, one can take the probabilities of one's measurement results to be equal to the probabilities of the corresponding values in fact obtaining. This means that

$$P(+_A \,\&\, +_B | \text{turn } A)$$

or the probability of getting + for the result at *A* and + for the result at *B*, given that one turns only the device at *A*, is just the probability that particle *A* has the property + in the +60° direction (since the device at *A* was turned and, hence, is oriented in the +60° direction) and particle *B* has the property + in the 0° direction (since the device at *B* was not turned and, hence, is still oriented in the 0° direction). Similarly

$$P(+_A \,\&\, +_B | \text{turn } B)$$

is the probability that particle *A* has the property + in the 0° direction and particle *B* has the property + in the −60° direction. And

$$P(+_A \,\&\, +_B | \text{turn } A \,\&\, \text{turn } B)$$

is the probability that particle *A* has the property + in the +60° direction and particle *B* has the property + in the −60° direction.

Suppose that the device at *A* is turned to the +60° orientation and the device at *B* is left in the 0° orientation. Since the two states in the eight above where this could happen (where *A* is + in the +60° direction and *B* is + in the 0° direction) are mutually exclusive, it follows from the assumption that the probability of one's measurement results is equal to the probabilities of the corresponding values in fact obtaining, together with the standard probability axioms, that the probability of getting a + result at both *A* and *B* is

[9] There are also a number of *implicit* assumptions in a typical Bell-type argument, and that is true here as well. It is typically assumed without comment, for example, that the observed nonclassical Bell correlations are not explained by a pre-established harmony between the value of the parameter that determines one's measurement outcomes and what observables one in fact chooses to measure. For their part, EPR would have been committed to such background assumptions. It was a key methodological principle for Einstein that one assume the physical world to be such that one can do meaningful experiments.

$$P(+_A \,\&\, +_B|\text{turn } A) = P(+-+;-+-) + P(+--;-++).$$

Similarly

$$P(+_A \,\&\, +_B|\text{turn } B) = P(++-;--+) + P(-+-;+-+).$$

And

$$P(+_A \,\&\, +_B|\text{turn } A \,\&\, \text{turn } B) = P(+--;-++) + P(++-;--+).$$

So

$$\begin{aligned} P(+_A \,\&\, +_B|\text{turn } A \,\&\, \text{turn } B) =& P(+_A \,\&\, +_B|\text{turn } A) + P(+_A \,\&\, +_B|\text{turn } B) \\ & - P(+-+;-+-) - P(-+-;+-+). \end{aligned}$$

But since all probabilities must be greater than or equal to zero

$$P(+_A \,\&\, +_B|\text{turn } A \,\&\, \text{turn } B) \leq P(+_A \,\&\, +_B|\text{turn } A) + P(+_A \,\&\, +_B|\text{turn } B).$$

This last expression is a Bell-type inequality that follows from the assumptions that went into the theorem.

For the angles specified here, quantum mechanics predicts that the probability on the left of the inequality is $\frac{1}{2}\sin^2(60°) = \frac{3}{8}$ and that the sum of the probabilities on the right is $\sin^2(30°) = \frac{1}{4}$. Since $\frac{3}{8}$ is not in fact less than or equal to $\frac{1}{4}$, the empirical predictions of the standard formulation of quantum mechanics are incompatible with the EPR assumptions. And since the standard predictions turn out to be right, no empirically adequate description of the world can satisfy the assumptions that go into the theorem. Hence, inasmuch as EPR would have agreed to each of the three assumptions here, their metaphysical intuitions are simply incompatible with the empirical properties of the physical world we inhabit.[10]

6.5 Quantum Property Attribution Redux

We are now in a position to summarize a few things that we have learned about property attribution in the standard theory.

[10] Note that the threat to locality is not addressed by simply moving to a field theoretic formulation of quantum mechanics. Field theory still must explain the observed EPR–Bell statistics. To do so, either one must assume non-local entangled field states that are non-locally disentangled on measurement as on the standard account or one must propose a clear alternative explaination.

To begin, on the standard interpretation of states, a system in state $|\phi\rangle_S$ only determinately has those properties for which $|\phi\rangle_S$ is an eigenstate, which is always an impoverished set. On the linear dynamics, an electron will always have *some* determinate property, but that property will virtually never be the property one sets out to measure. Hence, given how property attribution works, the linear dynamics alone cannot explain determinate measurement outcomes.

Quantum mechanics also allows a composite system to have properties that are not a function of the properties of its parts. Consider a two-particle system in the EPR state

$$\frac{1}{\sqrt{2}}(|\uparrow_x\rangle_A|\downarrow_x\rangle_B - |\downarrow_x\rangle_A|\uparrow_x\rangle_B).$$

Since the total spin of the composite system AB as described by the first term is zero and the total spin as described by the second term is zero, the composite system AB is in an eigenstate of having a total x-spin of zero. And since the state has the same form written in any other spin basis, the composite system AB is in an eigenstate of having a total spin in every other direction of zero as well. Consequently, the total x-spin (and the total spin in every other direction) of the composite system AB is in fact zero on the standard interpretation of states. But this is not because the sum of the x-spins of the two particles is zero.

Since the states of the two particles are entangled, neither has a determinate spin of any sort—indeed, neither particle even has a spin *state* to call its own. Hence the determinate properties of the composite system cannot be understood to be a function of the properties of its parts. Further, just considering spin properties for a moment, neither particle has a property that might be used to distinguish it from the other because *neither has any determinate properties to call its own*, yet there are in fact two particles on the standard interpretation of states.[11]

Finally, even when physical systems have determinate quantum states to call their own, those states may not individuate the physical systems. It can happen, for example, that the full physical state of two particles A and B is represented by precisely the same vector in Hilbert space $|\phi\rangle_A = |\phi\rangle_B$. In this case, each particle has *precisely the same determinate properties* and hence there are again no properties that might be used to distinguish the two particles.

6.6 EPR Morals

The empirical violation of Bell's original inequality and the Bell-type inequality we just considered means that no empirically adequate formulation of quantum

[11] Since each term is an eigenstate of there being two particles, the superposition is also an eigenstate of there being two particles.

mechanics can satisfy EPR's physical and philosophical intuitions. One might be tempted to conclude from that there is nothing wrong with quantum mechanics and everything wrong with EPR's old-fashioned understanding of the world. But this is too fast. While at least one of the assumptions EPR made must be wrong, there remains the question of *which* must be sacrificed and *precisely how*. Moreover, the standard formulation of quantum mechanics itself is not compatible with basic relativistic constraints, which is a serious problem given the foundational role that relativity plays as the other cornerstone of modern physics.

For many years, most physicists who thought about such things at all thought that Bell's theorem showed that one must give up EPR's reality and/or completeness assumptions. People typically did not think that locality or relativity was seriously threatened. While one might ultimately adopt such a position, one should reflect carefully on the explanatory sacrifice in doing so. On this view, there would be physical facts that one could predict with certainty (like the result of an x-spin measurement on a distant system) but no physical facts that explain why one can predict them (in this case, no determinate x-spin properties of that system).

As we will see, Bell himself was willing to give up locality, even in the strong sense of entertaining theories that were manifestly incompatible with the dynamical constraints of relativity, in order to keep a much weakened version of EPR's reality and completeness conditions. What led him to this view was that Bohmian mechanics already provided a clear example of precisely how one might do this *and* get the right empirical predictions.

Perhaps the best argument against the standard response to Bell's theorem is that even if one is willing to give up reality and/or completeness, it is difficult to formulate quantum mechanics in a way that is compatible with relativity. For its part, the standard collapse theory is manifestly incompatible with standard relativistic constraints. We will consider two very different ways to get a formulation of quantum mechanics that is arguably compatible with relativistic constraints when we discuss GRWf and various ways of understanding Everett's pure wave mechanics. But one does not get compatibility with relativity for free. Both of these approaches involve significant conceptual costs.

In the case of the standard theory, the collapse dynamics is the primary source of the incompatibility with relativity. If one could figure out how to get by with just the linear dynamics, then one would stand a much better chance of getting a theory that is compatible with relativistic constraints. But as we have seen, the standard interpretation of states requires the collapse dynamics—without collapses one typically does not have determinate measurement outcomes at all. So giving up the standard collapse dynamics would also require a radically different way of interpreting quantum-mechanical states.

That said, giving up the collapse dynamics and adopting a new interpretation of quantum-mechanical states is also *not sufficient* to get compatibility with relativity. Bohmian mechanics both gives up the collapse dynamics, consequently requiring a radically new way of interpreting quantum-mechanics states, *and* is incompatible with relativity.

But here we are getting ahead of the story again. Bohmian mechanics was not meant to address the incompatibility of the standard theory and relativity, or even to provide a more intuitive account of physical properties. It was meant to address the quantum measurement problem.

7
The Quantum Measurement Problem

7.1 Wigner's Friend

The philosopher Hillary Putnam tells of meeting Albert Einstein at Princeton for tea in 1953. The conversation turned to the conceptual difficulties with quantum mechanics. Einstein was most worried about the dynamical laws of the theory. As Putnam tells the story,

> What he said on that occasion was something like the following: "Look, I don't believe that when I am not in my bedroom my bed spreads out all over the room, and whenever I open the door and come in it jumps into the corner."
>
> (Putnam 2005, 624).

Einstein was right to worry.

The term *measurement* occurs in the standard collapse formulation of quantum mechanics as an undefined primitive term. In particular, the linear dynamics describes the evolution of a physical system when it is *not measured* and the collapse dynamics describes its evolution when it is *measured*. Insofar as it is unclear what counts as a measurement, the dynamics of the standard theory is ambiguous. But the problem is more subtle than this might suggest.

The Wigner's friend story describes a thought experiment where one observer makes a measurement of a quantum system then is herself measured by another observer. It can be thought of as a generalized version of Schrödinger's famous cat story.[1] It provides a good starting point for understanding the quantum measurement problem.

The Wigner's friend story was told by Eugene Wigner (1961), but Hugh Everett III (1956), one of Wigner's graduate students at Princeton, had already used it to characterize the conceptual problems faced by the standard formulation of quantum mechanics in a version of his PhD dissertation.[2] While Everett used the story to explain why the standard theory was inconsistent, Wigner used it to

[1] Schrödinger (1935) told his cat story to illustrate how easily it is for microscopic superposition to generate a superposition of ordinary properties of a macroscopic system—namely, he shows how one might easily end up with a cat in an entangled superposition of being alive and dead under the linear dynamics.

[2] John Wheeler was Everett's PhD advisor. While Everett's presentation is the earliest version of the Wigner's friend story I know, the basic structure of the story reflects von Neumann's (1932) discussion of when the collapse occurs in a sequence of measurement-like interactions. That said, von Neumann believed that one's choice of where to put the collapse did not matter empirically. That it does is

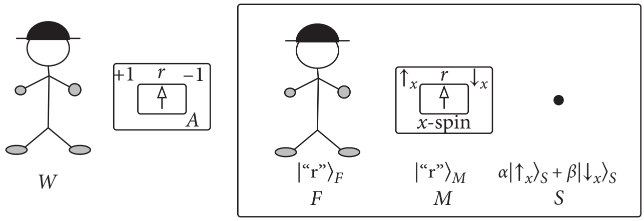


Figure 7.1. The Wigner's friend setup.

explain how he thought the standard theory worked. We will start with the story as Wigner told it, then discuss how Everett and Wigner each understood it.

Consider a friend *F*, with a measuring device *M*, who is ready to measure the *x*-spin of a system *S*. More specifically, suppose that the friend begins in a state $|``r"\rangle_F$ where she is ready to observe her measuring device and her measuring device begins in a state $|``r"\rangle_M$ where it is ready to measure the *x*-spin of *S* (as in Figure 7.1). Suppose further that the friend is a perfect observer of the measurement result indicated by the measuring device and that the measuring device is perfect in correlating the position of the pointer that represents its result with the *x*-spin of *S*. Finally, suppose that the measuring device correlates its pointer with the *x*-spin of *S* at time t_1 and that the friend looks at the result indicated by *M*'s pointer at time t_2.

The measuring device has three states where it makes a simple determinate report. In state $|``r"\rangle_M$ it is ready to make a measurement on *S*, and it indicates this fact with its pointer pointing at "ready." In state $|``\uparrow_x"\rangle_M$ it has completed its measurement and its pointer points at "*x*-spin up" indicating the result of the measurement. And in state $|``\downarrow_x"\rangle_M$ it has completed its measurement and its pointer indicates that the result was "*x*-spin down."

The friend similarly has three states where she makes a simple determinate report. We will understand these states in terms of her physical dispositions. In state $|``r"\rangle_F$ the friend is ready to look at the position of the pointer on *M* and to record the result of this observation. If the friend is in this state, she also has the sure-fire physical disposition to answer "yes" if asked whether she is ready to reliably observe the result indicated by the pointer on her measuring device. We will suppose that she then records the result of her measurement as a brain record. More specifically, in state $|``\uparrow_x"\rangle_F$ she has looked at the pointer on *M* and has recorded that it points at "*x*-spin up." In this state she has the sure-fire disposition to report "*x*-spin up" if she is asked what result she got. And in state $|``\downarrow_x"\rangle_F$ she has looked at the pointer, has recorded that it points at "*x*-spin down," and has the sure-fire disposition to report that it was "*x*-spin down." Insofar as the sure-fire

something that both Everett and Wigner made clear in their discussions of such nested measurements. See Wigner (1963) and (1981) for subsequent reflections on the measurement problem.

disposition of an observer to report a particular result indicates her beliefs, the friend *believes* that she in fact got whatever result she has the sure-fire disposition to report that she got.

Assuming ideal correlating interactions between the three systems, the composite system has the following two dispositions. If the object system begins in the state $|\uparrow_x\rangle_S$

$$\begin{aligned} |\psi(t_0)\rangle_{FMS} &= |\text{“}r\text{”}\rangle_F|\text{“}r\text{”}\rangle_M|\uparrow_x\rangle_S \\ &\downarrow \\ |\psi(t_1)\rangle_{FMS} &= |\text{“}r\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S \\ &\downarrow \\ |\psi(t_2)\rangle_{FMS} &= |\text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S. \end{aligned}$$

And if the object system begins in the state $|\downarrow_x\rangle_S$

$$\begin{aligned} |\psi(t_0)\rangle_{FMS} &= |\text{“}r\text{”}\rangle_F|\text{“}r\text{”}\rangle_M|\downarrow_x\rangle_S \\ &\downarrow \\ |\psi(t_1)\rangle_{FMS} &= |\text{“}r\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S \\ &\downarrow \\ |\psi(t_2)\rangle_{FMS} &= |\text{“}\downarrow_x\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S. \end{aligned}$$

So if the x-spin of the object system is determinate, the story is easy. The measuring device simply comes to indicate the x-spin of S; the friend ends up with a reliable brain record, and hence a reliable belief, regarding the measurement outcome; and she is prepared to reliably report the outcome of her measurement.

But suppose that the object system begins in a state that is not an eigenstate of x-spin. Say, the z-spin up state

$$|\uparrow_z\rangle_S = 1/\sqrt{2}(|\uparrow_x\rangle_S + |\downarrow_x\rangle_S). \tag{7.1}$$

The question of how the composite system evolves is now more subtle.

When the object system starts in an eigenstate of x-spin, we do not need to know which interactions should ultimately count as measurements because the collapse dynamics would not disturb a system that is already in an eigenstate of the observable being measured. But now it matters. To see why, let's start by supposing that the linear dynamics correctly describes both the interaction between M and S and the interaction between F and M.

Assuming the ideal correlating interactions just described, the linear dynamics predicts that the state of the composite system FMS will evolve as follows:

$$
\begin{aligned}
|\psi(t_0)\rangle_{FMS} &= |``r"\rangle_F|``r"\rangle_M\left[1/\sqrt{2}(|\uparrow_x\rangle_S + |\downarrow_x\rangle_S)\right] \\
&= 1/\sqrt{2}|``r"\rangle_F|``r"\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|``r"\rangle_F|``r"\rangle_M|\downarrow_x\rangle_S \\
&\downarrow \\
|\psi(t_1)\rangle_{FMS} &= 1/\sqrt{2}|``r"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|``r"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S \\
&\downarrow \\
|\psi(t_2)\rangle_{FMS} &= 1/\sqrt{2}|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S.
\end{aligned}
$$

Because the dynamics is linear, one simply calculates the action of the dynamics on the superposition by time-evolving each term separately using the composite system's dispositions and keeping the original coefficients.

The standard interpretation of states tells us that the final state here is one where the friend has no determinate measurement record whatsoever. This is reflected in the fact that if we asked her what measurement result she got, she would not have the determinate disposition to respond "x-spin up" nor would she have the determinate disposition to respond "x-spin down". Rather, by the linear dynamics, she would have the disposition to utter a symmetric *superposition* of these two reports. And, hence, insofar as her mental state supervenes on her physical state, she also fails to have the determinate belief that she got x-spin up or the determinate belief that she got x-spin down.[3] Indeed, since her record is entangled with both the position of *M*'s pointer and the x-spin of *S*, the friend does not even have a proper quantum-mechanical state to call her own.

In contrast, if we suppose that the nonlinear collapse dynamics correctly describes the interaction between *M* and *S*, then there would be a collapse between times t_0 and t_1 and the composite system would evolve as follows:[4]

$$
\begin{aligned}
& |\psi(t_0)\rangle_{FMS} = 1/\sqrt{2}|``r"\rangle_F|``r"\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|``r"\rangle_F|``r"\rangle_M|\downarrow_x\rangle_S \\
& \qquad\downarrow \\
\text{probability } 1/2: \; & |\psi(t_1)\rangle_{FMS} = |``r"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S \\
\text{probability } 1/2: \; & |\psi(t_1)\rangle_{FMS} = |``r"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S \\
& \qquad\downarrow \\
\text{probability } 1/2: \; & |\psi(t_2)\rangle_{FMS} = |``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S \\
\text{probability } 1/2: \; & |\psi(t_2)\rangle_{FMS} = |``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S.
\end{aligned}
$$

[3] In the strong sense in which we are using the notion, we will say that the mental state of an agent *supervenes* on her physical state if and only if her physical state can be understood as determining her mental state. This notion allows for some distance between mental and physical states. Specifically, an agent's mental state might supervene on her physical state without itself being in any way the state of a physical system.

[4] Note that we are assuming throughout, as the standard theory requires, that the linear dynamics describes the time-evolution of the composite system whenever there is no collapse of the state.

Hence the friend would determinately get one of the two possible measurement outcomes, each with equal probability. And, given that her mental state supervenes on her dispositions to report, she would determinately believe that she got that result.

If the collapse occurs instead when F interacts with M, then there would be a collapse between times t_1 and t_2 and the composite system would evolve as follows:

$$|\psi(t_0)\rangle_{FMS} = 1/\sqrt{2}|\text{“}r\text{”}\rangle_F|\text{“}r\text{”}\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|\text{“}r\text{”}\rangle_F|\text{“}r\text{”}\rangle_M|\downarrow_x\rangle_S$$
$$\downarrow$$
$$|\psi(t_1)\rangle_{FMS} = 1/\sqrt{2}|\text{“}r\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|\text{“}r\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S$$
$$\downarrow$$
$$\text{probability } 1/2: \ |\psi(t_2)\rangle_{FMS} = |\text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S$$
$$\text{probability } 1/2: \ |\psi(t_2)\rangle_{FMS} = |\text{“}\downarrow_x\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S.$$

While this is a different sequence of states, the friend would again end up with one or the other of the two possible determinate measurement records with equal probability and with the corresponding belief.

So regardless of whether the collapse occurs between M and S or between F and M, the state of the composite system FMS will ultimately either end up

$$|\text{up}\rangle_{FMS} = |\text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S$$

or

$$|\text{down}\rangle_{FMS} = |\text{“}\downarrow_x\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S,$$

each with probability 1/2. And, unlike the state that evolves from the linear dynamics alone, both of these describe the friend as having a perfectly determinate measurement record on the standard interpretation of states.

The collapse dynamics does two essential things in the standard theory. It explains how an observer ends up with a determinate measurement record given the standard interpretation of states, and it explains why these results are randomly distributed with the standard quantum statistics. But since the theory does not tell us what constitutes a measurement, we do not know *when* the collapse occurs in the chain of interactions between F, M, and S. Inasmuch as it is the collapse that is supposed to explain there being a determinate measurement result, it presumably must happen sometime before the friend is aware of the measurement result. But it might happen when the measuring device interacts with the object system or when the friend looks at the pointer on the measuring device. Or perhaps sometime else in a more detailed characterization of the chain of correlating interactions.

In order to avoid the problem of having to say precisely when the collapse occurs, von Neumann argued that it does not matter *empirically* whether one puts the collapse between M and S or between F and M.[5] In either case, one gets the state $|\text{up}\rangle_{FMS}$ with probability 1/2 and the state $|\text{down}\rangle_{FMS}$ with probability 1/2. But while it does not matter for the statistics for the result of the friend's x-spin measurement, it *does matter* for the full dynamical story and hence for the statistical results one would get for *other* observables. In short, the sequence of states of the composite system FMS depends on precisely when the collapse occurs, and different states have different empirical properties for measurements of the *composite system*. The Wigner's friend story is ultimately a story of *nested* measurements.

Suppose we want to know whether a collapse has happened at some point in the interactions between the three systems. Consider a Hermitian operator $\hat{A}$ that has

$$\begin{aligned}|{+1}\rangle_{FMS} &= 1/\sqrt{2}|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S \\ &= 1/\sqrt{2}(|\text{up}\rangle_{FMS} + |\text{down}\rangle_{FMS})\end{aligned}$$

as an eigenvector corresponding to the eigenvalue $\lambda = +1$ and

$$\begin{aligned}|{-1}\rangle_{FMS} &= 1/\sqrt{2}|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S - 1/\sqrt{2}|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S \\ &= 1/\sqrt{2}(|\text{up}\rangle_{FMS} - |\text{down}\rangle_{FMS})\end{aligned}$$

as an eigenvector corresponding to the eigenvalue $\lambda = -1$ (as in Figure 7.2). Rule 2 tells us that there is a physical observable A corresponding to $\hat{A}$.[6]

Suppose an external observer W measures the observable A of the composite system FMS (as in the figure 7.1). If the linear dynamics correctly describes the interactions between the three systems, then FMS will end up in the state $|{+1}\rangle_{FMS}$, which, as the notation suggests, is an eigenstate of A with eigenvalue +1.

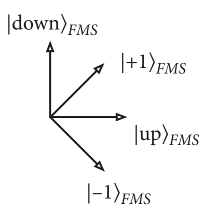


Figure 7.2. Eigenstates of the A-observable.

[5] See von Neumann (1955, 420–1 and 439–45) for his reflections.
[6] See Albert (1986), Albert and Barrett (1995), and Barrett (1999) for discussions of A-measurements and what they mean for the causal closure of worlds on Everettian formulations of quantum mechanics.

Hence if the linear dynamics correctly describes the evolution of the composite system, then an A-measurement of *FMS* will certainly yield the result $+1$. But if a collapse occurred somewhere in the sequence of interactions between the three systems (S, M, and F), then the final state of the composite system will be either $|\text{up}\rangle_{FMS}$ or $|\text{down}\rangle_{FMS}$. Since both $|\text{up}\rangle_{FMS}$ and $|\text{down}\rangle_{FMS}$ are at 45° angles to both $|+1\rangle_{FMS}$ and $|-1\rangle_{FMS}$, the norm-squared of the inner product of each of these states with $|+1\rangle_{FMS}$ is 1/2 and the norm-squared of the inner product of each with $|-1\rangle_{FMS}$ is 1/2. So, if a collapse has occurred at any point in the interactions between the three systems, one is no longer guaranteed to get the result $+1$ for an A-measurement on *FMS*; rather, the probability of finding that $A = +1$ is 1/2 and the probability of finding that $A = -1$ is 1/2.

Consequently, if one performs a sequence of A-measurements on composite systems like *FMS* after the internal interactions, one would be able to empirically distinguish between the state one gets by applying the linear dynamics to the composite system $|+1\rangle_{FMS}$ and either of the states that might result from the collapse dynamics $|\text{up}\rangle_{FMS}$ or $|\text{down}\rangle_{FMS}$. Similarly, there are other A-type interference observables that would tell one *when* in the chain of interactions between S, M, and F the collapse occurred, if a collapse occurs at all.

7.2 The Measurement Problem

We are now in a position to say what the quantum measurement problem is. The term *measurement* occurs in the standard collapse formulation of quantum mechanics as an undefined primitive term. Whether a system is measured or not determines when each of the theory's two dynamical laws obtains. Specifically, the deterministic linear dynamics describes the evolution of the state of a system when it is not measured and the random nonlinear collapse dynamics describes its evolution when it is measured. Since these two laws predict *physically incompatible states* when applied to the same interaction, one must specify strictly disjoint conditions for when each obtains for the theory to be logically consistent. Without this, the standard theory is at best *dynamically incomplete* on the most charitable reading. Further, since there are empirical consequences for when each of the two dynamical laws obtains, the theory is *incomplete in an empirically significant way*. If one happens to believe that measurement interactions are physical interactions between physical systems like any other (and why wouldn't they be?), and thus accurately described by the linear dynamics, then the theory is just *logically inconsistent*.

The sort of incompleteness involved in the measurement problem is different from the sort that EPR worried about. They worry that the theory does not characterize the values of all of the elements of reality at a time—that it does not specify the values of those quantities that can be determined with certainty while not disturbing the physical system. This is a worry regarding the synchronic incompleteness of the quantum-mechanical *state description*. Here the issue concerns the diachronic incompleteness of the dynamics. Since the theory's dynamics fails to specify how the quantum-mechanical state of a composite system will evolve, even if one grants the standard quantum state descriptions, the theory does not fully specify how such states evolve. This sort of incompleteness is particularly salient in the context of the Wigner's friend nested measurements. Here the theory cannot make any empirical predictions whatsoever for the results of A-type measurements of the composite system by dint of the fact that one does not even know what *quantum state* to ascribe to the composite system following the sequence of internal interactions.

Everett used his version of the Wigner's friend story, which he called an "amusing, but *extremely hypothetical* drama," to examine what he called *the question of the consistency* of the standard formulation of quantum mechanics (1956, 74–5). Since the standard theory's two dynamical laws predict incompatible states, he concluded that it was logically inconsistent and hence untenable. This is a much stronger verdict than that the theory is just dynamically incomplete. Everett was led to it by his firmly held conviction that observers and their measuring devices are physical systems like any other. If one supposes that observers and their measuring devices are composed entirely of simpler physical systems interacting with each other linearly, then interactions between measuring devices and other physical systems must always obey the linear dynamics just like every other interaction. And one immediately gets a logical contradiction if one's theory *ever* predicts a nonlinear collapse.

Since Everett believed that all physical systems, including sentient observers and measuring devices, always obey the linear dynamics, he believed that the composite system *FMS* would in fact end up, as required by the linear dynamics, in the state

$$1/\sqrt{2}|\text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|\text{“}\downarrow_x\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S$$

after the sequence of correlating measurement interactions between the systems. Consequently, he also believed that if the external observer W were to make an A-measurement of the composite system *FMS*, he would with certainty get the result $+1$.

What made the story "extremely hypothetical" is that Everett knew that it would be extremely difficult to make a reliable A-measurement on a macroscopic system like *FMS* in practice. Nevertheless, he held that one only has a satisfactory formulation of quantum mechanics if one can provide a consistent account of such nested measurements. *And* he believed that if such a measurement were ever made, one would find that there was in fact no collapse of *F*'s state, *M*'s state, or the state of the object system S at any point in the chain of interactions. But, if this is right, one is left with the puzzle of how to explain the friend's determinate measurement results (and the standard quantum statistics) at all in the context of the linear dynamics alone.

We will consider Everett's proposal for how to solve the measurement problem later. For now, it is important to get clear on precisely why A-measurements are extremely difficult.

7.3 Why A-Type Measurements are Difficult

Suppose an observer W wants to perform an A-measurement of the composite FMS to determine whether or not his friend caused a collapse when she performed her x-spin measurement and hence got one or the other of the two classically possible determinate records. But suppose that, unknown to W, there is a single particle p in the measuring device M that is positioned in such a way that (1) if the pointer moves to the right to indicate the result "x-spin down," it will push p from region 0 to region 1, and (2) if the pointer moves to the left to indicate the result "x-spin up," it will leave p unmoved in region 0 (as in Figure 7.3).

The composite system consisting of $FMSp$, then, has the following two dispositions:

$$
\begin{aligned}
|\psi(t_0)\rangle_{FMSp} &= |\text{“}r\text{”}\rangle_F|\text{“}r\text{”}\rangle_M|\uparrow_x\rangle_S|0\rangle_p \\
&\downarrow \\
|\psi(t_1)\rangle_{FMSp} &= |\text{“}r\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S|0\rangle_p \\
&\downarrow \\
|\psi(t_2)\rangle_{FMSp} &= |\text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S|0\rangle_p
\end{aligned}
$$

and

$$
\begin{aligned}
|\psi(t_0)\rangle_{FMSp} &= |\text{“}r\text{”}\rangle_F|\text{“}r\text{”}\rangle_M|\downarrow_x\rangle_S|0\rangle_p \\
&\downarrow \\
|\psi(t_1)\rangle_{FMSp} &= |\text{“}r\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S|1\rangle_p \\
&\downarrow \\
|\psi(t_2)\rangle_{FMSp} &= |\text{“}\downarrow_x\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S|1\rangle_p.
\end{aligned}
$$

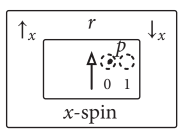


Figure 7.3. Environmental decoherence in an A-type measurement.

So, on the linear dynamics, the composite system would evolve as follows:

$$
\begin{aligned}
|\psi(t_0)\rangle_{FMSp} &= |\text{“}r\text{”}\rangle_F|\text{“}r\text{”}\rangle_M\left[\frac{1}{\sqrt{2}}(|\uparrow_x\rangle_S + |\downarrow_x\rangle_S)\right]|0\rangle_p \\
&= \frac{1}{\sqrt{2}}|\text{“}r\text{”}\rangle_F|\text{“}r\text{”}\rangle_M|\uparrow_x\rangle_S|0\rangle_p + \frac{1}{\sqrt{2}}|\text{“}r\text{”}\rangle_F|\text{“}r\text{”}\rangle_M|\downarrow_x\rangle_S|0\rangle_p \\
&\downarrow \\
|\psi(t_1)\rangle_{FMSp} &= \frac{1}{\sqrt{2}}|\text{“}r\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S|0\rangle_p + \frac{1}{\sqrt{2}}|\text{“}r\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S|1\rangle_p \\
&\downarrow \\
|\psi(t_2)\rangle_{FMSp} &= \frac{1}{\sqrt{2}}|\text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S|0\rangle_p + \frac{1}{\sqrt{2}}|\text{“}\downarrow_x\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S|1\rangle_p.
\end{aligned}
$$

Let's call this final state $|p\text{-super}\rangle_{FMSp}$. While it looks much like the final state in the original Wigner's friend story, it has very different physical properties.

Consider what would happen if observer W made an A-measurement of the composite system in state $|p\text{-super}\rangle_{FMSp}$. To do this, we need to write the state in terms of eigenstates of the observable A. The following two identities, relating expressions from our earlier discussion of the Wigner's friend story, are useful for the purpose at hand:

$$|\text{up}\rangle_{FMS} = \frac{1}{\sqrt{2}}(|{+1}\rangle_{FMS} + |{-1}\rangle_{FMS})$$

and

$$|\text{down}\rangle_{FMS} = \frac{1}{\sqrt{2}}(|{+1}\rangle_{FMS} - |{-1}\rangle_{FMS}).$$

These identities are readily checked by writing $|{+1}\rangle_{FMS}$ and $|{-1}\rangle_{FMS}$ in terms of $|\text{up}\rangle_{FMS}$ and $|\text{down}\rangle_{FMS}$, as at the end of section 7.1, then substituting the expressions above.

Using these identities, we can rewrite $|p\text{-super}\rangle_{FMSp}$ and simplify:

$$
\begin{aligned}
|p\text{-super}\rangle_{FMSp} &= \frac{1}{\sqrt{2}}|\text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S|0\rangle_p + \frac{1}{\sqrt{2}}|\text{“}\downarrow_x\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S|1\rangle_p \\
&= \frac{1}{\sqrt{2}}|\text{up}\rangle_{FMS}|0\rangle_p + \frac{1}{\sqrt{2}}|\text{down}\rangle_{FMS}|1\rangle_p
\end{aligned}
$$

$$= \frac{1}{\sqrt{2}}\left[\frac{1}{\sqrt{2}}(|{+1}\rangle_{FMS} + |{-1}\rangle_{FMS})|0\rangle_p + \frac{1}{\sqrt{2}}(|{+1}\rangle_{FMS} - |{-1}\rangle_{FMS})|1\rangle_p\right]$$
$$= \frac{1}{2}|{+1}\rangle_{FMS}|0\rangle_p + \frac{1}{2}|{-1}\rangle_{FMS}|0\rangle_p + \frac{1}{2}|{+1}\rangle_{FMS}|1\rangle_p - \frac{1}{2}|{-1}\rangle_{FMS}|1\rangle_p.$$

So, calculating quantum probabilities in the usual way, the probability that observer W will get the result $A = +1$ for his A-measurement is

$$|1/2|^2 + |1/2|^2 = 1/4 + 1/4 = 1/2$$

since this happens in the first and third terms. And the probability that he will get the result $A = -1$ is

$$|1/2|^2 + |-1/2|^2 = 1/4 + 1/4 = 1/2$$

since this happens in the second and fourth terms. The correlation between the position of M's pointer and particle p means that, instead of being guaranteed of getting the result $A = +1$, W gets each possible result about half the time. But these no-collapse statistics are precisely the same statistics that W would get if he had performed an A-measurement of FMS in either of the states $|\text{up}\rangle_{FMS}$ or $|\text{down}\rangle_{FMS}$ that might result from a *collapse* when F makes her x-spin measurement. So if the position of particle p becomes correlated to M's pointer, then the result of W's A-measurement does not tell him anything whatsoever regarding whether his friend caused a collapse of the quantum-mechanical state when she measured the x-spin of system S. Environmental decoherence destroys the particular interference effect that would be detected by an A-measurement *just as if a collapse had occurred when the friend F made her spin measurement.*[7]

The phenomenon here is precisely the same as in the two-path experiment when we put a particle on one of paths to record which path the electron traveled. In both cases, something in the environment becomes correlated with the state of the object system in a way that destroys the particular interference effect one set out to measure. The interference effect of getting the sure-fire result $A = +1$ for a measurement of FMS is destroyed if anything whatsoever becomes correlated with

[7] Similarly, the value of the observable A will be randomized if the friend's x-spin result gets correlated to anything whatsoever in her environment. Hence the friend cannot communicate her result to any external observer following an A-measurement without randomizing the value of A. As Albert (1992) has argued, however, there is a sense in which the external observer can communicate the result of his A-measurement to the friend without disturbing the friend's memory or the value of A. Hence, the friend alone may "know" the value of two non-commuting observables (as long as her thinking about the result of her of x-spin measurement does not in any way produce a new record of this result).

F's brain record or the position of *M*'s pointer or the x-spin of *S*. So the correlation between the position of *M*'s pointer and particle *p*'s position destroys this particular interference effect.

Since environmental decoherence makes it look like a collapse has occurred from the perspective of the interference effect detected by an A-measurement, it is tempting to imagine that environmental decoherence, like the collapse dynamics, explains why observers get determinate measurement results. The argument might go something like this. The collapse dynamics yields a quantum-mechanical state where there is a determinate x-spin record (either $|\text{up}\rangle_{FMS}$ or $|\text{down}\rangle_{FMS}$) and it destroys the interference effect of getting $A = +1$. Environmental decoherence destroys the interference effect of getting the sure-fire result $A = +1$. Hence environmental decoherence yields a state where there is a determinate x-spin measurement record. So we do not need the collapse dynamics since the linear dynamics and environmental decoherence already explain why one gets determinate measurement records.

But this is just sloppy. The collapse dynamics does indeed yield a quantum-mechanical state where there is a determinate x-spin record, either $|\text{up}\rangle_{FMS}$ or $|\text{down}\rangle_{FMS}$, instead of the state predicted by the linear dynamics

$$1/\sqrt{2}|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S.$$

Unlike the collapse dynamics, however, environmental decoherence involves a fully linear interaction between the system and its environment—one that does not yield a state where there is a determinate x-spin record. The interaction with p yields the state

$$|p\text{-super}\rangle_{FMSp} = \frac{1}{\sqrt{2}}|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S|0\rangle_p + \frac{1}{\sqrt{2}}|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S|1\rangle_p.$$

Hence, on the standard interpretation of states, environmental decoherence *does nothing whatsoever* to produce a determinate measurement record. Indeed, it produces a *more complicated* entangled superposition of incompatible measurement records. And even if one drops the standard interpretation of states, if one keeps state completeness, the friend *cannot* have a single determinate measurement record here since the resultant state $|p\text{-super}\rangle_{FMSp}$ fails to select one of the possible records over the other as the actual result of the friend's spin measurement.

Whenever environmental decoherence destroys one interference effect, it always creates new interference effects that one might in principle detect. There is an observable B of the larger composite system *FMSp* that has $|p\text{-super}\rangle_{FMSp}$ as an eigenstate corresponding to the eigenvalue $+1$ and an orthogonal state corresponding to the eigenvalue -1. A B-measurement would again allow the observer W to determine whether his friend's x-spin measurement had caused a collapse in the context of *this particular instance of environmental decoherence*. That said, a *B*-measurement would be even harder to make than an *A*-measurement since it involves a strictly larger system that one would have to be able to isolate and control. And the observer *W* would have to know that *B* was the right observable to measure to determine whether the interaction was in fact linear, which would require *W* to know exactly what sort of environmental interaction had occurred.

The upshot is that it would be extremely difficult to measure an observable that would determine whether or not the friend caused a collapse when she measured the *x*-spin of her object system. It would require one to be able to isolate and control every particle in the macroscopic system *FMS* to ensure that no interactions produced correlations that would destroy the interference effect one set out to measure of the composite system. Macroscopic systems contain many particles that one would need to track, and they typically couple strongly with their environments in complicated ways. This is why Everett called his version of the Wigner's friend story *extremely hypothetical*. But while environmental decoherence makes it difficult to perform nested measurements that show that the friend evolved linearly, this does not mean that environmental decoherence explains the friend's determinate spin result. Again, the full state of the composite system after environmental decoherence fails to select either possible record over the other as the actual result.

A satisfactory resolution of the quantum measurement problem requires one to be able to tell the Wigner's friend story consistently. And pointing out that a nested measurement experiment with macroscopic systems would be extremely difficult, or even technologically impossible, does nothing whatsoever to address the problem. A satisfactory formulation of quantum mechanics should tell us whether the final state of the Wigner's friend story is something like $|\text{up}\rangle_{FMS}$ or $|\text{down}\rangle_{FMS}$ or whether it is something like

$$1/\sqrt{2}|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S.$$

If it is the former, it needs to say when and why the state of the composite system collapsed. And if it is the latter, it needs to explain precisely how that state describes there being a determinate measurement record. If our best formulation of quantum mechanics cannot handle nested measurements, then we simply do not understand the physical world.

8
The Collapse of the Quantum State

8.1 Wigner's Solution

Eugene Wigner did not present his version of the friend story as a criticism of quantum mechanics. Rather, it was part of his explanation for how one should properly understand the theory. In particular, he used it to argue that the quantum-mechanical state of a system must collapse when the system is apprehended by a conscious mind.

Wigner held that a consistent formulation of quantum mechanics requires one to endorse a strong variety of mind–body dualism, something that he believed most physicists had already accepted.

> Until not many years ago, the "existence" of a mind or soul would have been passionately denied by most physical scientists. . . . There are [however] several reasons for the return, on the part of most physical scientists, to the Spirit of Descartes' "Cogito ergo sum". . . . When the province of physical theory was extended to encompass microscopic phenomena, through the creation of quantum mechanics, the concept of consciousness came to the fore again: it was not possible to formulate the laws of quantum mechanics in a consistent way without reference to consciousness. (1961, 168)

The worry about consistency here comes, as we saw in the last chapter, when one tries to say when collapses occur. Wigner believed that one can avoid the quantum measurement problem by opting for a strong Cartesian dualism.[1]

He argued for his proposed solution to the measurement problem as follows:

> The important point is that the impression which one gains at an interaction may, and generally does, modify the probabilities with which one gains the various possible impressions at later interactions. In other words, the impression

[1] Descartes famously held that the mind and body were metaphysically distinct objects. This is one of the central philosophical conclusions of his *Meditations on First Philosophy* (1641). The main argument is that since one can be certain that one exists as a thinking being (I am thinking, therefore I exist) while remaining uncertain whether one's body is an illusion (produced perhaps by an evil demon), one's mind and one's body cannot be the same thing. While it is unlikely that most physicists believed that quantum mechanics committed them to this sort of mind–body dualism, it is significant that Wigner thought they did. He was a co-recipient of the Nobel Prize in Physics two years later in 1963 for his work in quantum mechanics and in particular the application of symmetry principles to the theory.

one gains at an interaction, called also *the result of an observation*, modifies the wave function of the system. . . . [I]t is *the entering of an impression into our consciousness which alters the wave function* because it modifies our appraisal of the probabilities for different impressions which we expect to receive in the future. It is at this point that the consciousness enters the theory unavoidably and unalterably. (1961, 172–3)

But it is not just our *appraisal* of the probabilities for different future impressions that changes when the impression one gains enters our consciousness. Wigner required that a physical collapse of the state occurs whenever a conscious mind apprehends the state of a measured system and gains the impression of the measurement result. It is this change in the *physical state* that requires one to change one's appraisal of future probabilities.

