Undergraduate Lecture Notes in Physics

Travis Norsen

Foundations of Quantum Mechanics

An Exploration of the Physical Meaning of Quantum Theory

Springer

Undergraduate Lecture Notes in Physics

Undergraduate Lecture Notes in Physics (ULNP) publishes authoritative texts covering topics throughout pure and applied physics. Each title in the series is suitable as a basis for undergraduate instruction, typically containing practice problems, worked examples, chapter summaries, and suggestions for further reading.

ULNP titles must provide at least one of the following:

- An exceptionally clear and concise treatment of a standard undergraduate subject.
- A solid undergraduate-level introduction to a graduate, advanced, or non-standard subject.
- A novel perspective or an unusual approach to teaching a subject.

ULNP especially encourages new, original, and idiosyncratic approaches to physics teaching at the undergraduate level.

The purpose of ULNP is to provide intriguing, absorbing books that will continue to be the reader's preferred reference throughout their academic career.

Series editors

More information about this series at http://www.springer.com/series/8917

Travis Norsen

Foundations of Quantum Mechanics

An Exploration of the Physical Meaning of Quantum Theory

Springer

Travis Norsen
Department of Physics
Smith College
Northampton, MA
USA

ISSN 2192-4791 ISSN 2192-4805 (electronic)
Undergraduate Lecture Notes in Physics
ISBN 978-3-319-65866-7 ISBN 978-3-319-65867-4 (eBook)
DOI 10.1007/978-3-319-65867-4

Library of Congress Control Number: 2017949150

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Bohr always would go in for this remark, 'You cannot really explain it in the framework of space and time.' By God, I was determined I was going to explain it in the framework of space and time.

—John Slater

Preface

This textbook is intended as a lifeline to physics students (of either the traditional or the autodidactic variety) who have had some preliminary exposure to quantum mechanics but who want to actually try to make physical and conceptual *sense* of the theory in the same way that they have been trained and expected to do when learning about other areas of physics. Its main goals are (i) to help students appreciate and understand the concerns that people like Einstein, Schrödinger, and Bell have had with traditional formulations of the theory and (ii) to introduce students to the several extant formulations of quantum theory which purport to address at least some of the concerns and provide candidate accounts of what quantum theory might actually imply about how the micro-physical world works.

The book grew out of, and its structure in many ways reflects, the "special topics in physics" course on the *Foundations of Quantum Mechanics* that I taught at Smith College in the Spring of 2016. In this seminar-style course, students would read through each new chapter (and attempt a few of the end-of-chapter Projects that I recommended as appropriate pre-class exercises) prior to our weekly three-hour meeting. During our time together in class, we would discuss the more difficult concepts and derivations from the text, students would share their (sometimes only partial) solutions to the assigned pre-class projects (and we would discuss and complete those as needed), and then we would tackle some additional projects.

Not surprisingly, then, I envision the book being most straightforwardly useful for a similarly structured elective course in a physics department (or perhaps for a philosophy-of-physics course focused on the *Foundations of Quantum Mechanics* in a philosophy department). But the fact that the chapters were created as pre-class readings (as opposed to transcripts of "lecture notes") perhaps makes this, compared to most physics textbooks, unusually readable and accessible to individuals for whom it is not the textbook for any official course—e.g., interested physics students who are not lucky enough to find themselves in a department that offers an elective course on the *Foundations of Quantum Mechanics*, or just anyone with an interest in the puzzling and fascinating history, philosophy, and, really, *physics* of quantum physics.

The book begins with two introductory chapters. Chapter 1 ("Pre-Quantum Theories") introduces a number of important concepts and ideas in the context of classical physics theories such as Newtonian gravity and Maxwellian electrodynamics. Chapter 2 ("Quantum Examples") then provides a lightning overview of some quantum mechanical formalism and examples that serve as a foundation for later discussions. The level of these two chapters (as well as the rest of the book) reflects the background preparation I was able to expect for the students in my course at Smith College: they had taken a sophomore-level "modern physics" course including exposure to Schrödinger's equation and 1-D wave mechanics (but had not yet taken, or were in some cases taking concurrently, a junior-level "quantum mechanics" course using, for example, the text by Griffiths); similarly, they had seen Maxwell's equations before (in a 100-level introductory course and perhaps also a 300-level E&M course) and had a fairly strong prior exposure to vector calculus and differential equations. Still, the students found some of the material in the two introductory chapters quite fresh and challenging.

Readers who are missing one or more of the prerequisites I just mentioned (or readers who are pursuing physics outside of, or perhaps decades beyond, any organized undergraduate physics curriculum) should thus anticipate some struggle with some mathematical details in the first two chapters. However, I want to reassure people in this category that it will be OK, and that they should get what they can out of the first two chapters and press forward into the rest of the book. Let me explain my attitude here with an example. I don't think you can fully appreciate Bell's Theorem (the subject of Chap. 8) without digesting, in Chap. 1, the reasonableness of Bell's formulation of "locality" as a generalization of the specifically deterministic sort of local causality exhibited by Maxwellian electrodynamics (in contrast to Newtonian gravitation). But for readers for whom understanding the mathematical details is too big a stretch, it will suffice to merely accept *that* Bell's formulation purports to be a natural generalization of the important relativistic locality of classical E&M.

After the two introductory chapters, the book turns toward the first goal mentioned above: Helping students appreciate and understand the concerns that people like Einstein, Schrödinger, and Bell had with traditional formulations of quantum theory. We begin in Chap. 3 by studying "The Measurement Problem" which was most famously illustrated by Schrödinger's infamous cat and then emphasized and significantly clarified by Bell. Chapter 4 tackles "The Locality Problem" which was most famously brought out in the 1935 paper of Einstein, Podolsky, and Rosen—although, as we will discuss in detail, this canonical presentation does not perfectly capture Einstein's fundamental objection to the orthodox interpretation. Finally, Chap. 5 introduces "The Ontology Problem"—a concern that was intensely worrying to Schrödinger, Einstein, and others in the early days of quantum mechanics, but which has, unfortunately, been largely forgotten in the instrumentalist and anti-realist wake of the Copenhagen orthodoxy (and which, again unfortunately, remains under-appreciated even by certain schools of anti-Copenhagen quantum realism). One of the things I like best about this book is that it gives the ontology problem pride of place alongside the (more widely recognized) measurement and

locality problems as one of the "big three" concerns that clearheaded physicists should have in mind when they are evaluating and developing candidate theories.

Having thus surveyed the central problems that one would hope to see resolved, the book turns to reviewing and assessing the menu of available resolutions. We cover, in particular, what I consider the four most important perspectives on quantum mechanics that curious and intelligent physics students should understand. These include: in Chap. 6, "The Copenhagen Interpretation" (which is a self-confessed non-candidate for genuinely explaining micro-physics in an ordinary, realist way, but is of historical and sociological interest nevertheless since it has been the official, if only superficially understood and half-heartedly accepted, orthodoxy of the physics community for nearly a century); in Chap. 7, "The Pilot-Wave Theory" of de Broglie and Bohm; in Chap. 9, "The Spontaneous Collapse Theory" of Ghirardi, Rimini, Weber, and Pearle; and, finally, in Chap. 10, "The Many-Worlds Theory" of Everett. Chapter 8, on "Bell's Theorem," is a kind of sequel to Chap. 4 which explains the Earth-shattering advance that Bell was led to from his study of the pilot-wave theory.

The material in this second half of the book is, to a large but not perfect extent, organized historically. Thus, the Copenhagen Interpretation (largely developed in the 1930s) comes first, the pilot-wave theory (originally proposed by de Broglie in 1927 but then largely forgotten until Bohm resurrected the idea in 1952) comes second, then we turn to Bell's theorem of 1964 (which, as mentioned, was directly stimulated by Bell's contemplation of a seemingly troubling feature of the pilot-wave theory), and then the spontaneous collapse theories (which only began to be developed in the 1980s). Everett's many-worlds theory is presented last, despite the fact that Everett proposed it in 1957 (between Bohm's resurrection of the pilot-wave idea and Bell's presentation of his important theorem), both because the theory was not widely recognized as a serious candidate account of quantum phenomena until much more recently, and also because I think it is hard to recover from studying something rather surreal and focus on something rather more mundane!

Note that it might be slightly puzzling that the Copenhagen Interpretation is only covered (in Chap. 6) *after* we have reviewed the measurement, locality, and ontology problems (in Chaps. 3, 4, and 5, respectively)—this despite the fact that these "problems" were largely raised in *response* to the interpretive pronouncements of Bohr and Heisenberg and their colleagues. I structured things this way in part because I assume that students will already have been exposed, as part of a "modern physics" type course, to the basic Copenhagen philosophy of insisting on the *completeness* of the description in terms of wave functions alone (but also, paradoxically, denying the *reality* of wave functions) and then foreclosing further discussion as somehow scientifically inappropriate. So I thought students would be able to appreciate the somewhat-reactionary concerns of, for example, Einstein and Schrödinger, without any explicit prior discussion of the Copenhagen philosophy. In addition, I think having a clear sense of the critics' concerns can help motivate students to actually care about what, exactly, Bohr and Heisenberg said: Did they really assert what the critics reacted against, and did they have viable answers to the

criticisms? Finally, I thought that giving Bohr and Heisenberg the last word (after hearing from the critics) was a good way to try to maintain the neutrality that I have aimed at throughout the book—despite, perhaps obviously, not thinking very highly of the Copenhagen philosophy.

In the Smith College course, we went through these topics at a pace of one chapter per week. That left a couple of weeks at the end of the semester, during which the students each picked a topic they were individually interested in exploring further, did some independent reading and research, and then gave a presentation back to the class summarizing what they had learned and uncovered. This structure is reflected in the present book, which closes with an "Afterword" that tries to bring an (admittedly limited) element of closure to the covered topics by summarizing where things stand and then provides an informal laundry list of recommended topics for further exploration, including pointers to some more contemporary literature.

I attempt, though, even in the ten chapters of the book, to build bridges to the primary literature. There is, for example, extensive quoting from the published papers (as well as the private correspondence) of Einstein, Lorentz, Schrödinger, Bell, Bohr, Heisenberg, etc., and many of the end-of-chapter Projects invite students to read some accessible piece of primary literature and report on things they find interesting or surprising. Indeed, one of my goals with this book is to help students appreciate the extent to which their own confusions and concerns about quantum mechanics are not something to feel ashamed of (a feeling that is too-often the result of the "shut up and calculate" attitude that quantum physics professors frequently take toward the subjects we cover). Instead, students should feel proud that they can understand, and indeed in many cases will have anticipated without realizing it, concerns that were shared by some of the giants of twentieth-century physics—concerns that have unfortunately been suppressed and forgotten rather than adequately addressed. To capture the intended spirit of the book in this respect, I can do no better than quote from an email from my friend Kenny Felder who read drafts of most of the chapters:

> Reading [this], I have I think exactly the sense that you want me to have—or perhaps the meta-sense that you want me to have—in any case it's a wonderful sense that I really have never had before. I have the sense of a group of men who are very smart but perfectly human, right at the dawn of the quantum revolution, desperately trying to figure out what the experimental evidence is actually telling them. I see them throwing ideas around, trying and rejecting theories, alone and in correspondence with each other. And I get the sense that somewhere between them and us, that search for a coherent theory more or less evaporated—not because the questions were answered, but more because people kind of forgot about them—and you're trying to revitalize that quest. It's exciting!

Let me finally say something about the end-of-chapter "Projects" which I consider to be an essential component of the book, just as they were an essential component of the course it grew out of. Some of these are rather like traditional end-of-chapter exercises, which ask students, for example, to fill in gaps in derivations from the text or apply concepts introduced in the text to simple concrete examples. But many of the Projects are considerably more challenging and open-ended. For example, as

mentioned above, some invite students to read an article or essay that has been discussed in the text and report back on things they find interesting, surprising, or novel. Some projects invite students to use Mathematica or another programming language to create helpful visualizations or numerical solutions of difficult problems. There are even a few Projects (perhaps most suitable for students using the text in the context of a traditional course) asking students to interview a few physicists to get a sense of how real people think about some issue. It is hoped that the diversity and open-endedness of the Projects will allow students with many different backgrounds, technical abilities, and interests to stay actively engaged with the material (before, during, after, and/or without classroom time, as appropriate in each individual case).

Let me close by thanking Darby Bates, Jean Bricmont, Kira Chase, Kenny Felder, and Trevor Wright for reading, and providing significant helpful feedback on, earlier versions of at least some of the chapters. I also owe a more generalized debt of gratitude to my tireless and inspiring wife Sarah, and my kids Finn and Tate, for helping keep me grounded and happy—as well as to my wonderful parents, Steve and Carol, for believing and investing in me (especially by supporting my own undergraduate education at Harvey Mudd College, where my interest in the subject matter of this book began).

Northampton, USA Travis Norsen

Contents

1 Pre-Quantum Theories 1
1.1 Newtonian Mechanics 1
1.2 Maxwellian Electrodynamics 5
1.3 Locality 8
1.4 Bell's Formulation of "Locality" 13
1.5 Ontology 18
1.6 Measurement 22
1.7 Abstract Spaces 25
References 31

2 Quantum Examples 33
2.1 Overview 33
2.2 Particle-in-a-Box 36
2.3 Free Particle Gaussian Wave Packets 38
2.4 Diffraction and Interference 44
2.5 Spin 47
2.6 Several Particles 51
References 57

3 The Measurement Problem 59
3.1 The Quantum Description of Measurement 59
3.2 Formal Treatment 64
3.3 Schrödinger's Cat and Einstein's Bomb 69
3.4 Hidden Variables and the Ignorance Interpretation 74
3.5 Wrap-Up 79
References 85

4 The Locality Problem 87
4.1 Einstein's Boxes 87
4.2 EPR 96
4.3 Einstein's Discussions of EPR 100

4.4 Bohm's Reformulation . . . 104
4.5 Bell's Re-Telling . . . 107
References . . . 113

5 The Ontology Problem . . . 115
5.1 Complexity and Reality . . . 115
5.2 Configuration Space . . . 118
5.3 Ontology, Measurement, and Locality . . . 122
5.4 Schrödinger's Suggestion for a Density in 3-Space . . . 129
5.5 So Then What? . . . 133
References . . . 139

6 The Copenhagen Interpretation . . . 141
6.1 Bohr's Como Lecture . . . 142
6.2 Heisenberg . . . 148
6.3 Bohr on Einstein's Diffraction Example . . . 154
6.4 The Photon Box Thought Experiment . . . 160
6.5 Bohr's Reply to EPR . . . 166
6.6 Contemporary Perspectives . . . 169
References . . . 174

7 The Pilot-Wave Theory . . . 177
7.1 Overview . . . 178
7.2 Particle in a Box . . . 182
7.3 Other Single Particle Examples . . . 185
7.4 Measurement . . . 188
7.5 Contextuality . . . 194
7.6 The Many-Particle Theory and Nonlocality . . . 199
7.7 Reactions . . . 205
References . . . 212

8 Bell's Theorem . . . 215
8.1 EPRB Revisited . . . 215
8.2 A Preliminary Bell Inequality . . . 218
8.3 The Real Bell (and the CHSH) Inequality . . . 222
8.4 Experiments . . . 227
8.5 What Does It Mean? . . . 231
8.6 (Bell's) Locality Inequality Theorem . . . 236
References . . . 243

9 The Spontaneous Collapse Theory . . . 245
9.1 Ghirardi, Rimini, and Weber . . . 246
9.2 Multiple Particle Systems and Measurement . . . 254
9.3 Ontology, Locality, and Relativity . . . 259
9.4 Empirical Tests of GRW . . . 265
References . . . 271

10 The Many-Worlds Theory 273
10.1 The Basic Idea 274
10.2 Probability 280
10.3 Ontology 286
10.4 Locality 291
References 302

Afterword 303

Chapter 1
Pre-Quantum Theories

In this introductory chapter we review two theories from classical physics – Newtonian mechanics and Maxwellian electrodynamics – and use them to introduce a number of concepts (such as determinism, locality, ontology, measurement, and configuration space) that we will explore in the context of quantum mechanics in subsequent chapters.

1.1 Newtonian Mechanics

As a first example of a "pre-quantum theory" let's consider the picture of the universe formulated by Isaac Newton. The theory, in a nutshell, says that the physical world consists of *particles* interacting by means of *forces* which the particles exert on one another and which influence the particles' motions. About the particles, Newton wrote:

> ...it seems probable to me, that God in the Beginning form'd Matter in solid, massy, hard, impenetrable, moveable Particles, of such Sizes and Figures, and with such other Properties, and in such Proportion to Space, as most conduced to the End for which he form'd them; and that these primitive Particles being Solids, are incomparably harder than any porous Bodies compounded of them; even so very hard, as never to wear or break in pieces.... [A]ll material Things seem to have been composed of the hard and solid Particles above-mention'd, variously associated.... [1, pp. 400–2]

Newton's endorsement of the idea that observable macroscopic objects are composed of invisibly small, indestructible particles is a kind of bridge between the speculative notion of atomism that had been introduced by Ancient Greek philosophers such as Democritus, and the more scientific atomic theory of matter that grew out of chemistry and physics in the centuries following Newton.

Regarding the forces that these particles exert on one another, Newton wrote that

> Bodies act one upon another by the Attractions of Gravity, Magnetism, and Electricity; and these Instances shew the Tenor and Course of Nature, and make it not improbable but that

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_1

> there may be more attractive Powers than these. .... [W]e must learn from the Phaenomena of Nature what Bodies attract one another, and what are the Laws and Properties of the Attraction.... The Attractions of Gravity, Magnetism, and Electricity, reach to very sensible distances, and so have been observed by vulgar Eyes, and there may be others which reach to so small distances as hitherto escape Observation.... [1, p. 376]

Although he did not have any particular detailed theories about them, Newton thus anticipated the empirical quest to understand the short-range attractions and repulsions between particles that we now think of as responsible for micro-physical, chemical, and even biological processes. But of course Newton *did* have a rather well-worked-out theoretical account of the long-range *gravitational* interactions between particles.

According to Newton's law of universal gravitation, the gravitational force exerted *on* a particle of mass m_i located at position $\vec{r}_i$, *by* another particle of mass m_j located at position $\vec{r}_j$, is given by

$$\vec{F}_{i,j} = \frac{Gm_i m_j}{r_{ij}^2}\hat{r}_{ij} \tag{1.1}$$

where $r_{ij} = |\vec{r}_i - \vec{r}_j|$ is just the distance between the two particles and

$$\hat{r}_{ij} = \frac{\vec{r}_j - \vec{r}_i}{r_{ij}} \tag{1.2}$$

is a unit vector pointing along the line from $\vec{r}_i$ back toward $\vec{r}_j$. The gravitational force between two elementary particles, that is, is proportional to the product of the masses of the particles, inversely proportional to the square of the distance between them, and is directed back toward the particle exerting the force. The proportionality constant, G, which we now call "Newton's constant", was first measured by Cavendish about a century after Newton.

The total or net force on the i^{th} particle is then

$$\vec{F}_i^{\text{net}} = \sum_{j \neq i} \vec{F}_{i,j}. \tag{1.3}$$

(Note that here we ignore the existence of other, short-range forces and pretend for simplicity that the particles *only* interact gravitationally.) And of course it is this net force that influences the particle's trajectory through space in accordance with Newton's second law of motion:

$$\vec{F}_i^{\text{net}} = m_i \vec{a}_i. \tag{1.4}$$

Note that Newton's inverse square law, Eq. (1.1), also embodies Newton's third law: for every action there's an equal and opposite reaction. Or more precisely: if j exerts a force on i, then i necessarily also exerts a force on j, and these two forces (that they exert on each other) have equal magnitudes but precisely opposite directions. That is:

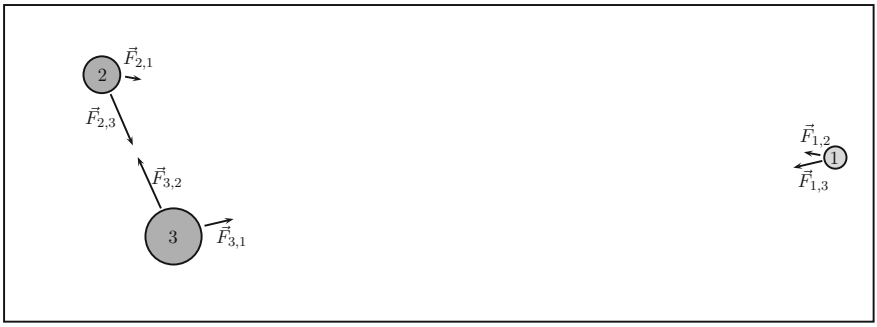


Fig. 1.1 Three massive bodies and the gravitational forces they exert on one another

$$\vec{F}_{i,j} = -\vec{F}_{j,i}. \tag{1.5}$$

It is nice to have some pictures to go along with all the equations, so in Fig. 1.1 I've illustrated some of these ideas by showing three particles (which one might think of as two stars forming a binary star system plus an orbiting planet) and the forces they exert on one another.

Note that the basic laws of Newtonian mechanics (both the expressions for the forces and also Newton's second law which describes how the particles respond to forces) are postulated as applying fundamentally to the elementary, microscopic "Particles" that Newton spoke of in the first block quote. It is perhaps not terribly surprising, but important and interesting nevertheless, that these same laws (properly understood) *also* turn out to apply to large macroscopic objects like stars and planets and apples. That is, in Newtonian mechanics, the applicability to macroscopic objects of (for example) the gravitational inverse square law and Newton's second law, are *theorems* which can be *derived* from the basic laws (understood as applying to the elementary Particles) rather than postulates. You are invited to consider this point further in some of the end-of-chapter Projects.

It is perhaps worth making more explicit that the long-range gravitational forces exerted on each particle depend, according to Eq. (1.1), on the *instantaneous* positions of the distant particles exerting the forces. There is nothing like a delay, for example, associated with some finite-speed propagation of the gravitational influence. The Newtonian gravitational forces, as described by Eq. (1.1), are thus *non-local*, by which we simply mean that they embody what Einstein would describe as a kind of "spooky action at a distance." Interestingly, though, Newton himself did not believe that this apparent non-locality should be taken seriously, as accurately capturing the true nature of gravitational interactions. In a famous 1693 letter to Richard Bentley, Newton wrote:

> It is inconceivable that inanimate brute matter should, without the mediation of something else which is not material, operate upon and affect other matter without mutual contact... That gravity should be innate, inherent, and essential to matter, so that one body may act upon another at a distance through a vacuum, without the mediation of anything else, by and

> through which their action and force may be conveyed from one to another, is to me so great an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it [2].

Newton thus evidently did not regard what we are here calling "Newtonian mechanics" as providing a complete and final description of gravitational interactions. Instead he seems to have regarded it as merely a starting point, justified by its success in accounting for the observed motions of planets (and comets and tides and falling apples and so on). And as we will see in the next section, physical theory did ultimately develop along the lines suggested here by Newton, with the "field" concept (and the associated removal of the troubling, if merely apparent, non-locality) that Faraday and Maxwell introduced into electro-magnetic theory (and then Einstein introduced into gravitational theory with his general theory of relativity).

Let us close this section with one last figure, which serves two functions: first, introducing the idea of a "space-time diagram" and, second, illustrating one last implication of the (perhaps merely apparent) non-locality of Newton's account of gravity.

A "space-time diagram" is just a graph of position versus time, but (by convention) with the time axis running vertically upward in the figure. Thus, a horizontal slice through the diagram represents the configuration of objects at some particular moment in time.[1] The curves representing the paths of objects "through space-time" are called "world-lines". This probably sounds fancier and deeper than it is; remember this is just a graph of position versus time, but turned sideways! Fig. 1.2 shows a space-time diagram for the same three-object system illustrated before. One sees the orbits of the two "stars" in the "binary star system" (about their mutual center of mass) as the double-helical world lines, with the world line for the (much lighter) "planet" suggesting a longer-period orbit around the "stars".

The particular time slice shaded in gray is meant to correspond to the situation depicted earlier, in Fig. 1.1. Note that, because of the dynamical non-locality mentioned before, there is a sense in which we must think of this particular "slicing" of space-time as somehow "objectively real" according to this theory. That is, in order for the equations of motion of the theory to be well-defined, there must be a real fact of the matter about which points, on the world lines of the other distant particles, are simultaneous with the point in question where we want to know the force.

To make this concrete, imagine a tilted slice through the same point on the world line of the "planet" that the slice in the Figure passes through. The tilted slice would intersect the world lines of the two "stars" at different points on their world lines, and therefore the magnitudes and directions of the gravitational forces exerted on the "planet" at that moment would be different. But the planet is going to move in some particular way, and this requires that some one of the possible ways of slicing up the

[1] That is: a particular moment in time from the point of view of some particular reference frame. Readers who have studied special relativity will recall that due to the relativity of simultaneity, a slice of space-time representing events occurring at the same time, but for a differently-moving observer, will appear tilted with respect to the one drawn in the Figure. This will be important shortly.

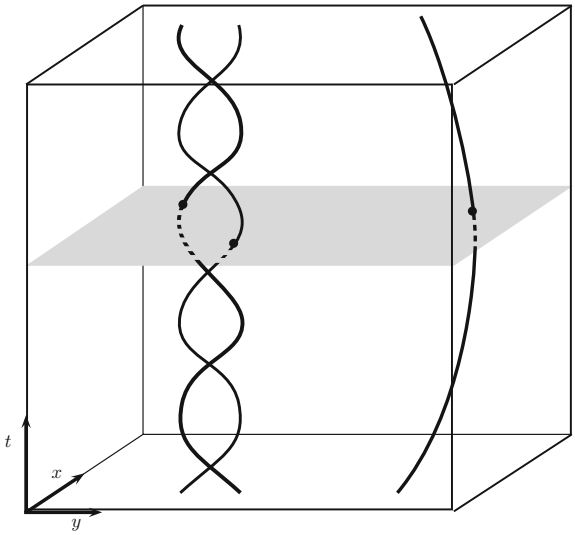


Fig. 1.2 A space-time diagram depicting the evolution, through time, of the binary-star-plus-planet system described earlier

space-time into "simultaneity slices" – the one that gives the forces that generate the planet's actual trajectory – is dynamically privileged, objectively correct.

But this idea of a special, dynamically privileged reference frame contradicts the fundamental principle of relativity theory, that all reference frames are equally valid. Or, as Einstein expressed this point,

> There is [according to relativity] no such thing as simultaneity of distant events; consequently there is also no such thing as immediate action at a distance in the sense of Newtonian mechanics [3, p. 61].

Newton's philosophical instincts may rightly have bristled at the idea of instantaneous action-at-a-distance, but Einstein's relativity theory provided, for the first time, a strong physics-based reason for denying the sort of non-locality suggested by (a naive reading of) Newton's law of universal gravitation.

1.2 Maxwellian Electrodynamics

Let us then turn to a second example "pre-quantum" theory – the theory of electrically charged particles interacting with electric and magnetic *fields*. We begin for simplicity with the case of electrostatics, which is basically parallel to Newtonian gravitation. According to *Coulomb's law*, electrically charged particles exert forces on each other, with the force exerted on charge i by charge j being

$$\vec{F}_{i,j} = -\frac{kq_i q_j}{r_{ij}^2}\hat{r}_{ij} \tag{1.6}$$

which is of course just like Newton's inverse square law for gravity but with the masses replaced by charges. (The minus sign out front is also different: whereas all masses are positive and gravitational forces are always attractive, the electrostatic force between two like charges – either both positive or both negative – is instead *repulsive*.) The net electric force on charge i is then

$$\vec{F}_i^{\text{net}} = \sum_{j \neq i} \vec{F}_{i,j}. \tag{1.7}$$

The concept of electric *field* is often first introduced as a kind of calculational tool: we define the electric field at some location $\vec{r}$ as the (net electric) force that a charged particle would feel if it were located at $\vec{r}$, divided by the charge q of that hypothetical charged particle. That is:

$$\vec{E}(\vec{r}) = \frac{\vec{F}^{\text{net}}(\vec{r})}{q} \tag{1.8}$$

(or, perhaps, to take care of a certain technical detail that need not concern us here, the same thing but in the limit as $q \to 0$). Thus, the electric field at each point $\vec{r}$ can be written as a sum over contributions from all the charged particles in the universe:

$$\vec{E}(\vec{r}) = \sum_{i=1}^{N} \frac{kq_i}{|\vec{r} - \vec{r}_i|^2} \hat{r}_i \tag{1.9}$$

where $\hat{r}_i$ is a unit vector pointing from the point $\vec{r}_i$ to the point $\vec{r}$ where we are calculating the electric field. This equation embodies what is called the principle of *superposition*, which basically means that if charged particle 1 produces a field $\vec{E}_1$ at some point (i.e., if *only* charged particle 1 were around, the electric field at that point would be simply $\vec{E}_1$) and charged particle 2 produces a field $\vec{E}_2$ at that point (i.e., if *only* charged particle 2 were around, the electric field at that point would be simply $\vec{E}_2$), then – with *both* 1 and 2 around – the field is just the sum $\vec{E}_1 + \vec{E}_2$. The concept of "superposition" plays an important role in quantum mechanics, so we highlight it here.

It follows from the definition of the electric field above that, if a particle with charge q is located at a point $\vec{r}$ where the electric field is $\vec{E}(\vec{r})$, it will feel a force

$$\vec{F} = q\vec{E}. \tag{1.10}$$

So far, the electric field should seem like a kind of pointless calculational middle-man: we say the force on a charged particle is determined by its charge and the electric field at its location, and then the electric field at its location is just defined as the superposition of a bunch of inverse-square-law contributions from all the other charged particles. Why not just eliminate the middle man and return to Eq. (1.6) with its apparent "spooky action at a distance"?

The answer is that there turns out to be compelling evidence to take the electric field $\vec{E}(\vec{r})$ – as well as the related magnetic field $\vec{B}(\vec{r})$ – *seriously*, not as mere calculational devices to help us compute the forces that the particles exert on each other, but as geniune physically-real *things* that can (for example) carry energy and momentum and other physical properties and so must actually *exist* in addition to the charged particles.

It will be helpful to remind ourselves about Maxwell's equations, which can be understood as telling us how the electric and magnetic fields change in response to the charged particles (and one another). To begin with we have "Gauss' Law"

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \tag{1.11}$$

where ρ is the electric charge density. For example, for a set of point charges q_i at positions $\vec{r}_i$, we would have $\rho(\vec{r}) = \sum_i q_i \delta^3(\vec{r} - \vec{r}_i)$. Gauss' Law should be understood as a re-formulation of Coulomb's Law, which tells us that the electric field around a point charge is radially outward and falls off in magnitude as the inverse square of the distance. (The ϵ_0 is just a constant, related to the constant k that appeared earlier in Coulomb's Law according to $k = \frac{1}{4\pi\epsilon_0}$.)

The second of Maxwell's equations is sometimes called "Gauss' Law for Magnetism." It reads

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{1.12}$$

which can be understood as saying that there are no "magnetic charges" (which produce radially-outward magnetic fields in the same way that electric charges produce radially-outward electric fields). Sometimes this is stated with the assertion that "there are no magnetic monopoles" or by noting that "magnetic field lines never begin or end but instead make closed loops".

But if there are no magnetic charges, why do magnetic fields exist at all? What produces them? The answer turns out to be: *moving* electric charges, i.e., electric *currents*. This is captured in Ampere's Law:

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{j} + \epsilon_0 \mu_0 \frac{\partial \vec{E}}{\partial t}. \tag{1.13}$$

The $\vec{j}$ on the right hand side is the electric current density. Again, for example, for a set of electric point charges moving with velocities $\vec{v}_i$, we can write the simple expression $\vec{j} = \sum_i q_i \vec{v}_i \delta^3(\vec{r} - \vec{r}_i)$. The μ_0 is another fundamental constant. So, actually, it is only the first term on the right hand side which corresponds to what I wrote just above, that magnetic fields are produced by moving charges. The second term (the so-called "displacement current" term that was famously added to Ampere's original expression by Maxwell) says that, in addition, *changing* electric fields can produce (or here sometimes one says "induce") magnetic fields. Changing electric fields, that is, in some sense act just like electric current in so far as they are able to give rise to magnetic fields.

The fourth and final Maxwell equation is Faraday's law, which says that changing magnetic fields can also give rise to ("induce") electric fields:

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}. \tag{1.14}$$

These last two equations taken together imply the existence of propagating ("electromagnetic") waves, in which a changing electric field at some point stimulates the appearance of a magnetic field at neighboring points, but the coming-into-existence of this magnetic field in turn induces the coming-into-existence of further electric fields at still-further neighboring points, and so on. It can be shown that these waves propagate with speed $1/\sqrt{\epsilon_0 \mu_0}$ which we identify as the speed of light, c:

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 3 \times 10^8 \text{ m/s}. \tag{1.15}$$

Together, as we have said, Maxwell's equations tell us how the electric and magnetic fields respond to (themselves, each other, and) the charged particles. To complete the formulation of the theory we also need to know how the charged particles respond to the fields. This information is contained in the so-called Lorentz Force Law, which says that a particle of charge q moving with velocity $\vec{v}$ at position $\vec{r}$ feels a force

$$\vec{F} = q\vec{E}(\vec{r}) + q\vec{v} \times \vec{B}(\vec{r}). \tag{1.16}$$

which determines the particle's trajectory according to Newton's second law.

To summarize, the overall picture of the universe according to Maxwellian electrodynamics involves particles moving through a kind of background "soup" (the fields). The fields influence the motion of the particles and can be thought of, just as Newton had suggested, as a kind of space-filling means or intermediary by which the particles influence one another. But the fields are physically real dynamical objects in their own right as well.

1.3 Locality

We have already described the (perhaps merely apparently) non-local character of Newton's theory of gravity and hinted at the idea that the situation is different in Maxwellian electrodynamics. Let us explore this further. I mentioned above that Maxwell's equations imply the existence of electromagnetic waves that propagate at the speed of light c. It is hardly obvious just looking at Maxwell's equations, but actually *all* electromagnetic interactions as such propagate at speed c (or, in some situations/senses, slower).

Let's try to extract this from the equations. To begin with, let's see how the wave equations for $\vec{E}$ and $\vec{B}$ follow from Maxwell's equations. Taking the curl of Eq. (1.14)

and using the vector identity $\vec{\nabla} \times (\vec{\nabla} \times \vec{V}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{V}) - \nabla^2 \vec{V}$ gives

$$\vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \nabla^2 \vec{E} = -\frac{\partial}{\partial t}\left(\vec{\nabla} \times \vec{B}\right). \tag{1.17}$$

We may then use Eqs. (1.11), (1.13), and (1.15) to arrive at

$$\nabla^2 \vec{E} - \frac{1}{c^2}\frac{\partial^2 \vec{E}}{\partial t^2} = \vec{\nabla}\left(\frac{\rho}{\epsilon_0}\right) + \frac{\partial}{\partial t}(\mu_0 \vec{j}). \tag{1.18}$$

In empty space, where $\rho = 0$ and $\vec{j} = 0$, this is simply the wave equation for $\vec{E}$:

$$\nabla^2 \vec{E} - \frac{1}{c^2}\frac{\partial^2 \vec{E}}{\partial t^2} = 0. \tag{1.19}$$

So far so good.

A similar process – beginning by taking the curl of Eq. (1.13) – gives

$$\nabla^2 \vec{B} - \frac{1}{c^2}\frac{\partial^2 \vec{B}}{\partial t^2} = -\mu_0 \vec{\nabla} \times \vec{j}. \tag{1.20}$$

So the magnetic field $\vec{B}$ also satisfies the wave equation

$$\nabla^2 \vec{B} - \frac{1}{c^2}\frac{\partial^2 \vec{B}}{\partial t^2} = 0 \tag{1.21}$$

in empty space. And note that the two kinds of waves (electric and magnetic) are necessarily coupled. So in empty space we have electromagnetic waves that propagate at the speed of light, c.

To understand exactly how electric charges (and moving charges, i.e., electric currents) affect the surrounding electric and magnetic fields, however, we should study Eqs. (1.18) and (1.20) in their full glory, including the source terms on the right hand sides. To begin with, think of these two equations as *six* equations, one for each Cartesian component of the fields. These six equations all have the same basic structure, which, for simplicity, we re-write here as follows:

$$\nabla^2 \psi(\vec{x}, t) - \frac{1}{c^2}\frac{\partial^2 \psi(\vec{x}, t)}{\partial t^2} = f(\vec{x}, t) \tag{1.22}$$

where the function $f(\vec{x}, t)$ represents the source term. The source term is in fact some time- or space- derivative of the charge density ρ or the current density $\vec{j}$ (or some combination of such things) but as it turns out the details don't matter so we will work in terms of the generic f.

Working out the detailed solution to Eq. (1.22) gets a little bit technical.[2] Here we will try to explain the overall idea and invite you to work through some of the details in the Projects.

First, it is possible to show that if the source f is concentrated at a single point ($\vec{x} = \vec{x}'$) in space and only pops into existence for an instant at $t = t'$, i.e., if

$$f(\vec{x}, t) = \delta^3(\vec{x} - \vec{x}\,')\,\delta(t - t'), \tag{1.23}$$

then

$$\psi_{\vec{x}',t'}(\vec{x}, t) = -\frac{1}{4\pi}\frac{\delta\left(t - \left[t' + \frac{|\vec{x}-\vec{x}'|}{c}\right]\right)}{|\vec{x} - \vec{x}'|} \tag{1.24}$$

is a solution of Eq. (1.22).[3]

What this says, in words, is that the point source at $(\vec{x}\,', t')$ gives rise to a non-zero field ψ only where the argument of the δ-function on the right hand side vanishes – i.e., only at positions $\vec{x}$ and times t satisfying

$$t = t' + \frac{|\vec{x} - \vec{x}'|}{c} \tag{1.25}$$

– i.e., only at positions $\vec{x}$ and times t which could be reached by a signal propagating out at the speed of light from the source at $\vec{x}'$ and t'. This set of events should be thought of as a growing spherical shell that propagates outward from the source point. But it is often referred to as the "future light cone" of the source point $(\vec{x}', t')$ because, when plotted in a space-time diagram (with one of the 3 spatial dimensions suppressed to make room for time!) the points where the field ψ is affected form a cone. See Fig. 1.3.

Now, of course, realistic sources $f(\vec{x}, t)$ are not concentrated at individual points in space-time. But any realistically distributed source can always be written as an integral (think: sum) over such point-sources. And then, by the principle of superposition, the total field that these sources produce is just the sum over the fields that would be produced by each of the point sources taken individually. That means, if you think about it, that the total field produced at some point $(\vec{x}, t)$ will involve contributions from all the sources present on the *past light cone* of $(\vec{x}, t)$, i.e., from all the locations in space-time that could "broadcast" an influence outward at the speed of light that just reaches $(\vec{x}, t)$.

See Fig. 1.4 for an illustration of the implied picture, of electric charges interacting in a locally causal way by means of influences propagating through the fields. In

[2] It is explained, for example, in Chap. 6 of Jackson [4].

[3] Note also that there is a second solution proportional to $\delta\left(t - \left[t' - |\vec{x} - \vec{x}'|/c\right]\right)$. Equation (1.24) is called the "causal" solution because it describes charges and currents affecting the surrounding fields in the future; we set aside here the other solution, which describes charges and currents affecting the surrounding fields in the past. Finally, if you want to pursue the mathematics in a little more detail, it will help to know that Eq. (1.24), a solution to the wave equation for a delta-function source, is called the "Green's function" for the wave equation.

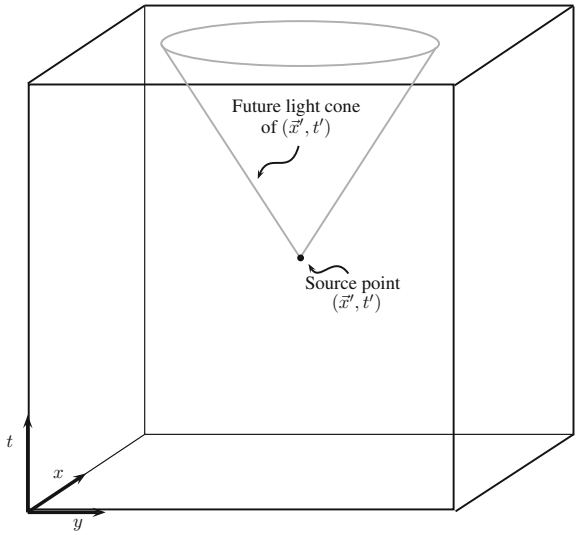


Fig. 1.3 A spatio-temporal point source $f(\vec{x}, t) = \delta^3(\vec{x} - \vec{x}')\delta(t - t')$ affects the field ψ at points on the future light cone of the source point $(\vec{x}', t')$

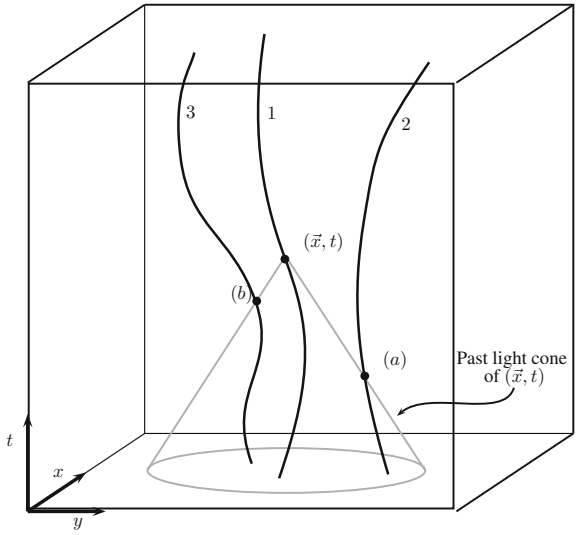


Fig. 1.4 What happens to particle 1 at the point $(\vec{x}, t)$ is determined, according to the Lorentz Force Law, by the fields $\vec{E}$ and $\vec{B}$ at that point. These, in turn, are determined by the sources at points on the past light cone of $(\vec{x}, t)$ – for example, as shown here, what particles 2 and 3 were doing at the points marked **a** and **b**

particular, if for example you want to know what particle 1 does at the point marked $(\vec{x}, t)$ in the Figure, this depends on the fields $\vec{E}(\vec{x}, t)$ and $\vec{B}(\vec{x}, t)$ at this point, which in turn are influenced by source terms on the past light cone, i.e., what particles 2 and 3 were doing at the points marked (a) and (b).

So far so good. Note, though, that in the language of differential equations, we have so far only been discussing the "particular solution" of Eq. (1.22) – that is, the contributions to the fields ψ which arise specifically from nonzero source terms f. But there is always in addition the so-called "complementary solution" – i.e., the solution

of the corresponding homogeneous problem. But this part of the general solution is simple, because the corresponding homogeneous problem is just the ordinary wave equation whose solutions are electromagnetic waves that propagate at the speed of light.

Anyway, the point here is that the "general solution" of Eq. (1.22) can be understood as the sum of two things: first, contributions from electric charges and currents at points on the past light cone of the point in question, and then second, contributions from "freely propagating" electromagnetic waves (which perhaps were themselves created in the more distant past by wiggling charges, or which perhaps instead have just always been around since the beginning of time).

To summarize, then, Maxwellian electrodynamics is a completely *local* theory: what *happens* at a given point in space-time depends exclusively on events lying on (or inside) the past light cone of that event. This is really just a fancy and formal way of saying that causal influences always propagate, according to this theory, at the speed of light (or slower[4]). And this is in contrast to Newtonian gravity (at least naively interpreted) in which, as we saw, what *happens* at a given point in spacetime depends on things that are happening *at that same moment* arbitrarily far away.

Note that the local character of Maxwellian electrodynamics makes it compatible with Einstein's relativity theory in a way that Newton's theory of gravity is not. According to relativity, simultaneity is relative; that is, according to relativity, there is simply no objective fact of the matter about what set of events are *simultaneous* with a given space-time point. And so, from the point of view of relativity, the Newtonian gravitational idea that the force on a certain particle at a certain moment in time depends on the *instantaneous* configuration of all the other particles in the universe, is literally meaningless. Whereas the idea in Maxwellian electrodynamics, that the force on a certain particle at a certain moment depends on what other particles were doing *earlier*, by exactly the amount required for a signal to propagate at the speed of light to the particle in question, *is* perfectly compatible with relativity because the speed of light is, for relativity, an invariant quantity.

And note that it is essentially the fixing of this problem with Newtonian gravitational theory – namely, removing the nonlocality and thereby making it compatible with relativity – that Albert Einstein accomplished in his *general* theory of relativity in 1915.

As a way of summarizing all of this discussion, here is a nice statement by Einstein:

> The success of the Faraday-Maxwell interpretation of electromagnetic action at a distance resulted in physicists becoming convinced that there are no such things as instantaneous

[4]How could a causal influence in electrodynamics ever go slower than light, given what we've been saying here? Well in some fundamental sense it can't. But it can zig-zag back and forth through space-time in such a way that each individual "zig" or "zag" goes at the speed of light, but the overall average speed between the beginning and the end is much slower. As an analogy, if you throw a ball at a wall and wait for it to bounce back and hit you on the head, there is a sense in which the causal influence always propagates at whatever speed the ball was moving... but another sense in which the average speed, from your throw to your getting bonked in the head, is zero. The same kind of thing is possible with charged particles replacing you and the wall, and electromagnetic influences propagating at the speed of light replacing the ball.

> action at a distance (not involving an intermediary medium) of the type of Newton's law of gravitation. According to the theory of relativity, action at a distance with the velocity of light always takes the place of instananeous action at a distance or of action at a distance with an infinite velocity of transmission. This is connected with the fact that the velocity c plays a fundamental role in this theory [5, p. 47].

1.4 Bell's Formulation of "Locality"

In the last section, we sketched the process by which one can extract – from the fundamental equations of Maxwellian electrodynamics – the theory's "relativistic local causality" (or just "locality" for short). The success of relativity theory, though, has strongly suggested that this sort of relativistic local causality (i.e., the idea that causal influences should never propagate faster than the speed of light) is a necessary property of *any* acceptable physical theory. So it will be useful, going forward, to have a "generic" formulation of this idea, i.e., a concept of relativistic local causality that is not tied to any particular theory like Maxwellian electromagnetism. We therefore discuss here the formulation provided by John Stewart Bell, whose contributions to the foundations of quantum theory will show up throughout this book.

Bell begins by noting that Maxwellian electrodynamics has the following property: a complete description of the state of all the fields and charges, on a time-like "slice" (Σ) across the past light cone of some event at $(\vec{x}, t)$, will *determine* what happens at $(\vec{x}, t)$. The image of a time-like "slice" is illustrated in Fig. 1.5. One should remember, though, that this language – "slice", suggesting a two-dimensional region

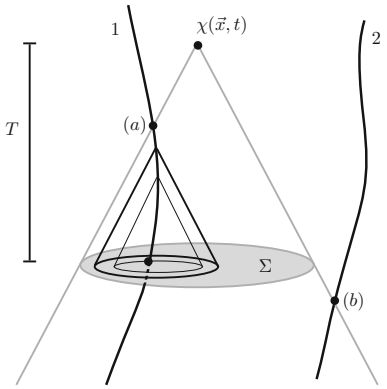


Fig. 1.5 In Maxwellian electrodynamics, a physical event χ at the point $(\vec{x}, t)$ – for example, the value of an $\vec{E}$ or $\vec{B}$ field, or the velocity of some particle that arrives there, or the electric charge density there – is uniquely determined by a complete specification ($\mathcal{C}_\Sigma$) of everything that's happening (i.e., the complete state of both fields and any charges and currents) in a horizontal "slice" through the backwards light cone of $(\vec{x}, t)$. The "slice" – Σ – is shown here as a shaded circle, although of course this really represents a three-dimensional spherical region of radius cT

that looks like the shaded circle in the Figure – is an artifact of the suppression of the third spatial dimension in these space-time diagrams. In fact, the shaded region in question represents a sphere of radius cT (where T is the time interval between the "slice" and the event in question at time t) centered at the event's spatial location $\vec{x}$.

Since, as we have argued in the previous section, causal influences in the theory always propagate at c (or slower), it's clear that all the causes of whatever happens at $(\vec{x}, t)$ must lie in this sphere. Nothing outside the sphere (at this earlier time) could *get* to $(\vec{x}, t)$ without propagating faster than light! But perhaps it is worth saying a little more about this to clarify how this idea connects with our earlier discussion.

To begin with, it is straightforward to see that there will be a contribution (corresponding to the complementary solution part of our earlier general solution) to fields at $(\vec{x}, t)$ from the electric and magnetic fields at the edge of Σ. In the three-dimensional picture, this corresponds to inward-propagating electromagnetic waves that will arrive at position $\vec{x}$ at time t. And then there will also be the contributions (corresponding to the particular solution from above) from electric charges and currents – the source terms – along the past light cone of $(\vec{x}, t)$. So, for example, in Fig. 1.5, what particle 1 is doing at the point marked (a) will influence what is happening at $(\vec{x}, t)$.

One might worry that "what particle 1 is doing at ... (a)" is not part of $\mathcal{C}_\Sigma$, our complete specification of events in Σ. But what particle 1 is doing at (a) is determined by $\mathcal{C}_\Sigma$. Think about it this way: what particle 1 is doing at (a) depends on where particle 1 was just prior to (a) and on the fields that influenced it at this earlier moment; these fields in turn depend on fields on the intersection of Σ with the past light cone of this earlier moment. See the thick black past-light-cone in the Figure. And then we can continue to step our way back along the world line of particle 1 in this same way: what it was doing at the earlier moment depends on where it was and what it was doing at an even-earlier moment, which in turn depends on the fields on the intersection of Σ with the past light cone of this even-earlier moment. (See the thin black past-light cone in the Figure.) And so on. You can then see how the whole structure of the world-line of particle 1 is determined, ultimately, by the state of particle 1 on Σ as well as the fields in a certain *part* of Σ that surrounds particle 1. And so, at the end of the day, *all* of the influences arriving at $(\vec{x}, t)$ – both direct ones and indirect ones – can indeed be traced back to physical facts about the states of the particles and fields on Σ. This is the sense in which any physical event $\chi(\vec{x}, t)$ is determined by $\mathcal{C}_\Sigma$ in Maxwellian electrodynamics.

(Note, by the way, that particles like 2 in the figure, whose world lines cross the past light cone of $(\vec{x}, t)$ prior to Σ are no problem: the influence of particle 2 on happenings at $(\vec{x}, t)$ coming from point (b) in the Figure are just "incoming electromagnetic waves" that we have already captured by including the state of the fields on the "edge" of Σ.)

In sum, by specifying the complete state of both fields and charged particles on the spacetime "slice" Σ, we have complete information about everything that is relevant to point $(\vec{x}, t)$ and therefore any physical fact pertaining to point $(\vec{x}, t)$ should be determined. We will formalize this by writing

$$\chi(\vec{x}, t) = f(\mathcal{C}_\Sigma) \tag{1.26}$$

where, again, χ is our generic name for some physical fact (like the value of some field or the velocity of some charged particle) at $(\vec{x}, t)$ and the right hand side means: some function of a complete specification of all the physical facts on Σ.

Equation (1.26) seems like it captures the idea of "relativistic local causality" for any *deterministic* theory. But remember, our goal here is to be completely generic. And as you probably already know, it has often been suggested that one of the lessons of quantum theory will turn out be that strict determinism must be abandoned. Bell proposed that we could modify the definition of locality, as follows, and arrive at something that still captures the idea of "relativistic local causality" but which is now applicable to both deterministic and indeterministic theories:

$$P\left[\chi_1 | \mathcal{C}_\Sigma\right] = P\left[\chi_1 | \mathcal{C}_\Sigma, \chi_2\right]. \tag{1.27}$$

This requires a bit of explanation. First of all, the left hand side means: the *probability* for some physical event χ_1 to occur at point 1, given a complete specification $\mathcal{C}_\Sigma$ of events on Σ (the "slice" through the past light cone). This is the same idea as before, but we just now speak of the probability of a certain event, rather than the event itself, since we don't want to presuppose that everything that happens is uniquely *determined* by events in the past light cone. (Of course, determinism is still allowed as a special case: in a deterministic theory, all of the probabilities will be either 1 or 0.) The right hand side is then meant to denote the probability assigned to the same event, but now with *both* $\mathcal{C}_\Sigma$ *and* some other event, χ_2 (which is at a point 2 which could not have been causally influenced by events in Σ) specified. See Fig. 1.6 for a picture.

After proposing (what we have here written as) Eq. (1.27), Bell writes:

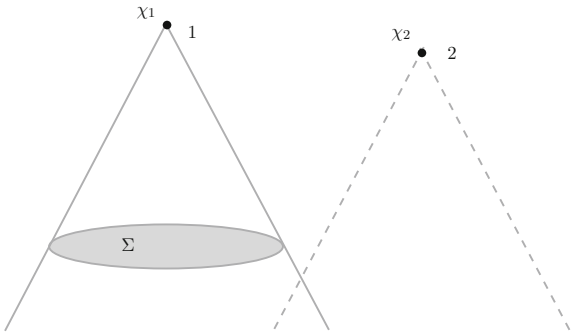


Fig. 1.6 According to Bell's generic definition of relativistic local causality, the probability of an event χ_1 at point 1, conditioned on a complete specification $\mathcal{C}_\Sigma$ of events in the region Σ (a "slice" across the backwards light cone of 1), should not be changed by specifying, in addition, some fact like χ_2 which cannot have been locally influenced by anything in Σ

> It is important that [Σ] completely shields off from 1 the overlap of the backward light cones of 1 and 2. And it is important that events in [Σ] be specified completely. Otherwise the traces in region 2 of causes of events in 1 could well supplement whatever else was being used for calculating probabilities about 1. The hypothesis is that any such information about 2 becomes redundant when [the state of things in Σ] is specified completely [6].

The importance of specifying events in Σ completely is pretty straightforward. If something is left out, then causal influences coming from the more distant past – influences which produce *correlations* between χ_1 and χ_2 – could result in χ_2 implying useful supplementary information (beyond that contained in the *incomplete* description of events in Σ).

But why is it so important that Σ should "shield" from 1 the overlapping past light cones of 1 and 2? In other words, why is it so important that the "other" event in Bell's formulation – χ_2 – be so far away that it could not have been causally influenced by events in Σ? See Fig. 1.7 and its caption for an explanation.

I hope that gives you a sense of why Eq. (1.27) seems to provide a good way of defining "relativistic local causality" in a completely general, a completely generic, way. It should come as no surprise that we will have occasion to use this formulation in later chapters.

Actually, a slight modification of Bell's formulation will also come in handy. To motivate this, one should understand that Bell's definition of locality basically amounts to the assertion that, once you provide a complete specification of events in a slice across its past-light-cone, the probability assigned to some physical event

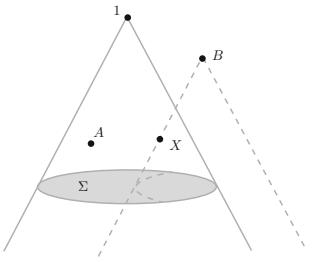


Fig. 1.7 In a non-deterministic theory, an event A may happen, despite not being determined to happen even by a complete specification of events in Σ, and then causally influence events at point 1. So specifying, in addition to $\mathcal{C}_\Sigma$, events like A that are in the past light cone of 1 but to the future of Σ, may indeed allow us to improve our predictions for events at 1. It might appear that, by contrast, events like B – which could be influenced by despite not being determined by $\mathcal{C}_\Sigma$ but which are outside the past light cone of 1 and hence could not locally influence events at point 1 – would *not* allow us to improve our predictions for events at 1. But this is incorrect: in an indeterministic theory, there might be an event X which is influenced (but not determined) by events in Σ, which then influences both B and happenings at 1. Specification of B can thus imply things about X which can in turn imply things about 1 that weren't already implied just on the basis of $\mathcal{C}_\Sigma$. This is why, in Bell's formulation of local causality, it is crucial that the other event (specification of which is not supposed to change the probability assigned to events at 1) should be outside the future light cone of Σ, i.e., this is why "[i]t is important that [Σ] completely shields off from 1 the overlap of the backward light cones of 1 and 2" [6]

should be *independent* of things happening outside the future-light-cone of that slice. Bell's formulation captures this independence by saying that the probability assigned to the event in question should not change depending on whether you do, or don't, specify such things.

But another way to capture this independence is by requiring that the probability be the same for any two different things that might be happening at this distant point:

$$P\left[\chi_1|\mathcal{C}_\Sigma, \chi_2\right] = P\left[\chi_1|\mathcal{C}_\Sigma, \chi_2'\right] \tag{1.28}$$

where χ_2 and χ_2' are two different possible versions of events at point 2 (in Fig. 1.6).

Let us illustrate all this with an example. It is not so interesting to take an example from Maxwellian electrodynamics, since that theory is manifestly local and hence respects both of our locality conditions: Bell's Equation (1.27) and our modification, Eq. (1.28). (You can try your hand at exploring this in the Projects if you are interested.) More interesting is seeing how our conditions can be used to diagnose the non-local character of Newtonian gravity. So let's take the following setup. There is one particle, of mass M, floating freely in empty space. There are two other particles, both of mass m, being held in place nearby (by some external non-gravitational forces) – one a certain distance to the left and the other the same distance to the right of our mass-M particle of interest. The symmetry implies that, if both of the mass m particles remain at rest, the net force on M will be zero and it will remain at rest. Whereas, if one of the mass-m particles is moved at the last second, the forces on M will no longer balance, and M will accelerate. See Fig. 1.8.

Thus, the probability for M to have an acceleration of zero at point 1 is different, depending on whether the mass-m particle on the right is moved, or not: Newtonian gravity explicitly violates Eq. (1.28). This is of course no surprise, but is a nice confirmation that Eq. (1.28) can be used to diagnose non-locality when it is present.

What about Eq. (1.27)? The problem is that, in a non-local theory, the left hand side, $P\left[\chi_1|\mathcal{C}_\Sigma\right]$, is simply not defined! Since the theory says the acceleration of the mass M particle depends on the locations of other distant particles at the moment in question, if we don't specify where the other distant particles are, there is no way for the theory to tell us what will (or might) happen to the particle in question. So although we expect that a genuinely local theory should respect both Eqs. (1.27) and (1.28), it is difficult to use Bell's formulation, Eq. (1.27), to explicitly diagnose the presence of non-locality in a theory. Our alternative formulation, Eq. (1.28), is nicer in this respect. It allows us to explicitly identify non-locality by seeing that the numbers on the two sides of the equation are different (rather than needing to try to compare something which isn't mathematically well-defined at all, to something that is). But whichever formulation we use, the non-locality of Newtonian gravity should be clear: the force on a particle (and hence its acceleration) depends on the location of distant particles at the exact moment in question.

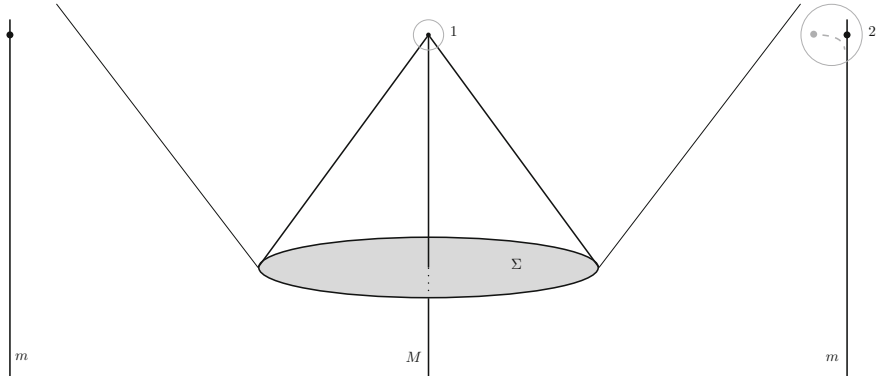


Fig. 1.8 A particle of mass M sits at rest between two particles of mass m that begin at equal distances from it in opposite directions. Let 1 be a space-time point through which the worldline of the mass-M particle will pass if everything remains as described. The complete specification, $\mathcal{C}_\Sigma$, of events on a slice, Σ, across the past-light-cone of 1, is rather simple: the particle is at rest at a certain point, and nothing else is going on! Take χ_1 to be the statement that the mass-M particle has an acceleration of zero at point 1. Suppose the mass-m particle on the *left* remains permanently fixed, but the mass-m particle on the *right* can either be *left* in place (call that χ_2), or pushed to the *left* (χ_2'). Then we have that $P\left[\chi_1|\mathcal{C}_\Sigma, \chi_2\right] = 1$. That is, if the mass-$m$ particle on the *right* remains fixed, the acceleration of the mass-M particle at point 1 will be zero with certainty. However, if the particle on the *right* is pushed to the *left*, the two gravitational forces on M will no longer add to zero, and M's acceleration at 1 will definitely not be zero: $P\left[\chi_1|\mathcal{C}_\Sigma, \chi_2'\right] = 0$. So we have a violation of Eq. 1.28

1.5 Ontology

"Ontology" is a fancy philosopher's word for "what really exists." In general, for the pre-quantum sorts of theories we've been looking at, "what really exists" according to the theory is fairly obvious and non-controversial. For example, according to Newtonian mechanics, the world is made of *particles* which move under the influence of (gravitational and probably other) forces they exert on one another. The picture is similar according to Maxwellian electrodynamics: the world is made of *particles* which interact with one another by means of the electric and magnetic *fields*. The ontology of Maxwellian electrodynamics, that is, includes particles and fields.

But there are often mathematically equivalent ways to formulate the basic laws of a theory, and this can sometimes raise questions about which formulation we should take seriously, as telling us in some relatively direct sense what really exists, physically, according to the theory.

For example, as you probably know, it is possible in electrodynamics to work with the so-called *potentials* (the scalar or electrostatic potential, and then the perhaps-less-familiar magnetic vector potential) instead of the *fields*. Let us briefly review some of this.

One of Maxwell's equations – the so-called "Gauss' Law for Magnetism" – reads $\vec{\nabla} \cdot \vec{B} = 0$. Since the divergence of any curl is identically zero, we can ensure that this equation is automatically satisfied if we introduce a magnetic vector potential $\vec{A}$ related to the magnetic field by

$$\vec{B} = \vec{\nabla} \times \vec{A}. \tag{1.29}$$

This allows us to re-write Faraday's Law as

$$\vec{\nabla} \times \left(\vec{E} + \frac{\partial \vec{A}}{\partial t} \right) = 0. \tag{1.30}$$

But then, because the curl of a gradient is identically zero, we can ensure that this equation is automatically satisfied if we introduce a "scalar potential" ϕ satisfying

$$\vec{E} + \frac{\partial \vec{A}}{\partial t} = -\vec{\nabla}\phi. \tag{1.31}$$

(The minus sign is just for convention/convenience.) Equivalently, if we write the electric field in terms of the scalar and vector potentials as

$$\vec{E} = -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t} \tag{1.32}$$

we guarantee the satisfaction of Faraday's Law. That then leaves the other two (non-homogeneous) Maxwell equations, which can now be re-written, in terms of ϕ and $\vec{A}$ as, respectively,

$$\nabla^2 \phi + \frac{\partial}{\partial t}(\vec{\nabla} \cdot \vec{A}) = -\frac{\rho}{\epsilon_0} \tag{1.33}$$

and

$$\nabla^2 \vec{A} - \frac{1}{c^2}\frac{\partial^2 \vec{A}}{\partial t^2} = -\mu_0 \vec{j} + \vec{\nabla}(\vec{\nabla} \cdot \vec{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t}). \tag{1.34}$$

These both look a little messy and complicated, but actually we have not yet taken full advantage of the freedom afforded by the potentials. In particular, since all that is really required of the potentials is that their various *derivatives* give the fields $\vec{E}$ and $\vec{B}$, there is an element of arbitrariness that we can leverage to make the equations look nicer.

For example, it's clear from the defining relation $\vec{B} = \vec{\nabla} \times \vec{A}$ that (since, again, the curl of a gradient is identically zero!) we could change $\vec{A}$ by the gradient of an arbitrary scalar function without affecting $\vec{B}$. That is,

$$\vec{A} \to \vec{A} + \vec{\nabla}\lambda \tag{1.35}$$

leaves $\vec{B}$ unchanged.

"But" (I can hear you saying) "changing $\vec{A}$ in this way *would* affect the electric field!" That's true, but we can "undo" that change by *also* requiring that, when $\vec{A}$ is shifted as in Eq. (1.35), we also shift the scalar potential ϕ as follows:

$$\phi \rightarrow \phi - \frac{\partial \lambda}{\partial t}. \tag{1.36}$$

It is easy to see that, if both ϕ and $\vec{A}$ are shifted in these ways, the fields $\vec{B}$ and $\vec{E}$ are unaffected. These "shifts" in the potentials are called "gauge transformations" and the idea is that there is a whole *equivalence class* of potentials (corresponding to all possible scalar functions λ) that correspond to the same field configurations.

This means that we are free to *choose* a particular set of potentials that makes some of the terms in Eqs. (1.33) and (1.34) disappear. We will briefly discuss two of these possible choices.

The first choice is to choose potentials satisfying the so-called "Lorentz gauge" condition:

$$\vec{\nabla} \cdot \vec{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} = 0. \tag{1.37}$$

With this choice, the wave equations satsified by ϕ and $\vec{A}$ take on the following particularly simple forms:

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\epsilon_0} \tag{1.38}$$

and

$$\nabla^2 \vec{A} - \frac{1}{c^2} \frac{\partial^2 \vec{A}}{\partial t^2} = -\mu_0 \vec{j}. \tag{1.39}$$

These are nice because they are precisely of the general form, Eq. (1.22), that we investigated earlier. It's clear, that is, that in Lorentz gauge, the effects of charges on the potentials propagate outward at the speed of light.

But here is another perfectly valid choice of gauges, the so-called "Coulomb gauge" in which

$$\vec{\nabla} \cdot \vec{A} = 0. \tag{1.40}$$

This turns out to imply the following dynamical equations (with source terms) for ϕ and $\vec{A}$:

$$\nabla^2 \vec{A} - \frac{1}{c^2} \frac{\partial^2 \vec{A}}{\partial t^2} = -\mu_0 \vec{j} + \frac{1}{c^2} \vec{\nabla} \frac{\partial \phi}{\partial t} \tag{1.41}$$

and

$$\nabla^2 \phi = -\rho / \epsilon_0. \tag{1.42}$$

Look in particular at the latter equation. It is just like Eq. (1.38), but with the propagation speed c set to infinity so that the term involving the second derivative with respect to time vanishes. So it implies that the scalar potential $\phi(\vec{x}, t)$ is determined by the *instantaneous* configuration of charges $\rho(\vec{x}, t)$ – in just the same way that the gravitational force on a given particle in Newtonian mechanics is determined by the instantaneous configuration of other particles. So the object ϕ – in Coulomb gauge

– is nonlocal! It changes *instantaneously*, without any speed-of-light delay, if some distant charge is wiggled.[5]

But so what? We should only be *bothered* by this kind of nonlocality, from the point of view of consistency with relativity, if we take ϕ as corresponding to something that is physically real. So does it? The usual answer is: no! The potentials ϕ and $\vec{A}$ are mere calculation devices – it is instead the fields $\vec{E}$ and $\vec{B}$ which we should take as directly corresponding to "physical stuff that really exists" according to the theory. To put this in terms of our notation from the previous section, it would be appropriate to let χ stand for (say) the value of $\vec{E}$ or $\vec{B}$ at some point, but it would not be appropriate to let χ represent the value of ϕ or $\vec{A}$ at some point. In order to function as intended, it is important that the χs in Bell's formulation of locality represent quantities that are endorsed, by the theory in question, as physically real.

Bell gives a memorable analogy here:

> The situation is further complicated by the fact that there *are* things which *do* go faster than light. British sovereignty is the classical example. When the Queen dies in London (long may it be delayed) the Prince of Wales, lecturing on modern architecture in Australia, becomes *instantaneously* King. (Greenwich Mean Time rules here.) And there are things like that in physics. In Maxwell's theory, the electric and magnetic fields in free space satisfy the wave equation
>
> $$\frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} - \nabla^2 \mathbf{E} = 0,$$
> $$\frac{1}{c^2}\frac{\partial^2 \mathbf{B}}{\partial t^2} - \nabla^2 \mathbf{B} = 0$$
>
> ...corresponding to propagation with velocity c. But the scalar potential, if one chooses to work in 'Coulomb gauge', satisfies Laplace's equation
>
> $$-\nabla^2 \phi = 0$$
>
> ...corresponding to propagation with infinite velocity. Because the potentials are only mathematical conveniences, and arbitrary to a high degree, made definite only by the imposition of one convention or another, this infinitely fast propagation of the Coulomb-gauge scalar potential disturbs no one. Conventions can propagate as fast as may be convenient. But then we must distinguish in our theory between what is convention and what is not [6].

This is a point that will occupy us considerably in the coming weeks. In quantum theory, which objects in the formalism are we supposed to take seriously, as corresponding to things that are physically real, and which are (like the scalar potential and British sovereignty!) bound up in some way with human knowledge or conventions? In particular, what is the ontological status of the quantum mechanical wave function?

[5] For a nice analysis of how, in Coulomb gauge, the electric field retains its local character even though the scalar potential is nonlocal, see Ref. [7].

1.6 Measurement

The two example theories we've been discussing are obviously regarded as good (if now slightly dated) scientific theories which have extremely favorable records of correctly predicting measured or observed phenomena in nature. It will be helpful to think a little bit about how, exactly, these theories achieve this status.

In the case of Newtonian mechanics, the situation is pretty straightforward. Newtonian mechanics is a theory about the motion of Particles – and, consequently, macroscopic assemblages of Particles which can in appropriate situations be treated as particles. Certain such particles are directly visible to us, so we can just check – by literally looking – to see if the particle is located where the theory says it will be located. If so then we say that the theory's prediction has been confirmed by observation.

For example, suppose you throw a ball up in the air from some initial height and with some initial speed. If you know something about the (say, gravitational and air drag) forces that will act on it, you can solve $\vec{F}_{\text{net}} = m\vec{a}$ for the ball and calculate, according to the theory, things like the maximum height the ball will reach and the time it will spend in the air before hitting the ground. In the case of the maximum height, the theory's prediction can perhaps be compared against a literal, perceptual observation: you just look and see how high the ball in fact goes before turning around and heading back down.

But this is, at best, pretty rough. A careful *measurement* of the ball's maximum height will require some additional sophistication. For example, you might set up some meter sticks and adjust your viewing perspective so that you can read off the maximum height to some precision by seeing exactly which mark on the meter stick lines up with the top of the ball at the moment it reaches its maximum height. One might for example also film the motion of the ball, with the meter stick in the background; looking through the individual frames later, one could identify the specific frame in which the ball reaches its maximum height and then make a more precise determination of that height using the image of the meter stick in the background. And of course more sophisticated techniques are also possible, but this indicates the overall pattern.

A measurement of, for example, the ball's time aloft will follow a similar pattern. Direct, unaided visual inspection perhaps gives some rough indication of the duration of the ball's flight, but a more precise measurement would involve additional care and equipment. For example, one might arrange for the ball's launch to trigger a stopwatch, whose second hand at that moment begins a rapid, steady rotation which is then triggered to halt when the ball strikes the ground. The final location of the second hand (which one may inspect at a convenient later time) then indicates the time the ball spent in the air.

The maximum height and time aloft of a launched ball are both things that are, in some sense, directly observable – we can "measure" them in a kind of rough and qualitative way by literally just watching the process unfold in real time. And then we have been discussing ways in which those rough perceptual observations

can be improved upon by using more sophisticated measuring equipment. There are, however, things that theories talk about which are not even in principle directly observable. In such cases, "measuring" or "observing" the fact in question (to, say, test a prediction of the theory) *requires* more sophisticated measuring equipment.

For example, suppose you want to test the Maxwellian electrodynamics prediction for what the electric field is between the two plates of a charged capacitor. You work out, based on the amount of charge that's present, etc., what the electric field should be according to the theory. How do you then measure this to see if the theory's prediction is correct? You certainly can't "just look and see".

But you can arrange for the electric field to leave its mark on something that you *can* just see. For example, you might stick a particle of mass m and electric charge q, at rest, at the place where you want to know the electric field. The particle will experience an electric force $q\vec{E}$ and will hence accelerate. If we can measure the particle's acceleration $\vec{a}$, we can then infer that the electric field was given by

$$\vec{E} = \frac{m}{q}\vec{a}. \tag{1.43}$$

Measuring the acceleration is, in turn, straightforward. For example, you could set a ruler next to the particle and observe the change $\Delta\vec{x}$ in its position after a short period of time Δt. Basic kinematics then tells us that

$$\vec{a} = \frac{2\Delta\vec{x}}{\Delta t^2}. \tag{1.44}$$

So, by measuring the position of a particle at two different times (and knowing the mass and charge of the particle) we are able to indirectly measure the electric field.

This puts us in a position to step back and see the point of this discussion. On the one hand, a theory that makes no testable predictions is in some sense clearly worthless. If you can't measure or observe *anything* that the theory says things about, the theory is fundamentally cut off from the structure of empirically-grounded knowledge and really has no meaningful content at all. But on the other hand, it is not the case that everything a theory says must somehow be directly observable. It is perfectly reasonable for theories to postulate the existence of "invisible" things (like electric and magnetic fields, atoms, neutrinos, etc.). But then the theory should provide a consistent account of the interactions of those postulated invisible things with other things (perhaps made of the invisible things!) that are visible, so that overall the theory connects with the given world of direct perceptual experience and thereby makes testable predictions.

As a kind of paradigmatic concretization of this idea, we will often speak of measurements having their outcomes registered in the position of a "pointer". One should think here of a sort of black-box measuring instrument, along the lines indicated in Fig. 1.9, whose internal mechanism provides a causal link between some physical quantity that is being (perhaps indirectly) measured (e.g., the time aloft of a ball, or the magnitude of the electric field at a certain point, or the energy of a neutrino, or ...)

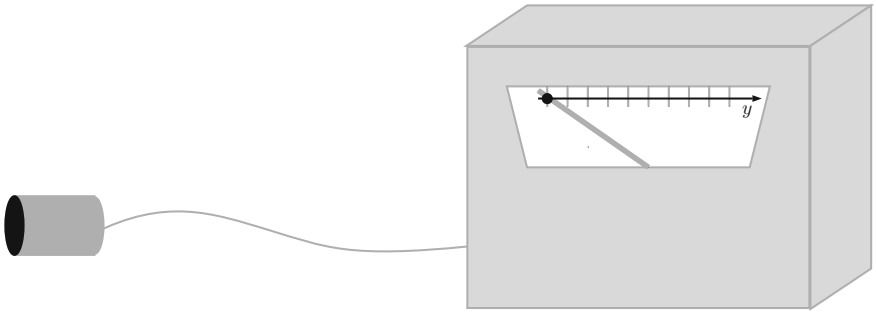


Fig. 1.9 A schematic measuring device whose probe end (on the *left*) can be arranged to interact with some (perhaps microscopic/invisible) system of interest. The outcome of the measurement is then registered by the position y of the device's pointer. This is a relatively accurate picture of how some real measuring devices work, but also captures in essentialized terms an important point about *any* kind of measurement: at the end of the day, the outcome is registered in the configuration of some directly-observable (macroscopic) object (e.g., the position of the hands of a stopwatch, the distribution of ink on a printout, etc.)

and the position of a pointer or needle that swings back and forth against a calibrated background to indicate the value in question.

This may seem like a very specialized kind of case, but actually it captures the essential idea of measurements quite generally. For example, the position of a ball (whose maximum height one wants to measure) is its own pointer. The second hand on the stopwatch functions as the pointer for measuring the ball's time aloft. The charged particle (whose final position allows one to determine its acceleration and hence the field that caused that acceleration) functions as the pointer for the measurement of the electric field that we discussed before. In general, *any* measurement must produce some effect in the configuration of some directly-observable macroscopic object, from whose final configuration we "read off" the result of the measurement. That is, any measurement must involve something that can be interpreted as a pointer.

From the point of view of assessing candidate theories, this discussion suggests several criteria. On the one hand, we should not demand too much from our theories. For example, we should not insist that everything postulated by a theory be somehow directly observable or measureable. Direct perception gives us *some* information about the structure of the world, but it does not give us *everything*; we should thus expect that, as science develops, theories should increasingly need to postulate invisible, microscopic objects, the empirical justification for which lies in the role of those postulated invisible objects in correctly accounting for the behavior of things (pointers!) that are directly visible.[6]

[6]Another way we might demand too much of a theory would be to demand that it not only account for the behavior of directly visible things like pointers, but that it somehow account for our conscious experiences of such things. The truth is that nobody understands how consciousness emerges and in particular how specific conscious experiences arise from the interactions among the external objects one perceives, one's perceptual apparatus (including eyes, brain, etc.), and the faculty of

On the other hand, we should also not demand too little from our theories. There *are* things – like the positions of "pointers" on lab equipment, whether or not a bomb has exploded, and the vitality of a certain cat – which are directly available in ordinary sense perception, and which any candidate account of the microscopic world should, in principle, be able to account for. Theories might validly postulate all kinds of crazy-sounding and counter-intuitive things, but at the end of the day, if a theory predicts the wrong thing for where pointers should point (or somehow cannot account for the existence of pointers that point at all) it cannot be correct.

1.7 Abstract Spaces

Let us raise one final point about the pre-quantum theories we've been discussing. In the case of both Newtonian mechanics and Maxwellian electrodynamics, everything we have talked about so far can be understood in terms of ordinary three-dimensional space (and/or four-dimensional space-time). The particles, for example, that are a central part of the ontology for these theories, "live" in three-dimensional space. As do the fields of Maxwellian electrodynamics. But it is worth pointing out that there are various mathematical re-formulations of some of these ideas, in which various sorts of "abstract spaces" are used.

For example, it is sometimes useful to use "phase space" to talk about the kinematics and dynamics of particles. For a single particle moving in three dimensional physical space, the phase space is a *six*-dimensional space whose axes correspond to the x, y, and z coordinates of the particle and also the three components of the particle's momentum: p_x, p_y, p_z. It is of course difficult to visualize a six-dimensional space, but if we consider a toy model system in which a particle is confined to move only along a single spatial dimension (say, x), then the phase space is two-dimensional (axes x and p_x) and we can easily visualize it.

Take, as a simple example, the case of a particle that is just moving inertially: $p_x =$ constant, so then $x = p_x t/m + x_0$. We can plot the "trajectory" of the particle through phase space; see Fig. 1.10.

A slightly more interesting case is a one-dimensional harmonic oscillator. The spatial coordinate x oscillates sinusoidally (say, about $x = 0$) and the momentum p_x also oscillates sinusoidally but out of phase with the position: the momentum is big (either positive or negative) when the position is zero, and vice versa. The "trajectory" of the particle through phase space is shown in Fig. 1.11. The "trajectory" is a closed orbit – an elliptical curve. See if you can figure out which direction the

(Footnote 6 continued)
consciousness that somehow emerges from brain structure and function (if, indeed, it is separable from these things at all). These are deep and difficult questions that are largely outside the scope of physics. As far as physics is concerned, if a theory accounts for the existence of macroscopic material objects which possess gross, coarse-grained properties consistent with what we are given in ordinary perception – i.e., if a theory gets the pointer positions right – we should regard it as perfectly acceptable and empirically adequate.

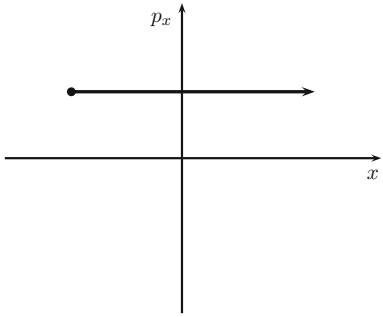
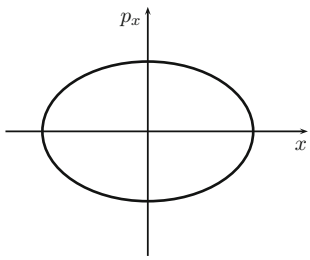


Fig. 1.10 The "trajectory" of a free particle (moving with constant momentum) through phase space: at $t = 0$ the particle begins, at a negative value of x, with a positive momentum which stays constant as the particle moves. So its "path" through phase space looks like the solid line and would continue indefinitely to the right as long as the particle continues moving with unchanged momentum

Fig. 1.11 The "trajectory" of a one-dimensional harmonic oscillator through phase space is a closed elliptical curve



phase space point moves around the ellipse. (Hint: when p_x is positive, is x increasing, or decreasing?)

Another type of "abstract space" that is sometimes useful to think about in the context of classical physics – and which we will confront again in the context of quantum mechanics – is "configuration space". For a single particle, configuration space is just the same as physical space: it is all the places the particle could be. But in a situation involving (say) two particles (moving, for simplicity, in one spatial dimension), there is the set of all places (call them x_1) where the first particle might be, and then also the set of all places (call them x_2) where the second particle might be. The configuration space is then the set of all possible configurations of the two particles jointly – that is, configuration space is a (here) two-dimensional space whose axes are x_1 and x_2.

Let's again consider two examples, one really simple and one at least slightly more interesting. Suppose, for example, that we have two particles moving in (the same) one-dimensional space. Except, let's say, they aren't actually moving, but are instead both just sitting there. In particular, particle 1 is sitting at a small positive value of x, and particle 2 is sitting at a somewhat larger negative value of x. Figure 1.12 shows this situation, both in (regular) physical space, and then again in configuration space.

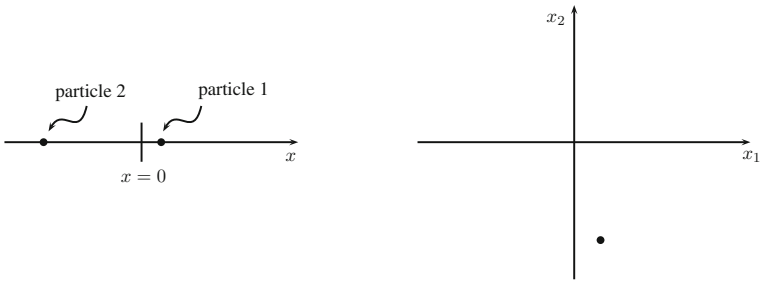


Fig. 1.12 On the *left* are two particles ("particle 1" and "particle 2") sitting at different points along the x-axis. This same situation is represented in the abstract configuration space as a single dot whose coordinates correspond to the positions (in physical space) of the particles

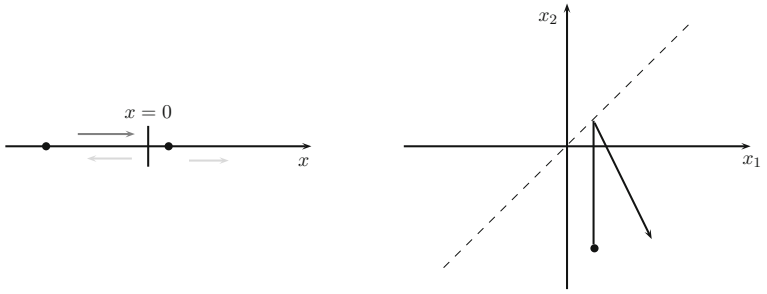


Fig. 1.13 The same two particles as before, but now particle 2 (which starts on the *left*) is given a kick at $t = 0$ so it moves to the *right* (its initial velocity is indicated by the *dark gray arrow*) until it collides with particle 1. The post-collision velocities are indicated with the *light gray arrows*. This same process is shown, on the *right*, in the abstract configuration space. Note the *dashed line* at $x_1 = x_2$ corresponding to the particles being at the same location, i.e., colliding

As a slightly more interesting example, suppose now that somebody comes in and gives particle 2 a kick so that it starts moving to the right. Eventually it runs into particle 1 and, let's suppose, they collide elastically. Suppose particle 1 has a greater mass than particle 2, so that after the collision particle 2 recoils back out to the left, whereas particle 1 drifts off slowly to the right. This process is depicted – both in physical space and in the abstract configuration space – in Fig. 1.13.

What is the point of discussing these abstract spaces? Well, as I said, they are sometimes useful ways to depict a certain physical process to gain some intuition. And there are even complete reformulations of Newtonian mechanics where things are formulated in terms of one of these abstract spaces – for example, if you've taken a course in classical mechanics, you have probably encountered "Hamiltonian mechanics" which is basically a way of re-writing $\vec{F}_{net} = m\vec{a}$ in terms of energy quantities. The basic dynamical equations in this Hamiltonian formulation of mechanics are first-order (in time) differential equations for the coordinates of a system in phase space. There is also something called the "Hamilton-Jacobi" formulation of mechanics which involves something like a time-dependent "field" on configuration space,

which influences or guides the actual configuration point (representing the positions of all the particles composing the system) through that space. We won't go into these things in any detail here, but it is good to be aware that they exist.

At this stage, the main take-home lesson from this discussion is this: don't confuse any of these abstract spaces with regular old physical space! For example, if you are studying a one-dimensional harmonic oscillator, and sketch its phase space trajectory as in Fig. 1.11, you should not ask: "What force provides the centripetal acceleration which holds the particle in this elliptical orbit?" Or similarly, you should not ask, about the particle collision process depicted in Fig. 1.13, "That dotted line in the picture that the particle bounced off of... what's it *made of*??" Those sorts of questions don't actually make any sense, and would seem to be based on simply forgetting that the space in question is an abstract one. If, in these kinds of situations, you want to know what is really going on, physically, you need to translate the abstract representation back into direct, literal, physical-space terms. For example, the dotted line in the previous Figure is not really a "thing" at all, but instead a kind of abstract representation of the strong repulsive forces that the two particles exert on each other if they get too close together in physical space. And the only force present in the case of the simple harmonic oscillator is the force exerted on the particle by a spring (or whatever) as it moves back and forth along a line – there is nothing like an elliptical orbit at all, if that means some literal material particle moving along a certain curved path through a two-dimensional physical space.

This of course all seems so clear and obvious as to be almost embarrassing to have to say. But as you'll see (especially in Chap. 5) confusion will arise around these kinds of issues when we get to quantum mechanics – to which we will turn very soon!

Projects:

(1.1) Show that a spherically-symmetric distribution of mass (e.g., a thin spherical shell composed of innumerable massive Particles) exerts the same gravitational force on an outside particle as the force that would be exerted by a single Particle with the same total mass as the shell and located at the center of the shell. Explain how this is an example of the theory explaining how and why a large (spherically symmetric) assemblage of Particles can be treated as a particle.

(1.2) Show that a rigid body (like a bunch of Particles glued together) will obey Newton's 2nd law ($\vec{F} = m\vec{a}$) if the individual particles do. Note that a rigid body can *rotate* in addition to moving translationally, so you should *not* assume that all the individual Particles have the same acceleration $\vec{a}$. Indeed, the main problem here is to figure out precisely how "the acceleration of the rigid body" can be defined to make something like Newton's 2nd law true. Hint: it might be helpful to consider the *center of mass* coordinate, $\vec{R} = \frac{1}{M_{\text{total}}} \sum_i m_i \vec{r}_i$, and its various time-derivatives. Explain how this is an example of the theory explaining how and why a large assemblage of Particles can be treated as a particle.

(1.3) Consider two equal-mass stars in a binary star system, each making a circular orbit about their mutual center-of-mass point. Now suppose that gravitational forces are given by Eq. (1.1), but with $\vec{r}_j$ being the position of the distant mass at a slightly earlier time (such that a signal emitted by it at that earlier time would just arrive at the mass in question now). Draw a careful diagram showing the gravitational forces acting on each star at some particular moment. Do the forces respect Newton's third law? Will the total momentum of the two-star system be conserved (assuming no forces act from the outside)? How about the total angular momentum? Can you think of another similar situation in which the total translational momentum would not be conserved? What do you make of all this?

(1.4) Give a simple example of a system of masses interacting via Newtonian gravitational forces, and show that/how the motion of the masses would be different if one used a different slicing of space-time into simultaneity slices. (That is, show that Newtonian mechanics with instantaneous, action-at-a-distance gravitational interactions is incompatible with relativity, since it requires a dynamically privileged notion of simultaneity.)

(1.5) The mathematical parallel between Newton's law of gravity and Coulomb's law of electrostatics suggests that a relativistic theory of gravity could be developed by, in effect, copying Maxwell's equations. Play around with this and see how far you can get. (Hint: the gravitational analog of the electric field $\vec{E}$ is the gravitational field $\vec{g}$, which has units of acceleration. So the gravitational analog of Gauss' Law should be something like $\vec{\nabla} \cdot \vec{g} \sim \rho_m$ where ρ_m is the mass density. What should the proportionality constant be in order to reproduce Newtonian gravity? After you figure out the gravitational analog to Gauss' Law, you can try to work out consistent gravitational analogs to the three other Maxwell equations as well!)

(1.6) Flesh out the equivalence between Coulomb's law and Gauss' law by explaining in detail how to solve for $\vec{E}$ in Gauss' law when $\rho = q\,\delta^3(\vec{x})$.

(1.7) Work through the details of deriving wave equations for $\vec{E}$ and $\vec{B}$ from Maxwell's equations. (Assume empty space, i.e., $\rho = 0$ and $\vec{j} = 0$.) Show that there are plane-wave solutions of the form $\vec{E}(\vec{x}, t) = \vec{E}_0 \sin(\vec{k} \cdot \vec{x} - \omega t)$, and similarly for $\vec{B}$. Are the waves transverse, or longitudinal? How do you know? What is the relationship between $|\vec{k}|$ and ω? What are the phase and group velocities of the waves in terms of ϵ_0 and μ_0?

(1.8) Let's try to understand better how Eq. (1.24) is a solution of the wave equation with a delta-function source. For simplicity, suppose the source point is at $\vec{x}' = 0$ and $t' = 0$ so that the differential equation in question reads

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = \delta^3(\vec{x}) \delta(t). \tag{1.45}$$

The solution ψ should be spherically symmetric, i.e., should be a function of r and t only. (a) Show that, for $r \neq 0$, any function of the general form

$$\psi(r,t) = \frac{g(t - r/c)}{r} \tag{1.46}$$

solves Eq. (1.45). (b) The correct function g should also satisfy Eq. (1.45) at $r = 0$. To ensure this, integrate both sides of Eq. (1.45) over a spherical volume of radius R centered at $r = 0$. Show that taking the $R \to 0$ limit then gives $-4\pi g(t) = \delta(t)$, i.e.,

$$\psi(r,t) = -\frac{1}{4\pi}\frac{\delta(t - r/c)}{r} \tag{1.47}$$

as claimed in the text.

(1.9) When he is presenting his formulation of locality, Bell gives an analogy to the boiling of an egg. You put the egg into the boiling water and set the timer for (say) 5 min. Then, 5 min later, "[t]he ringing of the alarm establishes the readiness of the egg." That is, the two events are *correlated*. However, the ringing of the alarm does not *affect* the egg. Bell explains: "if it is already given that the egg was nearly boiled a second before, then the ringing of the alarm makes the readiness no more certain." Draw a spacetime diagram; connect "the ringing of the alarm", "the readiness of the egg", and its being "already given that the egg was nearly boiled a second before" with the terminology χ_1, χ_2, and $\mathcal{C}_\Sigma$; and explain how the last sentence captures the locally causal character of the physical processes involved in this example.

(1.10) An unstable particle is heading for a particle detector which will "click" if the particle hits it. Given the state of the particle at some earlier time, suppose there is a 50% probability of its *not* decaying first and hence hitting the detector: $P\,[\text{click}|\mathcal{C}_\Sigma] = 1/2$. On the other hand, if the particle does decay before arriving at the detector, the decay products might themselves be detected and hence indicate that the original particle will not be detected by the original detector. So, for example, $P\,[\text{click}|\mathcal{C}_\Sigma, \chi_2] = 0$, where χ_2 denotes the successful detection of one of the decay products. See Fig. 1.14. Does the non-equality of the two conditional probabilities here imply a violation of Bell's locality condition? (One would hope not, since there is clearly nothing nonlocal happening here. On the other hand, this seems to be a case where information from outside the past light cone of the event in question, does affect the probabilities assigned to that event.) Explain.

(1.11) Make up an example, maybe along the lines of the example involving Newtonian gravity from the main text, to show that Maxwellian electrodynamics gets (correctly) diagnosed as "local" both by Bell's formulation and the modified formulation.

(1.12) Is there a way of choosing a gauge such that the vector potential $\vec{A}$ propagates with infinite speed, the way ϕ does in the Coulomb gauge? Explain.

(1.13) A standard introductory physics problem involves analyzing the "ballistic pendulum" in which a block of mass M hanging from a string of length L absorbs an incoming bullet of mass m moving at some unknown initial speed

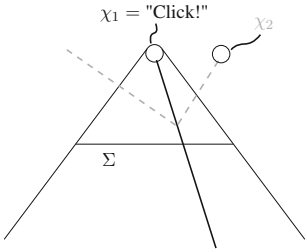


Fig. 1.14 Space-time diagram for the events described in Project 1.10

v_0. The bullet-and-block then swing up together, with the string eventually making some maximum angle θ with respect to the vertical. By observing θ one can thereby determine the initial speed of the bullet. Work out this relationship (i.e., solve for v_0 as a function of θ, m, M, and L) and explain how this method of measuring the bullet's speed fits into the general scheme introduced in the text in which the outcome is registered in some directly observable "pointer".

(1.14) Pick a measuring apparatus of interest to you (maybe something that you've used in a physics lab course or research, maybe the accelerometer in your iPhone, or just something else you're interested in) and learn more about how it actually works. Explain whether (and, assuming so, how) the actual device fits into the general scheme introduced in the text in which the outcome is registered in some directly observable "pointer".

(1.15) Two equal-mass gliders are floating on a (nearly) frictionless air track in an introductory physics classroom. The track is equipped with elastic bumpers on both ends, and the two gliders have elastic-collision attachments so they will bounce when they collide. One glider is initially at rest, while the other is given an initial velocity. Draw a space-time diagram showing world lines for both gliders as well as the two bumpers on the ends of the track. Now draw the "trajectory" of the two-glider system through the two-dimensional configuration space.

References

1. I. Newton, *Opticks* (Dover Publications Inc, New York, 1952)
2. J. Andrew, Newton's Philosophy, in *The Stanford Encyclopedia of Philosophy*, ed. by E.N. Zalta (2009), http://plato.stanford.edu/archives/win2009/entries/newton-philosophy/
3. A. Einstein, Autobiographical Notes, in *Albert Einstein: Philosopher-Scientist*, ed. by P.A. Schilpp (Harper and Row, New York, 1959)
4. J.D. Jackson, *Classical Electrodynamics*, 2nd edn. (Wiley, New York, 1975)
5. A. Einstein, *Relativity: The Special and the General Theory* (Penguin Classics, New York, 2006)
6. J.S. Bell, La Nouvelle Cuisine, reprinted in *Speakable and Unspeakable in Quantum Mechanics*, 2nd edn. (Cambridge University Press, Cambridge, 2004)
7. O.L. Brill, B. Goodman, Am. J. Phys. **35**, 832–837 (1967)

Chapter 2
Quantum Examples

In this chapter we review quantum theory (at the level of wave mechanics) and develop a toolbox of simple quantum mechanical examples that we will use, in the following chapters, to discuss a number of the issues raised in Chap. 1: locality, ontology, measurement, etc.

2.1 Overview

We begin with the (time-dependent) Schrödinger Equation,

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}\Psi. \tag{2.1}$$

For a single particle of mass m moving in one dimension, the Hamiltonian operator $\hat{H}$ is

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x) \tag{2.2}$$

so that the time-dependent Schrödinger Equation reads:

$$i\hbar\frac{\partial\Psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + V(x,t)\Psi(x,t). \tag{2.3}$$

We typically use this equation in the following sort of way: given some initial wave function $\Psi(x,0)$ (which we think of as having been created by a special *preparation* of the particle in question), we then solve Schrödinger's Equation to find the wave function at some later time t when some kind of observation or measurement of the particle occurs. A second basic postulate of the theory – the "Born rule" – is then introduced to tell us what $\Psi(x,t)$ implies about how the measurement will come

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_2

out. In its simplest form, corresponding to a measurement of the position x of the particle, the Born rule says that the probability of observing the particle at the point x is equal to the square of the wave function's modulus:

$$P(x) = |\Psi(x,t)|^2 . \tag{2.4}$$

(Well, technically, this $P(x)$ is a probability *density*, i.e., a probability-per-unit-length along the x axis. The precise statement is that the probability of finding the particle in a range of size dx near the point x is $P(x)dx = |\Psi|^2 dx$. Note also that we assume here that the wave function is properly normalized, i.e., $\int_{-\infty}^{\infty} |\Psi|^2 dx = 1$.)

If some property of the particle other than its position is measured (for example, its momentum or energy) then we will use the "generalized Born rule". This says that we should write $\Psi(x,t)$ as a linear combination of *eigenstates* of the operator corresponding to the property in question. That is, we should write

$$\Psi(x,t) = \sum_i c_i \Psi_i(x,t) \tag{2.5}$$

(or perhaps instead an integral if the property in question has a continuous spectrum) where Ψ_i is an eigenstate of the operator $\hat{A}$ with eigenvalue A_i:

$$\hat{A}\Psi_i(x,t) = A_i \Psi_i(x,t). \tag{2.6}$$

Then the probability that the measurement of A yields the value A_i is

$$P(A_i) = |c_i|^2. \tag{2.7}$$

As a simple example, suppose we measure the *momentum* of a particle at time t. The momentum operator is

$$\hat{p} = -i\hbar \frac{\partial}{\partial x} \tag{2.8}$$

whose eigenstates are the plane waves

$$\psi_p(x) = e^{ipx/\hbar}. \tag{2.9}$$

(Note that there's a bit of funny business about normalization here, but let's ignore that for now.) Suppose our wave function at time t is $\Psi(x,t) = \sqrt{2}\sin(kx)$. We can write this as a linear combination of the momentum eigenstates as follows:

$$\Psi(x,t) = \sqrt{2}\sin(kx) = \frac{1}{\sqrt{2}i} e^{i(\hbar k)x/\hbar} - \frac{1}{\sqrt{2}i} e^{i(-\hbar k)x/\hbar}. \tag{2.10}$$

This is a linear combination of two momentum eigenstates, with eigenvalues $p = +\hbar k$ and $p = -\hbar k$, and expansion coefficients $c_{+\hbar k} = 1/\sqrt{2}i$ and $c_{-\hbar k} = -1/\sqrt{2}i$

respectively. So evidently the probability that the momentum measurement has the outcome "$p = +\hbar k$" is $P(+\hbar k) = |1/\sqrt{2}i|^2 = 1/2$, and the probability that the momentum measurement has the outcome "$p = -\hbar k$" is $P(-\hbar k) = |-1/\sqrt{2}i|^2$ which is also $1/2$.

Note that the original Born rule (for position measurements) can be understood as a special case of the "generalized Born rule" if we take $\hat{A}$ to be the position operator $\hat{x}$ with delta functions as eigenstates:

$$\hat{x}\delta(x - x') = x'\delta(x - x'). \tag{2.11}$$

We can then write any arbitrary state $\Psi(x, t)$ as a linear combination of position eigenstates as follows:

$$\Psi(x, t) = \int \Psi(x', t)\delta(x - x')\, dx' \tag{2.12}$$

where the $\Psi(x', t)$ should be understood as the expansion coefficient, like c_i. Thus, according to the generalized Born rule, the probability for a position measurement to yield the value x' should be the square of the expansion coefficient, i.e.,

$$P_t(x') = |\Psi(x', t)|^2 \tag{2.13}$$

just like in the original statement of the Born rule.

But enough about measurement. For now I just want to make sure you had heard of this so that you understand, in some practical terms, what solving the Schrödinger equation is *for*. There are a few Projects at the end of the chapter that will help you practice using Born's rule and then we will return to discuss all of this more critically in Chap. 3.

For now, we return to Schrödinger's equation. Given an initial wave function $\Psi(x, 0)$, how does one actually solve it? Our standard technique will take advantage of the fact that there exist "separable" solutions of the form

$$\Psi_n(x, t) = \psi_n(x) f_n(t). \tag{2.14}$$

If we plug this ansatz into the Schrödinger equation we find that the function $\psi_n(x)$ should satisfy the "time-independent Schrödinger equation" (TISE),

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi_n(x)}{\partial x^2} + V(x)\psi_n(x) = E_n\psi_n(x) \tag{2.15}$$

where E_n is just a constant that we can think of as the energy of the solution in question. The function $f_n(t)$ in turn satisfies

$$i\hbar\frac{df_n(t)}{dt} = E_n f_n(t) \tag{2.16}$$

which we can solve once and for all right now:

$$f_n(t) = e^{-iE_n t/\hbar}. \tag{2.17}$$

Now we can explain our basic strategy. For a given potential energy function $V(x)$, we solve the TISE to find the "energy eigenstates" $\psi_n(x)$ and corresponding energy eigenvalues E_n. If we can find a way to write the given initial wave function as a linear combination of these "energy eigenstates", as in

$$\Psi(x, 0) = \sum_n c_n \psi_n(x) \tag{2.18}$$

then we can construct a solution of the full time-dependent Schrödinger equation by simply tacking the appropriate time-dependent $f_n(t)$ factor onto each term in the sum. That is:

$$\Psi(x, t) = \sum_n c_n \psi_n(x) e^{-iE_n t/\hbar}. \tag{2.19}$$

This is the basic technique we'll now illustrate with a couple of examples.

2.2 Particle-in-a-Box

Suppose that a particle is absolutely confined to a certain region of the x-axis but is "free" within that region. That is, suppose

$$V(x) = \begin{cases} 0 & \text{for } 0 < x < L \\ \infty & \text{otherwise} \end{cases} \tag{2.20}$$

which we can (only somewhat misleadingly) think of as the particle being confined to a length-L "box" as illustrated in Fig. 2.1.

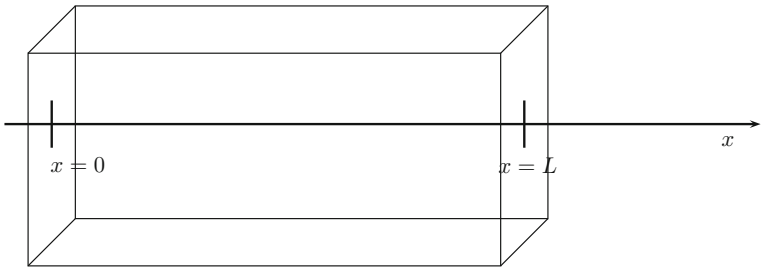


Fig. 2.1 The length-L "box" that our "particle in a box" is confined to

Outside the box, where $V = \infty$, we need $\psi = 0$. Inside the box, where $V = 0$, the TISE takes on the simple form:

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} = E\psi \tag{2.21}$$

whose solution is

$$\psi(x) = A\sin(kx) + B\cos(kx). \tag{2.22}$$

But since the potential V goes to infinity abruptly at $x = 0$ and $x = L$, the only way the TISE will be solved for all x (including $x = 0$ and $x = L$) is if $\psi(x) = 0$ at $x = 0$ and $x = L$. Requiring $\psi = 0$ at $x = 0$ means that we cannot have any of the cosine term, i.e., $B = 0$. And then requiring $\psi = 0$ at $x = L$ puts a constraint on the wave number k: an integer number of half-wavelengths must fit perfectly in the box, i.e., $k = k_n$ where

$$k_n = \frac{n\pi}{L} \tag{2.23}$$

with $n = 1, 2, 3, \ldots$. Good. So the "energy eigenfunctions" for the particle-in-a-box potential take the form

$$\psi_n(x) = \sqrt{\frac{2}{L}}\,\sin\left(\frac{n\pi x}{L}\right). \tag{2.24}$$

Note that the factor out front comes from requiring normalization: $\int_0^L |\psi_n(x)|^2 dx = 1$.

We can find the corresponding energy eigenvalues by plugging ψ_n into Eq. (2.15). The result is

$$E_n = \frac{\hbar^2\pi^2 n^2}{2mL^2}. \tag{2.25}$$

The general solution of the time-dependent Schrödinger equation can thus be written

$$\Psi(x,t) = \sum_{n=1}^{\infty} c_n\psi_n(x)e^{-iE_nt/\hbar}. \tag{2.26}$$

Let's illustrate with a simple example. Suppose the wave function of a particle in a box is given, at $t = 0$, by

$$\Psi(x,0) = \frac{1}{\sqrt{2}}\psi_1(x) + \frac{1}{\sqrt{2}}\psi_2(x). \tag{2.27}$$

This function is plotted in Fig. 2.2. Notice that there is constructive interference between ψ_1 and ψ_2 on the left hand side of the box, giving rise to a Ψ with a large modulus there, but (partial) destructive interference on the right. So at $t = 0$ the

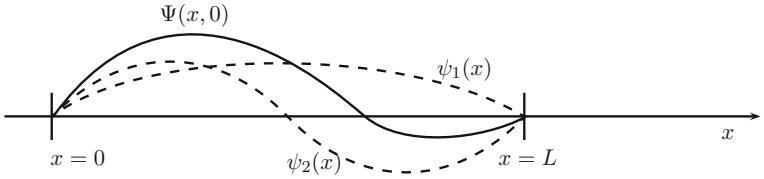


Fig. 2.2 The initial wave function (*solid curve*) for a particle-in-a-box that is in a superposition of the two lowest energy eigenstates (shown individually as *dashed curves*)

particle is much more likely to be found (if looked for!) on the left hand side of the box.

How, then, does Ψ evolve in time? Here we don't have to do any work to write the initial wave function as a linear combination of energy eigenstates. Equation (2.27) already gave it to us in that form! So then it is trivial to write down an equation for the wave function at time t:

$$\Psi(x,t) = \frac{1}{\sqrt{2}}\psi_1(x)e^{-iE_1t/\hbar} + \frac{1}{\sqrt{2}}\psi_2(x)e^{-iE_2t/\hbar} \tag{2.28}$$

or, writing everything out in full explicit glory,

$$\Psi(x,t) = \frac{1}{\sqrt{L}}\sin\left(\frac{\pi x}{L}\right)e^{-i\hbar\pi^2t/2mL^2} + \frac{1}{\sqrt{L}}\sin\left(\frac{2\pi x}{L}\right)e^{-4i\hbar\pi^2t/2mL^2}. \tag{2.29}$$

Notice that each term has an $e^{-iE_nt/\hbar}$ factor, but that the frequencies ($\omega_n = E_n/\hbar$) are different for the two terms. In particular, the frequency of the $n = 2$ term is four times as big as the frequency of the $n = 1$ term. And so, for example, in a time $t = T_1/2 = \pi/\omega_1 = \hbar\pi/E_1 = 2mL^2/\hbar\pi$ equal to half the period of the $n = 1$ factor (so that the $n = 1$ factor is -1), the $n = 2$ factor will have gone through two complete oscillations and will therefore be back to its original value of unity. At this time, the overall wave function will therefore look like the one shown in Fig. 2.3: there will now be constructive interference (and hence a high probability of finding the particle) *on the right*.

Thus, already in this simple example, we see an interesting non-trivial dynamics: the (probability of finding the) particle in some sense "sloshes back and forth" in the box.

2.3 Free Particle Gaussian Wave Packets

Let us now turn our attention to a second simple example – the "free particle". This is the same as the particle-in-a-box, but with the edges of the box (which in the previous section were at $x = 0$ and L) pushed back to $\pm\infty$. So in principle we could jump in

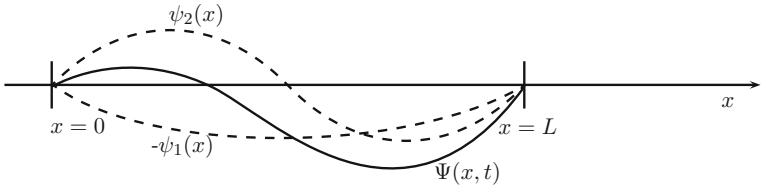


Fig. 2.3 The wave function for the same situation but after a time equal to half the period of the $n = 1$ state. The frequency of the $n = 2$ state is four times higher, so in the same amount of time that makes $e^{-iE_1t/\hbar} = -1$, we have that $e^{-iE_2t/\hbar} = +1$. So the $n = 2$ term in the superposition looks the same as it did at $t = 0$, but the $n = 1$ term is now "upside down". This produces destructive interference on the *left* and constructive interference on the *right*

by saying that the general solution to the TISE is just Eq. (2.22) again:

$$\psi(x) = A\sin(kx) + B\cos(kx) \tag{2.30}$$

but where now there is no reason that B needs to be zero, and no constraint at all on the wave number k.

This would be fine, actually, but it turns out to be a little nicer to instead use the so-called plane-wave states

$$\psi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx} \tag{2.31}$$

which are also perfectly good solutions of the free-particle TISE and which, as mentioned earlier, can be understood as eigenstates of the momentum operator $\hat{p} = -i\hbar\partial/\partial x$ with eigenvalue $p_k = \hbar k$. They are also of course energy eigenstates with

$$E_k = \frac{p_k^2}{2m} = \frac{\hbar^2 k^2}{2m}. \tag{2.32}$$

(Note that – like the "position eigenstates" we mentioned earlier in the chapter – these momentum eigenstates are not properly normalized, and indeed not technically normalizable at all! As long as we include the pre-factor of $\frac{1}{\sqrt{2\pi}}$ in our definition of the ψ_k states, however, their normalization is in a certain sense consistent with the normalization of the δ-function position eigenstates, and we won't run into trouble.)