On Wigner's understanding of the theory, the dynamical laws of the standard collapse formulation of quantum mechanics are sharpened as follows:

4. Laws of motion (Wigner):
   I (Wigner). Linear dynamics: *when no conscious mind apprehends* the state of a physical system S, it evolves in the standard deterministic, linear way: $\hat{U}(t_0, t_1)|\psi(t_0)\rangle_S$, where $\hat{U}$ is a unitary operator.
   II (Wigner). Nonlinear collapse dynamics: *when a conscious mind apprehends* the state of a physical system S, it will instantaneously collapse to an eigenstate of the observable being measured $|\chi\rangle_S$ with the standard quantum probabilities $|\langle\chi|\psi\rangle|^2$, where $|\psi\rangle_S$ is the initial state of the system.

He believed that this sharpening of the laws was "required" for the consistency of the theory, and he considered it to be the "simplest way out" of the quantum measurement problem (1961, 180).[2]

Wigner believed that both the interaction between the measuring device M and the object system S and the interaction between the friend's physical body F and the measuring device M in the Wigner's friend story were both perfectly ordinary physical interactions and, hence, correctly described by the usual deterministic linear dynamics. The composite *physical system*, then, evolves to the state

$$|\text{super}\rangle_{FMS} = 1/\sqrt{2}|\text{``}{\uparrow_x}\text{''}\rangle_F|\text{``}{\uparrow_x}\text{''}\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|\text{``}{\downarrow_x}\text{''}\rangle_F|\text{``}{\downarrow_x}\text{''}\rangle_M|\downarrow_x\rangle_S.$$

But, Wigner argued, this cannot be the final state of the composite system. Rather, in order for there to be a determinate result, the composite system *FMS* must be either

[2] This is a modification to rule 4 of the standard formulation. We will characterize alternative formulations of quantum mechanics by saying how they might be understood as modifications of the standard theory. Unless otherwise noted, the rest of the theory stays the same.

$$|\text{up}\rangle_{FMS} = |\text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S$$

or

$$|\text{down}\rangle_{FMS} = |\text{“}\downarrow_x\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S.$$

Wigner believed that were he to ask his friend the question “What did you feel about the result of your measurement before I asked you?,” the friend would certainly reply, “I told you already, I got the result [“$\uparrow_x$” or “$\downarrow_x$”]” as the case may be. That is, the friend would report that the result of her measurement “was already decided in [her] mind” before she was asked (1961, 176). Hence Wigner concluded:

> If we accept this, we are driven to the conclusion that the proper wave function immediately after the interaction of friend and object was already either [state $|\text{up}\rangle_{FMS}$] or [state $|\text{down}\rangle_{FMS}$] and not the linear combination [state $|\text{super}\rangle_{FMS}$]. . . . It follows that the being with a consciousness must have a different role in quantum mechanics than the inanimate measuring device. . . .
> (1961, 176–7)

While it is not logically inconsistent to deny that the friend is right in reporting that she already had a determinate measurement result before she was asked, Wigner took such an option to be unacceptable. He argued that to deny that the friend has the same sort of determinate experiences that we do “is surely an unnatural attitude, approaching solipsism, and few people, in their hearts, will go along with it” (1961, 177–8). Further, by a basic principle of charity, it must be when the friend herself apprehends the state, and not just when Wigner asks her what her result was, that the composite system collapses to a state where she has a determinate measurement record. Given that everything else in the story is just ordinary physical systems interacting in ordinary ways, the only *principled* place to put the collapse is where a nonphysical entity, the friend’s *mind*, forms an impression of the result. Here the friend simultaneously creates a determinate physical record, makes that record accurate by causing a collapse of the correlated measuring device and object system, and apprehends the value of the record. And all this happens in a way that generates measurement records that satisfy the standard quantum statistics.

In some sense, this proposal immediately solves the measurement problem by saying when collapses occur. But one only really knows precisely when collapses occur here if one knows precisely which physical systems are associated with minds and the conditions under which these minds apprehend measurement outcomes.

For his part, Wigner advertised this as a feature of his theory and not a flaw. Since minds affect the quantum-mechanical states of physical systems, they also affect the objective, observable properties of those systems. Namely, if it turns out that one's friend is not in fact conscious and hence does not cause a collapse, then the state of the composite system *FMS* would be

$$1/\sqrt{2}|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + 1/\sqrt{2}|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S,$$

and an *A* measurement of the composite system would certainly yield the result $+1$. But if the friend caused a collapse, then the final state of the composite system would be $|\text{up}\rangle_{FMS}$ or $|\text{down}\rangle_{FMS}$, so an *A* measurement of the composite system would yield result $+1$ or -1 each with probability $1/2$. The upshot is that, while they would be extraordinarily difficult to perform, there are at least in principle experiments that would determine what systems cause collapses, and hence what systems are conscious. Wigner took the fact that one now has an empirical way for determining what systems are conscious to be a virtue—the theory and our intuitions concerning which systems are associated with conscious minds are inter-testable.

By saying how he thought the friend story should go, Wigner provided a sharpening of the standard collapse theory and a proposed resolution to the measurement problem. His proposal has the virtue of saying more precisely when the linear dynamics and the collapse dynamics obtain. But it does not really pin this down until one knows which systems are in fact conscious. One could appeal to one's intuitions about which systems are conscious, but this is arguably not much better than appealing to one's intuitions concerning which systems should count as measuring devices on the standard theory.

The most salient problem is that Wigner's theory requires a strong variety of mind–body dualism where minds do not supervene on physical states but cause physical events. An observer's mental state is not determined by her physical state. Rather, her mental state *causes* her physical state to collapse whenever *an impression enters into her consciousness*, whatever that might mean. Wigner's proposal is a return to the spirit of Descartes, but insofar as one is committed to mental processes supervening physical processes this is not an attractive prospect.

Finally, one would like one's solution to the quantum measurement problem to suggest how one might reconcile quantum mechanics with the dynamical constrains of special relativity. But, of course, Wigner's proposal for when and how collapses occur is as incompatible with relativity as the standard formulation of quantum mechanics, and for precisely the same reasons.

8.2 GRW★

Giancarlo Ghirardi, Alberto Rimini, and Tullio Weber (GRW) (1985; 1986) showed how to formulate a purely physical theory that says when and why collapses occur without requiring a commitment to a strong variety of mind–body dualism. The thought is that one wants macroscopic systems to have determinate classical properties so that one can explain things like tables with determinate locations and the motion of the moon, but one also wants microscopic systems to obey the linear dynamics so that one can explain things like interference effects and the stability of matter. GRW accomplish this by changing the dynamical laws of the theory. In order to understand the main idea behind their proposal, we will start by considering a toy version of their theory we will call GRW*.

GRW* supposes that physical systems are composed of fundamental particles, and it describes the particles in terms of their wave functions. Specifically, the state of a particle p is given by a complex-valued function $|\psi(r)\rangle_p$ over the possible positions where the particle might be found. The state of a collection of N particles is given by their wave function over $3N$-dimensional configuration space.[3]

GRW* is just like the standard collapse formulation of quantum mechanics, except that one replaces the linear dynamics and the collapse dynamics by a *single hybrid dynamical law*:

4. Law of motion (GRW*): The state of every physical system S evolves in the standard deterministic, linear way $|\psi(t_1)\rangle_S = \hat{U}(t_0, t_1)|\psi(t_0)\rangle_S$ except that each particle in the system has a small probability λ per unit time of collapsing randomly to an eigenstate of position. The probability of a particle collapsing to a position in a specified region is given by the standard quantum probabilities.

It is as if an external observer measures the position of a randomly selected particle at a randomly selected time and collapses it to an eigenstate of position, except that it happens spontaneously without requiring an observer of any sort. GRW chose a collapse rate λ of about 10^{-16}/sec, or about once every 10^8 years. That makes collapses rare for microscopic systems containing just a few particles but frequent for macroscopic systems containing something in the order of Avogadro's number of 6.022×10^{23} particles. The idea is that microscopic systems will behave quantum-mechanically most of the time, and macroscopic systems will behave almost classically most of the time.

Consider the Wigner's friend story in the context of GRW*. Suppose that the initial state of the friend's object system S is

$$\alpha|\uparrow_x\rangle_S + \beta|\downarrow_x\rangle_S.$$

On the hybrid dynamics, the composite system evolves linearly until one of its constituent particles collapses. Suppose that time t_1 is just before the first particle in the composite system collapses. The system then starts by evolving as follows:

[3] See the discussion of the wave function and $3N$-dimensional configuration space in section 5.7.

$$
\begin{aligned}
|\psi(t_0)\rangle_{FMS} &= |\text{“}r\text{”}\rangle_F |\text{“}r\text{”}\rangle_M (\alpha|\uparrow_x\rangle_S + \beta|\downarrow_x\rangle_S) \\
&= \alpha|\text{“}r\text{”}\rangle_F |\text{“}r\text{”}\rangle_M |\uparrow_x\rangle_S + \beta|\text{“}r\text{”}\rangle_F |\text{“}r\text{”}\rangle_M |\downarrow_x\rangle_S \\
&\downarrow \\
|\psi(t_1)\rangle_{FMS} &= \alpha|\text{“}r\text{”}\rangle_F |\text{“}\uparrow_x\text{”}\rangle_M |\uparrow_x\rangle_S + \beta|\text{“}r\text{”}\rangle_F |\text{“}\downarrow_x\text{”}\rangle_M |\downarrow_x\rangle_S.
\end{aligned}
$$

Suppose the measuring device pointer is a macroscopic system (so that one can read it) that indicates the measurement result by the direction it is pointing. Since it is macroscopic, it contains something in the order of Avogadro's number of particles. Hence, it is very likely that one of the particles in the pointer, whose position is now correlated to the x-spin of system S, will collapse to an eigenstate of position. Consequently, GRW★ predicts that the state $|\psi(t_1)\rangle_{FMS}$ will be extremely unstable. In order to see the sense in which it is unstable, we will rewrite it in terms of the n particles that make up M's pointer:

$$
\begin{aligned}
|\psi(t_1)\rangle_{FMS} = \alpha|\text{“}r\text{”}\rangle_F |\text{“}\uparrow_x\text{”}\rangle_{m_1} |\text{“}\uparrow_x\text{”}\rangle_{m_2} \cdots |\text{“}\uparrow_x\text{”}\rangle_{m_n} |\uparrow_x\rangle_S \\
+ \beta|\text{“}r\text{”}\rangle_F |\text{“}\downarrow_x\text{”}\rangle_{m_1} |\text{“}\downarrow_x\text{”}\rangle_{m_2} \cdots |\text{“}\downarrow_x\text{”}\rangle_{m_n} |\downarrow_x\rangle_S,
\end{aligned}
$$

where m_k is the kth particle in the pointer, $|\text{“}\uparrow_x\text{”}\rangle_{m_k}$ is the state where m_k is in the region where the pointer indicates the result "x-spin up," and $|\text{“}\downarrow_x\text{”}\rangle_{m_k}$ is the state where it is in the region where the pointer indicates the result "x-spin down." Since n is *very* large, while the collapse rate for each particle is small, it is highly likely that *some* particle will randomly collapse to an eigenstate of position in the next fraction of a second. The hybrid dynamics tell us that the probability of it collapsing to an eigenstate of position in the region where the pointer indicates the result "x-spin up" is $|\alpha|^2$ and the probability of it collapsing to an eigenstate of position in the region where the pointer indicates the result "x-spin down" is $|\beta|^2$.

Suppose that the first particle to collapse is particle m_{217} (as in Figure 8.1). The probability that it will collapse to an eigenstate of position in the region where the pointer indicates the result "x-spin up" is $|\alpha|^2$. Suppose it does. Since particle m_{217} is now determinately in the "x-spin up" pointer region, the probability of finding it in the "x-spin down" region is zero. If one assumes that the positions of all of the particles that constitute the pointer are perfectly correlated, one can think of the collapse as multiplying the second term of the state $|\psi(t_1)\rangle_{FMS}$ of the composite system by zero. After renormalizing, the resultant state of the composite system is

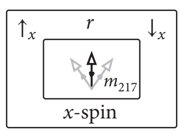


Figure 8.1. Unstable pointer superposition in GRW.

$$\begin{aligned}|\psi(t_{1.00001})\rangle_{FMS} &= |``r"\rangle_F|``\uparrow_x"\rangle_{m_1}|``\uparrow_x"\rangle_{m_2}\dots|``\uparrow_x"\rangle_{m_n}|\uparrow_x\rangle_S \\ &= |``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S.\end{aligned}$$

In this case, the friend will see the pointer determinately indicating x-spin up because its center of mass is in this region of ordinary space. Hence the friend will end up in the state

$$|\text{up}\rangle_{FMS} = |``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S.$$

Similarly, the probability that particle m_{217} will collapse to an eigenstate of position in the pointer region "x-spin down" is $|\beta|^2$, in which case the resultant state of the composite system is

$$\begin{aligned}|\psi(t_{1.00001})\rangle_{FMS} &= |``r"\rangle_F|``\downarrow_x"\rangle_{m_1}|``\downarrow_x"\rangle_{m_2}\dots|``\downarrow_x"\rangle_{m_n}|\downarrow_x\rangle_S \\ &= |``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S.\end{aligned}$$

In this case, the friend will end up in state

$$|\text{down}\rangle_{FMS} = |``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S$$

and see the pointer determinately indicating x-spin down because its center of mass is in this region of ordinary three-dimensional space.

So GRW* predicts the standard quantum statistics for the position of the macroscopic measuring-device pointer, and hence predicts determinate position records that satisfy the standard quantum statistics. And it does so without requiring anything nonphysical to cause the collapse of the composite system. GRW* predicts the right quantum statistics for the determinate position of M's pointer regardless of whether any conscious entity observes it.

On the hybrid dynamics, microscopic systems with relatively few particles will evolve linearly most of the time and will hence behave quantum mechanically, and the frequent collapses of the particles to eigenstates of position in macroscopic systems will keep their centers of mass close to eigenstates of position. And since the *expectation values* for position evolve classically under the linear dynamics, the expected positions of these centers of mass will evolve in an approximately classical way.

This is how GRW* explains the appearance of the classical, macroscopic world. But one only gets determinate physical measurement records, and hence determinate appearances, under special conditions.

Suppose one has an x-spin measuring device $M^\dagger$ that records the result of its measurement of the system S in the position of a single particle p instead of the position of a *macroscopic* pointer. One might think of p as a *microscopic* pointer. Assuming that $M^\dagger$ perfectly correlates the position of p with the x-spin of S, the linear dynamics predicts that the composite system $M^\dagger S$ will end up in the state

$$\alpha|``\uparrow_x"\rangle_{M^\dagger}|\uparrow_x\rangle_S + \beta|``\downarrow_x"\rangle_{M^\dagger}|\downarrow_x\rangle_S.$$

But since the position of only the particle p is correlated with the x-spin of S, GRW⋆ predicts that this state will be extremely stable. Hence, it is highly unlikely that the interaction will produce a determinate measurement record. For this reason, GRW⋆ would predict that there is likely no collapse of the state in the two-path experiment where we tried to record the position of the electron in the position of a single particle. Consequently, it would explain the destruction of the two-path interference effect *in that case* as the result of decoherence effects and not a collapse of the state of the electron.

There is a good argument that a measuring device with a pointer consisting of a single particle should not count as a measuring device at all. If one wanted to know where $M^\dagger$'s pointer was, one would need to read its position with a *second* measuring device $M^\ddagger$ with a pointer one could see. Suppose that $M^\ddagger$ perfectly correlates the position of its *macroscopic pointer*, and hence one that can be seen, with the position of p. By the linear dynamics, the resultant state will be

$$\alpha|``\uparrow_x"\rangle_{M^\ddagger}|``\uparrow_x"\rangle_{M^\dagger}|\uparrow_x\rangle_S + \beta|``\downarrow_x"\rangle_{M^\ddagger}|``\downarrow_x"\rangle_{M^\dagger}|\downarrow_x\rangle_S.$$

Inasmuch as $M^\ddagger$'s pointer contains on the order of Avogadro's number of particles, GRW⋆ predicts that *this* state will be extremely unstable since it is very likely that a collapse of one of the particles in $M^\ddagger$'s macroscopic pointer will collapse to an eigenstate of position leaving the composite system in either state

$$|``\uparrow_x"\rangle_{M^\ddagger}|``\uparrow_x"\rangle_{M^\dagger}|\uparrow_x\rangle_S$$

or state

$$|``\downarrow_x"\rangle_{M^\ddagger}|``\downarrow_x"\rangle_{M^\dagger}|\downarrow_x\rangle_S$$

with the standard quantum probabilities $|\alpha|^2$ and $|\beta|^2$ respectively. And this second interaction will very likely count as a successful measurement on the theory yielding a perfectly determinate measurement record. In each of the collapsed final states it will look like $M\dagger$ had a determinate measurement record all along and that the result that it recorded accurately represented the x-spin of S. But, of course, $M\dagger$ had no determinate measurement record and S had no determinate x-spin before $M^\ddagger$'s measurement.

This is something characteristic of measurement in GRW*. A measurement does not *reveal* the physical properties of an object system—it *creates* them.[4] When S starts in the state

$$\alpha|\uparrow_x\rangle_S + \beta|\downarrow_x\rangle_S$$

it does not have any determinate x-spin whatsoever. But when the position of a measuring device's macroscopic pointer becomes correlated to S's x-spin and a particle in that pointer collapses to an eigenstate of position, that *gives* S a determinate x-spin by giving the pointer's center-of-mass a determinate position.

That a pointer be macroscopic and well-correlated to the property of the object system being measured is, however, not by itself sufficient for there to be a determinate measurement record on GRW*. Suppose that one had a measuring device $M^\dagger$ that recorded the result of its x-spin measurement of S in the *z-spins* of each particle in a system P consisting of an Avogadro's number of spin-1/2 particles. Here the z-spins of the particles in P act as a sort of pointer—but not a very good one given how GRW* works. Suppose that the measurement interaction between $M^\dagger$ and S perfectly correlates $M^\dagger$'s pointer P with S's x-spin. The linear dynamics, then, predicts that the state of the composite system $M^\dagger + S$ will be

$$\alpha|\text{“}\uparrow_x\text{”}\rangle_{M^\dagger}|\uparrow_x\rangle_S + \beta|\text{“}\downarrow_x\text{”}\rangle_{M^\dagger}|\downarrow_x\rangle_S.$$

But here, since $M^\dagger$'s record is in terms of the z-spins of the particles of P and since collapses to eigenstates of position do nothing to spin, this state is perfectly stable under the GRW* dynamics. And it remains stable as long as no particle *positions* become correlated with the x-spin of S or the z-spins of the particles in P.

That said, inasmuch as $M^\dagger$'s record involves the z-spins of many particles, it may be difficult to keep the positions of particles in $M^\dagger$'s environment from becoming correlated with the z-spins of the particles in its pointer. If the z-spins of the particles in P become correlated with the positions of enough particles, then the entangled superposition of the composite system consisting of S, $M^\dagger$, and its environment will be unstable. A collapse of any particle whose position is well-correlated with any of the z-spins of the particles in P will collapse both the z-spin of that particle and the x-spin of S. This will consequently yield a determinate measurement record, and it will *accurately* represent the new-found and *new-made* x-spin of S.

Here environmental interactions that would just produce decoherence effects under the linear dynamics alone may also produce *collapses* on the GRW* hybrid dynamics. And these collapses would then yield determinate physical

[4] Note that this is also true for other collapse theories. On Wigner's theory, for example, there is typically no matter of fact about the value of the observable being measured until it is measured.

records—something that environmental decoherence cannot provide under the linear dynamics.

The measurement records that ultimately determine whether or not a theory is empirically adequate are those records that determine the experiences of human observers. For his part, Wigner simply stipulated that mental states are always determinate and cause collapses and then end up in states that agree with the physical records produced by the collapses. This immediately guaranteed determinate mental records on which one's *experience* might supervene. Further, given the details of the collapse dynamics, these records will exhibit the standard quantum statistics. The directness of this explanation, however, is also what renders it ad hoc. To the credit of the theory, since GRW* does not stipulate nonphysical minds with always determinate states, it must account for the determinate experience of observers *indirectly* by appealing to standard physical facts.

Suppose that Q is a physical property on which the mental records of an observer F in fact supervene. GRW* will very likely make F's mental records determinate and ensure that they are distributed in the standard quantum mechanics way if the value of Q is either *determined by* or *correlated with* the positions of a large number of particles after a measurement interaction. That is, one likely gets a determinate mental record in the theory whenever the physical record on which the mental record supervenes is in fact correlated with the positions of many particles. Note that GRW* does not treat all physical observables the same way—*position* plays a special role in explaining how the theory seeks to account for our experience.

That GRW* requires that one's measurement records ultimately supervene on the positions of particles is a version of the *preferred basis* problem. Since the Hilbert space formalism represents every physical observable of a system in the same way, there is at least an ad hoc feel to a dynamics that singles out one observable to make determinate, especially when position was chosen precisely because its being determinate also plausibly makes determinate those measurement records that we in fact take ourselves to have. Put the other way around, GRW did not choose to make the z-spin of a particular neutrino near the center of the spiral galaxy NGC 5457 determinate even though this was a perfectly coherent theoretical option—*that* would have done nothing whatsoever to explain our determinate experience of coffee cups and baseballs. Rather, they chose to make something determinate that arguably provides determinate measurement records given how we in fact record measurement outcomes and the sorts of interactions we believe typically obtain between those measurement records and their environments. This particular ad hoc selection of a preferred observable has the virtue of plausibly providing determinate measurement records. And choosing position as a physically preferred quantity is arguably less ad hoc than Wigner's proposal of simply stipulating that mental states are always determinate.

There are, however, two significant problems with choosing position as the dynamically preferred physical quantity. The first problem is technical. GRW⋆ requires particles to collapse to eigenstates of position, but there are no such eigenstates in the standard Hilbert space representation. The reason is that Hilbert space needs to be separable, which means it can have at most a countably infinite dimension. But since there are a continuously infinite number of possible positions, the Hermitian operator representing exact position would have to have a continuously infinite number of mutually orthogonal eigenvectors. This is only possible in a continuously infinite dimensional space, and such spaces are not suitable for representing quantum-mechanical states since there is typically no unique linear decomposition of a vector with respect to a given basis.

The second problem is physical. GRW⋆, like the standard theory and Wigner's theory, is not compatible with the constraints of relativity. Consider two particles A and B and two x-spin measuring devices, one particle and one measuring device at each end of the experiment M_A and M_B, that record their measurement results in the positions of macroscopic pointers. The composite system starts in the stable state

$$|\text{“}r\text{”}\rangle_{M_A}|\text{“}r\text{”}\rangle_{M_B}\left[\frac{1}{\sqrt{2}}(|\uparrow_x\rangle_A|\downarrow_x\rangle_B + |\downarrow_x\rangle_A|\uparrow_x\rangle_B)\right]$$

and, given the properties of good measuring devices, evolves by the linear dynamics to the unstable state

$$\frac{1}{\sqrt{2}}(|\text{“}\uparrow_x\text{”}\rangle_{M_A}|\text{“}\downarrow_x\text{”}\rangle_{M_B}|\uparrow_x\rangle_A|\downarrow_x\rangle_B + |\text{“}\downarrow_x\text{”}\rangle_{M_A}|\text{“}\uparrow_x\text{”}\rangle_{M_B}|\downarrow_x\rangle_A|\uparrow_x\rangle_B).$$

When the *first* particle in the pointer of either M_A or M_B collapses to an eigenstate of position, the composite system will collapse to either

$$|\text{“}\uparrow\text{”}\rangle_{M_A}|\text{“}\downarrow\text{”}\rangle_{M_B}|\uparrow_x\rangle_A|\downarrow_x\rangle_B$$

or

$$|\text{“}\downarrow\text{”}\rangle_{M_A}|\text{“}\uparrow\text{”}\rangle_{M_B}|\downarrow_x\rangle_A|\uparrow_x\rangle_B$$

with probability 1/2 for each. The first pointer particle to collapse disentangles the state of the composite system and instantaneously gives both its own pointer and the pointer of the distant system determinate positions that correspond, respectively, to the determinate x-spins it instantaneously gives the two particles A and B. Since there is no physical matter of fact concerning the *order* of spacelike-separated events, if M_A and M_B are sufficiently far apart and if they make their measurements at the nearly same time in the laboratory frame of reference, there will be no physical matter of fact concerning which pointer particle collapsed first. But that means that there can be no consistent dynamical story for how the systems disentangle if the two particles are described by a single wave function.

The incompatibility of GRW* and relativity can also be seen by the role played by the new physical collapse rate constant λ. Since relativistic clocks disagree in different inertial frames, one must choose a preferred inertial frame in order to even specify λ. But special relativity requires that there be no preferred inertial frame.

The incompatibility with relativity is particularly striking given the privileged dynamical role played by *exact position* in GRW*. If there were eigenstates of exact position, the expectation value of energy of such states would have to be unbounded. The reason for this has to do with the quantum-mechanical relationship between position and momentum. The quantum state representing the position of a particle p is given by a complex-valued function $|\psi(r)\rangle$ over all possible positions. The value of this function given the probability of finding the particle in a spacial region R is

$$P(R) = \int_R |\langle\psi(r)|\psi(r)\rangle|^2 dr.$$

Suppose that $|\psi(r)\rangle$ is a gaussian wave packet with a probability density given by $|\langle\psi(r)|\psi(r)\rangle|^2$. The same state can be expressed as a complex-valued function of momentum $|\phi(p)\rangle$ with a probability density given by $|\langle\phi(p)|\phi(p)\rangle|^2$. The two wave functions $|\psi(r)\rangle$ and $|\phi(p)\rangle$ are related to each other by a mathematical operation called a Fourier transform. This operation has two properties that matter here. First, if $|\psi(r)\rangle$ is gaussian, then so is $|\phi(p)\rangle$. Second, the narrower the gaussian wave packet $|\psi(r)\rangle$, the wider the gaussian wave packet $|\phi(p)\rangle$ and the other way around (as in Figure 8.2).[5] This means that the sharper particle p's position, the less sharp its momentum. And this means that the more determinate p's position, the higher the probability of finding it with a large momentum. In the limit as p's position becomes perfectly determinate, the expectation value of its momentum is unbounded, which means that the expectation value for its energy would also be unbounded. The upshot is that if any particle in a measuring-device pointer ever did collapse to a state where it had a perfectly precise position, the device would explode, destroying the universe. This renders GRW* incompatible with relativity insofar as one takes relativity to require the conservation of energy. Inasmuch as the universe is still here, it also means that GRW* is not even close to being empirically adequate.

[5] The variable p in the expressions here is particle p's momentum, and r is position.

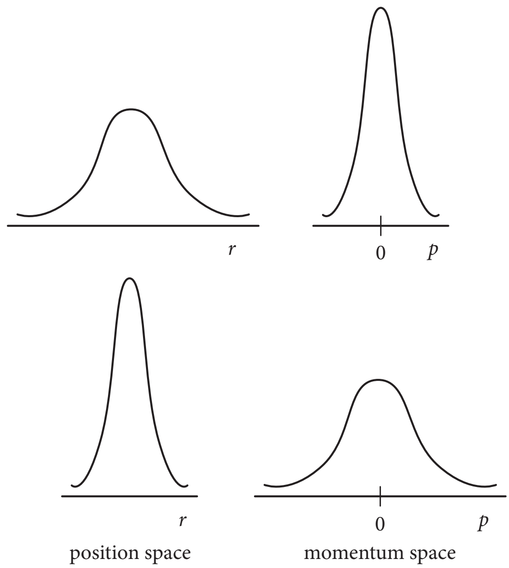


Figure 8.2. Wave functions in position and momentum space.

8.3 GRW

GRW explicitly recognized that GRW* is empirically unacceptable. The theory that they ultimately presented, which we will just call GRW, is similar to GRW* except that they replaced the dynamics of the standard formulation of quantum mechanics with a more complicated hybrid law:

4. Law of motion (GRW): The state of every physical system S evolves in the standard deterministic, linear way $|\psi(t_1)\rangle_S = \hat{U}(t_0, t_1)|\psi(t_0)\rangle_S$ except that each particle has a small probability λ per unit time of randomly collapsing to a narrow gaussian wave packet in position with width $1/\sqrt{\alpha}$. The probability of the wave packet being centered in a specified region R is given by the standard quantum probabilities.

In contrast with the GRW* dynamics, the state here evolves as if an observer made an *unsharp* quantum measurement of the position of a randomly selected particle at random times and caused a corresponding unsharp collapse to a narrow gaussian state that is *close* to an eigenstate of position. But, again, on the hybrid dynamics this happens at a particular moment to the wave function in $3N$-dimensional configuration space randomly and without the necessity of an observation.

The hybrid dynamics still fails to conserve energy, but GRW chose the width of the collapsed gaussian wave packet $1/\sqrt{\alpha}$ to be *wide enough*, given the particle collapse rate, to keep the violation of the conservation of energy low enough to explain why we have not seen it yet. They also chose the width of the gaussian to be *narrow enough* so that a collapsed system in that state might plausibly be said to have an *almost* determinate position (specifically, they took $1/\sqrt{\alpha} \approx 10^{-5}$cm). Their choice of $\lambda \approx 10^{-16}$/sec makes collapses rare for those systems we know behave linearly and common for those systems that we take to have determinate positions. Each of the two new physical parameters are carefully tuned to mesh with the empirical evidence we have so far. So while all this is clearly ad hoc, it works pretty well.

Consider this hybrid dynamics in the context of a measurement of the x-spin of a system S. Suppose S starts in the symmetric superposition

$$\frac{1}{\sqrt{2}}(|\uparrow_x\rangle_S + |\downarrow_x\rangle_S).$$

The linear dynamics, then, predicts that the state of the composite system $M + S$ will be

$$\frac{1}{\sqrt{2}}(|\text{“}\uparrow_x\text{”}\rangle_M |\uparrow_x\rangle_S + |\text{“}\downarrow_x\text{”}\rangle_M |\downarrow_x\rangle_S). \tag{8.1}$$

If M records its result in the position of a macroscopic pointer, then this state is unstable under rule 4(GRW) because of the choice of collapse rate. But this time, when the first pointer particle collapses, it collapses to a narrow gaussian wave packet instead of an eigenstate of position with the center of this wave packet determined by the usual quantum probabilities. Suppose that the first pointer particle to collapse is m_{17}. Its wave packet has equal chance here of ending up centered in region “$\uparrow_x$” and centered in region “$\downarrow_x$”. Suppose it collapses to a state that is centered in region “$\uparrow_x$”. Since the resultant state of m_{17} is a gaussian wave packet, it is still in a superposition of being in region “$\uparrow_x$” *and* in region “$\downarrow_x$”. It is just that the norm-squared of the amplitude associated with it being in region “$\downarrow_x$” is now very small. The effect of m_{17}'s collapse, then, is to multiply the second term of the state above by a factor that is just slightly greater than zero. After renormalizing, the resultant state is

$$|\text{grw}\rangle_{MS} = \alpha|\text{“}\uparrow_x\text{”}\rangle_M |\uparrow_x\rangle_S + \beta|\text{“}\downarrow_x\text{”}\rangle_M |\downarrow_x\rangle_S,$$

where $|\alpha|^2 \gg |\beta|^2$. Note that $|\text{grw}\rangle_{MS}$ is not an eigenstate of M's pointer indicating the result “$\uparrow_x$” or anything else. On the standard interpretation of states, this is a state where M is close to recording a determinate measurement result but doesn't, S is close to having a determinate x-spin but doesn't, and the two systems are close to being disentangled but aren't. So this is a state where M fails to record or indicate any determinate measurement outcome whatsoever on the standard interpretation of states.

This is known as the *tails problem*. Under plausible assumptions, $|\text{grw}\rangle_{MS}$ is *close* (in the Hilbert space metric determined by the inner product) to a state where the measurement record is determinately x-spin up. But if one wants to say that $|\text{grw}\rangle_{MS}$ *is* a state where M determinately indicates the result x-spin up, then one needs a new way to interpret quantum-mechanical states.

To this end, one might further tune the theory by replacing the standard interpretation of states with a new rule 3(GRW):

3. Interpretation of states (GRW): A system S has a determinate property if and only if it is *sufficiently close* to being in an eigenstate of having that property.

Here one can either choose a concrete standard of what one means by *sufficiently close* once and for all or, as with classical treatments of vague predicates, allow the standard to depend on the explanatory context at hand. One advantage of the latter is that it does not involve adding another physical parameter to the theory. If one wants to explain why one's measuring device does not explode, one might note that neither the pointer nor any of the particles that constitute it are in fact in eigenstates of position. But if one wants to explain why the pointer appears to record the determinate outcome x-spin up, one might argue that its center of mass is close enough to an eigenstate of position that it should count as having the corresponding determinate position.

That even classical predicates are typically *vague* for macroscopic systems gives some plausibility to the proposed revision of the standard interpretation of states. Consider the claim that the Eiffel Tower is in Paris. No one would take this to mean that *every atom of iron that was present in the original structure in 1889* is now in Paris. Some of the original Eiffel Tower atoms are almost certainly in New York, which does precisely nothing to undermine the every-day truth that the Eiffel Tower is in Paris. The predicate "is in Paris" is vague, but still useful. The fact that the Eiffel Tower is in Paris (even when some of its original atoms are in New York) explains why one sees the Eiffel Tower when one looks north-west down the Champ de Mars. The thought is that the vague claim that the pointer on a particular measuring device records the result x-spin up is similarly useful. The fact that the pointer indicates the result x-spin up (even when it is not in an exact eigenstate of position) explains why one sees it pointing at the result x-spin up.

One might, however, object that there is an important disanalogy between the two cases. The Eiffel Tower *appears* to be in Paris on classical grounds because *most of its (microscopic) parts* are simply and determinately in Paris. But GRW does not characterize any of the parts of the pointer as simply and determinately pointing at any particular measurement result. Because of the infinite tails of the gaussian wave packet, one's finest-grained description will always characterize the pointer and its parts as being in an entangled superposition of recording all possible measurement outcomes, being part of the motor of a taxi in Manhattan, being in orbit around α Centauri, etc.

The real issue, one might insist, is not the *language* of vague predicates. Rather, it is whether a pointer with an almost determinate position in the sense provided by GRW *looks like* a pointer with a fully determinate position. It concerns whether the theory correctly predicts our determinate *experience* of seeing macroscopic objects with perfectly determinate positions.

For its part, the theory says that a physical observer will typically end up in an entangled superposition of having recorded every possible measurement outcome. One of the records will get most of the quantum-mechanical amplitude, but that does not mean that it will *feel like* seeing a pointer with a fully determinate position. All it means *formally* is that the norm-squared of one coefficient on one term in the composite state is much larger than the other. The empirical adequacy of the theory presumably depends on our being able to say what one would *experience* in such a state. And *that* apparently has nothing to do with the language of vague predicates.

That said, it is not at all clear that we need the center of mass of a system to be *exactly* determinate in order to explain why it appears to have a determinate classical position. Indeed, inasmuch as the expectation value for energy would be unbounded under standard quantum assumptions, we have *never* seen a pointer or any other physical system in an exact eigenstate of position, and we never want to see it. Since the standard formulation of quantum mechanics does not even allow one to represent such states, if that theory *ever* explained the apparent determinate positions of macroscopic objects, then GRW arguably does as well. Further, it is not at all clear that we need a physical theory to tell us what it *feels like* to be in a particular physical state. It is, one might argue, enough for there to be *something physical* on which our determinate experience might plausibly be taken to *supervene*. And GRW arguably provides physical records that do *that*.

There is more to say about what one should want from a physical theory that purports to be empirically adequate and hence to explain our experience. This will be a recurring theme throughout the rest of the chapter and book. That it need only predict *something physical* on which our determinate experience might be taken to *supervene* is arguably too weak. We want the supervenience relation to be one that meshes with our beliefs concerning the sorts of physical facts that might in fact matter to experience. Starting with the next section, we will see how this plays out in a number of concrete formulations of quantum mechanics.

Like GRW*, GRW has the salient virtue of treating all physical processes, including measurement, in precisely the same way, but, unlike GRW*, it has the added virtue of predicting a relatively stable universe. GRW conclude that their theory "reproduces in a consistent way quantum mechanics for microscopic objects and classical mechanics for macroscopic objects, and provides the basis for a conceptually appealing description of quantum measurement" (Ghirardi et al. 1986, 491). But it also exhibits a number of unsatisfactory features.

Instead of the position being the preferred physical quantity as in GRW*, here it is *almost position with gaussian width* $1/\sqrt{\alpha}$. Inasmuch as choosing position as a preferred physical quantity and specifying a just-right collapse rate of λ was ad hoc, this choice makes the theory look still more ad hoc. Indeed, insofar as GRW *explicitly* constructed their theory and chose the collapse rate λ and the gaussian width $1/\sqrt{\alpha}$ so that microscopic systems will almost always behave linearly and hence exhibit distinctive quantum behavior and the centers-of-mass of macroscopic systems will almost always remain close to eigenstates of position, exhibit roughly classical behavior, and not heat up so much that we would have already noticed, the theory is manifestly ad hoc from the start.

In addition to being ad hoc, there is nothing subtle about the incompatibility of this version of GRW and relativity. Since clocks in different inertial frames disagree, one must choose a preferred inertial frame to specify the collapse rate λ; but here, since different inertial frames do not agree on measurements of length either, one must also choose a preferred inertial frame to specify the width $1/\sqrt{\alpha}$ of the collapsed wave packet.

Consider the EPR experiment again. A collapse of a single particle in either of the two pointers in GRW will instantaneously collapse the state of the composite system just as in the standard theory. Here it will give both pointers almost determinate positions; both particles almost determinate, almost opposite spins; and almost disentangle the systems. The statistical distribution of such records will be the same as predicted by the standard collapse theory. But like both the standard theory and GRW*, one must be able to specify the temporal order in which spacelike separate events occur in GRW in order to tell a consistent dynamical story, and special relativity says that there is simply no physical matter of fact concerning the temporal order of spacelike separated events. The incompatibility of GRW and relativity can also be seen in the violation of the conservation of energy, something that relativity arguably requires. In short, if one believes that special relativity is right, then this version of GRW can't be.

8.4 GRWr, GRWm, and GRWf

There are various ways of understanding GRW. The ontological commitments associated with each provide different accounts of experience.

Inasmuch as they take macroscopic systems to be constituted by particles, one might imagine that Ghirardi, Rimini, and Weber are committed to an ontology that includes particles. As they put it, a composite system is "a system of *N* particles" and the localization process (the spontaneous collapse of a particle) occurs "independently for each constituent of a many-particle system" (1986, 476). But this is at least somewhat misleading. The theory characterizes the state of the particles in terms of the wave function of the composite system in configuration space. Consequently, one cannot appeal to the *particle-like properties* of the particles to account for the experience of observers. Rather, the talk of particles ultimately just serves to characterize the degrees of freedom of the system and hence the structure of the state space. Physically, the theory provides nothing beyond the wave function of the composite system and how it evolves to account for our experience.[6] As a result, the most straightforward way to interpret GRW is arguably in terms of a direct wave function realism.

In parallel with the terminology we will use for other interpretations of the GRW formalism, we will call this GRWr.[7] Here one takes the wave function in configuration space to be a concrete physical object. As David Albert describes the view,

> The sorts of metaphysical objects that wave functions are, on this way of thinking, are (plainly) *fields*—which is to say that they are the sorts of of objects whose states one specifies by specifying the values of some set of numbers at every point in the arena in which they live, the sorts of objects whose states one specifies (in this case) by specifying the values of two numbers . . . at every point in the configuration space of the universe. (2013, 53)

The theory is just GRW understood as describing a field in configuration space as the only physically real thing. As Albert puts it, the world consists of "exactly one physical object—the universal wave function" (Albert 2013, 54).

This view may seem to be closely analogous with the Eiffel Tower example. Here one might imagine that a measuring device pointer indicates a particular result because *most of the field stuff* that constitutes the pointer in fact indicates that result after a measurement-like interaction. There is something right about this, but the situation is more subtle than this quick characterization might suggest.

The wave function $|\psi(r)\rangle_S$ is a complex-valued function over points r in a $3N$-dimensional configuration space, where N is the number of particle degrees of freedom in system S.[8] As a result, the wave function representing the world is a function in $3N$-dimensional configuration space, where N is the total number of particles in the universe, something many orders of magnitude greater than Avogadro's number. So if one takes $|\psi(r)\rangle_S$ to directly represent a physical object,

[6] On this view, particles cannot be understood as providing a primitive ontology for the theory. We will discuss the significance of this as we go.

[7] Inasmuch as this view might be understood as simply sharpening the ontological commitments of GRW themselves, one might take GRW and GRWr to be the same theory. Calling it GRWr here just marks the explicit interpretation of the wave function as a physically real field in configuration space.