Let's again focus on a concrete example: suppose that the wave function of a free particle is initially given by the Gaussian function

$$\Psi(x, 0) = Ne^{-x^2/4\sigma^2} \tag{2.33}$$

where N is a normalization constant. What, exactly, N is is not that important, but it will be a useful exercise to calculate it here. The idea is to choose N so that

$$1 = \int |\Psi(x,0)|^2\, dx = |N|^2 \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\, dx. \tag{2.34}$$

There is a cute trick to evaluate Gaussian integrals like that appearing here on the right hand side. Let's define the "standard Gaussian integral" as

$$J = \int_{-\infty}^{\infty} e^{-Ax^2}\, dx. \tag{2.35}$$

Then we can write J^2 as a double integral like this:

$$\begin{aligned} J^2 &= \left(\int_{-\infty}^{\infty} e^{-Ax^2}\, dx\right)^2 \\ &= \left(\int_{-\infty}^{\infty} e^{-Ax^2}\, dx\right)\left(\int_{-\infty}^{\infty} e^{-Ay^2}\, dy\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-Ax^2} e^{-Ay^2}\, dx\, dy. \end{aligned} \tag{2.36}$$

One can think of this as integrating a two-dimensional Gaussian function over the entire $x-y-$plane. But we can rewrite this integral using polar coordinates ($r^2 = x^2 + y^2$) as follows:

$$J^2 = \int_0^{\infty} e^{-Ar^2} 2\pi r\, dr. \tag{2.37}$$

But then this integral can be done (using a substitution, $u = Ar^2$, so $2\pi r\, dr = \pi\, du/A$) to give

$$J^2 = \frac{\pi}{A} \tag{2.38}$$

so that

$$J = \sqrt{\frac{\pi}{A}}. \tag{2.39}$$

Using this general result to simplify the right hand side of Eq. (2.34) gives

$$1 = |N|^2 \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\, dx = |N|^2 \sigma\sqrt{2\pi}. \tag{2.40}$$

so

$$|N| = \frac{1}{\sqrt{J}} = \frac{1}{\sqrt{\sigma\sqrt{2\pi}}}. \tag{2.41}$$

We might as well choose N to be real and positive, so now we know how to write a properly-normalized Gaussian initial wave function:

$$\Psi(x,0) = \frac{1}{\sqrt{\sigma\sqrt{2\pi}}} e^{-x^2/4\sigma^2}. \tag{2.42}$$

But of course the real question is: how does this state evolve in time?

To find out, we need to follow the general procedure: write the initial state $\Psi(x,0)$ as a linear combination of the energy eigenstates, then just tack on the appropriate $e^{-iEt/\hbar}$ factor to each term in the linear combination.

OK, so, first step: write the initial state as a linear combination of the energy eigenstates. Here there is a continuous infinity of energy eigenstates (parameterized by the wave number k) so the linear combination will involve an integral rather than a sum. It should look like this:

$$\Psi(x,0) = \int_{-\infty}^{\infty} \phi(k) \frac{e^{ikx}}{\sqrt{2\pi}}\, dk. \tag{2.43}$$

The (continuously infinite collection of!) numbers $\phi(k)$ are the "expansion coefficients". How do we find them? One way is to recognize that the last equation says: $\phi(k)$ is just the Fourier transform of $\Psi(x,0)$. So if that's a familiar thing, there you go! If not, though, here's how to extract them. This procedure is sometimes called "Fourier's trick". The idea is to use the fact that the different energy eigenstates (here, the plane waves) are *orthogonal* in the following sense: if you multiply one of them (k) by the complex conjugate of a different one (k') the result is oscillatory and its integral is zero – unless $k = k'$ in which case the product is just 1 and you get a giant infinity. Formally:

$$\int \left(\frac{e^{ikx}}{\sqrt{2\pi}}\right)\left(\frac{e^{ik'x}}{\sqrt{2\pi}}\right)^* dx = \delta(k-k'). \tag{2.44}$$

We can use this property to isolate the expansion coefficients $\phi(k)$ in Eq. (2.43). Just multiply both sides by $e^{-ik'x}/\sqrt{2\pi}$ and then integrate both sides with respect to x. The result of the x-integral on the right is a delta function that we can use to do the k-integral. When the dust settles, the result is

$$\phi(k') = \int \left(\frac{e^{-ik'x}}{\sqrt{2\pi}}\right) \Psi(x,0)\, dx. \tag{2.45}$$

So far we have avoided plugging in our Gaussian state for the initial wave function so this result is completely general. But let's now plug in Eq. (2.42) and proceed as follows:

$$\begin{aligned}
\phi(k) &= \frac{1}{\sqrt{2\pi}} \int e^{-ikx} \Psi(x,0)\, dx \\
&= \frac{N}{\sqrt{2\pi}} \int e^{-ikx} e^{-x^2/4\sigma^2}\, dx \\
&= \frac{N}{\sqrt{2\pi}} \int e^{-\frac{1}{4\sigma^2}(x^2+4ik\sigma^2 x)}\, dx \\
&= \frac{N}{\sqrt{2\pi}} \int e^{-\frac{1}{4\sigma^2}[x^2+4ik\sigma^2 x+(2ik\sigma^2)^2-(2ik\sigma^2)^2]}\, dx \\
&= \frac{N}{\sqrt{2\pi}} \int e^{-\frac{1}{4\sigma^2}[(x+2ik\sigma^2)^2-(2ik\sigma^2)^2]}\, dx \\
&= \frac{N}{\sqrt{2\pi}} e^{\frac{(2ik\sigma^2)^2}{4\sigma^2}} \int e^{-\frac{1}{4\sigma^2}(x+2ik\sigma^2)^2}\, dx \\
&= \frac{N}{\sqrt{2\pi}} e^{-k^2\sigma^2} \int e^{-x^2/4\sigma^2}\, dx
\end{aligned} \tag{2.46}$$

which is just another "standard Gaussian integral". Using our general formula to perform it, we arrive at:

$$\phi(k) = \sqrt{2} N \sigma\, e^{-k^2\sigma^2}. \tag{2.47}$$

Qualitatively, the important thing here is that the Fourier transform of a Gaussian (in x) is another Gaussian (in k). And notice in particular that the "width" of $\phi(k)$ is (something like) $1/\sigma$ – the inverse of the width of the original Gaussian in position space. This illustrates an important qualitative principle of Fourier analysis, that to make a sharply peaked function in position space requires superposing plane-waves with a very broad range of wave numbers, whereas you only need a narrower range of wave numbers to construct a more spread out function in position space. In the context of quantum mechanics this idea is intimately related to the Heisenberg uncertainty principle: the width of a wave packet in position space is essentially "the uncertainty in its position", Δx, whereas the "width" in k-space is (since $p = \hbar k$) "the uncertainty in its momentum" divided by $\hbar$. These being inverses of each other therefore means that the position uncertainty and the momentum uncertainty are inversely related: $\Delta x \sim \hbar/\Delta p$.

See the end-of-chapter Projects for some further (and more careful) exploration of this connection.

Let's step back and remember why we're doing all this math. We want to start with a nice Gaussian wave packet and see how it evolves in time according to Schrödinger's equation. To do that, we needed to first figure out how to write the initial Gaussian packet as a superposition of the energy eigenstates – here, the plane-wave states. That's what we've just accomplished! That is, we figured out that we can write

$$\Psi(x,0) = \int \phi(k) \frac{e^{ikx}}{\sqrt{2\pi}}\, dk \tag{2.48}$$

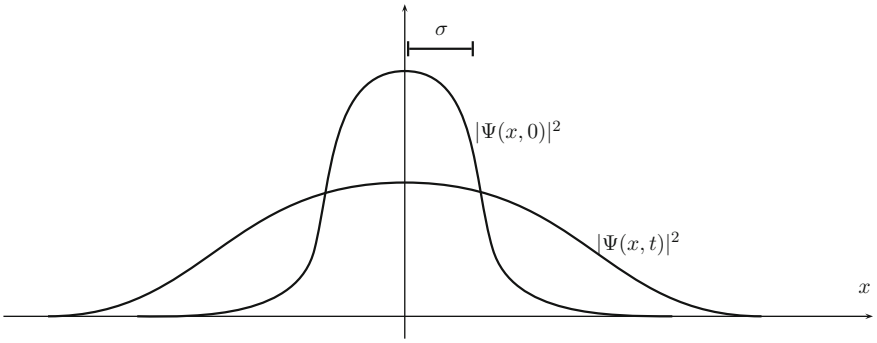


Fig. 2.4 A wave function that is a Gaussian with half-width σ at $t = 0$ spreads out in time

where

$$\phi(k) = \sqrt{2}N\sigma\, e^{-k^2\sigma^2}. \tag{2.49}$$

Now the whole reason we wanted to write $\Psi(x, 0)$ in this special form, is that doing so makes it easy to write down an equation for the state at a later time t: we just tack the $e^{-iEt/\hbar}$ factor on each term. So let's do that! The result is:

$$\begin{aligned}\Psi(x,t) &= \int \phi(k)\,\frac{e^{ikx}}{\sqrt{2\pi}}\,e^{-iE_k t/\hbar}\,dk \\ &= \frac{\sigma N}{\sqrt{\pi}}\int e^{-k^2\sigma^2}e^{ikx}e^{-i\hbar k^2 t/2m}\,dk.\end{aligned} \tag{2.50}$$

Now, with the same sort of massaging we did before ("completing the square" in the argument of the exponential, etc.) we can do this integral. I'll leave that as a Project if you want to go through it and just quote the result here:

$$\Psi(x,t) = N\frac{\sigma}{\sqrt{\sigma^2 + \frac{i\hbar t}{2m}}}\; e^{-\frac{x^2}{4(\sigma^2 + i\hbar t/2m)}}. \tag{2.51}$$

Phew!

This function is Gaussian-ish... You could think of it as a Gaussian with a time-dependent, complex width (whatever that means!). But if you multiply it by its complex conjugate, to get the probability density for finding the particle, that is definitely Gaussian:

$$P_t(x) = |\Psi(x,t)|^2 = \frac{N^2}{\sqrt{1 + \frac{\hbar^2 t^2}{4m^2\sigma^4}}}\exp\left[\frac{-x^2}{2\sigma^2\left(1 + \frac{\hbar^2 t^2}{4m^2\sigma^4}\right)}\right]. \tag{2.52}$$

Notice in particular that the *width* of this Gaussian (i.e., the uncertainty in the position of the particle) grows with time:

$$\Delta x(t) = \sigma\sqrt{1 + \frac{\hbar^2 t^2}{4m^2\sigma^4}}. \tag{2.53}$$

See Fig. 2.4 for an illustration. Initially (for times small compared to $2m\sigma^2/\hbar$) the width grows slowly, but then later (for times long compared to $2m\sigma^2/\hbar$ the width grows linearly in time. So the uncertainty in the position of the particle grows and grows as time evolves. Interestingly, the uncertainty in the momentum never changes: the first line of Eq. (2.50) can be understood as saying that the complex *phases* of the momentum "expansion coefficients" change with time, but their magnitudes stay the same. So the probability distribution for momentum values, and hence Δp, is independent of time. This makes sense, if you think about it, since we're talking about a free particle, i.e., a particle on which no forces act. Anyway, this nicely illustrates the fact that the Heisenberg uncertainty principle takes the form of an inequality: the product of Δx and Δp can be arbitrarily large, but there's a smallest possible value.

2.4 Diffraction and Interference

The spreading of an initial wave packet is closely related to the phenomenon of *diffraction*. Imagine, for example, a particle that is incident on a barrier with a slit: the barrier simply blocks/absorbs the part of the particle's wave function that hits it, so that just downstream of the barrier and along the direction transverse to the direction of propagation of the particle, the wave function has a packet-shaped profile like the ones we were discussing in the last section. And the evolution of the packet-shape with *position*, downstream of the slit, is (approximately) the same as the evolution of the one-dimensional packets (discussed in the previous section) with *time*. In particular, as we saw in the last section, the wave packet will *spread* in this transverse direction as it propagates downstream. This is the phenomenon, illustrated in Fig. 2.5, of *diffraction*.

Of course, for a literal slit (which absorbs everything that hits it, and transmits whatever part of the incident wave goes through the slit) the transverse profile of the wave function (just downstream of the barrier) would be something like this:

$$\Psi(x,0) = \begin{cases} N & \text{for } -\frac{L}{2} < x < \frac{L}{2} \\ 0 & \text{otherwise} \end{cases} \tag{2.54}$$

with the constant N evidently being $1/\sqrt{L}$ to ensure proper normalization. As it turns out, the sharp edges (at $x = -L/2$ and $x = L/2$) of this function produce a Fourier transform $\phi(k)$ that diverges at $k = 0$ and this makes it slightly tricky to work with. See the Projects for a work-around that allows one to deal with this situation.

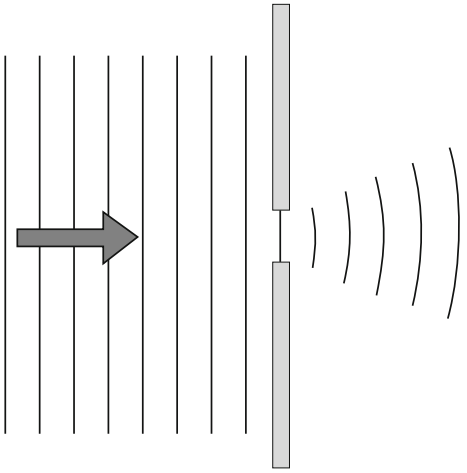


Fig. 2.5 An incident wave passes through a slit and diffracts

But just to understand the process conceptually, we can contemplate a "Gaussian slit", i.e., a barrier with a "transmission profile" (i.e., fraction of incident wave function that transmits rather than being absorbed) equal to a Gaussian. Then – basically by definition – the transverse profile of the beam just downstream of the barrier is a Gaussian, as in Eq. (2.42). If we make the approximation that the wave just steadily propagates to the right at some speed $v = \frac{\hbar k}{m}$ then we can relate the coordinate y along the direction of propagation to the time via $y = vt$. (This approximation is explained and developed further in the Projects.) And so we can immediately use, for example, Eq. (2.52) to write down an expression for the "intensity" (i.e., probability density) for finding the particle in the two-dimensional region behind the barrier:

$$I(x, y) \sim |\Psi(x, y)|^2 \sim \frac{1}{\sqrt{1 + \frac{\hbar^2 y^2}{4m^2 v^2 \sigma^4}}} e^{-x^2/2\left(\sigma^2 + \frac{\hbar^2 y^2}{4m^2 v^2 \sigma^4}\right)}. \tag{2.55}$$

I used Mathematica to make a nice density plot of this function; the result is shown in Fig. 2.6.

One of the nice reasons for setting this up, however, is that it provides a simple way to examine the structure of the wave function behind a *double slit* barrier. The classic two-slit interference pattern was first identified by Thomas Young as crucial evidence that light was a *wave*. And then of course the identification that "particles" (such as electrons) also exhibit interference, played and continues to play a crucial role in our understanding of the quantum nature of the sub-atomic realm.

So, then, imagine a barrier with not one but *two* "Gaussian slits" centered, say, at $x = a$ and $x = -a$. Then, the transverse profile of the wave function just behind the barrier will be given by

$$\Psi(x, 0) \sim \Psi_G(x - a, 0) + \Psi_G(x + a, 0) \tag{2.56}$$

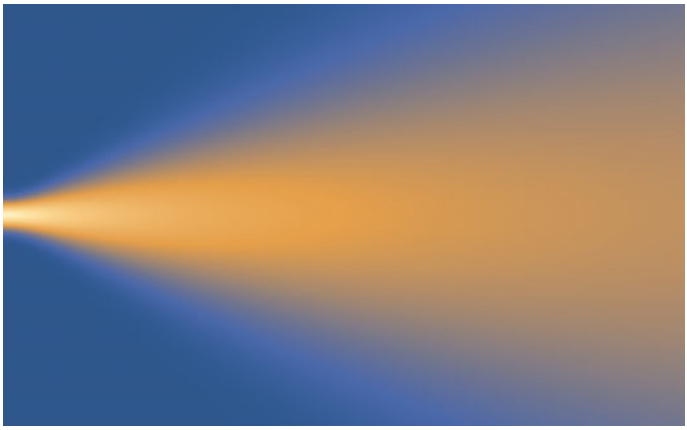


Fig. 2.6 Density plot of $|\Psi(x, y)|^2$ from Eq. (2.55) illustrating the intensity of a wave, diffracting as it propagates to the right having emerged from a "Gaussian slit" on the *left* edge of the image

i.e., a superposition of two Gaussians, one centered at $x = a$ and one centered at $x = -a$.

Each Gaussian term just spreads out in time in the way we analyzed in the previous section. (Formally, we can say that since the Schrödinger equation is *linear* in Ψ, the solution $\Psi(x, t)$ for this initial state – a superposition of two Gaussians – is just the corresponding superposition of the solutions for the two superposed terms individually.) Thus, using Eq. (2.51) twice (but with one small tweak each time) we can write

$$\Psi(x,t) \sim \Psi_G(x-a,t) + \Psi_G(x+a,t) \sim \frac{1}{\sqrt{\sigma^2 + \frac{i\hbar t}{2m}}} \left[e^{-\frac{(x-a)^2}{4(\sigma^2 + i\hbar t/2m)}} + e^{-\frac{(x+a)^2}{4(\sigma^2 + i\hbar t/2m)}} \right] \tag{2.57}$$

or, converting this into an expression for the wave function in the two-dimensional region in the way that we did before,

$$\Psi(x,y) \sim \frac{1}{\sqrt{\sigma^2 + \frac{i\hbar y}{2mv}}} \left[e^{-\frac{(x-a)^2}{4(\sigma^2 + i\hbar y/2mv)}} + e^{-\frac{(x+a)^2}{4(\sigma^2 + i\hbar y/2mv)}} \right]. \tag{2.58}$$

This is slightly messy to work with, but the idea qualitatively is that, as the two individual Gaussians begin to spread, they start to overlap. But then there can be either constructive or destructive interference depending on the relative *phases* in the region of overlap. For example, along the symmetry line, $x = 0$, the phases of the two terms will always match and we will therefore always have constructive interference, corresponding to a large value of $|\Psi|^2$, i.e., a high probability for the particle (if looked for) to be detected. But if we move a little bit to the side (say, in the positive x-direction) we are moving *toward* the central peak of one of the

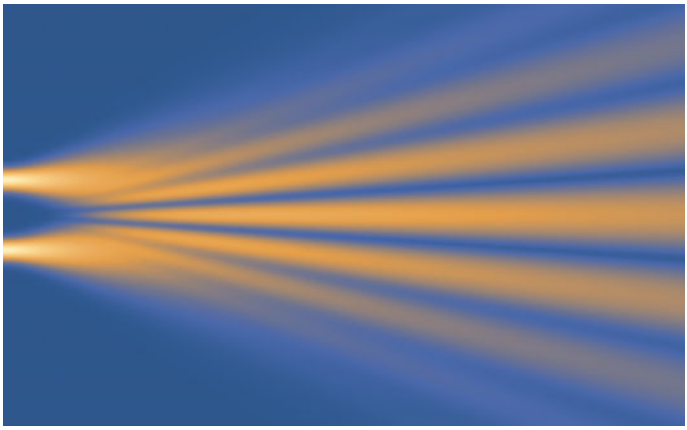


Fig. 2.7 Density plot of $|\Psi|^2$ from Eq. (2.58). A classic interference pattern emerges in the intensity of the wave downstream from the "double Gaussian slit" barrier at the *left* edge of the image

Gaussians and *away* from the central peak of the other, and so the phases of the two terms change at different rates, and eventually we find a spot where there is (at least for large y, nearly complete) destructive interference, corresponding to $|\Psi|^2 = 0$. Moving even farther in the positive x-direction eventually yields another spot where there is constructive interference, and so on.

The intensity pattern that results is shown in Fig. 2.7, which is again a Mathematica density plot of $|\Psi(x, y)|^2$, with $\Psi(x, y)$ given by Eq. (2.58). It is the classic two-slit interference pattern.

Of course, one should remember that such images of smoothly-distributed waves only tell half the story according to quantum mechanics. When an individual particle is looked for, it is not observed to be spread out like in these pictures; instead, it is found at some one particular spot, with the smooth $|\Psi|^2$ functions providing the probability distribution for the discrete sharp "hit". See, for example, in Fig. 2.8, the beautiful results of Tonomura *et al.* for a two-slit experiment with individual electrons and, in Fig. 2.9, the equally beautiful results of Dimitrova and Weis for a similar experiment using individual photons.

2.5 Spin

We will have occasion later to discuss measurements of the spin of (spin 1/2) particles. For such measurements, there are only two possible outcomes – "spin up" along the axis in question, or "spin down". This makes spin a very simple and elegant system to treat using the quantum formalism.

We can represent the "spin up along the z-axis" state this way:

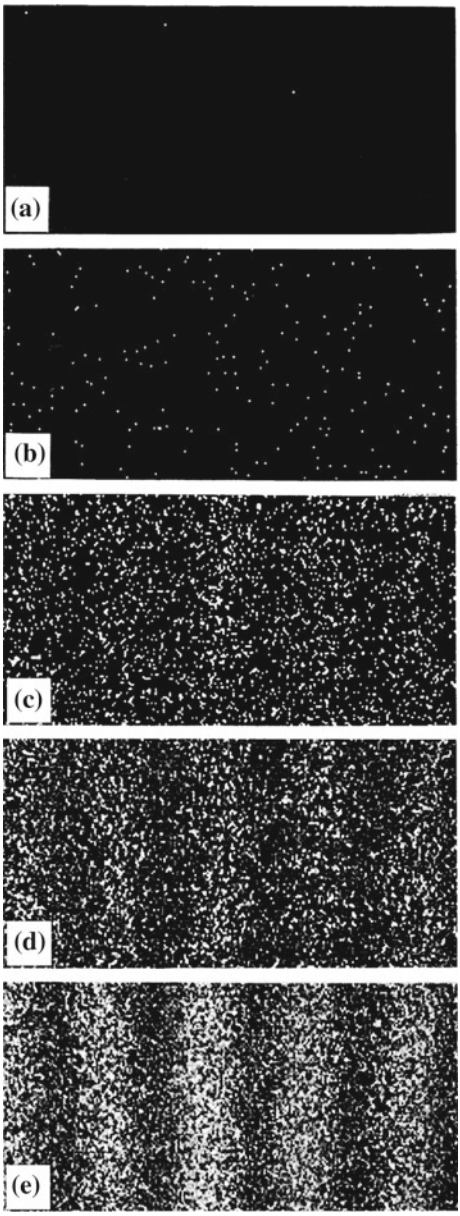


Fig. 2.8 Data from a double-slit experiment with electrons, in which electrons are sent through the apparatus one at a time. Each electron is found at a particular spot on the detection screen. The statistical pattern of spots – that is, the probability distribution for electron detection – builds up the classic two-slit interference pattern. (Reproduced from Tonomura et al., "Demonstration of single-electron buildup of an interference pattern" *American Journal of Physics* **57** (2), February 1989, pp. 117–120, our Ref. [1], with the permission of the American Association of Physics Teachers. http://aapt.scitation.org/doi/abs/10.1119/1.16104) See also Ref. [2]

$$\psi_{+z} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{2.59}$$

and the "spin down along the z-axis" state this way:

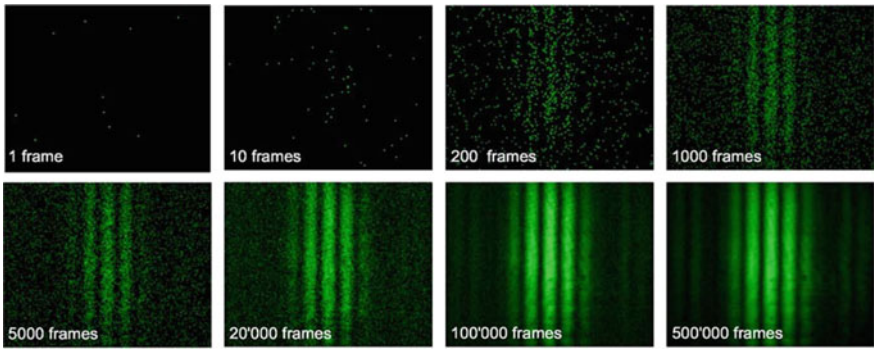


Fig. 2.9 Results of a similar experiment with "feeble light", i.e., individual photons. Just as with Tonomura's electrons, the measurement (here, using a CCD array) of the position of the photon always yields a definite, sharp location. The interference pattern is then realized in the statistical distribution of such individual sharp locations, after many photons are detected. (Reproduced from T.L. Dimitrova and A. Weis, "The Wave-Particle Duality of Light: A Demonstration Experiment," *American Journal of Physics* **76** (2008), pp. 137–142, our Ref. [3], with the permission of the American Association of Physics Teachers. http://aapt.scitation.org/doi/abs/10.1119/1.2815364)

$$\psi_{-z} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{2.60}$$

Note that these two-component vectors (technically "spinors") are the eigenvectors (with eigenvalues $+1$ and -1 respectively) of the spin-along-z operator, which can be represented as a two-by-two matrix:

$$\hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.61}$$

Of course, this is quantum mechanics, so the two eigenstates of $\hat{\sigma}_z$ are not the only possible states – instead they merely form a *basis* for the space of possible states. (Think of the spin up and spin down states here, ψ_{+z} and ψ_{-z}, as being like the energy eigenstates for the particle-in-a-box potential. These are not the only possible states! Instead, the general state is an arbitrary properly-normalized linear combination of them.) Here, a general state can be written as

$$\psi = c_+ \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_- \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} c_+ \\ c_- \end{pmatrix}. \tag{2.62}$$

The expansion coefficients c_+ and c_- should of course be understood to have their usual, generalized Born rule meanings: if a particle is in the spin state ψ and its spin along the z-axis is measured, the probability for the measurement to have outcome "spin up along z" is $P_+ = |c_+|^2$ whereas the probability for the measurement to have outcome "spin down along z" is $P_- = |c_-|^2$. And note that, since these are

the only two possible outcomes, the probabilities should sum to one. That is, proper normalization for the general spin state ψ requires $|c_+|^2 + |c_-|^2 = 1$.

Things get a little more interesting when we consider the possibility of measuring the spin of a particle along some axis other than the z-axis. We will only ever have occasion to care about measurements along axes in (say) the $x-z$-plane. The operator corresponding to spin measurements along the x-axis can be written

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2.63}$$

whose eigenvectors are

$$\psi_{+x} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{2.64}$$

(with eigenvalue $+1$ corresponding to "spin up" along the x-direction) and

$$\psi_{-x} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{2.65}$$

(with eigenvalue -1 corresponding to "spin down" along the x-direction).

The operator corresponding to spin measurements along an arbitrary direction $\hat{n}$ in the x-z-plane is

$$\hat{\sigma}_n = \hat{n} \cdot \hat{\vec{\sigma}} = \cos(\theta)\hat{\sigma}_z + \sin(\theta)\hat{\sigma}_x = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix} \tag{2.66}$$

where θ is the angle between $\hat{n}$ and $\hat{z}$ (toward $\hat{x}$). It is a simple exercise in linear algebra to show that the eigenvectors of this matrix can be written

$$\psi_{+n} = \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \tag{2.67}$$

(with eigenvalue $+1$, corresponding to "spin up along n") and

$$\psi_{-n} = \begin{pmatrix} \sin(\theta/2) \\ -\cos(\theta/2) \end{pmatrix} \tag{2.68}$$

(with eigenvalue -1, corresponding to "spin down along n"). Notice that, for $\theta = 0$, these states correspond to ψ_{+z} and ψ_{-z}, as they should, and similarly for $\theta = 90°$, they correspond to ψ_{+x} and ψ_{-x}.

Let's consider a concrete example to illustrate these ideas. Suppose a particle is prepared in the "spin up along n" state where n is a direction 60° down from the z-axis (toward the x-axis). That is, suppose

$$\psi_0 = \begin{pmatrix} \cos(30°) \\ \sin(30°) \end{pmatrix}. \tag{2.69}$$

Then, we are going to measure the spin of this particle along the z-axis. What is the probability that this z-spin measurement comes out "spin down"?

To answer this, as always, we have to write the given state as a linear combination of the eigenstates of the operator corresponding to the measurement that is to be performed. Here that means writing ψ_0 as a linear combination of ψ_{+z} and ψ_{-z}. But that is easy:

$$\psi_0 = \begin{pmatrix} \cos(30°) \\ \sin(30°) \end{pmatrix} = \cos(30°) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \sin(30°) \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{2.70}$$

So then we can read off that the probability of the measurement having outcome "spin down along z" is the square of the expansion coefficient on the "spin down along z" term, i.e.,

$$P_{-z} = \sin^2(30°) = \frac{1}{4}. \tag{2.71}$$

So if a whole beam of particles (all identically prepared in the state ψ_0) is sent into a Stern Gerlach device (with its axis aligned along the z direction), 25% of the particles will emerge having been deflected "down" and the remaining 75% will emerge having been deflected "up".

2.6 Several Particles

So far all of the examples we've considered involve only a single particle (and in particular its spatial or spin degrees of freedom). In situations involving two or more particles, the principles are the same, but there are some important new possibilities that will become important in subsequent chapters. Let us lay some of the groundwork here.

A crucial point is that, for an N-particle system, it is not the case that each of the N particles has its own wave function. Instead, there is a single wave function for the whole N-particle system. This wave function obeys the N-particle Schrödinger equation

$$i\hbar \frac{\partial \Psi(x_1, x_2, ..., x_N, t)}{\partial t} = \sum_{i=1}^{N} \frac{-\hbar^2}{2m_i} \nabla_i^2 \Psi(x_1, x_2, ..., x_N, t) + V(x_1, x_2, ..., x_N) \Psi(x_1, x_2, ..., x_N, t). \tag{2.72}$$

Note that the wave function $\Psi(x_1, x_2, ..., x_N, t)$ is a (time-dependent) function on the *configuration space* of the N-particle system: x_1 is the spatial coordinate of particle 1, x_2 is the spatial coordinate of particle 2, etc.

As an example, consider the case of two particles (which have identical m and which do not interact with each other) trapped in the box from Sect. 2.2. The time-dependent Schrödinger equation reads

$$i\hbar\frac{\partial\Psi(x_1,x_2,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x_1,x_2,t)}{\partial x_1^2} - \frac{\hbar^2}{2m}\frac{\partial^2\Psi(x_1,x_2,t)}{\partial x_1^2} + V(x_1)\Psi(x_1,x_2,t) + V(x_2)\Psi(x_1,x_2,t) \tag{2.73}$$

where V is just the "particle-in-a-box" potential, Eq. (2.20).

It is easy to show, by separation of variables, that there are solutions of the form

$$\Psi_{m,n}(x_1,x_2,t) = \sqrt{\frac{2}{L}}\sin\left(\frac{m\pi x_1}{L}\right)\sqrt{\frac{2}{L}}\sin\left(\frac{n\pi x_2}{L}\right)e^{-i(E_m+E_n)t/\hbar} \tag{2.74}$$

which are *products*: one of the one-particle energy eigenfunctions for particle 1, times one of the one-particle energy eigenfunctions for particle 2, and then with the usual time-dependent phase factor involving the energy, which is just the sum of the two one-particle energies.

If the two-particle quantum state is one of these product states, the wave function Ψ is formally a function on the two-particle configuration space, but there is an obvious sense in which each particle has its own definite state.

But, as usual in quantum mechanics, these states do not exhaust the possibilities – instead, they merely form a *basis* for the space of all possible wave functions. And that gives rise to the crucially-important concept of "entanglement". An "entangled" wave function (or quantum state) for several particles is simply one that is *not a product*. An entangled state of two particles, that is, *cannot* be written as "some wave function for particle 1" times "some wave function for particle 2". In entangled states, the individual particles really fail to have their own, individual, states.

Here is an example. Consider the two particles in the "box" potential, and suppose we are only interested in the situation at $t = 0$ (so we ignore time-dependence). One possible state for the two particles to be in is

$$\Psi_{1,2} = \psi_1(x_1)\psi_2(x_2) \tag{2.75}$$

corresponding to particle 1 being in the ground state and particle 2 being in the first excited state. Another possible state is

$$\Psi_{2,1} = \psi_2(x_1)\psi_1(x_2) \tag{2.76}$$

corresponding to particle 1 being in the first excited state and particle 2 being in the ground state. Neither of these states is particularly interesting or troubling since, for each of them, each particle has its own definite state (with a definite energy).

But here is another possible state that the two-particle system could be in:

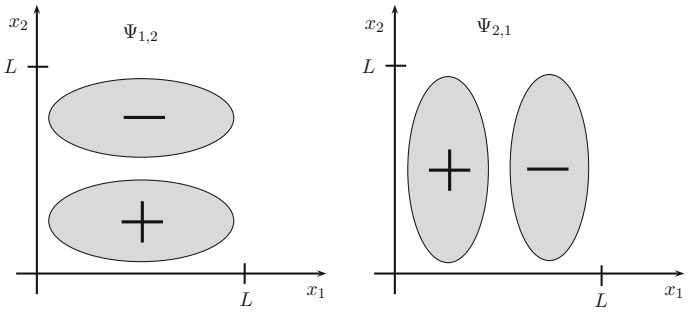


Fig. 2.10 The cartoon graph on the *left* indicates (in a very rough way) the structure (in the two-dimensional configuration space) of $\Psi_{1,2} \sim \sin(\pi x_1/L)\sin(2\pi x_2/L)$. This is the product of a function that is positive for all x_1 between 0 and L, but then switches from being positive for $0 < x_2 < L/2$ to being negative for $L/2 < x_2 < L$. So the product has a reasonably large magnitude in roughly the *grey-shaded* areas and is positive and negative in the regions indicated. The graph on the *right* indicates the structure of $\Psi_{2,1} \sim \sin(2\pi x_1/L)\sin(\pi x_2/L)$ in a similar way

$$\Psi_{ent} = \frac{1}{\sqrt{2}}\left(\Psi_{1,2} + \Psi_{2,1}\right) = \frac{1}{\sqrt{2}}\big[\psi_1(x_1)\psi_2(x_2) + \psi_2(x_1)\psi_1(x_2)\big]. \tag{2.77}$$

This is a *superposition* of (on the one hand) a state in which particle 1 is in the ground state and particle 2 is in the first excited state and (on the other hand) a state in which particle 2 is in the first excited state and particle 2 is in the ground state. So neither particle 1 nor particle 2 is in a state of definite energy at all. (Interestingly, though, this entangled two-particle state *is* an eigenstate of the total energy: the two particles definitely have a total energy of $E_1 + E_2$... there's just no particular fact of the matter about how this total energy is distributed between the two particles!)

It is perhaps helpful to practice visualizing these states in the two-particle configuration space. Figure 2.10 shows sketchy cartoon versions of the two states $\Psi_{1,2}$ and $\Psi_{2,1}$. Each wave function is positive in one part, negative in another, and zero between them.

The sum of these two states – the *entangled superposition* state in Eq. (2.77) – is shown in this same sketchy cartoon style in Fig. 2.11. There is (approximate) destructive interference in the upper-left and lower-right corners of the configuration space, and instead constructive interference in the lower-left and upper-right corners. So the state Ψ_{ent} has a large positive value in the lower-left corner, a large negative value in the upper-right corner, and is approximately zero elsewhere. Note that, since the probability of finding the particles at positions x_1 and x_2 is $|\Psi|^2$, this means that, if the two particles are in the state Ψ_{ent}, they are unlikely to be found at different locations: the upper-left and lower-right corners of the configuration space here correspond, respectively, to "particle 1 is on the left and particle 2 is on the right" and "particle 1 is on the right and particle 2 is on the left". These are precisely the regions of configuration space where Ψ has a small amplitude and hence the corresponding probability is small. On the other hand, the probabilities for finding

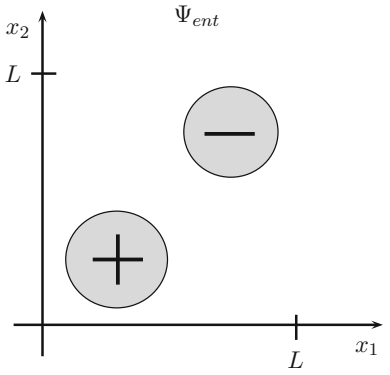


Fig. 2.11 The structure of the entangled state $\Psi_{ent} = \frac{1}{\sqrt{2}}\left(\Psi_{1,2} + \Psi_{2,1}\right)$ in configuration space. In the *upper-left corner* (i.e., $0 < x_1 < L/2$ and $L/2 < x_2 < L$) the two superposed terms have opposite sign and (partially) cancel out. The same thing happens in the *lower-right corner*. But in the *lower left corner* (i.e., $0 < x_1 < L/2$ and $0 < x_2 < L/2$) the two superposed terms have the same (positive) sign and hence add up to a function with a large (positive) value. The same thing happens in the *upper-right corner*, but with "positive" replaced by "negative"

both particles "on the left" and finding both particles "on the right" are high. So in some sense this particular entangled state is one in which neither particle has a definite energy, and of course neither particle has a definite position either, and yet there are certain *correlations* between them, i.e., certain *joint properties* that are more well-defined: the *total* energy of the two particles, for example, is perfectly definite, and it is extremely likely that the particles will be found to be near one another if their positions are measured.

That last sentence, by the way, should kind of blow your mind. So slow down and let it percolate for a while if you didn't already!

This example of two particles in a box has dealt exclusively with the spatial degrees of freedom of two particles. Note that it is also possible for the spin degrees of freedom of two particles to be entangled. For example, we might have two particles in the joint spin state:

$$\Psi_{singlet} = \frac{1}{\sqrt{2}}\left[\begin{pmatrix}1\\0\end{pmatrix}_1\begin{pmatrix}0\\1\end{pmatrix}_2 - \begin{pmatrix}0\\1\end{pmatrix}_1\begin{pmatrix}1\\0\end{pmatrix}_2\right]. \tag{2.78}$$

This can be understood as a superposition (with, for variety, a minus sign this time) of (on the one hand) a state in which particle 1 is "spin up along z" and particle 2 is "spin down along z", and then (on the other hand) a state in which particle 1 is "spin down along z" and particle 2 is "spin up along z". As in the previous example, neither particle individually has a definite spin state, but there are certain correlations between the particles' spins; for example, here, if the z-spins of both particles are measured, one cannot predict in advance whether it will be "particle 1 is spin up" and "particle 2 is spin down" (which joint outcome has probability 50%) or instead

"particle 1 is spin down" and "particle 2 is spin up" (which also has probability 50%)... but one can predict in advance, with 100% certainty, that the outcomes of the two spin measurements will be *opposite* – one "up" and one "down".

You can play around a little bit more with this entangled spin state in the Projects if you so choose. And then we will encounter it again soon when we discuss the famous argument of Einstein, Podolsky, and Rosen in Chap. 4.

Projects:

(2.1) For the example from the Particle-in-a-box section – with $\Psi(x, 0)$ given by Eq. (2.27) – calculate the probability that a measurement of the particle's position x at time t finds the particle on the left-hand-side of the box: $0 < x < L/2$.

(2.2) Use Mathematica or a similar software package to make nice movies of the exact evolution of the real and imaginary parts of $\Psi(x, t)$ given by Eq. (2.29).

(2.3) A particle in a box starts in the state $\Psi(x, 0) = 1/\sqrt{L}$. What is $\Psi(x, t)$? What is the probability that an energy measurement at time t yields the ground state energy?

(2.4) Show explicitly that Eq. (2.26) satisfies the time-dependent Schrödinger Equation.

(2.5) The uncertainty of some quantity A is defined as: $(\Delta A)^2 = \langle (A - \langle A \rangle)^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2$. Use this definition to calculate the exact uncertainty Δx of the position for the Gaussian wave packet given by Eq. (2.33). Note that, for example, $\langle x^2 \rangle = \int x^2 |\Psi(x)|^2 dx$. (Here's a clever way to do integrals of this form: $\int x^2 e^{-ax^2} dx = -\frac{\partial}{\partial a} \int e^{-ax^2} dx$.) Then calculate also the uncertainty Δk in the wave number using (2.47) and convert this into a statement about the uncertainty in the momentum. What, exactly, is the product of Δx and Δp? As it turns out, this Gaussian is a "minimum uncertainty wave packet" – meaning that the product of Δx and Δp for this state is the smallest the product can ever be. (But it can be and usually is bigger!) Summarize this fact by writing down an exact mathematical statement of the Heisenberg uncertainty principle.

(2.6) Work through the gory mathematical details of deriving Eq. (2.51) from Eq. (2.50). Or better, develop a general formula for Gaussian integrals of the form $\int e^{-Ak^2} e^{Bk}\, dk$ in terms of A and B. Then use the general formula to show how (2.51) follows from (2.50).

(2.7) Show explicitly that the $\Psi(x, t)$ in Eq. (2.51) solves the time-dependent Schrödinger Equation.

(2.8) Use Mathematica (or some similar package) to make some nice animations showing the time-evolution of $\Psi(x, t)$ for the initially Gaussian wave packet, Eq. (2.51). For example, what does the real part look like? The imaginary part? The modulus squared?

(2.9) Suppose the initial wave function is a position eigenstate: $\Psi(x, 0) = \delta(x - x')$. What is $\Psi(x, t)$? Note that this is a very useful result, since *any* initial wave function can be written as a linear combination of δ functions in a rather trivial way: $\Psi(x, 0) = \int \Psi(x', 0)\delta(x - x')\, dx$. And of

course the Schrödinger equation is linear, so $\Psi(x,t)$ is just that same linear combination of the time-evolved versions of $\delta(x-x')$, i.e., $\Psi(x,t) = \int \Psi(x',0)G(x,x',t)\,dx$, where $G(x,x',t)$ is just the wave function that $\Psi(x,0) = \delta(x-x')$ evolves into at time t. Use this alternative approach to re-derive our expression for $\Psi(x,t)$ for the initially Gaussian wave packet.

(2.10) Use the approach from Project 2.9 to write an expression for $\Psi(x,t)$ for a $\Psi(x,0)$ that is constant for $-L/2 < x < L/2$, and zero otherwise. This expression will have some divergence issues. But you should be able to show that in the $t \to \infty$ limit, a certain simplification allows you to derive a nice result for (what can be understood as) the probability density associated with (regular, non-Gaussian) single-slit diffraction (assuming the detection screen is far behind the slit). Make a nice graph.

(2.11) Let's try to understand the mathematics behind the idea, from Sect. 2.4, of trading out the t-dependence of our one-dimensional $\Psi(x,t)$, using $y = vt$, for a wave function that we interpret as a solution $\psi(x,y)$ of the two-dimensional TISE. Start with the Schrödinger Equation in two dimensions,

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{2mi}{\hbar}\frac{\partial}{\partial t}\right)\Psi(x,y,t) = 0, \tag{2.79}$$

and look for solutions of the form

$$\Psi(x,y,t) = \phi(x,y)e^{i(ky-\omega t)} \tag{2.80}$$

corresponding to a plane wave propagating in the y-direction, but with a slowly-varying y-dependent transverse profile ϕ. (a) Plug Eq. (2.80) into Eq. (2.79) and show that, for $\omega = \hbar k^2/2m$ and $2k\frac{\partial\phi}{\partial y} \gg \frac{\partial^2\phi}{\partial y^2}$, ϕ should satisfy

$$\frac{\partial^2\phi}{\partial x^2} + 2ik\frac{\partial\phi}{\partial y} = 0. \tag{2.81}$$

(b) Explain why the two conditions used in (a) are reasonable and what they mean physically. (c) Argue that, with $y \leftrightarrow vt$ (where $v = \hbar k/m$), Eq. (2.81) is just the one-dimensional time-dependent Schrödinger equation. [Note that this technique is called the "paraxial approximation".]

(2.12) A spin 1/2 particle is prepared in the state ψ_{-x} (spin down along x). We then perform a measurement of its spin along the same $\hat{n}$ direction used in the example in the text: 60° down from the z-axis (toward the x-axis). Find the probabilities for the two possible measurement outcomes.

(2.13) If a spin 1/2 particle is placed in a magnetic field $\vec{B}$, the spin-up and spin-down states (parallel to the magnetic field direction) have different energies, which one can capture with an appropriate Hamiltonian operator. For example, if the magnetic field is in the y-direction, we can write

$$\hat{H} = \hbar\omega\hat{\sigma}_y = \hbar\omega \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

where ω is a constant (with angular frequency units, hence the letter) that depends on the magnetic dipole moment of the particle and the strength of the field. Use this Hamiltonian operator in Eq. (2.1) to find out how the (spin) state of a particle, say initially in the state ψ_{+z}, evolves in time. (Hint: use the general method outlined in the chapter of solving the time-dependent Schrödinger equation, namely, first find the energy eigenstates, then write the initial state as a linear combination of energy eigenstates, then tack the appropriate time-dependent exponential factor onto each term in the linear combination.)

(2.14) For the "two particles in a box" system, construct an entangled state in which even the total energy of the two particles is not well-defined. Use Mathematica to make some density plots and/or movies showing how the state looks and how it evolves in time.

(2.15) The "two particles in a box" system is mathematically isomorphic to a "one particle in a two-dimensional box" system. Explain and contemplate.

(2.16) Re-write the "singlet" spin state for two spin 1/2 particles – Eq. (2.78) – in terms of the spin-up and spin-down *along the x-axis* states, ψ_{+x} and ψ_{-x}.

(2.17) Re-write the "singlet" spin state for two spin 1/2 particles – Eq. (2.78) – in terms of the spin-up and spin-down *along the n-axis* states, ψ_{+n} and ψ_{-n}.

References

1. A. Tonomura, J. Endo, T. Matsuda, T. Kawasaki, H. Ezawa, Demonstration of single-electron buildup of an interference pattern. Am. J. Phys. **57**(2), 117–120 (1989). February
2. Roger Bach, Damian Pope, Sy-Hwang Liou, Herman Batelaan, Controlled double-slit electron diffraction. New J. Phys. **15**, 033018 (2013)
3. T.L. Dimitrova, A. Weis, The wave-particle duality of light: a demonstration experiment. Am. J. Phys. **76**, 137–142 (2008)

Chapter 3
The Measurement Problem

In Chap. 2, we reviewed the mathematical formalism of quantum mechanics and practiced applying it to a number of concrete examples. In the present chapter, we will begin the process of stepping back and turning a critical eye toward the theory. In particular, in this chapter, we will look carefully at the curiously central role that the theory gives to the process of "measurement" and discuss the network of related concerns, centering on the infamous example of Schrödinger's Cat, that have come to be called "the measurement problem."

3.1 The Quantum Description of Measurement

Our discussion of the examples in the previous Chapter focused on solving Schrödinger's equation to understand how the quantum states of microscopic systems evolve in time, and then using Born's rule to connect these quantum states to probabilities for various possible measurement outcomes. Here, we want to emphasize and develop two additional aspects of Born's rule, and then step back and look at the quantum mechanical description of measurement and, really, the quantum mechanical description of the world as a whole.

The first new aspect of Born's rule that we need to stress is the so-called "collapse postulate". Recall that, according to Born's rule, we calculate the probabilities of different measurement outcomes as follows: first, write the quantum state Ψ (of the system we are measuring) as a linear combination of eigenstates of the operator corresponding to the property we are measuring, as in

$$\Psi = \sum_i c_i \Psi_i \tag{3.1}$$

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_3

where Ψ_i is an eigenstate of the operator $\hat{A}$ with eigenvalue A_i. Then the probability for the measurement to have outcome A_i is $P(A_i) = |c_i|^2$. That should be familiar and clear.

But it is an experimental fact that, if you *immediately repeat* a measurement – for example, you measure the energy of a particle and then immediately measure its energy again – you always get *the same result* for the second measurement that you got for the first. (Note how weird it would be if this weren't true. At very least, the word "measurement" would then seem quite inappropriate.) But then the consistent applicability of Born's rule to the two measurements implies that, by the time the second measurement occurs, the system must *be* in the eigenstate corresponding to the outcome of the first measurement. Only this (according to Born's rule) will ensure that the probability of seeing that same result again, in the second measurement, is 100%. This is the collapse postulate: when a measurement occurs, and has outcome A_n, the quantum state of the system being measured *ceases* to be whatever superposition it might have been previously, and "*collapses*" to the eigenstate Ψ_n whose eigenvalue is the realized outcome A_n.

Formally, for a measurement that begins at time t_1 and ends at time t_2, one has

$$\Psi(t_1) = \sum_i c_i \Psi_i \quad \rightarrow \quad \Psi(t_2) = \Psi_n. \tag{3.2}$$

where the right arrow indicates the time-evolution. The crucial point – and the reason the "collapse postulate" is a *postulate* – is that this is a very different sort of time-evolution than wave functions normally undergo, when they are evolving according to Schrödinger's equation.

The second new aspect of the Born rule then has to do with the idea of experiments having definite outcomes. So far we have just used these words abstractly, without really thinking in concrete physical terms about what we mean. So consider, for example, the two-slit experiment described in the last Chapter. A single electron (or photon) propagates through the apparatus, and we describe its state with a wave function. But then, at some point, the particle interacts with the "detection screen" that is being used to measure its position. This (as we have just been discussing) apparently causes the wave function of the particle to "collapse" – to switch from something that is spread out across (say) nearly the whole screen, to something more like a δ-function spike at a particular spot.

But something important happens to the screen, as well! If, for example, we think of the screen as a piece of film, a certain little spot on the film *changes color*. Or if we think of it as a CCD array, an electrical signal is produced which (say) results in the coordinates of the particular pixel that the particle "hit" being printed on a computer screen. Whatever the details, exactly, the point is that the measuring device itself is a physical thing, which undergoes some kind of observable physical change, that is intimately coupled with the change (described by the collapse posulate) in the state of the particle that happens at that same moment. The "measurement", in short, is a *physical interaction* between these two physical systems, which both change as a result.

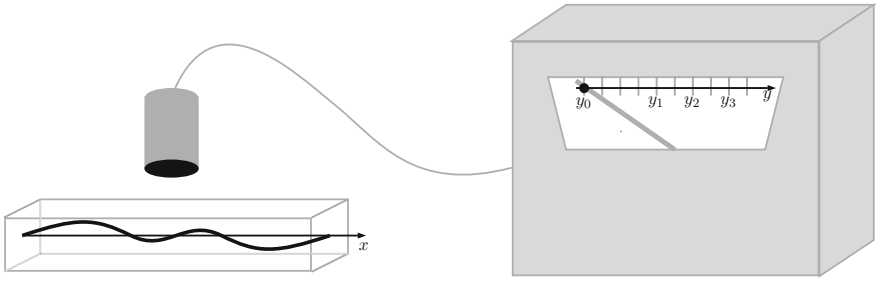


Fig. 3.1 The quantum particle-in-a-box (whose spatial degree of freedom is called x) is shown on the *left*; the curve is meant to indicate its wave function (though one should be careful not to take this picture too literally!). Then there is an energy-measurement device which will perform the measurement. The device has a macroscopic pointer, which we can idealize as a single, very heavy particle with horizontal coordinate y. Prior to the measurement-interaction, the pointer is sitting in its "ready" position (y_0); after the measurement interaction, the pointer will move to a new position which indicates the outcome of the measurement: y_1 will mean that the energy of the particle is E_1, etc

Let us set up a semi-realistic concrete example to consider, involving the "pointer" concept introduced in Chap. 1. Suppose our microscopic quantum system is a particle-in-a-box, which has been prepared so that its state is a superposition of the first few energy eigenstates. Say:

$$\psi_0 = c_1\psi_1 + c_2\psi_2 + c_3\psi_3 \tag{3.3}$$

where ψ_1 is the $n = 1$ (i.e., ground state) particle-in-a-box wave function, etc. And suppose that we are planning to wait a few seconds and then measure the energy of the particle-in-a-box. This measurement will be performed using some kind of energy-measuring apparatus, which we will treat schematically as a black box (with internal workings whose details we need not worry about too much) with a macroscopic pointer whose position, after the interaction, will indicate the result of the energy measurement. See Fig. 3.1.

So then the quantum mechanical description of the measurement process goes like this. The particle-in-a-box is described by a wave function, which starts as ψ_0 and then evolves according to the Schrödinger equation until the measurement is made. The pointer, meanwhile, starts out at position $y = y_0$ and just sits there (obeying $F_{\text{net}} = ma$ with $F_{\text{net}} = 0$) until the measurement is made.

At some point, the measurement interaction occurs and when the dust settles, the situation is now as follows: the wave function of the particle-in-the-box has collpased to one of the energy eigenfunctions ψ_n (with n being either 1, or 2, or 3); and the apparatus pointer has moved and is now pointing at the appropriate corresponding value, $y = y_n$. (See Fig. 3.2 for an illustration of one of the three possibilities.) With the interaction between the particle-in-the-box and the measuring apparatus now completed, each part – the particle-in-the-box and the apparatus pointer – continue evolving as normal (i.e., the wave function of the particle-in-the-box evolves

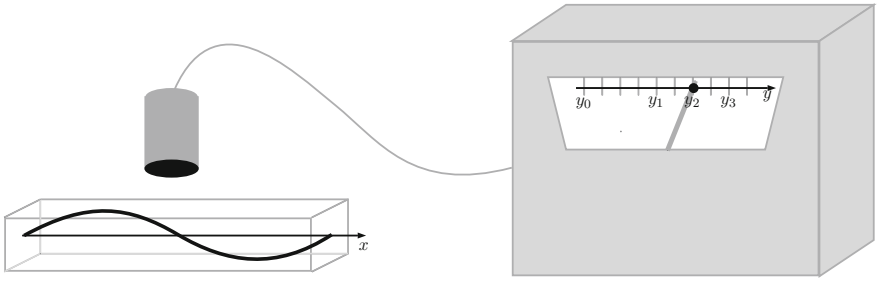


Fig. 3.2 One of the three possible post-measurement states of the particle-in-a-box and measurement apparatus pointer: the wave function of the PIB has "collapsed" to ψ_2 and the pointer has moved to position y_2, indicating that the energy measurement had outcome E_2

according to the Schrödinger equation, and the apparatus pointer again obeys $F_{\text{net}} = ma$ with, presumably, $F_{\text{net}} = 0$ again).

Now there are two curious and connected things about all of this, one dynamical and one more ontological.

The curious thing, dynamically, is that the measurement interaction seems to involve a violation of the "usual" dynamics for each of the interacting objects. We have already pointed out that the "collapse" that occurs to the wave function of the particle-in-the-box is not (or at least is not obviously, not apparently) something that is described by Schrödinger's equation. The wave function, during that brief time period, collapses *instead* of evolving in accordance with Schrödinger's equation. And then something similar occurs with the pointer, although the schematic nature of the example makes this a little harder to see. The claim here is meant to be something like: "$F_{\text{net}} = ma$ ceases to apply to the pointer when it suddenly swerves from y_0 to y_n for some particular value of n." But of course, that cannot be exactly true. The pointer is, after all, just the last step in a long causal chain. The (say, magnetic) force on it – which results in it moving some particular distance to the right during the course of the interaction – can be perfectly explained in terms of (say) electrical currents that flow through some wires inside the black box part of the apparatus. But those currents can, in turn, be perfectly explained, in standard classical Maxwellian electrodynamical terms, by (say) other electrical currents, in the cable, i.e., further upstream along the causal chain. And so on... but at *some* point, we must come to some change in some aspect of the physical state of the apparatus for which the usual classical dynamical rules do not provide an explanation. Otherwise, why would we need quantum mechanics at all?

This brings us to the other curious thing about this whole situation: the apparatus is being described in classical terms! We do not speak, for example, of "the wave function of the pointer," but instead of the pointer's *position*. The whole setup and description, that is, presupposes that the pointer (and presumably most of the contents of the entire black box, at least until we get far upstream into some murky meso- or micro-scopic stuff) is the kind of thing to which we can just unproblematically attribute definite, sharp, unambiguous, un-superposed, classical properties. The

picture, that is, seems to presume that quantum systems (described in terms of wave functions) exist *in addition to* classical systems (described, for example, in terms of particles with definite positions). If this kind of description is taken seriously, as faithfully capturing the true nature of the world being described, it implies for example that there are two fundamentally different types of particles in the world: those (like our particle-in-a-box here) which are "wavy" (i.e., which are properly described in terms of a spatially-spread-out wave function) and those (like the particle(s) composing the pointer here) which are "sharp" (i.e., which are properly described as having definite positions at all times).

At the broadest level, then, it seems like the implied quantum mechanical picture of the world goes something like the following. There is, to begin with, the familiar macroscopic "classical" realm in which things have definite properties and are described in clear, everyday terms. This macroscopic realm basically obeys the dynamical laws of classical mechanics. Then there is also a microscopic realm where our everyday classical intuitions don't apply and we must instead describe things using quantum mechanical wave functions which, of course, do not necessarily attribute definite properties to things: energy, momentum, position, etc., can all be "smeared out" in quantum superpositions, i.e., these properties can fail to have definite values in the way we would have expected classically.

And then finally there are the special rules describing the *interaction* of the macroscopic and microscopic realms. In particular, during a *measurement*, the quantum mechanical wave function describing the microscopic system fails, momentarily, to evolve in accordance with Schrödinger's equation, and instead *collapses* to one particular eigenstate of the operator corresponding to the physical property (energy or momentum or whatever) that is being measured. Which particular eigenstate? This is supposed to be irreducibly random and inexplicable, but the particular state that the quantum system collapses to is the one that corresponds to the particular measurement outcome that is displayed by the apparatus, as a result of its own process of jumping, inexplicably and in violation of the usual (here, classical) dynamical rules.

Stepping back, all of this should make one feel very uncomfortable. To begin with, there is a kind of schizophrenic division of the world into two "realms" (the microscopic quantum part, and the macroscopic classical part) which seem to have completely different ontologies and completely different dynamical laws. And then there are apparently special dynamical rules which come into play when the two realms interact, during a "measurement". And the situation is then made worse by the *vagueness* of the concept "measurement". If you say "during measurements, quantum wave functions momentarily cease to obey Schrödinger's equation and instead collapse" that is already weird and troubling, but it becomes downright *meaningless* if you can't specify *exactly* what kinds of physical processes count as "measurements". Bell put this particular point very sharply as follows:

> What exactly qualifies some physical systems to play the role of 'measurer'? Was the wave-function of the world waiting to jump for thousands of millions of years until a single-celled living creature appeared? Or did it have to wait a little longer, for some better qualified system ... with a Ph.D.? If the theory is to apply to anything but highly idealised laboratory operations, are we not obliged to admit that more or less 'measurement-like' processes are

> going on more or less all the time, more or less everywhere? Do we not have [quantum] jumping [i.e., collapse] then all the time? [1]

It indeed seems necessary to admit that "measurements" are ubiquitous, and occur even in places and times where there are no human experimenters. But it also seems hopeless to think that we will be able to give an appropriately sharp answer to the question of what, *exactly*, differentiates the 'ordinary' processes (where the usual dynamical rules apply) from the 'measurement-like' processes (where the rules momentarily change).

In an interview, Bell was once asked whether he thought the problems with quantum mechanics were philosophical or experimental. His answer is relevant here:

> I think there are *professional* problems. That is to say, I'm a professional theoretical physicist and I would like to make a clean theory. And when I look at quantum mechanics I see that it's a dirty theory. The formulations of quantum mechanics that you find in the books involve dividing the world into an observer and an observed, and you are not told where that division comes – on which side of my spectacles it comes, for example – or at which end of my optic nerve. You're not told about this division between the observer and the observed. What you learn in the course of your apprenticeship is that for practical purposes it does not much matter where you put this division; that the ambiguity is at a level of precision far beyond human capability of testing. So you have a theory which is fundamentally ambiguous... [2].

Stepping back, it begins to seem like we must have taken the quantum mechanical "recipe" – Born's rule and the collapse postulate in particular – too literally, somehow. Surely, for example, big macroscopic things like measuring devices and their pointers (not to mention cats and trees and planets) are just large collections of electrons and other microscopic particles. And so surely, if macroscopic stuff is literally made of lots and lots of little microscopic parts, then shouldn't a fully microscopic description suffice, at least in principle even if not in practice? In other words, shouldn't the familiar macroscopic world (meaning large objects with definite, classical properties, that at least approximately obey Newton's laws of motion) somehow *emerge* from the more fundamental quantum mechanical description, as opposed to being postulated at the fundamental level?

3.2 Formal Treatment

Let us explore this possibility in a more formal way here. (Doing this will help us understand the point Schrödinger meant to be making with his cat example, when we turn to that shortly.)

Take the example from the previous section – measuring the energy of a particle-in-a-box – but now let us attempt to use the microscopic type of quantum description (in terms of wave functions obeying Schrödinger's equation) for the entire setup, including the measuring apparatus. Suppose, as before, that the particle-in-a-box starts out in the state

$$\psi_0(x) = c_1\psi_1(x) + c_2\psi_2(x) + c_3\psi_3(x). \tag{3.4}$$

As for the pointer, previously we had been describing it classically and hence attributing to it some definite pre-measurement position y_0. But now we want to instead describe the pointer quantum mechanically, with a wave function. Let's say that the pointer in its "ready position" can be described by a Gaussian wave packet centered on the position y_0:

$$\phi(y) = Ne^{-(y-y_0)^2/4\sigma^2}. \tag{3.5}$$

At the moment (call it $t = 0$) when the measurement interaction begins, the joint wave function of the particle and pointer will be

$$\Psi_0(x, y) = \psi_0(x)\phi(y). \tag{3.6}$$

This quantum system will then evolve in time in accordance with the Schrödinger equation,

$$i\hbar\frac{\partial\Psi(x, y, t)}{\partial t} = \hat{H}\Psi(x, y, t). \tag{3.7}$$

But what is the Hamiltonian, $\hat{H}$? Evidently there will be three contributions. First, it will include the usual terms corresponding to the kinetic and potential energies of the particle-in-the-box, whose degree of freedom is "x":

$$\hat{H}_x = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x). \tag{3.8}$$

As we saw in Chap. 2, $\hat{H}_x$ gives each term in Eq. (3.4) a time-dependent complex phase, with distinct frequencies for the different terms.

Second, the overall Hamiltonian should include the kinetic energy of the pointer, whose degree of freedom is "y":

$$\hat{H}_y = -\frac{\hbar^2}{2M}\frac{\partial^2}{\partial y^2}. \tag{3.9}$$

As we saw in Chap. 2, $\hat{H}_y$ will tend to make the wave packet (describing here the state of the pointer) spread in time. But recall that our single pointer particle is just a schematic way of describing what is, in fact, some enormous macroscopic collection of elementary particles. We should thus probably attribute to our pointer particle a very large mass M. This means – recall Eq. 2.53 – that the packet will spread very slowly. Or even more simply, it warrants making the following approximation:

$$\hat{H}_y \approx 0. \tag{3.10}$$

The third (and here most important) contribution to the Hamiltonian will be the one describing the *interaction* of the particle and the pointer. Of course, in reality, this interaction would be quite indirect and quite complicated, mediated somehow

by all the *other* particles composing the apparatus. For our schematic treatment, though, all we really want to ensure is the following: *if* the particle were to start out definitely in an energy eigenstate ψ_n, then the pointer should move sideways by a distance proportional to E_n, the corresponding energy eigenvalue. That, after all, is what pointers on measuring devices *do*, rather by definition – they register the outcome of the measurement by their positions, and if the particle starts out in an energy eigenstate, then we know the outcome will definitely be that corresponding eigenvalue.

It turns out that an interaction Hamiltonian of the following form will achieve this:

$$\hat{H}_{int} = \lambda \hat{H}_x \hat{p}_y = -i\hbar\lambda \hat{H}_x \frac{\partial}{\partial y}. \tag{3.11}$$

Here λ is a constant; $\hat{H}_x$ is the energy operator for the particle in the box, i.e., $\hat{H}_x$ is the operator corresponding to the quantity that we are *measuring*; and $\hat{p}_y$ is the momentum operator for the pointer.

Let's see why this works. Suppose again that the particle starts out in an energy eigenstate, so that the particle-pointer initial wave function is

$$\Psi(x, y, 0) = \psi_n(x)\phi(y). \tag{3.12}$$

Now for simplicity assume that λ (which describes the strength of the interaction) is very large, so that – during the period in which the interaction is occurring – we can ignore any other terms in the overall Hamiltonian. Then the Schrödinger equation reads

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}_{int}\Psi = -i\hbar\lambda \hat{H}_x \frac{\partial \Psi}{\partial y}. \tag{3.13}$$

Simplifying a bit gives

$$\frac{\partial \Psi}{\partial t} = -\lambda \hat{H}_x \frac{\partial \Psi}{\partial y}. \tag{3.14}$$

One sees that the t- and y-dependencies of Ψ are coupled, but the variable x is not involved. We may thus assume that $\Psi(x, y, t)$ remains proportional to $\psi_n(x)$ and hence remains an eigenstate of $\hat{H}_x$, so that the previous equation simplifies to

$$\frac{\partial \Psi}{\partial t} = -\lambda E_n \frac{\partial \Psi}{\partial y}. \tag{3.15}$$

It is then straightforward to check that the solution is

$$\Psi(x, y, t) = \psi_n(x)\phi(y - \lambda E_n t). \tag{3.16}$$

Suppose the interaction lasts until a time $t = T$. Then the quantum state of the particle-pointer system at the end of the interaction is evidently

$$\Psi(x, y, T) = \psi_n(x)\phi(y - \lambda E_n T) \tag{3.17}$$

which can be understood as follows: the particle-in-the-box remains in the n^{th} energy eigenstate, and the pointer remains a Gaussian wave packet but whose center has *moved*, to the right, a distance $\lambda E_n T$ so that it is now centered at $y_0 + \lambda E_n T$ which we can identify as y_n – the final location of the pointer when it registers the n^{th} energy value.

To summarize, the (admittedly weird-looking) interaction Hamiltonian, Eq. (3.11), does exactly the job we wanted: it makes the pointer *move* a distance proportional to the energy of the particle when the particle actually begins with a particular, definite energy. So it seems like this model – with this interaction Hamiltonian – provides a schematic, but still faithful, way of capturing all the complicated physical interactions that in fact couple the particle-in-the-box to the apparatus pointer in this kind of situation.

(A technical aside: if you are worried that we dropped something important, either by ignoring $\hat{H}_x$ during the measurement interaction, or by setting the mass of the pointer to infinity and hence ignoring $\hat{H}_y$, you shouldn't be. Including $\hat{H}_x$ would only have the effect of giving an extra factor $e^{-iE_n t/\hbar}$ in Eq. (3.17) – a meaningless overall phase. And if we had included a term corresponding to the kinetic energy of the pointer, this would only have the effect of making the pointer wave packet spread slightly during the course of the interaction. This also doesn't really change anything important, so the approximations made above really do seem to capture what is essential. You are invited to explore this in the Projects if you want to.)

So far so good. But of course we are not so interested in the special case where the particle starts out with a definite energy. We want to know what happens when the particle starts out in the superposition state, Eq. (3.4), and we try to treat the measurement interaction fully quantum mechanically – i.e., without bringing in any *ad hoc* extra *postulates* about exceptions to the Schrödinger evolution, the pointer always having a classical position, etc. Remember, the hope is that, if we just use the purely microscopic part of quantum mechanics – wave functions obeying Schrödinger's equation – to describe the entire interaction between the two systems, everything will work out the way we want it to: the final quantum state will attribute an approximately-definite position to the pointer, the wave function of the particle-in-a-box will be one of the energy eigenstates, etc.

But, sadly, our hope is immediately dashed. It is very easy to see – from the fact that the overall Schrödinger equation is *linear* – that with

$$\hat{H} = \hat{H}_{int} \tag{3.18}$$

and

$$\Psi(x, y, 0) = \left(\sum_i c_i \psi_i(x)\right)\phi(y) \tag{3.19}$$

the wave function at time T (when the measurement interaction ceases) is

$$\Psi(x, y, T) = \sum_i c_i \ \psi_i(x) \ \phi(y - \lambda E_i T). \tag{3.20}$$

This represents an *entangled superposition* of several states, in each of which the particle has a definite energy and the pointer's position is slightly fuzzy but centered on a definite position corresponding to the energy of the particle. But of course, the energy of the particle, and the post-measurement position of the pointer, are *different* in each of the superposed states. And that is seriously problematic. The particle-in-a-box does not end up in a particular energy eigenstate at all, and (worse!) the pointer is not localized around any particular one of (what we thought of previously as) its possible final positions.

To summarize: if you try to treat the measurement process using just the microscopic part of quantum mechanics, it simply doesn't give you what you want, which is some kind of explanation for the emergence of one definite outcome (pointer position). Instead of somehow *resolving* the initial ambiguity in the energy of the particle, and thereby causing the particle to end up with a definite energy and the pointer to end up in a definite place, the interaction between the particle-in-the-box and the pointer *infects* the pointer with its quantum ambiguity!

Now you might think that this result is simply a consequence of our having treated the measurement process so schematically. Perhaps a more detailed, more realistic, quantum mechanical description of the measuring apparatus and its interaction with the quantum system, would yield the desired result? Unfortunately it is easy to see that this cannot possibly work. You can include as many of those intermediate degrees of freedom as you want, making the description as realistic and as accurate as you like, and still it won't make any difference to the outcome. Here's why. Suppose we include another intermediate degree of freedom, called z, so that now the initial state is something like

$$\Psi(x, y, z, 0) = \left(\sum_i c_i \psi_i(x) \right) \phi(y) \chi_0(z). \tag{3.21}$$

Schrödinger's equation reads

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H} \Psi. \tag{3.22}$$

Since we mean to be describing a *measurement* of the particle's energy, we demand that the Hamiltonian $\hat{H}$ have the property that if

$$\Psi(x, y, z, 0) = \psi_n(x) \phi(y) \chi_0(z) \tag{3.23}$$

then

$$\Psi(x, y, z, T) = \psi_n(x) \phi(y - \lambda E_n T) \chi_n(z). \tag{3.24}$$

That is, we demand that the position of the pointer should move, by an amount proportional to E_n, when the particle starts out in an energy eigenstate with energy

E_n. (And the intermediate degree of freedom ends up in some associated state.) But then it is obvious, again from the linearity of the Schrödinger equation, that in the general case where the particle starts out in a superposition of energies, as in Eq. (3.21), it ends up in the entangled superposition state

$$\Psi(x, y, z, T) = \sum_i c_i \psi_i(x) \phi(y - \lambda E_i T) \chi_i(z). \tag{3.25}$$

The intermediate degree of freedom just gets infected with the superposition, too. Including it – or, indeed, including as many such intermediate degrees of freedom as you might want and thereby making the description as complete and accurate and unschematic as you might want – changes nothing. Our simple schematic treatment brought out the essential and inevitable problem already.

3.3 Schrödinger's Cat and Einstein's Bomb

The most famous illustration of the problem described in the last section was presented by Schrödinger in the last section of his 1935 paper, "The present situation in quantum mechanics." [3] Actually, in that paper, Schrödinger illustrates the problem several times, with several different examples. Here, for instance, is his discussion of the case of a radioactive nucleus emitting an alpha particle:

> [the ψ-] function has provided quite intuitive and convenient ideas, for instance the 'cloud of negative electricity' around the nucleus, etc. But serious misgivings arise if one notices that the uncertainty affects macroscopically tangible and visible things, for which the term 'blurring' seems simply wrong. The state of a radioactive nucleus is presumably blurred in such degree and fashion that neither the instant of decay nor the direction, in which the emitted α-particle leaves the nucleus, is well-established. Inside the nucleus, blurring doesn't bother us. The emerging particle is described, if one wants to explain intuitively, as a spherical wave that continuously emanates in all directions from the nucleus and that impinges continuously on a surrounding luminescent screen over its full expanse. The screen however does not show a more or less constant uniform surface glow, but rather lights up at *one* instant at *one* spot – or, to honor the truth, it lights up now here, now there, for it is impossible to do the experiment with only a single radioactive atom. If in place of the luminescent screen one uses a spatially extended detector, perhaps a gas that is ionised by the α-particles, one finds the ion pairs arranged along rectilinear columns, that project backwards on to the bit of radioactive matter from which the α-radiation comes (C.T.R. Wilson's cloud chamber tracks, made visible by drops of moisture condensed on the ions) [3].

The idea here is illustrated in Fig. 3.3. Schrödinger's point is that, according to quantum mechanics, the α particle emitted by a radioactive nucleus is not going in any particular direction. Instead, the theory describes it as coming out of the nucleus in a superposition of all possible directions, which is mathematically equivalent to a spherically symmetric, outward-propagating wave function. It is easy to understand how a flash – at a particular location on the screen – could be created by an α particle that had been emitted in a particular direction, namely, toward that particular spot.

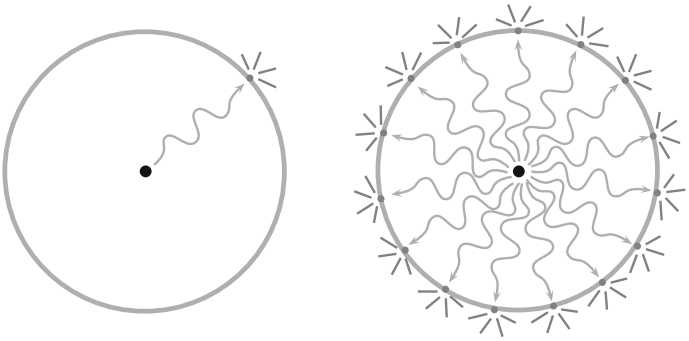


Fig. 3.3 A radioactive nucleus emits an alpha particle which then causes a visible scintillation on a surrounding circular detection screen. The left panel shows that, if the alpha particle is emitted in a particular direction, the scintillation will occur at a spot on the screen in that same direction. But if both the alpha particle and the detection screen are treated quantum mechanically, the linearity of Schrödinger's equation implies that if the alpha particle is emitted in a superposition of different directions (say, a spherically symmetric wave function propagating outward) the final quantum state will be an entangled superposition involving terms with scintillations in all directions on the screen. That is, the microscopic quantum dynamics cannot explain how it is that the flash is seen to occur at just some one particular spot on the screen

But if the emission of the alpha particle is really somehow spherically symmetric, there seems to be nothing in the microscopic part of quantum mechanics to break the symmetry and explain the flash occuring at a particular spot.

The situation here can be described quantum mechanically in the same way we described the measurement process in the last section. Take ψ_n to be the wave function of an α-particle that has a reasonably sharply-defined propagation direction θ_n, and take ϕ_0 to be the wave function of a photo-luminescent screen on which no flashes have yet appeared. Then the idea is that the overall wave function (for the α particle and screen jointly) would evolve, under Schrödinger's equation, as follows:

$$\psi_n \phi_0 \rightarrow \psi_n \phi_n \tag{3.26}$$

where ϕ_n is the wave function for the photo-luminescent screen with a bright flash at angle θ_n.

But then this immediately implies – from the linearity of Schrödinger's equation – that if (as is in fact the case in this kind of situation) the wave function of the α particle is a (say, spherically symmetric) superposition

$$\psi_{sph} \sim \sum_i \psi_i \tag{3.27}$$

the state will evolve, under Schrödinger's equation, as follows:

$$\psi_{sph}\phi_0 \rightarrow \sum_i \psi_i \phi_i . \tag{3.28}$$

This is of course an entangled superposition of states; each term in the superposition has the alpha particle being emitted in a particular direction and the screen flashing at a particular point, but the superposition as a whole includes such terms corresponding to all possible directions. No one particular direction is picked out, either for the alpha particle *or for the flash*. As Schrödinger points out, though, this democratic wave function does not seem to correspond appropriately to what we actually observe in this kind of case, which is a flash at some particular definite location: "The screen however does not show a more or less constant uniform surface glow, but rather lights up at *one* instant at *one* spot" [3].

Schrödinger immediately continues the discussion by describing the case of the famous cat:

> One can even set up quite ridiculous cases. A cat is penned up in a steel chamber, along with the following diabolical device (which must be secured against direct interference by the cat): in a Geiger counter there is a tiny bit of radioactive substance, *so* small, that *perhaps* in the course of one hour one of the atoms decays, but also, with equal probability, perhaps none; if it happens, the counter tube discharges and through a relay releases a hammer which shatters a small flask of hydrocyanic acid. If one has left this entire system to itself for an hour, one would say that the cat still lives *if* meanwhile no atom has decayed. The first atomic decay would have poisoned it. The ψ-function of the entire system would express this by having in it the living and the dead cat (pardon the expression) mixed or smeared out in equal parts [3].

To put some formalism to this, we would say that, in the course of the hour, the single radioactive atom evolves (according to Schrödinger's equation) into an equally-weighted superposition of decayed and not-yet-decayed states:

$$\psi_0 \rightarrow \frac{1}{\sqrt{2}}\,\psi_{\text{decayed}} + \frac{1}{\sqrt{2}}\,\psi_{\text{not decayed}}. \tag{3.29}$$

If we treated the surrounding equipment *classically*, and applied the collapse postulate, we would say that when the atom interacts with (say) the Geiger counter, this interaction triggers a collapse and it thereby becomes unambiguously the case that *either* the atom has decayed and the Geiger counter has clicked, *or* the atom has not decayed and the Geiger counter has not clicked. Then, a classical interaction (mediated by the hammer and flask of acid) between the Geiger counter and the cat would result in the cat definitely being alive if the Geiger counter did not click, and the cat definitely being dead if the Geiger counter did click.

However, if we instead treat the surrounding apparatus quantum mechanically, we find that the final state is something like the following:

$$\Psi_f = \frac{1}{\sqrt{2}}\,\psi_{\text{decayed}}\,\phi_{\text{shattered}}\,\chi_{\text{dead}} + \frac{1}{\sqrt{2}}\,\psi_{\text{not decayed}}\,\phi_{\text{intact}}\,\chi_{\text{alive}}. \tag{3.30}$$

This is a superposition of two states: (i) a state in which the atom is decayed, the hammer is down and the flask of poison is shattered, and the cat is dead; and (ii) a state in which the atom is not decayed, the hammer is up and the flask of poison is intact, and the cat is alive. In particular, this is not a state in which there is any definite fact of the matter about whether the cat is dead or alive. The poor cat is, in Schrödinger's words, "mixed or smeared out" between living and dead.

As a bit of historical context, it is perhaps interesting to note that in the weeks leading up to Schrödinger's submitting the paper containing the cat example, he was exchanging letters with Einstein. And in one of his letters to Schrödinger (from August 8, 1935), Einstein suggested an example that illustrates the same basic point illustrated by the cat:

> The system is a substance in chemically unstable equilibrium, perhaps a charge of gunpowder that, by means of intrinsic forces, can spontaneously combust, and where the average life span of the whole setup is a year. In principle this can quite easily be represented quantum-mechanically. In the beginning the ψ-function characterizes a reasonably well-defined macroscopic state. But, according to your equation, after the course of a year this is no longer the case at all. Rather, the ψ-function then describes a sort of blend of not-yet and of already-exploded systems. Through no art of interpretation can this ψ-function be turned into an adequate description of a real state of affairs; [for] in reality there is just no intermediary between exploded and not-exploded [4, p. 78].

So perhaps the basic idea of the "Schrödinger's cat" example actually started with Einstein? This is again suggested by a later letter in which Einstein seems to get slightly confused and mixes the two examples together in an amusing way:

> Dear Schrödinger,
>
> .... I am as convinced as ever that the wave representation of matter is an incomplete representation of the state of affairs, no matter how practically useful it has proved itself to be. The prettiest way to show this is by your example with the cat (radioactive decay with an explosion coupled to it). At a fixed time parts of the ψ-function correspond to the cat being alive and other parts to the cat being pulverized.
>
> If one attempts to interpret the ψ-function as a complete description of a state, independent of whether or not it is observed, then this means that at the time in question the cat is neither alive nor pulverized. But one or the other situation would be realized by making an observation.
>
> If one rejects this interpretation then one must assume that the ψ-function does not express the real situation but rather that it expresses the contents of our knowledge of the situation. This is Born's interpretation, which most theorists today probably share. But then the laws of nature that one can formulate do not apply to the change with time of something that exists, but rather to the time variation of the content of our legitimate expectations.
>
> Both points of view are logically unobjectionable; but I cannot believe that either of these viewpoints will finally be established.
>
> There is also the mystic, who forbids, as being unscientific, an inquiry about something that exists independently of whether or not it is observed, i.e., the question as to whether or not the cat is alive at a particular instant before an observation is made (Bohr). Then both interpretations fuse into a gentle fog, in which I feel no better than I do in either of the previously mentioned interpretations, which do take a position with respect to the concept of reality.

> I am as convinced as ever that this most remarkable situation has come about because we have not yet achieved a complete description of the actual state of affairs.
>
> Of course I admit that such a complete description would not be observable in its entirety in the individual case, but from a rational point of view one also could not require this....
>
> Best regards from
>
> Yours, A. Einstein [5, pp. 35-6]

In any case, whoever deserves credit for originating the example, it is nice to know that Schrödinger and Einstein agreed about what it established. Here, for example, is Einstein – again writing to Schrödinger – now on Sept 4, 1935, just after Schrödinger submitted his manuscript:

> ...your cat shows that we are in complete agreement concerning our assessment of the character of the current theory. A ψ-function that contains the living as well as the dead cat just cannot be taken as a description of a real state of affairs. To the contrary, this example shows exactly that it is reasonable to let the ψ-function correspond to a statistical ensemble that contains both systems with live cats and those with dead cats [4, p. 84].

For Einstein and Schrödinger, then, the cat/bomb example strongly suggested that, rather than providing direct, literal, complete descriptions of physical systems, quantum mechanical wave functions should instead be understood as describing our incomplete knowledge – our ignorance – about the physical states of these systems. As Einstein puts it here, the wave function (attributing, typically, a *range* of possible values to various system properties) should not be understood as a complete description of an individual system, but should instead be understood as characterizing a statistical ensemble of systems with some variation among the individual members of the ensemble. We will discuss this alternative viewpoint in more detail in the following section.

But, after all this talk of dead cats and exploding bombs, let's close this section with a slightly-happier image of Schrödinger's cat as re-envisaged by Bell. In Bell's re-telling – See Fig. 3.4 – the poison which is either released, or not, is replaced by

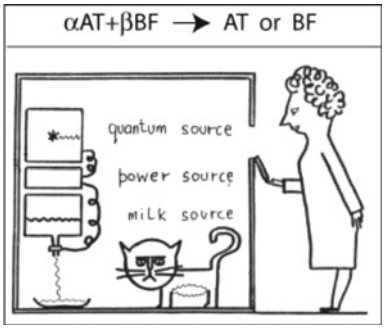


Fig. 3.4 Bell's version of Schrödinger's cat. The state of the radioactive nucleus ("*A*" for "not decayed" and "*B*" for "decayed") becomes entangled with the delivery (or not) of milk into the cat's dish and thereby also with the size of the cat's stomach ("*T*" for "thin" and "*F*" for "fat"). From Ref. [6]. Figure © IOP Publishing. Reproduced with permission. All rights reserved. https://doi.org/10.1088/1751-8121/40/12/S02

a portion of milk that is either released, or not, into a dish, for the cat to drink. If the nucleus decays, the cat gets fed and ends up fat – whereas if the nucleus does not decay, the cat does not get fed and ends up thin. But in either case he survives!

3.4 Hidden Variables and the Ignorance Interpretation

In recent popular culture, Schrödinger's cat has become a kind of symbol or emblem of the weirdness of quantum mechanics. Many people therefore have the impression that Schrödinger thought the cat really *would* end up "mixed or smeared out" between live and dead, and that one simply had to accept this as true despite its incomprehensibility. But that is completely wrong. In fact, Schrödinger intended the cat thought experiment as a *reductio ad absurdum* of the idea that quantum mechanical wave functions provide *complete* descriptions of physical systems. The idea was that, in this kind of case, quantum mechanics implies something that is (at least in the opinion of Schrödinger and Einstein) *obviously wrong*. The pointer does *not* in fact end up in some kind of superposition of different locations, but rather points to one particular spot. The screen around the radioactive nucleus "does not show a more or less constant uniform surface glow, but rather lights up at *one* instant at *one* spot". And the cat is most certainly either (fully, definitely) alive or (fully, definitely) dead – not both alive and dead, "mixed or smeared out in equal parts." [3]

Schrödinger summarizes his point as follows just after presenting the cat example:

> It is typical of these cases that an indeterminacy originally restricted to the atomic domain becomes transformed into macroscopic indeterminacy, which can then be *resolved* by direct observation. That prevents us from so naively accepting as valid a 'blurred model' for representing reality. In itself it would not embody anything unclear or contradictory. There is a difference between a shaky or out-of-focus photograph and a snapshot of clouds and fog banks [3].

He continues shortly after by noting that, with the cat example,

> we saw that the indeterminacy is not even an actual blurring, for there are always cases where an easily executed observation provides the missing knowledge [3].

I would summarize Schrödinger's point here this way. If we describe the entire measurement process using the microscopic part of quantum mechanics, the theory tells us that the measuring apparatus (or some other macroscopic object like the cat) ends up in its own ambiguous, superposed state. But *we know this cannot be a complete description of the state of such things* since direct observation reveals that such objects are always in perfectly definite states. Therefore, at least when it is used to describe the state of macroscopic things, the quantum mechanical description *cannot be complete*: the ambiguity of quantum superposition must (as Einstein also remarked in the letter quoted in the previous section) refer to *our ignorance* about which of several possibilities is in fact realized, as opposed to describing an objective blurring in the physical state of the object itself. This, I think, is the point of the intriguing sentence "There is a difference between a shaky or out-of-focus photograph and a snapshot of

clouds and fog banks." He is suggesting that quantum mechanical wave functions, thought of as describing or depicting the objective physical states of things, are *not* like (sharp, in-focus) photographs of clouds – i.e., faithful reproductions of things which are themselves, objectively, smeared out and fuzzy. Instead, he means to suggest, we should understand quantum mechanical wave functions as like "shaky or out-of-focus photograph[s]" of objects that are, in themselves, perfectly sharp. In this kind of case, the smeared out or fuzzy character does not pertain to the object described, but is instead a kind of failing or imperfection in the reproduction process.

But surely there is no fundamental distinction between microscopic and macroscopic systems (the latter, after all, literally being made of the former). This, I think, is the point of including, in the cat example, the detailed description of the intermediate parts of the mechanism – the causal chain – whereby the state of the nucleus is coupled to the state of the cat. Surely, Schrödinger invites us to think, there is no particular spot along this continuous chain between micro and macro where it would make sense to draw a sharp line and say "different dynamical laws start applying *here*". But if the micro and the macro must be treated uniformly, and if the quantum mechanical description (in terms of wave functions) of macroscopic systems (like cats and pointers) is not complete, then surely this is also the case for microscopic systems.

If that's right, then, for example, when we say that the particle-in-the-box is in a superposition of several different energy eigenstates,

$$\psi_0 = c_1\psi_1(x) + c_2\psi_2(x) + c_3\psi_3(x), \tag{3.31}$$

what this must *mean* is that the particle is *either* in the state $\psi_1(x)$ *or* the state $\psi_2(x)$ *or* the state $\psi_3(x)$... we're just not sure which one! It's not that the energy of the particle is somehow blurred or indefinite – rather, it's only our *knowledge* which is blurred or indefinite. The energy is *uncertain* (in the literal sense, meaning "unknown to us") but it is perfectly sharp, some one definite value or another, in reality. And then, if the energy is measured, we simply *find out* what the energy was all along.

Or similarly, when we say that the wave function of the emitted alpha particle is spherically symmetric, what this *means*, according to this viewpoint, is just that we have no idea which direction the alpha particle is going. The subjective probability distribution we would assign to its direction is spherically symmetric, but the thing itself isn't! The alpha particle itself, on this view, is already moving in some one particular direction – we don't know which one, but it is perfectly definite in reality all the same. Seeing a flash at some particular spot on the surrounding screen is then not a big mystery and not a proof that the microscopic quantum dynamics is wrong... it's simply the way we *find out* which direction the alpha particle was going all along.

In its simplest (or, one might say, most naive possible) form, this view might be called the "ignorance interpretation of superposition". I think it should be admitted that it has a certain alluring reasonableness. Indeed, for some people reading this, it may be the view that you have had in mind all along! By getting rid of any "quantum fuzziness" at the root, down at the microscopic scale, the "ignorance interpretation of superposition" totally pre-empts the difficulty, illustrated by Schrödinger's cat, of

amplifying the fuzziness up to a macroscopic scale where, apparently, it conflicts with our direct experience of the world.

Notice also that the ignorance interpretation provides a beautifully simple resolution of our earlier worries about the collapse postulate: if wave functions are not really descriptions of the physical states of systems at all, but instead descriptions of the state of our *knowledge* of those systems, then there is nothing remotely problematic about wave functions collapsing. The collapse is simply an updating of our knowledge, when we get new data! So for example when we measure the energy of the particle-in-the-box and its wave function collapses from $\psi_0 = c_1\psi_1 + c_2\psi_2 + c_3\psi_3$ to, say, ψ_2, this was not a dynamical process, a change in the physical state of the particle, at all. The measurement simply *reveals* something that was there all along but unknown to us. We simply learn the value of the particle's energy, which we did not know before, and so update ψ – our "knowledge catalog" – accordingly. All the worries about inconsistent dynamical rules, and ambiguities about which ones should apply when, and so on – all of these simply evaporate if we adopt the ignorance interpretation.

This was essentially the view put forward by Max Born in 1926. Born's view morphed significantly as it became bound up with the Copenhagen Interpretation that we will discuss in Chap. 6, but the *original* "Born interpretation" was nothing but the ignorance interpretation we have discussed here. (See Sect. 2.4 of Ref. [7] for a nice overview.) And this view continued to enjoy support (from those who resisted the Copenhagen orthodoxy) in subsequent decades. For example, in an essay written in 1949, just a few years before his death in 1955, Einstein seemed to again advocate something along these lines:

> Within the framework of statistical quantum theory there is no such thing as a complete description of the individual system. More cautiously it might be put as follows: The attempt to conceive the quantum-theoretical description as the complete description of the individual systems leads to unnatural theoretical interpretations, which become immediately unnecessary if one accepts the interpretation that the description refers to ensembles of systems and not to individual systems. In that case the whole 'egg-walking' performed in order to avoid the 'physically real' becomes superfluous. There exists, however, a simple psychological reason for the fact that this most nearly obvious interpretation is being shunned. For if the statistical quantum theory does not pretend to describe the individual system (and its development in time) completely, it appears unavoidable to look elsewhere for a complete description of the individual system; in doing so it would be clear from the very beginning that the elements of such a description are not contained within the conceptual scheme of the statistical quantum theory. With this one would admit that, in principle, this scheme could not serve as the basis of theoretical physics. Assuming the success of efforts to accomplish a complete physical description, the statistical quantum theory would, within the framework of future physics, take an approximately analogous position to the statistical mechanics within the framework of classical mechanics. I am rather firmly convinced that the development of theoretical physics will be of this type; but the path will be lengthy and difficult [8, p. 671].

So this point of view not only seems quite sensible but also seems to have a strong pedigree.

But, unfortunately, the ignorance interpretation – at least in its simplest form – cannot possibly be right. It would seem to imply, for example, that in the double

slit experiment described in the last chapter, in the middle of which the particle's quantum state is a superposition of "going through slit 1" and "going through slit 2"

$$\psi = \frac{1}{\sqrt{2}}\left[\psi_{\text{slit 1}} + \psi_{\text{slit 2}}\right], \tag{3.32}$$

there is nevertheless some fact of the matter about which slit the particle *really* went through: *either* slit 1 *or* slit 2. But if each particle is really just a literal particle which has a perfectly definite (if sometimes unknown) position, it is (to understate it) very difficult to understand how the subsequent particle locations could form an *interference pattern*. The interference pattern strongly – indeed, I think, conclusively – establishes that the quantum wave function is really something physical, something real, not just a description of our incomplete state of knowledge.

In addition, there are a number of rigorous mathematical theorems proving that it is impossible to assign definite values to quantum properties in, at least, the naive way suggested by the ignorance interpretation. Here we briefly indicate the flavor of these so-called "no hidden variable theorems" with a simple example. (See the Projects for two additional examples discussed already by Schrödinger in his 1935 paper.)

Consider the case of a single spin-1/2 particle whose spin might be measured along the $\hat{z}$, $\hat{x}$, or $\hat{n}$ directions (where $\hat{n}$ is in the $x - z$ plane and halfway between $\hat{x}$ and $\hat{z}$, i.e., 45° away from both). The operators corresponding to the particle's spin along these three directions are

$$\hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.33}$$

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{3.34}$$

and

$$\hat{\sigma}_n = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{\sqrt{2}}\left[\hat{\sigma}_x + \hat{\sigma}_z\right]. \tag{3.35}$$

Now if the particle has a definite spin along one of these directions (i.e., if its wave function is an eigenstate of one of the three operators) it will be in a superposition of "spin-up" and "spin-down" with respect to the other two directions. Thus, according to the idea that the wave function provides a complete description of the physical state of the particle, a particle can never possess a definite value of spin along all three of these directions at once.

But, according to the ignorance interpretation of superposition, the fact that the quantum state is a superposition does not mean that the particle doesn't have a definite value of spin. So we contemplate the possibility that particles have definite spin values, s_x, s_y, and s_n, along all three directions at once. These values would in general not all be known at once, although presumably the idea is that a measurement

of the corresponding quantity will simply reveal the value in question. Since (let's say) such measurements have outcomes $+1$ (meaning "spin-up") or -1 (meaning "spin-down"), we should assign either the value $+1$ or -1 to each of s_x, s_y, and s_n.

Furthermore, there is some reason to expect that the values of the spin along these three directions should obey the same mathematical relationships that are obeyed by the corresponding operators: the quantum mechanical expectation values, for example, will have this relationship, and it is plausible to suspect that, in the context of a "hidden variable" theory like we are contemplating here, the average values will have this relationship because each individual set of possible values has this relationship. Equation (3.35) would then lead us to demand that

$$s_n = \frac{1}{\sqrt{2}}\left[s_x + s_z\right]. \tag{3.36}$$

But it is immediately obvious that it will be mathematically impossible to assign values $+1$ or -1 to s_x, s_z, and s_n in accordance with Eq. (3.36): the quantity in square brackets on the right will be -2, 0, or $+2$, and none of these, divided by $\sqrt{2}$ is $+1$ or -1.

The argument just sketched is closely related to the first "no hidden variables" proof, given by John von Neumann in 1932. Historically, von Neumann's argument convinced many people that the type of ignorance interpretation favored by Schrödinger and Einstein was untenable, and thus provided a nudge in the direction of the Copenhagen interpretation (which we will study in greater depth in Chap. 6). As it turns out, though, von Neumann's argument is rather impotent. This was first pointed out in 1935 by Grete Hermann, but her critique tragically failed to gain any traction in the physics community. [7] Several decades later, John Stewart Bell independently tackled the question of whether "hidden variable" theories had been mathematically refuted; he would later describe von Neumann's proof as "not merely false but *foolish*!" [9].

The reason for this harsh assessment has to do with the requirement that the "hidden variables" should obey the same mathematical relationships as their corresponding quantum mechanical operators – a requirement which perhaps makes sense for *commuting* operators, but which is completely unmotivated for sets of *non-commuting* operators like $\hat{\sigma}_x$, $\hat{\sigma}_z$, and $\hat{\sigma}_n$.

In more recent decades, "no hidden variables" proofs have been found whose assumptions are somewhat more reasonable. Getting into the details here would take us too far afield, but suffice it to say it is seriously problematic to think that one can understand all measurements as simply revealing the values of properties that are, while unknown, perfectly definite and independent of the measurement procedure itself. The naive ignorance interpretation, that is, is really not tenable. (Crucial references here include Bell's paper "On the problem of hidden variables in quantum mechanics" [10] and Mermin's review article [9].) As we will see in Chap. 7, however, there does exist a perfectly viable "hidden variable" theory with

a crucial property called "contextuality" (meaning, in a nutshell, that some *but not all* measurements simply reveal the pre-existing values of properties) that allows it to elude all of these impossibility theorems!

3.5 Wrap-Up

For now, let us try to recap, summarize, and package what we've seen in this Chapter. It will probably be helpful, for example, to step back and try to get clear on the answer to the following question:

What, exactly, is "the measurement problem"?

It is admittedly a little confusing, because this phrase is used to refer to several inter-related things, all of which we have touched on here.

To begin with, sometimes "the measurement problem" refers to the fact that the postulates of textbook quantum mechanics include statements about "measurements" (and their outcomes), even though "measurement" is a very fuzzy and human concept. That is, it simply is not clear exactly which set of physical interactions or processes in nature should count as "measurements". And so, until or unless this is somehow clarified, it simply isn't clear exactly what the theory is even saying. This is the point Bell had in mind when he wrote that "[t]he concept of 'measurement' becomes so fuzzy on reflection that it is quite surprising to have it appearing in physical theory *at the most fundamental level*." [11] Or, as he put it elsewhere – somewhat less diplomatically – "conventional formulations of quantum theory, and of quantum field theory in particular, are unprofessionally vague and ambiguous" [12].

Then there is a closely-related, second, meaning to (or aspect of) "the measurement problem". Even if the notion of "measurement" were somehow given a clear and precise meaning – even if, that is, a sharp boundary were somehow drawn between "measurements" and "non-measurements" so that it became unambiguous when to apply which part of the quantum formalism – there would still be something unbelievable about the idea that there are these two fundamentally distinct types of processes. Equivalently (since the difference between the two supposedly distinct types of processes has to do with whether a microscopic system is, or is not, interacting appropriately with something from the other, macroscopic, "realm") there is something unbelievable about the idea that the world is fundamentally "split" into these two distinct "realms". Surely a proper fundamental theory should describe the *entire* universe in a coherent, unified way. And so, in this aspect, "the measurement problem" refers to the failure of standard quantum theory to provide such a unified description. Bell often remarked that quantum mechanics involved what he described as a "shifty split". For example:

> There can be no question then of identifying the quantum system S with the whole world W. There can be no question – without changing the axioms – of getting rid of the shifty split. Sometimes some authors of 'quantum measurement' theories seem to be trying to do just that. It is like a snake trying to swallow itself by the tail. It can be done – up to a point.

> But it becomes embarrassing for the spectators even before it becomes uncomfortable for the snake [13].

And Schrödinger as well had already criticized the idea that "measurement" was somehow a special, dynamically distinct kind of process, with its own special dynamical rules. In quantum mechanics, he wrote,

> any *measurement* suspends the law that otherwise governs continuous time-dependence of the ψ-function and brings about in it a quite different change, not governed by any law but rather dictated by the result of the measurement. But laws of nature differing from the usual ones cannot apply during a measurement, for objectively viewed it is a natural process like any other, and it cannot interrupt the orderly course of natural events. Since it does interrupt that of the ψ-function, the latter ... can *not* serve ... as an experimentally verifiable representation of an objective reality [3].

The idea that, "objectively viewed", "measurement ... is a natural process like any other" perfectly captures this second aspect of "the measurement problem."

And then, finally, "the measurement problem" also sometimes denotes the theory's apparent inability to provide sensible results when it is modified in the obvious way in response to the criticisms of the previous paragraphs. This is the aspect that is illustrated by Schrödinger's cat. If we refuse to accept the "fractured universe" implied by the most straightforward reading of the quantum formalism, the easiest way to try to solve *that* problem is to simply get rid of the postulates about "measurement" (and the separately-presupposed classical "realm") and retain just the microscopic part of the theory. In this modified understanding of the theory, *everything* will be described in terms of wave functions obeying Schrödinger's equation *always*, and so, to be sure, we have a coherent, unified worldview. But the problem – as we saw – is that this worldview simply doesn't seem to be right. Detection screens do not "show a more or less constant uniform surface glow", pointers on measuring devices are never blurry, and cats are never observed to be "mixed or smeared out in equal parts" of living and dead.

Sometimes this last aspect of "the measurement problem" is expressed by noting that the theory does not seem to be able to explain the occurence of definite measurement results. That is fine as far as it goes, but it can also be confusing or misleading. The full, original theory – with "collapse postulate" and all – certainly has no difficulty explaining the occurrence of definite measurement results! Indeed, they are right there in the postulates of that version of the theory! But that is precisely the problem: those rules appear to have been implausibly put in by hand, to avoid embarrassment, and (for the reasons I summarized in what I described as the first two aspects of "the measurement problem") it seems impossible to take them seriously as fundamental physical laws.

At the end of the day, the measurement problem is probably best understood as the problem of understanding the seemingly paradoxical relation of the collapse postulate to the rest of quantum theory. On the one hand, it seems impossible to include the collapse postulate in the axioms of the theory and still regard the theory as providing a fundamental account of the microscopic world. On the other hand, it seems impossible to eliminate the collapse postulate from the dynamical axioms of

the theory, either by simply jettisoning it (and letting wave functions evolve according to the Schrödinger equation all the time) or by interpreting it (and the wave function generally) as pertaining not to reality itself but only to our knowledge of reality.

In later Chapters, we will explore some concrete proposals for resolving (or dissolving) the measurement problem. One of these – Everett's "Many Worlds Interpretation", the subject of Chap. 10 – attempts to retain the idea of the quantum descriptions of reality (in terms of wave functions alone, obeying Schrödinger's equation always) being complete. Another – the "Pilot-Wave Theory" of de Broglie and Bohm, the subject of Chap. 7 – proposes to supplement the wave function with additional ("hidden") variables that resolve the dilemmas posed by Schrödinger, but in a way that avoids the "no hidden variable" theorems discussed here. A third proposal – the "Spontaneous Collapse" theory discussed in Chap. 9 – attempts to unify the Schrödinger equation and the collpase postulate to give a single uniform dynamical description that can be applied coherently at all scales. And then there is also a philosophical perspective – the "Copenhagen Interpretation" of Bohr and Heisenberg, discussed in Chap. 6 – that urges us to reject Schrödinger's worries as somehow baseless and meaningless.

Before turning to these proposals, however, we explore in the following two Chapters two additional problems that seem to afflict textbook quantum theory.

Projects:

3.1 Show explicitly that Eq. (3.20) indeed satisfies the Schrödinger equation with $\hat{H} = \hat{H}_{int} = \lambda \hat{H}_x \hat{p}_y$.

3.2 Suppose, in our schematic formal treatment of the measurement of the energy of a particle-in-a-box, we use the more complete Hamiltonian operator

$$\hat{H} = \hat{H}_x + \hat{H}_y + \hat{H}_{int} \tag{3.37}$$

(with M finite so $\hat{H}_y$ cannot just be ignored). What now is the solution to the Schrödinger equation with $\Psi(x, y, 0)$ still given by Eq. (3.6)?

3.3 Sketch some configuration space cartoons – in the style of Figs. 2.10 and 2.11 – to illustrate the evolution of the wave function (for the particle-in-a-box + pointer system) from Sect. 3.2.

3.4 In Chap. 2, we saw a simple example of measuring the momentum of a particle whose wave function was $\psi_0(x) = \sqrt{2}\sin(kx)$. Set up a formal (purely microscopic) quantum description of the measurement process: assume a "pointer" degree of freedom y, which starts in a Gaussian state centered at $y = 0$. What interaction Hamiltonian is appropriate for coupling the post-interaction pointer position to the particle's momentum? What is the final quantum state $\Psi(x, y, T)$ at the end of the interaction?

3.5 A particle is in the following superposition of position eigenstates:

$$\psi(x) = \frac{1}{\sqrt{2}}\left[\delta(x - a) + \delta(x + a)\right] \tag{3.38}$$

where the "δ"s are Dirac delta functions. The position of this particle is to be measured by a position measuring apparatus (also described quantum mechanically) whose pointer (with degree of freedom "y") should indicate the particle's position, $x = +a$ or $x = -a$, by moving to the right or to the left, respectively. What interaction Hamiltonian $\hat{H}_{int}$ will accomplish this? If the interaction turns on at $t = 0$ (and any contributions to the total Hamiltonian other than $\hat{H}_{int}$ are negligible) what is the wave function $\Psi(x, y, t)$ at time t? You should show explicitly that your answer is a solution of Schrödinger's equation. Finally, sketch/indicate the time-evolution of $\Psi(x, y, t)$ in configuration space.

3.6 Schrödinger says that "[i]nside the nucleus, blurring doesn't bother us" [3]. Why not? Why is "blurring" a problem for macroscopic things like pointers, but not a problem for microscopic things like nuclei?

3.7 Here is another simple "no hidden variables" argument that Schrödinger gives in Ref. [3]. Suppose that a quantum mechanical particle actually has a definite position $\vec{r}$ and a definite momentum $\vec{p}$ (even though these two quantities cannot be simultaneously known). Then it will have an angular momentum of magnitude $|\vec{L}| = |\vec{r} \times \vec{p}|$. Suppose that a measurement of the angular momentum magnitude simply reveals this pre-existing value. Now note that, by varying the origin with respect to which we measure the position, $\vec{r}$ – and therefore also $|\vec{L}|$ – can take on any value in a whole continuous spectrum. Explain how this is inconsistent with the (quantized!) measurement outcomes for angular momentum measurements, and therefore why the quantities $\vec{r}$, $\vec{p}$, and $\vec{L}$ cannot possess pre-existing definite values (satisfying the relation $\vec{L} = \vec{r} \times \vec{p}$) which are simply revealed by measurements.

3.8 Here is yet another simple "no hidden variables" argument from Ref. [3]. Consider a quantum mechanical simple harmonic oscillator, with energy operator (Hamiltonian) $\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2$. Suppose the oscillator is in its ground state with energy $E = \frac{1}{2}\hbar\omega$. The naive sort of "hidden variable" theory (associated with the ignorance interpretation of superposition) would say that this state describes an ensemble of individual systems, all with energy E, but different values of x and p satisfying $E = \frac{p^2}{2m} + \frac{1}{2}m\omega^2x^2$. Explain why this is not straightforwardly possible. (Hint: if the assumption is that position measurements simply reveal the actual pre-existing value of x, the Born rule implies that arbitrarily large values of $|x|$ are represented in the ensemble.)

3.9 One of the main ideas of this chapter is that there is no hope of introducing a sharply defined notion of dynamical collapse, such that the Schrödinger evolution and the other kind of evolution each apply in their own well-defined and non-overlapping spheres. But there is one idea for sharply defining such a boundary; it was proposed (or at least considered) by Eugene Wigner (and is perhaps somewhat widespread in more popular accounts of QM, for example the weird movie "What the bleep do we know"). The idea is that wave function collapse happens when physical matter interacts with *mind*. So, for example, in the Schrödinger's cat case, the wave function obeys the linear Schrödinger

equation when the radioactive atom is in the process of decaying, and when it is interacting with the Geiger counter, which in turn interacts with the hammer which interacts with the vial of poison which interacts with the cat ... all of this ends up in the superposition described in the text ... until the moment some human opens the box and becomes *consciously aware* of the result, at which point the involvement of her *mind* (presumably, to be specific, the interaction of her mind with her brain) collapses the wave function, for the whole physical system up through and including the brain, down to one or the other of the definite results. What do you think of this idea? Is it a good possible solution to the measurement problem, or utter nonsense, or what? (See Wigner's essay "Remarks On the Mind-Body Question" [14].)

3.10 In his "Reply to criticisms" Einstein gives a nice one-particle version of a Schrödinger's Cat type argument:

> If our concern is with macroscopic masses (billiard balls or stars), we are operating with very short de Broglie waves, which are determinative for the behavior of the center of gravity of such masses. This is the reason why it is possible to arrange the quantum-theoretical description for a reasonable time in such a manner that for the macroscopic way of viewing things, it becomes sufficiently precise in position as well as in momentum. It is true also that this sharpness remains for a long time and that the quasi-points thus represented behave just like the mass-points of classical mechanics. However, the theory shows also that, after a sufficiently long time, the point-like character of the ψ-function is completely lost to the center of gravity-co-ordiantes, so that one can no longer speak of any quasi-localisation of the centers of gravity. The picture then becomes, for example in the case of a single macro-mass-point, quite similar to that involved in a single free electron.
>
> If now, in accordance with the orthodox position, I view the ψ-function as the complete description of a real matter of fact for the individual case, I cannot but consider the essentially unlimited lack of sharpness of the position of the (macroscopic) body as *real*. On the other hand, however, we know that, by illuminating the body by means of a lantern ... we get a (macroscopically judged) sharp determination of position. In order to comprehend this I must assume that the sharply defined position is determined not merely by the real situation of the observed body, but also by the act of illumination. This is again a paradox.... The spook disappears only if one relinquishes the orthodox standpoint, according to which the ψ-function is accepted as a complete description of the single system [8].

Work out some quantitative estimates of the time durations involved in this kind of case. For example, consider the center-of-mass coordinate of a billiard ball. Suppose, at $t = 0$, it is described quantum mechanically by a Gaussian wave function of width one nanometer. How long would it take for the wave function to spread to a width of order, say, a meter? How long would it take the position of, say, a planet to become implausibly fuzzy?

3.11 Read Schrödinger's cat paper, Ref. [3], and report on anything you find interesting that wasn't already covered here.