[8] As discussed in section 5.7.

the *only* physical object, then it is an object that lives in an unimaginably high-dimensional space, not in ordinary three-dimensional space. It is *this object* that we are seeing when we observe pointers, baseballs, coffee cups, and such. Indeed, we are ourselves also just manifestations of this single, high-dimensional object. One is left, then, with the task of providing a dynamical story for how this object gives rise to observers who seem to experience everyday three-dimensional macroscopic objects situated in ordinary three-dimensional space. That one can do precisely that should seem plausible. Given how the wave function evolves in GRW, there is clearly something on which the standard quantum measurement records might supervene. But the story one tells here requires some care.

In order to account for an observer's experience of *seeing* an ordinary three-dimensional object where she sees it, one needs a special sort of supervenience relation. Specifically, one needs to suppose that one's experience *roughly supervenes in a particular way* on the observer's brain-record degrees of freedom of the full wave function in 3*N*-dimensional configuration. One needs a supervenience relation that says something like: whenever the quantum state is within some cutoff distance ε of a brain-record eigenstate in Hilbert space, then one has a determinate experience corresponding to that eigenstate. This special supervenience relation provides an interpretational principle that ties physical states to experience. What one is *seeing* when one looks at an ordinary extended object here are coarse-grained features of the wave function in 3*N*-dimensional configuration space that arise from the correlations between the physical degrees of freedom.

On GRWr all one has to explain one's experience is an evolving field in a high-dimensional space. It is the correlations in the degrees of freedom of the 3*N*-dimensional wave function that produce the *illusion* of ordinary three-dimensional objects in ordinary three-dimensional space behaving as we find them. Space *seems* to be three-dimensional because of the structure of these correlations. Specifically, it is because the Hamiltonian treats the degrees of freedom in 3*N*-dimensional configuration space as if they described a collection of *N* particles interacting in three dimensions.[9] The *determinate experience* of a three-dimensional world inhabited by ordinary three-dimensional objects in GRWr results from (1) the precise structure of correlations in the 3*N*-dimensional wave function under the GRW dynamics, and (2) the special supervenience relation that relates mental states and the 3*N*-dimensional wave function.[10] But there are other ways to understand GRW.

On GRWm one starts with the wave function in configuration space under the GRW dynamics but understands it as coding for a continuous distribution of

[9] The Hamiltonian gives the energy properties of a physical system—in this case the entire physical world. In so doing, it determines the linear operator that shows up in the standard quantum dynamics by describing how the particle degrees of freedom are dynamically related.

[10] See Albert (2013) and (2015) for discussions regarding how one might account for experience in GRW-type theories. See also Maudlin (2019) for a critical discussion of configuration space realism.

matter *in ordinary three-dimensional space*.[11] Specifically, one uses the wave function in configuration space to calculate the marginal mass distribution associated with each particle in three-dimensional space then sums over the mass densities for all *N* particles.[12] This gives a description of the physical world in terms of a continuous distribution of matter at each time. We and everything we see are the result of this continuous matter sloshing about in ordinary three-dimensional space.

Inasmuch as this view involves a matter field in ordinary space, it might also seem to be closely analogous to the Eiffel Tower example. Here one might imagine that one sees the measurement pointer where one does because it is just a pointer-shaped bump in the distribution of matter and one simply sees the location of the bump when one looks for the pointer. But again, the situation is more subtle than it might at first appear.

Here one needs to spell out the details of the dynamical story that ties one's brain record of the position of the pointer to the mass density that constitutes the pointer. Inasmuch as this story will involve details of how the wave function evolves in 3*N*-configuration space, it is unclear that the account of experience one ends up with is significantly more intuitive than on GRWr. The point is that one is not *directly seeing* the pointer on GRWm. Observation is mediated either by the universal wave function in 3*N*-dimensional configuration space or by something like a brute-fact law for the evolution of the matter density that codes for the dynamical information in the wave function. In neither case is this particularly intuitive. We will return to this issue of intuitiveness in connection with GRWf since it arguably has all of the virtues of GRWm plus a few extra.

On GRWf the wave function in configuration space determines the probability densities of *flashes* in ordinary 3-space, and it is constellations of such flashes that explain the appearance of macroscopic objects.[13] One might calculate the location of flashes by evolving of the wave function in configuration space under the standard GRW dynamics, then associating a flash with the spacetime point at the center of the gaussian wave packets that are generated by the random collapses. Since flashes just occupy spacetime locations, they are not the sort of things that have energies. Hence, one might argue, there is in fact no actual violation of conservation of energy.[14]

Allori, Goldstein, Tumulka, and Zanghì describe how GRWf accounts for our experience:

[11] This view was proposed by Benatti et al. (1995). See Allori (2013) for a description of the approach and a comparison with other options.

[12] See Allori, Goldstein, Tumulka, and Zanghì (2012) for a prescription for how to do this.

[13] The inspiration for this view can be traced to John Bell's (1987, 181–95) discussion of quantum interference phenomena. See also Tumulka (2007) and Allori (2013, 67–8) for descriptions of this approach.

[14] That said, insofar as the flashes constitute pointers that indicate the same measurement outcomes as pointers in GRW, it would nevertheless *appear* that there was a small violation of conservation of energy—precisely the violation predicted by standard GRW.

> [F]or a reasonable choice of the parameters of the GRWf theory, a cubic centimeter of solid matter contains more than 10^8 flashes per second. That is to say that large numbers of flashes can form macroscopic shapes, such as tables and chairs. That is how we find an image of our world in GRWf. (2012, 6)

But if one takes the flash ontology seriously, flashes are more than just the manifest image of the world—there is a sense in which they *are* the world. On this account, physical objects are *constituted* by constellations of flashes. There is, however, an important sense in which we never *see* the flashes that constitute an extended object.

The flashes that constitute an object are not themselves the manifest image of the object. Inasmuch as the flashes do not show up in the theory's dynamics, they do not *do* anything. Specifically, they do not cause an observer's experience of the positions, shapes, or anything else of the extended three-dimensional objects they constitute. Flashes just occupy spacetime points, distributed with the standard quantum statistics.

To be sure, the theory predicts that there will be a *statistical correlation* between the flashes that constitute an extended object and the flashes that constitute a good observer's brain records of that object. So if one explains one's determinate experience by supposing that it supervenes on the value of one's *brain-record flashes*, one might infer probabilities concerning the position or shape of an extended object from one's experience. But one's experience does not supervene on the flashes that constitute the *extended objects* one takes oneself to see. Rather, it supervenes on the flashes that constitute one's *brain records*, assuming that one can make good sense of that. Again, the correlation between the two is mediated by either the wave function in 3*N*-dimensional configuration space or by something like a brute-fact law that codes for the dynamical information in the wave function and determines the correlated locations of the flashes.

That flashes have no dynamical role in GRWf is an important feature of the theory. It allows for a sort of dynamical compatibility between GRWf and relativity. Since flashes just occupy spacetime points and do not do anything dynamically and since there is nothing non-relativistic about the *statistical distribution* of flashes, one just needs a relativistic description of the evolution of the wave function to get a relativistic formulation of GRWf.[15]

Roderick Tumulka (2007) has shown how to do this. Getting a relativistic description of the evolution of the wave function involves a trick. Consider a situation where we want to represent a pair of distant particles in something like an EPR state. Rather then there being a simple matter of fact regarding their

[15] The distribution of flashes is described by the standard quantum statistics, which do not allow one to send superluminal signals nor do they select any preferred inertial frame. See Bell (1987, 52–62) for one discussion.

composite state or the state of either particle alone, the trick is to allow the wave function of the composite system to depend on the choice of a spacelike hypersurface.[16] Since the wave function is hypersurface-dependent, there is typically no simple matter of fact regarding whether a GRW collapse has occurred at a time. While the state of the composite system depends on one's choice of inertial frame, the states associated with different inertial frames (associated with different families of hyperplanes) can be related in such a way that they are all compatible with the standard quantum statistics for the location of flashes under the GRW dynamics. This is possible because the standard quantum statistics do not themselves select a preferred inertial frame. So the collection of hyperplane-dependent wave functions does not determine a preferred inertial frame precisely because they are hyperplane-dependent, and the statistical distribution of flashes does not determine a preferred inertial frame either. Since all one has are the hyperplane-dependent wave functions and the spacetime flashes, and since neither of these determines an inertial frame, such a formulation of GRWf might be understood as satisfying the dynamical constraints of relativity.

There are two things to note. First, this does not mean that GRWf is *local*. There is a sense in which the standard quantum statistics themselves characterize nonlocal correlations. Specifically, while a hypersurface-dependent formulation of GRWf is dynamically compatible with relativity, the distribution of flashes still violates the Bell inequalities. In this sense, a sense weaker than dynamical incompatibility with relativity, the theory is nonlocal just by dint of being empirically adequate. Second, since there is typically no simple matter of fact concerning the wave function of a system at a time, this means that the flashes must determine the values of the local measurement records. On this view, one might argue, the collection of hypersurface-dependent wave functions just amounts to a part of the statistical *recipe* for where to put flashes. And again, any particular distribution of flashes over spacetime is perfectly compatible with the dynamical constraints of relativity.

While such a theory is dynamically compatible with relativity in the sense just described, there is also a sense in which it is not a relativistic theory at all. One might expect a truly relativistic formulation of quantum mechanics to predict the phenomena that are taken to be characteristic of relativistic quantum mechanics, phenomena like particle creation and annihilation, how photons are found to scatter from a single free electron, etc. One should want *novel* relativistic quantum phenomena to result naturally from imposing relativistic constraints on quantum mechanics. One should not have to put such phenomena in later by hand.[17] The

[16] This idea has a long tradition among people who want to formulate a collapse theory that does not require one to select a preferred inertial frame. See for example Aharonov and Albert (1981). Note that the wave function remains a function in 3*N*-configuration space—it is just (typically) a *different function* for each choice of spacelike hyperplane.

[17] One would, for example, like to be able to *derive* something like the Klein–Nishina (1929) scattering formula from one's relativistic formulation of quantum mechanics. This consideration also

explanation of the phenomena should follow from the relativistic dynamics in the context of quantum mechanics. Settling for compatibility with relativistic constraints by simply stipulating hypersurface-dependent states is nothing like that.[18]

8.5 Empirical Ontology and Experience

We have seen how GRW might be formulated in terms of a physically real field in configuration space, a matter distribution in ordinary three-dimensional space, and flashes at spacetime points. Such options represent different choices of an *empirical ontology*. One's choice of empirical ontology specifies the basic structure of the physical world and hence the entities that one can appeal to in explaining experience.

A recurring issue over the next few chapters concerns the sort of empirical ontology that would allow for a satisfactory account of experience. One approach is to require that the empirical ontology be a *primitive ontology*—an ontology of ordinary objects or fields in three-dimensional space that directly explain our experience.[19] The thought is that if a theory has a primitive ontology, then one can think of the objects of the theory as providing a *manifest image of the world* that simply corresponds to our experience of a three-dimensional world with ordinary three-dimensional objects in motion. Indeed, this is precisely how Allori, Goldstein, Tumulka, and Zanghì characterize how flashes in GRWf explain our experience (2012, 6). Here the primitive ontology is taken to be a suitable surrogate for specifying what is directly observable in the theory.

But we have seen that the situation is significantly more subtle than this might suggest. Given that we do not directly see the flashes that constitute everyday objects on GRWf one needs an account of how it is that the primitive ontology, being what it is (in this case flashes situated in spacetime), makes determinate our most immediately accessible records. That the ontology is *primitive* is supposed

applies in the context of formulations of Bohmian mechanics that seek to account for relativistic phenomena by sticking them in by hand. Everettian formulations of quantum mechanics are arguably the best positioned to explain genuinely relativistic phenomena. But, as we will see over the next two chapters, it is nontrivial for such theories to explain even determinate measurement records.

[18] That said, it may be possible to make some progress here. Tumulka's characterization of the state of a relativistic system is given by a multi-time wave function with one time variable for each of the N particle degrees of freedom. The evolution of the multi-time wave function is given by N linear equations that must satisfy a consistency condition. This condition is automatically satisfied for Hamiltonians describing non-interacting particles, but it is very difficult to satisfy for interacting particles in even the simplest of situations. But there is reason to suppose that the prospects are better for the state of a multi-time quantum field. In addition to allowing for a consistent account of particle-like interactions, such a formulation might allow one to capture something like particle creation and annihilation phenomena. See Lienert, Petrat, and Tumulka (2017) for a discussion.

[19] See Allori (2013) for a description and motivation of this approach.

to help in providing this account, but, insofar as the flashes that constitute the objects we take ourselves to see are themselves not *directly observable* and are mediated by either the 3*N*-dimensional wave function or something like a brute-fact law that codes for the same information, it is unclear why a primitive ontology of three-dimensional objects is automatically preferable to an empirical ontology more generally, perhaps one that like GRWr appeals to the wave function in configuration space directly.

In both GRWr and GRWf one has *something* on which mental states might be take to supervene, but in neither theory is the explanation particularly intuitive. Whether one prefers the account of experience provided by GRWr over the account provided by GRWf turns on whether one finds it more plausible that mental states roughly supervene on brain-record degrees of freedom of the wave function (as required by GRWr) or that they supervene on the positions of brain-record flashes (as required by GRWf). *Primitive ontologists* prefer GRWf because the flashes can be understood as objects with ordinary spatial locations. But one can easily imagine someone who prefers the dynamical and/or metaphysical economy of GRWr.

Ultimately, one wants an *empirical ontology* that provides records on which one's experience might plausibly be taken to supervene. Unsurprisingly, there are competing intuitions regarding precisely what this should be and, hence, what the best formulation of quantum mechanics and the most plausible account of experience will look like. We will return to discuss competing empirical ontologies and their relative virtues again when we consider Bohmian mechanics and variants of that theory (Chapters 11 and 12), but we will begin in the next chapter with the general issue of what it should even *mean* for a physical theory to be empirically adequate.

GRW also faces a less subtle empirical issue, one that we should get out of the way before moving on. If physical objects really do appear to warm up as predicted by the standard GRW dynamics, then that would be a remarkable empirical success for the theory. It would make the worries we have expressed along the way seem entirely beside the point. But we arguably already have good inductive reason to believe that the nonlinear behavior predicted by GRW never happens.

Whenever we have been technologically able to measure an interference effect that would determine whether a system evolves linearly or collapses, we have always found that it evolves linearly. This was true for the simplest physical systems, and it has proven true as we have learned how to measure interference effects involving increasingly more complex systems.[20] We are in the possession then of a sort of inductive argument on the complexity of the physical systems we have observed that they always evolve linearly. This does not immediately rule

[20] See Fein et al. (2019) for a description of recent experiments.

out the GRW formulation. GRW themselves explicitly chose a collapse width and collapse rate so that their theory was not ruled out by what had been observed to that time. And, while the region for choosing these parameters has been significantly reduced by careful experimental work since then, there is still some room to make a choice. That the choice is clearly ad hoc is why I would bet against the novel predictions of the theory. It could be right, but I do not believe it.

There is also a significant philosophical issue at hand, one that concerns how one understands physical theories more generally. When one goes from the standard collapse theory to GRW, one changes the dynamical laws of the theory, but, as we have seen, one must also change how one *interprets* the theory by giving up the standard eigenvalue–eigenstate link. This illustrates why it is virtually impossible to separate a physical theory from its interpretation.

The various GRW-type theories we have considered explain determinate measurement outcomes differently and their explanations are all radically different from the standard theory's. The special dynamical role that position plays in GRW goes hand in hand with how states are interpreted in GRWr, GRWm, and GRWf. The mathematical specification of the physical theory and the physical interpretation of the mathematical objects *work together* to make predictions and provide explanations. Inasmuch as physical theories provide explanations and those explanations depend on how one understands the theory, one's interpretational principles are an essential part of one's physical theory.

9

Pure Wave Mechanics

9.1 Everett's Solution to the Measurement Problem

Hugh Everett III developed his relative-state formulation of quantum mechanics while a graduate student in physics at Princeton University. His doctoral thesis (1957a) was accepted in March 1957 and a paper covering essentially the same material was published in July of the same year (1957b).[1] DeWitt and Graham (1973) later published Everett's longer, more detailed description of the theory in a collection of papers on the topic. The published version was revised from a longer draft thesis that Everett (1956) had given John Wheeler, his Ph.D. adviser, in January 1956 under the title "Wave Mechanics Without Probability."[2] While Everett always favored the description of the theory as presented in the longer thesis, Wheeler, in part because of Bohr's disapproval of Everett's critical approach, insisted on the revisions that led to the much shorter thesis that Everett ultimately defended.

Everett's proposal for solving the quantum measurement problem was to simply drop the collapse dynamics (rule 4II) from the standard formulation of quantum mechanics and take the resulting physical theory to provide a complete and accurate description of all physical systems in the context of all possible physical interactions. Everett called the resulting theory *pure wave mechanics*. When supplemented with the notions of *relative states* and *typical branches*, he called it the *relative-state formulation* of pure wave mechanics.[3]

There is a sense in which dropping the collapse dynamics from the standard formulation of quantum mechanics immediately solves the quantum measurement problem. Since there is only one dynamical law, there is no longer a threat of a contradiction and the measurement process is a physical process just like any other. Further, as mentioned at the end of the last chapter, the fact that the linear dynamics has always been found to correctly describe the time-evolution of a

[1] See Barrett and Byrne (2012, 173–96) for the text of the short thesis with notes and historical and conceptual introductions. References to Everett's work here follow the page numbering of that collection. See also Byrne (2010) for a biography of Everett and Osnaghi, Freitas, and Freire (2009) for a description of the reception of Everett's work.

[2] See Barrett and Byrne (2012, 72–172) for the text of the long thesis with notes.

[3] See Barrett and Byrne (2012) and Barrett (2014) for detailed descriptions of Everett's relative-state formulation of quantum mechanics. See Barrett (2015) and (2017) for more details regarding the arguments in this and the next chapter.

system whenever we have been able to test its predictions provides good empirical reason for dropping the collapse dynamics.

Simply dropping the collapse dynamics, however, immediately leads to two new problems. Without the collapse dynamics, one needs another way of accounting for *determinate measurement records* and for *quantum probabilities*. After all, quantum mechanics is only empirically adequate because it predicts determinate measurement records distributed with the standard quantum statistics.

Everett believed that he could *deduce* the standard statistical predictions of quantum mechanics, the predictions that depend on rule 4II in the standard collapse formulation, in terms of the subjective experiences of observers who are themselves treated as ordinary physical systems within pure wave mechanics. He described his proposed deduction as follows:

> We shall be able to introduce into [pure wave mechanics] systems which represent observers. Such systems can be conceived as automatically functioning machines (servomechanisms) possessing recording devices (memory) and which are capable of responding to their environment. The behavior of these observers shall always be treated within the framework of wave mechanics. Furthermore, we shall deduce the probabilistic assertions of Process 1 [rule 4II] as subjective appearances to such observers, thus placing the theory in correspondence with experience. We are then led to the novel situation in which the formal theory is objectively continuous and causal, while subjectively discontinuous and probabilistic. While this point of view thus shall ultimately justify our use of the statistical assertions of the orthodox view, it enables us to do so in a logically consistent manner, allowing for the existence of other observers. (1956, 778)

His project, then, was to show that the memory records of an observer as described by quantum mechanics without the collapse dynamics would agree with those predicted by the standard formulation with the collapse dynamics. This ultimately amounted to showing that a *typical relative observer*, modeled as a physical automaton within pure wave mechanics, would have fully determinate *relative measurement records* that were *statistically distributed* as if they resulted from collapses with the standard quantum probabilities. Everett's notion of a typical relative observer associated with a typical branch of the quantum-mechanical state played a central role in his explanation of quantum statistics.

To see how this was supposed to work, let's return to the Wigner's friend story but consider it as Everett told it. Suppose that the friend F, her measuring device, and her object system begin in the state[4]

$$|r\rangle_F |r\rangle_M (\alpha|\uparrow_x\rangle_S + \beta|\downarrow_x\rangle_S).$$

[4] The notation r indicates that the associated system is ready to make a measurement. As the states we consider become increasingly complicated, we will simplify notation where possible.

Given the properties of a good measuring device and observer, the linear dynamics predicts that the resultant state will be

$$|\text{everett}\rangle_{FMS} = \alpha|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + \beta|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S.$$

Everett insisted that since there are no collapses of the quantum-mechanical state, the external observer W is right to believe that the composite system FMS is in this entangled superposition of having recorded mutually incompatible results. Further, he believed that an A-type measurement would empirically show that this was the case. As he put this point more generally:

> It is . . . improper to attribute any less validity or "reality" to any element of a superposition than any other element, due to [the] ever present possibility of obtaining interference effects between the elements. All elements of a superposition must be regarded as simultaneously existing. (1956, 150)

But he also knew that it would be extraordinarily difficult to actually perform an A-type interference measurement to show that this is the state that in fact obtained. This is what he meant when he said that the story was "extremely hypothetical." Nevertheless, he believed that for the sake of the empirical consistency of quantum mechanics the entangled superposition $|\text{everett}\rangle_{FMS}$ must be the final state of the composite system.

That said, Everett also believed that it would *appear* to the friend that she had a perfectly ordinary and fully determinate measurement record if she were in the state $|\text{everett}\rangle_{FMS}$. One sense in which this is true is that, under plausible assumptions, an ideal friend would have the sure-fire disposition to *report* that she has a perfectly ordinary measurement result.[5]

9.2 The Bare Theory

Let's start with the question of what the world would look like to the friend if there were no collapses on measurement. To do so, we will consider the dispositions of an ideal observer in the context of the *bare theory*. The bare theory is just the standard collapse formulation of quantum mechanics but without the collapse dynamics (rule 4I). And we will suppose that we can read the friend's experiences and beliefs from what she would report about her physical measurement records.

[5] There are several quite different ways to reconstruct Everett's formulation of quantum mechanics. In this chapter we will try to stay as close as possible to what he actually said. We will discuss alternative Everettian formulations of quantum mechanics in the next chapter.

The bare theory has only the eigenvalue–eigenstate link (rule 3) as an interpretational principle. Everett adds a number of auxiliary assumptions to the bare theory to get his relative-state formulation. But since it is simple and has a number of suggestive properties, the bare theory is a good place to start. The bare theory is one way of understanding what Everett meant by pure wave mechanics.

The bare theory's suggestive properties will each correspond to an *absolute* property of an observer in Everett's relative-state formulation. Absolute properties are properties that hold for *almost all* of an observer's relative states in the norm-squared amplitude measure. We will return to this point later. We will start by discussing each of the bare theory's suggestive properties in turn.

9.2.1 Determinate result

The first suggestive property of the bare theory concerns why it might seem that there is a determinate measurement result when there isn't one. The entangled superposition $|\text{everett}\rangle_{FMS}$ predicted by the bare theory does not describe the friend as having a determinate measurement record on the eigenvalue–eigenstate link, but there is a sense in which it describes the friend as *believing* that she has a perfectly ordinary, fully determinate measurement result—something similar to what one would expect from a collapse of the quantum-mechanical state.

Suppose that F is reliable at reporting that she has a determinate measurement record when she does and that she is reliable at reporting that she does not have a determinate measurement record when she does not. Suppose further that F's sure-fire dispositions to report in fact reliably indicate her experiences and beliefs. It is the assumption that an observer's experience supervenes on her physical dispositions in a straightforward way that allows Everett to deduce appearances from physical states.

Consider what happens if W asks F whether she had a measurement record *before* F measures the x-spin of a system S that is x-spin up. Let $|r\rangle_W$ be the state where W is ready to ask the question and $|r,\ r\rangle_F$ be the state where F is ready to answer the question and ready to make the measurement. Given the dispositions of a reliable friend, the composite physical system would evolve as follows:

$$|r\rangle_W |r,\ r\rangle_F |r\rangle_M |\uparrow_x\rangle_S$$
$$\downarrow$$
$$|\text{“no”}\rangle_W |\text{“no”}, r\rangle_F |r\rangle_M |\uparrow_x\rangle_S.$$

Here F says “no” since she has not performed the measurement yet, and W hears “no” because that's what F said because that's what she believes.

Consider what happens if *F* performs her measurement and gets one of the two results, *then W* asks *F* whether she got a perfectly ordinary, fully determinate result. If *F* got the result *x*-spin up, then the composite system would evolve as follows:

$$|r\rangle_W |r,\ \text{“}\uparrow_x\text{”}\rangle_F |\text{“}\uparrow_x\text{”}\rangle_M |\uparrow_x\rangle_S$$
$$\downarrow$$
$$|\text{“yes”}\rangle_W |\text{“yes”},\ \text{“}\uparrow_x\text{”}\rangle_F |\text{“}\uparrow_x\text{”}\rangle_M |\uparrow_x\rangle_S.$$

Here *F* says “yes” since she in fact got a determinate result, and *W* hears “yes” because that’s what *F* said because that’s what she believes. The composite system would also have the disposition:

$$|r\rangle_W |r,\ \text{“}\downarrow_x\text{”}\rangle_F |\text{“}\downarrow_x\text{”}\rangle_M |\downarrow_x\rangle_S$$
$$\downarrow$$
$$|\text{“yes”}\rangle_W |\text{“yes”},\ \text{“}\downarrow_x\text{”}\rangle_F |\text{“}\downarrow_x\text{”}\rangle_M |\downarrow_x\rangle_S.$$

Here *F* says “yes” since she in fact got a determinate result, and *W* hears “yes” because that’s what *F* said because that’s what she believes.

Now suppose instead that the composite system starts in the entangled superposition represented by $|\text{everett}\rangle_{FMS}$ when *W* asks his question. Given the linear dynamics, the composite system evolves as follows:

$$\begin{aligned}
&|r\rangle_W(\alpha|r,\ \text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S + \beta|r,\ \text{“}\downarrow_x\text{”}\rangle_F|r,\ \text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S)\\
&\quad = \alpha|r\rangle_W|r,\ \text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S + \beta|r\rangle_W|r,\ \text{“}\downarrow_x\text{”}\rangle_F|r,\ \text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S\\
&\quad \downarrow\\
&\alpha|\text{“yes”}\rangle_W|\text{“yes”},\ \text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S + \beta|\text{“yes”}\rangle_W|\text{“yes”},\ \text{“}\downarrow_x\text{”}\rangle_F|r,\ \text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S\\
&\quad = |\text{“yes”}\rangle_W(\alpha|\text{“yes”},\ \text{“}\uparrow_x\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S + \beta|\text{“yes”},\ \text{“}\downarrow_x\text{”}\rangle_F|r,\ \text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S).
\end{aligned}$$

Here *F* answers the question with “yes” not because she has a determinate result but because the linear dynamics *requires* her to report that she has a determinate result given her dispositions to report so in the two cases where she does in fact have one or the other of the possible determinate measurement records. So, assuming that her reports in fact accurately describe her experience and beliefs, it will *appear* to the friend that she has a perfectly ordinary, fully determinate measurement result when she doesn’t.

This is what it *feels like* to be Wigner’s friend. She will *believe* that the pointer on *M* indicates either the determinate result “$\uparrow_x$” or the determinate result “$\downarrow_x$”, that precisely one of these two options is realized, and that it is as sharply and precisely realized as in the two simpler cases. And if *W* asks her if the description of her cognitive state in the last sentence is accurate, she will report that it is. The upshot is that it will look like a collapse has occurred and produced a single determinate result when the friend is in fact in an entangled superposition of having recorded contradictory measurement results.

Assuming then that the observer's sure-fire report of her experience in fact represents her experience, the linear dynamics produces the *illusion* of a perfectly ordinary, fully determinate measurement result when there isn't one. But this is a curious sort of experience. It is an experience of there being determinate content to a measurement record when the record in fact has no determinate content. Assuming that she believes precisely what she has the determinate disposition to report, the friend believes that she got a determinate result of either x-spin up or x-spin down since she is in an eigenstate of agreeing to this characterization. But it is also the case that there is no *particular* spin result that she determinately believes that she got since she is not in an eigenstate of reporting that she got x-spin up and she is not in an eigenstate of reporting that she got x-spin down.

In his argument that the friend must cause a collapse of the state, Wigner appealed to a principle of charity. The friend must cause a collapse and hence ultimately have a perfectly determinate physical record because Wigner thought that he himself would in similar circumstances. He thought that he knew what it would be like to be in the friend's position, and he thought that it was obvious that he would not be in an entangled superposition of having recorded incompatible results. But the determinate-result property suggests that if one were in fact in an entangled superposition of having incompatible measurement results, it would look like one had a determinate measurement record.

9.2.2 Repeatability

It will also *appear* to the ideal friend that her measurement result is repeatable. Suppose F repeats her measurement of S without disturbing it between measurements then W asks F if she got the same result for each measurement. If F is reliable when S is in an eigenstate of x-spin, then the composite system has the disposition:

$$
\begin{gathered}
|r\rangle_W |r,\ \text{``}\uparrow_x\text{''},\ r\rangle_F |\text{``}\uparrow_x\text{''},\ r\rangle_M |\uparrow_x\rangle_S \\
\downarrow \\
|r\rangle_W |r,\ \text{``}\uparrow_x\text{''}, \text{``}\uparrow_x\text{''}\rangle_F |\text{``}\uparrow_x\text{''}, \text{``}\uparrow_x\text{''}\rangle_M |\uparrow_x\rangle_S \\
\downarrow \\
|\text{``yes''}\rangle_W |\text{``yes''},\ \text{``}\uparrow_x\text{''}, \text{``}\uparrow_x\text{''}\rangle_F |\text{``}\uparrow_x\text{''}, \text{``}\uparrow_x\text{''}\rangle_M |\uparrow_x\rangle_S.
\end{gathered}
$$

Here F reports that her results agree since she in fact got the determinate result x-spin up both times she measured S. And the disposition:

$$|r\rangle_W|r,\ \text{“}{\downarrow_x}\text{”},\ r\rangle_F|\text{“}{\downarrow_x}\text{”},\ r\rangle_M|{\downarrow_x}\rangle_S$$
$$\downarrow$$
$$|r\rangle_W|r,\ \text{“}{\downarrow_x}\text{”},\text{“}{\downarrow_x}\text{”}\rangle_F|\text{“}{\downarrow_x}\text{”},\text{“}{\downarrow_x}\text{”}\rangle_M|{\downarrow_x}\rangle_S$$
$$\downarrow$$
$$|\text{“yes”}\rangle_W|\text{“yes”},\ \text{“}{\downarrow_x}\text{”},\text{“}{\downarrow_x}\text{”}\rangle_F|\text{“}{\downarrow_x}\text{”},\text{“}{\downarrow_x}\text{”}\rangle_M|{\downarrow_x}\rangle_S.$$

Here *F* reports that her results agree since she in fact got the determinate result *x*-spin down both times she measured *S*.

Now suppose that the composite system *FMS* starts in the state $|\text{everett}\rangle_{FMS}$, *F* repeats her *x*-spin measurement, then *W* asks *F* if she got the same result both times she measured *S*. By the linear dynamics the composite system evolves as follows

$$|r\rangle_W(\alpha|r,\ \text{“}{\uparrow_x}\text{”},\ r\rangle_F|\text{“}{\uparrow_x}\text{”},\ r\rangle_M|{\uparrow_x}\rangle_S + \beta|r,\ \text{“}{\downarrow_x}\text{”},\ r\rangle_F|\text{“}{\downarrow_x}\text{”},\ r\rangle_M|{\downarrow_x}\rangle_S)$$
$$\downarrow$$
$$|r\rangle_W(\alpha|r,\ \text{“}{\uparrow_x}\text{”},\text{“}{\uparrow_x}\text{”}\rangle_F|\text{“}{\uparrow_x}\text{”},\text{“}{\uparrow_x}\text{”}\rangle_M|{\uparrow_x}\rangle_S + \beta|r,\ \text{“}{\downarrow_x}\text{”},\text{“}{\downarrow_x}\text{”}\rangle_F|\text{“}{\downarrow_x}\text{”},\text{“}{\downarrow_x}\text{”}\rangle_M|{\downarrow_x}\rangle_S)$$
$$\downarrow$$
$$|\text{“yes”}\rangle_W(\alpha|\text{“yes”},\ \text{“}{\uparrow_x}\text{”},\text{“}{\uparrow_x}\text{”}\rangle_F|\text{“}{\uparrow_x}\text{”},\text{“}{\uparrow_x}\text{”}\rangle_M|{\uparrow_x}\rangle_S + \beta|\text{“yes”},\ \text{“}{\downarrow_x}\text{”},\text{“}{\downarrow_x}\text{”}\rangle_F|\text{“}{\downarrow_x}\text{”},\text{“}{\downarrow_x}\text{”}\rangle_M|{\downarrow_x}\rangle_S).$$

Here *F* reports that her two *x*-spin results agree. So the linear dynamics predicts that it will appear to *F* that she got determinate results for each of her two *x*-spin measurements and that the determinate results agree when she in fact got no determinate results at all.

The upshot of the determinate-result property and the repeatability property together is that if *W* asks *F* whether it appears that a collapse has occurred leaving her object system in an eigenstate of the observable being measured, she will say "yes". So assuming that her experience agrees with her dispositions to report, it will *appear* to the friend that a collapse has occurred and produced a perfectly ordinary, fully determinate, and repeatable result when no collapse has occurred and the friend in fact has no determinate repeatable result.

9.2.3 Intersubjective agreement

The illusions predicted by the bare theory allow for a sort of intersubjective agreement as well. Suppose that *W* now decides to check the *x*-spin of *S* himself then we ask him whether his result agrees with the result that *F* got. Suppose that *W* is reliable when *S* has a determinate *x*-spin and *F* has a determinate measurement record of that *x*-spin. The composite system has the disposition

$$|r,\ r\rangle_W |\text{“}{\uparrow_x}\text{”}\rangle_F |\text{“}{\uparrow_x}\text{”}\rangle_M |{\uparrow_x}\rangle_S$$
$$\downarrow$$
$$|r,\ \text{“}{\uparrow_x}\text{”}\rangle_W |\text{“}{\uparrow_x}\text{”}\rangle_F |\text{“}{\uparrow_x}\text{”}\rangle_M |{\uparrow_x}\rangle_S$$
$$\downarrow$$
$$|\text{“yes”},\ \text{“}{\uparrow_x}\text{”}\rangle_W |\text{“}{\uparrow_x}\text{”}\rangle_F |\text{“}{\uparrow_x}\text{”}\rangle_M |{\uparrow_x}\rangle_S$$

in which W reports that his result agrees with F's result because they both got the determinate result x-spin up when they measured S, and the disposition

$$|r,\ r\rangle_W |\text{“}{\downarrow_x}\text{”}\rangle_F |\text{“}{\downarrow_x}\text{”}\rangle_M |{\downarrow_x}\rangle_S$$
$$\downarrow$$
$$|r,\ \text{“}{\downarrow_x}\text{”}\rangle_W |\text{“}{\downarrow_x}\text{”},\rangle_F |\text{“}{\downarrow_x}\text{”}\rangle_M |{\downarrow_x}\rangle_S$$
$$\downarrow$$
$$|\text{“yes”},\ \text{“}{\downarrow_x}\text{”}\rangle_W |\text{“}{\downarrow_x}\text{”}\rangle_F |\text{“}{\downarrow_x}\text{”}\rangle_M |{\downarrow_x}\rangle_S$$

in which W reports that his result agrees with F's result because they both got the determinate result x-spin down when they measured S.

By the linear dynamics, then, if the initial state of FMS starts in the state $|\text{everett}\rangle_{FMS}$, the composite system evolves as follows:

$$|r,\ r\rangle_W (\alpha |\text{“}{\uparrow_x}\text{”}\rangle_F |\text{“}{\uparrow_x}\text{”}\rangle_M |{\uparrow_x}\rangle_S + \beta |\text{“}{\downarrow_x}\text{”}\rangle_F |\text{“}{\downarrow_x}\text{”}\rangle_M |{\downarrow_x}\rangle_S)$$
$$\downarrow$$
$$\alpha |r,\ \text{“}{\uparrow_x}\text{”}\rangle_W |\text{“}{\uparrow_x}\text{”}\rangle_F |\text{“}{\uparrow_x}\text{”}\rangle_M |{\uparrow_x}\rangle_S + \beta |r,\ \text{“}{\downarrow_x}\text{”}\rangle_W |\text{“}{\downarrow_x}\text{”},\rangle_F |\text{“}{\downarrow_x}\text{”}\rangle_M |{\downarrow_x}\rangle_S$$
$$\downarrow$$
$$\alpha |\text{“yes”},\ \text{“}{\uparrow_x}\text{”}\rangle_W |\text{“}{\uparrow_x}\text{”}\rangle_F |\text{“}{\uparrow_x}\text{”}\rangle_M |{\uparrow_x}\rangle_S + \beta |\text{“yes”},\ \text{“}{\downarrow_x}\text{”}\rangle_W |\text{“}{\downarrow_x}\text{”}\rangle_F |\text{“}{\downarrow_x}\text{”}\rangle_M |{\downarrow_x}\rangle_S.$$

So when we ask W whether he got the same x-spin result for S as F did he will end up in a superposition of saying "yes" and saying "yes" which sounds precisely like "yes."

On the assumption that W believes what he reports, this means that W will believe that he got precisely the same x-spin result as F. They will agree that a collapse has occurred and they will believe that they agree regarding the result of that collapse. But no collapse has occurred and there are in fact no determinate x-spin measurement records.

There is a sense in which it will *appear* to the observers whose behavior is treated within the framework of pure wave mechanics that there are determinate measurement results. But, again, this is subtle. It does not appear to the friend that her result was x-spin up, and neither does it appear that the result was x-spin down; rather, it appears that the result was *either* x-spin up or x-spin down. One might think of this as a *disjunctive experience*. It is an experience devoid of ordinary determinate phenomenal content inasmuch as it is not the determinate experience of x-spin up and it is not the determinate experience of x-spin down.

9.2.4 Relative frequency and randomness

One can also get results concerning the appearance of a random process with the standard quantum probabilities from just the deterministic linear dynamics. These properties, however, only apply in the limit as an observer performs an infinite number of measurements.

Consider an automaton M that is ready to make and record the results of an infinite series of measurements on each of an infinite series of systems S_k in the initial state

$$\alpha|\uparrow_x\rangle_{S_k} + \beta|\downarrow_x\rangle_{S_k} \tag{9.1}$$

(as in Figure 9.1).[6] Suppose that M interacts with each system in turn and perfectly correlates its kth memory register with the x-spin of system S_k by the linear dynamics. This will produce an increasingly complicated entangled superposition of different sequences of measurement outcomes. After one measurement, the state of M and S_1 in the determinate-record basis will be

$$\alpha|\text{“}\!\uparrow\!\text{”}\rangle_M|\uparrow\rangle_{S_1} + \beta|\text{“}\!\downarrow\!\text{”}\rangle_M|\downarrow\rangle_{S_1}. \tag{9.2}$$

After two measurements, the state of M, S_1, and S_2 will be

$$\begin{aligned}\alpha^2|\text{“}\!\uparrow\uparrow\!\text{”}\rangle_M|\uparrow\rangle_{S_1}|\uparrow\rangle_{S_2} &+ \alpha\beta|\text{“}\!\uparrow\downarrow\!\text{”}\rangle_M|\uparrow\rangle_{S_1}|\downarrow\rangle_{S_2}\\ &+ \beta\alpha|\text{“}\!\downarrow\uparrow\!\text{”}\rangle_M|\downarrow\rangle_{S_1}|\uparrow\rangle_{S_2} + \beta^2|\text{”}\!\downarrow\downarrow\!\text{”}\rangle_M|\downarrow\rangle_{S_1}|\downarrow\rangle_{S_2}.\end{aligned} \tag{9.3}$$

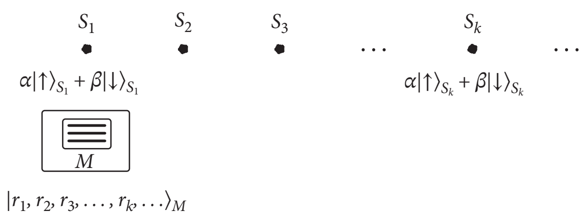


Figure 9.1. Experimental setup for relative frequency and randomness results.

[6] We will simplify notation in the following discussion by suppressing the subscripts on the x-spin properties and both the subscripts and quotation marks on the measurement records.

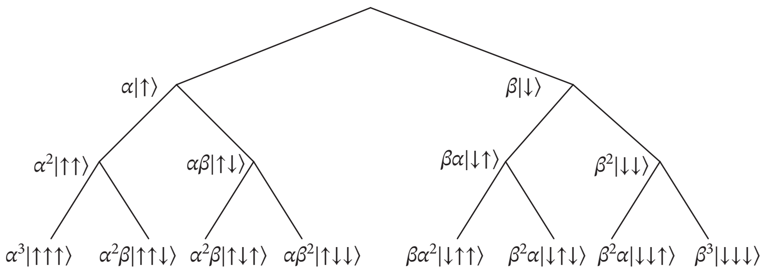


Figure 9.2. Branches after the first three steps.

Figure 9.2 illustrates the first three steps of the branching process and the amplitudes associated with each branch.

Using this setup, Everett (1956, 126–7) briefly sketched an argument for something we might call the *relative frequency* property:

Relative Frequency *For any $\delta > 0$ and $\varepsilon > 0$, there exists a k such that, after k measurements, the sum of the norm-squared of the amplitude associated with each branch where the distribution of spin-up results is within ε of $|\alpha|^2$ and the distribution of spin-down results is within ε of $|\beta|^2$ is within δ of* 1.