3.12 In a Stern–Gerlach experiment, one can think of the *position* of the particle as the "pointer" that indicates the outcome of the spin measurement. Suppose, for example, a spin 1/2 particle begins in the product state

$$\Psi_0 = \psi_{+z}\,\phi(z) \tag{3.39}$$

where ψ_{+z} is the spin eigenstate ("spin-up along the z-axis") and $\phi(z)$ is a Gaussian wave packet. (The z axis here is the one along which the Stern–Gerlach apparatus has a non-uniform magnetic field, i.e., the direction along which the beam of incoming particles will be split.) What does the wave function evolve into during the course of the experiment? Sketch a diagram. What if the initial state is instead

$$\Psi_0 = \psi_{+x}\,\phi(z) \tag{3.40}$$

and it is still the z-component of the spin that is being measured? Sketch another diagram and discuss the relationship to the examples discussed in the Chapter. (Could the role of the two properties be reversed? That is, could the spin be considered as a pointer indicating the position along the z-axis? Discuss.)

3.13 In his essay "The Problem of Measurement" [15], Wigner discusses an example that is now part of many introductory textbook explanations of spin: a beam of particles (with, say, initial spin state ψ_{+x}) is sent through a Stern–Gerlach device to measure the z-spin. As discussed in the previous Project, this results in two sub-beams that are spatially separated (transverse to the direction of propagation of the particles). But now suppose some additional magnets are added, which have the effect of re-combining the two beams. The recombined beam is then sent through another Stern–Gerlach device, this time oriented in the x-direction. (Draw a picture to keep track of all this!) If you think the particle's passage through the z-oriented S-G device constitutes a *measurement* of the particle's z-spin, you would say that the particle's wave function collapses in this intermediate stage. Discuss what you would then expect to see in the subsequent x-spin measurement. In fact, *all* particles in this kind of situation are observed to emerge from the final x-spin measurement as spin-up along x. Discuss the implications of this and relate it to the other examples from the Chapter.

3.14 A beam of spin-1/2 particles is sent through a Stern–Gerlach device aligned along the x-axis. Those particles which emerge spin-up along the x-axis then enter another Stern–Gerlach device aligned along the z-axis. What happens, and how would an advocate of the ignorance interpretation explain the results? Now suppose we allow particles emerging from the second S-G device as spin-up along the z-axis to enter a third S-G device, oriented parallel to the x-axis. What happens? Can an advocate of the ignorance interpretation explain these results? How?

3.15 In Chap. 2 I described – as something that should "kind of blow your mind" – a two-particle entangled state in which neither particle has a definite energy, but the two-particle system does have a definite total energy. The discussion in Chap. 3 should help you understand better exactly how one needs to be understanding quantum descriptions in order for this kind of situation to be

interesting. Would this sort of entangled state be at all mind-blowing to someone who adopted the ignorance interpretation of superposition?

3.16 One difference between the Schrödinger evolution of the wave function, and the collapse of the wave function, is that the former is deterministic while the second is supposed to be irreducibly random. Sometimes it is claimed that people (like Schrödinger and Einstein) who had problems with quantum mechanics really just had problems with accepting irreducible randomness, i.e., the failure of determinism. (Think here, for example, of Einstein's famous and oft-quoted remark "God does not play dice".) To what extent do you think it is accurate to say that the (supposed) "measurement problem" is really just based on a philosophical insistence on pure determinism?

References

1. J.S. Bell, Against 'Measurement', reprinted in *Speakable and Unspeakable in Quantum Mechanics*, 2nd edn. (Cambridge University Press, Cambridge, 2004)
2. P.C.W. Davies, R. Brown (eds.), *The Ghost in the Atom*, interview with J.S. Bell (Cambridge University Press, Cambridge, 1986)
3. E. Schrödinger, The present situation in quantum mechanics. Naturwissenschaften **23** (1935), translated by J. Trimmer, in *Proceedings of the American Philosophical Society*, vol. 124, 10 October 1980 (1980), pp. 323–338
4. A. Fine, *The Shaky Game* (University of Chicago Press, Chicago, 1996)
5. I. Born, trans., *The Born–Einstein Letters* (Walker and Company, New York, 1971)
6. J.S. Bell, The trieste lecture of John Stewart Bell, transcribed by A. Bassi, G.C. Ghirardi. J. Phys. A: Math. Theor. **40**, 2919–2933 (2007)
7. M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974)
8. A. Einstein, Reply to criticisms, in *Albert Einstein: Philosopher-Scientist*, ed. by P.A. Schilpp (1949)
9. N.D. Mermin, Hidden variables and the two theorems of John Bell. Rev. Mod. Phys. **65**, 803–815
10. J.S. Bell, On the problem of hidden variables in quantum mechanics. Rev. Mod. Phys. **38**, 447–452 (1966); reprinted in *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 2004)
11. J.S. Bell, Quantum mechanics for cosmologists, in *Quantum Gravity 2*, ed. by C. Isham, R. Penrose, D. Sciama (Clarendon Press, Oxford, 1981), pp. 611–637; reprinted in *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 2004)
12. J.S. Bell, Beables for QFT, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 2004)
13. J.S. Bell, Against 'measurement', in *62 Years of Uncertainty: Erice, 5–14 August 1989* (Plenum Publishers, New York); reprinted in *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 2004)
14. E. Wigner, Remarks on the mind-body question, in *Symmetries and Reflections: Scientific Essays*, ed. by W. Moore, M. Scriven (Indiana University Press, Indiana, 1967)
15. E. Wigner, The problem of measurement. Am. J. Phys. **31**, 6 (1963)

Chapter 4
The Locality Problem

Chapter 3 focused on Schrödinger's argument that quantum mechanical wave functions (evolving always in accordance with Schrödinger's own equation) cannot be understood as providing *complete* descriptions of the physical states of individual systems. In this chapter we focus on a second argument, due largely to Einstein, for the same conclusion – namely, the *incompleteness* of the quantum mechanical description of physical reality.

4.1 Einstein's Boxes

In the discussion period of an important international scientific conference in 1927, Einstein made what would turn out to be just the first of several important and related arguments purporting to prove that there is a contradiction between the idea that quantum mechanics already provides (with wave functions alone) complete descriptions of physical states, and the idea of "locality" that we reviewed in Chap. 1.

In this early argument, Einstein begins by asking us to consider a single particle (an electron, say) incident on a narrow slit, behind which there is a curved detection screen as indicated in Fig. 4.1. Behind the slit, the electron will diffract as we saw in Chap. 2, resulting in essentially spherical Schrödinger waves propagating toward the screen. Of course, each individual electron that is fired in will eventually be detected at some distinct point on the screen. Einstein's comments, which I quote here at length, focus on the apparent conflict between the spreading spherical wave and the distinct point of eventual detection.

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_4

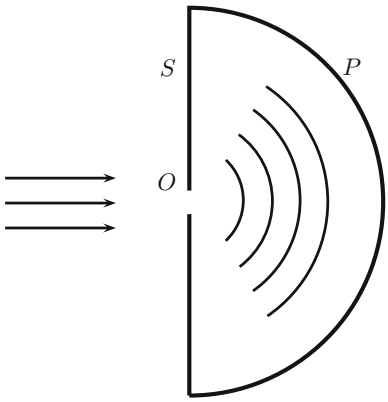


Fig. 4.1 A single electron approaches a narrow slit (O) in a screen (S). Downstream of the slit, the wave function diffracts and spreads more or less evenly over a curved detection screen (P)

One can take two positions towards the theory with respect to its postulated domain of validity, which I wish to characterise with the aid of a simple example.

Let S be a screen provided with a small opening O [see Fig. 4.1] and P a hemispherical photographic film of large radius. Electrons impinge on S in the direction of the arrow.... Some of these go through O, and because of the smallness of O and the speed of the particles, are dispersed uniformly over the directions of the hemisphere, and act on the film.

Both ways of conceiving the theory now have the following in common. There are de Broglie waves, which impinge approximately normally on S and are diffracted at O. Behind S there are spherical waves, which reach the screen P and whose intensity at P is responsible for what happens at P.

We can now charaterise the two points of view as follows.

1. Conception I. – The de Broglie - Schrödinger waves do not correspond to a single electron, but to a cloud of electrons extended in space. The theory gives no information about individual processes, but only about the ensemble of an infinity of elementary processes.

2. Conception II. – The theory claims to be a complete theory of individual processes. Each particle directed towards the screen, as far as can be determined by its position and speed, is described by a packet of de Broglie - Schrödinger waves of short wavelength and small angular width. This wave packet is diffracted and, after diffraction, partly reaches the film P in a state of resolution.

According to the first, purely statistical, point of view $|\psi|^2$ expresses the probability that there exists at the point considered *a particular* particle of the cloud, for example at a given point on the screen.

According to the second, $|\psi|^2$ expresses the probability that at a given instant *the same* particle is present at a given point (for example on the screen). Here, the theory refers to an individual process and claims to describe everything that is governed by laws.

The second conception goes further than the first, in the sense that all the information resulting from I results also from the theory by virtue of II, but the converse is not true. It is only by virtue of II that the theory contains the consequence that the conservation laws are valid for the elementary process; it is only from II that the theory can derive the result of the experiment of Geiger and Bothe, and can explain the fact that in the Wilson [cloud] chamber the droplets stemming from an α-particle are situated very nearly on continuous lines.

But on the other hand, I have objections to make to conception II. The scattered wave directed towards P does not show any privileged direction. If $|\psi|^2$ were simply regarded as

> the probability that at a certain point a given particle is found at a given time, it could happen that *the same* elementary process produces an action *in two or several* places on the screen. But the interpretation, according to which $|\psi|^2$ expresses the probability that *this* particle is found at a given point, assumes an entirely peculiar mechanism of action at a distance, which prevents the wave continuously distributed in space from producing an action in *two* places on the screen.
>
> In my opinion, one can remove this objection only in the following way, that one does not describe the process solely by the Schrödinger wave, but that at the same time one localises the particle during the propagation. I think Mr de Broglie is right to search in this direction. If one works solely with the Schrödinger waves, interpretation II of $|\psi|^2$ implies to my mind a contradiction with the postulate of relativity [1].

Einstein's "Conception 2" is of course just the idea that quantum wave functions provide complete descriptions of the physical states of individual particles. If, on this view, the wave function is spread out across some (hemispherical) region of space, then the particle itself is literally smeared out across that region. (To use Schrödinger's phrase from the last Chapter, it is like a cloud or fog bank.)

The idea of "Conception 1", on the other hand, is that the wave function does *not* provide a complete description of an individual electron, but is instead a kind of collective description of a large ensemble of individual electrons with different individual properties. This is closely related to the "ignorance interpretation of superposition" we discussed in the last chapter. The simplest (and probably wrong) possibility along these lines is the idea that electrons really are like classical particles which follow definite trajectories through space. If, for example, we sent a million particles through, one at a time, they would each follow (say) some different path between the slit and the screen, with $|\psi(x)|^2$ representing the fraction of the million trajectories that go near a given point x. The wave function thus provides us with information about the *probability* for a given electron to be found somewhere, but this smeared-out probabilistic information is certainly not able to tell us the exact trajectory of any given particle.

Einstein acknowledges that certain empirical observations seem to support Conception 2. But then he argues that Conception 2 conflicts with the principle of locality – the point we want to focus on in this Chapter. The argument seems to be roughly as follows. Suppose Conception 2 is correct, i.e., suppose that each individual particle really is smeared out across the whole hemispherical region, with each part of the cloud evidently possessing the power to perhaps trigger a "flash" at the corresponding point on the screen. But there is always only and exactly one flash for a given electron that is sent through. So it must be, on this view, that when a certain bit of the cloud manages to trigger a "flash" at a certain point on the screen, the *rest* of the cloud instantaneously loses its potency. This is essentially just the idea that is described formally in the collapse postulate: when a position measurement is made, the wave function of the electron collapses to a position eigenstate (i.e., it goes to zero at all points where the successful detection did not occur). Einstein is really just pushing us to consider the implications of this if we take the wave function not just as some kind of incomplete information catalog, but as a faithful and full description of a kind of spread-out physical field or cloud. An interaction between one part of

the cloud and a measuring device at the location of that one part can dramatically change the structure – the "intensity" – of the cloud at distant locations.

But, Einstein argues, this conflicts with the principle of locality: as soon as the "flash" is triggered somewhere, the field/cloud must suddenly change ("collapse") at other locations in order to ensure that no *additional* flashes are produced elsewhere. And note that the effect must be *instantaneous*. If, for example, the message to other pieces of the cloud – saying something like "Urgent! A flash has already been produced elsewhere! So don't make a flash!" – propagated out at the speed of light, there would be a chance that the message would arrive too late: a single electron might thus sometimes produce two (or more) flashes. Since this is never in fact observed to occur, it must therefore be that – again assuming Interpretation 2 is correct – the signal propagates out infinitely fast. But this, of course, is supposed to be impossible according to the idea of "locality" that is especially strongly implied by Einstein's own relativity theory.

And so, Einstein suggests, if the only two possibilities are Conception 1 and Conception 2, we must adopt Conception 1 since Conception 2 contradicts the relativistic notion of local causality, i.e., no-faster-than-light-action-at-a-distance.

Let us review a couple of other formulations of the same basic argument to make sure its structure is clear.

Einstein gives a similar (but slightly simpler) example in a letter he wrote to Schrödinger in 1935. He asks Schrödinger to consider a ball that is placed inside a box into which a partition is then inserted, so that the ball is either on the left or on the right. But suppose that we cannot see inside the box and things are arranged so that it is impossible to tell which side the ball is in fact on. Suppose further that the two halves of the box are then separated (again without looking inside or determining in any other way which half contains the ball) and carried to distant locations, where they are finally opened and their contents examined. As in his discussion at the 1927 Solvey Conference, Einstein suggests that there are two possible ways to understand what is going on:

> Now I describe a state of affairs as follows: *the probability is* $1/2$ *that the ball is in the first box*. Is this a complete description?
>
> NO: A complete description is: the ball *is* (or is not) in the first box. That is how the characterization of the state of affairs must appear in a complete description.
>
> YES: Before I open them, the ball is by no means in *one* of the two boxes. Being in a definite box only comes about when I lift the covers. This is what brings about the statistical character of the world of experience, or its empirical lawfulness. Before lifting the covers the state [of the distant box] is *completely* characterized by the number $1/2$, whose significance as statistical findings, to be sure, is only attested to when carrying out observations [2, p. 69].

Note that the NO and YES alternatives map exactly onto Conception 1 and Conception 2 from the 1927 discussion. According to the NO view (and Conception 1), the description of the state of the system in terms of probabilities is incomplete, there being, in reality, an actual fact of the matter about the location of the particle. According to the YES view (and Conception 2), the description in terms of probabilities is complete because the actual fact of the matter regarding the location of the ball (particle) only comes into existence with the act of measurement – the particle

is, prior to observation, a kind of cloud that is in fact spread 50/50 between the two half-boxes.

For the case of a literal (classical, macroscopic) ball, the YES view is not very plausible, and Einstein asserts that "the man on the street would only take the [NO] interpretation seriously." But for a single electron, the YES view is essentially the standard claim that quantum mechanics already provides a complete description of physical states. Einstein didn't accept this view and wanted to argue that the NO view was the correct one not only for the classical particle but for the electron as well. He constructs that argument, just as in 1927, by bringing in the idea of locality:

> *My* way of thinking is now this: properly considered, one cannot [refute the completeness doctrine, i.e., Conception 2, i.e., the YES view] if one does not make use of a supplementary principle: the 'separation principle.' That is to say: 'the second box, along with everything having to do with its contents, is independent of what happens with regard to the first box (separated partial systems).' If one adheres to the separation principle, then one thereby excludes the [YES] point of view, and only the [NO] point of view remains, according to which the above state description is an *incomplete* description of reality, or of the real states [3].

The principle of locality, that is, seems to imply that we can *find out* the contents of the distant box, without physically affecting it (or its contents) at all, merely by examining the contents of the nearby box. This seems to force on us the following dilemma: either (i) the distant box already did or didn't contain the particle (in which case the earlier statement that there was a 50% probability of its being found there is revealed as decidedly incomplete) – or (ii) the act of examining the contents of the *nearby* box instantaneously affects the physical contents of the *distant box* (changing it from a half-cloud with 50% potency to create a full-fledged particle, to either *nothing* or a full-fledged particle). But this latter option indeed seems to imply a violation of the relativistic notion of locality, i.e., seems to imply a kind of instantaneous action-at-a-distance.

Heisenberg, interestingly, presented a nice version of Einstein's argument in which he points out that it could be re-formulated in terms of a single photon that impinges on a half-silvered mirror (see Fig. 4.2):

> ...one other idealized experiment (due to Einstein) may be considered. We imagine a photon which is represented by a wave packet built up out of Maxwell waves. It will thus have a certain spatial extension and also a certain range of frequency. By reflection at a semi-transparent mirror, it is possible to decompose it into two parts, a reflected and a transmitted packet. There is then a definite probability for finding the photon either in one part or in the other part of the divided wave packet. After a sufficient time the two parts will be separated by any distance desired; now if an experiment yields the result that the photon is, say, in the reflected part of the packet, then the probability of finding the photon in the other part of the packet immediately becomes zero. The experiment at the position of the reflected packet thus exerts a kind of action (reduction of the wave packet) at the distant point occupied by the transmitted packet, and one sees that this action is propagated with a velocity greater than that of light [4, p. 39].

This is particularly interesting because Heisenberg – one of the creators and advocates of the "orthodox completeness doctrine" – seems here to concede that Einstein's argument really does establish the nonlocality of quantum theory, at least if

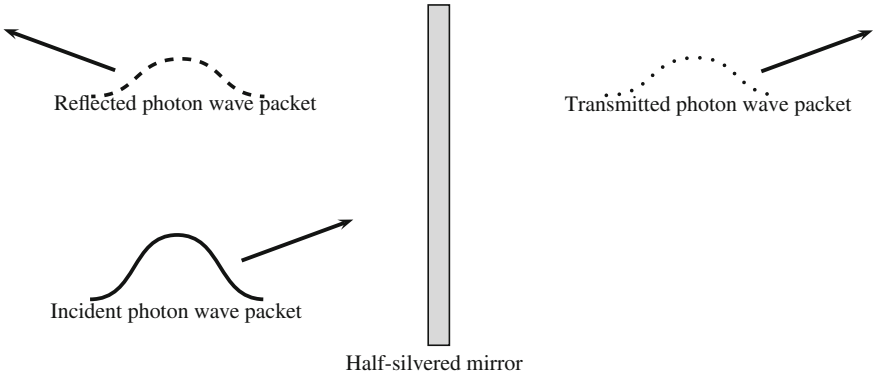


Fig. 4.2 Heisenberg's suggested setup for Einstein's Boxes argument: a single photon (*black*) is incident on a half-silvered mirror (i.e., a beam-splitter); a transmitted wave packet (*dotted*) contains 50% of the total probability for the particle to be detected, and a reflected packet (*dashed*) contains the other 50% of the total probability. As Heisenberg writes, the detection of the particle "at the position of the reflected packet thus exerts a kind of action (reduction of the wave packet) at the distant point occupied by the transmitted packet, and one sees that this action is propagated with a velocity greater than that of light [4]"

one assumes that it indeed provides *complete* descriptions of physical systems. But, as he goes on to state, Heisenberg doesn't believe that this implies any conflict with relativity theory: "However, it is also obvious that this kind of action can never be utilized for the transmission of signals so that it is not in conflict with the postulates of the theory of relativity." [4] It is surely correct that the nonlocality here cannot be used for superluminal communication. But the idea that this makes it compatible with relativity is quite dubious. Einstein, for example, obviously disagreed: he, apparently, thought that relativity prohibited instantaneous-action-at-a-distance as such, not merely that which can be somehow used by humans to build a telephone. (The question of compatibility with relativity will arise in several later Chapters as well. We set it aside here so as to focus on Einstein's argument that the completeness of the quantum mechanical description implies nonlocality... which Einstein, at least, regarded as implying "a contradiction with the postulate of relativity" [1].)

Here, finally, is one last statement of the "Einstein's Boxes" argument, this time as formulated in 1964 in a book by Louis de Broglie:

> Suppose a particle is enclosed in a box B with impermeable walls. The associated wave Ψ is confined to the box and cannot leave it. The usual interpretation asserts that the particle is 'potentially' present in the whole of the box B, with a probability $|\Psi|^2$ at each point. Let us suppose that by some process or other, for example, by inserting a partition into the box, the box B is divided into two separate parts B_1 and B_2 and that B_1 and B_2 are then transported to two very distant places, for example to Paris and Tokyo. The particle, which has not yet appeared, thus remains potentially present in the assembly of the two boxes and its wave function Ψ consists of two parts, one of which, Ψ_1, is located in B_1 and the other, Ψ_2, in B_2. The wave function is thus of the form $\Psi = c_1\Psi_1 + c_2\Psi_2$, where $|c_1|^2 + |c_2|^2 = 1$.

> The probability laws of wave mechanics now tell us that if an experiment is carried out in box B_1 in Paris, which will enable the presence of the particle to be revealed in this box, the probability of this experiment giving a positive result is $|c_1|^2$, whilst the probability of it giving a negative result is $|c_2|^2$. According to the usual interpretation, this would have the following significance: because the particle is present in the assembly of the two boxes prior to the observable localization, it would be immediately localized in box B_1 in the case of a positive result in Paris. This does not seem to me to be acceptable. The only reasonable interpretation appears to me to be that prior to the observable localization in B_1, we know that the particle was in one of the two boxes B_1 and B_2, but we do not know in which one, and the probabilities considered in the usual wave mechanics are the consequence of this partial ignorance. If we show that the particle is in box B_1, it implies simply that it was already there prior to localization. Thus, we now return to the clear classical concept of probability, which springs from our partial ignorance of the true situation. But, if this point of view is accepted, the description of the particle given by the customary wave function Ψ, though leading to a perfectly *exact* description of probabilities, does not give us a *complete* description of the physical reality, because the particle must have been localized prior to the observation which revealed it, and the wave function Ψ gives no information about this.
>
> We might note here how the usual interpretation leads to a paradox in the case of experiments with a negative result. Suppose that the particle is charged, and that in the box B_2 in Tokyo a device has been installed which enables the whole of the charged particle located in the box to be drained off and in so doing to establish an observable localization. Now, if nothing is observed, this negative result will signify that the particles is not in box B_2 and it is thus in box B_1 in Paris. But this can reasonably signify only one thing: the particle was already in Paris in box B_1 prior to the drainage experiment made in Tokyo in box B_2. Every other interpretation is absurd. How can we imagine that the simple fact of having observed *nothing* in Tokyo has been able to promote the localization of the particle at a distance many thousands of miles away? [5]

That is a very nice summary of the argument.

Now that we have (hopefully) made the "Einstein's Boxes" argument fairly clear in a qualitative way, let us try to make it a little more formally rigorous using Bell's formulation of "locality" from Chap. 1. This is of course somewhat anachronistic in the sense that Bell did not propose this definition of "locality" until 1976 (or, in its final version, 1990). So Einstein, for example, certainly never presented his argument in exactly this form. Still, it will help us to understand exactly the structure of the argument and the role of locality in particular.

The overall setup is illustrated in Fig. 4.3. The event "A" refers to the examination of the contents of the left half-box. If the particle is indeed found in the left half-box, we will denote this $A = +1$, whereas if that box is found to be empty we will denote this $A = 0$. Similarly, the events $B = +1$ and $B = 0$ will refer, respectively, to the finding and not-finding of a particle in the right half-box when it is opened and its contents are examined. In the Figure, the quantum mechanical wave function $\psi = \psi_L + \psi_R$ describing the state of the particle-in-the-boxes is depicted as the dashed and dotted packets, each containing 50% of the total probability associated with eventually finding the particle. (Note that here ψ_L and ψ_R are not separately normalized. For example, $\int |\psi_L|^2 dx = 1/2$.)

Now recall that Bell's definition of locality requires us to compare the probabilities assigned to an event like (for example) $A = +1$, when events from region B are, and aren't, conditioned upon. The definition also involves a complete specification

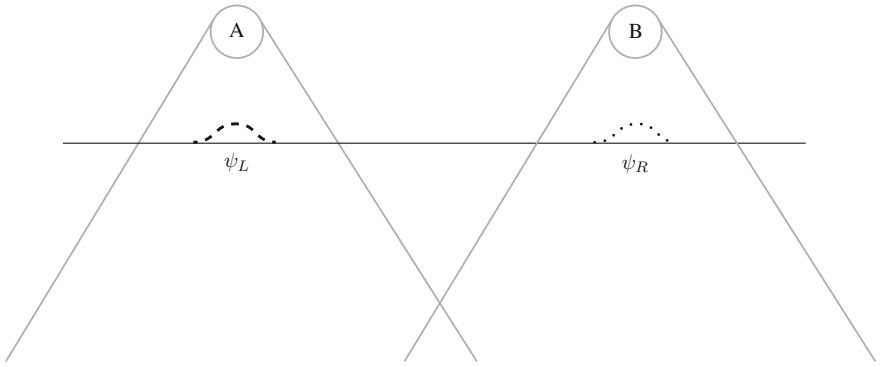


Fig. 4.3 Space-time diagram of the "Einstein's Boxes" setup. The events "A" and "B" represent observations in which the half-boxes are opened and their contents examined. The *dashed* (ψ_L) and *dotted* (ψ_R) curves represent the parts of the particle's quantum mechanical wave function ($\psi = \psi_L + \psi_R$) contained in the separated half-boxes. At the time corresponding to the horizontal *black line*, this wave function ψ is supposed, according to orthodox quantum mechanics, to provide a complete description of the state of the particle

of events $\mathcal{C}_\Sigma$ in a "slice" Σ across the backwards light cone of A. Here, Σ can just be the intersection of the horizontal black line in the figure with the backwards light cone of A. $\mathcal{C}_\Sigma$ then evidently includes (for the particle-in-the-box), the wave packet, ψ_L, contained in the left half-box, as well as all of the physical details about the left half-box itself, the transportation method, the observer and particle-detection apparatus, etc. It is clear, though, that according to quantum theory, we can say, for example, that even if all of these complicated details were specified, there would still just be an irreducible 50% probability assigned to the event $A = +1$. That is:

$$P\big[\, A = +1 \,|\, \mathcal{C}_\Sigma \,\big] \;=\; \frac{1}{2}. \tag{4.1}$$

However, consider now the event B – the examination of the contents of the right half-box. The particle is either found there ($B = +1$) or not ($B = 0$). But in either case, specifying the outcome of this other observation changes the probability assigned to the event $A = +1$. For example:

$$P\big[\, A = +1 \,|\, \mathcal{C}_\Sigma,\; B = +1 \,\big] \;=\; 0. \tag{4.2}$$

That is, the probability of finding the particle in the left half-box *given that it is found in the right half-box* is *zero*. And, similarly, the probability of finding the particle in the left half-box given that it is *not* found in the right half-box is *one*:

$$P\big[\, A = +1 \,|\, \mathcal{C}_\Sigma,\; B = 0 \big] \;=\; 1. \tag{4.3}$$

So we have a clear *violation* of Bell's locality condition,

$$P\big[\, A \,|\, \mathcal{C}_\Sigma \,\big] \;\neq\; P\big[\, A \,|\, \mathcal{C}_\Sigma, B \,\big], \tag{4.4}$$

as well as a violation of the modified condition,

$$P\big[\, A \,|\, \mathcal{C}_\Sigma,\; B \,\big] \;\neq\; P\big[\, A \,|\, \mathcal{C}_\Sigma,\; B' \,\big]. \tag{4.5}$$

Events from region B affect the probability assigned to A, even when the physical state of a (properly situated) slice across the past light cone of A has been specified (by assumption) completely.

What do we make of this? Basically it is just bringing out the fact that the collapse of the wave function violates locality, if we take the wave function seriously as representing a kind of physically-real field. The collapse provides a mechanism whereby one measurement (here, the examination of the contents of the right half-box) influences the state of a distant system (here, the contents of the left half-box) and thereby influences the probabilities assigned to events that are influenced by that system (here, the examination of the contents of the left half-box).

We must be absolutely clear, though, that this in no way establishes the real existence of nonlocal causal influences in nature! Instead, it merely establishes the existence of nonlocal causal influences in a certain *theory*, namely, the version of quantum theory according to which wave functions provide complete descriptions of the physical states of microscopic systems. A crucial assumption in our analysis, that is, was the assumption that the quantum mechanical wave function ψ_L provided a complete specification of the contents of the left half-box! We could – rather obviously – avoid the violation of locality by considering instead a different theory, according to which a complete specification of the contents of the half-boxes – at the time corresponding to the horizontal black line in the Figure – attributes the particle definitely to one, or the other, of the half-boxes. This is precisely what Einstein was suggesting when he said, at the end of his remarks in 1927, that one should "not describe the process solely by the Schrödinger wave, but [should in addition localise] the particle during the propagation." [1]

What the Einstein's boxes argument shows, then, is that we face a dilemma between "locality" and "completeness". If quantum mechanical wave functions provide complete descriptions of microscopic systems, then the theory must violate locality in order to make correct statistical predictions. In short, completeness implies that the collapse of the wave function is a physical, dynamical process, which conflicts with relativistic locality. On the other hand, we could preserve locality by denying the completeness doctrine and considering instead a hidden variable theory in which, even when the wave function involves a superposition between the particle being in the left and the right half-boxes, the particle is in fact (although unbeknownst to us) already in one place or the other. (The pilot-wave theory of de Broglie and Bohm, the subject of our Chap. 7, is a hidden variable theory of just this sort.)

4.2 EPR

In 1935, Einstein co-authored with Boris Podolsky and Nathan Rosen the most famous criticism of the idea that quantum mechanical wave functions provide complete descriptions of physical states. The argument of the EPR paper has the same basic structure as that of the earlier and simpler "boxes" type argument: if quantum mechanics is complete this would imply a violation of locality. Or equivalently: if we believe in the principle of relativistic local causality, we must reject the completeness doctrine.

It was discovered only rather recently that the entire text of the EPR paper – "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" [6] – was written by Podolsky (after conversations with Einstein and Rosen) and sent in for publication before Einstein had even seen the manuscript. Einstein wrote, in a private letter to Schrödinger, that the main point of the argument had not been made very clear: "the essential thing is, so to speak, smothered by the formalism [2]." So we should be a bit cautious about treating the EPR paper as providing an accurate presentation of Einstein's views. But the paper is so famous and so important that we will review it rather carefully. The subsequent section then discusses some of Einstein's own later expressions of the same basic argument.

Here is the abstract of the EPR paper, which lays out the argument to be presented:

> In a complete theory there is an element corresponding to each element of reality. A sufficient condition for the reality of a physical quantity is the possibility of predicting it with certainty, without disturbing the system. In quantum mechanics in the case of two physical quantities described by non-commuting operators, the knowledge of one precludes the knowledge of the other. Then either (1) the description of reality given by the wave function in quantum mechanics is not complete or (2) these two quantities cannot have simultaneous reality. Consideration of the problem of making predictions concerning a system on the basis of measurements made on another system that had previously interacted with it leads to the result that if (1) is false then (2) is also false. One is thus led to conclude that the description of reality as given by a wave function is not complete [6].

The structure here is a bit convoluted, so let us delve in and try to understand it better.

The explanation of what it means for a theory to be "complete" seems clear and uncontroversial. In the paper, EPR elaborate what they describe as a necessary condition for calling a theory "complete": "*every element of the physical reality must have a counterpart in the physical theory*". The overall goal of the paper will thus be to establish the existence of *more* elements of reality than have counterparts in quantum wave functions. In particular, the argument can be understood as an attempt to establish that a single particle can have *both* a definite momentum *and* a definite position, something that is forbidden in quantum mechanics since the position and momentum operators do not commute. (This implies that there is no wave function that is simultaneously an eigenstate of both position and momentum.)

In order to try to establish the existence of these properties, EPR require a "sufficient condition for the reality of a physical quantity". As they elaborate in the main text, this criterion is as follows:

> *If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity* [6].

As we will discuss in Chap. 6, this criterion became the focal point of Bohr's attempt to rebut the EPR argument. But it has always seemed perfectly valid to me. In any case, it is a little hard to understand the idea, so let's think it through with a simple concrete example.

Suppose someone hands you a shoebox with a wine glass in it and you want to determine whether the wine glass is shattered, or intact. One way of doing this might be to shake the box vigorously and listen for the tinkling sound of shattered pieces of glass hitting one another. But if the question is whether the glass in the box was shattered *originally*, when the box was first handed to you, this method does not really work: the act of shaking the box may very well result in an originally intact glass shattering! So hearing tinkling glass pieces inside would *not* effectively establish that the glass had already been broken prior to your shaking. The shaking itself may have brought that shattered state about.

By contrast, suppose (to complicate the scenario slightly) that the glass came to be in the box by the following procedure. There were two glasses on the shelf; one of them was perfectly intact, and one of them was already broken. Then your trusted friend flipped a coin and thereby randomly selected one of the glasses to seal up in the box; suppose we are certain that he was extremely careful so that, if the intact glass was the one selected for inclusion in the box, the glass was not broken during the act of putting it in the box. The second glass is then left on the shelf and the cupboard door is closed. Now in this situation, another method of determining the state of the glass in the box presents itself: simply open the cupboard and see which glass is there! If the intact glass is there in the cupboard, it must be the already-shattered glass that is in the box, and vice versa. This way of determining the contents of the box – in which we never interact directly with the box or its contents at all – ensures that the determined state of the glass in the box faithfully represents the true original state of the glass. We preclude the possibility that our act of determining the state has somehow affected and changed the state. This is the basic scheme that EPR will use to try to show that (in a certain special situation) a particle can be said to possess simultaneously definite values of both position and momentum.

EPR thus consider the following situation. Suppose two particles have interacted and gotten into an entangled state but then spatially separated so they are now far apart from one another. See Fig. 4.4. Assuming the particles are well-separated in regard to their y coordinates, we then focus our attention on the degrees of freedom x_1 and x_2. Suppose in particular that the particles are in the following entangled state:

$$\Psi(x_1, x_2) = \delta(x_1 - x_2) = \int \delta(x_1 - x)\delta(x_2 - x)\, dx \tag{4.6}$$

Pictured in the two-dimensional configuration space, this state is a "ridge" along the diagonal line $x_1 = x_2$. One can think of it as a superposition of states, over all possible values of x, in which both particles are definitely located at position x. That is, the

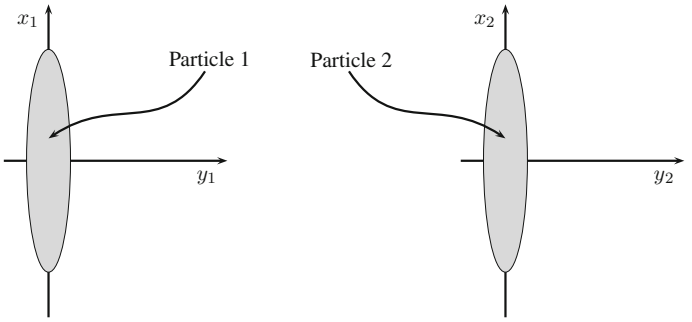


Fig. 4.4 Two particles which have previously interacted are spatially separated but remain in an entangled state. In particular, the spatial degrees of freedom x_1 and x_2 are entangled. We assume that, say, the part of the quantum state associated with the y coordinates is a simple product of two well-separated wave packets, centered (say) at $y_1 = 0$ and $y_2 = 0$ for the coordinate systems shown here. (Note that we use distinct coordinate systems for the two particles such that, for example, $y_1 = 0$ and $y_2 = 0$ are perhaps a million miles apart from one another!) So the particles are entangled (in so far as their positions along x are concerned) but they are unambiguously well-separated in space in regard to their y positions

state does not attribute a definite position to particle 1 or to particle 2 – both particles are maximally smeared out. But they are smeared out in a perfectly correlated way: measurement of the position x_1 of particle 1 immediately tells us the position x_2 of particle 2 because (even though neither x_1 nor x_2 has a well-defined value prior to such measurements) x_1 and x_2 are definitely equal to one another.

But this means it is possible to determine the position of particle 2 *indirectly* – without disturbing the physical state of particle 2 at all – by measuring the position x_1 of its distant entangled partner. And so, by the reality criterion, it follows that the distant particle must already *have* a definite position even when no such position is attributed to it by the pre-measurement wave function, Eq. (4.6). And note that this is already sufficient to show that that pre-measurement wave function did not provide a complete description of the state of the two particles: particle 2 *has* a definite position, but the wave function doesn't tell us about this at all.

EPR, however, go farther. The two-particle wave function can be re-written in this alternative (but mathematically equivalent) form

$$\Psi(x_1, x_2) = \delta(x_1 - x_2) = \frac{1}{2\pi} \int e^{ik(x_1 - x_2)}\, dk \;=\; \frac{1}{2\pi} \int e^{ikx_1} e^{-ikx_2}\, dk. \qquad (4.7)$$

The last expression can be understood as saying that the state is a superposition – over all possible values of k – of states in which particle 1 has momentum $p_1 = \hbar k$ and particle 2 has momentum $p_2 = -\hbar k$. So the state can also, alternatively, be understood as a state in which neither particle has any definite momentum value, but the momenta of the two particles are perfectly (anti-) correlated: $p_1 = -p_2$.

But this means it is possible to determine the momentum of particle 2 *indirectly* – without disturbing it at all – by measuring the momentum p_1 of its distant entangled partner. And so, by the reality criterion, it follows that the distant particle must already *have* a definite momentum even when no such momentum is attributed to it by the pre-measurement wave function. It is, in short, the same story again with momentum as it was before with position. So in addition to possessing a definite position (about which the pre-measurement wave function was silent), particle 2 apparently *also* possesses a definite *momentum* (about which the pre-measurement wave function was also silent). So that wave function provides, at best, a decidedly – a doubly – incomplete description of the state of the particle. There are at least these two physical properties, position and momentum, which in reality have sharp well-defined values, for which there is no corresponding element in the theoretical description. And in a way it's even worse than that, for by establishing the real existence of both position and momentum (for the one distant particle) EPR show not only that the particular wave function in Eq. (4.6) fails to provide a complete description, but that no wave function possibly could provide a complete description. For there is, simply, no such thing as a wave function that is simultaneously a position and momentum eigenstate.

That is the essential argument. But I have explained it here in my own words, and my version doesn't appear to correspond perfectly to the logical structure of the EPR paper's abstract. Let us try to understand that. First of all, what should we make of this disjunction (from the paper's abstract), "that either (1) the quantum-mechanical description of reality given by the wave function is not complete or (2) when the operators corresponding to two physical quantities do not commute the two quantities cannot have simultaneous reality"? This sounds very complicated but is actually quite trivial. Suppose two operators (for example position and momentum) fail to commute. Clearly, either the corresponding physical properties (1) *can* have simultaneous reality, or (2) *cannot* have simultaneous reality. If they *can*, then quantum mechanics is necessarily incomplete, because there is no wave function that is simultaneously an eigenstate for (i.e., there is no wave function that simultaneously attributes definite real values to) the two properties in question. So the trivial disjunction I wrote two sentences back is equivalent to the one from the EPR text.

Now, in the paper, EPR continue the argument as follows. Having established the disjunction between (1) and (2) just discussed, they write:

> Starting then with the assumption that the wave function does give a complete description of the physical reality, we arrived at the conclusion that two physical quantities, with noncommuting operators, can have simultaneous reality. Thus the negation of (1) leads to the negation of the only other alternative (2). We are thus forced to conclude that the quantum-mechanical description of physical reality given by wave functions is not complete.

Where and what, exactly, is the argument described in the first sentence here? It is again somewhat obscure and confusing, but actually this is just the essential argument we reviewed before. The way it is presented in the text is along the following lines. Thinking of the state as $\Psi = \int \delta(x_1 - x)\delta(x_2 - x)\, dx$ it is obvious that, if we measure the position of particle 1 and find it, say, at $x_1 = X$, the state of the two-particle system *collapses* to $\delta(x_1 - X)\delta(x_2 - X)$ which (being a product state) implies that we can

attribute the following wave function to particle 2:

$$\psi_2 = \delta(x_2 - X). \tag{4.8}$$

On the other hand, thinking of the (2-particle pre-measurement) state as $\Psi = \frac{1}{2\pi}\int e^{ikx_1}e^{-ikx_2}\,dk$ it is obvious that, if we measure the momentum of particle 1 and find, say, $p_1 = P$, the state of the two-particle system *collapses* to $\frac{1}{2\pi}\,e^{iPx_1/\hbar}\,e^{-iPx_2/\hbar}$ which (being a product state) implies that we can attribute the following wave function to particle 2:

$$\psi_2 = e^{i(-P/\hbar)x_2}. \tag{4.9}$$

EPR write: "Thus, *it is possible to assign two different wave functions ... to the same reality* (the second system after the interaction with the first)."

But then, *assuming that wave functions provide a complete description of the state of the particle*, i.e., assuming the negation of statement (1) from before, we have – from Eq. (4.8) – that particle 2 has a definite position, and – from Eq. (4.9) – that particle 2 has a definite momentum. Which indeed contradicts statement (2) from before.

Thus, the way it is presented in the actual EPR paper, the argument has the following extremely convoluted structure: either (1) or (2), but denying (1) requires one to also deny (2), and so one cannot consistently deny (1); that is, one must accept (1). That is, to be sure, logically valid. But it is also needlessly convoluted. The heart of the argument is simply the idea that, for spatially-separated but appropriately-entanged pairs of particles, we can determine, with certainty, the value of some property of one of the particles without actually *messing with it* at all, but by instead messing with its entangled partner and using the correlations built into the entangled state to infer something about the undisturbed particle. It is just like the example of the wine glass in the box. No wonder Einstein thought Podolsky's version of the argument was unnecessarily confusing!

4.3 Einstein's Discussions of EPR

I mentioned in the last section that although the paper grew out of discussions between Einstein, Podolsky, and Rosen, Podolsky actually wrote the EPR paper and submitted it for publication before Einstein had had a chance to see it or comment. And Einstein was somewhat frustrated and disappointed with how it came out. We began to see in the last section how Podolsky's version of the argument seemed needlessly convoluted, and undoubtedly that is part of what frustrated Einstein. But Einstein also specifically remarked that the main point had been "smothered". What was this main point that got buried in Podolsky's write-up?

Almost certainly it was the concept of "locality" which, as we have already seen in the discussion of the simpler "boxes" type arguments, was quite central to Einstein's

thinking about this kind of situation, but which hardly appears explicitly in the EPR paper itself. It is, though, implied in the application of the "reality criterion". Why do we think that, if we only make an actual measurement on particle 1, particle 2 is not disturbed at all? Well, evidently, because the two particles are spatially separated – they are distant from one another – and surely nothing we do *here* can have an immediate effect *over there*. That, presumably, is the idea – that is, it is only if we make the locality assumption that we are entitled to actually *apply* the reality criterion in the case at hand. But this is not made very clear. The closest we come, in the actual EPR paper, to an explicit mention of "locality" is in the penultimate paragraph of the paper:

> One could object to this conclusion [that quantum mechanics is incomplete] on the grounds that our criterion of reality is not sufficiently restrictive. Indeed, one would not arrive at our conclusion if one insisted that two or more physical quantities can be regarded as simultaneous elements of reality *only when they can be simultaneously measured or predicted*. On this point of view, since either one or the other, but not both simultaneously, of the quantities [momentum and position of particle 2] can be predicted, they are not simultaneously real. This makes the reality of [particle 2's momentum and position] depend upon the process of measurement carried out on the first system, which does not disturb the second system in any way. No reasonable definition of reality could be expected to permit this [6].

Here EPR seem to be anticipating the objection that, as we will see, Bohr makes against their argument: you cannot measure *both* the position *and* the momentum of the nearby particle, so you cannot determine both of these properties for the distant particle. You can only do one or the other. I think EPR are right to point out that this objection would make the real state of particle 2 "depend [nonlocally!] upon the process of measurement carried out on the first system." It doesn't matter whether you *do* in fact determine the position or momentum of particle 2 by measuring the corresponding property of particle 1; the mere fact that you *could* do so implies that those distant properties exist. Locality implies that the state of the distant particle is unaffected by what happens here – even the choice of whether or not to in fact go ahead with a certain kind of measurement.

So, I think, EPR here make a good and valid point, but it remains unfortunate that this is practically the only place they stress the notion of locality. As mentioned earlier, establishing the incompleteness of quantum mechanical descriptions doesn't even require establishing that the distant particle has *both* a definite position *and* a definite momentum. Its merely possessing, say, a definite position – when its state is described by the entangled wave function, Eq. (4.6) – is completely sufficient. And the point is, locality plays a crucial role already in the argument that a single such property, for the distant particle, can be established. Podolsky should have made all of this clearer.

Anyway, to round out our understanding of the EPR argument, we consider finally some of Einstein's own commentaries on the argument and related issues.

To begin with, in the same 1935 letter to Schrödinger from which I earlier quoted his remarks about the "boxes" example, Einstein writes:

> The preceding [boxes] analogy corresponds only very imperfectly to the quantum mechanical example in the [EPR] paper. It is, however, designed to make clear the point of view that

> is essential to me. In quantum mechanics one describes a real state of affairs of a system by means of a normed function ψ of the coordinates (of configuration space). The temporal evolution is uniquely determined by the Schrödinger equation. One would now very much like to say the following: ψ stands in a one-to-one correspondence with the real state of the real system. The statistical character of measurement outcomes is exclusively due to the measuring apparatus, or the process of measurement. If this works, I talk about a complete description of reality by the theory. However, if such an interpretation doesn't work out, then I call the theoretical description 'incomplete'.... [2, p. 71]

Einstein continues:

> Now what is essential is exclusively that [the wave functions, relating to the distant particle, that arise from different kinds of measurements on the nearby particle] are in general different from one another. I assert that this difference is incompatible with the hypothesis that the ψ description is correlated one-to-one with the physical reality (the real state). After the [particles separate], the real state of [the two particle system] consists precisely of the real state of [particle 1] and the real state of [particle 2], which two states have nothing to do with one another. *The real state of [particle 2] thus cannot depend upon the kind of measurement I carry out on [particle 1].* ('Separation hypothesis' from above.) But then for the same state of [particle 2] there are two (in general arbitrarily many) equally justified $\psi_{[2]}$, which contradicts the hypothesis of a one-to-one or complete description of the real states [3].

This passage has the virtue of making clearer the exact sense in which Einstein understood the "incompleteness" of the quantum mechanical description: a failure of the one-to-one correspondence between real states and theoretical descriptions.

In his auto-biographical contribution to the collection *Albert Einstein: Philosopher Scientist* from 1949, Einstein also gave an extensive discussion of the EPR-type argument for quantum incompleteness. We quote it at length here:

> Physics is an attempt conceptually to grasp reality as it is thought independently of its being observed. In this sense one speaks of 'physical reality'. In pre-quantum physics there was no doubt as to how this was to be understood. In Newton's theory reality was determined by a material point in space and time; in Maxwell's theory, by the field in space and time. In quantum mechanics it is not so easily seen. If one asks: does a ψ-function of the quantum theory represent a real factual situation in the same sense in which this is the case of a material system of points or of an electromagnetic field, one hesitates to reply with a simple 'yes' or 'no'; why? What the ψ-function (at a definite time) asserts, is this: What is the probability for finding a definite physical magnitude q (or p) in a definitely given interval, if I measure it at time t? The probability is here to be viewed as an empirically determinable, and therefore certainly as a 'real' quantity which I may determine if I create the same ψ-function very often and perform a q-measurement each time. But what about the single measured value of q? Did the respective individual system have this q-value even before the measurement? To this question there is no definite answer within the framework of the [existing] theory, since the measurement is a process which implies a finite disturbance of the system from the outside; it would therefore be thinkable that the system obtains a definite numerical value for q (or p), the measured numerical value, only through the measurement itself. For the further discussion I shall assume two physicists, A and B, who represent a different conception with reference to the real situation as described by the ψ-function.
>
> A. The individual system (before the measurement) has a definite value of q (i.e., p) for all variables of the system, and more specifically, *that* value which is determined by a measurement of this variable. Proceeding from this conception, he will state: The ψ-function is no exhaustive description of the real situation of the system but an incomplete description; it expresses only what we know on the basis of former measurements concerning the system.

> B. The individual system (before the measurement) has no definite value of q (i.e., p). The value of the measurement only arises in cooperation with the unique probability which is given to it in view of the ψ-function only through the act of measurement itself. Proceeding from this conception, he will (or, at least, he may) state: the ψ-function is an exhaustive description of the real situation of the system.
>
> We now present to these two physicists the following instance: There is to be a system which at the time t of our observation consists of two partial systems S_1 and S_2, which at this time are spatially separated and (in the sense of the classical physics) are without significant reciprocity. The total system is to be completely described through a known ψ-function ψ_{12} in the sense of quantum mechanics. All quantum theoreticians now agree upon the following: If I make a complete measurement of S_1, I get from the results of the measurement and from ψ_{12} an entirely definite ψ function ψ_2 of the system S_2. The character of ψ_2 then depends upon *what kind* of measurement I undertake on S_1.
>
> Now it appears to me that one may speak of the real factual situation of the partial system S_2. Of this real factual situation, we know to begin with, before the measurement of S_1, even less than we know of a system described by the ψ-function. But on one supposition we should, in my opinion, absolutely hold fast: the real factual situation of the system S_2 is independent of what is done with the system S_1, which is spatially separated from the former. According to the type of measurement which I make of S_1, I get, however, a very different ψ_2 for the second partial system.... Now, however, the real situation of S_2 must be independent of what happens to S_1. For the same real situation of S_2 it is possible therefore to find, according to one's choice, different types of ψ-function. (One can escape from this conclusion only by either assuming that the measurement of S_1 (telepathically) changes the real situation of S_2 or by denying independent real situations as such to things which are spatially separated from each other. Both alternatives appear to me entirely unacceptable.)
>
> If now the physicists, A and B, accept this consideration as valid, then B will have to give up his position that the ψ-function constitutes a complete description of a real factual situation. For in this case it would be impossible that two different types of ψ-functions could be co-ordinated with the identical factual situation of S_2.
>
> The statistical character of the present theory would then have to be a necessary consequence of the incompleteness of the description of the systems in quantum mechanics, and there would no longer exist any ground for the supposition that a future basis of physics must be based upon statistics [7].

This passage, I think, makes very clear: (i) Einstein's belief that the *randomness* of quantum mechanics is in fact not inherent to nature, but is instead a result of the incompleteness of the theory's descriptions of nature; (ii) his hope that a future theory might *complete* the quantum descriptions and thereby restore the principle of *determinism* to physical theory; and (iii) his *reasons*, based in particular on the locality principle to which he stresses we must "absolutely hold fast", for this belief and this hope.

I want to stress, in particular, that Einstein's belief that a proper theory would restore determinism – captured in his oft-quoted remark "God does not play dice" – was in no way based on a philosophical unwillingness to contemplate fundamental, irreducible randomness. Instead it was based on his argument that consistency with relativistic locality required one to reject the claim that quantum mechanical wave functions provide complete descriptions of the physical states of the systems they describe.

Einstein gave another, much briefer, summary of his position in his other contribution (called "Reply to Criticisms") to the same 1949 book:

> By this way of looking at the matter it becomes evident that the paradox forces us to relinquish one of the following two assertions:
>
> 1. the description by means of the ψ-function is complete.
>
> 2. the real states of spatially separated objects are independent of each other.
>
> On the other hand, it is possible to adhere to (2) if one regards the ψ-function as the description of a (statistical) ensemble of systems (and therefore relinquishes (1)). However, this view blasts the framework of the 'orthodox quantum theory' [8].

In that same essay, Einstein states: "I am, in fact, firmly convinced that the essentially statistical character of contemporary quantum theory is solely to be ascribed to the fact that this [theory] operates with an incomplete description of physical systems." [8]

In this section, I have tried to stress that, for Einstein, establishing the real existence of (for example) *both* the position *and* the momentum of a distant particle (by making one or the other measurement on the nearby particle) was not really necessary to establish the incompleteness of the quantum mechanical descriptions of states. (He wrote to Schrödinger that whether the EPR argument establishes the reality of, for example, both position and momentum "ist mir *wurst*" – roughly, "I couldn't care less!" [2]) Instead, his preferred argument merely pointed out that one's choice of measurements to make on the nearby particle produced (via wave function collapse) different wave functions for the distant particle; and so, if the actual state of that distant particle remains unaffected by such measurements on the nearby particle, we have a failure of completeness in the sense of one-to-one correspondence between wave functions and physical states.

But I also don't want to make it appear that Einstein's views were totally different from those presented in our summary of the EPR paper itself. Indeed, it seems that – although he did not think this was necessary to establish the conflict between locality and completeness – Einstein did accept that the EPR-type argument does establish that, if locality is true, the distant system must already possess (for example) both position and momentum. For example, in a 1938 letter to Tanya Ehrenfest, Einstein wrote "Here, however, I [cannot] reconcile myself to the following, that a manipulation undertaken on A has an influence on B; thus I see myself required to suppose, as actually or physically realized at B, everything relating to measurement outcomes on B that can be predicted with certainty, on the basis of some measurement or other undertaken on A." [2, p. 63]

4.4 Bohm's Reformulation

In 1951, the young physicist David Bohm published a textbook on quantum theory. The book is somewhat unusual in emphasizing conceptual questions and containing lengthy prose discussions (in addition to the more standard mathematical presenta-

tion). The book is also of interest because it treats the subject in a very orthodox, Copenhagen way; but shortly after its publication, Bohm met with Einstein, who evidently convinced Bohm of the inadequacy of this approach (and in particular convinced him of the "incompleteness" of the orthodox quantum state descriptions). And in the following year, 1952, Bohm would produce (or really, because similar ideas had been proposed, but then prematurely abandoned, 25 years earlier by de Broglie, reproduce) a fully-worked-out "hidden variable theory" that we will discuss in depth in Chap. 7.

For our present purposes, though, we want to focus on Bohm's 1951 presentation of "The Paradox of Einstein, Podolsky, and Rosen" which marks an important (if also seemingly minor and merely technical) advance that would play an important role in Bell's Theorem, the subject of our Chap. 8.

Here is Bohm's lead-in to the discussion:

> What the authors [of EPR] wished to do with their criteria for reality was to show that the above interpretation of the present quantum theory is untenable and that the wave function cannot possibly contain a complete description of all physically significant factors (or 'elements of reality') existing within a system. If their contention could be proved, then one would be led to search for a more complete theory, perhaps containing something like hidden variables, in terms of which the present quantum theory would be a limiting case [9, p. 612].

And then here is Bohm's presentation of the EPR argument, re-framed in terms of the *spins* of a pair of spin 1/2 particles:

> We have modified the experiment somewhat [compared to the way it was presented in the actual EPR paper], but the form is conceptually equivalent to that suggested by them, and considerably easier to treat mathematically.
>
> Suppose that we have a molecule containing two atoms in a state in which the total spin is zero and that the spin of each atom is $\hbar/2$. Roughly speaking, this means that the spin of each particle points in a direction exactly opposite to that of the other, insofar as the spin may be said to have any definite direction at all. Now suppose that the molecule is disintegrated by some process that does not change the total angular momentum. The two atoms will begin to separate and will soon cease to interact appreciably. Their combined spin angular momentum, however, remains equal to zero, because by hypothesis, no torques have acted on the system.
>
> Now, if the spin were a classical angular momentum variable, the interpretation of this process would be as follows: While the two atoms were together in the form of a molecule, each component of the angular momentum of each atom would have a definite value that was always opposite to that of the other, thus making the total angular momentum equal to zero. When the atoms separated, each atom would continue to have every component of its spin angular momentum opposite to that of the other. The two spin-angular-momentum vectors would therefore be correlated. These correlations were originally produced when the atoms interacted in such a way as to form a molecule of zero total spin, but after the atoms separate, the correlations are maintained by the deterministic equations of motion of each spin vector separately, which bring about conservation of each component of the separate spin-angular-momentum vectors.
>
> Suppose now that one measures the spin angular momentum of any one of the particles, say No. 1. Because of the existence of correlations, one can immediately conclude that the angular-momentum vector of the other particle (No. 2) is equal and opposite to that of No. 1. In this way, one can measure the angular momentum of particle No. 2 indirectly by measuring the corresponding vector of particle No. 1.

> Let us now consider how this experiment is to be described in the quantum theory. Here, the investigator can measure either the x, y, or z component of the spin of particle No. 1, but not more than one of these components, in any one experiment. Nevertheless, it still turns out as we shall see that whichever component is measured, the results are correlated, so that if the same component of the spin of atom No. 2 is measured, it will always turn out to have the opposite value. This means that a measurement of any component of the spin of atom No. 1 provides, as in classical theory, an indirect measurement of the same component of the spin of atom No. 2. Since, by hypothesis, the two particles no longer interact, we have obtained a way of measuring an arbitrary component of the spin of particle No. 2 without in any way disturbing that particle. If we accept the definition of an element of reality ... suggested by ERP, it is clear that after we have measured σ_z for particle 1, then σ_z for particle 2 must be regarded as an element of reality, existing separately in particle No. 2 alone. If this is true, however, this element of reality must have existed in particle No. 2 even before the measurement of σ_z for particle No. 1 took place. For since there is no interaction with particle No. 2, the process of measurement cannot have affected this particle in any way. But now let us remember that, in each case, the observer is always free to reorient the apparatus in an arbitrary direction while the atoms are still in flight, and thus to obtain a definite (but unpredictable) value of the spin component in any direction that he chooses. Since this can be accomplished without in any way disturbing the second atom, we conclude that ... precisely defined elements of reality must exist in the second atom, corresponding to the simultaneous definition of all three components of its spin. Because the wave function can specify, at most, only one of these components at a time with complete precision, we are then led to the conclusion that the wave function does not provide a complete description of all elements of reality existing in the second atom.
>
> If this conclusion were valid, then we should have to look for a new theory in terms of which a more nearly complete description was possible [9].

As mentioned, Bohm himself became convinced of the need for such a "new theory" – and indeed produced one! – in the following year. But at the time of this writing he continued to accept the Copenhagen philosophy, according to which the EPR argument is not valid. We will review the Copenhagen philosophy (and in particular Bohr's reply to the EPR argument) in Chap. 6.

For now, let's just focus on the technical aspects of Bohm's reformulation of the EPR scenario in terms of the spins of two spin 1/2 particles. Using the notation of Chap. 2, in which for example "ψ^1_{+z}" denotes a state in which particle 1 is "spin up" along the z-direction, the state of total spin zero described by Bohm is the following:

$$\Psi = \frac{1}{\sqrt{2}}\left[\psi^1_{+z}\psi^2_{-z} - \psi^1_{-z}\psi^2_{+z}\right]. \tag{4.10}$$

This is a superposition of (one the one hand) a state in which particle 1 is "spin up" along z and particle 2 is "spin down" along z and (on the other hand) a state in which particle 1 is "spin down" and particle 2 is "spin up". The minus sign (i.e., the relative phase between the two terms in the superposition) turns out to be important. The qualitatively similar superposition with a "+" sign is also a state in which the z-component of the total spin is zero, but (unlike the state with the minus sign) the *magnitude* (squared) of the total spin is not zero. Indeed, this other state – with the "+" sign – naturally goes together with the "both particles are spin up" and "both particles are spin down" states to form a so-called *triplet* of states (with one unit of

total spin angular momentum, but with the z-component being -1, 0, and $+1$ in the three states).

By contrast, the state in Eq. (4.10) is sometimes called the *singlet* state because it alone has zero units of total spin angular momentum – and therefore zero for its z-component and indeed also all other components. I will leave the exploration of some of the mathematical aspects of this state for you to work through in the Projects.

But it should be clear, at least, that this EPR-Bohm state allows a simplified version of an EPR-type argument in the following way. The state in Eq. (4.10) is not an eigenstate of σ_z for particle 2. It is instead an entangled superposition of two states in which σ_z for particle 2 has two distinct values ($+1$ and -1). So according to the usual completeness doctrine, particle 2 has no definite value of σ_z when the state of the two particles is given by Eq. (4.10). However, by measuring σ_z of particle 1, we can determine – seemingly without disturbing particle 2 in any way – σ_z for particle 2: if particle 1 turns out to be spin-up, then particle 2 is spin-down, and vice versa. *After* such a measurement on particle 1, it thus seems clear that particle 2 *has* a definite z-spin. Then, either it had that definite z-spin value all along – in which case the quantum mechanical state description of Eq. (4.10) is revealed as having been incomplete – or its z-spin value only crystallized, from some earlier "blurry" state, as a result of the measurement on particle 1. But this latter possibility involves a kind of "spooky action-at-a-distance", i.e., a violation of local causality. We thus have to either accept the non-locality (and face the seemingly daunting task of trying to reconcile it with relativity) or abandon the completeness doctrine.

4.5 Bell's Re-Telling

John Bell – whose seminal 1964 theorem will be the subject of Chap. 8 – wrote extensively on the foundations of quantum theory and the EPR argument in particular. One of his presentations in particular is so amusing and beautiful and clear that I cannot help but include it here. It is from a paper with the intriguing title "Bertlmann's socks and the nature of reality". The paper begins:

> The philosopher on the street, who has not suffered a course in quantum mechanics, is quite unimpressed by Einstein–Podolsky–Rosen correlations. He can point to many examples of similar correlations in everyday life. The case of Bertlmann's socks is often cited. Dr. Bertlmann likes to wear two socks of different colours. Which colour he will have on a given foot on a given day is quite unpredictable. But when you see that the first sock is pink you can be already sure that the second sock will not be pink. Observation of the first, and experience of Bertlmann, gives immediate information about the second. There is no accounting for tastes, but apart from that there is no mystery here. And is not the EPR business just the same? [10]

Bell then reviews Bohm's reformulation of the EPR setup, with the two entangled spin-1/2 particles, including also a nice discussion of the difficulty of understanding the results of individual Stern–Gerlach spin measurements in terms of classical

magnetic dipoles one of whose pre-existing components is simply revealed by the measurement. Bell continues:

> Phenomena of this kind made physicists despair of finding any consistent space-time picture of what goes on on the atomic and subatomic scale. Making a virtue of necessity, and influenced by positivistic and instrumentalist philosophies, many came to hold not only that it is difficult to find a coherent picture but that it is wrong to look for one – if not actually immoral then certainly unprofessional. Going further still, some asserted that atomic and subatomic particles do not *have* any definite properties in advance of observation. There is nothing, that is to say, in the particles approaching the [Stern–Gerlach] magnet, to distinguish those subsequently deflected up from those subsequently deflected down. Indeed even the particles are not really there. [Note: to help prevent the reader from getting lost in quotes within quotes, passages that Bell quotes from other authors are *italicized* in the remainder of this block quote as well as the following one. In particular, the following italicized passages are quotations from Peterson, Heisenberg, Zilsel, Pauli, and Born.]
>
> For example, [Bohr's colleague Peterson recalled that] *Bohr once declared when asked whether the quantum mechanical algorithm could be considered as somehow mirroring an underlying quantum reality: 'There is no quantum world. There is only an abstract quantum mechanical 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'.*
>
> And for Heisenberg ...*in the experiments about atomic events we have to do with things and facts, with phenomena that are just as real as any phenomena of daily life. But the atoms or the elementary particles are not as real; they form a world of potentialities or possibilities rather than one of things or facts.*
>
> And [Zilsel recollects] *Jordan declared, with emphasis, that observations not only* **disturb** *what has to be measured, they* **produce** *it. In a measurement of position, for example, as performed with the gamma ray microscope, 'the electron is forced to a decision. We compel it* **to assume a definite position***; previously it was, in general, neither here nor there; it had not yet made its decision for a definite position... If by another experiment the* **velocity** *of the electron is being measured, this means: the electron is compelled to decide itself for some exactly defined value of the velocity... we ourselves produce the results of measurement'.*
>
> It is in the context of ideas like these that one must envisage the discussion of the Einstein–Podolsky–Rosen correlations. Then it is a little less unintelligible that the EPR paper caused such a fuss, and that the dust has not settled even now. It is as if we had come to deny the reality of Bertlmann's socks, or at least of their colours, when not looked at. And as if a child had asked: How come they always choose different colours when they *are* looked at? How does the second sock know what the first has done?
>
> Paradox indeed! But for the others, not for EPR. EPR did not use the word 'paradox'. They were with the man in the street in this business. For them these correlations simply showed that the quantum theorists had been hasty in dismissing the reality of the microscopic world. In particular Jordan had been wrong in supposing that nothing was real or fixed in that world before observation. For after observing only one particle the result of subsequently observing the other (possibly at a very remote place) is immediately predictable. Could it be that the first observation somehow fixes what was unfixed, or makes real what was unreal, not only for the near particle but also for the remote one? For EPR that would be an unthinkable 'spooky action at a distance'. To avoid such action at a distance they have to attribute, to the space-time regions in question, *real* properties in advance of observation, correlated properties, which *predetermine* the outcomes of these particular observations. Since these real properties, fixed in advance of observation, are not contained in quantum formalism, that formalism for EPR is *incomplete*. It may be correct, as far as it goes, but the usual quantum formalism cannot be the whole story [10].

That, I submit, is as clear as anyone will ever make the EPR argument, and it would seem appropriate to end the Chapter on that note.

But in the paper Bell goes on to discuss another aspect of Einstein's worries about the quantum theory, and I think it will be very illuminating to include this here as well:

> It is important to note that to the limited degree to which *determinism* plays a role in the EPR argument, it is not assumed but *inferred*. What is held sacred is the principle of 'local causality' – or 'no action at a distance'. Of course, mere *correlation* between distant events does not by itself imply action at a distance, but only correlation between the signals reaching the two places. These signals, in the idealized example of Bohm, must be sufficient to *determine* whether the particles go up or down. For any residual undeterminism could only spoil the perfect correlation.
>
> It is remarkably difficult to get this point across, that determinism is not a *presupposition* of the analysis. There is a widespread and erroneous conviction that for Einstein determinism was always *the* sacred principle. The quotability of his famous 'God does not play dice' has not helped in this respect. Among those who had great difficult in seeing Einstein's position was Born. Pauli tried to help him in a letter of 1954:
>
> *...I was unable to recognize Einstein whenever you talked about him in either your letter or your manuscript. It seemed to me as if you had erected some dummy Einstein for yourself, which you then knocked down with great pomp. In particular, Einstein does not consider the concept of 'determinism' to be as fundamental as it is frequently held to be (as he told me emphatically many times)... he* **disputes** *that he uses as a criterion for the admissibility of a theory the question: 'Is it rigorously deterministic?' ... he was not at all annoyed with you, but only said you were a person who will not listen.*
>
> Born had particular difficulty with the Einstein–Podolsky–Rosen argument. Here is his summing up, long afterwards, when he edited the Born-Einstein correspondence:
>
> *The root of the difference between Einstein and me was the axiom that events which happen in different places A and B are independent of one another, in the sense that an observation on the state of affairs at B cannot teach us anything about the state of affairs at A.*
>
> Misunderstanding could hardly be more complete. Einstein had no difficulty accepting that affairs in different places could be correlated. What he could not accept was that an intervention at one place could *influence*, immediately, affairs at the other.
>
> These references to Born are not meant to diminish one of the towering figures of modern physics. They are meant to illustrate the difficulty of putting aside preconceptions and listening to what is actually being said. They are meant to encourage *you*, dear listener, to listen a little harder [10].

Bell then closes this section of his paper by quoting the following "summing-up by Einstein himself", which is from Einstein's 1948 *Dialectica* essay:

> If one asks what, irrespective of quantum mechanics, is characteristic of the world of ideas in physics, one is first of all struck by the following: the concepts of physics relate to a real outside world.... It is further characteristic of these physical objects that they are thought of as arranged in a space-time continuum. An essential aspect of this arrangement of things in physics is that they lay claim, at a certain time, to an existence independent of one another, provided these objects 'are situated in different parts of space.'
>
> The following idea characterizes the relative independence of objects far apart in space (A and B): external influence on A has no direct influence on B...
>
> There seems to me no doubt that those physicists who regard the descriptive methods of quantum mechanics as definitive in principle would react to this line of thought in the

> following way: they would drop the requirement ... for the independent existence of the physical reality present in different parts of space; they would be justified in pointing out that the quantum theory nowhere makes explicit use of this requirement.
>
> I admit this, but would point out: when I consider the physical phenomena known to me, and especially those which are being so successfully encompassed by quantum mechanics, I still cannot find any fact anywhere which would make it appear likely that (that) requirement will have to be abandoned.
>
> I am therefore inclined to believe that the description of quantum mechanics ... has to be regarded as an incomplete and indirect description of reality, to be replaced at some later date by a more complete and direct one [11].

And that seems like an entirely fitting way to close this chapter.

Projects:

4.1 What is the Bothe-Geiger experiment that Einstein mentions in his 1927 Solvay remarks? Do a little research and report back. (Hint: it relates to Compton scattering and something called the Bohr–Kramers–Slater or "BKS" theory, which was a kind of pre-cursor to the formal quantum theory that eventually developed.)

4.2 In the text, our application of Bell's formulation of "locality" to the Einstein's Boxes argument consisted of showing that quantum theory (with the completeness assumption and with the collapse postulate) violates locality. One could also, however, put the same pieces together in a slightly different way – showing that quantum theory (with the completeness assumption but *not* the collapse postulate) implies, if you assume locality, that there should be a nonzero probability for detecting the same one particle *twice*, once in the left half-box and once again in the right half-box. Explain carefully how this argument would go.

4.3 Show that the commutator $\left[\hat{x},\, \hat{p}\right] = \hat{x}\hat{p} - \hat{p}\hat{x}$ of $\hat{x}$ and $\hat{p}$ is the constant $i\hbar$. Hint: let $\left[\hat{x},\, \hat{p}\right]$ act on an arbitrary function $f(x)$, using $\hat{x} = x$ and $\hat{p} = -i\hbar\frac{d}{dx}$, and show that you get $i\hbar f(x)$.

4.4 Prove that, if the commutator of two operators $\hat{A}$ and $\hat{B}$ is the (nonzero) constant c, then there cannot exist a state ψ which is a simultaneous eigenstate of both $\hat{A}$ and $\hat{B}$. Use this, along with the results of Project [4.3], to argue that in quantum mechanics there cannot be a state which attributes a sharp value to position and momentum simultaneously – a point that was crucial in the EPR argument.

4.5 Give a careful summary of the argument for incompleteness that Einstein gives in his "Autobiographical Notes" (quoted in Sect. 4.3).

4.6 One assumption of all these EPR-type arguments, that is sometimes taken for granted and not given the attention it maybe deserves, is the assumption that the statistical predictions of quantum mechanics in the relevant situations are actually *correct*. For example, in the "boxes" type argument, it is assumed that, indeed, each particle will only be found at one place later. An early experiment by Ádám, Jánossy, and Varga (ÁJV) attempted to test this prediction, but their results were not very conclusive. Here is a bit of description, though, from John Clauser, who re-did a more convincing version of the experiment in the 1970s:

> As an original heretic to the standard religion, [Schrödinger] persuaded *Ádám, Jánnosy and Varga* (ÁJV) to actually perform [this] experiment[:] two independent photodetectors are placed respectively in the transmitted and reflected beams of a half-silvered mirror. If photons have a particle-like character, i.e., if their detectable components are always spatially bounded and well localized, then photons impinging on the half-silvered mirror will not be split in two at this mirror. On the other hand, if they are purely wave-like in nature, ... then they can and will be split into two independent classical wave packets at this mirror. This fact then implies that if they are purely wave-like (in this classical sense), then the two detectors will show coincidences when a single temporally localized photon is directed at said mirror. One of these independent wave packets will be transmitted to illuminate the first detector, and the other will be reflected to illuminate the second detector, and both detectors will then have a finite probability of detecting the same photon (classical wave packet). This latter possibility, however, violates the predictions [of quantum theory] which prohibits such coincidences. ÁJV thus searched for anomalous coincidences between photomultiplier tubes that viewed the reflected and transmitted beams behind a half -silvered mirror [12].

Take a look at Clauser's experimental paper, Ref. [13], reporting the results of his later version of this kind of experiment. Summarize his experimental setup and findings.

4.7 Read the actual EPR paper [6] and report back, sharing any insights, confusions, and/or questions.

4.8 Recall the quantum mechanical collapse postulate: when a measurement of some observable is made, the quantum state changes suddenly and discontinuously into the particular eigenstate (of that observable) corresponding to the actually-realized outcome of the measurement. The theme of Chap. 3 could perhaps be summarized by saying that there are two different ways one could interpret this collapse rule, and each seems problematic: first, if you think of the collapse as describing a real physical change in the state of the system, i.e., as a dynamical process, this seems impossible to reconcile with the *normal* system dynamics (namely Schrödinger's equation); whereas, second, if you think of the collapse as describing a mere updating of information, i.e., as describing a change in our *knowledge* not implying any change in the physical state of the thing described, then the quantum state descriptions are revealed as obviously incomplete. Explain how the EPR dilemma between locality and completeness can be understood in these same terms, with the two horns of the dilemma corresponding exactly to the two views (dynamic vs. epistemic) one might take to wave function collapse. What, exactly, does the EPR argument then *add* to the arguments from Chap. 3? How is it different or better?

4.9 Illustrate the EPR scenario on a space-time diagram showing (i) the preparation of the two-particles at some central source, (ii) the separating of the two entangled particles, and (iii) the measuring equipment that will perhaps be used to measure some property of one or both properties. Can Bell's formulation of "locality" be used to rehearse a more formally rigorous version of the argument, along the lines of what we did in the Chapter with the Einstein's Boxes argument? If so, explain how; or if not, explain why not. (Note: this is a bit of a

trick question, for reasons that will be discussed in the following Chapter. But it is still well worth thinking about here.)

4.10 Re-write the "EPR-Bohm state" (that is, the spin-singlet state of the two spatially-separated spin 1/2 particles), Eq. (4.10), in terms of the states ψ_{+x} and ψ_{-x}, whose relation to the states ψ_{+z} and ψ_{-z} are explained in Chap. 2. Use your re-writing to argue that not only, as explained in the text, a measurement of σ_z on particle 1 provides an indirect determination of the value of σ_z of particle 2, but that also a measurement of σ_x on particle 1 provides an indirect determination of the value of σ_x of particle 2. Discuss how and whether and under what assumptions this all implies that the distant particle must possess (contrary to what you'd say if you believe the quantum state provides a complete description) definite values of *both* σ_x *and* σ_z.

4.11 Re-write the EPR-Bohm spin state in terms of the states ψ_{+n} and ψ_{-n} from Chap. 2. Thus show that the singlet state takes the same mathematical form for spin components along *any* arbitrary axis, and that therefore the argument from the text, establishing the real existence of σ_z for the distant particle (which argument was already generalized to σ_x in Project 4.10) applies to *all* possible axes.

4.12 Suppose two spin $1/2$ particles are prepared in the EPR-Bohm spin state, Eq. (4.10). Now suppose that the spin of one of the particles (say, particle 1) is measured along the z-direction, and the spin of the other particle (2) is measured along the direction $\hat{n}$ (in the $x-z$-plane and making an angle θ with respect to the z-axis). What are the probabilities for the four possible joint outcomes (i.e., "particle 1 is spin-up/spin-down along z and particle 2 is spin-up/spin-down along n)? To answer this, write the EPR-Bohm state as a linear combination of four terms of the form $\psi^1_{\pm z}\psi^2_{\pm n}$ and read off the probabilities as the absolute squares of the coefficients. Finally, compute the "correlation coefficient", defined here as the expected value of the *product* of the two outcomes, taking spin-up/spin-down as $+1/-1$:

$$C = (+1)(+1)P_{++} + (+1)(-1)P_{+-} + (-1)(+1)P_{-+} + (-1)(-1)P_{--}. \quad (4.11)$$

Does the correlation coefficient make sense in various limiting cases like $\theta = 0$?

4.13 Sometimes the EPR argument (say, in the Bohm version in terms of spin) is explained as follows: "you can determine *both* σ_x and σ_z for the same one particle at the same time, by measuring one of these quantities *directly* (i.e., by actually measuring it on that particle) and then by also measuring the other quantity *indirectly* (i.e., by actually measuring that same quantity on the *other* particle and then attributing the opposite value to the particle in question)." The conclusion is then something like: "...so both of these properties must be real (or, at least, you can learn more about them than is supposed to be allowed by the uncertainty principle) and QM is incomplete." Is this a good argument? Discuss its merits and its relation to the actual EPR argument.

4.14 Here is a paragraph from the Wikipedia page on the "EPR Paradox" (grabbed on Jan 7, 2016):

> While EPR felt that the paradox showed that quantum theory was incomplete and should be extended with hidden variables, the usual modern resolution is to say that due to the common preparation of the two particles (for example the creation of an electron-positron pair from a photon) the property we want to measure has a well defined meaning only when analyzed for the whole system while the same property for the parts individually remains undefined. Therefore if similar measurements are being performed on the two entangled subsystems, there will always be a correlation between the outcomes resulting in a well defined global outcome, i.e., for both subsystems together. However, the outcomes for each subsystem separately at each repetition of the experiment will not be well defined or predictable. This correlation does not imply any action of the measurement of one particle on the measurement of the other, therefore it doesn't imply any form of action at a distance. This modern resolution eliminates the need for hidden variables, action at a distance or other structures introduced over time in order to explain the phenomenon.

Pretend that you are Einstein, magically transported to the present day, with both internet access and too much free time. Write a few paragraphs that you would post on the Wikipedia discussion page, explaining to the other contributors how and why the "modern resolution" described here is inadequate, and clarifying your original argument.

4.15 In Bell's re-telling of the EPR argument, he stresses that determinism is "not assumed, but *inferred*." This may be somewhat confusing since the notion of "determinism" did not play much of a role in our earlier presentations of the EPR argument. What, exactly, is the role of "determinism" in the argument? Is Bell correct?

4.16 Provide a detailed explanation of the sort of theoretical model which could explain the quantum mechanical predictions for the Einstein's Boxes scenario in a perfectly local way. Then do the same for the EPR-Bohm scenario.

References

1. Einstein's remarks from Solvay 1927, translated in Bacciogallupi and Valentini, *Quantum Theory at the Crossroads*, pp. 485–487, http://arxiv.org/pdf/quant-ph/0609184.pdf
2. A. Fine, *The Shaky Game* (University of Chicago Press, Chicago, 1986)
3. D. Howard, Einstein on locality and separability. Stud. Hist. Phil. Sci. **16**, 171–201 (1985)
4. W. Heisenberg, *The Physical Principles of the Quantum Theory* (Dover Publications, New York, 1949), p. 39
5. L. de Broglie, *The Current Interpretation of Wave Mechanics: A Critical Study* (Elsevier Publishing Company, Amsterdam, 1964)
6. A. Einstein, B. Podolsky, N. Rosen, Can quantum mechanical description of reality by considered complete? Phys. Rev. **47**, 777–780 (1935)
7. A. Einstein, Autobiographical notes, in *Albert Einstein: Philosopher-Scientist*, ed. by P.A. Schilpp (Harper and Row, New York, 1949)
8. A. Einstein, Reply to criticisms, in *Albert Einstein: Philosopher-Scientist*, ed. by P.A. Schilpp (Harper and Row, New York, 1949)
9. D. Bohm, *Quantum Theory* (Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1951)
10. J.S. Bell, Bertlmann's socks and the nature of reality, *Speakable and Unspeakable in Quantum Mechanics*, 2nd edn. (Cambridge University Press, Cambridge, 2004)

11. A. Einstein, Quantum Mechanics and Reality. Dialectica (1948)
12. J.F. Clauser, *Quantum [Un]speakables: From Bell to Quantum Information* (Springer, Berlin, 2002)
13. J.F. Clauser, Experimental distinction between the quantum and classical field-theoretic predictions for the photoelectric effect. Phys. Rev. D **9**(4), 853–860 (1974)

Chapter 5
The Ontology Problem

The previous two chapters reviewed two important arguments against the idea that quantum mechanics provides *complete* descriptions of physical reality. Here we discuss a couple of other related concerns which might be summarized by the general question: even leaving aside the question of whether or not the description is complete, what kind of physical thing – what ontology, exactly – could the quantum wave function possibly represent, and how would that representation work?

5.1 Complexity and Reality

Every student of quantum mechanics learns that the wave function Ψ is *complex*: it has, in general, both a real part and an imaginary part. This is of course not surprising given the explicit appearance of the imaginary quantity $i = \sqrt{-1}$ in the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi. \tag{5.1}$$

Some textbooks make a big deal of the fact that Ψ is complex – arguing that (unlike complex numbers that are sometimes used as a matter of convenience in for example classical electrodynamics) the complexity of the quantum wave function is somehow a deep and fundamental requirement with profound implications. Indeed, one of these alleged implications is that the quantum mechanical wave function Ψ *cannot* represent a "physically real field" in the way, for example, that the classical electric and magnetic fields, $\vec{E}$ and $\vec{B}$ are supposed to. The classic *Quantum Physics* text by Eisberg and Resnick, for example, states:

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_5

> The fact that wave functions are complex functions should not be considered a weak point in the quantum mechanical theory. Actually, it is a desirable feature because it makes it immediately apparent that we should not attempt to give to wave functions a physical existence in the same sense that water waves have a physical existence. The reason is that a complex quantity cannot be measured by any actual physical instrument. The 'real' world (using the term in its nonmathematical sense) is the world of 'real' quantities (using the term in its mathematical sense) [1, p. 134].

If such a view is correct, then evidently we would have to reject the idea that the quantum mechanical wave function provides a complete description of physical reality: if Ψ cannot represent something physically real at all, because it is not a mathematically real function, then certainly it cannot provide a faithful and full representation!

But I do not think this type of argument is convincing at all. It is simply false that one is somehow *required* to use complex numbers. For example, one could always break the quantum wave function apart into its real and imaginary parts:

$$\Psi(x,t) = f(x,t) + i\, g(x,t) \tag{5.2}$$

(where f and g are now *real* functions). Then, by plugging this ansatz into Schrödinger's equation and taking real and imaginary parts, we can break Schrödinger's (complex) equation into two coupled real equations:

$$-\hbar\frac{\partial g}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2 f + Vf \tag{5.3}$$

and

$$\hbar\frac{\partial f}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2 g + Vg. \tag{5.4}$$

We could think of this set of equations as perhaps something like Maxwell's equations, which couple together the dynamics of the two (perfectly real!) fields $\vec{E}$ and $\vec{B}$.

It should be clear that one can always do this: a single complex field is mathematically equivalent to two coupled real fields. So the most the Eisberg/Resnick-type argument could establish is that, if one wants to regard the quantum wave function as representing something physically real, one would have to interpret it as representing these two coupled fields.

This may sound somewhat contrived and artificial, but actually there is an even closer parallel to all of this in Maxwellian electromagnetism that is worth pointing out. Take, for simplicity, the case of electromagnetic fields propagating in empty space (where the charge density ρ and the current density $\vec{j}$ both vanish). The four Maxwell equations are then

$$\vec{\nabla}\cdot\vec{E} = 0 \qquad \text{and} \qquad \vec{\nabla}\cdot\vec{B} = 0 \tag{5.5}$$

and then also

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \qquad \text{and} \qquad \vec{\nabla} \times \vec{B} = \frac{1}{c^2}\frac{\partial \vec{E}}{\partial t}. \tag{5.6}$$

Now notice that we can rewrite these four equations – and indeed make them look a little simpler and nicer! – by re-writing them in terms of the *complex* quantity

$$\vec{F} = \vec{E} + ic\vec{B} \tag{5.7}$$

which is sometimes called the Riemann–Silberstein vector. (Note that the "c" is put into the definition so the units of the two terms on the right hand side are the same.) In terms of this quantity, it is easy to see that the first pair of Maxwell equations can be re-expressed as:

$$\vec{\nabla} \cdot \vec{F} = 0. \tag{5.8}$$

The real and imaginary parts, respectively, reproduce the two entries in Eq. (5.5). Now the cool thing is that the two entries in Eq. (5.6) can also be reproduced by writing

$$\vec{\nabla} \times \vec{F} = \frac{i}{c}\frac{\partial \vec{F}}{\partial t}. \tag{5.9}$$

You should take a second and check that, indeed, the real and imaginary parts of this equation correspond exactly to the two entries in Eq. (5.6).

One can even rearrange the last equation into the following form:

$$i\hbar\frac{\partial \vec{F}}{\partial t} = \hbar c\vec{\nabla} \times \vec{F} \tag{5.10}$$

whose structure is rather like that of Schrödinger's equation:

$$i\hbar\frac{\partial \vec{F}}{\partial t} = \hat{H}\vec{F} \tag{5.11}$$

where, evidently, the Hamiltonian operator here is

$$\hat{H} = \hbar c\,\vec{\nabla} \times \ . \tag{5.12}$$

Indeed, the complex vector $\vec{F}$ can (with some caveats) be understood as a kind of "quantum wave function for the photon." But we will not pursue this interesting connection any further here.

Our point is instead just that the dynamical equations – that is, Maxwell's equations – for the fields $\vec{E}$ and $\vec{B}$ can be re-expressed in an elegant form by combining $\vec{E}$ and $\vec{B}$ into a single, complex-valued quantity, $\vec{F}$. Yet nobody, I think, regards this as some kind of proof that $\vec{F}$ cannot correspond to anything physically real. It does! Its real part – $\vec{E}$ – corresponds to the physically real electric field, and its imaginary part

– $\vec{B}$ – corresponds to the physically real magnetic field. We should therefore allow this same flexibility of mathematical representation in the case of the quantum wave function Ψ and remain open to the possibility that, despite being complex-valued, it represents a physically real (or perhaps more than one coupled physically real) field.

5.2 Configuration Space

There is, however, a different mathematical fact about the quantum wave function Ψ that leads to a much more difficult and troubling question about the ontology it might conceivably describe: Ψ is a function on *configuration space*. So if Ψ represents one or more *fields* – things somehow like the electromagnetic fields $\vec{E}$ and $\vec{B}$ – the fields would be very, very unusual because they would live, not in ordinary three-dimensional physical space, but rather a higher-dimensional and seemingly purely abstract space.

This was, interestingly, a worry that arose almost immediately when Schrödinger first invented (and wrote down his dynamical equation for) the wave function. In 1926, for example, Schrödinger sent copies of his papers to a number of his colleagues and asked for their comments. Hendrik Lorentz, in his very first reply, praised Schrödinger's work for its physical/intuitive comprehensibility (compared to the more purely mathematical, and hence physically obscure, "matrix mechanics" which had previously been developed by Heisenberg and others). But Lorentz also raised a number of concerns about Schrödinger's wave mechanics, and the very first of these concerns had to do with the fact that the wave was not a wave in physical space but instead the abstract configuration space:

> Dear Colleague,
>
> I am finally getting around to answering your letter and to thanking you very much for kindly sending me the proof sheets of your three articles, all of which I have in fact received. Reading these has been a real pleasure to me. Of course the time for a final judgment has not come yet, and there are still many difficulties, it seems to me, about which I shall get to speak immediately. But even if it should turn out that a satisfactory solution cannot be reached in this way, one would still admire the sagacity that shows forth from your considerations, and one would still venture to hope that your efforts will contribute in a fundamental way to penetrating these mysterious matters.
>
> I was particularly pleased with the way in which you really construct the appropriate matrices and show that these satisfy the equations of motion. This dispels a misgiving that the works of Heisenberg, Born, and Jordan, as well as Pauli's, had inspired in me: namely, that I could not see clearly that in the case of the H-atom, for example, a solution of the equations of motion can really be specified. With your clever observation that the operators q and $\frac{\partial}{\partial q}$ commute or do not commute with each other in a similar way to the q and p in the matrix calculation, I began to see the point. In spite of everything it remains a marvel that equations in which the q's and p's originally signified coordinates and momenta, can be satisfied when one interprets these symbols as things that have quite another meaning, and only remotely recall those coordinates and momenta.
>
> If I had to choose now between your wave mechanics and the matrix mechanics, I would give the preference to the former, because of its greater intuitive clarity, so long as one only

> has to deal with the three coordinates x, y, z. If, however, there are more degrees of freedom, then I cannot interpret the waves and vibrations physically, and I must therefore decide in favor of matrix mechanics. But your way of thinking has the advantage for this case too that it brings us closer to the real solution of the equations; the eigenvalue problem is the same in principle for a higher dimensional q-space as it is for a three dimensional space [2, p. 43–44].

Note in particular the suddenness – the immediate finality – with which Lorentz simply dismisses the possibility that a function on configuration space could represent a physically real field: "I cannot interpret the waves and vibrations physically...."

Indeed, it is interesting that Lorentz does not just say that, in the case of two (or more) particles where the configuration space is 6- (or more) dimensional, he is confused about how to interpret Schrödinger's wave function. Instead, he says he would prefer in this case to go back to the physically obscure matrix mechanics. We can only speculate about exactly what he meant, and this was admittedly only a first impression, but one gets the feeling that he preferred to have no intuitive physical interpretation available at all, rather than one which was so obviously absurd as to suggest the existence of fields/waves in an unphysical, abstract space.

Einstein expressed a similar concern about Schrödinger's wave function in letters from this same period. Here are some excerpts, all quoted in Ref. [3]:

- "Schrödinger's conception of the quantum rules makes a great impression on me; it seems to me to be a bit of reality, however unclear the sense of waves in n-dimensional q-space remains". (May 1, 1926, to Lorentz)
- "Schrödinger's works are wonderful – but even so one nevertheless hardly comes closer to a real understanding. The field in a many-dimensional coordinate space does not smell like something real." (June 18, 1926, to Ehrenfest)
- "The method of Schrödinger seems indeed more correctly conceived than that of Heisenberg, and yet it is hard to place a function in coordinate space and view it as an equivalent for a motion. But if one could succeed in doing something similar in four-dimensional space, then it would be more satisfying." (June 22, 1926, to Lorentz)
- "Of the new attempts to obtain a deeper formulation of the quantum laws, that by Schrödinger pleases me most. If only the undulatory fields introduced there could be transplanted from the n-dimensional coordinate space to the 3 or 4 dimensional!" (August 21, 1926, to Sommerfeld)
- "Schrödinger is, in the beginning, very captivating. But the waves in n-dimensional coordinate space are indigestible..." (August 28, 1926, to Ehrenfest)
- "The quantum theory has been completely Schrödingerized and has much practical success from that. But this can nevertheless not be the description of a real process. It is a mystery." (February 16, 1927, to Lorentz)

Even Schrödinger himself admitted quite openly that, as a function on configuration space, the wave function can't really be understood as corresponding to some kind of physically real wave. In the abstract of "Wave Mechanics," his contribution to the 1927 Solvay conference, he wrote:

> Of course this use of the q-space is to be seen only as a mathematical tool, as it is often applied also in the old mechanics; ultimately ... the process to be described is one in space and time [4, p. 447].

In the body of the paper he elaborates on the crucial question:

> What does the ψ-function mean now, that is, *how does the system described by it really look like in three dimensions*? Many physicists today are of the opinion that it does not describe the occurrences in an individual system, but only the processes in an ensemble of very many like constituted systems that do not sensibly influence one another and are all under the very same conditions. I shall skip this point of view since others are presenting it. I myself have so far found useful the following perhaps somewhat naive but quite concrete idea. The classical system of material points does not really exist, instead there exists something that continuously fills the entire space and of which one would obtain a 'snapshot' if one dragged the classical system, with the camera shutter open, through *all* its configurations, the representative point in q-space spending in each volume element $d\tau$ a time that is proportional to the *instantaneous* value of $\psi\psi^*$. (The value of $\psi\psi^*$ for only *one* value of the argument t is thus in question.) Otherwise stated: the real system is a superposition of the classical one in all its possible states, using $\psi\psi^*$ as 'weight function' [4, p. 453].

The first view that Schrödinger mentions here – according to which the ψ function "does not describe ... an individual system" but instead characterizes "an ensemble of very many like constituted systems" – is the idea, argued for in the previous two chapters, that the wave function does *not* provide a complete description of physical reality.

But in contrast to these interpretations Schrödinger here suggests an alternative view in which physical reality really is faithfully described by the wavy, spread-out wave function. He speaks of "something that continuously fills the entire space" and then suggests that one could perhaps understand "the real system [as] a superposition of the classical one in all its possible states, using $\psi\psi^*$ as a 'weight function'."

For a single particle, whose wave function $\psi(\vec{x}, t)$ lives in ordinary, physical, three-dimensional space, one can understand this idea as saying that the "particle" is really a cloud whose density is given by the square of the wave function. For example, one could characterize the electron in terms of a mass density

$$\rho_m(\vec{x}, t) = m\,|\psi(\vec{x}, t)|^2 \tag{5.13}$$

or an electric charge density

$$\rho_e(\vec{x}, t) = e\,|\psi(\vec{x}, t)|^2 \tag{5.14}$$

where m and e are the total mass and charge of the electron. One often sees – in, for example, Chemistry textbooks – pictures of the "electron cloud" surrounding the nucleus for, say, different states of the Hydrogen atom. Such pictures invite you to think of the wave function in the way that Schrödinger was suggesting here.

The problem, of course, is that this interpretation doesn't make any sense as soon as one has a quantum system with more than one particle in it. Then the wave function Ψ is a function on configuration space, and so the "charge density"

$$\rho \sim |\Psi|^2 \tag{5.15}$$

would also be a charge density in this high-dimensional, abstract space. And that, to use Einstein's phrase, simply "does not smell like something real".

It is sometimes difficult for people to fully grasp the nature of the problem associated with the fact that the quantum mechanical wave function (for a system of more than 2 particles) is a function on configuration space. Basically the problem is that, mathematically, the wave function is – like the electric and magnetic fields of classical electrodynamics – a function of continuous spatial degrees of freedom which satisfies a dynamical evolution equation with the general structure of a *wave equation*, i.e., a partial differential equation relating spatial and temporal partial derivatives. Mathematically, in short, the wave function seems to look and act like a *field*. But unlike the familiar and unproblematic electric and magnetic fields, $\vec{E} = \vec{E}(\vec{x}, t)$ and $\vec{B} = \vec{B}(\vec{x}, t)$, we cannot ask for the value of the wave function at a point in three-dimensional space at a particular time: it is not $\Psi = \Psi(\vec{x}, t)$ but rather $\Psi = \Psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_N, t)$. So if the wave function describes a *field*, it would not appear to be a *physical* field in the sense that we are accustomed to thinking about $\vec{E}$ and $\vec{B}$ from electrodynamics. So if the wave function provides a description of some physically real things or stuff, the description must be in some sense indirect, abstract.

But whereas a single point in 3N-dimensional configuration space can be easily understood as an abstract representation of the configuration of N particles in 3-dimensional physical space (that was the kind of thing we discussed when "configuration space" came up in Chap. 1), it is hard to see what kind of thing a *field* in configuration space might be an abstract representation *of*. That is, when we are dealing with a multi-particle wave function in quantum mechanics, it is simply not clear, the way it is clear when we use configuration space in the context of the classical mechanics of particles, what we are talking about!

Of course, one apparent possibility (the one mentioned first by Schrödinger) would be that the wave function simply describes our incomplete *knowledge* of the state of a set of (literal, pointlike) particles. On this view, the wave function isn't a physical thing (like a field) at all – the wave function, that is, is epistemic, not ontological. As discussed already in Chap. 3, this interpretation is very difficult to reconcile with, for example, the existence of 2-slit interference. Still, it is helpful to have in mind as a possible way of avoiding the "ontology problem" which, it would seem, is going to afflict any theory that takes the quantum mechanical wave function seriously, as providing a direct and literal description of some kind of physically real thing. This problem would seem to be particularly worrisome for the orthodox viewpoint according to which the wave function provides a *complete* description of the physical state of the system being described. The *ontology* of orthodox quantum mechanics, that is, seems entirely mysterious: the one kind of thing that we can straightforwardly understand the wave function as describing (namely, a physical field living in the abstract, high-dimensional configuration space) seems unacceptably bizarre and absurd and, indeed, seems not really to be a legitimate physical "thing" at all.

Perhaps we can thus summarize "the ontology problem" as follows: in quantum mechanics, there simply is nothing in the theory other than the wave function Ψ with which to describe the physical state of a microscopic system; but it simply is not clear how the wave function Ψ might be understood as describing some material structures in three-dimensional physical space. Put simply, it is just not at all clear, from the mathematical formulation of the theory, what sort of physical things quantum mechanics might be *about*.

In Sect. 5.4 we will consider an old idea of Schrödinger's for trying to address the question: what kind of (field-like) physical reality might the wave function be an abstract representation *of*? But first, let us develop a bit further our thinking about the nature of the ontology problem in general.

5.3 Ontology, Measurement, and Locality

To help flesh out the problem and to emphasize its fundamentality, let's explore the connections between the ontology problem and the other two problems we discussed in the previous two chapters.

Recall first the measurement problem. Here is a quick summary of the way we presented the problem in Chap. 3, using the concrete example of the particle-in-a-box whose energy is to be measured by a device which will indicate the result with the position of a pointer:

> If the particle-in-a-box starts out in an energy eigenstate (such that it *has* a definite pre-measurement energy E_n), the Schrödinger equation evolution of the particle-pointer system proceeds according to
>
> $$\psi_n(x)\ \phi(y) \quad \rightarrow \quad \psi_n(x)\ \phi(y - \lambda E_n T). \tag{5.16}$$
>
> This is unproblematic since the final state $\Psi(x, y, T) = \psi_n(x)\,\phi(y - \lambda E_n T)$ describes the particle (still) being in a state of definite energy E_n and describes the pointer as having moved a definite distance $(\lambda E_n T)$ past its initial position, i.e., correctly and unambiguously indicating that the energy of the particle was E_n.
>
> However, if the particle-in-a-box instead starts out in a superposition of energy eigenstates, the Schrödinger evolution of the particle-pointer wave function looks like this:
>
> $$\left[\sum_i c_i\,\psi_i(x)\right]\phi(y) \quad \rightarrow \quad \sum_i c_i\ \psi_i(x)\,\phi(y - \lambda E_i T). \tag{5.17}$$
>
> This is highly problematic since the final state does not seem to attribute any particular position to the pointer; instead, the pointer is in an (entangled) superposition of many distinct locations, and this simply doesn't seem to correspond to the observed behavior of real pointers.

Our main goal in Chap. 3, following Schrödinger, was to emphasize the difficulty of understanding macroscopic superpositions. It was from this point of view that we stressed the problematic character of the final state in Eq. (5.17), as against the comparatively unproblematic final state in Eq. (5.16).

But now we'd like to cycle back and ask: is the final state in Eq. (5.16) really so unproblematic? When we said so before, we simply took for granted that we could understand a product state like

$$\Psi(x, y, T) = \psi_n(x)\,\phi(y - \lambda E_n T) \tag{5.18}$$

as saying "the particle-in-a-box is in the nth energy eigenstate" *and* "the pointer is a distance $\lambda E_n T$ to the right of its ready position". But what is really going on when we interpret this state this way?

First of all, we are assuming that when the overall wave function is a product, we can simply "peel apart" the factors and take them individually as describing the states of the individual sub-systems in question. As straightforward as this is in the case of a product state, though, it is simply not possible in general: as soon as the overall state fails to be a product, nothing like this straightforward mathematical "peeling apart" procedure is possible. So we should probably be suspicious of the propriety of this procedure in general, i.e., even in the special kind of situation where it is technically possible.

Second, our descriptions of the physical meanings of the individual factors were a little bit abstract. For example, to say that the particle-in-a-box has a certain energy is really just to say that an appropriate energy measuring device will respond in a certain way if it is allowed to interact with the particle. A more direct and literal description of the state of the particle (in terms of constitutive rather than dispositional properties, one might say) would instead just say: its wave function has a certain spatial structure, namely, that of $\psi_n(x)$.

The overall point is that if we simply take the quantum mechanical description literally, then the state of the particle-pointer system is just given by its wave function Ψ. And this, taken at face value, means that there are some regions of the abstract (here, two-dimensional) configuration space where Ψ has "high intensity". This is depicted, in Fig. 5.1, for the final states $\Psi(x, y, T)$ associated with the two cases from Eqs. (5.16) and (5.17). And then the point is that, even in the simpler case depicted in the left panel, which we previously regarded as relatively unproblematic, there is a question about how this wave function corresponds to the kinds of physical objects we thought we were describing. This high intensity region of configuration space – the "blob" colored grey in the left panel of Fig. 5.1 – doesn't exactly look like a particle-in-a-box (in some particular state) and a pointer (with some reasonably sharp location), each moving in a one-dimensional physical space.

To be clear, the claim here is not that this problem is insoluble. The obvious response would be to insist that the "blob" in configuration space *does* represent the state of our two particles, but the representation is somehow indirect or abstract. After all, the single dot in the right-hand panel of Fig. 1.12, back in Chap. 1, didn't really "look like" the two particles depicted, in ordinary physical space, in the left-hand panel of that same Figure. And yet we have no trouble understanding how the one thing can perfectly well represent the other. Isn't it just the same here with our quantum mechanical wave functions?

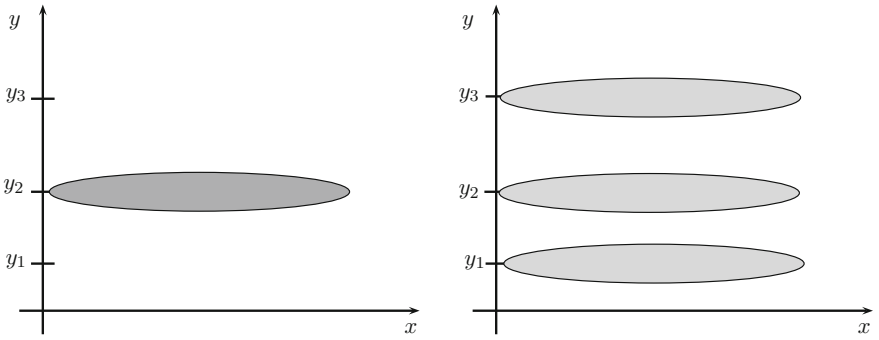


Fig. 5.1 "Configuration space cartoons" showing the intensity of $\Psi(x, y, T)$ for the two cases discussed in the text. The *left panel* corresponds to Eq. (5.16), in which the particle-in-a-box (whose energy is being measured) begins in the nth energy eigenstate (here the case $n = 2$ is depicted) and the pointer's final position is y_n (so, here, y_2). The right panel corresponds to Eq. (5.17), in which the particle-in-a-box begins in a superposition of several energy eigenstates, and so the pointer ends up in an entangled superposition involving several different positions (here y_1, y_2, and y_3). The novel point being developed in the present chapter is that, in addition to the difficulties associated with the wave function depicted in the *right panel*, there is a deeper kind of problem: even the single blob in the *left panel* does not, on its face, "look like" two particles (the particle-in-the-box and the pointer) moving in one spatial dimension. The relationship between the quantum mechanical wave function, and some ontology of objects in three-dimensional space, remains obscure

The point is: it might well be. But unlike the case of classical mechanics, where we *started out* with a clear ontology of particles (moving and interacting in three-dimensional space) and then constructed abstract representations, like configuration space, to describe these particles in a new way; here, in the quantum mechanical case, we *only* have, as yet, the abstract representation. We don't yet know what kind of reality, what sort of physically real objects or stuff, in three-dimensional space, the wave functions might be abstract representations *of*.

Pointing out that quantum mechanics suffers from an "ontology problem" is thus largely a plea for help: anyone who says that quantum mechanical wave functions should be taken seriously, as corresponding in some sense to physical reality (as opposed, for example, to our incomplete knowledge), should be asked to explain in concrete, mundane detail how that alleged correspondence works. They should tell us what sorts of things (particles?) or stuff (fields??) quantum mechanics is *about*, and clarify in precise mathematical detail the relationship between those things (and/or that stuff) and quantum mechanical wave functions. Until or unless this is done, I think we have to admit that the connection with the three-dimensional reality of direct experience remains puzzling, not only for the final state described in Eq. (5.17), but also – already – for the state described in Eq. (5.16).

Let us then turn to exploring how the "ontology problem" relates to the other of the two big worries we explored previously: the "locality problem".

The locality problem, recall, was that – if one regards quantum mechanical wave functions as providing complete descriptions of physical states – then quantum

mechanics evidently violates the relativistic notion of local causality. We developed Bell's careful formulation of local causality in Chap. 1 and then showed explicitly, in Chap. 3, how local causality is violated by ordinary quantum mechanics (assumed complete) in the "Einstein's Boxes" scenario. But you might have noticed that we never applied our explicit formulation of local causality to the two-entangled-particle EPR scenario in the same way.

The reason for this has to do with the ontology problem. Recall that, in Bell's formulation of locality, we need to compare the probabilities assigned to some event, conditioned on a *complete specification of the physical state on a slice across the past light cone of that event*, when some distant event is, and is not, also specified. In the EPR-Bohm situation, in which the two particles are jointly in the spin singlet state

$$\Psi = \frac{1}{\sqrt{2}} \left[\psi^1_{+z} \psi^2_{-z} - \psi^1_{-z} \psi^2_{+z} \right], \tag{5.19}$$

for example, we might ask whether

$$P\,[\,A\,|\,\mathcal{C}_\Sigma\,] \;=\; P\,[\,A\,|\,\mathcal{C}_\Sigma,\; B\,] \tag{5.20}$$

where A is, say, the event "particle 1 comes out spin-up along the z-direction", B is the event "particle 2 comes out spin-up along the z-direction", and $\mathcal{C}_\Sigma$ is a complete specification of the physical state in region Σ indicated in Fig. 5.2.

Intuitively, one wants to say that locality is violated, because, according to quantum mechanics (assumed complete) there is just an irreducible 50/50 probability for a z-spin measurement on particle 1 to come out up/down. So the probability on the left hand side of Eq. (5.20) is 50%. Whereas if we specify also B – that a z-spin measurement on particle 2 comes out spin-up – then the probability assigned to particle 1 being spin-up along the z-direction, i.e., the right hand side of Eq. (5.20) is instead zero. So locality is violated, just as in the one-particle Einstein's Boxes scenario.

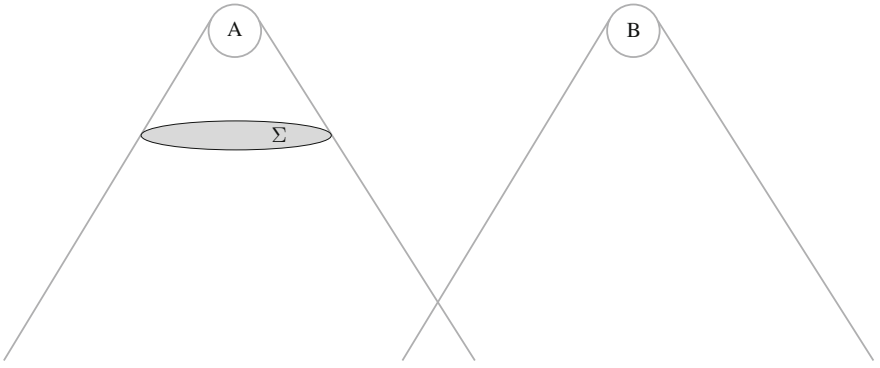


Fig. 5.2 Space-time diagram for the attempted application of Bell's formulation of local causality to the Einstein–Podolsky–Rosen scenario

But is it really so clear that, for example, $P\,[A|\mathcal{C}_\Sigma] = 50\%$? What, exactly, is $\mathcal{C}_\Sigma$ here? Again, intuitively, one wants to say: it is the complete quantum mechanical description of the state of particle 1, i.e., the wave function of particle 1. But when particles 1 and 2 are in an entangled state, *there is simply no such thing as* "the wave function of particle 1". One cannot "peel apart" the two-particle state into two one-particle states when the two-particle state is entangled.

This may be slightly obscure since, in the EPR-Bohm scenario, we are largely suppressing the spatial degrees of freedom of the two particles and focusing on the (more abstract) spin degrees of freedom. But (as we saw already in Chap. 2) it is perfectly possible for the spatial degrees of freedom of two particles to be entangled. Indeed, you will recall that the original EPR argument was framed in terms of an example involving the entanglement of the spatial degrees of freedom of two particles. Here is another example that is a little better suited to our immediate needs here: suppose that particle 1 is definitely in a room a million miles to the West, either on the left side of the room (ψ_L^W) or on the right side of the room (ψ_R^W). Similarly, particle 2 is in a different room, a million miles to the East, either on the left side of that room (ψ_L^E) or the right side of that room (ψ_R^E). Then suppose in particular that the two particles are in the following entangled state

$$\Psi(x_1, x_2) \;=\; \frac{1}{\sqrt{2}}\left[\,\psi_L^W(x_1)\;\psi_L^E(x_2) \;+\; \psi_R^W(x_1)\;\psi_R^E(x_2)\,\right] \tag{5.21}$$

which can be understood as a superposition of "both particles are on the left sides of their respective rooms" and "both particles are on the right sides of their respective rooms".

It should be clear that, if the quantum mechanical state of the two particles jointly is given by Eq. (5.21), it is impossible to assign a one-particle wave function to either particle alone. If you have even the slightest doubt about this, take a minute and try to work out what you think "the wave function of particle 1" is, for example.[1]

The problem here is of course just the problem we've been focusing on in this chapter: quantum mechanical wave functions for multiple-particle systems are something like fields, but in an abstract configuration space rather than ordinary three-dimensional physical space. So it is simply not meaningful to, for example, consider "the part" of such a wave function that describes goings-on in a particular region of space such as Σ in Fig. 5.2. There is no such "part" because the wave function doesn't live in (ordinary, three-dimensional, physical) space to begin with.