And he even more briefly sketched (1956, 127–8) an argument for something we might call the *randomness* property:

Randomness *The sum of the norm-squared of the amplitude associated with each branch where the sequence of relative records satisfies all standard criteria for being random goes to one as the number of measurements gets large.*

It follows from the relative frequency and randomness properties that M will approach a state where it has the sure-fire disposition to report that its measurement results were randomly distributed with the standard quantum probabilities $|\alpha|^2$ and $|\beta|^2$ for x-spin up and x-spin down respectively.[7]

Consequently, treating observers as automata, the bare theory predicts that it will *appear* to an observer that she gets perfectly ordinary, fully determinate, and repeatable measurement results, and in the limit it will appear to her that these results are randomly distributed with the standard quantum probabilities. In other words, there is a sense in which *the linear dynamics alone* predicts that it will appear to an observer, treated within the framework of pure wave mechanics, that there

[7] Both of these results can in fact be proven. See Hartle (1968) for an early discussion of the relative frequency result and Barrett (1999) for a proof of both results in the context of the bare theory.

are random collapses of the quantum-mechanical state precisely as predicted by the standard collapse dynamics (rule 4II).

9.3 Pure Wave Mechanics with just the Standard Interpretation of States

The determinate-result, repeatability, and intersubjective-agreement properties are how the bare theory seeks to address the determinate-result problem. While the observer will typically not in fact have ordinary determinate measurement records, she will *believe* that she does. The relative frequency and randomness properties are how it seeks to address the probability problem. It predicts that it will *appear* to an observer that her measurement records are randomly distributed with the standard quantum probabilities in the limit as she performs an infinite number of measurements when she in fact has no determinate measurement records.

Pure wave mechanics has the virtue of being extremely simple. Further, there is no collapse dynamics to conflict with the linear dynamics or with relativistic constraints. To make pure wave mechanics compatible with relativity, one would just need to put the linear dynamics into a form that is compatible with relativistic constraints, something that is typically taken to provide no special difficulties.[8]

Everett's relative-state formulation, however, suggests that he had a different sort of account of experience in mind from what the bare theory provides. It is not difficult to see why one might want something more.

While one might argue that there is an exotic sense in which the bare theory predicts the same experience as the standard collapse theory, there is also a straightforward sense in which it does not. For its part, the standard theory does not predict that the friend will be under the *illusion* that she got a perfectly ordinary determinate x-spin result. It simply predicts that she will in fact get a perfectly ordinary determinate x-spin result. Further, the standard theory does not predict that an observer will converge to a state where she will be under the *illusion* that her results were randomly distributed with the standard relative frequencies in the limit as she performs an infinite number of measurements. It predicts that she will get a determinate result for every measurement and that over time her actual results will with an increasingly high probability exhibit the standard quantum statistics. These represent significant differences between the empirical predictions of the bare theory and the standard theory.

[8] While there is currently no relativistic field theory that satisfactorily addresses the measurement problem, the difficulty here has more to do with explaining determinate records and less to do with providing a relativistic formulation of the linear dynamics. That said, there is a tension between even the linear dynamics and relativistic constraints. See Albert (2013) for one discussion. The problem is particularly clear if one believes that there is a simple matter of fact concerning whether two systems are entangled. See Barrett (2014) for an extended discussion of this point.

A central issue in assessing alternative formulations of quantum mechanics concerns *what it should mean* for a physical theory to be empirically adequate. The present distinction is between explaining why one in fact got a particular measurement result and explaining why one believes that one got *some* determinate result when one in fact did not get any particular result at all. If one takes a theory to be empirically adequate only if it explains the empirical evidence that we take ourselves to have pre-theoretically, then the bare theory is not empirically adequate. Rather than explain what we take ourselves to see in the laboratory, it explains why we don't in fact see any such thing but nevertheless believe we do.

But this is not to say that we have good empirical evidence for rejecting the bare theory. For all we know it might describe our world. But if it does, then we are under the illusion that we have definite measurement outcomes when we don't, and that poses a serious methodological problem. If the bare theory is true, then the physical world is such that one can never have reliable empirical evidence for accepting the bare theory. We will say that such a theory is *empirically incoherent*.[9] Since the bare theory is empirically incoherent, if it is true, then our world violates the preconditions for the possibility of empirical inquiry. It is a remarkable feature of the quantum measurement problem that it has led us to take seriously empirically incoherent theories.

9.4 The Relative-State Formulation of Pure Wave Mechanics

While Everett explicitly appealed to the bare theory's formal properties, his strategy for explaining experience was more straightforward. Rather than argue that one is under the illusion that one has determinate results when one in fact doesn't, he argued that pure wave mechanics could be understood as representing the actual statistical content of an observer's experience. More specifically, he argued that his *relative-state* formulation of pure wave mechanics was both *consistent* and *empirically faithful*.[10]

Everett believed that pure wave mechanics immediately solved the quantum measurement problem by getting rid of the potential conflict between the linear dynamics and the collapse dynamics. He took the key to explaining determinate records and probabilities in pure wave mechanics to be the distinction between *absolute* and *relative states*. This distinction was based on the *the principle of the fundamental relativity of states*. He explained the principle as follows:

[9] See Barrett (1996) and (1999) for discussions of the bare theory and the notion of empirical incoherence.

[10] See Barrett (2011a) and (2015) for discussions of what Everett meant in requiring that a satisfactory formulation of quantum mechanics be empirically faithful and the relationship between this notion and empirical adequacy.

> There does not, in general, exist anything like a single state for one subsystem of a composite system. Subsystems do not possess states that are independent of the states of the remainder of the system, so that the subsystem states are generally *correlated* with one another. One can arbitrarily choose a state for one subsystem, and be led to the relative state for the remainder. Thus we are faced with a fundamental *relativity of states*, which is implied by the formalism of composite systems. It is meaningless to ask the absolute state of a subsystem—one can only ask the state relative to a given state of the remainder of the subsystem.[11]

Relative states might be thought of as introducing a *branch indexical* to the theory. Just as the state of a physical system may differ *at different times*, the state of a system here may differ *at different branches*. It is in the relative branch states that one finds an observer's determinate (but relative) measurement records. We will first reflect on how the branch indexical works, then how Everett thought of it.

Consider a physical system S that is composed of two subsystems A and B. The *absolute state* of S is just its ordinary quantum-mechanical state, and, as usual, it is represented by a vector $|\psi\rangle_S$ of unit length in a Hilbert space $\mathcal{H}$. If the absolute state of S can be written in the form

$$|\psi\rangle_S = \sum a_i |\psi_i\rangle_A |\chi_i\rangle_B$$

then $|\psi_k\rangle_A$ is the state of subsystem A *relative* to the state of B being $|\chi_k\rangle_B$. This allows one to also distinguish between absolute and relative properties. The system S has an *absolute property* if and only if it is in an eigenstate of having that property. If $|\psi_k\rangle_A$ is the state of subsystem A *relative* to the state of B being $|\chi_k\rangle_B$ and if $|\psi_k\rangle_A$ is an eigenstate of property P and $|\chi_k\rangle_B$ is an eigenstate of property Q, then subsystem A has property P *relative* to subsystem B having property Q. Everett called the terms in the superposition $\sum a_i |\psi_i\rangle_A |\chi_i\rangle_B$ *elements* or *branches* of the state of S.

After she has performed her measurement of system S, the state of Wigner's friend F can be written as

$$\alpha |\text{“}\!\uparrow_x\!\text{”}\rangle_F |\text{“}\!\uparrow_x\!\text{”}\rangle_M |\uparrow_x\rangle_S + \beta |\text{“}\!\downarrow_x\!\text{”}\rangle_F |\text{“}\!\downarrow_x\!\text{”}\rangle_M |\downarrow_x\rangle_S.$$

While F does not have an absolute result to her measurement in this state, she got the result x-spin up *relative* to M's pointer indicating x-spin up and S being x-spin up and she got the result x-spin down *relative* to M's pointer indicating x-spin down and S being x-spin down. When written in the determinate-record basis, then, the

[11] Everett presents this as the central interpretational principle with just slightly different words and italics in both the long and short versions of his thesis (1956, 103) and (1957, 180). The quotation here follows the latter.

state of the composite system here has two *branches*, one where F got x-spin up and one where she got x-spin down.

Everett took the relative records associated with each branch to explain the friend's experience. As he put it in conversation with Abner Shimony some years later, "Each individual branch looks like a perfectly respectable world where definite things have happened" (Barrett and Byrne 2012, 276).[12] One can get a sense of what Everett had in mind when he said this by considering what he said about what we called the bare theory's suggestive properties.

When Everett is describing the determinate result property for a state like $|\text{everett}\rangle_{FMS}$, he says

> [I]n each *element* of the superposition . . . , the object-system state is a particular eigenstate of the observer, and *furthermore the observer-system state describes the observer as definitely perceiving that particular system state*. It is this correlation that allows one to maintain that a measurement has been performed.
>
> (1956, 121)

So the friend got a determinate measurement result because each relative friend state describes her as having a determinate result. If F repeats her x-spin measurement, she ends up in a state like

$$\alpha|``\uparrow_x", ``\uparrow_x"\rangle_F|``\uparrow_x", ``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + \beta|``\downarrow_x", ``\downarrow_x"\rangle_F|``\downarrow_x", ``\downarrow_x"\rangle_M|\downarrow_x\rangle_S.$$

Regarding the repeatability property, Everett says:

> Again, we see that each element [in the resulting superposition] describes a system eigenstate, but this time also describes the observer as having obtained the *same result* for each of the two observations. Thus for each separate state of the observer in the final superposition, the result of the observation was repeatable, even though different for different states. This repeatability is, of course, a consequence of the fact that after the observation the *relative* system state for a particular observer state is the corresponding eigenstate. (1956, 122)

So the friend gets the same result for each of the two measurements because she gets the same result within each relative state.

The difference between the relative-state formulation and the bare theory is that here the friend actually gets perfectly determinate *relative results*. The determinate experience of getting the result x-spin up is explained by the relative state where she records x-spin up and the determinate experience of getting the result x-spin

[12] This quotation is the only place I know where Everett himself directly associates branches with worlds.

down is explained by the relative state where she records x-spin down. Further, the friend repeats the result x-spin up relative to S being x-spin up and she repeats the result x-spin down relative to S being x-spin down.

So while the explanation of determinate experience here is different, the suggestive properties of the bare theory are still relevant. One way to think of this is that by telling us what absolute properties obtain for a composite system, they tell us what a *typical relative observer* will experience. That there will appear to be a collapse is an *absolute* property of the observer. As we have seen, the composite system is simply in an eigenstate of the observer believing that she has a determinate result. But, in the present case, this means that every *relative* observer will believe she got a perfectly determinate measurement result. Of course, *what* result a particular relative observer believes she got is a relative fact. That every relative observer believes that she got a determinate result, then, is explained by the fact that each relative observer in fact has a perfectly determinate *relative* record. This relative result characterizes the content of that relative observer's experience.

The situation is only slightly more subtle in the case of the limiting suggestive properties. Regarding the state that one gets in the limit as an observer measures an infinite sequence of systems, Everett explains that

> [I]f we consider the sequences to become longer and longer (more and more observations performed) *each* memory sequence of the final superposition will satisfy any given criterion for a randomly generated sequence, generated by the independent probabilities [$|\alpha|^2$ and $|\beta|^2$], except for a set of total measure which tends toward zero as the number of observations becomes unlimited. Hence all averages of functions over *any* memory sequence, including the special case of frequencies, can be computed from the probabilities [$|\alpha|^2$ and $|\beta|^2$], except for a set of memory sequences of measure zero. We have therefore shown that the statistical assertions of [the collapse process] will appear to be valid to *almost all* observers described by separate elements of the superposition . . . in the limit as the number of observations goes to infinity. (1956, 127)

The measure that Everett has in mind in referring to *almost all* relative observers is the norm-squared of the coefficient associated with each term in the superposition of memory sequences. He stipulated this as a measure of typicality over branches of the quantum-mechanical state. While it is not generally true that most branches *by count* will exhibit the standard quantum statistics, it is true that the sum of the norm-squared of the coefficients on those terms that do exhibit the standard quantum statistics will go to one as the number of measurements gets large. Hence, a typical relative observer *in the norm-squared coefficient sense of typical* will record measurement results that exhibit the standard quantum statistics. And in the limit *almost all* relative observers, in the norm-squared amplitude measure, will exhibit the standard quantum statistics. This was the single fact that Everett thought was most significant in understanding his theory.

Everett argued that a similar result holds for measurements that are less than perfect. And from this he concluded that a typical relative observer will have the subjective experience of random collapses with the standard quantum probabilities. This was his promised deduction of the predictions of the standard collapse theory in terms of the subjective experience of observers treated in the framework of pure wave mechanics.

While pure wave mechanics is a deterministic theory that says nothing whatsoever about probability, Everett wanted to provide *some sense* in which it might be understood as predicting the standard quantum probabilities. His argument amounts to this: inasmuch as the relative measurement records of a *typical* relative observer will exhibit the standard quantum statistics, it will *appear to a typical relative observer* that the standard quantum probabilities obtain.

Note that Everett did not seek to deduce quantum probabilities *over branches*. The probability associated with each branch is always just one. Rather, he wanted to derive the appearance of the standard quantum probabilities from the statistics of records *within a typical branch*. Within a *typical branch*, in his norm-squared amplitude sense of *typical*, it will appear that the measurement outcomes were determined by the standard quantum probabilities. This is the significance of the relative frequency and randomness properties of the bare theory for Everett.[13]

9.5 Everett's Empiricism

Instead of trying to explain precisely what a relative observer *is* metaphysically, Everett adopted a throughly empiricist attitude toward the theory. As he put the point: "There can be no question of which theory is 'true' or 'real'—the best that one can do is reject those theories which are *not* isomorphic to sense experience" (Barrett and Byrne 2012, 253). He just wanted a theory that was consistent and empirically faithful. He explained what he meant by this in a letter to the physicist Bryce DeWitt, who would later become a strong supporter of pure wave mechanics:

> First, I must say a few words to clarify my conception of the nature and purpose of physical theories in general. To me, any physical theory is a logical construct (model), consisting of symbols and rules for *their* manipulation, *some* of whose elements are associated with elements of the perceived world. If this association

[13] This differs from the tradition of Everettian attempts to derive probabilistic expectations from pure wave mechanics. See Deutsch (1999), Saunders (2010), and Wallace (2010b) for recent examples of this tradition.

> is an isomorphism (or at least a homomorphism) we can speak of the theory as correct, or as *faithful*. The fundamental requirements of any theory are logical consistency and correctness in this sense.[14]

In the long version of his PhD thesis, he further explained that "[t]he word *homomorphism* would be technically more correct, since there may not be a one-one correspondence between the model and the external world" (1956, 169).

So let's consider the two conditions. The relative-state formulation of pure wave mechanics is clearly consistent. It is just the bare theory with an added distinction between absolute and relative states and a specified typicality measure over branches. One might think of it as the standard theory without rule 4II but with a stronger version of rule 3 that includes a description of what relative states are and the specification of what it means for a relative state to be typical. Since there is no collapse, there is nothing that might be inconsistent in the theory.

For Everett, a theory was empirically faithful, and hence adequately explained experience, if and only if there was a homomorphism between its model and the world as experienced. The relative-state formulation is empirically faithful because an observer's actual experience can be *found* as a relative sequence of records associated with a *typical* relative observer. While most relative sequences will not agree with a given observer's actual experience, Everett took every element of the superposition to be equally real in the operational sense that one could always, at least in principle, perform an *A*-type interference measurement that would detect the presence of other branches, branches with relative records that one would not take to characterize one's actual measurement results.

Importantly, Everett did not derive the standard *probabilistic predictions* of quantum mechanics regarding chance events, and he knew it. He clearly and repeatedly insisted that there were no probabilities in pure wave mechanics. This is also reflected in the early title of his thesis as *Wave Mechanics without Probability*. Rather than claim that his theory predicted the same chance events as predicted by the nonlinear collapse dynamics, Everett showed that it would *appear to a typical relative observer* that there had been such events. It was to this end that he argued that an observer's relative measurement records in a typical branch would appear to be randomly distributed with the standard quantum probabilities. His account appealed to notions that one simply does not have in pure wave mechanics or the bare theory. Everett added a model of observers as automata to tie experience to physical records. He added the notion of relative states to characterize relative sequences of records. And he added a typicality measure to say *which* relative sequences of records matter for experience.

[14] This passage is from Everett's letter to DeWitt dated 31 May 1957. See Barrett and Byrne (2012, 252–6) for the full text of the letter and commentary.

While the typicality measure that Everett chose to add to pure wave mechanics has a number of natural properties, it also has an ad hoc flavor to it. The norm-squared amplitude measure over branches is just the ordinary quantum probably measure over outcomes. And since the standard theory predicts that one's records will probably satisfy the standard quantum statistics, it is unsurprising that a typical observer's records *in precisely the same measure* will satisfy the standard quantum statistics. That said, Everett can use whatever typicality measure he wants in his claim that the relative records in a typical branch will exhibit the standard quantum statistics as long as he explicitly specifies it as a part of his relative-state formulation.

The auxiliary assumptions Everett added to pure wave mechanics to get the relative-state formulation pay off in providing an account of an observer's measurement results in terms of the *relative* records of a *typical* observer. And since the dynamics are still purely linear, it provides an appropriate starting point for formulating a relativistic theory. The interpretational apparatus is a bit more involved, but it remains a very simple theory.

Some have worried that the relative-state formulation requires one to choose a preferred physical basis.[15] Each branch describes a relative observer with determinate relative records but only if one writes the state of the composite system in a special determinate-record basis. That said, inasmuch as Everett just wanted a theory that satisfies his condition for being empirically faithful, he arguably does not need a *physically* preferred basis. For the theory to be empirically faithful, one just needs to be able to *find* an observer's actual experience associated with a typical sequence of relative records. One finds it in a branch of the state of the composite system written in the determinate-record basis, but that need not require that there be anything *physically* privileged about that basis. This squares with both Everett's view that there is no privileged way to individuate branches and his operationalist conception of his theory.[16]

The most significant problem with the relative-state formulation is that empirical faithfulness is a very weak standard of empirical adequacy. Its weakness can been seen in the fact that while it is empirically faithful, it does not predict that an observer will *probably* record the standard quantum statistics or should *expect* to record the standard quantum statistics. To be sure, the theory predicts that a typical branch, in the norm-squared amplitude sense of typical, will exhibit the standard quantum statistics, but it does not predict that *an observer's future branch* will be typical or probably typical or should be expected to be typical, or anything like that.

[15] See Barrett (1999) for a discussion of this worry.

[16] See Everett's principle of the relatively of states. See also his marginal notes in reply to John Bell for his view that there is no physically preferred basis (Barrett and Byrne 2012, 287).

The relative-state formulation tells us that the sequence of records in a typical branch will appear to be determined by the standard quantum probabilities, but it does not tell us that a typical branch is *probable* right now or that one should *expect* one's future records to be determined by the relative sequence of a typical branch. One could add an assumption like that to the theory, something we will discuss in the next chapter, but Everett did not do so. A move like that would require *probabilities over branches* and *something for those probabilities to describe*.

Pure wave mechanics does not make probabilistic predictions even when supplemented with relative states. Inasmuch as a probability is a measure over possibilities where precisely one is realized and insofar as all possibilities are realized in pure wave mechanics, Everett is right to insist that there can be no probabilities associated with alternative relative sequences of measurement records. Similarly, any understanding of typicality that somehow involves any sort of *selection* of a typical relative sequence of records is incompatible with pure wave mechanics inasmuch as the theory describes no such thing. Neither can the typicality measure represent an expectation of the standard quantum statistics obtaining for one's actual relative sequence of measurement records since all such sequences are equally actual in the model. Insofar as the theory describes anything happening, it describes *everything* happening, so there is no particular sequence of measurement records that might be taken as *probable* or *expected* in any standard sense of the terms. We will have more to say about this in the next chapter.

For the purpose of understanding Everett's own views, it is important to be clear that he does not solve the probability problem by showing that measurement records that satisfy the standard quantum statistics are probable or to be expected. Rather, he specifies a measure, then shows that in that weighting over branches almost all of the weight ends up associated with branches where the relative observer has a sequence of records that satisfy the standard quantum statistics. That is what it means when he says that the theory is *empirically faithful*. Since neither pure wave mechanics nor the relative-state formulation assigns probabilities *over* branches, they do not make probabilistic predictions for future events and, hence, do not make the same empirical predictions as the standard collapse theory.

10
Many Worlds and Such

10.1 Extending Pure Wave Mechanics

While pure wave mechanics does not explain determinate measurement records or quantum probabilities on its own, its simplicity and manifest consistency make it a good starting point for constructing a formulation of quantum mechanics that avoids the measurement problem. If one adds the standard interpretation of states, as in the bare theory, one can explain why an observer would *believe* that she has determinate measurement records in special situations when she in fact has no such records. If one further adds the distinction between absolute and relative states and specifies a typicality measure over branches, one can explain why a *typical relative observer* will have relative records that exhibit the standard quantum statistics. But such accounts of experience are more exotic than compelling.

One would like to have a theory that simply predicts the determinate measurement records we in fact take ourselves to have with the standard forward-looking quantum probabilities. If so, we need stronger auxiliary assumptions than Everett made. But we also want to keep the theory as simple and plausible as possible. Ideally, one would want to add the most modest auxiliary assumptions one can to pure wave mechanics that deliver an account of experience that one can in good conscience accept.

Since the post-measurement states predicted by pure wave mechanics typically do nothing whatsoever to explain the presence of determinate measurement records, one must add something that says the sense in which the physical state represents determinate measurement records. Similarly, since pure wave mechanics says nothing at all about probabilities, if one wants one's theory to make probabilistic predictions, then one must add something that allows one to derive probabilities of the sort one requires. And insofar as one takes one's *physical theory* to explain determinate measurement records and quantum probabilities, these auxiliary assumptions should be specified along with the rest of the theory.

There has been significant disagreement regarding how to best understand Everett's formulation of quantum mechanics. This disagreement has been driven at least in part by the desire for an account of experience that is both richer and more straightforward than those provided by the bare theory or the relative-state formulation.

Since the various Everettian formulations of quantum mechanics are often only loosely related to Everett's own views, we will focus here on just evaluating each proposed formulation on its own merits. Each supplements pure wave mechanics with different auxiliary assumptions which, in turn, suggest different explanations of determinate measurement records and quantum probabilities. The auxiliary assumptions often involve metaphysical commitments. These commitments typically do real explanatory work in accounting for determinate records and the statistics they exhibit.

We will start with a few examples of many-worlds interpretations. We will consider some of the problems they encounter and how they might be addressed. We will then discuss a number of closely related theories.

10.2 Splitting-Worlds

While Everett himself did not present his theory as a many-worlds interpretation, this is how the physicist Bryce DeWitt (1971) and his student Neill Graham (1973) understood it. We will call their approach the splitting-worlds theory.

The splitting-worlds theory postulates a quantum-mechanical state for the whole universe that always evolves linearly without collapsing. As the state evolves, the universe splits "into a multitude of mutually unobservable but equally real worlds, in each of which every good measurement has yielded a definite result and in most of which the familiar statistical quantum laws hold" (DeWitt and Graham 1973, v).

The idea is that there is a separate, but equally real, physical world corresponding to each branch of the quantum-mechanical state. When a measurement is made, an observer's world splits and each of the resultant worlds contains a different physical copy of the observer. Each copy of the observer has a perfectly determinate measurement record, but different copies typically have different records.[1]

There are a few things to note to get started. First, since this splitting process is supposed to create an objectively real world for each branch, it now matters to the *physics* how one decomposes the state into branches. Hence the theory requires that one choose a physically preferred basis or some rule for specifying such a basis. Second, the insistence that worlds are mutually unobservable differs from Everett's own view as he explicitly held that it was always possible in principle to observe quantum interference between branches. Finally, if one associates each branch with one world, it is typically not true that most branches *by count* will

[1] In his popular account of the splitting-worlds theory, DeWitt (1970) described it as being held by Everett, Wheeler, and Graham. There was, however, never a single theory that Everett, Wheeler, and Graham would all have endorsed. For his part, as we will see, Everett strongly disagreed with both DeWitt and Graham.

exhibit the standard quantum statistics. Rather, as we saw in the last chapter, it is only most branches *in the norm-squared amplitude measure* that will exhibit the standard quantum statistics. We will return to this last point in a moment.[2]

DeWitt described his reaction when he first considered Everett's theory: "I still recall vividly the shock I experienced on first encountering this multiworld concept. The idea of 10^{100+} slightly imperfect copies of oneself all constantly splitting into further copies, which ultimately become unrecognizable, is not easy to reconcile with common sense" (DeWitt and Graham 1973, 161). But he felt better about this after Everett sent him a letter explaining how the theory itself explains why he would not feel any splitting process. Everett's argument might be summarized as follows. To detect such a split, one would need something that could be thought of as a record of the splitting process or the existence of alternative worlds produced by the process. Such records are not produced in typical measurement interactions. For his part, Everett knew that one could only record the production of branches corresponding to different measurement outcomes if one performed something like an A-type measurement of the observer whose measurement interaction caused the split. But, as he noted in the long version of his thesis, while not impossible, that would be extremely difficult.

The splitting-worlds theory might be understood as the standard formulation of quantum mechanics without the collapse dynamics (rule 4II) but with a new interpretational rule for quantum-mechanical states. The idea is to replace the standard interpretation of states (rule 3) with something like the following:

3. Interpretation of states (splitting worlds): There is a basis $\mathcal{B}$ such that when the state of the universe $|\psi\rangle_U$ is written in $\mathcal{B}$ each term in the expansion describes an actual physical world in the quantum state characterized by that term and all measurement records are in fact determinate in every world. The standard interpretation of quantum states then applies to the state of each world.

One wants to choose the basis $\mathcal{B}$ to be one that would make measurement records determinate in every world. When a measurement is made, then the universe splits, and there is a different, but perfectly determinate, measurement outcome in each of the resulting worlds.

Consider a universe consisting of just Wigner's friend and her measuring device and object system. Choose a basis $\mathcal{B}$ that makes determinate whatever physical quantity the measuring device M uses (as a pointer variable) to indicate the result of its x-spin measurement. Suppose that the state of the composite system FMS is initially

[2] As we will see, DeWitt and Graham ultimately associated many worlds with each Everett branch. For the basic splitting-worlds theory described here, we will suppose that there is just one world for each branch.

$$|``r"\rangle_F|``r"\rangle_M(\alpha|\uparrow_x\rangle_S + \beta|\downarrow_x\rangle_S).$$

Written in the determinate-record basis, this state has just one term, so there is just one world. In this world the friend F is ready to look at the pointer on the measuring device, the measuring device M is ready to measure the x-spin of S, and S is in a superposition of x-spin up and x-spin down.

When M measures the x-spin of S, the composite system evolves by the linear dynamics to

$$\alpha|``r"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + \beta|``r"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S.$$

Here there are two terms, so there are now two worlds. The interaction between M and S caused the split. In one world the friend F is ready to observe M's pointer, M's pointer indicates the result x-spin up, and S is determinately x-spin up. In the other the friend F is ready to observe M's pointer, M's pointer indicates the result x-spin down, and S is determinately x-spin down (as in Figure 10.1).

When F looks at M's pointer, the composite system evolves by the linear dynamics to

$$\alpha|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + \beta|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S.$$

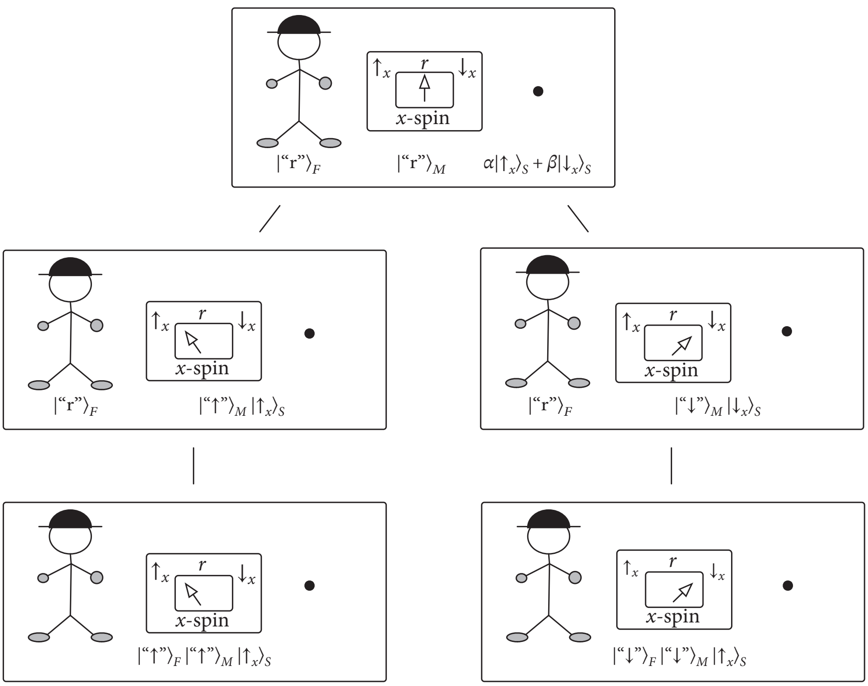


Figure 10.1. Worlds resulting from a spin measurement.

Here there are still two worlds. In one the friend sees the pointer indicating the result x-spin up and the object system is x-spin up, and in the other the friend sees the pointer indicating the result x-spin down and the object system is x-spin down.

Since there is nothing in the theory so far that tells one how to identify worlds across time, one might add an auxiliary assumption that says that these are the *same* two worlds we had after *M*'s interaction with *S* as indicated by the lines connecting the frames in Figure 10.1. In that case, one might think of the friend as getting the x-spin result she does in each world *because* the pointer is already indicating a determinate result in that world. On this view, the measuring device indicates a perfectly determinate measurement record in each world and the friend just checks to see what it is.

The explanation of why an observer gets a determinate measurement outcome, then, is entirely straightforward. Each observer gets a determinate measurement outcome for the simple reason that each has a fully determinate measurement record in each world she inhabits. But this explanation comes at a cost.

The preferred basis $\mathcal{B}$ determines when worlds split. In the Wigner's friend story as just told, we chose the determinate physical quantity to be one that makes *M*'s *pointer reading* determinate in each world. On that choice, the world splits when *M*'s pointer position becomes correlated with *S*'s x-spin. But if we chose, instead, a preferred basis that makes *F*'s *brain record* determinate directly, the world would not split until *F*'s brain state becomes correlated with the reading of *M*'s pointer. In this case, the splitting-worlds story would look like the sequence of events in Figure 10.2. This is a very different sequence of physical events from the case where one uses *M*'s pointer reading to determine what worlds there are.

One wants worlds to split whenever one needs them to split to explain determinate measurement records. In other words, one wants worlds to split in the splitting-worlds theory precisely when one would want a collapse to occur on the standard collapse theory. Indeed, the splitting worlds theory is very much like a collapse theory with a different collapsed world for each possible measurement outcome. The upshot is that choosing the preferred basis $\mathcal{B}$ here is exactly as difficult as providing a satisfactory specification of when collapses occur in the standard theory. But if we knew how to do that, we would be able to solve the quantum measurement problem in the context of the standard collapse theory.

Because world-splitting events are essentially collapse events that produce multiple worlds, the splitting-worlds theory is also incompatible with relativistic constraints. In place of having to say precisely when an object system collapses in a frame-invariant way, one now has to say precisely when worlds split. If anything, the situation is *worse* here as one also has the issue of what it could even *mean* for worlds to split in a relativistic theory. If spacetime itself is supposed to split, one must presumably specify another super-spacetime in which to tell *that* dynamical story.

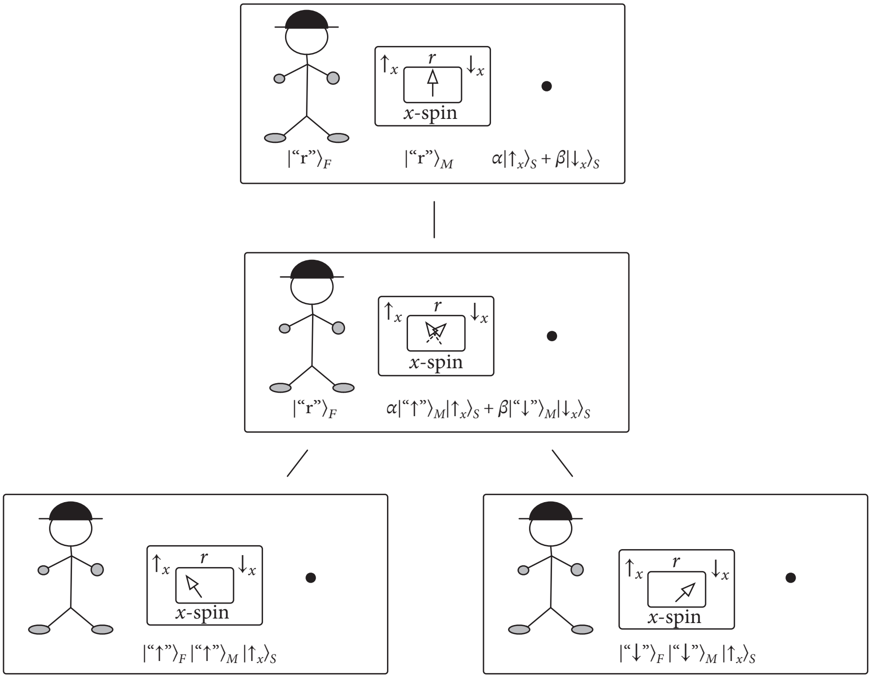


Figure 10.2. Worlds if splits occur later in the measurement process.

Another problem with the splitting-worlds theory concerns probability. While the right choice of preferred basis will explain determinate measurement records, the theory does not predict the right quantum probabilities for an observer's future measurement records. Insofar as each copy of the friend has an equal claim to being the *future* friend, the probability that F gets x-spin up is simply one. And so is the probability that she gets x-spin down. But that's the wrong answer.

One can distinguish here between two closely related problems involving quantum statistics. The first concerns how to get the *probabilistic prediction* that F will find herself in the world where she records x-spin up with probability $|\alpha|^2$ and will find herself in the world where she records x-spin down with probability $|\beta|^2$. This problem involves *forward-looking* probabilities. Part of this problem is that F *will find herself in each world*. The second problem concerns why one should expect the records in one's world *at a particular time* to exhibit the standard quantum statistics. This problem involves accounting for one's current records. Which problem one finds more pressing depends on the sort of explanation of quantum statistics one wants. Since the second problem is easier, let's start with it.

Everett showed that the relative records in a typical branch, in his norm-squared amplitude measure sense of typical, will exhibit the standard quantum statistics. In the splitting-worlds theory this means that the physical records in a typical world *in this measure* will exhibit the standard quantum statistics. So to explain why one should expect one's current records to exhibit the standard quantum statistics one just needs to explain why one should expect one's world to be typical in this sense. Put this way, it might seem like an easy task, so easy that it is hardly worth mentioning that there is a task at all. But it requires care.

That something is typical does not mean that it is probable. Pennies are typically made of zinc. But that does not logically entail that the penny sitting on the right-hand side of my desk is probably made of zinc. To infer *that* one would need to know something like that the penny was randomly selected in an unbiased measure from the set of all pennies. If so, this is a highly nontrivial property of the penny. Similarly, if one wanted to explain why one should *expect* the standard quantum statistics to hold for the relative records in one's world, then one would need to know something like that one's world was randomly selected in the norm-squared-amplitude measure over branches. Even to say that and have it make sense, one would need a way to determine what worlds there are and would need to be able clearly say what it could possibly mean for a world to be randomly selected as the one that is actually experienced.

The splitting-worlds formulation, like pure wave mechanics, says nothing about quantum probabilities. Consequently, one would need to add further auxiliary assumptions to the splitting-worlds theory to get anything like the standard quantum probabilistic predictions or the expectation of the standard quantum statistics. There is nothing wrong with doing so, but one only has a clear account if one understands where the probabilities come from and what they are probabilities of.

10.3 Probability and Typical Worlds

DeWitt initially described the splitting-worlds formulation of quantum mechanics as we have characterized the theory, but this is not the version that DeWitt and Graham ultimately favored. A short digression on their favored formulation of the theory will help to clarify what one would have to add to pure wave mechanics to get probabilistic predictions.[3] Again, we will distinguish between forward-looking probabilities and synchronic probabilities that explain our current records.

DeWitt and Graham did not believe that Everett had provided a satisfactory account of quantum probabilities. They thought that the problem was with his typicality measure over branches. Putting it in terms of worlds, Everett had shown that a typical world *in the norm-squared amplitude measure sense of typical* would exhibit the standard quantum statistics. While that is true, what bothered DeWitt and Graham was that it is usually not the case that *most worlds by simple count* will

[3] See Barrett (2017) for a more detailed discussion of typicality and probability in many-worlds formulations of quantum mechanics.

exhibit the standard quantum statistics. As Graham put it, "it is extremely difficult to see what significance [Everett's] measure can have when its implications are completely contradicted by a simple count of the worlds involved" (1973, 236).

What DeWitt and Graham wanted was a theory where *most worlds* exhibited the standard quantum statistics. Only then did they believe that one would be able to provide a satisfactory explanation for why one should expect one's measurement records to exhibit the standard quantum statistics. To this end, Graham stipulated a rule for how worlds split such that the *number of worlds* that exhibit a particular outcome after a measurement interaction was proportional to the square of the coefficient on the term describing that outcome. This ad hoc splitting procedure ensured that almost all worlds by simple count will exhibit the standard quantum statistics.[4]

Observers have determinate measurement outcomes on this version of the splitting-worlds theory because measurement records are determinate in each world—there are just a lot more worlds here than in the basic splitting-worlds theory. But having more worlds does not provide any advantage in addressing the probability problem. Seeing why will help to clarify the relation between typicality measures and probabilities.[5]

To begin, we know what Everett thought of DeWitt's criticism and Graham's proposal from his marginal notes on his personal copy of the DeWitt–Graham (1973) anthology. Where DeWitt (1971, 185) claimed that Everett's typicality argument was unsatisfying, Everett wrote in the margin "only to you!" And where Graham (1973, 236) claimed that Everett's typicality measure was unmotivated, Everett wrote in the margin "bullshit."[6] Everett knew that he could stipulate any measure of typicality he wanted over branches, then use that measure to make claims about the properties of typical branches. Since the norm-squared amplitude measure over branches is just the standard quantum probably measure over outcomes, his choice has a special feel to it. But that's fine as long as one explicitly acknowledges the particular choice of measure in one's theory.

That DeWitt and Graham believed that making something true of most branches *by simple count* was the only satisfactory way to explain why an observer should expect his results to exhibit the standard quantum statistics tells us something about the type of explanation they were looking for and the auxiliary

[4] Graham (1973) presented his "two step" splitting procedure as an extension of what he called Everett's "one step" account of measurement. On Graham's account, one introduces a third macroscopic apparatus that mediates between a microscopic system and a macroscopic observer, then that apparatus comes to thermodynamical equilibrium with its environment. The partition of branches that Graham subsequently stipulated for the resulting state differs from the determinate-record partition that Everett derived by assuming that the linear dynamics fully described the measurement interaction.

[5] See Goldstein (2012) for one approach to the relationship between typicality and probability in physics. See Hemmo and Shenker (2015) for a discussion of the relationship between typicality and probability in Bohmian mechanics. See Kent (2010) for additional worries concerning recent many-worlds formulations of quantum mechanics.

[6] See Barrett and Byrne (2012, 365–6) for photocopies of these pages.

assumptions they were willing to make to get it. To begin, DeWitt and Graham took the talk of physically real worlds somehow inhabited by copies of observers seriously, arguably much more seriously than Everett himself did inasmuch as he did not present his theory in terms of splitting worlds.[7] Second, they seem to have wanted standard forward-looking probabilities, something that Everett does not get with his conclusions regarding the statistical properties of typical sequences of relative measurement records at a specified time. Finally, their intuitions were apparently guided by a *principle of indifference* that says something like: if there are n possibilities, then, in the absence of other information, one should assign probability $1/n$ to each possibility.[8]

Given this, an account of quantum probabilities might go something like the following. Most quantum worlds *by simple count* exhibit the standard quantum statistics on the DeWitt-Graham specification for how to partition the universal state into worlds. Since, by the principle of indifference, it is equally likely that one would find oneself in each such world, one should with high probability expect to find that one's actual measurement results exhibit the standard quantum statistics.

But there are several problems with this. First, it is unclear precisely what such quantum probabilities are probabilities of. Since there are no chance events nor uncertainties stipulated by pure wave mechanics itself, one needs to add something to the theory to which probabilities might refer. Even if one adds a preferred partition of worlds, inasmuch as there is a copy of the observer in each of the post-measurement worlds, one would need to explain why the probability of finding the observer in any particular world is something other than one.

As suggested by the language we have been using, one might try to make sense of the probabilities in a many-worlds theory as *self-location* probabilities. The idea is to add auxiliary assumptions to the theory that allow one to show that one should with high probability expect *to find oneself* in a world where one's measurement results exhibit the standard quantum statistics, whatever that might mean.[9]

The extent to which this strategy works depends on the details of one's metaphysical commitments and the sort of explanation one wants. If I believe that I will in fact inhabit all future worlds, then my forward-looking probability of each possible *future* measurement outcome is simply one. But if there were a law that said that my synchronic self-location probabilities were equal to the norm-squared amplitude measure associated with each world, I might explain why I expect my *current* records to exhibit the standard quantum statistics.