[1]Experts may object here that, although one cannot assign a one-particle *wave function* to either particle in this kind of situation, one may nevertheless describe the state of each particle separately using something called a reduced density matrix. This is in some sense true but is nevertheless ultimately unhelpful. The reduced density matrix for particle 1 is a formal way of expressing that particle 1 is either on the left, or the right, side of its room, with 50/50 probability. And then similarly for particle 2. But then, crucially, the two reduced density matrices together fail to capture everything that is implied by Eq. (5.21); in particular, the *correlation* between the positions of the two particles (namely the fact that either both are on the left or both are on the right), is lost. So it is simply not true that the two reduced density matrices jointly capture, in the form of state descriptions for the two particles separately, the full state of the two-particle system as given in Eq. (5.21).

So what does this mean, with respect to the question of whether quantum mechanics (assumed complete) is a local theory? One might think that, by showing that we cannot cleanly apply Bell's formulation of locality to diagnose the theory as non-local in the EPR scenario, we leave the door open to the claim that perhaps the theory is, after all, consistent with relativistic local causality. But that is wrong, for several reasons.

First, we shouldn't forget that we were already able to diagnose quantum mechanics (assumed complete) as non-local in the simpler, one-particle "Einstein's Boxes" scenario, where the weirdness associated with multi-particle wave functions never came up. So if the question is just "Is ordinary quantum mechanics (assumed complete) a local theory?" that question has already been decisively and conclusively answered in the negative before we even get around to considering the EPR scenario.

Second, "not cleanly diagnosable as non-local" is not the same as "local". Indeed, in the spirit of Pauli's memorable phrase "not even wrong", it would probably be most accurate to summarize the situation by describing quantum mechanics here as "not *even* non-local". Remember what "locality" means: the causal influences that objects, moving and interacting in three-dimensional space, exert on one another, always propagate at the speed of light or slower. A theory which fails to provide a clear ontology of objects moving and interacting in three-dimensional space – for example, a theory which posits something like a physical field that lives in a more abstract space – doesn't even rise to the level of making the question, of whether causal influences always propagate at the speed of light or slower, or not, meaningful. In this sense I think it would be appropriate to say that ordinary quantum mechanics (assumed complete) is even less local than, for example, our paradigmatic example of a non-local theory: Newtonian mechanics with instantaneous gravitational forces. Whereas Newtonian mechanics provides a perfectly coherent "local ontology" (of objects, in this case particles, that move and interact in three-dimensional space) but posits "dynamical non-locality" (in the form of the gravitational interactions), quantum mechanics does not even appear to provide a coherent local ontology.

And third, it *is* possible to cleanly diagnose quantum mechanics as violating a modified version of Bell's formulation, as follows. Suppose we accept Bell's formulation as reasonable, at least for the kinds of theories to which it can be cleanly applied. This means we accept, if the two probabilities in Eq. (5.20) are different, when $\mathcal{C}_\Sigma$ is a complete specification of the physical state of region Σ in Fig. 5.2 according to some candidate theory, that the theory is non-local. But then surely if, in place of $\mathcal{C}_\Sigma$, we conditionalize both probabilities on *even more information* – say, a complete specification of the physical state of region Σ and some other stuff besides – we must still conclude that the theory is non-local if the two probabilities are different. That is, we may take

$$P\,[\,A\,|\,\mathcal{C}\,] \;=\; P\,[\,A\,|\,\mathcal{C},\;B\,]\,, \tag{5.22}$$

where A and B are situated as in Fig. 5.3 and $\mathcal{C}$ *includes* (but is not necessarily restricted to providing merely) a complete specification of events in Σ, as a *necessary condition* for locality.

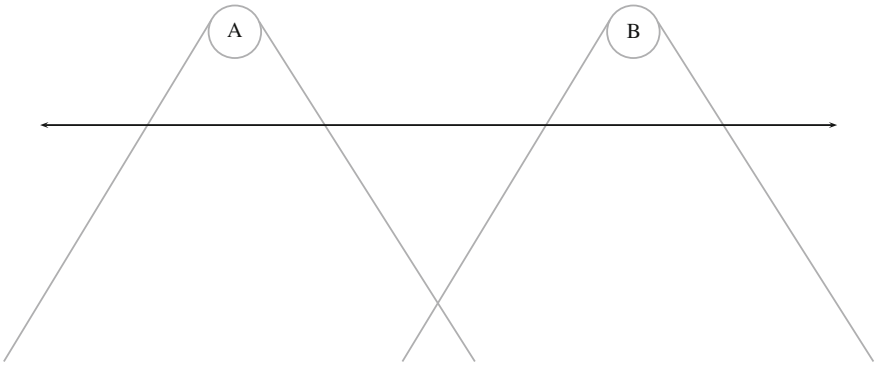


Fig. 5.3 Space-time diagram for the attempted application of Bell's formulation of local causality to the Einstein–Podolsky–Rosen scenario

In particular, we may let $\mathcal{C}$ denote a complete specification of events *throughout the entire universe*, at some moment in time prior to the events A and B, i.e., on the unbounded "slice" indicated by the horizontal black line in Fig. 5.3. If Eq. (5.22) is violated, for some theory, even with this hugely expanded $\mathcal{C}$, then surely the theory should be considered non-local.

The nice thing about this modified version of Bell's formulation is that multi-particle quantum mechanical wave functions are at least well-defined at particular moments in time. That is, we needn't any longer worry about the impossibility of extracting, from an entangled two-particle wave function, the "part" that pertains to a certain region of space; instead, we can just take the theory at face value and accept that, somehow, the ontologically mysterious two-particle wave function $\Psi(x_1, x_2, t)$ provides a complete description of the state of the two particles in question at time t. And so Ψ (supplemented, as appropriate, by any relevant macroscopic goings-on, but these play no important role here and will be suppressed for simplicity) can play the role of $\mathcal{C}$ in Eq. (5.22), which thus reduces to

$$P\,[\,A\,|\,\Psi\,]\;=\;P\,[\,A\,|\,\Psi,\;B\,]\,. \tag{5.23}$$

But these two probabilities are simply not equal, in just exactly the intuitive way we sketched at the beginning of this discussion: if A refers to particle 1 emerging as "spin-up" along the z-direction, and B refers to particle 2 emerging as "spin-up" along the z-direction, then the probability on the left hand side is 50% whereas the one on the right hand side is zero. So our necessary condition for locality is violated, and we must conclude that the theory is non-local.

Hopefully this long digression about the status of quantum mechanics, with respect to our concept of relativistic local causality, has illuminated an important loose end from Chap. 4. But of course the real point of this discussion, in the context of the present chapter, is to stress the problematical character of multi-particle quantum mechanical wave functions, taken as somehow providing complete descriptions of the physical states of those multi-particle systems.

5.4 Schrödinger's Suggestion for a Density in 3-Space

So far in this chapter we have been exploring the nature and fundamentality of the ontology problem. The (ultimately untenable) idea that the wave function is purely epistemic, with the ontology of the theory just being ordinary literal particles (like in classical mechanics), was mentioned as at least one possible way of eluding the problem. In this section, we turn to the one other remotely plausible proposed solution that I know of. I think, at the end of the day, this proposed solution is also unsatisfactory, for reasons that we will discuss. Nevertheless, it is worth exploring, because having some relatively clear and concrete ideas in mind, for the sort of three-dimensional physical reality that quantum mechanical wave functions might be (complete) descriptions of, will help clarify the nature of the difficulty and will put us in a better position to appreciate some more sophisticated proposals that we'll explore in subsequent chapters.

In the letter he wrote back to Lorentz (in response to Lorentz's letter that was quoted back in Sect. 3.2) Schrödinger proposed an answer to the puzzle about the wave function Ψ (and hence any density functions proportional to $|\Psi|^2$) being a function on the $3N$-dimensional configuration space:

> My dear Professor Lorentz,
>
> You have rendered me the extraordinary honor of subjecting the train of thought in my latest papers to a profound analysis and criticism on eleven closely written pages. I cannot find words with which to thank you sufficiently for this precious gift that you have thereby made to me; I am deeply distressed that I have made such excessive demands on your time in this way.....
>
> 1. You mention the difficulty of projecting the waves in q-space, when there are more than three coordinates, into ordinary three dimensional space and of interpreting them physically there. I have been very sensitive to this difficulty for a long time but believe that I have now overcome it. I believe, (and I have worked it out at the end of the third article), that the physical meaning belongs not to the quantity itself but rather to a *quadratic* function of it. *There* [i.e., in the article] I chose [a somewhat more complicated quadratic function of ψ]. *Now* I want to choose more simply $\psi\bar{\psi}$ [$= |\psi|^2$], that is, the square of the absolute value of the quantity ψ. If we now have to deal with N particles, then $\psi\bar{\psi}$ (just as ψ itself) is a function of $3N$ variables or, as I want to say, of N three dimensional spaces, R_1, R_2,..., R_N. Now first let R_1 be identified with the real space and integrate $\psi\bar{\psi}$ over R_2, R_3, ..., R_N; second, identify R_2 with the real space and integrate over R_1, R_3, ..., R_N; and so on. The N individual results are to be added after they have been multiplied by certain constants which characterize the particles (their charges, according to the former theory). I consider the result to be the electric charge density in real space [2, pp. 55–56].

That is, Schrödinger's original idea for physically interpreting the meaning of the wave function – even in cases of $N \geq 2$ particles where the wave function is a function on $3N$-dimensional configuration space – is to use the wave function to construct a (mass or) charge density for each particle separately, and to then add these together to get the total (say) charge density.

Let us express formally this idea that Schrödinger wrote in words. For an N-particle system with wave function

$$\Psi = \Psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_N, t) \tag{5.24}$$

we may construct the electric charge density of the ith particle this way:

$$\rho_i(\vec{x}, t) = q_i \int |\Psi|^2 \, \delta^3(\vec{x}_i - \vec{x}) \, d^3x_1 \, d^3x_2 \cdots d^3x_N \tag{5.25}$$

where q_i is the charge of the ith particle. Note that we integrate over the coordinates of all N particles here, but the δ-function makes the integration over the coordinates of the ith particle trivial. So the expression here is equivalent to saying, as Schrödinger expressed it in the letter: for the ith particle, *only* integrate over the coordinates associated with all of the *other* particles; this gives a function of $\vec{x}_i$ (and of course t); now regard this as a function on physical space, i.e., just identify these three remaining coordinates with the three coordinates in physical space, i.e., $x_i = x$, $y_i = y$, and $z_i = z$.

The total charge density for all N particles is then defined by simply summing the charge densities for the individual particles:

$$\rho_{\text{total}}(\vec{x}, t) = \sum_{i=1}^{N} \rho_i(\vec{x}, t). \tag{5.26}$$

And it is this total charge density which Schrödinger proposes as the physical reality described by the wave function Ψ.

Schrödinger eventually gave up on this picture, for reasons which have to do with another of the worries raised by Lorentz in his original letter: Schrödinger's equation implies that wave packets *spread*, as we saw in Chap. 2, and so it turns out that, as time evolves, these nice little "clouds" – lumps of nonzero charge density – would diffuse into an increasingly blurry haze that doesn't seem to correspond to the relatively sharp macroscopic world that, as we saw in Chap. 3, Schrödinger was at pains to make sure to capture in the fundamental theory. And the situation appears to be even worse when we consider the possibility of (entangled) superpositions of macroscopically distinct states, as illustrated for example by Schrödinger's cat. (Appearances, however, might in this case turn out to be misleading, as we will see in Chap. 10 when we examine the "many-worlds" interpretation of Hugh Everett.)

For now, though, let us set aside these sorts of concerns about whether Schrödinger's suggestion – for interpreting the wave function Ψ as describing a "density of stuff" in three-dimensional, physical space – is ultimately viable. If we are going to understand the wave function Ψ as providing a complete description of some continuous, field-like ontology, *something* like Schrödinger's suggestion will be necessary, so we should try to understand the suggestion as carefully as possible.

A concrete example should help. Let's therefore consider a one-dimensional "box" in which particles can be confined – but suppose, like the box that appeared in the last chapter, this box has been split in half and the two halves have been carried to distant locations. For definiteness, suppose in particular that each half box has a width L

and the two boxes are separated by a distance d:

$$V(x) = \begin{cases} 0 & \text{for } 0 < x < L \\ 0 & \text{for } d < x < d + L \\ \infty & \text{otherwise} \end{cases} \tag{5.27}$$

Now, for example, a particle that is confined to the (half-) box on the left might have wave function

$$\psi_L(x) = \begin{cases} \sqrt{\frac{2}{L}} \sin\left(\frac{\pi x}{L}\right) & \text{for } 0 < x < L \\ 0 & \text{otherwise} \end{cases} \tag{5.28}$$

while a particle that is instead confined to the (half-) box on the right might have wave function

$$\psi_R(x) = \begin{cases} \sqrt{\frac{2}{L}} \sin\left(\frac{\pi(x-d)}{L}\right) & \text{for } d < x < L + d \\ 0 & \text{otherwise} \end{cases} . \tag{5.29}$$

And then of course it is possible that a particle might find itself *split* between the two (half-) boxes, i.e., in a superposition of being on the left and being on the right:

$$\psi_{L+R} = \frac{1}{\sqrt{2}} \left[\psi_L(x) + \psi_R(x)\right] . \tag{5.30}$$

Note that, for a single particle in these various states, Schrödinger's electric charge density acts the way one would naively expect: if the particle's quantum state is ψ_L then that particle's charge is smeared throughout (but is exclusively contained in) the left box; if the particle's quantum state is ψ_R then the particle's charge is smeared throughout (but exclusively contained in) the right box; and if the particle's quantum state is ψ_{L+R} the particle's charge is half in the left box and half in the right box.

But of course our goal here is to consider situations in which there are now *two* (or more, but two will suffice) particles involved. Suppose, for simplicity, that we have two particles with identical electric charges, q, but that the particles are distinguishable. (Then we don't need to worry about the Pauli Exclusion Principle, the symmetry/anti-symmetry properties of the two-particle wave function under exchange, etc.) For example, perhaps particle 1 (whose coordinate we call x_1) is an electron and particle 2 (whose coordinate we call x_2) is a muon.

Then let us consider three possible quantum states – call them A, B, and C – that these two particles might be in:

- In state A

$$\psi_A = \psi_L(x_1)\, \psi_R(x_2) \tag{5.31}$$

  particle 1 is definitely in the (half-) box on the left, and particle 2 is definitely in the (half-) box on the right.
- In state B

$$\psi_B = \psi_{L+R}(x_1)\,\psi_{L+R}(x_2) \tag{5.32}$$

particle 1 is in a superposition of being on the left and being on the right (it is "smeared out" evenly between the two half-boxes), and so is particle 2.

- Finally, state C

$$\psi_C = \frac{1}{\sqrt{2}}\left[\,\psi_L(x_1)\,\psi_R(x_2) \;+\; \psi_R(x_1)\,\psi_L(x_2)\,\right] \tag{5.33}$$

is an entangled superposition of (on the one hand) a state in which particle 1 is on the left and particle 2 is on the right, and (on the other hand) a state in which particle 1 is on the right and particle 2 is on the left.

These three different quantum states are sketched, in the two-dimensional configuration space, in Fig. 5.4. And it should be clear – both from the mathematical definitions of the three states and from the pictures in the Figure – that the states are indeed not all the same: there are real, measurable things that would be different in the three different cases.

For example, if we prepare a large ensemble of particle pairs in state A, and then measure their positions, we will always find particle 1 on the left and particle 2 on the right. In particular, note that we will never find the two particles in the same place! Each particle has a distinct position (the same for all elements in the ensemble) and so the positions of the two particles are perfectly correlated.

If we instead prepare a large ensemble of particle pairs in state B and then measure their positions, we will find that particle 1 is found on the left half the time and on the right half the time, and the same for particle 2, and the positions of the two particles are totally uncorrelated: 25% of the time we will find both particles on the left, 25% of the time we will find particle 1 on the left and particle 2 on the right, 25% of the

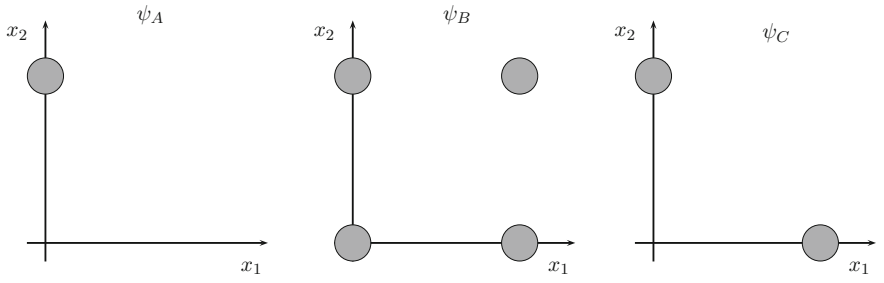


Fig. 5.4 Three different quantum states that a pair of particles might be in, schematically represented in the two-dimensional configuration space. (The *grey circles* represent regions in configuration space where the wave function is nonzero.) In the state ψ_A, particle 1 is on the left ($x_1 \approx 0$) and particle 2 is on the right ($x_2 \approx d$). In the state ψ_B, both particles are "smeared" between being on the *left* and on the *right*. State ψ_C, finally, is a superposition of "particle 1 is on the *left* and particle 2 is on the *right*" and "particle 1 is on the *right* and particle 2 is on the *left*"

time we will find particle 1 on the right and particle 2 on the left, and 25% of the time we will find both particles on the right.

And things are different yet again if we now prepare a large ensemble of particle pairs in state C and measure their positions. It is again the case (as with state B) that particle 1 is found half of the time on the left and half of the time on the right, and the same is again true for particle 2 as well, but now the positions are perfectly correlated: whenever particle 1 is found on the right, particle 2 is found on the left, and vice versa. The particles are somehow definitely in different places, even though neither particle has a definite location!

Now, for our purposes here, what is interesting about all this is the following. These three genuinely different, empirically distinguishable quantum states *all produce the exact same charge distribution* according to Eq. (5.26). That is, the "physical reality" of these three states – if Schrödinger's early idea about how to interpret the physical meaning of the wave function is correct – is the same. But this, I submit, simply cannot be correct. The three states are different – *physically* different – as proved by the fact that the outcome statistics for various kinds of measurements are different for the three states.

And so, it seems, Schrödinger's idea cannot be correct, or at least cannot be the whole story. It provides a nice way of understanding how the wave function (on configuration space) might describe some kind of material stuff in three-dimensional physical space. But, in mathematical terms, one loses a lot of information by "projecting down" the wave function Ψ into the charge density field ρ. The correspondence, that is, is many-to-one in the sense that there are very different – and meaningfully different, physically different – wave functions that correspond to the same charge density. So Schrödinger's suggestion is a nice try, and as we will see in later chapters it has a role to play in some more sophisticated candidate theories, but it apparently cannot just be the case that Schrödinger's charge density ρ *is*, all alone, the ontology of the world described by quantum mechanics.

5.5 So Then What?

The obvious follow-up question to the previous sentence would be: "OK, so if quantum mechanical wave functions aren't really, or aren't just, descriptions of something like Schrödinger's charge density ρ, what *are* they descriptions of?" You would probably therefore expect this next section to start introducing some other possible ideas, for the ontology of the quantum world, one of which maybe turns out to be right, or at least viable. But unfortunately I don't know of any other possible ideas. If there is some other way of taking quantum mechanical wave functions seriously, as somehow more or less directly describing some kind of field or fields in regular physical space, I don't know about it.

Well, actually, there is another idea on this issue that is sometimes suggested. I think it is based on a confusion, and is not viable at all, but perhaps it is worth mentioning here if only to pre-empt ongoing confusion. The idea is that the ontology

problem somehow magically goes away when we remember that our best current theory is not non-relativistic quantum mechanics but rather quantum field theory (QFT). The idea is that, perhaps unlike non-relativistic quantum mechanics (NRQM), QFT provides a straightforward and obvious and unproblematic ontology, namely, one of *fields* (in physical space).

But this is simply not true. Fields with definite configurations play exactly the same role in QFT that particles with (for example) definite positions play in non-relativistic quantum mechanics. That is, it is possible, in QFT, to write down a quantum state which can be interpreted as describing a field with a definite configuration, just as it is possible, in NRQM, to write down a quantum state (namely, a δ function) which can be interpreted as describing a particle with a definite position. But just as a generic quantum state for a single particle in NRQM will *not* attribute any particular position to the particle, so a generic quantum state for a field in QFT will *not* attribute any particular configuration to the field. That is, just as a typical quantum state for a single particle in NRQM can be understood as a superposition over a continuous infinity of different positions (think $\psi(x, t) = \int \psi(a, t)\, \delta(x - a)\, da$), so a typical quantum state for a field in QFT can be understood as a superposition over a continuous infinity of different field configurations. So unless one openly rejects the idea that quantum states provide complete state descriptions and adopts a naive (and, of course, ultimately untenable) "ignorance interpretation" of quantum states one simply cannot claim that QFT describes fields which always possess definite configurations.

There are several reasons for confusion about this. One is that, in the way that QFT is traditionally presented, one does not typically deal with generic quantum states, but instead focuses almost exclusively on (asymptotic) initial and final states corresponding to various incoming or outgoing particles in a scattering experiment, and, in a certain sense, these states can be understood in terms of fields with definite configurations. The analog in NRQM would be focusing on calculating the probability that, if a particle starts at a certain position x_0 at $t = 0$, it will be detected at some other position x_f at $t = T$. One can see why focusing exclusively on this kind of case might reinforce the belief that it is perfectly viable to think that, according to quantum mechanics, particles always have definite positions. A similar thing is happening when people get the idea that QFT is just unproblematically about fields (in physical space) with evolving, but always unproblematically definite, configurations.

A second and deeper reason for the confusion, though, is just that most people have not really thought carefully about these kinds of issues, even in the context of NRQM. Perhaps they tend to think exclusively about one-particle examples, and so have in mind an ontology of single-particle waves running around through physical space. Or perhaps they do hold some naive version of the "ignorance interpretation", according to which the ontology is something like classical (i.e., literal) particles, with wave functions providing only some kind of very incomplete description of their states. Or perhaps they don't have any particular ontological picture in mind, but are instead happy to just play games with mathematical symbols without thinking about (and without even acknowledging that *someone* should think about) what the symbols correspond to in physical reality. In any case, and whatever the ultimate reasons, most physicists have simply not appreciated or accepted that there is some

problem associated with understanding what NRQM wave functions might describe exactly – and so they are open to the (in fact rather ridiculous) suggestion that there is definitely no such problem in quantum field theory.

So where does that leave us? If, as "realist" and literal-minded physicists like Einstein and Schrödinger seem to have assumed, the idea of quantum mechanics providing *complete* descriptions of physical states means that in some sense physical reality "looks exactly like" wave functions, that view seems very difficult to maintain. The measurement problem shows that wave functions (obeying Schrödinger's equation all the time) seem unable to capture the definite outcomes that we always observe in measurements; Einstein et al. pointed out that the assumption of completeness appears to generate a direct conflict with the idea of relativistic local causality; and now we have seen that even leaving these other worries aside it is simply not clear how one might regard wave functions as directly and faithfully describing a three-dimensional physical reality at all, since wave functions are (in general) functions on an abstract, multi-dimensional configuration space.

As mentioned before, we will return to some of the ideas of this chapter when we study Everett's many-worlds version of quantum theory in Chap. 10, and also when we study the so-called "spontaneous collapse" theories in Chap. 9. We will also encounter, in Chap. 7, another candidate version of quantum theory according to which wave functions alone are *not* the whole story and which attempt to give perfectly clear descriptions of physical processes in three-dimensional space in terms of the thing that is postulated to exist *in addition to* the wave function. But before turning to those alternative theories, we will explore, in the next chapter, the so-called "Copenhagen interpretation" and try to understand better the point of view that people like Einstein and Schrödinger were reacting against when they complained that a coherent description of micro-physical reality was needed, but not provided by ordinary quantum mechanics.

Projects

5.1 Decompose the (complex) Schrödinger equation into two (real) equations by substituting in $\Psi = Re^{iS/\hbar}$ and then separating the real and imaginary parts of the resulting equation. (This will hardly be obvious, but it turns out these two resulting equations are rather interesting. One of them can be understood as expressing local conservation of probability, and the other turns out to be exactly the Hamilton–Jacobi equation from classical mechanics, but with an extra – somehow purely quantum – term in the potential energy. You might be interested to google "Hamilton–Jacobi equation" if you haven't encountered it before, and see what that is all about. But really the point of this question is just to let you practice decomposing a single equation involving a complex quantity into two equations involving two real quantities.)

5.2 Show that there exist plane-wave solutions, of Eq. (5.10), of the form $\vec{F}(\vec{x}, t) = \vec{F}_0\, e^{i\vec{k}\cdot\vec{x}-\omega t}$. What, exactly, can $\vec{F}_0$ be? (Hint: don't forget about the additional constraint imposed by Eq. (5.8).) Identify the type of polarization (linear? circular?) associated with your solution.

5.3 Take the dot product of $\vec{F}^*$ (the complex conjugate of $\vec{F}$) with Eq. (5.10). Then write down a second equation which is the dot product of $\vec{F}$ with the complex conjugate of Eq. (5.10). Now subtract your two equations and show that you get

$$\frac{\partial}{\partial t}\left(\vec{F}^* \cdot \vec{F}\right) = -ic\vec{\nabla} \cdot \left(\vec{F} \times \vec{F}^*\right). \tag{5.34}$$

This is an important result in classical electromagnetism which you may (or may not!) have seen before. Put it in more familiar terms by using $\vec{F} = \vec{E} + ic\vec{B}$ to re-express it in terms of $\vec{E}$ and $\vec{B}$, and interpret the result. (Hint: it has something to do with the Poynting vector, $\vec{S} = \frac{1}{\mu_0}\vec{E} \times \vec{B}$. and the electromagnetic field energy density, $\rho = \frac{\epsilon_0}{2}E^2 + \frac{1}{2\mu_0}B^2$.) For extra credit, what does this correspond to in regular non-relativistic quantum mechanics? (If you're not sure, you could follow the same procedure and see what happens: multiply Schrödinger's equation by Ψ^*, then multiply the complex conjugate of Schrödinger's equation by Ψ, then subtract and simplify...)

5.4 Consider the two-particle entangled state given in Eq. (5.21). Draw a configuration space cartoon showing the regions of the two-dimensional configuration space where this state has nonzero intensity.

5.5 Consider the "Einstein's Boxes" scenario from the point of view of Schrödinger's suggested interpretation of ψ in terms of charge density. In particular, suppose an electron is "split" between two well-separated half-boxes. What is its electric charge density? What is the electric charge density after the half-boxes are opened and the electron is found (complete!) in one half-box or the other? Explain, finally, how one can apply Bell's formulation of "locality" to diagnose this theoretical description as non-local.

5.6 Consider a case of measurement (described quantum mechanically) from the point of view of Schrödinger's suggested interpretation of ψ. Take the example from Chap. 3: a one-dimensional particle-in-a-box in a superposition of energy eigenstates has its energy measured and recorded in the final position of a "pointer" so that the final quantum state is the right hand side of Eq. (5.17). Calculate and sketch a picture of the electric charge density after the measurement interaction.

5.7 Maybe, instead of adding up all the one-particle charge densities from Eq. (5.25) to get the total charge density of Eq.(5.26), we should instead regard each particle's individual charge density ρ_i as physically real. (Physical reality, on this view, would not consist of just one undifferentiated total charge density, but would instead involve distinct charge densities for each particle.) Does this view survive the argument made in Sect. 5.4? In particular, are the "physical realities" implied by the three states (ψ_A, ψ_B, and ψ_C) all the same on this view?

5.8 Interview one of your other physics teachers and ask her whether she agrees with Bohr that the wave function provides a complete description of quantum systems. (She will probably say yes, but you never know.) Then ask her to elaborate by explaining, for the case of the electron in a Hydrogen atom in

the ground state, how she pictures the electron. (Presumably she will describe something like the "cloud" picture implied by taking $|\psi|^2$ as a density-of-stuff, but, again, you never know.) Finally ask her how to understand what physical reality is described by an entangled two-particle state like the one in Eq. (5.21). Summarize her viewpoint.

5.9 In a nice historical paper on "Schrödinger's route to wave mechanics" [5] Linda Wessels explains that "In the case of a single classical particle ψ could be interpreted as a wave function describing a matter wave. For a system of n classical particles, however, ψ was a function of $3n$ spatial coordinates and therefore described a wave in a $3n$-dimensional space that could not be identified with ordinary physical space. To give his theory a wave interpretation Schrödinger would either have to show how the ψ in $3n$-dimensional space determined n waves in 3-dimensional space, or reformulate the theory so that it would yield directly the required n wave functions." She then adds, citing a 1962 interview with Carl Eckart conducted by John Heilbron: "The obvious solution would be to rewrite the equations of wave mechanics so that even for a system of several 'particles', only three-dimensional wave functions would be determined. C. Eckart has reported that at one time he attempted this and remarked that *it was something that initially 'everybody' was trying to do*." (Emphasis added.) It might be a little intimidating to know that "everybody" tried something already, and nobody succeeded... but how hard can it be, really? See if you can come up with some other way or ways to convert Schrödinger's $\Psi(x_1, x_2, \ldots, x_N)$ into N "single particle" waves. Test your ideas out on examples like the one from Sect. 5.4.

5.10 In one version of the EPR-type argument we discussed in the previous chapter, Einstein pointed out that if "completeness" means a one-to-one correspondence between physically real states and quantum wave functions, then (assuming locality!) QM is not complete. One could also summarize the discussion of Sect. 5.4 by saying that there is a failure of one-to-one correspondence between physical states (as understood according to Schrödinger's suggestion) and wave functions. (Namely: the three different wave functions discussed there would all correspond to the same one physical state.) Compare and contrast these two different applications of the "one-to-one correspondence" idea. What exactly is being argued for in the two cases and how do the arguments relate?

5.11 Consider the two electrons in a diatomic Hydrogen *molecule*. What exactly do chemists and physicists think the two electrons are doing, and how does this relate to their joint wave function? For example, find a "picture" of a Hydrogen molecule in a chemistry or physics book (or online). (Or interview a physicist or chemist, and ask them to draw a Hydrogen molecule.) What exactly is pictured? Is it Schrödinger's early suggestion? Something else? Summarize and explain.

5.12 One weird idea that has been suggested, as a way of dealing with "the ontology problem", is to take the $3N$-dimensional configuration space seriously, as the fundamental physical space. (Evidently N here should be what we would

ordinarily think of as the total number of particles in the universe.) Then there is no longer any problem understanding how the wave function could directly and faithfully and completely describe what's real. On this view, physical reality would consist of something like a (complex valued) field (or maybe two coupled real-valued fields!) in this 3N-dimensional space. The three-dimensional world would then be somehow "emergent" from this more basic reality. What do you think of this idea?

5.13 It was mentioned at the end of Sect. 5.4 that one of Schrödinger's reasons for abandoning his early suggestion (about the physical meaning of Ψ) had to do with the inevitable spreading-out of wave packets we first encountered in Chap. 2. But, you might wonder, isn't (for example) the electron in a Hydrogen atom *bound*? Won't its wave function remain forever localized around the proton without spreading? So... is there really a problem with spreading, or not? (Hint: consider both the electron and the proton in the Hydrogen atom. A complete answer will involve constructing an explicit one-dimensional toy model of a Hydrogen atom and depicting the time-evolution of its quantum state in the two-dimensional configuration space.) What if the Hydrogen atom is bound to some other atoms in a molecule?

5.14 Show that, as claimed in the text, the "three genuinely different, empirically distinguishable quantum states [ψ_A, ψ_B, and ψ_C] *all produce the exact same charge distribution* according to Eq. (5.26)."

5.15 The main conclusion of Sect. 5.4 was that Schrödinger's charge density ontology cannot be correct, or at least cannot be the whole story. Maybe supplementing the charge density with some additional properties would do the trick? What sort of thing would you need to *add* to the ontology to give appropriate *physical* differences between the states described by ψ_A, ψ_B, and ψ_C?

5.16 Two particles of identical charge q are in the same length-L box, in the entangled state

$$\psi(x_1, x_2) = \frac{1}{\sqrt{2}} \left[\psi_1(x_1)\, \psi_2(x_2) + i\, \psi_3(x_1)\, \psi_1(x_2)\right] \tag{5.35}$$

where the ψ_ns are the usual particle-in-a-box energy eigenstates. According to Schrödinger's early suggestion about the physical meaning of the wave function, Eq. (5.25), what is the charge density $\rho_1(x)$ associated with particle 1? Calculate it explicitly. Will $\rho_1(x)$ change in time?

5.17 Consider the following version of (something like) the EPR argument:

> A pair of spatially-separated particles is in the entangled spin state
>
> $$\Psi = \frac{1}{\sqrt{2}} \left[\psi^1_{+z}\psi^2_{-z} - \psi^1_{-z}\psi^2_{+z}\right]. \tag{5.36}$$
>
> If one measures the spin (say, along the z-direction) of particle 1, the two particle state will, according to QM, collapse – either to $\psi^1_{+z}\psi^2_{-z}$ or to $\psi^1_{-z}\psi^2_{+z}$. That is, after and as a result of the measurement on particle 1, the spin state of particle 2 will become either ψ^2_{+z} or ψ^2_{-z}. But prior to the measurement the spin state of particle 2 was neither

of these. Therefore, the spin state of particle 2 has been (non-locally) affected by the measurement on (the distant) particle 1.

What, from the point of view of the issue raised in this chapter, is not quite right about this argument? Do you think that the not-quite-right-ness of this argument suggests that, contra EPR, QM is actually a local theory? Explain.

5.18 What is "the ontology problem"? Summarize in your own words.

References

1. R. Eisberg, R. Resnick, *Quantum Physics*, 2nd edn. (Wiley, New York, 1985)
2. K. Przibram, *Letters on Wave Mechanics*, Martin Klein, trans (Philosophical Library, NY, 1967)
3. D. Howard, Nicht Sein Kann Was Nicht Sein Darf, or the Prehistory of EPR, 1909–1935: Einstein's Early Worries about the Quantum Mechanics of Composite Systems, in *Sixty-Two Years of Uncertainty*, ed. by A.I. Miller (Plenum Press, New York, 1990)
4. G. Bacciogallupi, A. Valentini, *Quantum Theory at the Crossroads*, http://arxiv.org/pdf/quant-ph/0609184.pdf
5. L. Wessels, Schrödinger's route to wave mechanics. Stud. Hist. Philo. Sci. Part A **10**(4), 311–340 (1979)

Chapter 6
The Copenhagen Interpretation

The Copenhagen interpretation of quantum mechanics is the set of ideas, about how the theory should be understood, that was chiefly developed by Niels Bohr in collaboration with various colleagues, most notably Werner Heisenberg, in the 1920s and 1930s. Bohr's philosophy rapidly achieved the status of a kind of orthodoxy within the physics community, with early dissenters (such as Einstein and Schrödinger) being typically dismissed with charges of senility, and occasional critics from later decades (such as Bohm and Bell and Everett) being regarded practically as heretics, sinners against the true and proper nature of science. It became commonplace for proponents of the Copenhagen interpretation to insist that there was, in fact, no logically viable alternative to it at all, and authors of quantum mechanics textbooks continue, to the present day, to pay universal (if typically brief) lip service to Bohr's philosophy.

All of that said, however, the question of what, precisely, the Copenhagen interpretation *says* is surprisingly controversial. It has been joked that there are as many different versions of the Copenhagen interpretation as there are physicists who claim to follow it, and even scholars who study Bohr's writings in detail tend to come up with radically different interpretations of what he says and means. And yet, despite this unclarity, there is somehow nevertheless a fairly clear dichotomy between Bohr's actual views (whatever they were exactly) and the shallow, pragmatic version of them that students typically absorb from their textbooks and teachers.

The present chapter therefore begins with a rather long "guided tour" of several of the most relevant and important papers by Bohr and Heisenberg, so that readers can start to develop some direct acquaintance with their actual views. In the middle part of the chapter, we look at Bohr's analysis of several thought experiments that Einstein had proposed by way of criticizing the Copenhagen approach. Only then, toward the end of the chapter, will we step back and discuss (more briefly) the typical contemporary understanding of the Copenhagen interpretation, and how it is viewed by both its proponents and its critics.

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_6

6.1 Bohr's Como Lecture

Bohr's "Como Lecture" was first delivered at a celebration for Alexander Volta in Como, Italy, in the fall of 1927 and was subsequently published in *Nature* the following year [1]. Its actual title was "The Quantum Postulate and the Recent Development of Atomic Theory" and it provides an illuminating summary of Bohr's philosophical interpretation of the first several years of the development of quantum mechanics.

Bohr cuts right to the chase in the first paragraph:

> The quantum theory is characterised by the acknowledgment of a fundamental limitation in the classical physical ideas when applied to atomic phenomena. The situation thus created is of a peculiar nature, since our interpretation of the experimental material rests essentially upon the classical concepts. Notwithstanding the difficulties which hence are involved in the formulation of the quantum theory, it seems, as we shall see, that its essence may be expressed in the so-called quantum postulate, which attributes to any atomic process an essential discontinuity, or rather individuality, completely foreign to the classical theories and symbolized by Planck's quantum of action [1].

Already here we see a central theme of Bohr's Copenhagen philosophy, concerning the tension between (i) the supposed necessity of our continuing to use "the classical concepts" and (ii) the limitations of these concepts in capturing the uniquely quantum processes. Bohr elaborates in the following paragraph:

> This [quantum] postulate implies a renunciation as regards the causal space-time co-ordination of atomic processes. Indeed, our usual description of physical phenomena is based entirely on the idea that the phenomena concerned may be observed without disturbing them appreciably. This appears, for example, clearly in the theory of relativity, which has been so fruitful for the elucidation of the classical theories. As emphasised by Einstein, every observation or measurement ultimately rests on the coincidence of two independent events at the same space-time point. Just these coincidences will not be affected by any differences which the space-time co-ordination of different observers otherwise may exhibit. Now the quantum postulate implies that any observation of atomic phenomena will involve an interaction with the agency of observation not to be neglected. Accordingly, an independent reality in the ordinary physical sense can neither be ascribed to the phenomena nor to the agencies of observation. After all, the concept of observation is in so far arbitrary as it depends upon which objects are included in the system to be observed. Ultimately every observation can of course be reduced to our sense perceptions. The circumstance, however, that in interpreting observations use has always to be made of theoretical notions, entails that for every particular case it is a question of convenience at what point the concept of observation involving the quantum postulate with its inherent 'irrationality' is brought in [1].

I would summarize this by saying that, according to Bohr, our everyday, classical notions of external physical reality (for example, assigning definite "states" to various objects, talking about the causal interactions of objects in space and time, etc.) tacitly rely on the assumption that the act of observation can be taken as having no (or at least negligible) influence on the objects being observed. Whereas, in the quantum realm, "observation ... will involve an interaction with the agency of observation not to be neglected." The act of observation, in short, *disturbs* the state of the observed object in an ineliminable and unpredictable way, thus rendering it impossible to acquire knowledge of the (pre-existing, undisturbed) state of the object and, indeed,

thereby rendering talk of such "pre-existing states" empirically meaningless. We must therefore take a more holistic perspective on the broader system comprising both the "observer" and the "observed system" and recognize that any sharp division between them (such as would be implied in analyzing the interaction in terms of separate systems, each with its own well-defined state, interacting) is an arbitrary construct – one which we perhaps cannot avoid imposing in discussing and reporting our observations, but one which nevertheless in some fundamental sense distorts the real situation.

The overall line of reasoning here resonates with a general philosophical framework (that was quite popular at the time) called "positivism" (and/or sometimes the closely-related idea of "operationalism"), one of whose essential points was the idea that meaningful assertions must be *verifiable* by direct observation. To use a slightly silly and unfair example, just to try to clarify the idea, a positivist might claim that it is literally meaningless to speculate about what happens to the light inside your refrigerator when the door is closed. Since there is (or rather: assuming it was somehow the case that there is) no way to observe how light or dark it is inside the refrigerator when the door is shut (because observing this requires opening the door!), it is literally meaningless to even speculate about it, and any such speculative talk should be dismissed from rational scientific discourse as worthless and "metaphysical".

Bohr's perspective here also recalls that of the famous 18th century German philosopher, Immanuel Kant, who influentially argued that we are fundamentally cut off from true (so-called "noumenal") reality because, in effect, our minds are hard-wired to categorize the incoming sensory information in certain ways. We are thus aware, by ordinary means, only of the so-called "phenomenal" world – i.e., the world of appearances, of things-as-processed-by-us whose true natures must remain forever inaccessible. One commentator, Henry Krips, has suggested that Bohr can be understood as continuing a trend initiated by 19th century thinkers who "physiologiz[ed] the Kantian conception of observation" by locating the supposedly distorting process not in the mind, but in the physiology of perception. According to Krips,

> Bohr extended this position by proposing that the 'external procedures' that affect the forms of sensible intuition include the processes of observation themselves. Thus Bohr stood at the end of a long historical trajectory: Kant conceived the apparatus of observation as an inner mental faculty, analogous to a pair of spectacles that mediated and in particular gave form to and interpreted raw sense impressions. Neo-Kantians projected the interpretative aspect of vision outwards, reconceiving it as a bodily, and specifically physiological process. Bohr took this further by including observation as [affecting] not merely what we see but also the terms in which we describe it [2].

In any case, though, and whatever the historical precedents, for Bohr the fundamental lesson of the quantum theory was that there is a kind of inherent "graininess" and unpredictability to interactions, including the interactions between "external object and "observer" (or "measuring apparatus"). Such interactions supposedly imply a finite, non-negligible, and uncontrollable *disturbance*, of the "external object", whenever we try to observe it. And so we are cut off from the possibility of scientifically meaningful descriptions of the microscopic world, for just the same reasons that (in

the silly example of the last paragraph) we are cut off from scientifically meaningful descriptions of the state of illumination inside a closed refrigerator: the very act of trying to *verify* any hypothesis about the state of the object in question, *disturbs* its state and thereby undercuts the attempted verification. This is why Bohr speaks of a "*renunciation*" of the applicability of our classical concepts.

Bohr continues to explain that

> This situation has far-reaching consequences. On one hand, the definition of the state of a physical system, as ordinarily understood, claims the elimination of all external disturbances. But in that case, according to the quantum postulate, any observation will be impossible, and, above all, the concepts of space and time lose their immediate sense. On the other hand, if in order to make observation possible we permit certain interactions with suitable agencies of measurement, not belonging to the system, an unambiguous definition of the state of the system is naturally no longer possible, and there can be no question of causality in the ordinary sense of the word. The very nature of the quantum theory thus forces us to regard the space-time co-ordination and the claim of causality, the union of which characterises the classical theories, as complementary but exclusive features of the description, symbolising the idealisation of observation and definition respectively. Just as the relativity theory has taught us that the convenience of distinguishing sharply between space and time rests solely on the smallness of the velocities ordinarily met with compared to the velocity of light, we learn from the quantum theory that the appropriateness of our usual causal space-time description depends entirely upon the small value of the quantum of action as compared to the actions involved in ordinary sense perceptions. Indeed, in the description of atomic phenomena, the quantum postulate presents us with the task of developing a 'complementarity' theory the consistency of which can be judged only by weighing the possibilities of definition and observation [1].

Here we first encounter Bohr's fundamental concept of "complementarity". In this paragraph, he describes, as "complementary", the causal and space-time perspectives on events. His point is that, whereas in the context of classical physics it is taken for granted that both perspectives are simultaneously applicable and indeed classical descriptions by definition provide causal accounts of spatio-temporal events, the two perspectives are mutually exclusive in the quantum realm. Meaningful attribution of precise spatial and temporal coordinates to events requires, for example, careful position measurements. But such measurements, as we have seen, imply physical interactions which disrupt the causal processes which might otherwise, in the absence of said interactions, have been taking place.

For Bohr, the "causal" description means one taking account of energy and momentum conservation. So the "complementarity" between space-time and causal descriptions arises specifically from the fact that position measurements imply an interaction involving unpredictable momentum exchange between the system in question and the position-measuring apparatus. One thus sees an intimate connection (which we will continue to explore as this chapter develops) between Bohr's view that "space-time" and "causal" descriptions are mutually exclusive, and Heisenberg's important discovery that position and momentum (as well as time and energy) jointly obey an uncertainty (or indeterminacy) principle.

Bohr saw a similar sort of complementarity between the wave and particle pictures of light (and, subsequently, electrons), controversy about which had given rise to quantum mechanics in the first place:

> This view is already clearly brought out by the much-discussed question of the nature of light and the ultimate constituents of matter. As regards light, its propagation in space and time is adequately expressed by the electromagnetic theory. Especially the interference phenomena *in vacuo* and the optical properties of material media are completely governed by the wave theory superposition principle. Nevertheless, the conservation of energy and momentum during the interaction between radiation and matter, as evident in the photoelectric and Compton effect, finds its adequate expression just in the light quantum idea put forward by Einstein. As is well known, the doubts regarding the validity of the superposition principle on the one hand and of the conservation laws on the other, which were suggested by this apparent contradiction, have been definitely disproved through direct experiments. This situation would seem clearly to indicate the impossibility of a causal space-time description of the light phenomena. On one hand, in attempting to trace the laws of the time-spatial propagation of light according to the quantum postulate, we are confined to statistical considerations. On the other hand, the fulfilment of the claim of causality for the individual light processes, characterised by the quantum of action, entails a renunciation as regards the space-time description. Of course, there can be no question of a quite independent application of the ideas of space and time and of causality. The two views of the nature of light are rather to be considered as different attempts at an interpretation of experimental evidence in which the limitation of the classical concepts is expressed in complementary ways.
>
> The problem of the nature of the constituents of matter presents us with an analogous situation. The individuality of the elementary electrical corpuscles is forced upon us by general evidence. Nevertheless, recent experience, above all the discovery of the selective reflection of electrons from metal crystals, requires the use of the wave theory superposition principle in accordance with the ideas of L. de Broglie. Just as in the case of light, we have consequently in the question of the nature of matter, so far as we adhere to classical concepts, to face an inevitable dilemma, which has to be regarded as the very expression of experimental evidence. In fact, here again we are not dealing with contradictory but with complementary pictures of the phenomena, which only together offer a natural generalisation of the classical mode of description. In the discussion of these questions, it must be kept in mind that, according to the view taken above, radiation in free space as well as isolated material particles are abstractions, their properties on the quantum theory being definable and observable only through their interaction with other systems. Nevertheless, these abstractions are, as we shall see, indispensable for a description of experimental evidence in connexion with our ordinary space-time view [1].

Here Bohr stresses that each side of the wave-particle duality has a secure foundation in experimental evidence: for light, for example, the continuous space-time propagation as described by Maxwell's equations is required to account for interference phenomena, whereas Einstein's "light quantum" (i.e., "light particle" or "photon") picture is necessary to account for phenomena such as the photoelectric effect and Compton scattering. Bohr's view seems to be that if we take either picture too seriously – i.e., if we take either picture as capturing, fully and finally, the true physical nature of light – we would have a clear contradiction with some aspect of the experimental evidence which can only be described by the alternative picture. So, for Bohr, we must not take either picture fully seriously: the contradiction is merely "apparent". Yet, simultaneously, we must take both pictures quite seriously, in the sense that only together do they allow us to understand the totality of experimental evidence. The two pictures, that is, are mutually exclusive (in the sense that, taken as capturing the full truth, they contradict one another) and yet jointly exhaustive (in the sense that we need both, together, to capture all aspects of the phenomena revealed by observation).

It is interesting, here, to compare Bohr's view with another possible interpretation of the wave-particle duality. Einstein, for example, considered (during this same period) a "pilot-wave" model of photons, in which the wave-particle "duality" is taken quite literally: each individual "photon" in this model consists of a literal point particle (carrying the energy) which is guided (or piloted) by a surrounding wave obeying Maxwell's equations. We will explore this type of model further (but for massive particles rather than photons) in Chap. 7. For now, the point is just that there seem to exist various ways that one might consider really reconciling – *unifying* – the aspects that Bohr considers "complementary". Doing this of course requires modifying the classical concepts. For example, in this pilot-wave model of photons, there is still a wave obeying Maxwell's equations, but its role is completely different – instead of actually being the seat of light's energy and momentum, it is a kind of behind-the-scenes "ghost", pushing and pulling the associated photon particle.[1] And similarly, although there is a particle with a definite trajectory, it does not obey the familiar dynamical laws of Newtonian mechanics, but instead something completely novel which gives rise to all sorts of unexpected and surprising motions (e.g., when a photon reflects from a mirror, the particle stops and sits still for some time some distance in front of the mirror!). In any case, this example illustrates, I think, the attitude that people like Einstein and Schrödinger had toward the "apparent contradiction" Bohr mentions here. They, like Bohr, saw the conflicts as pointing to inadequacies in the existing models. But they took for granted that it should be possible to build new theories – new pictures of microscopic reality with new associated dynamical laws – that would unify and explain *all* available experimental evidence.

But Bohr would have none of this. For Bohr, the classical models may be "abstractions" (whose domain of applicability we stretch when we use them to describe the microscopic world), but they are *necessary* – almost "hard-wired" in a kind of Kantian sense – abstractions that can not and/or should not be abandoned, modified, or replaced. For Bohr, the quantum theory was not so much an attempt to accurately describe microscopic reality (this being supposedly impossible, for the philosophical reasons we have been sketching) but was rather a formal and precise mathematical scheme to referee disputes between complementary (i.e., individually inadequate but still jointly necessary) perspectives.

Here again the Heisenberg uncertainty relations are crucial and central. As Bohr explains,

> ...in the classical theories any succeeding observation permits a prediction of future events with ever-increasing accuracy, because it improves our knowledge of the initial state of the system. According to the quantum theory, just the impossibility of neglecting the interaction with the agency of measurement means that every observation introduces a new uncontrollable element. Indeed, it follows from the above considerations that the measurement of the positional coordinates of a particle is accompanied not only by a finite change in the dynamical variables, but also the fixation of its position means a complete rupture in the causal description of its dynamical behaviour, while the determination of its momentum always implies a gap in the knowledge of its spatial propagation. Just this situation brings

[1] Einstein literally called it a "Gespensterfeld", ghost-field.

> out most strikingly the complementary character of the description of atomic phenomena which appears as an inevitable consequence of the contrast between the quantum postulate and the distinction between object and agency of measurement, inherent in our very idea of observation [1].

We will hear more about the connection between the uncertainty principle and the Copenhagen interpretation from Heisenberg himself, in the following section.

Before turning to that, however, here is one last excerpt from Bohr's Como lecture, in which he discusses Schrödinger's wave mechanics and echoes some of the issues we reviewed in the previous chapter:

> ...Schrödinger has expressed the hope that the development of the wave theory will eventually remove the irrational element expressed by the quantum postulate and open the way for a complete description of atomic phenomena along the line of the classical theories. In support of this view, Schrödinger, in a recent paper (...) emphasises the fact that the discontinuous exchange of energy between atoms required by the quantum postulate, from the point of view of the wave theory, is replaced by a simple resonance phenomenon. In particular, the idea of individual stationary states would be an illusion and its applicability only an illustration of the resonance mentioned. It must be kept in mind, however, that just in the resonance problem mentioned we are concerned with a closed system which, according to the view presented here, is not accessible to observation. In fact, wave mechanics ... represents a symbolic transcription of the problem of motion of classical mechanics adapted to the requirements of quantum theory and only to be interpreted by an explicit use of the quantum postulate. ....
>
> The symbolical character of Schrödinger's method appears not only from the circumstance that its simplicity ... depends essentially upon the use of imaginary arithmetic quantities. But above all there can be no question of an immediate connexion with our ordinary conceptions because the 'geometrical' problem represented by the wave equation is associated with the so-called co-ordinate [i.e., configuration] space, the number of dimensions of which is equal to the number of degrees of freedom of the system, and hence in general greater than the number of dimensions of ordinary space. Further, Schrödinger's formulation of the interaction problem ... involves a neglect of the finite velocity of propagation of the forces claimed by relativity theory.
>
> On the whole, it would scarcely seem justifiable, in the case of the interaction problem, to demand a visualisation by means of ordinary space-time pictures. In fact, all our knowledge concerning the internal properties of atoms is derived from experiments on their radiation or collision reactions, such that the interpretation of experimental facts ultimately depends on the abstractions of radiation in free space, and free material particles. Hence, our whole space-time view of physical phenomena, as well as the definition of energy and momentum, depends ultimately upon these abstractions. In judging the applications of these auxiliary ideas we should only demand inner consistency, in which connexion special regard has to be paid to the possibilities of definition and observation [1].

Bohr's description of Schrödinger's waves as "symbolical" – on (largely) the grounds that, as waves in *configuration* space, the wave functions clearly cannot be taken seriously as physically real – is particularly interesting. It should be becoming clear that, whatever exactly Bohr and his colleagues may have meant when they made claims implying the *completeness* of the quantum mechanical description, it was not exactly the same kind of thing that Einstein and Schrödinger meant by this same word, and

against which their objections were made.[2] For Bohr and the other Copenhagenists, the completeness of quantum mechanics did not mean that the theory provides a literal and direct and exhaustive description of the physical states of external objects. Indeed, as we have seen, for Bohr, the essential lesson of quantum theory is precisely that such an exhaustive description is, for supposedly deep philosophical reasons, impossible to achieve and thus inappropriate to seek. For Bohr, "completeness" is used instead in an epistemological or semantic sense (rather than the realist or descriptive sense we have largely assumed in earlier chapters) – something less along the lines of "no aspect of objective reality has been missed" and instead more along the lines of "you can't rationally ask for anything more (than this formal refereeing between the complementary classical concepts) without lapsing into meaningless, unscientific, metaphysical talk".

This point of view will become somewhat clearer when we review Bohr's analysis of some concrete examples. But first let's consider the Copenhagen interpretation as explained by its second-most-important proponent, Werner Heisenberg.

6.2 Heisenberg

In Heisenberg's writings, one finds an overall agreement with the perspectives of Bohr. But Heisenberg is a little simpler and a little more practical – a little less careful and a lot less philosophically grandiose – in his way of expressing himself. Let us begin here by giving the overall flavor of Heisenberg's style by quoting, "rapid-fire," some excerpts from his essay on "The History of Quantum Theory" [4]:

- "... from this time on ... the physicists learned to ask the right questions.... What were these questions? Practically all of them had to do with the strange apparent contradictions between the results of different experiments. How could it be that the same radiation that produces interference patterns, and therefore must consist of waves, also produces the photoelectric effect, and therefore must consist of particles? How could it be that the frequency of the orbital motion of the electron in the atom does not show up in the frequency of the emitted radiation? .... Again and again one found that the attempts to describe atomic events in the traditional terms of physics led to contradictions."
- "Gradually, during the early twenties, the physicists became accustomed to these difficulties, they acquired a certain vague knowledge about where trouble would occur, and they learned to avoid contradictions. .... This was not sufficient to form a consistent general picture of what happens in a quantum process, but it changed the minds of the physicists in such a way that they somehow got into the spirit of quantum theory."