If one wanted to make sense of the standard *forward-looking* quantum probabilities in a splitting-worlds theory, one would need to add something to the metaphysics that would allow one to talk sensibly about precisely one future

[7] See Barrett (2011b) for a discussion of Everett's metaphysical views.

[8] There are a number of discussions of the difficulties one encounters in applying the principle of indifference to branches. See, for examples, Barrett (1999, 168–73) and Greaves (2007).

[9] Self-location probabilities are a recurring topic in Saunders, Barrett, Kent, and Wallace (eds) (2010). See also Vaidman (2012).

world being somehow realized for an observer at the expense of the others. Then, depending on how one sets it up, such probabilities might be understood in terms of chances or subjective degrees of belief concerning a perfectly ordinary matter of fact—namely, the world one will in fact inhabit. To do *this*, one must specify both how to calculate forward-looking probabilities and, importantly, precisely what they are probabilities *of*.

With care, there are a number of ways that one can get the standard forward-looking quantum probabilities for a many-worlds theory. On the many-threads theory, something we will discuss in detail in section 10.6, worlds do not split.[10] Rather, there is one world for every possible complete world history, or thread, and a probability measure over these possible complete worlds. One way to picture this set of worlds is by associating a thread to each possible way one might travel through the complete branching structure described by the splitting-worlds theory. The theory then assigns the standard quantum probabilities to each such world history. From the perspective of a particular observer, these are the probabilities that each is the world she in fact inhabits. This provides epistemic probabilities at a time that she can use to calculate forward-looking probabilities as she conditions on observed properties of the world she inhabits. Since neither she nor her world splits, she can make perfectly clear sense of forward-looking self-location quantum probabilities as epistemic probabilities regarding future events in the world she in fact inhabits.

In contrast, on a many-worlds formulation where worlds split and there is no way to identify an observer with any single future self, an observer must assign a probability of one to each of her future selves if she assigns forward-looking probabilities at all. Because of this, when people talk about self-location probabilities, they typically have synchronic probabilities rather than forward-looking probabilities in mind. After one makes a measurement and causes a split, a relative observer finds herself in a world. She can *then* ask regarding the probability that she is in fact in a world where the result was, say, x-spin up. But, inasmuch as such probabilities do not tell one what one should *expect* as the result *before* the measurement, they do not capture standard quantum predictions.

DeWitt and Graham seem to have wanted an explanation that addressed the question of what measurement statistics one should *expect*. It is unsurprising that they would inasmuch as it is precisely such forward-looking probabilistic predictions that make quantum mechanics empirically attractive in the first place. But if so, they clearly failed to deliver it. Graham sensed as much:

> Thus we conclude that values of the relative frequency near [the standard predictions] will be found in the majority of Everett worlds of the apparatus and observer. If we assume our own world to be a "typical" one, then we may expect

[10] See Barrett (1999) for a detailed description of this approach and how it relates to other no-collapse theories. As we will see, one also gets perfectly good forward-looking probabilities from no-collapse theories like Bohmian mechanics (Chapter 11) and the single- and many-minds theories (section 10.5).

> a human or mechanical observer to perceive relative frequencies in accord with the [standard predictions]. Why we should be able to assume our own world to be typical is, of course, itself an interesting question, but one that is beyond the scope of the present paper. (1973, 252)

One needs at least *this* assumption to get from talk of typical worlds to talk of quantum *probabilities* or *expectations*.

A second problem with the DeWitt and Graham line of argument is their implicit assumption of a principle of indifference. It is clearly not the case as a simple matter of rationality that if there are *n* possibilities, then the probability of each is $1/n$. Consider the two possibilities (1) the sun will explode tomorrow and (2) the sun will not explode tomorrow. It is simply false that the probability of each is 1/2. This is why an ignorance clause is typically included in such principles: if there are *N* possibilities *and if one has no other information*, then the probability one should assign to each possibility is $1/n$. But, even so, there is no reason at all to require a rational agent to assign probabilities in this way.[11]

For something to be a principle of reason, there should be negative consequences of some sort for not heeding it. A good Bayesian, however, might assign any set of prior probabilities that are coherent (i.e. satisfy the standard axioms of probability theory) and non-dogmatic (i.e. not zero or one) to her *n* hypotheses without fear of finding herself committed to a Dutch book or failing to respond appropriately to relevant evidence. The probabilities that a rational agent *should* assign are always simply whatever coherent, non-dogmatic credences *she in fact has*. These might be evenly distributed over the possibilities at hand or not. That there is no failure of *reason* if the agent's de facto priors are not evenly distributed over a given partition is evident from the fact that nothing bad will happen as a result.

Closely related, it is entirely unclear how one might understand a principle of indifference as having any specific content. One can only assign probability $1/n$ to each possibility if one knows what the possibilities are. A principle of indifference, then, only has content when supplemented with the choice of a specific partition. The philosopher Bas van Fraassen's (1989) cube factory story illustrates the point. Suppose one only knows that a factory produces cubes with a side between 0 and 2 meters. If one considers side length, one might imagine that a principle of indifference requires that one take the probability of a randomly selected cube having a side between 0 and 1 to be 1/2 since *side lengths* range from 0 to 2. If one considers volume, one might imagine that a principle of indifference requires that one take the probability of a randomly selected cube having a volume between 0 and 1 to be 1/4 since *volumes* range from 0 to 4. But since having a side length between 0 and 1 is the same thing as having a volume between

[11] The principle of indifference goes back to at least Pierre-Simon Laplace. See Hàjek (2011) for a brief introduction to the history of the principle. See Zabell (2016) for an example of early worries regarding its status as a principle of reason.

0 and 1, the different partitions, each perfectly natural given different interests, yield inconsistent probability assignments.

The moral is that there is no content to a principle of indifference without a specified partition and what partition one finds natural depends on one's interests. Inasmuch as each is tailored to a specific application, any particular principle of indifference is a poor candidate for a general principle of reason. One might assign even priors over whatever set of possibilities one currently takes as salient, but I can see no plausible argument for a *principle of reason* that compels one to do so.

Returning to the issue at hand, since pure wave mechanics specifies no canonical decomposition of the full state, it cannot, by itself, tell one what set of branches one should use in assigning each branch probability $1/n$. This problem is made particularly acute by the fact that choosing a determinate-record basis in order to guarantee that each relative observer has determinate measurement records then assigning each branch probability $1/n$ typically yields the *wrong* quantum probabilities anyway. DeWitt and Graham must assume *both* a principle of indifference *and* a special, preferred way to individuate branches to even get started in deriving the standard quantum probabilities. Stipulating precisely the right number of duplicate worlds to get the right quantum probabilities on a principle of indifference that one finds intuitively attractive is manifestly ad hoc.[12]

Finally, it is entirely unclear why one should expect one's *intuitions* concerning what one finds to be rational or natural to be relevant to the laws of physics. Even if one had an intuitively compelling principle for assigning objective prior probabilities to physical possibilities, there is no good reason to trust it here. Quantum phenomena themselves vividly illustrate the unreliability of our physical intuitions. If one accepts such a principle, it must be as a part of one's commitment to the particular formulation of quantum mechanics it serves. We will see how this works in a concrete case when we consider the distribution postulate in Bohmian mechanics (Chapter 11).

There is a long tradition of people adding principles to pure wave mechanics that would allow one to derive the standard quantum probabilities over branches. A number of such derivations pass through a step where it is argued that one should be indifferent among a specified set of branches with equal quantum amplitudes.[13] Different derivations appeal to different assumptions here, but any

[12] Indifferent priors are sometimes suggested as a way of representing a complete lack of initial information. But judgments of one's lack of information are also relative to one's interests. As suggested by the box-factory example, if one is interested in information regarding side length, one partition may seem more natural; if one is interested in information regarding volume, another partition may. Similarly, in the context of pure wave mechanics, if one is interested in information regarding determinate measurement records, one decomposition of the state may seem more natural; if one is interested in information regarding another observable, say energy, another decomposition may. And, as we just saw, the application of a principle of indifference over different partitions typically yields incoherent probability assignments.

[13] See, for example, Wallace (2010b) and Sebens and Carroll (2015). Inasmuch as they require that nontrivial auxiliary assumptions derive quantum probabilities, the suggestion is that they should be

argument that gets to this step is making assumptions that are together equivalent to assuming something like a special principle of indifference. The more direct one's appeal to a principle of indifference, the quicker the argument, but, for the reasons we have just been considering, also the less compelling the argument.

Returning to Everett's project for a moment, insofar as he was not concerned to derive probabilities *over branches*, he had little use for a principle of indifference. In order to account for appearances *within a typical branch*, he just needed to specify a typicality measure over branches. One would need something like a principle of indifference to infer probabilities over branches from typicalities over branches.[14]

If one wants probabilities *over branches*, one needs (1) something that says how to assign probabilities over branches and (2) something that says what the probabilities concern. To get a theory that makes the *standard quantum predictions* one further needs a way to assign forward-looking probabilities to possible measurement outcomes.

10.4 Decohering Worlds

Choosing the special preferred basis in the context of the splitting-worlds theory is as difficult as saying when collapses occur in the standard collapse theory. The decohering-worlds formulation of quantum mechanics seeks to address this problem by providing a rule that determines what worlds there are given (1) the physical interactions between the systems of interest and (2) a specified level of description of those systems. This provides a way to characterize alternative sets of post-measurement worlds that each contain determinate measurement records at a specified level of description.[15]

Worlds on this view are *emergent*. How one characterizes them depends on one's level of description, but they are nevertheless fully real entities. David Wallace gives an analogy between worlds and tigers. Just as one might understand tigers as emergent patterns, or structures, *within* the states of a classical microphysical theory, one should understand worlds as emergent patterns, or structures, *within* the global quantum-mechanical state (2010a, 56). Such worlds, like medium-sized objects in a microphysical theory, might be individuated differently at different levels of description. At a particular level of description one might identify worlds

specified along with the rest of the physical theory. We will say more about Wallace's approach in particular when we discuss decohering worlds in the next section.

[14] Let's call such an auxiliary assumption, one connecting typicalities to probabilities, a ⋆-principle. A ⋆-principle might be thought of as a sort of generalized principle of indifference. One needs to supplement one's theory with a ⋆-principle to assign probabilities to branches. But one also needs to say what such probabilities are probabilities *of*.

[15] See Saunders (2010) and Wallace (2012) for descriptions of the decohering worlds formulation of quantum mechanics. The decohering worlds formulation is closely related to the decohering histories. See Griffiths (1984) and Gell-Mann and Hartle (1990) for descriptions of this approach.

with substructures of the quantum state. As Wallace puts it, they are "mutually dynamically isolated structures instantiated within the quantum state, which are structurally and dynamically 'quasiclassical' " (2010a, 70). Because of the constant possibility of interference effects between branches and the contingency on one's level of description, one might put it more precisely by saying that worlds are quasi-isolated, quasiclassical structures *at an appropriately specified level of description*.

Proponents of the decohering worlds formulation like to think of it as nothing more than Everett's pure wave mechanics.[16] It is, however, better characterized as pure wave mechanics with at least two auxiliary assumptions. One is a rule for how to individuate worlds given the interactions between physical systems, and the other is a rule for what probabilities are and how to assign them to worlds. The rules for individuating states and assigning probabilities are typically presented as self-evident given the dynamics of the theory or as resulting directly from basic principles of reason. Regarding the former, the thought is that emergent worlds are just *found* in the decohering structure described by the theory. Regarding the latter, the thought is that probabilities over these worlds are *determined* by basic principles of rational action that one would arguably want to adopt if one were committed to decohering worlds as an accurate description of the physical world.[17]

There is good methodological reason to include whatever auxiliary assumptions are central to one's physical explanations in the description of one's physical theory. The specification of what constitutes a physical world and how such worlds are individuated is meant to do real explanatory work in accounting for determinate measurement records. And since pure wave mechanics itself says nothing whatsoever regarding quantum probabilities, the details of one's understanding of probabilities in decohering worlds have real explanatory work to do as well. In short, tracking the auxiliary assumptions one needs to explain determinate measurement records and whatever is meant to count as quantum probabilities will help us be clear concerning what is doing the explaining and, hence, what one might want to include in one's specification of the theory itself.

One might take the rule for how to understand and individuate worlds as having something like the following form:

3. Interpretation of states (decohering worlds): There is a branching structure $\mathcal{B}(t)$ that is associated with the linearly evolving state of the universe $|\psi(t)\rangle_U$. This branching structure is determined at a specified level of description L by the decohering interactions between physical systems. Each branch

[16] While it is quite different from what Everett had in mind, proponents of the decohering worlds formulation typically ascribe it to Everett. They further adopt his goal of deducing the standard quantum probabilities from pure wave mechanics alone. But unlike Everett, they typically want self-location probabilities *over* worlds rather than quantum statistics *within* a branch.

[17] See Wallace (2010b) for a careful description of how probabilities over branches are derived. We will have more to say about this in a moment.

in $\mathcal{B}(t)$ represents a world at level of description L. The properties of this world are determined by the local state associated with that branch.

On this view, the world will split (at a specified level of description) whenever the state of a superposed object system becomes correlated to the state of a recording property of a measuring device that decoheres with its environment at the specified level of description. Since decoherence is always a matter of degree, one might think of the level of description as specifying how much and what sort of decoherence one requires in order to feel comfortable saying that the worlds have split given the systems and properties one finds salient at the given level of description. This depends on what one wants to explain. If one is just interested in a coarse-grained specification of something like the position of a macroscopic pointer, typical decoherence effects will guarantee determinate measurement records in each roughly specified, decoherence-induced branch. The observer's determinate experience of *seeing* the pointer indicating a particular result is to be explained by the fact that she occupies a world where the pointer in fact indicates that result at an appropriate level of description.[18]

Whenever the environment is correlated with the value of a physical record, the record will be determinate at a level of description appropriate for characterizing the value of the record. Inasmuch as actual measurement records are meant to be read, it will be hard to keep macroscopic records from decohering in precisely that parameter used to indicate the measurement result. Hence most systems that one would think of as indicating a measurement result, systems like the pointer on a real measuring device, will indicate a determinate measurement result in each world at a level of description appropriate for describing macroscopic systems.

To see how this works, let's return to the Wigner's friend story but consider it in a less idealized context. On the decohering-worlds formulation, worlds begin to split relative to the a macroscopic-pointer level of description when the position of the pointer on the friend's measuring device M becomes correlated with the x-spin of S. Since the pointer is a macroscopic system, in any everyday-world version of the story its environment will very quickly become correlated with its position. Environmental decoherence dynamically selects the approximate-pointer-position basis as privileged since it is a property to which the environment correlates at the specified level of description. Including the environment E, the resultant state will be something like

$$\alpha|\text{“}r\text{”}\rangle_F|\text{“}\uparrow_x\text{”}\rangle_E|\text{“}\uparrow_x\text{”}\rangle_M|\uparrow_x\rangle_S + \beta|\text{“}r\text{”}\rangle_F|\text{“}\downarrow_x\text{”}\rangle_E|\text{“}\downarrow_x\text{”}\rangle_M|\downarrow_x\rangle_S$$

when written in the environmentally selected basis. Here there are two roughly defined worlds—one where S is x-spin up and M records x-spin up and the other

[18] See Wallace (2012, 46–102) for a description of the process by which approximate decohering worlds form at a specified level of description.

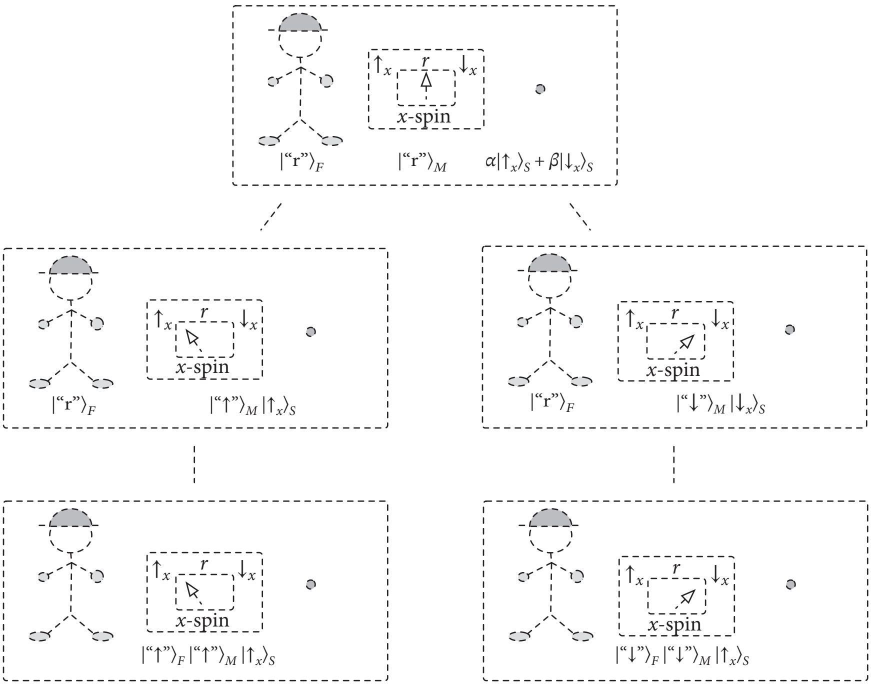


Figure 10.3. Rough decohering worlds resulting from a spin measurement.

where S is x-spin down and M records x-spin down. Then after F looks at the pointer, the state will be something like

$$\alpha|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_E|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + \beta|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_E|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S,$$

again with two roughly defined worlds—one where one where S is x-spin up and M and F record x-spin up and the other where S is x-spin down and M and F record x-spin down (as in Figure 10.3). The dotted lines indicate that the worlds are dependent on one's level of description and, even so, a bit fuzzy—indeed, fuzzy enough that there is typically no simple matter of fact about how many worlds there are. But more on this in a moment.

Because of the interaction with the environment, the final state above is not an eigenstate of the usual Wigner's friend A observable. If W were to make an A-measurement, he would get the results $+1$ and -1 with approximately equal probability just as when a single particle is correlated with the pointer position.[19] That is, W would not get the interference effect of always getting the result $+1$ and hence not detect interference between the branches. Hence, in this specific

[19] The result $+1$ is still slightly favored as long as the correlation with the environment is less than perfect.

and limited sense, it would seem to W that a collapse had occurred in the box. Of course, there is another A-type observable of the composite system *FEMS* that has the final state above as an eigenstate with eigenvalue $+1$. Call this observable B. If W measured B it would show that there was in fact no splitting of worlds into *causally isolated* entities. But measuring B would be yet more difficult than measuring A since one would necessarily need to include the environment in the measurement and the environment will itself decohere.

But it is not the *difficulty* in making an A-type measurement that explains the experience of a determinate measurement outcome. Observers typically do not even try to make such measurements. Rather, the explanation is that there is a *level of description* appropriate for characterizing the relevant properties of the macroscopic systems and the properties of these systems at this level of description are such that each copy of the friend inhabits a different world and each world presents the friend with a determinate measurement record at that level of description.

In order to explain the friend's determinate experience, one wants to say that there are two friends inhabiting different worlds at the end of the measurement and that each has a different, yet perfectly determinate, measurement record. The decohering-worlds formulation provides this, but only at a specified level of description. If one wants to capture Everett's understanding of the Wigner's friend story, one must also be able to say that, at *another* level of description, there is just one world inhabited by a single friend who is part of the composite system that is in an eigenstate of an A observable and would, hence, exhibit the inference effect of yielding $+1$ as the result of the A-measurement if it were ever performed. That one cannot in fact measure it does nothing whatsoever to change the fact that the linear dynamics *requires* an interference outcome if one could measure it. Both of these facts can be accommodated by decohering worlds, but at different levels of description.

The decohering worlds formulation, then, provides *a sense in which* one might *at a specified level of description* understand the post-measurement state as one where there are two friends, each with a determinate measurement record. It is *this* that is meant to explain the experience of a determinate measurement outcome. There is *also a sense in which* one might *at another level of description* understand the post-measurement state as one where there is just one friend without any particular determinate measurement record. It is *this* that would explain why the composite system is in an eigenstate of a A-type observable if one could ever make such a measurement. How many worlds there are and the states of those worlds depends on one's level of description, and one chooses a level of description appropriate to what one wants to explain.

Macroscopic systems will evolve *almost classically* in each world at a level of description appropriate for describing the interactions between such systems. This is because under the linear dynamics the state of each world will evolve classically as long as the branch that characterizes that world does not split or interfere with another branch, and decohering worlds are precisely those for which one should expect not to see splitting or interference effects between branches on typical interactions at the appropriate level of description. Part of the point of saying *almost* classically is that the worlds are virtually never perfectly isolated at any level of description.

In order to understand this as more than just a discussion of linguistic convention, one needs to take seriously the emergent metaphysics. One wants to be able to say something like that the friend *sees* the result x-spin up because there *really is* a copy of the friend seeing that result at a level of description appropriate for describing friends and their experience. The point is just that whether one gets a satisfactory explanation of what it is like to live in such a physical world will depend on the details of the metaphysical story of emergence one tells and how one understands experience in that story.

The branching process by which worlds emerge is always a matter of degree. It begins as environmental decoherence progressively destroys interference effects at one's specified level of description. For everyday physical systems, one would expect the environment never to be perfectly correlated with the decohering property of one's object system, hence one should always expect some interference between branches, and hence some nonclassicality, even at a level of description aimed at characterizing properties of macroscopic systems. Further, inasmuch as decoherence is always a matter of degree, there is never a simple matter of fact concerning how many worlds there or what the properties of those worlds are.[20] Such questions must always be asked relative to a specified level of description. Since there is always a measurement that would in principle detect interference effects between different worlds that are individuated at one level of description, there is always a level of description at which the worlds did not split at all.

Of course, decoherence effects alone cannot explain why an observer gets a determinate measurement record. Here they are being used to specify a rule for how and when worlds split at a given level of description. The explanatory work is being done by the emergent metaphysics. It is the fact that the observer inhabits a real world with a determinate measurement result at an appropriate level of description for talking about such things that is meant to explain why she really sees that result. And since the level of description one requires for an explanation of macroscopic behavior immediately selects a rough decomposition of the global state into worlds, the decohering-worlds formulation is arguably a significant improvement over the metaphysical story provided by the splitting-worlds theory discussed earlier.

[20] As Wallace puts it "*there is actually no such thing as the number of branches*" (2012, 120). See also Wallace (2012, 99–102).

The *metaphysical* interpretation of the quantum formalism plays an essential role here. Level of description *talk* does not address the determinate experience problem. That the formalism provides a rough sense with which one might describe the observer as having a determinate record at a given level of description does nothing to explain the observer actually having a determinate experience corresponding to that record. There are typically copies of the observer at that level of description that have *incompatible records*, and there are typically other levels of description that describe the observer as *not having any determinate records at all*, indeed levels of description that fail even to characterize there being a determinate observer. The explanation for why a given observer *in fact* has a specified determinate experience is given by the metaphysical commitment to there actually *being* an emergent observer with that record. Again we see that the formulation of quantum mechanics and one's interpretation of the formulation are not independent—metaphysics plays an *essential* role in the theory's explanations.

This leaves us with the problem of accounting for probabilities. If one wants probabilities over decohering worlds, then inasmuch as each world is equally real, one needs a rule that assigns something like self-location probabilities to each of the decohering worlds at a specified level of description. This requires further auxiliary assumptions.

Wallace (2010b) shows how one might derive such a rule in the context of the decohering-worlds formulation supplemented with a set of ten auxiliary assumptions concerning how a rational agent would behave in the context of a world described by such a theory.[21] The idea is to show that an agent who satisfies those assumptions would act as if the standard quantum probabilities obtained. While the auxiliary assumptions are presented as basic principles of reason, since some of them concern such things as an agent's preferences when faced with a *superposition* of options, they go well beyond the assumptions of classical decision theory. As a result, this is manifestly not a case of deriving the standard quantum probabilities from pure wave mechanics alone. Rather, this approach is best understood as showing how one might try to make sense of rational choice *if one were already committed to the truth of the decohering-worlds formulation of pure wave mechanics*. There is nothing wrong with making such auxiliary assumptions, but since they are essential to explaining quantum probabilities and since quantum probabilities represent the empirical content of the theory, there is a good methodological argument for including them in one's specification of the physical theory.

Given appropriate auxiliary assumptions, one can get synchronic self-location probabilities on the decohering-worlds formulation without much difficulty, but it is much less clear how to get the standard forward-looking quantum-mechanical probabilities. The problem is that since every copy of one's future self is fully realized in an emergent world at an appropriate level of description, there is a

[21] Wallace's derivation is a carefully executed version of the decision-theoretic approach to probability in Everettian quantum mechanics introduced by David Deutsch (1999).

straightforward sense in which the forward-looking self-location probability associated with each possible measurement outcome is just one. There is nothing here that the standard quantum probabilities might concern. Without forward-looking probabilities, one does not have the usual quantum predictions concerning what one should expect regarding future measurement outcomes. Hence one ultimately lacks the standard grounds for rational action understood in the standard way.

While it encounters significant conceptual problems, the decohering-worlds formulation has two notable virtues. First, since splitting is just a matter of decoherence at a level of description, this theory, like pure wave mechanics, can be described in a way that is compatible with relativistic constraints.[22] Further, unlike GRW-type theories and Bohmian mechanics, the decohering-worlds formulation does not require any special choice of preferred physical observable. Rather, like pure wave mechanics itself, the formalism provides a framework theory that can be applied to almost any basic characterization of the physical world.

10.5 Single-Mind and Many-Minds Theories

Albert and Loewer's (1988) single- and many-minds theories provide one way to make clear what one means by both synchronic and future-directed self-locating probabilities. The single-mind theory might be thought of as pure wave mechanics with the usual linear dynamics and with the following stipulation and auxiliary dynamical rule:

4*. Mental dynamics (single mind): Every physical observer *F* is associated with a mind that always has a determinate mental state $\mathcal{M}_F$. The observer's mind is associated with a branch in *F*'s physical state written in the determinate-record basis. The physical records represented by the branch determine the mental state $\mathcal{M}_F$. When the branch splits, the observer's mind is randomly associated with one of the resulting branches, with probabilities proportional to the norm-squared of their respective coefficients, and the new branch determines *F*'s new mental state.

On this rule an observer's mental state is randomly determined by the usual quantum probabilities each time a measurement-type interaction involving *F*'s physical records occurs.

Consider the Wigner's friend story. The friend's mental state $\mathcal{M}_F$ starts in the "ready to make a measurement" state, then, after *F* physically interacts with *M*, the friend's mind is randomly associated with the first term of the state resulting from the linear dynamics

[22] It is, however, unclear that one could introduce standard forward-looking probabilities and preserve this virtue.

$$\alpha|``\uparrow_x"\rangle_F|``\uparrow_x"\rangle_M|\uparrow_x\rangle_S + \beta|``\downarrow_x"\rangle_F|``\downarrow_x"\rangle_M|\downarrow_x\rangle_S$$

with probability $|\alpha|^2$ and with the second term with probability $|\beta|^2$. In the first case, the friend's mental state $\mathcal{M}_F$ records the result "x-spin up" and in the second case it records the result "x-spin down." More generally, the probability of an observer getting a particular result is the standard quantum probability of getting that result conditional on the branch the observer's mind is associated with before the measurement.[23]

The physical observer will typically be in an entangled superposition of having recorded mutually incompatible results. Since it will always be in principle possible to detect interference between branches, every branch will be physically real in Everett's operational sense. In particular, an A-measurement of the composite system *FMS* will indicate that there is no collapse on measurement. But the friend's mental state will always represent her as having recorded a particular, fully determinate result. From her perspective, it will *appear* that a random collapse occurred with the standard quantum probabilities. And, unlike in the bare theory, the observer will have a perfectly ordinary mental record of the outcome given by her mental state $\mathcal{M}_F$.

In contrast with the situation in most many-worlds theories, understanding standard forward-looking probabilities here is easy. The probability of an observer getting a particular result is just the chance that her mind will in fact end up associated with the branch representing the corresponding physical record. While the strong mind–body dualism involved in the single-mind theory may strike one as wildly implausible, the explanation of probabilities is perfectly clear. The metaphysics does real explanatory work by saying what the quantum probabilities are probabilities *of*. Here they are simply descriptive of the stochastic evolution of the observer's mental state $\mathcal{M}_F$.

There is a sense in which the mental dynamics of the single-mind theory is fully determined by the suggestive properties of the bare theory alone *if one requires that an observer's mental state agree with what she will have the sure-fire disposition to report*. In particular, the bare theory's limiting properties tell us that the observer will approach a state where she has the sure-fire disposition to report that her results were randomly distributed with the standard quantum-mechanical relative frequencies. For this report to be in fact true, her sequence of mental records must be randomly determined by the standard quantum probabilities precisely as specified by the single-mind theory's mental dynamics. The suggestive properties of the bare theory similarly place very strong constraints on the determinate records one might introduce to *any* no-collapse formulation of quantum mechanics.[24]

[23] The single-mind theory might be thought of as a hidden-variable theory where the hidden variable is an observer's actual mental state.

[24] This includes placing dynamical constraints on hidden-variable theories like Bohmian mechanics (Chapter 11). See Barrett (1999) for a discussion of this and of the "transcendental" strategy for extending the bare theory more generally.

The single-mind theory might be understood as a no-collapse hidden-variable formulation of quantum mechanics where the observer's mental state $\mathcal{M}_F$, rather than a physical parameter like position as in Bohmian mechanics (Chapter 11), is the so-called hidden variable that explains her determinate measurement records. Among its virtues is that, unlike other hidden-variable theories, the single-mind theory requires no *physically* preferred basis. Rather, the preferred basis is just part of the rule that connects physical states to mental states. The single-mind theory also clearly provides determinate measurement records in terms of the observers' mental states, and hence directly explains the observer's experience. But this direct explanation comes at a cost.

This way of getting determinate measurement records and forward-looking probabilities is blatantly ad hoc. Since an observer's mental state does not supervene on her physical state, like Wigner's solution to the measurement problem, the single-mind theory requires a strong variety of mind–body dualism. Indeed, the particular way that supervenience fails here leads to a sort of solipsism.

Consider the Wigner's friend story again. Suppose that *F*'s mental state records x-spin up as her measurement outcome. Suppose that *W* then measures the x-spin of *S*. The mental dynamics predicts that *W* will get the mental-state result x-spin down with probability $|\beta|^2$. Suppose he does. Here *F* and *W* will have incompatible mental records for the same "physical" property, but they will never know it. When they compare their results, by means of a *physical* process described by the linear dynamics, *F* and *W* will end up believing that they got precisely the same result because of the bare theory's intersubjective agreement property. So they will be under the illusion that their experiences agree when they in fact do not.

This type of illusion is less pernicious than the sort predicted by the bare theory. At least here the observers really get determinate measurement results with perfectly ordinary determinate content, and they get these results with the standard forward-looking probabilities. Nevertheless, since there is no reliable means for them to share their experiences, if the theory is in fact true, an observer could only check its predictions against *her own* experience. One might say that the theory is *empirically coherent*, but only with respect to one's own experience.

Albert and Loewer proposed their many-minds theory specifically to address this failure of supervenience. The many-minds theory associates a continuous infinity of minds with each physical observer. One might think of the theory as pure wave mechanics supplemented with the following stipulation and auxiliary dynamics:

4⋆. Mental dynamics (many minds): Every physical observer *F* is associated with a continuous infinity of minds. Each of these minds is independently associated with a branch in *F*'s determinate-record basis and always has the corresponding determinate mental state. When the branch that a particular mind is associated with splits, that mind is randomly associated with one of the resulting branches with probabilities determined by the norm-squared of their respective coefficients.

Here the mental state of each mind evolves just as in the single-mind theory, only there are more of them. The (global) observer experiences all possible measurement outcomes in the sense that *some* of her minds experience each, but each (local) mind has the fully determinate experience of a particular measurement outcome. It is the local minds that are meant to explain ordinary experience.

Now with probability one there will be *some* mind of *W* that gets the same result as any particular one of *F*'s minds. So agreement between *F* and *W* is, in this curious sense, agreement between sentient agents. One will be able to determine the *proportion* of an observer's minds with each mental state from the amplitudes associated with the branches of the observer's physical state. But associating each observer with a continuous infinity of minds seems a high price to pay for this exotic and relatively weak variety of supervenience.

On both the single-mind and many-minds theories, the *physical state* evolves in a linear way that is compatible with the constraints of special relativity. For what it's worth, inasmuch as the proportion of an observer's minds with each possible mental state supervenes on her physical state, one might take this *weak sort of global mental state* (but not the local states of the individual minds) to evolve in a way that is compatible with relativistic constraints on the many-minds theory.

Whether or not one thinks that the sort of supervenience one gets provides a sufficient motivation for the theory, the many-minds formulation is instructive inasmuch as it illustrates how one might construct a *non-splitting* many-worlds theory. The thought is to suppose that the complete history of the world *as it appears to each of a particular observer's minds* is in fact the history of a non-splitting, physically distinct world. As each mind evolves, it traces a path through the branching structure determined by the evolution of the global quantum-mechanical state written in the determinate-record basis. Each path, or thread, is then understood as the history of a non-splitting physical world. This is the idea behind the *many-threads* formulation of quantum mechanics.

10.6 Many Threads and Many Maps

The *many-threads* formulation of quantum mechanics provides another way to make sense of both synchronic and forward-looking probabilities. The theory might be thought of as pure wave mechanics with the following rule:

4†. World dynamics (many threads): There is an infinite set of non-splitting worlds. The history of each world corresponds to a path through the branching structure determined by the evolution of the global quantum state written in the determinate-record basis $\mathcal{B}$. The prior probability of each world being *ours* is given by the standard quantum probability assigned to its history.

The quantum-mechanical state, its linear evolution, the preferred basis, and the norm-squared amplitude measure might be understood here as together providing a procedure for specifying the set of possible complete histories and a probability measure over that set. The particular sequence of measurement records we experience is explained by the fact that those are the actual physical records of the world we inhabit. The prior probability of each possible world history being ours is given by the standard quantum probabilities. And, as we learn further properties of our world, we condition on this evidence to update our degrees of belief concerning future events.[25]

Since there is no branching process to copy an observer, there is no special problem in assigning forward-looking probabilities to future measurement results. Probabilities of future events are simply degrees of belief concerning the features of the world history one takes oneself to in fact inhabit.

At root, the many-threads formulation of quantum mechanics just consists in a procedure for specifying a set of physically possible worlds along with their complete histories and a probability measure over the set of world histories that indicates the prior probability of each being ours.[26]

If one wants something that looks more like a relativistic formulation of quantum mechanics, one might generalize the notion slightly and consider a *many-maps* theory. Such a theory consists in a set of complete spacetime maps of events (which might be given in terms of something like particle locations, mass densities, or flashes) and an epistemic probability measure over that set of maps. Each map would give the full *spacetime* history of a possible physical world, and the measure over the set of maps might be thought of as the prior probability that each presents the history of our world.

We have already seen something very like a many-maps theory. Tumulka's (2007) relativistic formulation of GRWf might be understood as using hypersurface-dependent wave functions to provide a rule for distributing flashes in spacetime. On this view, the theory might be understood as a specified set of possible spacetime maps of flashes together with an epistemic probability distribution over the set representing the prior probability of each world being the one we in fact inhabit. One's determinate measurement records are then taken to supervene on local constellations of flashes in the map that is descriptive of our world.

Similarly, a hidden-variable theory like Bohmian mechanics (Chapter 11) might be understood as a many-threads theory where there is one world history for each

[25] As described in the last section, the many-threads theory might be thought as the many-minds theory where the trajectory of each mind determines the complete history of a physically possible world. See Barrett (1999) and (2005b) for more detailed discussions of the many-threads and many-maps approaches.

[26] The worlds are individuated by their full histories and never split. As suggested by the wording here, quantum probabilities are stipulated by the theory and are naturally thought of as epistemic. The way that one individuates the possible worlds might involve decoherence considerations. This type of theory is sometimes called a *many-histories* formulation of quantum mechanics. See Griffiths (1984) and Gell-Mann and Hartle (1990) for examples of this approach.

possible initial particle configuration relative to the wave function and where the distribution postulate provides an epistemic probability measure of each world history in fact being ours. But this can be looked at the other way around as well. Since we only need one world, *our world*, to explain our actual experience, the other worlds in a many-threads theory are in an important sense superfluous. Given this, one might think of a many-threads theory as a hidden-variable theory that gives epistemic probabilities for all possible evolutions of the hidden variable, only one of which describes our actual world.

It is clear what the quantum probabilities associated with future events concern on a many-threads or many-maps interpretation of quantum mechanics. They simply represent the epistemic probability that those events will occur in *our* world. More generally, if one adopts a metaphysics of *non-splitting* worlds, the standard quantum probabilities can be understood as degrees of belief concerning what the future looks like in the world one in fact inhabits.

Inasmuch as a many-maps theory just consists of a set of possible spacetime maps and an epistemic probability measure over that set, one might argue that it is compatible with relativity by dint of the fact that are no dynamical laws that might be incompatible with the dynamical constraints of relativity. But this is a particularly cheap variety of compatibility, and, as we discussed in connection with Tumulka's relativistic formulation of GRWf, that a theory is compatible with relativistic constraints does not mean that it is relativistic theory in the sense of predicting relativistic quantum phenomena. One might try to reverse-engineer such relativistic phenomena as pair-production and -annihilation as events in the context of a many-maps formulation of quantum mechanics, but one should want such phenomena to follow from the dynamical constraints of relativity in the context of quantum mechanics. Pasting relativistic phenomena into a theory after the fact is bound to look as ad hoc as it in fact is.

The explanatory weakness of the approach goes deeper. Many-threads and many-maps formulations fail to provide any dynamical explanations at all. The explanation for every event is just *because that's what happens here and now in the physical world we inhabit.* Of course, a *typical world* in the sense given by the epistemic probability measure associated with a many-maps theory can be expected to exhibit the standard quantum statistics, but it should be clear by this point why there is nothing particularly virtuous in that. That said, depending on one's sense of the relative explanatory costs, one might be willing to give up dynamical explanation in order to get the sort of account of probability that one wants. It is presumably a significant virtue that one can understand probabilities in a perfectly straightforward way as epistemic probabilities concerning self-location in worlds with complete histories in the context of a many-threads or many-maps theory.[27]

[27] The situation here is closely analogous to the one faced by those Bohmians who specify Bohmian mechanics (Chapter 11) in terms of a typicality measure then aim to get probabilities from typicalities.

While a many-threads or many-maps formulation of quantum mechanics avoids the sort of mind–body dualism of something like the single- and many-minds theories, this also comes at a cost. The determinate-record basis in a many-threads or many-maps theory is *physically* preferred. As in the splitting-worlds theory, one's choice of $\mathcal{B}$ determines what physical properties will in fact be determinate in each of the physically possible worlds. Insofar as one's choice of $\mathcal{B}$ is designed to make actual measurement records determinate, stipulating a just-right set of physical facts looks blatantly ad hoc. Of course, one might try to take some of the sting out of this by using decoherence considerations to individuate the threads or maps at a specified level of description.

10.7 Epistemological, Pragmatic, and Information-Theoretic Interpretations

On the many-threads formulation the quantum-mechanical state is just used to get an epistemic probability measure over possible physical histories. In a similar spirit, there is a class of information-theoretic formulations of quantum mechanics where the standard quantum-mechanical state is understood as representing an agent's epistemic or informational state. On this sort of theory, the collapse of the state simply represents the fact that the observer has *learned* something about the system measured. There are a number of such theories, and each works a little differently.[28]

Suppose observer F is prepared to measure the x-spin of a system S in an eigenstate of z-spin. On an epistemological interpretation, that F assigns S the state

$$1/\sqrt{2}(|\uparrow_x\rangle_S + |\downarrow_x\rangle_S)$$

before the measurement indicates that F *does not know* what the measurement result will be and assigns unbiased probabilities over each possibility as indicated by the norm-squared coefficients. That she assigns $|\downarrow_x\rangle_S$ after getting x-spin down

Here one argues that physical worlds that are typical in a sense specified by one's theory should be expected to exhibit the standard quantum phenomena, then one supplements the theory with auxiliary assumptions that allow one to infer that one should expect to find oneself in a world that is typical in just this sense. For examples of this approach see Bricmont et al. (eds) (2001). Here, as in the Everettian tradition, one cannot derive probabilities from typicalities without auxiliary assumptions. And since these assumptions play a salient explanatory role in the theory (by explaining ones probabilistic expectations), the clearest strategy is to add the auxiliary assumptions to one's specification of the theory.

[28] See Fuchs (2010), Healey (2012), and Bub (2015) for examples in this tradition. See also Chiribella and Spekkens (eds) (2016) for a discussion of information-theoretic approach and generalized probabilistic formulations of quantum mechanics.

as the result indicates that she now *knows* what she would get if she were to measure the x-spin of the system again.

Inasmuch as quantum states represent the epistemological states of observers, states are indexed to observers on this approach. In the Wigner's friend story, the external observer W might assign the state

$$\frac{1}{\sqrt{2}}|\text{``}\uparrow_x\text{''}\rangle_F|\uparrow_x\rangle_S + \frac{1}{\sqrt{2}}|\text{``}\downarrow_x\text{''}\rangle_F|\downarrow_x\rangle_S$$

to the composite system $F+S$ when F herself assigns the state $|\text{``}\downarrow_x\text{''}\rangle_F|\downarrow_x\rangle_S$ to that system. In this case, the state from W's perspective indicates that he does not yet know what result F got but does believe that an A-measurement of $F+S$ would yield the result $+1$ in agreement with his description of the state as this particular entangled superposition.

On an epistemological interpretation, one might take an observer's beliefs to concern the values of physical quantities that are not fully described by the quantum-mechanical state or, more simply, to concern expectations for the results of future measurements. In the first case, the theory might be thought of as a sort of hidden-variable theory. In the second, it looks more like a predictive algorithm.

There are a number of puzzles such theories face. One of these concerns self-measurement.[29] If F performs an x-spin measurement on S and gets x-spin down, then the state of $F+S$ from her perspective is $|\text{``}\downarrow_x\text{''}\rangle_F|\downarrow_x\rangle_S$. If W performs an A-measurement on $F+S$ and gets the result $+1$, then the state of $F+S$ from his perspective is the entangled superposition above. But if F performs an A-measurement on $F+S$ in this situation, then the linear dynamics requires that she also gets the result $+1$ in which case the state of $F+S$ from her perspective will be both $|\text{``}\downarrow_x\text{''}\rangle_F|\downarrow_x\rangle_S$ *and* the entangled superposition above. The agent, then, is left with an *inconsistent* epistemic state.[30]

One might address the inconsistency here by using *systems* rather than *agents* to index quantum states. Along these lines, F might maintain logical consistency by assigning the state $|\downarrow_x\rangle_S$ to the system S but, like W, assign the entangled superposition above to the system $F+S$. But inasmuch as quantum states are indexed to systems and not agents on such a proposal, they are not well characterized as simply representing the epistemic state of agents.[31]

[29] See Albert (1992) for an extended discussion of self-measurement.

[30] See Hagar and Hemmo (2006) for other considerations along these lines.

[31] There is also a class of theorems that cause problems for specific epistemological formulations of quantum mechanics. The most famous of these is the PBR theorem (Pusey, Barrett, and Rudolph 2011 and 2012). That there is always some way to get around such theorems, however, is evident in the observation that any consistent formulation of quantum mechanics that purports to describe the physical world might be reinterpreted as giving merely statistical information regarding what one should expect regarding the results of future measurements.

A more general and telling problem with such formulations is that we have formulations of quantum mechanics that are just as good predictively but that aim to be more directly descriptive of the physical world. A theory that can be taken as descriptive of physical states and processes provides a richer variety of explanation than one that explicitly does not purport to do so. The attractiveness of epistemological, pragmatic, and information-theoretic interpretations of quantum mechanics, then, depends on one believing that those formulations of quantum mechanics that we have formulated with the aim of describing the physical world are ultimately unsatisfactory *even when just understood as predictive algorithms*. It is difficult to see the argument for taking such a line.

11
Bohmian Mechanics

11.1 Bohm's Theory

Bohmian mechanics (1952) is a no-collapse hidden-variable theory. The quantum-mechanical state is represented by a wave function that evolves according to the linear dynamics. But particles also always have fully determinate positions. And they move in a way that is fully determined by the evolution of the wave function. While it will take a while to explain precisely what this means, one might think of the theory as pure wave mechanics but where the actual particle configuration at a time selects a single effective branch of the total wave function. This selection then, in turn, determines the effective properties of all physical systems.

In its simplest form Bohm's theory is characterized by the following principles:[1]

1. Representation of states: The complete physical state of a system S at time t is given by the wave function $|\psi(q,t)\rangle_S$ over configuration space and a point in configuration space $Q(t)$.
2. Representation of observables: All observations are ultimately measurements of position. The value of a measurement record is determined by the wave function $|\psi(q,t)\rangle_S$ and the configuration $Q(t)$. Specifically, the empirical content of the record is given by the effective wave function selected by the configuration.[2]
3. Interpretation of states: The position of every particle is always determinate and is given by the configuration $Q(t)$.
4. Laws of motion:

   I. Linear dynamics: The wave function evolves in the standard unitary way. In the simplest case this is given by

$$i\hbar\frac{\partial|\psi(q,t)\rangle_S}{\partial t} = \hat{H}|\psi(q,t)\rangle_S.$$

[1] This description of Bohm's theory follows Bell (1987, 127) rather than Bohm's (1952) original quantum-potential description.

[2] Presentations of Bohmian mechanics typically just say something along the lines of the first sentence of this rule. We will consider here what the second and third sentences mean and how they follow from the details of how measurement works in Bohmian mechanics. See Dürr, D., S. Goldstein, and N. Zanghì (1992) for an an introduction to the idea of an effective wave function.

II. Particle dynamics: Particles move according to

$$\frac{dQ_k(t)}{dt} = \frac{1}{m_k} \frac{\mathrm{Im}(\langle \psi(q,t)|\nabla_k|\psi(q,t)\rangle)}{\langle \psi(q,t)|\psi(q,t)\rangle}\bigg|_{Q(t)}$$

where m_k is the mass of particle k and Q is the current configuration.

5. Distribution postulate: The prior epistemic probability of the configuration $Q(t_0)$ being in region R of configuration space at an initial time t_0 is given by the standard quantum probability

$$\int_R |\langle \psi(q,t_0)|\psi(q,t_0)\rangle|^2 dr.$$

6. Composition: One composes the wave function associated with different systems and properties using the tensor product in the standard way.

Recall that configuration space for a system S is a $3N$-dimensional space where N is the number of particles in S. Each point in configuration space, then, determines the three-dimensional position of each particle in S. The usual quantum-mechanical state representing particle positions can be given as a complex-valued function over configuration space. But configuration space plays a more central, conceptual role here.

In Bohmian mechanics, both the wave function $|\psi(q,t)\rangle_S$ and the current particle configuration $Q(t)$ evolve in configuration space. Since the integral of the probability density $|\langle \psi(q,t_0)|\psi(q,t_0)\rangle|^2$ over configuration space is always one, one might think of probability as a compressible fluid. As the wave function evolves deterministically according to the standard linear dynamics, rule 4I (Bohm), the probability fluid flows about in configuration space just as a high-dimensional, compressible fluid might. The auxiliary particle dynamics, rule 4II (Bohm), characterizes the point representing the full particle configuration $Q(t)$ as being carried along by the probability current in configuration space as if it were a massless particle. The evolution of this point in $3N$-dimensional configuration space, in turn, determines how the N particles of the system move in ordinary three-dimensional space.

This picture of the wave function as a field in configuration space that pushes around the current configuration is crucial to understanding the theory. As John Bell, a strong proponent of Bohmian mechanics, put it:

> *No one can understand this theory until he is willing to think of ψ as a real objective field rather than just a "probability amplitude." Even though it propagates not in 3-space but in 3N-space.* (1987, 128)[3]

[3] The italics here are Bell's. They indicate the importance he placed on this passage. Because dynamical stories are most naturally told in configuration space in Bohmian mechanics, some proponents of

The quantum probabilities specified by rule 5 can be thought of as epistemic. Given the dynamics, if the epistemic probability density for the particle configuration is ever given by the standard quantum probabilities $|\psi(q, t)|^2$, then it will continue to be so until one makes an observation. Since one can think of the configuration as being carried by the probability current, the epistemic probability of the configuration being in a region R changes precisely as the amount of probability fluid in R changes. After a measurement, the posterior epistemic probabilities are given by the standard quantum probabilities conditional on the value of one's record. This distribution is given by the norm-squared of the *effective wave function* given the record.[4] Since the distribution postulate stipulates that one assign a prior probability density $|\psi(q, t_0)|^2$ to the particle configuration at time t_0, Bohmian mechanics predicts the standard quantum probabilities for the distribution of particles.[5] One might want to weaken the distribution postulate so that it just says that an initial state is *typical* in the measure over configuration space given by $|\psi(q, t)|^2$, but then, as we saw in the discussion of Everettian quantum mechanics, one would need to add something else to the theory to get from talk of typical states to probabilistic predictions.[6]

In Bohmian mechanics, particle positions are always determinate, and it is this that is supposed to explain experience. For his part, Bell believed that getting the right "positions of things" was both necessary and sufficient to explain experience. He explained how observation works in the theory as follows:

> The fundamental interpretative rule of the model is just that [the particle position] is the real position of the particle at time t, and that observation of position will yield this value. Thus the quantum statistics of position measurements, the probability density $[|\langle\psi(q, t_0)|\psi(q, t_0)\rangle|^2]$ is recovered immediately. But many other measurements reduce to measurements of position. For example, to "measure the spin component σ_x" the particle is allowed to pass through a Stern–Gerlach magnet and we see whether it is deflected up or down, i.e. we observe position at a subsequent time. Thus the quantum statistics of spin measurements are also reproduced, and so on. (1987, 34)

the theory take the real physical world to consist of the full particle configuration and a field represented by the full wave function in configuration space. See, for example, Albert (2013) and (2015). For this variety of wave-function realist, the appearance of ordinary three-dimensional objects occupying locations in three-dimensional space is an emergent illusion, akin to the sort of illusions predicted by GRWr, generated by the evolution of the particle configuration and this high-dimensional field.

---

4 See Dürr, Goldstein, and Zanghì (1993) for a discussion of the equivariance of the standard quantum probability distribution.

5 Note that Bohmian mechanics violates both directions of the standard eigenvalue–eigenstate link. Every particle always has a determinate position even when it is not in an eigenstate of position, and it almost never is. And, as we will see (p. 197), it is possible for the wave function to represent a particle as being in an eigenstate of one location when the particle is in fact somewhere else.

6 For this sort of typicality approach in Bohmian mechanics and discussions of it see Dürr, Goldstein, and Zanghì (1992), Callender (2007), Goldstein (2012), and Hemmo and Shenker (2015).

Given this description, one might imagine that the theory explains experience because we directly see where things are—that the real positions of things are simply given, and that we, hence, have immediate epistemic access to those positions. Then, since the theory predicts the standard quantum probabilities for the positions of things, it makes the standard quantum predictions insofar as every measurement is in fact ultimately a measurement of position. Hence, one might suppose, an ontology of *things with positions* immediately explains our experience in the context of the theory.

But observation in Bohmian mechanics is more subtle than this description might suggest. Observing the position of something is an indirect process, and it never quite tells one where the thing is. One cannot *simply see* where an electron is after it passes though a Stern–Gerlach device or at any other time. The issue is not that electrons are small and hard to detect. It is that if one ever did know the precise position of any object, or even just knew it more precisely than allowed by the standard quantum probabilities, then one would be able to predict its future behavior more precisely than the standard quantum probabilities allow. The empirical adequacy of Bohmian mechanics depends on the fact that we *do not* have epistemic access to the precise position of anything.

Determinate particle positions play a role in explaining determinate records, and hence experience, but not by directly representing the empirical content of those records. Rather, determinate positions explain determinate records in Bohmian mechanics by selecting an effective wave function or branch of the quantum-mechanical state.

We will get at what this means in two steps. First, we will consider what the theory predicts regarding the motion of a particle in a spin device of the sort described by Bell. Then we will consider what it means to observe the position of something. This will illustrate how contextual properties like spin work, the nature of measurement records, and how the theory explains experience.

11.2 Basic Spin Experiments

Consider an x-spin experiment of the sort Bell describes. A few preliminary reflections will help to set up the experiment.

In Bohmian mechanics the precise physical geometry of one's measuring device matters. We will consider a particle passing through a Stern–Gerlach magnet where the inhomogeneous magnetic field is oriented in the x-direction.

For the purpose of describing these experiences, we will suppose that localized wave packets are spherically symmetric with a constant diameter and a uniform probability density. This assumption makes the dynamical stories we tell easier to picture. One can drop this idealizing assumption once one has a basic sense of how the theory works and replace it with something more like realistic gaussian wave packets.

Since configuration space for a single particle is ordinary three-dimensional space, the state of an electron is given by its position in ordinary three-dimensional space and a wave function defined over ordinary three-dimensional space.

Finally, as we have seen, the wave packet associated with an electron can be represented as a tensor product of a spin component and a position component. In Bohmian mechanics, one might think of the position component as describing the *density of the probability fluid* and the spin component as characterizing the *flavor of the wave packet*.

We will first consider how the electron's wave function behaves, then how the electron itself moves. Suppose that the electron's initial wave packet is $|\uparrow_x\rangle_e|0\rangle_e$ (as in Figure 11.1(a)). This might be thought of as representing an x-spin-up-flavored, spherically symmetric, three-dimensional droplet of probability fluid in region 0. The spin flavor of the wave packet determines how it will move in the x-spin device. Given the standard setup of the x-spin device, the x-spin-up-flavored wave packet $|\uparrow_x\rangle_e|0\rangle_e$ would be deflected up to region up_x and hence evolve to $|\uparrow_x\rangle_e|\text{up}_x\rangle_e$. In contrast, the x-spin-down-flavored packet $|\downarrow_x\rangle_e|0\rangle_e$ would be deflected down to region down_x and hence evolve to $|\downarrow_x\rangle_e|\text{down}_x\rangle_e$ (as in Figure 11.1(b)).

The distribution postulate tells us that the initial position of the electron is determined by the standard quantum probabilities. In this case that means that the electron is inside its associated wave packet with probability one. Since we are supposing that the probability density is uniform in the packet, the electron has an equal probability of being in the *top half* and *bottom half* of its spherical wave packet on this side view, and an equal probability of being in the *right half* and *left half* if one were looking at the wave packet from a top-down perspective, etc. Further, by volume, the electron has an equal chance of being in the top half of the *top half* and the bottom half of the *top half* and an equal chance of being in the top half of the *bottom half* and the bottom half of the *bottom half* of the wave packet, etc. In short, while we know with probability one that the electron is in its associated wave packet, we do not know where it is in that wave packet beyond the probability densities given by the wave function.

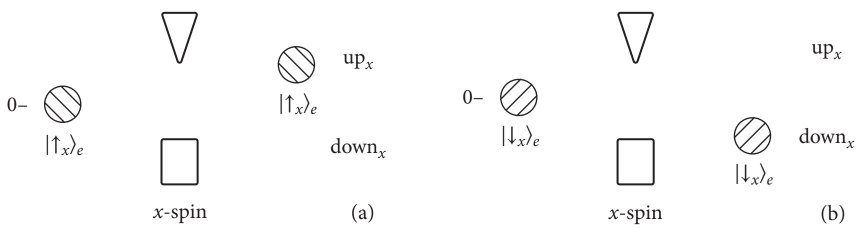


Figure 11.1. How the wave function is deflected by Stern–Gerlach magnets.

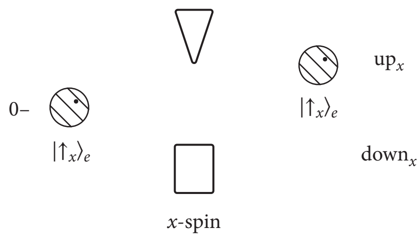


Figure 11.2. How the wave function moves the electron.

Suppose that the electron happens to begin in the *top half* of an x-spin-up-flavored wave packet as indicated in Figure 11.2. Given the standard setup, the wave packet will evolve from $|\uparrow_x\rangle_e|0\rangle_e$ to $|\uparrow_x\rangle_e|\mathrm{up}_x\rangle_e$. For its part, the electron will move as if it were a massless particle being pushed along by the probability current in configuration space. Since, for a single particle, configuration space is just ordinary three-dimensional space, the electron will be carried by its wave packet to region up$_x$ as the wave packet itself moves there. And if the electron's initial wave packet were $|\downarrow_x\rangle_e|0\rangle_e$, the wave packet would evolve to $|\downarrow_x\rangle_e|\mathrm{down}_x\rangle_e$ carrying the electron along with it to region down$_x$.

What it means for an electron to be x-spin up on this account is that it is associated with an x-spin-up-flavored wave packet. Such an electron moves like an x-spin up particle because its x-spin-up-flavored wave packet moves as one would expect an x-spin up particle to move and carries the electron along with it. Spin properties are not intrinsic. Rather, they are contextual properties that particles have by virtue of their being associated with a particular flavor of wave packet.[7]

Now consider how a z-spin up wave packet would behave in an x-spin device. Suppose again that the wave packet starts in region 0. Since

$$|\uparrow_x\rangle_e|0\rangle_e$$
$$\downarrow$$
$$|\uparrow_x\rangle_e|\mathrm{up}_x\rangle_e$$

and since

$$|\downarrow_x\rangle_e|0\rangle_e$$
$$\downarrow$$
$$|\downarrow_x\rangle_e|\mathrm{down}_x\rangle_e,$$

[7] Most of the usual physical properties one might consider are *contextual* in Bohmian mechanics. Contextual properties are not robust intrinsic properties. Part of what this means is that the value of the contextual property one gets on measurement depends on precisely how one performs the measurement. But since Bohmian mechanics makes the standard quantum statistical predictions for measurement outcomes, that most properties are contextual poses no empirical problem whatsoever for the theory.

by the linearity of the wave function dynamics

$$
\begin{gathered}
|\uparrow_z\rangle_e|0\rangle_e = \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|0\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|0\rangle_e \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|\text{up}_x\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|\text{down}_x\rangle_e.
\end{gathered}
$$

This means that the initial z-spin up wave packet $|\uparrow_z\rangle_e|0\rangle_e$ *will split symmetrically* into two x-spin-flavored wave packets. The x-spin-up-flavored wave packet $|\uparrow_x\rangle_e|\text{up}_x\rangle_e$ will end up in region up_x and the x-spin-down-flavored wave packet $|\downarrow_x\rangle_e|\text{down}_x\rangle_e$ will end up in region down_x (as in Figure 11.3).

Suppose that the electron starts located in the *top half* of the initial z-spin up wave packet as illustrated in Figure 11.4. As that wave packet moves toward the magnetic field, it will carry the electron to the right. Since the wave packets have uniform probability density, when the x-spin up and x-spin down wave packets begin to split, the electron will initially feel no probability currents in the up or down direction. It will just continue moving directly to the right. But as the x-spin packets move apart, the electron will eventually only feel the probability current from one of the two packets. When this happens the motion of the electron will be fully determined by the motion of that wave packet.

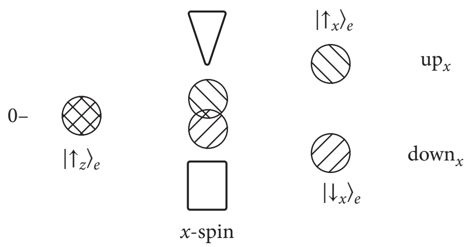


Figure 11.3. How a z-spin wave function evolves in an x-spin device.

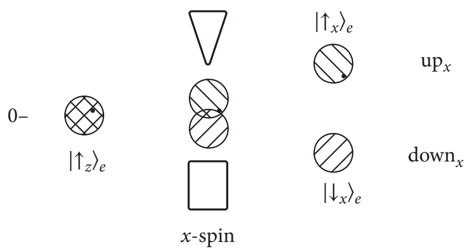


Figure 11.4. How an (initially) z-spin electron moves in an x-spin device.

Because of the symmetry of the wave packets and how they come apart in the x-spin device, the probability current in the x-direction on the plane that is perpendicular to the x-direction and through the middle of the initial wave packet will always be zero. An electron that starts in the *top half* of the initial z-spin up wave packet will end up in the x-spin up wave packet and will be deflected toward region up_x. And it will continue to act like an x-spin up electron as long as it remains associated with the x-spin-up-flavored wave packet. In this case, the electron's *effective wave function* will be $|\uparrow_x\rangle_e|\text{up}_x\rangle_e$. Similarly, if an electron started in the *bottom half* of the initial z-spin up wave packet, it would end up in the x-spin down wave packet and be deflected toward region down_x. And it would continue to act like an x-spin down electron as long as it remained associated with the x-spin-down-flavored wave packet. In this case, the electron's effective wave function will be $|\downarrow_x\rangle_e|\text{down}_x\rangle_e$.

The evolution of the electron's wave function and position are fully deterministic. The theory predicts the standard quantum probabilities as epistemic probabilities. We do not know where the electron will end up since we do not know where it started. By the distribution postulate, the probability of the electron initially being in the *top half* and the probability of it being in the *bottom half* of the initial z-spin wave packet are each 1/2, so the probability of it ending up x-spin up and the probability of it ending up x-spin down are also each 1/2. This is why a z-spin up electron has an equal probability of ending up x-spin up or x-spin down in an x-spin device. The story for an electron initially associated with a z-spin down wave packet proceeds along precisely the same lines.

On the idealized assumptions we are making, an electron that started precisely in the center of the initial z-spin wave packet would end up stranded in a region of zero wave function support since it would never feel any probability currents in the x-direction. One direction of the eigenvalue–eigenstate link is routinely violated by the theory since particles always have exact determinate positions while their wave functions are rarely if ever precise eigenstates of position. This shows one way that the other direction might be violated.

Now consider what happens if we change the inhomogeneous magnetic field so that the x-spin up wave packet $|\uparrow_x\rangle_e|0\rangle_e$ is deflected toward region down_x and the x-spin down wave packet is deflected toward region up_x (as in Figure 11.5).[8] If the electron begins in the *top half* of the initial z-spin up wave packet, it will again be deflected toward region up_x, but now this means that it is associated with the x-spin *down* wave packet. Hence the electron will behave like an x-spin down electron. So for a given initial state of the electron, the details of how we set up the x-spin device determine its resulting x-spin. But, however one sets up the device, the epistemic probability of getting each of the two possible x-spin results is 1/2.

[8] This can be done either by flipping the field gradient or by exchanging the poles on the magnets.

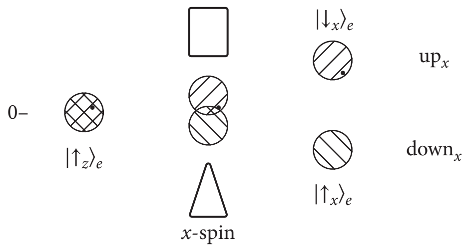


Figure 11.5. The contextuality of spin results.

Back to the original setup, let's consider more precisely how the electron moves in Figure 11.4. Since the two x-spin wave packets that make up the initial z-spin wave packet are separating in the x-direction before the electron itself starts to move in the x-direction, the relative position of the electron in its final wave packet is different from its relative position in its initial wave packet. More specifically, if the electron starts in the top half of the *top half* of its initial z-spin up wave packet, it will end up in the *top half* of its final x-spin up wave packet. But if it starts in the bottom half of the *top half* of its initial wave packet, it will end up in the *bottom half* of its final x-spin up wave packet. Similarly, if the electron starts in the top half of the *bottom half* of its initial wave packet, it will end up in the *top half* of its final x-spin down wave packet. And if it starts in the bottom half of the *bottom half* of its initial wave packet, it will end up in the *bottom half* of its final x-spin down wave packet.[9]

The electron in Figure 11.4 starts in the bottom half of the *top half* of its initial wave packet, so it ends up in the *bottom half* of its final x-spin up wave packet. This shift in the relative position of the electron in its wave packet is critically important to the probabilistic predictions of the theory. It is what explains why one cannot predict the motion of the electron in a *second* x-spin device *after* it passes through an intervening z-spin device.

Suppose we send a z-spin up electron through an x-spin device and suppose that the electron begins in the bottom half of the *top half* of its initial wave packet (as in Figure 11.6). From our side-view perspective, the electron will end up in the *bottom half* of the resulting x-spin up wave packet and will be deflected with this wave packet up to up$_x$. If we immediately send the electron through a second x-spin device, it would again be deflected in the x-spin up direction. Indeed, as long as it is associated with the x-spin up wave packet, it will act like an x-spin up electron.

[9] For more realistic wave packets where the probability density is neither uniform nor bounded, the electron's motion is more complicated, but the idea is the same. In that case, one would think of the electron's relative position in terms of its position relative to the total probability of the wave packet.

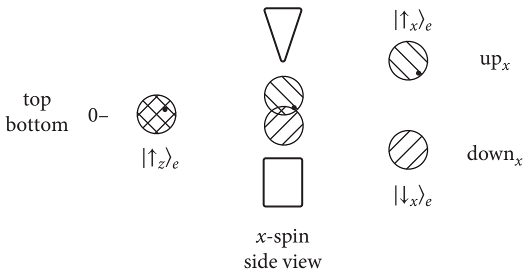


Figure 11.6. Alternating spin measurements (initial side view).

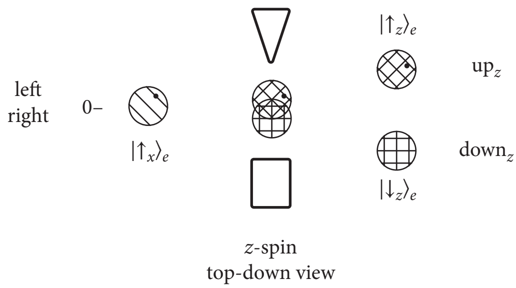


Figure 11.7. Alternating spin measurements (top-down view).

But suppose we send this x-spin up electron through a z-spin device before sending it through a second x-spin device (as in Figure 11.7). Since the magnetic field of the z-spin device is at a right angle to the magnetic field of the x-spin device, consider the motion of the wave packet and electron from the top-down perspective. The z-spin device splits the initial x-spin up wave packet into a z-spin up packet that is deflected left to region up$_z$ and a z-spin down packet that is deflected right to region down$_z$. What matters for the electron's motion here is whether it starts in the *left half* or *right half* of the initial x-spin up packet. If it starts in the *left half* (as in Figure 11.7), then it is carried by the z-spin up wave packet to region up$_z$.

Now suppose we send this z-spin up electron through a *second* x-spin device (as in Figure 11.8). Returning to the side view, the electron is still in the *bottom half* of the wave packet where it was left after its x-position was shifted by the first x-spin device. The second x-spin device will split the z-spin wave packet into an x-spin up wave packet and an x-spin down wave packet, but this time the electron will be deflected down because it is in the *bottom half* of the z-spin up packet—because that is where it was left after passing through the first x-spin device as it started in the bottom half of the *top half* of the initial z-spin wave packet at the very beginning of the experiment (as in Figure 11.6). By the distribution postulate,

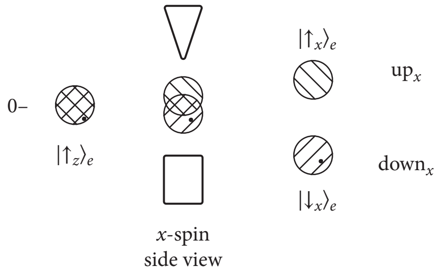


Figure 11.8. Alternating spin measurements (final side view).

the electron had an equal chance of starting in the top half of the *top half* or the bottom half of the *top half* of the initial wave packet by volume, so the probability that it will end up with each of the possible x-spins after the second x-spin device is 1/2. The shift in the position of the electron *relative to its wave packet* that was produced by the first x-spin device is why it behaves randomly in a second x-spin device *after* passing through an intervening z-spin device.

11.3 Interference and the Two-Path Experiment

Bohmian mechanics explains interference effects by predicting how interactions between wave packets in configuration space affect the motions of particles. Considering precisely how this works will also help to set up the discussion of what it means to *observe the position of a particle* in the theory.

Consider the two-path experiment with x-spin. Since there is only one particle, configuration space is ordinary three-dimensional space. We will describe the evolution of the electron's wave function (as in Figure 11.9). We will then consider what this means for the motion of the electron.

An initial x-spin up wave packet in the two-path experiment would evolve as follows:

$$\begin{gathered} |\uparrow_x\rangle_e|0\rangle_e \\ \downarrow \\ |\uparrow_x\rangle_e|A\rangle_e \\ \downarrow \\ |\uparrow_x\rangle_e|0'\rangle_e \end{gathered}$$

and the probability currents generated by the motion of the wave packet would carry an electron associated with it from region 0 along path A to region $0'$. Similarly, an initial x-spin down wave packet would evolve as follows:

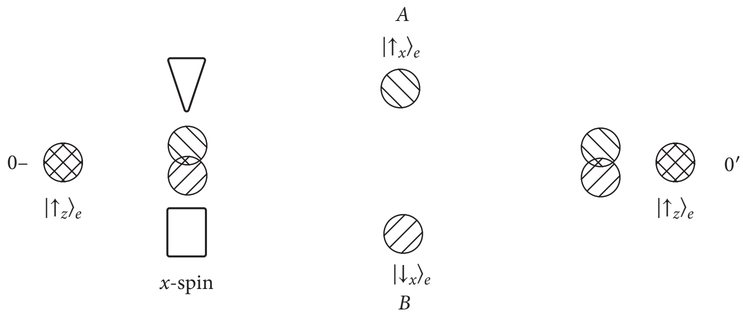


Figure 11.9. How the wave function evolves in the two-path experiment.

$$
\begin{gathered}
|\downarrow_x\rangle_e|0\rangle_e \\
\downarrow \\
|\downarrow_x\rangle_e|B\rangle_e \\
\downarrow \\
|\downarrow_x\rangle_e|0'\rangle_e,
\end{gathered}
$$

carrying an associated electron from region 0 along path B to region $0'$.

So, if the electron begins associated with a z-spin up wave packet, by the linearity of the dynamics, the wave function will evolve as follows:

$$
\begin{aligned}
&|\uparrow_z\rangle_e|0\rangle_e = \\
&\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|0\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|0\rangle_e \\
&\qquad\qquad\downarrow \\
&\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B_x\rangle_e \\
&\qquad\qquad\downarrow \\
&\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|0'\rangle_e + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|0'\rangle_e \\
&= |\uparrow_z\rangle_e|0'\rangle_e.
\end{aligned}
$$

Here the initial wave packet *splits* and the x-spin up packet travels path A to region $0'$ and the x-spin down packet travels path B to region $0'$. The trajectory of the electron depends on its position in the initial z-spin up packet.

If the electron is in the *top half* of the initial z-spin up wave packet (as in Figure 11.10), it will be carried by the x-spin up wave packet along path A, and it will behave like an x-spin up electron as long as it is associated with an x-spin-up-flavored wave packet. Meanwhile, the empty x-spin down wave packet travels

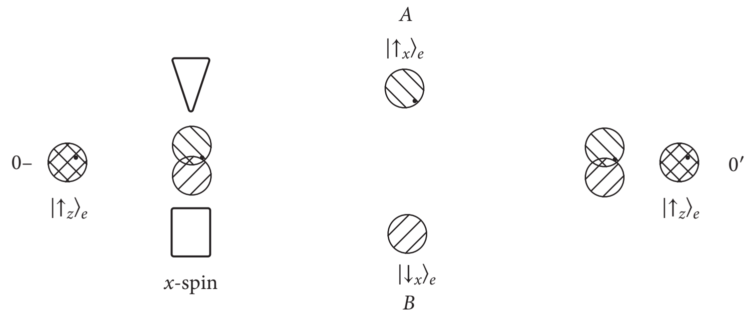


Figure 11.10. How the electron moves in the Bohmian two-path experiment.

path B to region 0′. If the two wave packets come together in the same symmetric way that they came apart, the electron will again be associated with a z-spin up wave packet and will thus act like a z-spin up particle. It will also go back to its initial relative position in its wave packet. And if the electron starts in the *bottom half* of the initial z-spin up wave packet it will travel path B, act like an x-spin down electron along the way, then act like a z-spin up electron again when the two wave packets come together in region 0′. This is how a z-spin up particle exhibits a determinate x-spin if one checks it while it is on path A or B but returns to acting z-spin up in region 0′.

This story is very different from the one we told in the context of the standard theory. It is also very different from the one that one would tell in Wigner's theory, GRW, or a decohering-worlds theory. In those theories, the electron is in a superposition of traveling path A and traveling path B and hence does not determinately travel either path. Here the electron determinately travels precisely one of the two paths and the wave function travels both. We know that the wave function travels both paths because of the interference effect that produces z-spin up dispositions in region 0′, what happens when we block one of the paths, and the effect of a total-of-nothing box on either path.

If a barrier gently blocks the path that the electron travels, then it will stop the wave packet that guides the electron and hence stop the electron. If one checks the spin of the blocked electron, one will find that it exhibits the x-spin property matching the wave-packet flavor associated with that path. A barrier on the path that the electron does not travel will stop the empty wave packet that takes that path. The electron and the wave packet that guides it will get to region 0′, but it will not meet up with the wave packet that was blocked on the other path, so it will continue to be associated with a wave packet with the determinate x-spin flavor matching the path the electron took and hence continue to act like an x-spin up or an x-spin down particle (as the case may be) when it gets to region 0′. The same thing would happen if one or the other of the two x-spin wave packets were slowed down or diverted or if one path were longer than the other or if anything

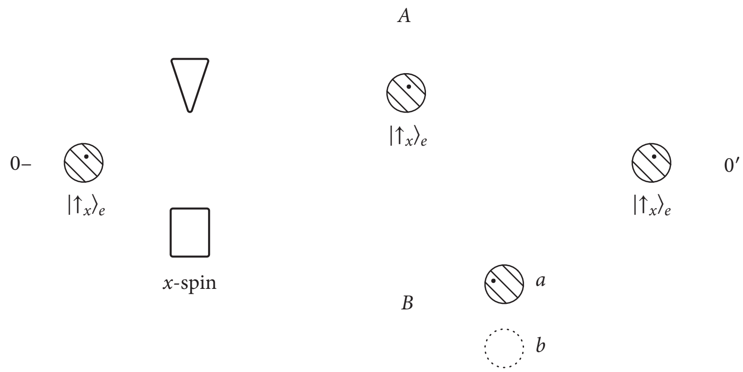


Figure 11.11. An x-spin up electron in the two-path experiment with a recording particle.

else happened to prevent the two packets from overlapping at the same time in region $0'$. One only gets the interference effect of having a z-spin up electron in region $0'$ if the two x-spin wave packets come together at the same place and time so that the electron is again associated with a z-spin-up-flavored wave packet.

Placing a total-of-nothing device on one of the two paths does not change the spin flavor of the wave packet that goes through it, but it changes the spin flavor of the final wave packet in region $0'$ from z-spin up to z-spin down when the two x-spin wave packets come together. Note that here, as in the basic two-path experiment, the empty wave packet only makes an empirical difference if it comes to overlap the actual particle configuration (here the position of the electron) and hence changes the effective wave function of the system (here the electron's effective wave packet). Understanding the role of the effective wave function is the key to understanding how measurement works in the theory.

11.4 Measurements and Records

In order to consider what it means to *measure the position* of the electron in the theory, we will introduce a second system to record the result of the measurement. A measuring device in Bohmian mechanics is just a physical system that produces a position record of a property of a system by correlating that property with the position of the record. In the simplest case, one might record the position of one particle in the position of another. But even with just two particles, one can no longer tell the dynamical story in ordinary three-dimensional space as both the composite system's wave function and configuration live in six-dimensional configuration space. Understanding how observation works in Bohmian mechanics requires a dynamical story in configuration space.

Suppose we set things up in the two-path experiment so that the path traveled by the electron e is recorded in the position of a particle p. To this end, we might, as earlier, put particle p in region a and arrange its interaction with e so that p moves from region a to region b if and only if e takes path B. Since particles only move in response to the changes in the wave function, one characterizes the interaction between the two particles by saying how the wave function of the composite system evolves.

Given the recording dispositions that we want for p, if e starts x-spin up (as in Figure 11.11), we want the composite system to evolve as follows:

$$\begin{gathered} |\uparrow_x\rangle_e|0\rangle_e|a\rangle_p \\ \downarrow \\ |\uparrow_x\rangle_e|A\rangle_e|a\rangle_p \\ \downarrow \\ |\uparrow_x\rangle_e|A\rangle_e|a\rangle_p \\ \downarrow \\ |\uparrow_x\rangle_e|0'\rangle_e|a\rangle_p. \end{gathered}$$

And, just slightly more interestingly, if e starts x-spin down (as in Figure 11.13), we want the composite system to evolve as follows:

$$\begin{gathered} |\downarrow_x\rangle_e|0\rangle_e|a\rangle_p \\ \downarrow \\ |\downarrow_x\rangle_e|B\rangle_e|a\rangle_p \\ \downarrow \\ |\downarrow_x\rangle_e|B\rangle_e|b\rangle_p \\ \downarrow \\ |\downarrow_x\rangle_e|0'\rangle_e|b\rangle_p. \end{gathered}$$

Consider how the two-particle configuration evolves, and hence how the ordinary positions of the two particles evolve, in each of these cases. The coordinates that matter most to the dynamical story are e's x-position and p's x-position, so those are the coordinates we will track in the figures. We will consider the two simple cases first.

Suppose that the two-particle configuration starts associated with the wave packet $|\uparrow_x\rangle_e|0\rangle_e|a\rangle_p$ (as in Figure 11.11). This puts the electron e starting in region 0 and particle p starting in region a. The interaction between the two particles can be seen in configuration space as illustrated in figure 11.12. As the wave packet evolves in configuration space, the probability current pushes the two-particle

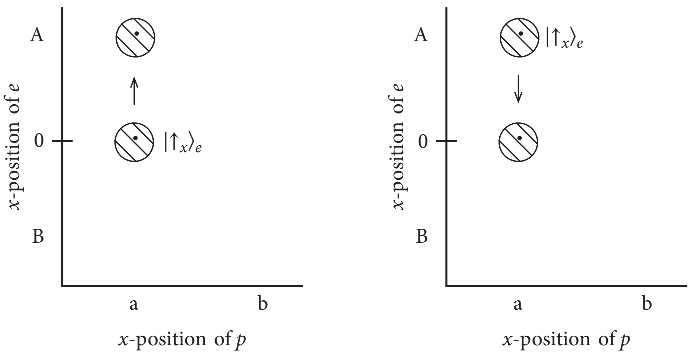


Figure 11.12. An x-spin up electron and recording particle in configuration space.

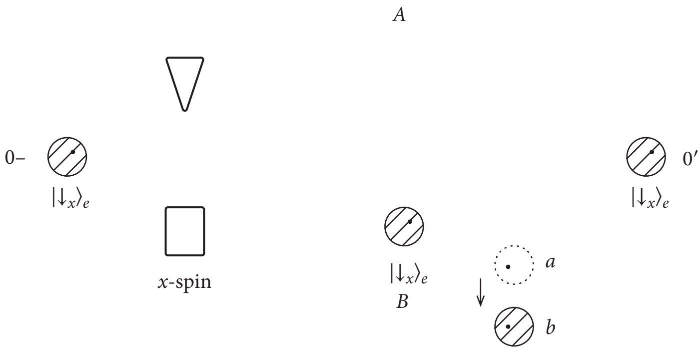


Figure 11.13. An x-spin down electron with a recording particle.

configuration in such a way that e moves from the center position in region 0 up along path A then back down to the center at region $0'$. Since there are no probability currents in particle p's x-direction (or any of its other coordinates) it simply stays where it is. In this case, the recording particle p does not move at all, reliably indicating that e took path A and is effectively x-spin up.

Suppose that the configuration starts associated with the wave packet $|{\downarrow_x}\rangle_e|0\rangle_e|a\rangle_p$ (as in Figure 11.13). Again, e starts in region 0 and p starts in region a. As the wave packet evolves in configuration space (as in Figure 11.14), the probability current pushes the two-particle configuration in such a way that e moves from region 0 down along path B, then p moves from region a to region b, then e moves back up to the center at region $0'$. In this case, the recording particle p moves from region a to region b, reliably indicating that e took path B and is effectively x-spin down.

Now consider what happens when the configuration starts associated with the wave packet $|{\uparrow_z}\rangle_e|0\rangle_e|a\rangle_p$ (as in Figure 11.15). By the linearity of the dynamics, the wave function of the composite system evolves in configuration space as follows:

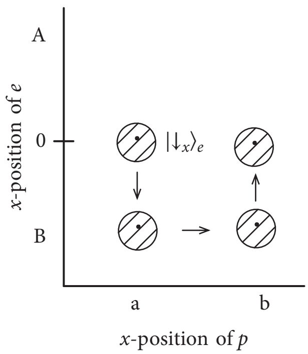


Figure 11.14. The x-spin down electron and recording particle in configuration space.

$$
\begin{gathered}
|\uparrow_z\rangle_e|0\rangle_e|a\rangle_p = \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|0\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|0\rangle_e|a\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e|a\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e|b\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|0'\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|0'\rangle_e|b\rangle_p.
\end{gathered}
$$

Note that, unlike the situation with just one particle, the final wave packet here is not z-spin up flavored.

Here the x-spin device splits the initial wave packet in configuration space. Consider two cases.

In Figure 11.15 the two-particle configuration begins in the *top half* of the configuration wave packet in e's x-position. On the auxiliary dynamics, the two-particle configuration will be carried along by the probability current in configuration space as if it were a massless particle. By symmetry considerations just like those in the single particle case, this means that the configuration will be picked up by the x-spin-up-flavored wave packet. And this means that e will take path A and will end up effectively x-spin up and p will not move from region a (as in Figure 11.16). Here p reliably records that e took path A.

Now suppose instead that the configuration begins in the *bottom half* of the wave packet in e's x-position (as in Figure 11.17). Here this means that the

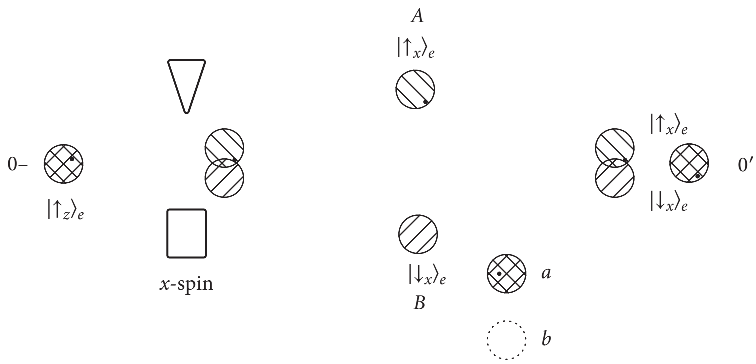


Figure 11.15. A *z*-spin up electron in the top half with a recording particle.

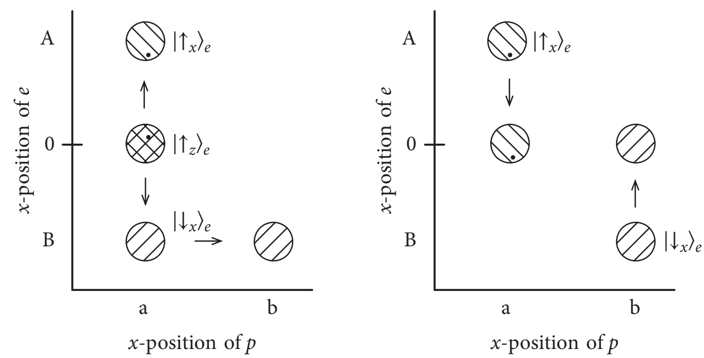


Figure 11.16. A *z*-spin up electron with a recording particle in configuration space.

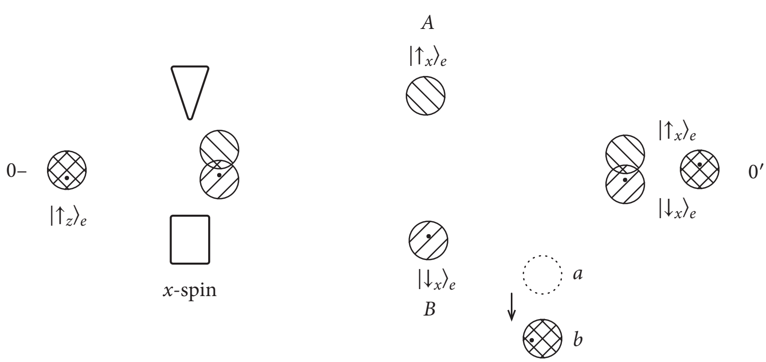


Figure 11.17. A *z*-spin up electron in the bottom half of its wave packet with a recording particle.

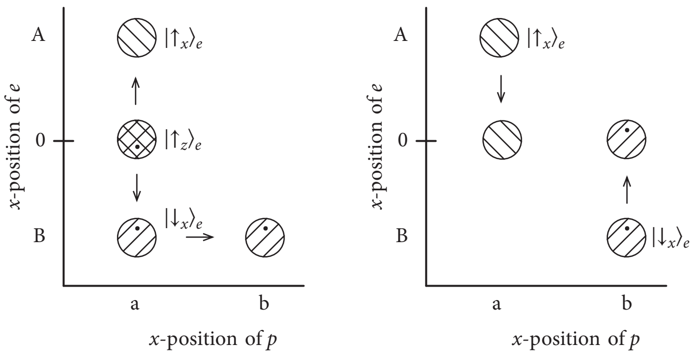


Figure 11.18. A z-spin up electron in the bottom half of its wave packet with a recording particle in configuration space.

configuration will be picked up by the x-spin-down-flavored wave packet, which means that particle e will take path B and will end up effectively x-spin down and particle p will move from region a to region b (as in Figure 11.18). Here p reliably records that e took path B.

Note that while the two x-spin wave packets end up in the same region of three-dimensional space, *they do not overlap in configuration space*. Hence the configuration remains associated with the x-spin-up-flavored wave packet in the first case (as in Figures 11.15 and 11.16) and with the x-spin-down-flavored wave packet in the second case (as in Figures 11.17 and 11.18). The position of particle p reliably records *both* the path taken by e in each case and that the electron ends up effectively x-spin up in the first case and effectively x-spin down in the second. More precisely, the record in the first case tells us that the *effective wave function* for the two particles is $|\uparrow_x\rangle_e|0'\rangle_e|a\rangle_p$ and the record in the second case tells us that the *effective wave function* for the two particles is $|\downarrow_x\rangle_e|0'\rangle_e|b\rangle_p$.

It is in this precise sense that the value of a measurement record in Bohmian mechanics indicates the branch of the wave function selected by the actual configuration. The *content* of the record is given by *the effective wave function of the composite system*. To be sure, the value of the record tells one something about the actual configuration that one did not already know, but it never tells one precisely where any particle is. If one knew anything more regarding the positions of particles than what is given by the effective wave function, then one would be able to make empirical predictions that violate the standard quantum probabilities.

11.5 Surreal Trajectories and Decoherence

The very existence of a position record can significantly affect the behavior of the measured system. Consider the same two-path experimental setup with the recording particle p but this time we will allow the wave packets to pass through

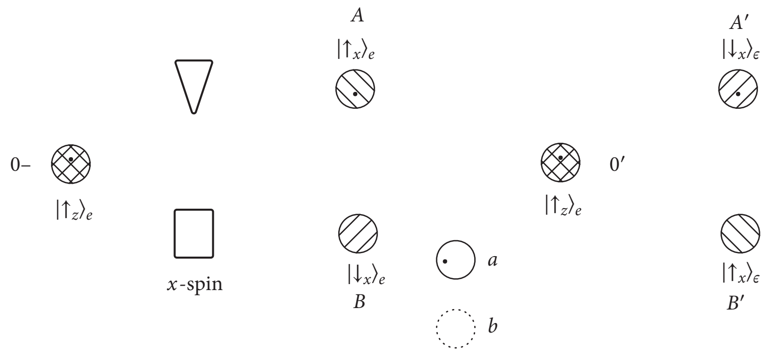


Figure 11.19. Electron trajectory if there is no interaction between *e* and *p*.

each other in region 0′ so that the x-spin down wave packet continues to region A' and the x-spin up wave packet continues to region B'. The electron starts associated with a z-spin up wave packet in region 0 and the recording particle starts in region a with positions as indicated in Figure 11.19.

First consider how the electron behaves if there is no interaction whatsoever between *e* and *p and hence no measurement record is made of the path traveled by the electron*. Since the two particles start in separable states and since they do not interact, we could just consider how *e*'s state evolves in ordinary three-dimensional space, but we will keep the two-particle configuration space representation for the purpose of comparison.

Here the wave function of the composite system evolves as follows:

$$
\begin{gathered}
|\uparrow_z\rangle_e|0\rangle_e|a\rangle_p = \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|0\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|0\rangle_e|a\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e|a\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e|a\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|0'\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|0'\rangle_e|a\rangle_p \\
= |\downarrow_z\rangle_e|0'\rangle_e|a\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|B'\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|A'\rangle_e|a\rangle_p.
\end{gathered}
$$

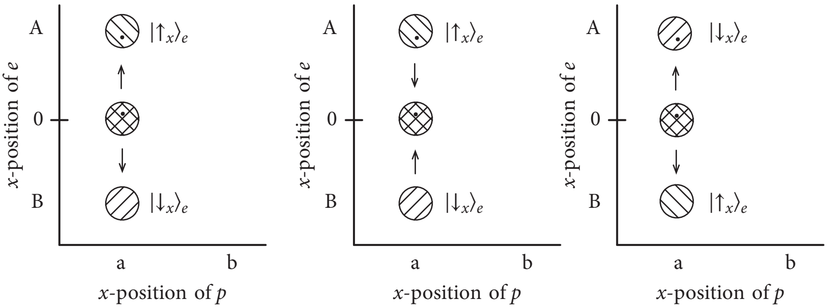


Figure 11.20. The electron's surreal trajectory in configuration space.

The initial wave packet splits, but since there is no correlation between *e*'s position and spin and *p*'s position, the state ends up separable in region $0'$.

Since there is no interaction, the recording particle *p* does not move at all and no record is made. In configuration space the evolution looks like Figure 11.20. The electron, however, does a number of curious things on this trajectory.

The electron starts in the bottom half of the *top half* of its wave packet, so when the wave packet splits, it ends up in the *bottom half* of the *x*-spin up wave packet and is carried along path *A* by the associated probability currents. But since there is no interaction with *p*, the two *x*-spin wave packets end up overlapping in configuration space when they reach region $0'$. This produces a *z*-spin up wave packet in region $0'$. And the probability current shifts *e*'s relative position in its wave packet back to where it started at the beginning of the experiment as the *x*-spin wave packets come together. The *x*-spin wave packets then *pass through each other* in region $0'$. Since *e* is in the *top half* of the wave packet again, it gets picked up by the wave packet headed toward region A' again, but this time it is the *x*-spin *down* packet. The electron is carried to region A' by the associated probability current and ends up effectively *x*-spin down.

Here the electron follows what is sometimes called a *surreal trajectory*.[10] It starts *z*-spin up in region 0, becomes *x*-spin up while passing through the magnets, travels along path *A*, becomes *z*-spin up again in region $0'$, bounces out of this *classically field-free* region, becomes *x*-spin down, and ends up in region A'. This bouncing behavior and the spontaneous change in spin in region $0'$, which

[10] Englert, Scully, Süssmann, and Walther (1992) argue that surreal trajectories pose an empirical problem for Bohmian mechanics in as much as such trajectories are never seen. But, as we will see, the theory itself explains why they are never seen. In short, they go away whenever one looks for them. A more recent line of argument against Bohmian mechanics is given by Frauchiger and Renner (2018). They use an extended version of the Wigner's friend story to argue that Bohmian mechanics violates the standard quantum predictions (something they call assumption Q). Their argument, however, involves a misunderstanding of how state preparation and measurement work in a no-collapse theory like Bohmian mechanics. In brief, as Lazarovici and Hubert (2018) show, Frauchiger and Renner do not keep track of the full quantum state of the systems involved in their story.

contains no classical fields whatsoever, violates both conservation of momentum and conservation of angular momentum. That we never see anything like this was once given as an argument against Bohmian mechanics.[11] But the theory itself explains why we never see such bounces.

To see the electron bounce, one would need to see that it was initially on path A *then* see that it ends up in region A'. But if one sees it on path A, then the theory predicts that the electron *does not bounce*. It is the measurement record of it being on path A that keeps it from bouncing.

Consider the experiment again, but this time we will allow p to record the position of e. In this case, the wave function of the composite system evolves as follows:

$$
\begin{gathered}
|\uparrow_z\rangle_e|0\rangle_e|a\rangle_p = \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|0\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|0\rangle_e|a\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e|a\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|A\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|B\rangle_e|b\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|0'\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|0'\rangle_e|b\rangle_p \\
\downarrow \\
\frac{1}{\sqrt{2}}|\uparrow_x\rangle_e|B'\rangle_e|a\rangle_p + \frac{1}{\sqrt{2}}|\downarrow_x\rangle_e|A'\rangle_e|b\rangle_p.
\end{gathered}
$$

Because of the correlation between e's position and spin and p's position, this dynamical story involves *both* particles in an essential way, so it must be told in configuration space (as in Figure 11.21).

Since the two-particle configuration starts in the *top half* of the wave packet in e's x-direction, it becomes associated with the x-spin up wave packet when the z-spin wave packet splits in configuration space (as in Figure 11.21). The two-particle configuration is then carried by the probability current in configuration space in such a way that the electron travels along path A and p stays in region a. But this time when the electron enters region $0'$, the two wave packets do not

[11] Englert et al. (1992) thought that this was a serious problem for Bohmian mechanics since we never in fact see such trajectories. Bohmian mechanics, however, predicts that one would never see such trajectories since measurement records in position prevent such behavior.

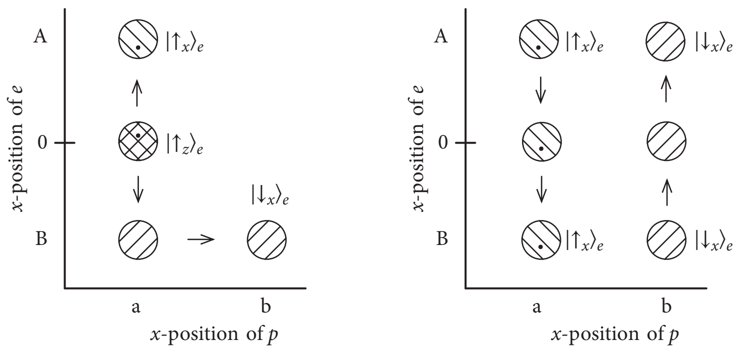


Figure 11.21. The effect of a measurement record in configuration space.

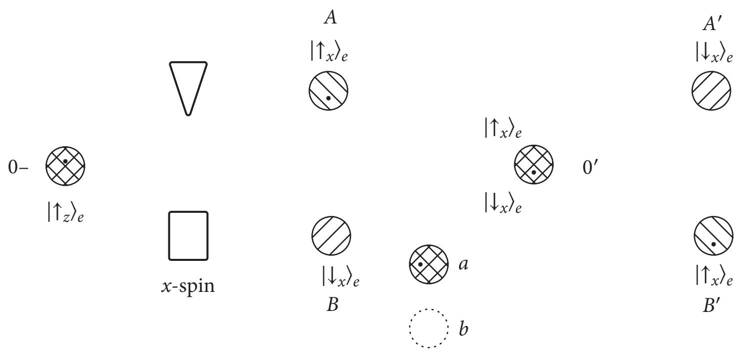


Figure 11.22. The effect of a measurement record in ordinary space.

overlap in configuration space and hence do not produce a z-spin up wave packet. This is because the interaction where p records e's position shifts the x-spin down wave packet in configuration space in p's x-direction so that the two wave packets miss each other. But, since they miss each other, the two-particle configuration only feels the probability current from the x-spin up wave packet and is carried to region B' in e's x-direction, which means that e moves through region $0'$ *without bouncing* and ends up effectively in region B' and still effectively x-spin up. And the electron will continue to behave like an x-spin up particle as long as the configuration selects an effective wave function that is x-spin-up-flavored for e.

In ordinary 3-space this trajectory looks like Figure 11.22.[12] Note that even though p does not move in this case, the fact that e took path A and is hence effectively x-spin up is recorded by p being at location a because the interaction between the two particles correlates p's position with both e's position and x-spin as represented by the wave function of the composite system. The two-particle

[12] Contrast this with the surreal trajectory (in Figure 11.19) when there is no interaction between e and p.

configuration then selects an effective x-spin for e and a *corresponding* record for p. Of course, if the electron begins in the *bottom half* of the initial wave packet, it would take path B and p *would* move to region b indicating that e is in fact on path B.

In either case, *the very existence of the measurement record* prevents the electron from bouncing out of the field-free region $0'$. This is why we never see surreal trajectories. While momentum is not conserved in Bohmian mechanics, *observed momentum* is. The existence of the record also provides the electron with *stable* effective spin properties. Here the electron stays x-spin up even when passing through region $0'$.

In order to get interference effects, the wave packets and the two-particle configuration must overlap at the same time *in configuration space*. The only way that this can happen here is if the measurement record is erased by carefully disentangling p's position from e's position *and* x-spin. This would undo the shift in the x-spin down wave packet in p's x-direction. Then one might put the two x-spin wave packets back together in e's position and get back to a z-spin up wave packet and the associated z-spin up interference behavior for e. The degrees of freedom involved here make this difficult to accomplish.

That position records stabilize the physical properties of the object system and lead to classical behavior by destroying interference effects provides an important role for environmental decoherence in Bohmian mechanics. Insofar as interactions between a system and its environment produce (unintentional) position records of the system's properties, they stabilize the properties of the system and promote its classical behavior. Since the properties of a macroscopic system that we think of as classical are typically recorded in the positions of systems in its environment, such systems and properties will indeed behave classically in Bohmian mechanics.

11.6 How the Theory Explains Experience

We can now say what one sees when one observes the position of a particle in Bohmian mechanics. Suppose that the recording particle p in the two-path experiment moves to region b. This does not tell us precisely where the electron is. It might be anywhere in the wave packet that traveled path B.[13] Rather, it tells us *which wave packet* the configuration (and hence the electron) is associated with. That is, the empirical content of the record, what one can deduce from the value of the record, is given by *the effective wave function selected by the current*

[13] Indeed, it could be anywhere where this wave packet provides positive wave function support. And, for more realistic wave packets, that means that it could be virtually anywhere. That said, for a well-designed experiment the electron is probably on (or very nearly on) path B.

particle configuration. This is what an observer has epistemic access to given her measurement record. In this precise sense, this is what she *sees*.

If Bohmian mechanics correctly describes the physical world and if one ever had direct epistemic access to the positions of particles, one would be able to make predictions more precisely than allowed by the standard quantum probabilities. The very empirical adequacy of the theory depends on the fact that one never directly observes the precise position of anything.

Particle positions do not typically by themselves determine the content of a measurement record. Given the contextual nature of measurement, the same particle positions might represent different measurement outcomes by selecting different-flavored wave packets. And the wave function does not typically determine the content of a measurement record by itself either. The wave function typically fails to specify any particular outcome. It is the particle configuration and the wave function *together* that determine the value of the record. The effective wave function selected by the particular configuration both determines the empirical content of the record and tracks the dispositional properties of the measured system. For a formulation of quantum mechanics to explain our experience, it needs to characterize something that exhibits the standard quantum statistics and on which one might plausibly take one's experience to supervene. In Bohmian mechanics this is the effective wave function selected by the current particle configuration.

11.7 EPR and Relativity

Bohmian mechanics is manifestly incompatible with the constraints of relativity. The dynamics makes essential use of configuration space to explain the standard quantum statistics by means of the correlated motions of distant particles. The very idea of configuration space is incompatible with relativity. The point in configuration space $Q(t)$ represents the positions of all the particles in a system at time t no matter how far apart they may be. To make any sense of this at all one would need to choose a preferred inertial frame, which would violate a basic principle of relativity.

Consider how the theory explains the results of an EPR experiment.[14] Suppose one has two particles e_a and e_b in the EPR state

$$\frac{1}{\sqrt{2}}(|\uparrow_x\rangle_{e_a}|\downarrow_x\rangle_{e_b} + |\downarrow_x\rangle_{e_a}|\uparrow_x\rangle_{e_b}) \tag{11.1}$$

where particle e_a is located at A on Earth with one observer and particle e_b is located at B in orbit about α Centauri with another. The observer at A will measure

[14] This story follows Albert (1992, 156–60).

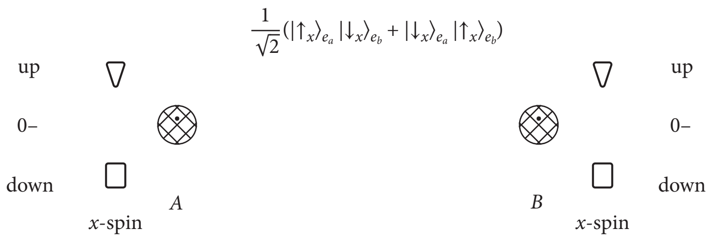


Figure 11.23. EPR setup.

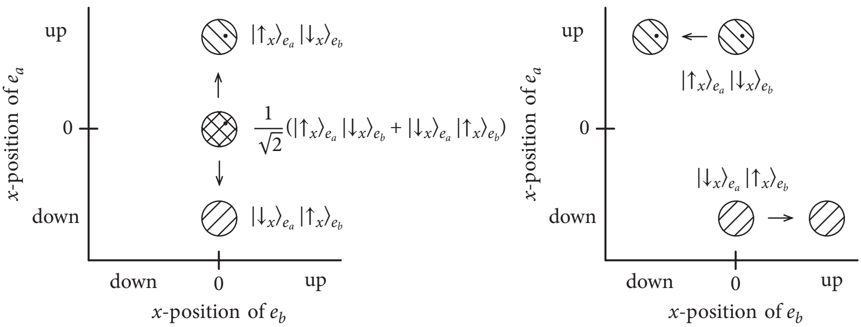


Figure 11.24. Trajectory in configuration space if *A*'s measurement is first.

the x-spin of e_a at noon 1 January 2050 (in the laboratory inertial frame) and the observer at *B* will measure the x-spin measurement of e_b at noon 1 January 2050 plus one second (in the laboratory inertial frame), each on their respective particle. The experimental setup is illustrated in Figure 11.23.

Since these two measurement events are spacelike separated, there is no matter of fact about which is first. There are inertial frames where observer *A*'s measurement is first (as in the laboratory frame) and there are others where *B* performs her measurement first.

Since the spin properties of the two particles are entangled, we have no choice but to tell the dynamical story in configuration space. The initial state of the two particles is given by an EPR-flavored wave packet in six-dimensional configuration space and a point in that configuration space representing the positions of the two particles. Suppose that the two-particle configuration is such that each particle is in the *top half* of the wave packet in its x-direction (as in Figures 11.23 and 11.24). Again, we will just track the x-positions of the two particles in configuration space.

Consider an inertial frame where *A*'s measurement occurs first (as in Figure 11.24). When she performs her measurement, the linear dynamics requires the initial EPR-flavored wave packet to split in configuration space in e_a's x-direction. Given the usual way of setting up the magnets so that x-spin up wave packets are deflected up and x-spin down wave packets are deflected down, the

$|\uparrow_x\rangle_{e_a}|\downarrow_x\rangle_{e_b}$-flavored wave packet is deflected up in e_a's x-direction, and the $|\downarrow_x\rangle_{e_a}|\uparrow_x\rangle_{e_b}$-flavored wave packet is deflected down in e_a's x-direction. Since the two-particle configuration starts in the *top half* of the EPR-flavored wave packet in e_a's x-direction, the configuration ends up associated with the $|\uparrow_x\rangle_{e_a}|\downarrow_x\rangle_{e_b}$ wave packet, which means that e_a is deflected up. When B measures the x-spin of e_b, the $|\uparrow_x\rangle_{e_a}|\downarrow_x\rangle_{e_b}$ wave packet is deflected down in e_b's x-direction, and the $|\downarrow_x\rangle_{e_a}|\uparrow_x\rangle_{e_b}$ wave packet is deflected up in e_b's x-direction. Since the configuration is associated with the $|\uparrow_x\rangle_{e_a}|\downarrow_x\rangle_{e_b}$ wave packet, e_b is deflected down. In this case, the configuration selects an effective wave function such that e_a ends up effectively an x-spin up electron and e_b ends up effectively an x-spin down electron. That the two electrons exhibit opposite x-spin is good since spin is anti-correlated in the EPR state.

Now consider an inertial frame where B's measurement occurs first (as in Figure 11.25). Here the linear dynamics requires the initial EPR-flavored wave packet to split in configuration space in e_b's x-direction. So the $|\uparrow_x\rangle_{e_a}|\downarrow_x\rangle_{e_b}$-flavored wave packet is deflected down in e_b's x-direction, and the $|\downarrow_x\rangle_{e_a}|\uparrow_x\rangle_{e_b}$-flavored wave packet is deflected up in e_b's x-direction. Since the two-particle configuration starts in the *top half* of the wave packet in e_b's x-direction, this time it ends up associated with the $|\downarrow_x\rangle_{e_a}|\uparrow_x\rangle_{e_b}$ wave packet, so this time e_b is deflected *up*. When A measures the x-spin of e_a, the $|\uparrow_x\rangle_{e_a}|\downarrow_x\rangle_{e_b}$ wave packet is deflected up in e_a's x-direction, and the $|\downarrow_x\rangle_{e_a}|\uparrow_x\rangle_{e_b}$ wave packet is deflected down in e_a's x-direction. But now since the configuration is associated with the $|\downarrow_x\rangle_{e_a}|\uparrow_x\rangle_{e_b}$ wave packet, e_a is deflected *down*. In this case, the configuration selects an effective wave function such that e_a ends up effectively an x-spin down electron and e_b ends up effectively an x-spin up electron. Again, it is good that the x-spins are opposite. But it is not good that the theory predicts *different x-spin results* depending on the temporal order of the measurements at A and B.

Since there is no matter of fact regarding the temporal order of spacelike separated events in special relativity, there can be no physical facts that depend on the

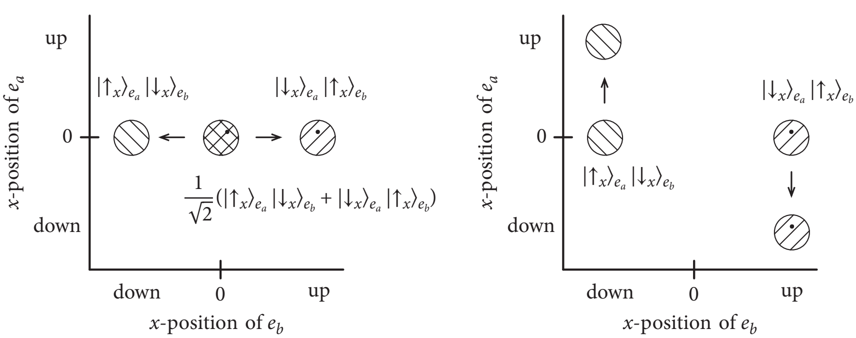


Figure 11.25. Trajectory in configuration space if B's measurement is first.

temporal order of such events. But the outcomes of EPR measurements *depend on the temporal order of the spacelike separated measurements*. So Bohmian mechanics is dynamically incompatible with relativity. That the dynamical incompatibility between Bohmian mechanics and relativity is not subtle is illustrated by the fact that one would be able to send superluminal messages if one knew the initial particle configuration in the present experiment.

Suppose that observer A performs her measurement first. If she knows that she is making the first measurement and if she knows that the two-particle configuration is in the *top half* of the wave packet in e_a's x-direction, then she would know that e_a will deflect in the up direction no matter how she orients the inhomogeneous magnetic field on her device. This means that (in the standard orientation) her measurement will associate the configuration with the $|\uparrow_x\rangle_{e_a}|\downarrow_x\rangle_{e_b}$ wave packet. This would make e_a effectively x-spin up and e_b effectively x-spin down. So when observer B gets around to checking her particle (in the standard orientation), it will be deflected down. But by changing the orientation of her field, in just the way we considered when we discussed contextual properties earlier, observer A can arrange things so that the $|\downarrow_x\rangle_{e_a}|\uparrow_x\rangle_{e_b}$ wave packet is the one that is deflected in the up direction at her end of the experiment. The particle e_a will again be deflected up, but now e_a is effectively x-spin down and e_b is effectively x-spin up. So when observer B gets around to checking her particle, it will be deflected *up*.

The upshot is that observer A can instantaneously control the motion of particle e_b no matter how far away it is if she knows the relative position of e_a in the initial EPR-flavored wave packet. If the distribution postulate is satisfied, then she does not know the relative position of e_a in its wave packet. That said, she is still instantaneously affecting e_b's motion by the field orientation she chooses. It is just that she does not know how.

That Bohmian mechanics is nonlocal in a way that renders it flatly incompatible with relativity is an essential feature of the theory. It uses such nonlocal interactions to get the right anti-correlation statistics here and the right EPR–Bell statistics more generally.

11.8 Virtues and Vices

Bohmian mechanics is a no-collapse formulation of quantum mechanics that makes the standard forward-looking probabilistic predictions for the positions of particles. And since it treats measurement interactions precisely the same way that it treats other physical interactions, it resolves the quantum measurement problem by providing a consistent account of nested measurements.

In the context of the Wigner's friend experiment, Bohmian mechanics predicts that the friend, her measuring device, and the particle she is measuring will all evolve according to the linear dynamics. The configuration of the composite system will select an effective wave function that represents a perfectly determinate x-spin result for the friend. Because this is an effective wave function for the entire composite system and since the friend's brain state is correlated with the pointer position which is correlated to the x-spin of the object system, all of the parts of the composite system will behave in a way that is consistent with the result. And if an external observer were ever able to make an A-measurement of the composite system, he would get the result $+1$ with probability one.[15] Inasmuch as the wave function always evolves linearly, Bohmian mechanics is in principle empirically distinguishable from a collapse formulation of quantum mechanics like Wigner's theory or GRW.

One might worry that Bohmian mechanics, unlike pure wave mechanics, commits one to a *particle* ontology. But the Bohmian *approach* can be applied to almost any choice of fundamental physical ontology, making the general hidden-variable approach a sort of framework theory. We will see a concrete example of how this works when we discuss Bell's Bohmian field theory in the next chapter. The problem is that while the Bohmian approach can be used to provide a quantum-mechanical dynamics for any physical observable, we don't know what observable it *should* be talking about.[16] It is in this sense that it encounters a version of the preferred-basis problem.

Proponents of Bohmian mechanics often argue that since particles with ordinary three-dimensional positions constitute a primitive ontology, they are precisely the sort of thing that one needs to provide a satisfactory account for experience. But, as we have seen, measuring the position of something in Bohmian mechanics is more subtle than directly seeing its location in ordinary three-dimensional space. Indeed, the theory allows for a plausible argument that the appearance of ordinary things in ordinary three-dimensional space is best understood as a sort of illusion generated by the way the $3N$-dimensional configuration selects an effective component of the wave function in $3N$-dimensional configuration space. Regardless of whether one calls it an illusion, inasmuch as the content of an observer's experience is given by the effective wave function selected by the total configuration of her measuring system, the primitive ontology of particles with ordinary three-dimensional positions is not what explains the observer's experience.

Rather than being directly observable things that single-handedly constitute our manifest image of the world, the particles in Bohmian mechanics play the role of

[15] This is the configuration that would get all of the probability when one correlates the pointer position on an A-measuring device with the A-property of the composite system *FMS*.

[16] As we will see in the next chapter, the problem is not that one must choose *particle position* to be a privileged physical quantity to make a Bohm-type theory work. The problem, rather, is that one must choose *some privileged physical quantity to be determinate* and that quantity must, by dint of its being determinate, account for *all* of our physical records. Not only do we not know what quantity to choose, but any just right choice for explaining our experience is bound to look ad hoc given the other choices one might have made.

selecting a single branch of the quantum-mechanical state as the effective wave function that is in fact realized on measurement. The content of one's measurement record is given by the effective wave function, which, in turn, provides something on which one's experience might plausibly supervene.[17]

One other connection before moving on. As mentioned in the discussion of the many-threads formulation (Chapter 10), Bohmian mechanics may be thought of as a sort of non-splitting, many-worlds theory. Specifically, to get a Bohmian many-threads theory, fix the wave function for the world at some initial time t_0, then consider a different non-splitting physical world corresponding to each possible initial particle configuration with a history given by how that configuration would evolve under the Bohmian deterministic dynamics.[18] One might take the theory to consist of this set of possible world histories and the measure over this set given by the probability that the distribution postulate assigns to the initial configuration that characterizes each world. These probabilities are the prior probabilities that each possible world is the world we in fact inhabit. One updates by conditioning on what one learns of the actual world. This picture captures the idea that Bohmian mechanics might be thought of as pure wave mechanics with a new variable, the total configuration, that selects the actual branch at each time and hence determines a complete history.

Arguably, the most significant problem with Bohmian mechanics is its manifest dynamical incompatibility with relativity. The theory relies on strongly nonlocal interactions that violate relativistic constraints to explain things like the EPR–Bell statistics. Bell thought that one might lessen the worry over the theory's incompatibility with relativity somewhat by considering a Bohmian field theory. We now turn to that.

[17] The suggestion here is in contrast to the view that particle configuration is superfluous in Bohmian mechanics and that the theory is hence no better off than pure wave mechanics. See Brown and Wallace (2005) for such an argument. By selecting a *single* effective wave function, the Bohmian particle configuration does something that pure wave mechanics cannot—it provides *precisely one physical record* as a possible target for experiential supervenience.

[18] See section 10.6 for a characterization of the many-threads theory.

12

Empirical Ontology and Explanation

12.1 The Explanatory Work of Metaphysics

The quantum measurement problem encourages us to consider what it is that we should expect from a satisfactory physical theory. We like to think of our best physical theories as descriptive of the physical world in some strong sense, but none of the formulations of quantum mechanics we have considered stands out as the one that clearly has things right. While even our best physical theories are always provisional, the situation here seems significantly worse.

The alternative formulations of quantum mechanics we have considered often suggest radically different metaphysical commitments, even when they are formulations of the same basic theory. GRW comes in wave function (GRWr), mass density (GRWm), and flash (GRWf) versions. Pure wave mechanics may be reformulated in terms of a single world with illusions of determinate records, a single world with relative facts, splitting worlds, decohering worlds, a single world with one mind for each observer, a single world with many minds for each observer, many threads, many maps, or physical hidden variables. Some formulations of quantum mechanics are stochastic, others are deterministic. Some are purely physical, others require one to embrace a strong form of mind–body dualism.

But quantum mechanics failing to provide anything like a canonical metaphysics does not mean that metaphysics doesn't matter. As we have seen repeatedly, the metaphysical commitments one associates with a particular formulation plays an essential role in explaining what measurement records are and why they can be expected to exhibit the standard quantum statistics.[1]

Since the same basic theoretical framework will often provide different explanations given different metaphysical commitments, insofar as one individuates theories by the explanations they provide, it makes sense to take alternative metaphysical interpretations of a basic theoretical framework as representing

[1] See Lewis (2016) for a more general discussion of the ways that one's metaphysical commitments, may matter to one's understanding of quantum mechanics and quantum mechanics may matter to one's metaphysical commitments, including issues such as causation, part–whole relations, and freewill. See also Maudlin (2019) for a series of inventive arguments involving the metaphysical implications of quantum mechanics. Some of these contrast with the arguments we have considered over the last few chapters and hence provide a broader perspective on the metaphysical options one has for understanding quantum mechanics.

alternative physical theories.[2] The thought is that one is only clear about the descriptive content of one's theory when one is clear about how it explains our experience, and one is only clear about *that* in the context of a specific set of metaphysical commitments.

Explanations that appeal to metaphysical commitments often take the form of descriptive stories. Such stories may say why an event occurs, or why it should be expected, or how it is physically possible. Richer explanatory stories may characterize mechanisms, describe how events are caused, or account for the existence of emergent entities or structures of a particular sort. It is in this way that our metaphysical commitments are in service of our best understanding of the physical world and the explanations that go with it.

In the standard collapse theory and Wigner's theory, probabilities are the result of fundamental chance events. But in theories like Bohmian mechanics, where there are no chance events, probabilities are purely epistemic. While they are alternative formulations of the same basic theory, GRWr, GRWm, and GRWf describe quite different physical worlds and, hence, explain the *object* of one's experience differently. Wigner's commitment to a strong mind–body dualism is essential to his account of how observers end up with determinate measurement records and why their records exhibit the standard quantum statistics. Albert and Loewer's many-minds and single-minds theories also involve a commitment to a strong variety of mind–body dualism, but since minds do not cause physical events, these theories provide a very different account of determinate records and quantum probabilities. While the single- and many-minds theories are no-collapse theories, the sort of dualism they embrace makes it particularly easy to say what future-looking self-location probabilities might consist in. Worlds split on the splitting-worlds and decohering-worlds theories and they do not split on the many-threads theory, which, again, matters for the sort of quantum probabilities one can make sense of on the two approaches. The emergent nature of worlds and that there is no simple matter of fact about how many worlds there are on the decohering-worlds theory means that the account of determinate experience provided by the theory is necessarily associated with a level of description. This is not the case in the splitting-worlds theory, where one chooses a physically preferred basis once and for all. In Bohmian mechanics, what it means to have a determinate measurement record and why such records can be expected to be

[2] Along these lines, one might be tempted to insist that two physical theories are equivalent if and only if they provide the same explanations. But this does not settle the general issue of theoretical equivalence without agreement concerning what it is for two theories to provide the same explanations. Indeed, this strategy pushes, I think, in the wrong direction. I do not take there to be plausible necessary and sufficient conditions for theoretical equivalence. Rather, it seems, questions of theoretical equivalence often turn on whether alternative explanations are the same or different *given the explanatory purposes at hand*. And here metaphysical talk can play the role both of expressing the explanatory purposes at hand (one might, for example, seek a causal account of a particular type) and fulfilling those purposes (in this case, providing a causal narrative of the sort desired).

distributed according to the standard quantum statistics is a result of the special role played by position in the theory. But, back to where we started, there is nothing at all special about position in the standard collapse theory or Wigner's refinement of it.

This quick list of philosophical refections is just a reminder that the metaphysical commitments one associates with a physical theory are essential to the explanations the theory provides. This is why meaningful arguments over the metaphysical commitments of a particular formulation of quantum mechanics are arguments over alternative varieties of physical explanation. Since none of the formulations of quantum mechanics that have been proposed so far satisfy all the explanatory demands one might reasonably have for a satisfactory physical theory, dogmatic adherence to any particular formulation with its associated metaphysical commitments makes little sense. But, again, that does not mean that there is nothing at stake in the choice—there are real trade-offs in how alternative formulations of quantum mechanics characterize the physical world and, hence, in how they account for our experience. With this in mind, we will consider one last example of ontology and explanation.

12.2 Beables and Experience

John Bell was an ardent proponent of Bohmian mechanics, which went hand in hand with his sense that the proper path to explaining experience is by getting the "positions of things" right. But this does not mean that he was entirely satisfied with the standard Bohmian metaphysics of particles in motion.

Part of Bell's worry concerned the compatibility of Bohmian mechanics with relativistic constraints. As he put the problem, "When the cogency of Bohm's reasoning is admitted, a final protest is often this: it is all nonrelativistic" (1987, 171). Bell believed that this protest was best addressed by showing how to construct a Bohmian field theory. He thought that, while it would still be incompatible with relativity, such a theory would show that one could nevertheless do real physics of the sort that one might have thought could only be done in a truly relativistic theory. But he also wanted a theory that ultimately involved a metaphysics of things with positions, since he felt that would best secure a satisfactory account of experience. This explanatory commitment to a particular sort of metaphysics influenced how Bell set about constructing his field theory.

The *local beable* that Bell chose to make determinate was *fermion number density* at each point on a spatial lattice. The always-determinate fermion number density here is analogous to the always-determinate particle configuration in Bohmian mechanics—it is the metaphysics of fermion density in spacetime regions that is supposed to explain why measurements yield determinate outcomes. As Bell explained it:

> The beables of the theory are those elements which might correspond to elements of reality, to things which exist. Their existence does not depend on observation. Indeed observation and observers must be made out of beables.
>
> (1987, 174)

The *local beables* here are the *primitive ontology* of the theory. The thought is that one's experience supervenes on determinate measurement records which in turn supervene on the local state of the physical world which is composed of elements of the primitive ontology associated with the theory. If the world described by the theory predicts measurement records that exhibit the standard quantum statistics, then the theory accounts for our experience. One might take the account to be better or worse depending on the plausibility of our experience in fact supervening on such records.

On Bell's Bohmian field theory (1987, 174–7), $n(t)$ is the fully determinate fermion-number configuration at time t. It is a point in a fermion-number configuration space that assigns a field value to every point on the spatial lattice at time t. The linear dynamics describes the time-evolution of the quantum-mechanical part of the state

$$\frac{d}{dt}|t\rangle = -iH|t\rangle$$

where $|t\rangle$ is a wave function over fermion-number configuration space. The auxiliary dynamics then describes how the determinate local field values evolve as the wave function evolves. There is some flexibility in the choice of dynamics that will not make an empirical difference to the statistical predictions of the theory. Bell stipulated that the transition probability from field configuration m to field configuration n over time interval dt is dtT_{nm} where

$$\begin{aligned} T_{nm} &= J_{nm}/D_m, \\ J_{nm} &= \sum_{qp} 2\,\mathrm{Re}\,\langle t|nq\rangle\langle nq| - iH|mp\rangle\langle mp|t\rangle, \end{aligned}$$

and

$$D_m = \sum_q |\langle mq|t\rangle|^2$$

if $J_{nm} > 0$. Otherwise, $T_{nm} = 0$.

One can think of the dynamics for $n(t)$ as describing a random walk through a *field-configuration space*, where each point in the space represents a complete specification of the fermion number density at every point in the lattice representing ordinary three-dimensional space. The number of particles at every point in ordinary three-dimensional space then will always be determinate. Bell showed that if the epistemic probability over possible fermion-number configurations is ever equal to the standard quantum probabilities, then one should always expect it to be so on this dynamics. Among other things, this means that one will get the standard EPR–Bell statistics for nonlocally entangled fermion-number-density field states.

Bell expressed the potential worry that the "the very sharpness of this reformulation [of quantum field theory] brings into focus some awkward questions." First, he noted that the theory "does not inspire complete happiness" since "there is nothing unique about the choice of fermion number density as basic local beable . . . [w]e could have others instead, or in addition" (1987, 178). This is the preferred basis problem again. The problem in part concerns what metaphysical commitments will provide the most compelling account of experience. The worry is that one might choose the *wrong* physical quantity to make determinate.

In standard Bohmian mechanics one supposes that experience is explained by determinate particle positions. In Bell's Bohmian field theory one supposes that it is fermion number density that explains experience. Any argument for one being the one and only *canonical choice* of local beable serves to undermine one's arguments for the other. And there are an infinite number of other options that Bell might have considered.

This fact is worth a moment of reflection. One can provide a Bohm–Bell type dynamics for most any physical quantity Q that one wants to be always determinate.[3] Jereon Vink (1993) gave perhaps the best illustration of how to do this. Vink's approach was based on Bell's, but his aim was different. Vink's idea was to make *all* physical quantities determinate then to use the generalized Bohm–Bell–Vink dynamics to describe their evolution. But it is a consequence of the Kochen–Specker theorem (1967) that one cannot make every physical quantity determinate in such a way that their possessed values both exhibit the standard quantum statistics and satisfy the standard functional relationships between the quantities—like that the value of *position-squared* is the square of the value of *position*.[4] The upshot is that if one makes all physical quantities determinate, then it is unclear what these determinate quantities are supposed to represent. On the other hand, one might argue, if one chooses to make only one privileged physical quantity determinate, then one has already lost the functional relationships that provide the usual understanding of that quantity.[5] And, more salient to the present

[3] Along these lines, one might take Q-theory as a Bohmian schema theory where Q is the physical parameter that describes one's primitive ontology and hence explains the determinate values of one's local physical records.

[4] See chapter 11 of Earman (1986) for a discussion of Kochen–Specker and related theorems. That one cannot make all physical quantities simultaneously determinate while preserving their functional relationships clearly poses a significant problem for anyone like EPR who takes physical quantities to be real (as described earlier, in Chapter 6). As Earman argues, it is unclear that one has a choice between this sort of naive realism and locality inasmuch as there is reason to doubt that one can entirely satisfy either classical intuition.

[5] See Barrett (2005a) for a discussion of Vink's proposal and the virtues and costs to implementing it.

example, if one does feel justified in choosing just one always-determinate beable to make determinate, what should it be?

While Bell chose to make a field variable determinate in order to get a theory that might be made compatible with relativity at some level, he also knew what *kind of account of experience* he wanted when he chose fermion number density as the single, always-determinate beable. Given his commitment to the sort of explanation Bohmian mechanics gives for experience, he held that "[w]hat is essential is to be able to define the positions of things, including the positions of instrument pointers or (the modern equivalent) of ink on computer output" (1987, 175). He settled on stipulating an always-determinate fermion number density as one way to accomplish this. While Bell recognized a number of options, he reported that "I do not see how this choice can be made experimentally significant, so long as the final results of experiments are defined so grossly as by the positions of instrument pointers, or of ink on paper" (1987, 178). That is, insofar as fermion density number can be expected to account for the determinate positions of things like pointer positions and ink on paper, it explains our experience.

Bell's choice of local beable was arguably a good one for explaining experience, but it is clearly not *canonical*—there are, as he clearly noted, others that would serve as well. For his part, Bell's approach to the problem was pragmatic. He argued for determinate particle position being the right choice in Bohmian mechanics and for determinate fermion number density being the right choice for his field theory, and for flashes being the right choice in GRW.[6] He wanted something that (1) provided the strongest variety of compatibly one could get with relativistic considerations and (2) would plausibly give macroscopic objects effectively determinate positions, banking on his sense that determinate positions are enough to explain all of our experience.

Whether or not one shares Bell's guiding intuition, it is clearly not the case that any choice of beable will provide a compelling account of experience. Taking the total angular momentum of all of the cows in Oberwinkl (a small town near Salzburg) as the one and only always-determinate beable would presumably do little to explain our experience generally. The plausibility of one's account of experience depends on the details of one's empirical ontology. And, as we saw in the context of Bohmian mechanics, the precise way that it depends on one's empirical ontology can be subtle.

Bell chose *fermion number density* as determinate because he thought that this would make the *positions of things* (i.e. pointers and ink on paper) determinate and hence explain our determinate experience. But, just as in Bohmian mechanics, one cannot know more about the fermion number density than allowed by the standard quantum probabilities given the current wave function. This means that

[6] For Bell's views regarding the flash ontology in GRW, see Bell (1987, 181–95) and Allori (2013, 67–8).

one does not *directly see* the fermion number densities here any more than one directly sees the precise positions of particles in the basic theory. Rather, by an analogous argument, the *empirical content* of a good measurement record in Bell's field theory is given by the effective wave function selected by the current fermion number density. We will return to this in a moment.

There is, of course, something curious about the very idea of *choosing* what physical quantity to make determinate. It is presumably a matter of physical fact, not explanatory convenience, whether particle position or fermion number density or some other physical property is in fact metaphysically determinate. If possible, one would like to know what the real ontology of the world is, how it behaves, precisely how perception works, and hence what really determines our experience. This presumably has little to do with choosing an empirical ontology that satisfies our metaphysical intuitions.

The second awkward thing Bell noted was that while his theory agrees with standard quantum field theory empirically, the dynamics relies on the choice of a preferred inertial frame and is, hence, incompatible with relativity.[7] On his dynamics, spacelike separated fermion number densities evolve simultaneously just like spacelike separated particle positions do on standard Bohmian mechanics. Regarding this problem, Bell reported that "[I] am unable to prove, or even formulate clearly, the proposition that a sharp formulation of quantum field theory, such as that set out here, must disrespect serious Lorentz invariance. But it seems to me that this is probably so" (1987, 179). The problem again is that Bohmian field theory, just like standard Bohmian mechanics, explains the EPR–Bell statistical correlations in a way that is dynamically incompatible with relativistic constraints.[8]

12.3 Metaphysics and Empirical Adequacy

The ontology one associates with a theory allows one to say what counts as a measurement record in the theory. How compelling one's explanation of experience is depends on the plausibility that our experience in fact supervenes on records so constituted.

[7] Just like Bohmian mechanics, for example, Bohmian field theory does not allow for superluminal *signaling* but does allow for superluminal *effects*. Again, it is only because one does not know the actual initial field configuration that one cannot send signals.

[8] For a selection of approaches to Bohmian field theory and tentative suggestions for relativistic formulations of Bohmian mechanics see Dürr, Goldstein, Tumulka, and Zanghì (2005); Dürr, Goldstein, Norsen, Struyve, and Zanghì (2013); Vink (2017); and Deckert, Esfeld, and Oldofredi (2018). While not relativistic in a standard sense, these proposals illustrate how one might incorporate relativistic phenomena like particle creation and annihilation into a hidden-variable theory. That the EPR–Bell correlations can be explained in a way that is compatible with relativistic constraints in GRWf is an arguable advantage over Bohmian mechanics. While it lacks a dynamics, the many-maps formulation is the closest that I know to something like this for Bohmian mechanics.

We want a formulation of quantum mechanics that we can take as empirically adequate. That means that the associated empirical ontology needs to be one on which we can take our actual experience to supervene. To do this, the theory needs to predict determinate records that satisfy the standard quantum statistics. Given that we often intuitively think of our physical records as supervening on the positions and motions of nearby objects, one might suppose that the empirical ontology must be a primitive ontology to be satisfactory, but empirical adequacy comes in different types and degrees depending on how one's theory characterizes records and on precisely how one's experience is taken to supervene on them.

One might take a physical theory to be *weakly adequate* if one can understand it as describing a world that contains *anything whatsoever* on which one's experience might be taken to supervene; and more *strongly adequate* if one can take it to describe a world that contains something on which one's experience might *plausibly supervene* given one's background commitments concerning what physical facts are actually relevant to experience. The distinction here is not canonical. Rather, it is a matter of the *sense* and the *degree* to which the theory allows one to plausibly model how an observer might end up with the measurement records she takes herself to have, given our background assumptions about how we experience the world (which may come from cognitive science, physiology, common sense, etc.).

Suppose that a physical theory predicts the existence of a book in an underground vault in New York City that accurately describes my full experience. Such a theory would automatically be *weakly adequate* no matter what else it predicted since it provides *something* on which my experience might be taken to supervene. But if the content of the book is the only thing on which my experience might supervene, then the theory is *only* weakly adequate. Given my background commitments concerning what in fact determines my experience, it is implausible that it is determined by descriptions in a distant underground vault no matter how detailed and accurate they may be. Indeed, given my background beliefs, it is unclear to me how I might even have epistemic access to the content of the book.

As a limiting case, a theory that described a physical world with such a book might be taken to be weakly adequate inasmuch as it would contain something on which my actual experience might supervene (the descriptions in the book) *even if the world contains nothing but the book*. But given what I in fact take as relevant to my experience, a world containing a book of my experiences and nothing else is entirely unacceptable as a world that really accounts for my experience. Again, one wants one's physical theory to describe a world on which one's measurement records, or more directly experience, might plausibly be taken to supervene *given one's background beliefs about the relationship between mental and physical states*. Of course, one's background beliefs might ultimately require some tuning, but there are also limits to plausible explanation.

The general issue here concerns what it should mean for a physical theory to be empirically adequate. The suggestion is that there is no canonical set of necessary and sufficient conditions for what should count as an adequate account of experience. Rather, what one requires for empirical adequacy depends on the explanatory trade-offs presented by the theory one is considering. On this view, the empirical adequacy of a theory is a matter of cost–benefit analysis and compromise.

The upshot is that a weakly adequate theory might allow for impressive predictions but provide entirely unsatisfactory explanations of experience. This is worth a bit more detail. For strong adequacy it is not enough that there merely be something on which one's experience might supervene. One wants a theory and associated ontology that predicts measurement records that can plausibly be taken to determine the empirical content of one's experience given those facts that we take to be relevant to experience. In quantum mechanics this can be expected to involve the theory telling us how measurement interactions produce records on which the empirical content of our experience might supervene given one's background assumptions concerning how experience works. So, among other things, one needs to consider what the theory itself allows one to know. Being able to model a *situated observer* in one's theory is arguably a precondition for the possibility of providing a plausible explanation for how the observer has epistemic access to measurement outcomes.[9]

One can *guarantee* a direct account of determinate experience by choosing the right *empirical ontology*. In the single-mind theory, Albert and Loewer simply stipulate that an observer's mental state is always determinate then randomly associate it with a branch of the wave function with the standard quantum probabilities.[10] But this is where cost–benefit considerations kick in—it is the very effectiveness of this proposal that marks it as ad hoc. One gets a direct account of determinate experience but only at the cost of violating general explanatory commitments regarding how our experience in fact supervenes on physical states.

Back to the vault example, it is not impossible that the book determines my experience. Rather, it is *implausible* that it does given my background commitments about what physical facts might be relevant to my experience. Given

[9] This is closely related to the notion of empirical coherence discussed earlier. Note that a purely epistemic, pragmatic, or information-theoretic formulation of quantum mechanics, one that just provides a predictive algorithm and does not even aim to describe physically situated observers, does not face the threat of empirical incoherence. But this cannot be considered a virtue when it is only because the theory is not descriptively strong enough to even make sense of the threat. Such theories are never strongly empirically adequate. Indeed, even what we are calling a weakly adequate theory arguably exhibits a stronger sort of empirical adequacy than a predictive algorithm. The former at least seeks to describe a physical world that would explain our experience. See Ismael (2007) for a discussion of the notion of a situated observer. Here our explanatory aims are significantly more modest.

[10] Note that this provides an immediate account of experience without there being a primitive ontology in the usual three-dimensional object sense.

everyday, pre-theoretic notions about how perception works, one might want a formulation of quantum mechanics that predicts determinate properties for the objects in my visual field or, perhaps more directly, a sequence of brain states that can be understood as representing a determinate sequence of perceptions that one would take to agree with one's experience.

Proponents of alternative formulations of quantum mechanics often differ in their background commitments regarding the explanatory demands a satisfactory account of experience should satisfy. We ultimately want to account for both quantum phenomena and our everyday experience of a world that appears to be constituted by three-dimensional macroscopic objects with determinate classical properties. But, as we have seen, there are many ways that one might go about providing such an account and not all of these involve a *primitive* ontology.

Bell thought that making the positions of things determinate, then having those positions exhibit the standard quantum statistics provided the most plausible explanation of experience. This is the idea behind a primitive ontology. One insists that an empirically adequate physical theory should account for our experience of a world that appears to be constituted by three-dimensional macroscopic objects with definite properties in terms of an ontology consisting of basic entities in motion in ordinary three-dimensional space or spacetime that provides the manifest image of the world.[11] The strategy involves two thoughts. The first is that our experience of the physical world is properly characterized in terms of the apparent motion of ordinary three-dimensional macroscopic objects, and the second is that the apparent motion of ordinary three-dimensional macroscopic objects is properly taken to supervene on the positions of more basic entities in ordinary three-dimensional space. The basic entities of a primitive ontology are usually taken to be something like particles with always-determinate positions (as in Bohmian mechanics), flashes at spacetime points (as in GRWf), or fermion densities at spacetime points (as in Bell's field theory). While flashes and fermion densities already require one to go beyond the requirement that a primitive ontology must consist in objects in motion, they still involve spatially situated things and, hence, might be taken to be close enough to the intuitive mark.

There is something intuitively attractive in having a theory with a primitive ontology. Such a picture is familiar from classical mechanics. But how this works in a formulation of quantum mechanics can be subtle and not particularly well-connected with the intuitions that led one to favor one's choice of primitive ontology in the first place.[12]

[11] See Allori (2013).

[12] And just because a strategy is intuitive and familiar does not mean that it is the only option. Another salient option here is Albert's (2013) proposal of taking configuration space to be basic and accounting for three-dimensional space and its objects as emergent entities.

12.4 Empirical Ontology and Experience

Consider again how one gets determinate measurement records from the positions of things and what these records mean in Bohmian mechanics. The theory tells us that one never sees precisely where a particle, or anything else, is. Rather, in the two-path experiment, the record "*a*" indicates that the particle configuration has selected $|\uparrow_x\rangle_e|A\rangle_e|a\rangle_p$ as the effective wave function. Inasmuch as our experience supervenes on such records, it supervenes on the effective wave function selected by the $3N$-dimensional particle configuration. That the effective wave function is what one most directly sees in Bohmian mechanics is flatly incompatible with the intuition that the theory provides a satisfactory account of experience *because* of its primitive ontology. Indeed, that the theory's ontology is primitive is arguably irrelevant to its account of experience.

Because particle configuration satisfies the standard quantum statistics in Bohmian mechanics, so does the effective wave function it selects. Since there is *something* on which our experience might supervene, Bohmian mechanics is at least weakly empirically adequate. Further, given that our experience supervenes on the effective wave function, the theory provides a plausible account for how an observer might have *precisely the limited epistemic access to the positions of things described by the standard quantum statistics*. The degree to which the theory is *strongly adequate* depends on how plausible one finds it that our experience in fact supervenes on such records.

While position plays an important role in how Bohmian mechanics accounts for our experience, the theory also provides a dynamical explanation for how position records make correlated physical properties *effectively determinate*. When one looks for the electron in the two-path experiment, the record "*a*" indicates that the configuration has selected $|\uparrow_x\rangle_e|A\rangle_e|a\rangle_p$ as the effective wave function. This means that p is effectively in region a, e is effectively on path A, and e is effectively x-spin up. So if, contrary to our pre-theoretic intuitions, our most immediately accessible perceptual records are in terms of x-spin we are still fine as long as the position of something gets correlated to x-spin over the course of a typical measurement. Since pretty much everything would get correlated to the position of something in a typical measurement of a macroscopic system, particle position is a good tentative candidate for one's empirical ontology. And, for the same reason, so is fermion density.

Providing *something* on which our experience might supervene is like getting the descriptions in the vault right. But Bohmian mechanics does significantly more than that. It explains *how* and *the extent to which* we have epistemic access to the positions of the objects we measure. It also explains how properties other than position might be *effectively determinate*. Such *dynamical stories* provide a stronger, richer variety of empirical adequacy than a theory that just provides something on which one's experience might be taken to supervene.[13]

The sense and degree to which a theory and its associated empirical ontology might be taken to be strongly empirically adequate depends on how well the theory models our epistemic access to the world, all things considered. Strong and weak empirical adequacy differ in the plausibility and explanatory richness of the account of experience provided by the theory. We want the strongest variety of empirical adequacy we can get given our background theoretical and explanatory commitments.

12.5 Philosophical Morals

In order to be empirically adequate, quantum mechanics must be counterintuitive. This is reflected in the fact that each formulation of quantum mechanics that we have considered violates basic common-sense and philosophical intuitions. There are two consequences. One is methodological, the other is epistemological.

The methodological consequence is one cannot trust one's common-sense or philosophical intuitions to get the world right. While our pre-theoretic intuitions guide our evaluation of alternative formulations of quantum mechanics, the physical world is ultimately counterintuitive. One might evaluate one's theoretical options with the aim of deciding which is most attractive on balance. But since alternative formulations of quantum mechanics exhibit both conceptual virtues and vices, our intuitions, even if they could be trusted to track the truth, which they clearly cannot, will never deliver an unambiguous winner.

This brings us to the epistemological moral. All of the formulations of quantum mechanics that we have considered are somehow descriptive of the physical world. If nothing else, each predicts the quantum phenomena we have observed so far that fall within the descriptive domain of the particular formulation. But the striking differences between how each describes the physical world show how little this means. While there is likely some sense in which each formulation is *approximately true*, the striking differences between how they describe the physical world illustrate just how little we know regarding the *sense in which* they might be approximately true.[14] Given our epistemic situation, the best strategy is to keep all of the serious theoretical options on the table while we aim for something better.

[13] One does the same thing in GRWr, GRWf, or GRWm when one shows how the dynamics predicts wave functions, flashes, or mass distributions that exhibit appropriately correlated measurement records that capture the standard quantum statistics and hence allow for successful action.

[14] See Barrett (2003) (2008) for discussions of what it might mean to take a physical theory like quantum mechanics to be approximately true.

With respect to the proper choice of empirical ontology and debates concerning such things as primitive ontology and configuration space realism more specifically, there is arguably not much to be gained by trying to stipulate an intuitively preferred metaphysics up front. Rather, given our epistemic situation, one might more profitably adopt a flexible view of the matter without trying to stipulate how our experience ought to supervene on the physical world once and for all.

But a pragmatic approach does not mean that anything goes. One wants a theory that allows for a rich and compelling account of experience and hence can be taken as strongly adequate. But strong adequacy is a matter of degree and always depends on one's background theoretical and explanatory commitments. One might consequently expect our judgments concerning the degree to which a theory is strongly adequate to be influenced by what we learn as we try to formulate more compelling theories.

Given the difficulty of the task at hand, one might take a formulation of quantum mechanics to be a significant achievement if it is logically consistent and at least weakly empirically adequate over nonrelativistic phenomena. We have considered a number of ways to get at least *that*. What we should want now is a formulation of quantum mechanics that we can clearly understand as being strongly adequate over relativistic phenomena.

APPENDIX A

A Formal Characterization of Hilbert Space

A Hilbert space is a separable vector space with an inner product satisfying the following axioms:[1]

1. The sum of any two vectors $|\phi\rangle$ and $|\chi\rangle$ in $\mathcal{H}$ is also a vector $|\phi\rangle + |\chi\rangle$ in $\mathcal{H}$. Vector addition satisfies the following properties:
   a. $|\phi\rangle + |\chi\rangle = |\chi\rangle + |\phi\rangle$.
   b. $(|\phi\rangle + |\chi\rangle) + |\psi\rangle = |\phi\rangle + (|\chi\rangle + |\psi\rangle)$.
   c. There is a zero vector $|0\rangle$ such that $|\phi\rangle + |0\rangle = |\phi\rangle$.
   d. For every vector $|\phi\rangle$ in $\mathcal{H}$, there is an inverse $-|\phi\rangle$ in $\mathcal{H}$ such that $|\phi\rangle + (-|\phi\rangle) = |0\rangle$.
2. A vector $|\phi\rangle$ in $\mathcal{H}$ multiplied by a complex number α is also a vector $\alpha|\phi\rangle$ in $\mathcal{H}$. Scalar multiplication has the following properties:
   a. $\alpha(\beta|\phi\rangle) = (\alpha\beta)|\phi\rangle$.
   b. $1|\phi\rangle = |\phi\rangle$.
   c. $\alpha(|\phi\rangle + |\chi\rangle) = \alpha|\phi\rangle + \alpha|\chi\rangle$.
   d. $(\alpha + \beta)|\phi\rangle = \alpha|\phi\rangle + \beta|\phi\rangle$.
3. For vectors in $\mathcal{H}$, there is an inner product $\langle\phi|\chi\rangle$ with the following properties:
   a. $\langle\phi|(|\chi\rangle + |\psi\rangle) = \langle\phi|\chi\rangle + \langle\phi|\psi\rangle$.
   b. $(\alpha\langle\phi|)|\psi\rangle = \alpha\langle\phi|\psi\rangle$.
   c. $\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^*$, where $*$ represents the complex conjugate.
   d. $\langle\phi|\phi\rangle \geq 0$, and $\langle\phi|\phi\rangle = 0$ only if $|\phi\rangle = 0$.
4. $\mathcal{H}$ is continuous and its dimension is at most countably infinite.

A complex number has the form $a + bi$ where a and b are real numbers and $i = \sqrt{-1}$. The complex conjugate of $a + bi$ is $(a + bi)^* = a - bi$.[2] The last axiom above characterizes two topological properties of a Hilbert space. The space is *separable* if and only if its dimension is at most countably infinite. If the space is separable, then there is a natural way to represent each vector as a linear combination of the elements of an orthonormal basis. This allows one to calculate quantum-mechanical probabilities in the usual way.

On the standard formulation of quantum mechanics, the unit-length elements of Hilbert space represent physical states. Linear operators on the Hilbert space represent observable physical properties and dynamical laws. In particular, Hermitian operators represent observable properties, and the linear dynamics is represented by a family of linear operators.

[1] An early characterization of the Hilbert space representation of quantum mechanical states is found in von Neumann (1932, 1955).

[2] *Wolfram MathWorld* is an online resource for basic mathematical notions. It contains short articles on most of the mathematical notions we discuss here. As a start, a brief overview of complex numbers can be found at <http://mathworld.wolfram.com/ComplexNumber.html>.

Bibliography

Aharonov, Y. and D. Bohm (1959). "Significance of Electromagnetic Potentials in Quantum Theory." *Physical Review* 115: 485–91.

Aharonov, Y. and D. Z. Albert (1981). "Can We Make Sense out of the Measurement Process in Relativistic Quantum Mechanics?" *Physical Review D* 24: 359–70.

Albert, David, Z (2015). *After Physics*. Cambridge, MA: Harvard University Press.

Albert, David, Z (2013). "Physics and Narrative," in *Reading Putnam*, Maria Baghramian (ed.) (2013). New York: Routledge. 225–36.

Albert, David, Z (2013). "Wave Function Realism," in Ney and Albert (eds) (2013) 52–7.

Albert, David, Z (1992). *Quantum Mechanics and Experience*. Cambridge, MA: Harvard University Press.

Albert, David, Z (1986). "How to Take a Photograph of Another Everett World." *Annals of the New York Academy of Sciences: New Techniques and Ideas in Quantum Measurement Theory* 480: 498–502.

Albert, David, Z and Jeffrey A. Barrett (1995). "On What It Takes to Be a World." *Topoi* 14(1): 35–7.

Albert, David, Z and Barry Loewer (1988). "Interpreting the Many Worlds Interpretation." *Synthese* 77, 195–213.

Allori, Valia (2013). "Primitive Ontology and the Structure of Fundamental Physical Theories," in Ney and Albert (2013) 58–75.

Allori, Valia, Sheldon Goldstein, Roderich Tumulka, and Nino Zanghì (2012). "Predictions and Primitive Ontology in Quantum Foundations: A Study of Examples" (version 2). arXiv.org, https://arxiv.org/abs/1206.0019. Accessed 7 September 2017.

Bacciagaluppi, G. and A. Valentini (2009). "Quantum Theory at the Crossroads: Reconsidering the 1927 Solvay Conference" (version 2). Online at https://arxiv.org/abs/quant-ph/0609184. Accessed 6 September 2017.

Barrett, Jeffrey A. (2018). "Everett's Relative-State Formulation of Quantum Mechanics." *The Stanford Encyclopedia of Philosophy* (Winter 2018 edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/entries/qm-everett/.

Barrett, Jeffrey A. (2017). "Typical Worlds." *Studies in the History and Philosophy of Modern Physics* 58: 31–40.

Barrett, Jeffrey A. (2015). "Pure Wave Mechanics and the Very Idea of Empirical Adequacy." *Synthese* 192(10): 3071–104.

Barrett, Jeffrey A. (2014). "Entanglement and Disentanglement in Relativistic Quantum Mechanics." *Studies in History and Philosophy of Modern Physics* 48(B): 168–74.

Barrett, Jeffrey A. (2011a). "On the Faithful Interpretation of Pure Wave Mechanics." *British Journal for the Philosophy of Science* 62(4): 693–709.

Barrett, Jeffrey A. (2011b). "Everett's Pure Wave Mechanics and the Notion of Worlds." *European Journal for Philosophy of Science* 1(2): 277–302.

Barrett, Jeffrey A. (2008). "Approximate Truth and Descriptive Nesting." *Erkenntnis* 68: 213–24.

Barrett, Jeffrey A. (2005a). "The Preferred Basis Problem and the Quantum Mechanics of Everything." *British Journal for the Philosophy of Science* 56(2): 199–220.

Barrett, Jeffrey A. (2005b). "Relativistic Quantum Mechanics Through Frame-Dependent Constructions." *Philosophy of Science* 72: 802–13.

Barrett, Jeffrey A. (2003). "Are Our Best Physical Theories Probably and/or Approximately True?" *Philosophy of Science* 70: 1206–18.

Barrett, Jeffrey A. (2000). "The Persistence of Memory: Surreal Trajectories in Bohm's Theory." *Philosophy of Science* 67(4): 680–703.

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

Barrett, Jeffrey A. (1996). "Empirical Adequacy and the Availability of Reliable Records in Quantum Mechanics." *Philosophy of Science* 63(1): 49–64.

Barrett, Jeffrey A. and Peter Byrne (eds) (2012). *The Everett Interpretation of Quantum Mechanics: Collected Works 1955–1980 with Commentary*, Princeton, NJ: Princeton University Press.

Barrett, Jeffrey A. and Simon Huttegger (2019). "Quantum Randomness and Underdetermination" (forthcoming in *Philosophy of Science*).

Baym, Gordon (1969). *Lectures on Quantum Mechanics*. Menlo Park, CA: Benjamin/ Cummings.

Becker, Adam (2018). *What is Real: The Unfinished Quest for the Meaning of Quantum Physics*. New York: Basic Books.

Bell, J. S. (1987). *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.

Benatti, Fabio, GianCarlo Ghirardi, and Renata Grassi (1995). "Describing the Macroscopic World: Closing the Circle within the Dynamical Reduction Program." *Foundations of Physics* 25: 5–38.

Bohm, David (1952). "A Suggested Interpretation of Quantum Theory in Terms of 'Hidden Variables,'" Parts I and II. *Physical Review* 85: 166–79, 180–93.

Born, M. (1926). "Zur Quantenmechanik der Stovorgänge." *Zeitschrift für Physik* 37: 863–7.

Bricmont, J., D. Dürr, M. C. Galavotti, G. Ghiradi, F. Petruccione, and N. Zanghì (eds) (2001). *Chance in Physics: Foundations and Perspectives*. Lecture Notes in Physics. Berlin: Springer.

Brown, H. R. and Wallace, D. (2005). "Solving the Measurement Problem: de Broglie-Bohm Loses Out to Everett." *Foundations of Physics* 35: 517–40.

Bub, J. (2015). "The Measurement Problem from the Perspective of an Information-Theoretic Interpretation of Quantum Mechanics." *Entropy* 17: 7374–86.

Bub, J. (1997). *Interpreting the Quantum World*. Cambridge: Cambridge University Press.

Byrne, Peter (2010). *The Many Worlds of Hugh Everett III: Multiple Universes, Mutual Assured Destruction, and the Meltdown of a Nuclear Family*. Oxford: Oxford University Press.

Callender, Craig (2007). "The Emergence and Interpretation of Probability in Bohmian Mechanics." *Studies in History and Philosophy of Modern Physics* 38: 351–70.

Chiribella, Giulio and Robert Spekkens (eds) (2016). *Quantum Theory: Informational Foundations and Foils*. Dordrecht: Springer.

Cushing, James T. (1994). *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*. Chicago: University of Chicago Press.

Deckert, Dirk-Andre, Michael Esfeld, and Andrea Oldofredi (2018). "A Persistent Particle Ontology for QFT in Terms of the Dirac Sea." *British Journal for the Philosophy of Physics*, forthcoming.

Descartes, Rene (1641). *Meditations on First Philosophy* in *The Philosophical Writings of Descartes: Volume 2* John Cottingham, Robert Stoothoff, Dugald Murdoch (translators) (1999). Cambridge: Cambridge University Press, 1–62.

Deutsch, David (1999). "Quantum Theory of Probability and Decisions." *Proceedings of the Royal Society of London* A455: 3129–37.

DeWitt, Bryce S. (1970). "Quantum Mechanics and Reality." *Physics Today* 23: 3035.

DeWitt, Bryce S. (1971). "The Many-Universes Interpretation of Quantum Mechanics," in B. d'Espagnat (ed.) *Foundations of Quantum Mechanics*. New York: Academic Press. Reprinted in DeWitt and Graham (1973) 167–218.

DeWitt, Bryce S. and R. Neill Graham (eds) (1973). *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton, NJ: Princeton University Press.

Dirac, P. A. M. (1930). *The Principles of Quantum Mechanics*. First edition. Oxford: Oxford University Press.

Dirac, P. A. M. (1957). *The Principles of Quantum Mechanics*. Fourth edition. Oxford: Oxford University Press.

Dürr, D., S. Goldstein, T. Norsen, W. Struyve, and N. Zanghì (2013). "Can Bohmian Mechanics be Made Relativistic?" *Proceedings of the Royal Society A* 470: 2162.

Dürr, D., S. Goldstein, R. Tumulka, and N. Zanghì, (2005). "Bell-Type Quantum Field Theories." *Journal of Physics A* 38(4): R1–R43.

Dürr, D., S. Goldstein, and N. Zanghì (1993). "A Global Equilibrium as the Foundation of Quantum Randomness." *Foundations of Physics* 23(5): 721–38.

Dürr, D., S. Goldstein, and N. Zanghì (1992). "Quantum Equilibrium and the Origin of Absolute Uncertainty." *Journal of Statistical Physics* 67(5/6): 843–907.

Earnshaw, Samuel (1842). "On the Nature of the Molecular Forces which Regulate the Constitution of the Luminiferous Ether." Trans. Camb. Phil. Soc., 7, 97–112.

Einstein, A., B. Podolsky, and N. Rosen (1935). "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review* 47: 777–80.

Englert, B. G., M. O. Scully, G. Süssmann, and H. Walther (1992). "Surrealistic Bohm Trajectories." *Zeitschrift für Naturforschung* 47a, 1175–86.

Everett, Hugh III (1956). "The Theory of the Universal Wave Function," in Barrett and Byrne (eds) (2012), 72–172.

Everett, Hugh III (1957a). "On the Foundations of Quantum Mechanics." Thesis submitted to Princeton University, March 1, 1957 in partial fulfillment of the requirement for the Ph.D. degree. See notes in Barrett and Byrne (eds) (2012), 173–96 for a comparison with Everett (1957b).

Everett, Hugh III (1957b) " 'Relative State' Formulation of Quantum Mechanics." *Reviews of Modern Physics* 29: 454–62, in Barrett and Byrne (eds) (2012), 173–96.

Earman, John (1986). *A Primer on Determinism*. University of Western Ontario Series on Philosophy of Science, Dordrecht: D. Riedel.

Fein, Yaakov Y. and Geyer, Philipp and Zwick, Patrick and Kiałka, Filip and Pedalino, Sebastian and Mayor, Marcel and Gerlich, Stefan and Arndt, Markus (2019) "Quantum superposition of molecules beyond 25 kDa" *Nature Physics.* doi:10.1038/s41567-019-0663-9.

Frauchiger, Daniela and Renato Renner (2018). "Quantum Theory Cannot Consistently Describe the Use of Itself." arXiv:1604.07422v2.

Fuchs, Christopher A. (2010). "QBism, the Perimeter of Quantum Bayesianism," manuscript at arxiv.org. https://arxiv.org/abs/1003.5209. Accessed 29 Sept 2017.

Geiger, Hans (1908). "On the Scattering of α-Particles by Matter." *Proceedings of the Royal Society of London A* 81(546): 174–7.

Geiger, Hans and Ernest Marsden (1913). "The Laws of Deflexion of α Particles through Large Angles." *Philosophical Magazine* Series 6. 25(148): 604–23.

Gell-Mann, M. and J. B. Hartle (1990). "Quantum Mechanics in the Light of Quantum Cosmology," in *Complexity, Entropy, and the Physics of Information*, Proceedings of the Santa Fe Institute Studies in the Sciences of Complexity, vol. VIII, W. H. Zurek (ed.) Addison-Wesley, Redwood City, CA, 425–58.

Gerlach, W. and O. Stern (1922a). "Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld." *Zeitschrift für Physik* 9: 349–52.

Gerlach, W. and O. Stern (1922b). "Das magnetische Moment des Silberatoms." *Zeitschrift für Physik* 9: 353–355.

Gerlach, W. and O. Stern (1922c). "Der experimentelle Nachweis des magnetischen Moments des Silberatoms." *Zeitschrift für Physik* 8: 110–11.

Ghirardi, G.C., A. Rimini, and T. Weber (1985). "A Model for a Unified Quantum Description of Macroscopic and Microscopic Systems." *Quantum Probability and Applications*, L. Accardi et al. (eds), Berlin: Springer.

Ghirardi, G.C., A. Rimini, and T. Weber (1986). "Unified Dynamics for Microscopic and Macroscopic Systems." *Physical Review* D 34: 470.

Gilton, Marian J. R. (2016). "Whence the Eigenstate–Eigenvalue Link?" *Studies in History and Philosophy of Modern Physics* 55: 92–100.

Goldstein, Sheldon (2012). "Typicality and Notions of Probability in Physics," in Y. Ben-Menahem and M. Hemmo (eds) *Probability in Physics*, 59–71.

Graham, N. (1973). "The Measurement of Relative Frequency," in DeWitt and Graham (eds) (1973).

Greaves, Hillary (2007). "Probability in the Everett Interpretation." *Philosophy Compass* 2/1: 109–28. 10.1111/j.1747-9991.2006.00054.x.

Griffiths, R. (1984). "Consistent Histories and the Interpretation of Quantum Mechanics." *Journal of Statistical Physics* 36: 219–72.

Hagar, Amit and Meir Hemmo (2006). "Explaining the Unobserved Why Quantum Mechanics Aint Only About Information." *Foundations of Physics* 36(9): 1295–323.

Hartle, J. B. (1968). "Quantum Mechanics of Individual Systems." *American Journal of Physics* 36(8): 704–12.

Hàjek, A. (2011). "Interpretations of Probability." *Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.) https://plato.stanford.edu/entries/probability-interpret/.

Healey, Richard (2012). "Quantum Theory: A Pragmatist Approach." *British Journal for the Philosophy of Science* 63(4): 729–71.

Hemmo, Meir and Orly Shenker (2015). "Probability and Typicality in Deterministic Physics." *Erkenntnis* 80(3): 575–86.

Ismael, J. T. (2016). *How Physics Makes Us Free*. New York: Oxford University Press.

Ismael, J. T. (2007). *The Situated Self*. New York: Oxford University Press.

Kent, Adrian (2010). "One World Verses Many: The Inadequacy of Everettian Accounts of Evolution, Probability, and Scientific Confirmation," in Saunders et al. (eds) (2010) 307–54.

Klein, O. and Y. Nishina (1929). "Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac." *Z. Phys.* 52(11–12): 853 and 869.

Kochen, S. and E. P. Specker (1967). "On the Problem of Hidden Variables in Quantum Mechanics." *Journal of Mathematics and Mechanics* 17: 59–87.

Lazarovici, Dustin and Mario Hubert (2018). "How Quantum Mechanics Can Consistently Describe the Use of Itself. " arXiv:1809.08070v2.

Lewis, Peter J. (2016). *Quantum Ontology: A Guide to the Metaphysics of Quantum Mechanics*. Oxford and New York: Oxford University Press.

Lienert, Matthias, Sören Petrat, and Roderich Tumulka (2017). "Multi-Time Wave Functions." arXiv:1702.05282v1.

Maudlin, Tim W. E. (2019). *Philosophy of Physics: Quantum Theory*. Princeton University Press: Princeton, NJ.

Ney, Alyssa and David Z. Albert (2013). *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*. Oxford: Oxford University Press.

Newton, Isaac (1687). *The Mathematical Principles of Natural Philosophy*. Andrew Motte translation.

Osnaghi, Stefano, Fabio Freitas, and Olival Freire Jr. (2009). "The Origin of the Everettian Heresy." *Studies in History and Philosophy of Modern Physics* 40: 97–123.

Pais, Abraham (1986). *Inward Bound*. Oxford: Oxford University Press.

Pais, Abraham (1982). *Subtle is the Lord*. Oxford: Oxford University Press.

Petersen, Aage (1963). "The Philosophy of Niels Bohr." *Bulletin of the Atomic Scientists* Sept: 8–14.

Planck, M. (1900). "Zur Theorie des Gesetzes der Energieverteilung im Normalspectrum." *Verhandlungen der Deutschen Physikalischen Gesellschaft* 2: 237–45.

Pusey, M. F., J. Barrett, and T. Rudolph (2011). "The Quantum State Cannot Be Interpreted Statistically." arXiv:1111.3328v1.

Pusey, M. F., J. Barrett, and T. Rudolph (2012). "On the Reality of the Quantum State." *Nature Physics* 8(6): 475–8. arXiv:1111.3328.

Putnam, Hillary (2005). "A Philosopher Looks at Quantum Mechanics (Again)." *Brit. J. Phil. Sci.* 56: 615–34.

Quine, W. V. (1980). *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Saunders, Simon (2010). "Many Worlds? An Introduction," in Saunders et al. (eds) (2010) 1–49.

Saunders, Simon (2010). "Chance in the Everett Interpretation," in Saunders et al. (eds) (2010) 181–205.

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

Schrödinger, Erwin (1926). "On the Relation between the Quantum Mechanics of Heisenberg, Born, and Jordan, and That of Schrödinger," in Erwin Schrödinger, *Collected Papers on Wave Mechanics*, third English edition (1982). New York: Chelsea Publishing Company, 45–61. Originally in *Annalen der Physik* (4) vol. 76 (1926).

Schrödinger, Erwin (1935). "Die gegenwärtige Situation in der Quantenmechanik." *Naturwissenschaften* 23(48): 807–12.

Sebens, C. T. and S. M. Carroll (2015). "Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics." arXiv:1405.7577.

Shor, Peter W. (1996). "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer." arXiv: quant-ph/9508027v2.

Susskind, Leonard and Art Friedman (2014). *Quantum Mechanics: The Theoretical Minimum*. New York: Basic Books.

Susskind, Leonard and Art Friedman (2017). *Special Relativity and Classical Field Theory: The Theoretical Minimum*. New York: Basic Books.

Thomson, J. J. (1904). "On the Structure of the Atom: An Investigation of the Stability and Periods of Oscillation of a Number of Corpuscles Arranged at Equal Intervals around the Circumference of a Circle; with Application of the Results to the Theory of Atomic Structure." *Philosophical Magazine*. Sixth. 7(39): 237–65. doi:10.1080/14786440409463107.

Tumulka, Roderich (2007). "The Point Processes of the GRW Theory of Wave Function Collapse." arXiv:0711.0035 [math-ph]. Accessed 11 September 2017.

Vaidman, Lev (2012). "Probability in the Many-Worlds Interpretation of Quantum Mechanics," in Yemima Ben-Menahem and Meir Hemmo (eds), *Probability in Physics (The Frontiers Collection XII)* Berlin and Heidelberg: Springer, 299–311.

van Fraassen, Bas (1989). *Laws and Symmetry*. Oxford: Clarendon Press.

Vink, J. C. (2018). "Particle Trajectories for Quantum Field Theory," *Foundations of Physics* 48: 209–36.

Vink, J. C. (1993). "Quantum Mechanics in Terms of Discrete Beables." *Physical Review A* 48: 1808–18.

von Neumann, J. (1932). *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer.

von Neumann, J. (1955). *Mathematical Foundations of Quantum Mechanics*. Princeton, NJ: Princeton University Press. Translated by R. Beyer.

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

Wallace, David (2010a). "Decoherence and Ontology," in Saunders et al. (eds) (2010) 53–72.

Wallace, David (2010b). "How to Prove the Born Rule," in Saunders et al. (eds) (2010) 227–63.

Weisstein, Eric (ed.) (2017). *Wolfram MathWorld*. http://mathworld.wolfram.com/, accessed 5 September 2017.

Wheeler, John A. and Wojciech H. Zurek (eds) (1983). *Quantum Theory and Measurement*. Princeton, NJ: Princeton University Press.

Wigner, Eugene (1961). "Remarks on the Mind–Body Problem," in I. J. Good (ed.) *The Scientist Speculates*, New York: Basic Books, 284–302. Reprinted in Wheeler and Zurek (eds) (1983) 168–81.

Wigner, Eugene (1963). "The Problem of Measurement." *American Journal of Physics* 31: 6–15.

Wigner, Eugene P. (1970). "On Hidden Variables and Quantum Mechanical Probabilities." *American Journal of Physics* 38: 1005–9.

Wigner, Eugene (1976). "Interpretation of Quantum Mechanics," in Wheeler and Zurek (eds) (1983) 260–314.

Wilson, Mark (2015). "What is Classical Mechanics Anyway?" Forthcoming in Physics Avoidance (an abridged version to appear in *The Oxford Handbook in Philosophy of Physics*, R. Batterman (ed.) (Oxford: Oxford University Press, forthcoming).

Zabell, Sandy (2016). "Johannes von Kriess *Principien*: A Brief Guide for the Perplexed." *Journal for General Philosophy of Science* 47: 131–50.

Index