[2]The "completeness" claim was made at least as early as 1927 when Max Born and Heisenberg declared: "We maintain that quantum mechanics is a complete theory; its basic physical and mathematical hypotheses are not further susceptible of modifications" [3].

- "The strangest experience of those years was that the paradoxes of quantum theory did not disappear during this process of clarification; on the contrary, they became even more marked and more exciting."
- "The two experiments – one on the interference of scattered light and the other on the change of frequency of the scattered light – seemed to contradict each other without any possibility of compromise."
- "But in what sense did the new formalism describe the atom? The paradoxes of the dualism between wave picture and particle picture were not solved; they were hidden somehow in the mathematical scheme."
- "The electromagnetic waves were interpreted not as 'real' waves but as probability waves, the intensity of which determines in every point the probability for the absorption ... of a light quantum by an atom at this point."
- "The probability wave ... meant a tendency for something. It was a quantitative version of the old concept of 'potentia' in Aristotelian philosophy. It introduced something standing in the middle between the idea of an event and the actual event, a strange kind of physical reality just in the middle between possibility and reality."
- "Bohr considered the two pictures – particle picture and wave picture – as two complementary descriptions of the same reality. Any of these descriptions can be only partially true, there must be limitations to the use of the particle concept as well as of the wave concept, else one could not avoid contradictions. If one takes into account those limitations which can be expressed by the uncertainty relations, the contradictions disappear."

Leaving aside their very different manners of expression, the biggest difference between the views of Heisenberg and Bohr is probably on this point of the wave function (i.e., "probability wave") representing some kind of at least half- or proto-real thing. Heisenberg's view here is much closer to the spirit of the view that, for example, Einstein had criticized – as seemingly in conflict with the principle of locality – in his "boxes" type arguments. We will return to this point later in the chapter.

For now, let us turn to one of Heisenberg's more careful attempts to articulate the Copenhagen philosophy. The first paragraph of his essay (written later, in the 1950s) on "The Copenhagen Interpretation of Quantum Theory" [5] contains a nice summary of the ideas we reviewed in the previous section:

> The Copenhagen interpretation of quantum theory starts from a paradox. Any experiment in physics, whether it refers to the phenomena of daily life or to atomic events, is to be described in the terms of classical physics. The concepts of classical physics form the language by which we describe the arrangement of our experiments and state the results. We cannot and should not replace these concepts by any others. Still the application of these concepts is limited by the relations of uncertainty. We must keep in mind this limited range of applicability of the classical concepts while using them, but we cannot and should not try to improve them [5].

The last sentence there, to me, captures the essence of the Copenhagen interpretation: because we must continue to use the classical concepts, even while acknowledging their limitations, we must in some deep sense renounce the goal of attempting to

understand and explain quantum phenomena in the clear, consistent, unified, literal way that had always been aimed at in "classical" physics.

Heisenberg elaborates his view about the nature of the quantum mechanical wave function later in the same article:

> ...it is useful to compare the procedure for the theoretical interpretation of an experiment in classical physics and in quantum theory. In Newton's mechanics, for instance, we may start by measuring the position and the velocity of the planet whose motion we are going to study. The result of the observation is translated into mathematics by deriving numbers for the co-ordinates and the momenta of the planet from the observation. Then the equations of motion are used to derive from these values of the co-ordinates and momenta at a given time the values of these co-ordinates ... at a later time, and in this way the astronomer can predict the properties of the system at a later time. He can, for instance, predict the exact time for an eclipse of the moon.
>
> In quantum theory the procedure is slightly different. We could for instance be interested in the motion of an electron through a cloud chamber and could determine by some kind of observation the initial position and velocity of the electron. But this determination will not be accurate; it will at least contain the inaccuracies following from the uncertainty relations and will probably contain still larger errors due to the difficulty of the experiment. It is the first of these inaccuracies which allows us to translate the result of the observation into the mathematical scheme of quantum theory. A probability function is written down which represents the experimental situation at the time of the measurement, including even the possible errors of the measurement.
>
> This probability function represents a mixture of two things, partly a fact and partly our knowledge of a fact. It represents a fact in so far as it assigns at the initial time the probability unity (i.e., complete certainty) to the initial situation: the electron moving with the observed velocity at the observed position; 'observed' means observed within the accuracy of the experiment. It represents our knowledge in so far as another observer could perhaps know the position of the electron more accurately. The error in the experiment does – at least to some extent – not represent a property of the electron but a deficiency in our knowledge of the electron. Also this deficiency of knowledge is expressed in the probability function.
>
> In classical physics one should in a careful investigation also consider the error of the observation. As a result one would get a probability distribution for the initial values of the co-ordinates and velocities and therefore something very similar to the probability function in quantum mechanics. Only the necessary uncertainty due to the uncertainty relations is lacking in classical physics.
>
> When the probability function in quantum theory has been determined at the initial time from the observation, one can from the laws of quantum theory calculate the probability function at any later time and can thereby determine the probability for a measurement giving a specified value of the measured quantity. We can, for instance, predict the probability for finding the electron at a later time at a given point in the cloud chamber. It should be emphasized, however, that the probability function does not in itself represent a course of events in the course of time. It represents a tendency for events and our knowledge of events. The probability function can be connected with reality only if one essential condition is fulfilled: if a new measurement is made to determine a certain property of the system. Only then does the probability function allow us to calculate the probable result of the new measurement. The result of the measurement again will be stated in terms of classical physics.
>
> Therefore, the theoretical interpretation of an experiment requires three distinct steps: (1) the translation of the initial experimental situation into a probability function; (2) the following up of this function in the course of time; (3) the statement of a new measurement to be made of the system, the result of which can then be calculated from the probability function. For the first step the fulfillment of the uncertainty relations is a necessary condition. The second

> step cannot be described in terms of the classical concepts; there is no description of what happens to the system between the initial observation and the next measurement. It is only in the third step that we change over again from the 'possible' to the 'actual' [5].

This passage raises a number of questions about how Heisenberg's views relate to Bohr's views as well as to the worries discussed in previous chapters. I'll invite you to think about some of these issues in the Projects.

Heisenberg's positivist philosophy is also on display in this essay. For example, in discussing the idea of electrons orbiting nuclei in atoms, he remarks: "one can never observe more than one point in the orbit of the electron; therefore, there is no orbit in the ordinary sense" [5]. What the electron does between observations is thus dismissed not merely as unknowable (and thus not meaningful to speak of) but as altogether non-existent. Indeed, this kind of inference – from unknowability to unreality – pushes beyond mere positivism and recalls the idealist philosophy of, for example, Bishop George Berkeley, who famously decreed "esse est percipi" – "to be, is to be perceived". The extent to which this sort of anti-realism, about (at least) the microscopic quantum realm, should be considered an official part of the Copenhagen doctrine, is one of those controversial issues about which there is no real consensus.

Heisenberg's continuing elaboration provides an illuminating perspective on Bohr's concept of "complementarity":

> Actually we need not speak of particles at all. For many experiments it is more convenient to speak of matter waves; for instance, of stationary matter waves around the atomic nucleus. Such a description would directly contradict the other description if one does not pay attention to the limitations given by the uncertainty relations. Through the limitations the contradiction is avoided. The use of 'matter waves' is convenient, for example, when dealing with the radiation emitted by the atom. By means of its frequencies and intensities the radiation gives information about the oscillating charge distribution in the atom, and there the wave picture comes much nearer to the truth than the particle picture. Therefore, Bohr advocated the use of both pictures, which he called 'complementary' to each other. The two pictures are of course mutually exclusive, because a certain thing cannot at the same time be a particle (i.e., substance confined to a very small volume) and a wave (i.e., a field spread out over a large space), but the two complement each other. By playing with both pictures, by going from the one picture to the other and back again, we finally get the right impression of the strange kind of reality behind our atomic experiments [5].

Once again, not only in the style of the writing, but also in some of the content of his remarks, one senses that Heisenberg's understanding of "complementarity" is a little more easy-going and pragmatic than Bohr's. For example, one doubts that Bohr would agree with Heisenberg's statement that the 'matter wave' picture "comes much nearer to the truth" in its description of the electrons orbiting a nucleus in an atom. This kind of (perhaps inadvertent) concession to the existence of some "real truth" about such things leaves Heisenberg, I think, much more open to the kinds of criticisms we reviewed in the last few chapters. Whereas Bohr's dense prose functions more effectively as an impenetrable barrier against such attacks.

Heisenberg returns to the theme of anti-realism (about unobserved microscopic phenomena) in his comments on the double-slit experiment:

> We assume that a small source of monochromatic light radiates toward a black screen with two small holes in it. The diameter of the holes may be not much bigger than the wave length of the light, but their [separation] will be very much bigger. At some distance behind the screen a photographic plate registers the incident light. If one describes this experiment in terms of the wave picture, one says that the primary wave penetrates through the two holes; there will be secondary spherical waves starting from the holes that interfere with one another, and the interference will produce a pattern of varying intensity on the photographic plate.
>
> The blackening of the photographic plate is a quantum process, a chemical reaction produced by a single light quanta. Therefore, it must also be possible to describe the experiment in terms of light quanta. If it would be permissible to say what happens to the single light quantum between its emission from the light source and its absorption in the photographic plate, one could argue as follows: The single light quantum can come through the first hole or through the second one. If it goes through the first hole and is scattered there, its probability for being absorbed at a certain point of the photographic plate cannot depend upon whether the second hole is closed or open. The probability distribution on the plate will be the same as if only the first hole was open. If the experiment is repeated many times and one takes together all cases in which the light quantum has gone through the first hole, the blackening of the plate due to these cases will correspond to this probability distribution. If one considers only those light quanta that go through the second hole, the blackening should correspond to a probability distribution derived from the assumption that only the second hole is open. The total blackening, therefore, should just be the sum of the blackenings in the two cases; in other words, there should be no interference pattern. But we know this is not correct, and the experiment will show the interference pattern. Therefore, the statement that any light quantum must have gone *either* through the first *or* through the second hole is problematic and leads to contradictions. This example shows clearly that the concept of the probability function does not allow a description of what happens between two observations. Any attempt to find such a description would lead to contradictions; this must mean that the term 'happens' is restricted to the observation [5].

One may have questions about why, in the case of an electron in an atom, "the wave picture comes much nearer to the truth", whereas in the case of a particle traversing a double-slit apparatus the wave does not in any sense provide a realistic "description of what happens". And of course one may also have philosophical concerns about the idea that nothing happens beyond that which is observed. Here I will just point out that, in addition to such philosophical concerns, one might also have (to use Bell's terminology from Chap. 3) "professional" concerns about this idea: if, according to quantum mechanics, physical reality (what "happens") is restricted, somehow, to observation, shouldn't we insist on a sharp definition of "observation", i.e., a clear discrimination between those interactions which do, and those which do not, count as "observations" and thereby give rise to real physical "happenings"? Otherwise the theory's account of what is real would necessarily remain "unprofessionally vague and ambiguous". But of course, such a concern presupposes something that Heisenberg and Bohr apparently did not accept – namely, that it is the proper goal of a physical theory to provide a clear and unambiguous account of what is real.

In terms of the quantum mechanical formalism, the question about the precise meaning of "observation" becomes the question of how to understand, and when precisely to apply, the postulate of wave function collapse. About this Heisenberg writes:

> The observation itself changes the probability function discontinuously; it selects of all possible events the actual one that has taken place. Since through the observation our knowledge of the system has changed discontinuously, its mathematical representation also has undergone the discontinuous change and we speak of a 'quantum jump'.
>
> Therefore, the transition from the 'possible' to the 'actual' takes place during the act of observation. If we want to describe what happens in an atomic event, we have to realize that the word 'happens' can apply only to the observation, not to the state of affairs between two observations. It applies to the physical, not the psychical act of observation, and we may say that the transition from the 'possible' to the 'actual' takes place as soon as the interaction of the object with the measuring device, and thereby with the rest of the world has come into play; it is not connected with the act of registration of the result by the mind of the observer. The discontinuous change in the probability function, however, takes place with the act of registration, because it is the discontinuous change of our knowledge in the instant of registration that has its image in the discontinuous change of the probability function [5].

One can see here, again, how Heisenberg's formulations invite some of the objections we have discussed previously. For example, if the change in the quantum state (induced by observation) merely represents a change in our knowledge of the system, doesn't that imply that the observation is simply revealing a fact about the observed system which was perfectly definite (though unknown) prior to the observation, such that the (earlier) quantum mechanical description was simply incomplete?

But on the other hand, we also begin to appreciate the very different underlying philosophical perspective that immunized Bohr and Heisenberg against such objections: if, for example, reference to unknown or unobserved elements of physical reality is literally meaningless, then the clean division between the epistemic and ontological interpretations of wave function collapse dissolves and the incompleteness objection loses its force.

Heisenberg continues, addressing (what would later become) Bell's objection that the vagueness and arbitrariness of the division of the world implied by the distinction between "observation processes" and "regular processes":

> It has been said that we always start with a division of the world into an object, which we are going to study, and the rest of the world, and that this division is to some extent arbitrary. It should indeed not make any difference in the final result if we, e.g., add some part of the measuring device or the whole device to the object and apply the laws of quantum theory to this more complicated object. It can be shown that such an alteration of the theoretical treatment would not alter the predictions concerning a given experiment. This follows mathematically from the fact that the laws of quantum theory are for the phenomena in which Planck's constant can be considered as a very small quantity, approximately identical with the classical laws. But it would be a mistake to believe that this application of the quantum theoretical laws to the measuring device could help to avoid the fundamental paradox of quantum theory.
>
> The measuring device deserves this name only if it is in close contact with the rest of the world, if there is an interaction between the device and the observer. Therefore, the uncertainty with respect to the microscopic behavior of the world will enter into the quantum-theoretical system here just as well as in the first interpretation. If the measuring device would be isolated from the rest of the world, it would be neither a measuring device nor could it be described in the terms of classical physics at all. ....
>
> Certainly quantum theory does not contain genuine subjective features, it does not introduce the mind of the physicist as a part of the atomic event. But it starts from the division of the

> world into the 'object' and the rest of the world, and from the fact that at least for the rest of the world we use the classical concepts in our description. This division is arbitrary and historically a direct consequence of our scientific method; the use of the classical concepts is finally a consequence of the general human way of thinking. But this is already a reference to ourselves and ... so ... our description is not completely objective [5].

Heisenberg thus rather explicitly acknowledges the criticisms of the theory that started our discussion of the measurement problem in Chap. 3: the theory indeed "starts from the division of the world into the 'object' and the rest of the world", this "division is arbitrary", and the systems on the two sides of the division are to be described in radically different theoretical terms.

But, again, on the other hand, it also becomes increasingly clear that Heisenberg does not see any of this as some sort of fatal flaw in the way that the critics (Schrödinger, Einstein, Bell, etc.) did. For Heisenberg, the theory is simply not an attempt to provide a literal, realistic description of the world. Its structure – and in particular the fact that it necessarily divides the world into two realms which are described very differently – should instead be understood as having merely an epistemological significance, growing out of the nature of "our scientific method" and "the general human way of thinking". The theory, in short, should be understood less as an attempt to provide an objective description of nature, and more as a kind of practical algorithm (with few if any ontological commitments) for making empirical predictions.

From Heisenberg's point of view, then, the criticisms of the critics are largely misplaced – even though, he would have to admit, the theory does fail to provide the kind of literal, direct description of physical processes that the critics ultimately wanted. According to the Copenhagen philosophy, however, this is no kind of deficiency in the quantum theory. Instead, from the point of view of Bohr and Heisenberg, the flaw lies in the misplaced demands of the critics: what they want, according to the Copenhagen point of view, is unattainable and indeed at odds with the nature of human scientific knowledge, so they are simply wrong to want it.

Let's give Heisenberg the final word:

> ...it has sometimes been suggested that one should depart from the classical concepts altogether and that a radical change in the concepts used for describing the experiments might possibly lead back to a nonstatical, completely objective description of nature.
>
> This suggestion, however, rests upon a misunderstanding. The concepts of classical physics are just a refinement of the concepts of daily life and are an essential part of the language which forms the basis of all natural science. Our actual situation in science is such that we *do* use the classical concepts for the description of the experiments, and it was the problem of quantum theory to find theoretical interpretation of the experiments on that basis. There is no use in discussing what could be done if we were other beings than we are [5].

6.3 Bohr on Einstein's Diffraction Example

In the last two sections, we've attempted to give a broad philosophical overview of the Copenhagen interpretation, closely grounded in the writings of Bohr and Heisenberg. We discussed, again in very abstract terms, some of the ways in which their views

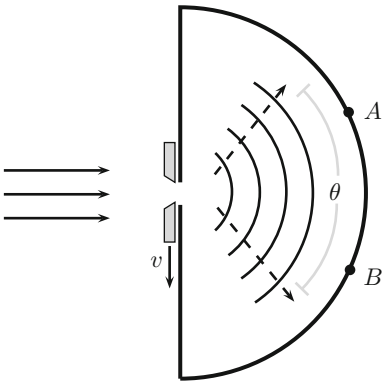


Fig. 6.1 Updated version of Bohr's illustration of Einstein's diffraction example (compare to the earlier Fig. 4.1). The diffracted wave has an angular spread θ implying appreciable probability for the particle to localize at many different points including A and B. Einstein's argument was that the (anti-) correlation between seeing the particle at A and seeing it at B implied either a kind of spooky action at a distance, or that the particle had a definite location all along (such that the description in terms of a diffracting wave was incomplete). In his version of the setup, Bohr also includes a moveable aperture (the gray trapezoids) which can be slid in place as shown – so there is a single slit of width a – or moved down at speed v to block the slit

seem to relate to the sorts of criticisms we reviewed in the previous three chapters. But Bohr, especially, engaged very directly with the critics, especially Einstein, on several example scenarios where their different points of view can be seen to clash in much more concrete terms. In this section (and the two following ones) we will thus turn to further elucidating the Copenhagen interpretation in the context of a few of these concrete examples.

We begin with Bohr's discussion of the example Einstein raised at the 1927 Solvay conference (which we discussed in Chap. 4 in the "Einstein's Boxes" section). Bohr discusses this in his beautifully written and rightly famous 1949 reminiscence, "Discussion with Einstein on Epistemological Problems in Atomic Physics" [6].

Bohr begins his discussion by summarizing Einstein's example as follows:

> To illustrate his attitude, Einstein referred ... to the simple example illustrated by [Fig. 6.1], of a particle (electron or photon) penetrating through a hole or a narrow slit in a diaphragm placed at some distance before a photographic plate. On account of the diffraction of the wave connected with the motion of the particle and indicated in the figure by the thin lines, it is under such conditions not possible to predict with certainty at what point the electron will arrive at the photographic plate, but only to calculate the probability that, in an experiment, the electron will be found within any given region of the plate. The apparent difficulty, in this description, which Einstein felt so acutely, is the fact that, if in the experiment the electron is recorded at one point A ... then it is out of the question of ever observing an effect of this electron at another point (B), although the laws of ordinary wave propagation offer no room for a correlation between two such events.
>
> Einstein's attitude gave rise to ardent discussions.... [which] centered on the question of whether the quantum-mechanical description exhausted the possibilities of accounting for observable phenomena or, as Einstein maintained, the analysis could be carried further and, especially, of whether a fuller description of the phenomena could be obtained by bringing into consideration the detailed balance of energy and momentum in individual processes [6].

The first paragraph seems like a perfectly good summary of Einstein's arguments (although the role of "locality" is perhaps not adequately stressed). But Bohr doesn't seem to have understood Einstein's argument (that, if one assumes locality, a full description of the physical state of the system must include more facts than are contained in the quantum description) as the primary issue here. Instead, Bohr focuses on analyzing the suggestion that, by monitoring the recoil of the diaphragm, one might improve one's ability to predict where the particle might eventually be detected: for example, assuming the incident particle and the diaphragm have no initial vertical momentum, then if (prior to the particle's arrival at the screen) one observes that the diaphragm has acquired (say) a *downward* momentum, it must be (assuming momentum conservation) that the particle has deflected *upward*, toward (say) point A rather than point B.

Recall from Chap. 4 that, according to Einstein, the application of the locality concept to this example requires that one "not describe the process solely by the Schrödinger wave, but that at the same time one localises the particle during the propagation." What Einstein meant to be arguing for, that is, is the claim that the particle *has* a definite location (say, near A or near B) even before it is observed. For Einstein, this reality claim would stand independently of whether or not the pre-measurement location of the particle could be (in some indirect sense, as for example by monitoring the recoil of the diaphragm) determined, and, indeed, whether or not the inclusion of the particle trajectory in one's theoretical description would change the operational predictions. For Bohr, though, the idea of a physical reality which is unobservable and/or irrelevant to theoretical predictions is a kind of contradiction in terms. So it makes sense to some degree that Bohr interpreted Einstein as arguing that it should be possible to improve (beyond what is allowed by ordinary quantum theory) one's practical ability to predict where the particle will hit the screen. This explains why Bohr's analysis of Einstein's diffraction example focuses on defending the self-consistency of the limitations placed on the theory's predictive accuracy by the Heisenberg uncertainty formulas.

That analysis proceeds as follows. Suppose the incoming particle has momentum $p = h/\lambda$ and the slit (when open) has width a. Then the particle will acquire, assuming it passes the slit, an uncertainty in its transverse position $\Delta q \approx a$. Then the standard relation for the angular width of a diffraction pattern ($\theta \approx \lambda/a$) implies that there is an uncertainty in the transverse component of the particle's momentum of order

$$\Delta p \approx \theta \cdot p \approx \frac{h}{\Delta q} \tag{6.1}$$

which is just the usual Heisenberg uncertainty relation. As Bohr notes: "This result could, of course, also be obtained directly by noticing that, due to the limited extension of the wave-field at the place of the slit, the component of the wave-number parallel to the plane of the diaphragm will involve a latitude $\Delta k \approx (1/a) \approx (1/\Delta q)$."

Now, if we suppose that the shutter opens the width-a slit only for a time Δt, the wave packet will have a spread of frequencies of width

$$\Delta\nu \approx \frac{1}{\Delta t} \tag{6.2}$$

which then implies, using the usual quantum energy-frequency relation $E = h\nu$, an uncertainty in the particle's energy of order

$$\Delta E \approx h\,\Delta\nu \approx h/\Delta t. \tag{6.3}$$

This is again the usual (energy-time) Heisenberg uncertainty relation.

Bohr then raises the question of where these latitudes in the particle's momentum and energy come from. That is, if ΔE and Δp represent the expected sizes of *changes* in the energy and momentum of the particle as it traverses the slit – and if the total energy and total momentum of an isolated system are strictly conserved – where do the new contributions to the energy and momentum of the particle come from? Bohr explains:

> From the point of view of the laws of conservation, the origin of such latitudes entering into the description of the state of the particle after passing through the hole may be traced to the possibilities of momentum and energy exchange with the diaphragm or the shutter. In the reference system [of the Figure], the velocity of the diaphragm may be disregarded and only a change of momentum Δp between the particle and the diaphragm needs to be taken into consideration. The shutter, however, which leaves the hole opened during the time Δt, moves with a considerable velocity $v \approx a/\Delta t$, and a momentum transfer Δp involves therefore an energy exchange with the particle, amounting to $[\Delta E = \int vF(t)dt \approx] v\,\Delta p \approx (1/\Delta t)\Delta q\,\Delta p \approx h/\Delta t$, being just of the same order of magnitude as the latitude ΔE given by [Eq. (6.3)] and, thus, allowing for momentum and energy balance [6].

With all of that laid out, Bohr then turns to confront Einstein's concerns (as he understood them) more directly. If, for example, the position and momentum of the diaphragm itself are carefully controlled and monitored, and conservation of energy and momentum are assumed in the interaction between the diaphragm and the passing particle, could one predict the subsequent behavior of the particle more accurately than would be allowed according to the quantum description?

> The problem raised by Einstein was now to what extent a control of the momentum and energy transfer, involved in a location of the particle in space and time, can be used for a further specification of the state of the particle after passing through the hole. Here, it must be taken into consideration that the position and the motion of the diaphragm and the shutter have so far been assumed to be accurately co-ordinated with the space-time reference frame. This assumption implies, in the description of the state of these bodies, an essential latitude as to their momentum and energy which need not, of course, noticeably affect the velocities, if the diaphragm and the shutter are sufficiently heavy. However, as soon as we want to know the momentum and energy of these parts of the measuring arrangement with an accuracy sufficient to control the momentum and energy exchange with the particle under investigation, we shall, in accordance with the general indeterminacy relations, lose the possibility of their accurate location in space and time. We have, therefore, to examine how far this circumstance will affect the intended use of the whole arrangement and, as we shall see, this crucial point clearly brings out the complementary character of the phenomenon [6].

Bohr means here that, up to now, we have assumed that the diaphragm is rigidly fixed in place so that, for example, its velocity is exactly zero and its position (the location

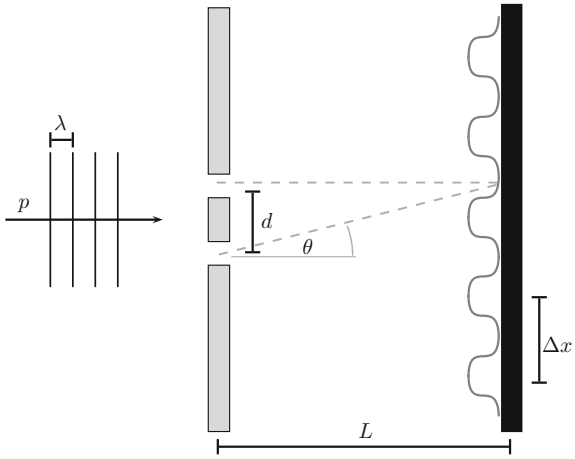


Fig. 6.2 The setup of the two-slit experiment discussed by Bohr in Ref. [6]. The *dashed grey lines* indicate two paths, differing in angle by θ, which a particle might take to some point on the screen

of the slit that the particle goes through) is precisely known. To make this concrete, one might imagine that the diaphragm structure is physically bolted down to the solid earth. But such bolts would allow and indeed necessitate a physical interaction, by means of which arbitrarily large quantities of energy and momentum might be exchanged between the diaphragm and the earth. So by bolting the diaphragm down (and thus precisely fixing its spatial location) we lose all control of its energy and momentum – and hence lose any ability to infer, from some hypothetical later measurement of its energy or momentum, any further information about the location of the now-distant particle.

Of course, by unbolting the diaphragm and, say, letting it glide freely along a frictionless track (running vertically in the Figure), we could remove the ability of the diaphragm to exchange energy and momentum with the earth, and thereby recover the ability to infer, from a later measurement of the energy or momentum of the diaphragm, something about the energy or momentum of the now-distant particle. But then, by the uncertainty principle, the spatial location of the diaphragm (and hence that of the particle) would become completely undefined. Thus, in a broad qualitative sense, Einstein's idea, as understood by Bohr, seems doomed.

Bohr proceeds to develop a closely-related example in which these ideas can more easily be analyzed quantitatively:

> The importance of considerations of this kind was, in the course of the discussions, most interestingly illuminated by the examination of an arrangement [involving a] diaphragm with two parallel slits, as is shown in [Fig. 6.2]. If a parallel beam of electrons (or photons) falls from the left [we shall] observe on the plate an interference pattern indicated by the [dark grey curve]. With intense beams, this pattern is built up by the accumulation of a large number of individual processes, each giving rise to a small spot on the photographic plate, and the distribution of these spots follows a simple law derivable from the wave analysis. The same distribution should also be found in the statistical account of many experiments performed with beams so faint that in a single exposure only one electron (or photon) will arrive at the photographic plate at some spot.... Since, now, as indicated by the [dashed lines], the momentum transferred to the ... diaphragm ought to be different if the electron

> was assumed to pass through the upper or the lower slit ..., Einstein suggested that a control of the momentum transfer would permit a closer analysis of the phenomenon and, in particular, to decide through which of the two slits the electron had passed before arriving at the plate [6].

Einstein, that is, had the idea that by carefully monitoring the position and momentum of the diaphragm (with, now, two slits in it), one could infer (from the final location at which the particle hits the detection screen) which slit the particle had gone through, because the momentum transfer between the particle and the diaphragm would need to have been slightly different in the two cases.

Bohr then presents the following rebuttal of Einstein's idea. The incident particle has momentum $p = h/\lambda$. The key idea here is that the momentum transfer between the particle and the diaphragm will be different, depending on which slit the particle goes through. For simplicity, in the Figure we have shown the case where there is *no* vertical momentum transfer if the particle goes through the top slit, whereas if the particle goes through the bottom slit it must bend upward by angle θ which implies it acquires a vertical momentum component of magnitude $p\sin(\theta) \approx (h/\lambda)(d/L)$ where we have written $\sin(\theta)$ in terms of the slit-spacing d and screen-distance L shown in the figure.

Now the idea is supposed to be that, by monitoring the vertical momentum of the diaphragm, we can determine, after the particle has passed, which slit it must have gone through. This will require that we can discriminate between the case in which the particle goes through the upper slit and the case in which the particle goes through the lower slit. But this requires that the uncertainty ΔP on the vertical momentum of the *diaphragm* be less than the difference between the momenta imparted to it when the particle goes through the different slits:

$$\Delta P \leq \frac{h}{\lambda}\frac{d}{L}. \tag{6.4}$$

But then, if we apply Heisenberg's uncertainty principle *to the diaphragm* we see that it must also have an uncertainty in its vertical position satisfying

$$\Delta Q \geq \frac{h}{\Delta P} \geq \frac{\lambda \cdot L}{d}. \tag{6.5}$$

This, as it turns out, is a very interesting result, because it is exactly the distance Δx between interference fringes on the screen. Therefore, in order to be able to determine which slit the particle went through by subsequently monitoring the vertical momentum of the diaphragm, the vertical position of the diaphragm must be uncertain by an amount greater than the fringe spacing on the screen:

$$\Delta Q \geq \Delta x. \tag{6.6}$$

But this obviously means that the interference pattern will be washed out: from one particle to the next, the probability distribution for the particle to hit the screen will shift up and down randomly by a distance as big as the spacing between the

interference fringes. And so the statistical pattern that builds up will no longer display the characteristic two-slit interference pattern.

Bohr summarizes the implications as follows:

> This point is of great logical consequence, since it is only the circumstance that we are presented with a choice of *either* tracing the path of the particle *or* observing interference effects, which allows us to escape from the paradoxical necessity of concluding that the behaviour of an electron or a photon should depend on the presence of a slit in the diaphragm through which it could be proved not to pass. We have here to do with a typical example of how the complementary phenomena appear under mutually exclusive experimental arrangements ... and are just faced with the impossibility, in the analysis of quantum effects, of drawing any sharp separation between an independent behaviour of atomic objects and their interaction with the measuring instruments which serve to define the conditions under which the phenomena occur [6].

The example thus shows not only that one cannot determine more details about the particle than is permitted according to the Heisenberg uncertainty principle, but also demonstrates the complementarity of the wave and particle descriptions: by adjusting the experimental arrangement in a way that allows an unambiguous determination of the particle's path, its wave character (namely, the appearance of interference) is thereby suppressed.

6.4 The Photon Box Thought Experiment

In his "Discussions with Einstein" essay, Bohr goes on to present another famous thought experiment that Einstein had proposed at the 1930 Solvay Conference:

> As an objection to the view that a control of the interchange of momentum and energy between the objects and the measuring instruments was excluded if these instruments should serve their purpose of defining the space-time frame of the phenomena, Einstein brought forward the argument that such control should be possible when the exigencies of relativity theory were taken into consideration. In particular, the general relationship between energy and mass, expressed in Einstein's famous formula
>
> $$E = mc^2$$
>
> should allow, by means of a simple weighing, to measure the total energy of any system and, thus, in principle to control the energy transferred to it when it interacts with an atomic object.
>
> As an arrangement suited for such purpose, Einstein proposed the device indicated in [Fig. 6.3], consisting of a box with a hole in its side, which could be opened or closed by a shutter moved by means of a clock-work within the box. If, in the beginning, the box contained a certain amount of radiation and the clock was set to open the shutter for a very short interval at a chosen time, it could be achieved that a single photon was released through the hole at a moment known with as great accuracy as desired. Moreover, it would apparently also be possible, by weighing the whole box before and after this event, to measure the energy of the photon with any accuracy wanted, in definite contradiction to the reciprocal indeterminacy of time and energy quantities in quantum mechanics [6].

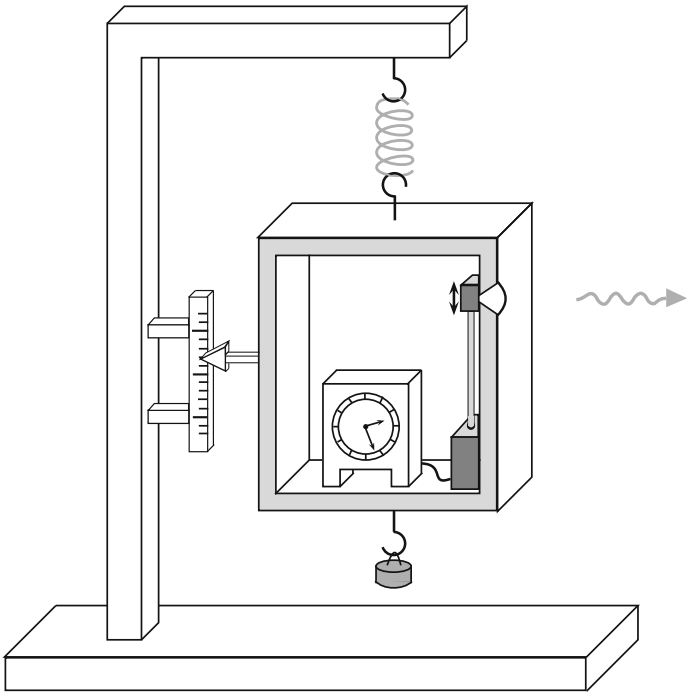


Fig. 6.3 Einstein's photon box setup as discussed by Bohr in Ref. [6]. An alarm clock inside the box triggers, at a pre-determined time as registered by the clock, a mechanical apparatus which briefly opens a shutter, allowing a single photon to escape to the *right*. The box hangs from a spring so that the weight of the box can be read from the scale *on the left*. After the photon is released, additional weight can be hung from the bottom of the box until the pointer returns to its original position, thus allowing a determination of the energy E of the escaped photon

So the idea is as follows: the clock mechanism inside the box will open the shutter at, say, time t, for a short duration Δt. During this period, a single photon will emerge from the aperture toward the right. The photon will be represented quantum mechanically as a wave packet of temporal duration Δt. This implies, by the standard energy-time uncertainty formula, that the energy of the photon will be "fuzzy" by an amount $\Delta E \geq h/\Delta t$.

Einstein's idea, however, was that the energy of the photon could be determined with arbitrary accuracy, after its emission, by carefully weighing the box from which the photon had emerged. And so the energy of the (now distant!) photon can be determined – and must hence be physically well-defined – to an accuracy greater than its quantum mechanical uncertainty ΔE. And so the quantum mechanical description must be incomplete.

According to Bohr, however, "it became clear ... that this argument could not be upheld." For, as Bohr goes on to explain, the process of weighing the box is itself subject to uncertainty principle constraints:

> The weighing of the box may ... be performed with any given accuracy Δm by adjusting the balance to its zero position by means of suitable loads. The essential point is now that any determination of this position with a given accuracy Δq will involve a minimum latitude Δp in the control of the momentum of the box connected with Δq by the [Heisenberg uncertainty principle]. This latitude must obviously again be smaller than the total impulse which, during the whole interval T of the balancing procedure, can be given by the gravitational field to a body with a mass Δm, or
>
> $$\Delta p \approx \frac{h}{\Delta q} < T \cdot g \cdot \Delta m$$
>
> where g is the gravity constant. The greater the accuracy of the reading q of the pointer, the longer must, consequently, be the balancing interval T, if a given accuracy Δm of the weighing of the box with its content shall be obtained.

Then, in a masterful judo-like move of using one of Einstein's greatest accomplishments against him, Bohr notes that

> according to general relativity theory, a clock, when displaced in the direction of the gravitational force by an amount of Δq, will change its rate in such a way that its reading in the course of a time interval T will differ by an amount ΔT given by the relation
>
> $$\frac{\Delta T}{T} = \frac{1}{c^2} g \Delta q.$$
>
> By comparing [the last two equations] we see, therefore, that after the weighing procedure there will in our knowledge of the adjustment of the clock be a latitude
>
> $$\Delta T > \frac{h}{c^2 \Delta m}.$$
>
> Together with [the formula $\Delta E = \Delta m\, c^2$ coming from Einstein's famous equation], this relation again leads to
>
> $$\Delta T \cdot \Delta E > h$$
>
> in accordance with the indeterminacy principle. Consequently, the use of the apparatus as a means of accurately measuring the energy of the photon will prevent us from controlling the moment of its escape [6].

Although the physics involved is rather more complicated, the conclusion here is essentially identical to what Bohr said in the analysis of the diffraction and interference experiments from the previous section: using the photon box in the intended way to determine, after its emission, both the energy and release-time of the photon, would require *mutually exclusive* experimental procedures. In particular, an accurate determination of the release-time of the photon requires that the box (with its internal clock) be held rigidly fixed in the background gravitational field; such fixation, though, allows energy transfer (with the earth or whatever the box is fixed to) and thus precludes subsequently inferring, from the weight of the box, the energy of the now-distant photon. And conversely, leaving the box free to oscillate vertically, so that the energy of the emitted photon can be reliably inferred, means that (due to the general relativistic time-dilation effect) the reading of the clock can no longer reliably indicate the release-time of the photon.

Thus – if it had indeed been Einstein's claim that the energy E and emission time T of the photon could both be determined, by subsequent examination of the

box, to an accuracy greater than should be allowed by the Heisenberg uncertainty relations – then Bohr has shown that in fact, no, this is not after all possible. As long as we consistently apply the Heisenberg uncertainty principle to all elements of the physical system under examination (including the measuring equipment!) it seems to turn out that the uncertainty principle cannot be beaten. Bohr regards this as the essential proof for the completeness of the quantum mechanical description. And since in supposedly refuting Einstein's objection, Bohr had used Einstein's own general relativity against him, this episode is widely regarded as a rhetorical triumph for Bohr and the Copenhagen philosophy.

But was this, after all, Einstein's claim?

Interestingly, Bohr reports that, around 1933

> Einstein was far from satisfied and with his usual acuteness had discerned new aspects of the situation which strengthened his critical attitude. In fact, by further examining the possibilities for the application of the balance arrangement, Einstein had perceived alternative procedures which, even if they did not allow the use he originally intended, might seem to enhance the paradoxes beyond the possibilities of logical solution [6].

Such comments by Bohr have given rise to a widespread suggestion that, between 1930 and 1935 (when the EPR paper finally appeared) Einstein responded to his supposed defeats (in 1927 and 1930) by fundamentally changing his approach to criticizing the developing Copenhagen orthodoxy. In particular, according to this viewpoint, Einstein finally came to grips with the internal consistency of the theory and began to explore instead the "new aspects" that would appear explicitly in the 1935 EPR paper.

Recall from Sect. 4.1, however, the centrality of the concept of locality to Einstein's concerns as he expressed them already in 1927. As we have seen, Bohr's 1949 recapitulation of the 1927 discussions seem to completely omit the aspects of Einstein's concerns that render them quite in line with the later EPR argument. So one begins to suspect that the "new aspects" (of Einstein's thinking) which Bohr recognized only after 1930, were not new at all; instead, Bohr had simply failed to understand them prior to this point.

In regard to the photon box thought experiment, this suspicion would imply that the concept of locality played, somehow, a more central and important role than is apparent in Bohr's analysis. In particular, one suspects that, for Einstein, it was crucial that (after some time) the emitted photon is *spatially separated* from the box. Locality would then seem to imply that our choice of which measuring procedure to implement on the box, could have no effect on the physical state of the distant photon. One suspects, that is, that for Einstein locality was the crucial assumption warranting inference from "we could measure *either* of two complementary properties on the nearby system" to "*both* of the corresponding properties must exist for the distant system".

The suspicion is strongly confirmed by the contents of a letter written to Bohr by Paul Ehrenfest (with whom Einstein had discussed the dialogues from the 1930 meeting shortly afterwards). We quote here philosopher Don Howard's description of this entire episode:

> At center stage in the Einstein-Bohr encounter at the 1930 Solvay meeting was Einstein's well-known photon box thought experiment. A box containing a photon has an opening covered by a shutter that is activated by a timer attached to a clock inside the box by means of which we could accurately time the emission of the photon from the box. The whole box is suspended by a spring by means of which arrangement we could weigh the box both before and after the photon's emission with whatever accuracy we desire, thus determining the photon's energy via the mass-energy equivalence relation. As Bohr tells the story, Einstein introduced the photon-box thought experiment for the purpose, yet again, of exhibiting violations of Heisenberg indeterminacy. Simply perform both measurements: weigh the box to fix the emitted photon's energy and open the box to check the clock and fix the time of emission. Bohr tells us that, at first, Einstein had him completely stumped. He could find no flaw in Einstein's reasoning. Only in the wee hours of the morning did it come to him. Ironically, general relativity would save quantum mechanics, specifically the general relativistic effect of a gravitational field on clock rates. A quick calculation showed Bohr that the change in the box's mass when the photon is emitted changes, in turn, its vertical location in the earth's gravitational field, and that the effect of the latter change on the rate of the clock in the box induces precisely the uncertainty in the clock's rate needed to ensure satisfaction of the Heisenberg indeterminacy principle. Bohr uses general relativity against Einstein to save quantum mechanics! A wonderful story. But is it true?
>
> Einstein seems to have thought that they were arguing about something else. We know this from a letter that Paul Ehrenfest wrote to Bohr in July 1931, after a visit with Einstein in Berlin. Ehrenfest and Einstein seem to have had a long and thorough chat about the debate with Bohr at the previous fall's Solvay meeting. Ehrenfest reports to Bohr a most surprising comment from Einstein:
>
> *He [Einstein] said to me that, for a very long time already, he absolutely no longer doubted the uncertainty relations, and that he thus, e.g., had BY NO MEANS invented the 'weighable light-flash box' (let us call it simply L-F-box) 'contra uncertainty relation,' but for a totally different purpose.* [Ehrenfest to Bohr, 9 July 1931]
>
> What was that totally different purpose? It was nothing other than an anticipation of Einstein's later argument for the incompleteness of quantum mechanics.
>
> As Ehrenfest explains to Bohr, Einstein's idea was this. Let the photon leave the box and be reflected back from a great time and distance, say one-half light year. At about the time when the photon is reflected, we can either weigh the box or check the clock, making possible our predicting either the exact time of the photon's return or its energy (literally, its color), which is to say that, depending upon which measurement we choose, we ascribe a different theoretical state to the photon, one with definite energy, one entailing a definite time of arrival. Crucial is the fact that the event of performing the measurement on the box – weighing it the second time or checking the clock – is [spatially] separated from the event of the photon's distant reflection, because then our choice of a measurement to perform can have no effect on the real state of affairs of the photon, meaning that the photon's real state of affairs when it returns will be one and the same, regardless of the measurement we performed on the box. This is all just quantum mechanics, in Einstein's view. But then quantum mechanics has associated two different theoretical states with one real state of affairs, which is possible only if the quantum theory's state descriptions are incomplete [7].

So it definitely appears that Einstein's early (pre-1935) arguments were simply not understood properly by Bohr. (And it is curious that Bohr's later account of these early discussions did not attempt to correct the misunderstanding, but instead reinforced it for posterity.) In any case, though, it represents progress that, by around 1933, Bohr

(in his re-telling) began to recognize and more directly confront the "new aspects" of Einstein's arguments.

Here is Bohr's description of these "new aspects" in the context of the photon box thought experiment:

> Einstein had pointed out that, after a preliminary weighing of the box with the clock and the subsequent escape of the photon, one was still left with the choice of either repeating the weighing or opening the box and comparing the reading of the clock with the standard time scale. Consequently, we are at this stage still free to choose whether we want to draw conclusions either about the energy of the photon or about the moment when it left the box. Without in any way interfering with the photon between its escape and its later interaction with other suitable measuring instruments, we are, thus, able to make accurate predictions pertaining *either* to the moment of its arrival *or* to the amount of energy liberated by its absorption. Since, however, according to the quantum-mechanical formalism, the specification of the state of an isolated particle cannot involve both a well-defined connection with the time scale and an accurate fixation of the energy, it might thus appear as if this formalism did not offer the means of an adequate description [6].

Now that, for sure, captures the concern that Einstein seems to have had in mind all along, and would eventually appear most famously and explicitly in the 1935 EPR paper. Note in particular the exact parallel to the EPR reasoning: by measuring one property of the box, we can determine an exact value for the corresponding property of the distant photon; on the other hand, by instead measuring a different property of the box, we can determine an exact value for a different property of the distant photon; but since our measurements on the box must – by the locality assumption – have no effect on the physical state of the distant photon, the possibility of our determining either of these properties implies that both properties have, already, sharp (if unknown) values. And so the quantum mechanical formalism (which precludes such simultaneous value assignments) must be incomplete.

So how, then, does Bohr respond to this early version of the EPR argument?

> Once more Einstein's searching spirit had elicited a peculiar aspect of the situation in quantum theory, which in a most striking manner illustrated how far we have here transcended customary explanation of natural phenomena. Still, I could not agree with the trend of his remarks.... In my opinion, there could be no other way to deem a logically consistent mathematical formalism as inadequate than by demonstrating the departure of its consequences from experience or by proving that its predictions did not exhaust the possibilities of observation, and Einstein's argumentation could be directed to neither of these ends. In fact, we must realize that in the problem in question we are not dealing with a *single* specified experimental arrangement, but are referring to *two* different, mutually exclusive arrangements. In the one, the balance together with another piece of apparatus like a spectrometer is used for the study of the energy transferred to the photon; in the other, a shutter regulated by a standardized clock together with another apparatus of similar kind, accurately timed relatively to the clock, is used for the study of the time of propagation of a photon over a given distance. In both these cases, as also assumed by Einstein, the observable effects are expected to be in complete conformity with the predictions of the theory.
>
> The problem again emphasizes the necessity of considering the *whole* experimental arrangement, the specification of which is imperative for any well-defined application of the quantum-mechanical formalism [6].

I think it is probably safe to say that one will either regard this response as satisfying, or not, depending on the extent to which one's philosophical attitudes align with those of Bohr, or Einstein, respectively.

In the next section, we will continue to explore Bohr's responses to Einstein's concerns by considering Bohr's official response to the 1935 paper of Einstein, Podolsky, and Rosen.

6.5 Bohr's Reply to EPR

Let us then finally turn to Bohr's response to the actual EPR paper of 1935. It is of historical interest that, according to the later recollection of Bohr's close colleague Rosenfeld, the EPR paper was an "onslaught" which "came down upon us as a bolt from the blue." Rosenfeld reports that "as soon as Bohr had heard my report of Einstein's argument, everything else was abandoned" as they dedicated themselves to rebutting the argument [8].

So, after the days and weeks of careful thinking, how did Bohr respond? Early on in the essay, Bohr reviews the idea that the impossibility of attributing definite properties to measured systems arises from the uncontrollable physical disturbance of their states during the physical interaction with the measuring apparatus:

> The apparent contradiction in fact discloses only an essential inadequacy of the customary viewpoint of natural philosophy for a rational account of physical phenomena of the type with which we are concerned in quantum mechanics. Indeed the *finite interaction between object and measuring agencies* conditioned by the very existence of the quantum of action entails – because of the impossibility of controlling the reaction of the object on the measuring instruments if these are to serve their purpose – the necessity of a final renunciation of the classical ideal of causality and a radical revision of our attitude towards the problem of physical reality [9].

But surely the intention of the EPR argument was precisely to neutralize this "disturbance" defense, by separating the measurement event from the particle to which properties are being attributed. If, in the EPR case, the definite inferred properties of the distant particle in any sense arise, newly, as a result of the measurement on the nearby partner, this would be the very sort of nonlocal causation that EPR regarded as unacceptable or, more precisely, in conflict with relativity's prohibition on faster-than-light causation.

Bohr seems to only partially appreciate this, and his response is thus notoriously difficult to understand. He insists that the EPR criterion of reality (inside of which, remember, the crucial concept of locality was buried) "contains – however cautious its formulation may appear – an essential ambiguity when it is applied to the actual problems with which we are here concerned" [9]. Here is his detailed statement of the alleged ambiguity:

> From our point of view we now see that the wording of the above-mentioned criterion of physical reality proposed by Einstein, Podolsky, and Rosen contains an ambiguity as regards the meaning of the expression 'without in any way disturbing a system.' Of course there

> is in a case like that just considered no question of a mechanical disturbance of the system under investigation during the last critical stage of the measuring procedure. But even at this stage there is essentially the question of *an influence on the very conditions which define the possible types of predictions regarding the future behavior of the system.* Since these conditions constitute an inherent element of the description of any phenomenon to which the term 'physical reality' can be properly attached, we see that the argumentation of the mentioned authors does not justify their conclusion that quantum-mechanical description is essentially incomplete. On the contrary this description, as appears from the preceding discussion, may be characterized as a rational utilization of all possibilities of unambiguous interpretation of measurements, compatible with the finite and uncontrollable interaction between the objects and the measuring instruments in the field of quantum theory. In fact, it is only the mutual exclusion of any two experimental procedures, permitting the unambiguous definition of complementary physical quantities, which provides room for new physical laws, the coexistence of which might at first sight appear irreconcilable with the basic principles of science. It is just this entirely new situation as regards the description of physical phenomena, that the notion of *complementarity* aims at characterizing [9].

It seems that Bohr is agreeing with EPR that a "mechanical disturbance" – that is, a nonlocal causal influence on the distant particle – is unacceptable. But, he seems to say, there is another kind of influence – what one commentator [10] has described as a "semantic disturbance"... not, apparently, a physical influence per se, but instead an influence on what we can *say* about the distant system.

Here is what Bell would write, later, about Bohr's response:

> While imagining that I understand the position of Einstein, as regards the EPR correlations, I have very little understanding of the position of his principal opponent, Bohr. Yet most contemporary theorists have the impression that Bohr got the better of Einstein in the argument and are under the impression that they themselves share Bohr's views. As an indication of those views I quote a passage from his reply to Einstein, Podolsky and Rosen. It is a passage which Bohr himself seems to have regarded as definitive, quoting it himself when summing up much later. Einstein, Podolsky and Rosen had assumed that '...if, without in any way disturbing a system, we can predict with certainty the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity'. Bohr replied: '...the wording of the above mentioned criterion... contains an ambiguity as regards the meaning of the expression "without in any way disturbing a system". Of course there is in a case like that just considered no question of a mechanical disturbance of the system under investigation during the last critical stage of the measuring procedure. But even at this stage there is essentially the question of *an influence on the very conditions which define the possible types of predictions regarding the future behaviour of the system* [so] their argumentation does not justify their conclusion that quantum mechanical description is essentially incomplete ... This description may be characterized as a rational utilization of all possibilities of unambiguous interpretation of measurements, compatible with the finite and uncontrollable interaction between the objects and the measuring instruments in the field of quantum theory'.
>
> Indeed I have very little idea what this means. I do not understand in what sense the word 'mechanical' is used, in characterizing the disturbances which Bohr does not contemplate, as distinct from those which he does. I do not know what the italicized passage means – 'an influence on the very conditions...'. Could it mean just that different experiments on the first system give different kinds of information about the second? But this was just one of the main points of EPR, who observed that one could learn *either* the position *or* the momentum of the second system. And then I do not understand the final reference to 'uncontrollable interactions between measuring instruments and objects', [as] it seems just to ignore the essential point of EPR that in the absence of action at a distance, only the first system could

> be supposed disturbed by the first measurement and yet definite predictions become possible for the second system. Is Bohr just rejecting the premise – 'no action at a distance' – rather than refuting the argument? [11].

Bell, that is, suggests reading Bohr as conceding (despite his explicit denial of a specifically "mechanical" disturbance) that there is a non-local action-at-a-distance at work in this situation, according to quantum mechanics.

Einstein could also do no better than this same uncomfortable understanding of Bohr's response. In his 1949 commentary, he wrote:

> Of the 'orthodox' quantum theoreticians whose position I know, Niels Bohr's seems to me to come nearest to doing justice to the problem. Translated into my own way of putting it, he argues as follows:
>
> If the partial systems A and B form a total system which is described by its Ψ-function $\Psi(AB)$, there is no reason why any mutually independent existence (state of reality) should be ascribed to the partial systems A and B viewed separately, *not even if the partial systems are spatially separated from each other at the particular time under consideration*. The assertion that, in this latter case, the real situation of B could not be (directly) influenced by any measurement taken on A is, therefore, within the framework of quantum theory, unfounded and (as the paradox shows) unacceptable [12].

From the point of view of a "realist" such as Einstein – meaning simply someone who believes in the existence of an external physical world that is what it is independent of any observation and which observation is ultimately observation *of* – Bohr's reply to EPR will always remain deeply unsatisfying. Yet we must remember that from the point of view of the Copenhagen philosophy, it is precisely, at the end of the day, this assumption of "realism" which is being challenged. Heisenberg, for example, wrote that

> the idea of an objective real world whose smallest parts exist objectively in the same sense as stones or trees exist, independently of whether or not we observe them ... is impossible.... [13]

Bohr, similarly, insisted that in quantum mechanics we meet

> in a new light the old truth that in our description of nature the purpose is not to disclose the real essence of the phenomena but only to track down, so far as it is possible, relations between the manifold aspects of our experience [14].

Bohr stressed repeatedly this point that physical theories must not aim at describing some independent, objective physical reality:

> The entire formalism is to be considered as a tool for deriving predictions, of definite or statistical character, as regards information obtainable under experimental conditions described in classical terms and specified by means of parameters entering into the algebraic or differential equations.... These symbols themselves are not susceptible to pictorial interpretation [15].

And according to Bohr's colleague Aage Petersen, when Bohr was once asked whether the theory could in any sense be understood as describing an objective reality, Bohr replied

> 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 [16].

Bohr's dismissal of the EPR argument for incompleteness may indeed be the only coherent and rational response given this deeper philosophical point of view.

6.6 Contemporary Perspectives

I mentioned in the introduction of this chapter that the Copenhagen philosophy achieved a kind of orthodox status in the 1930s and has essentially held this position to the present day. But I think most physics students who have been exposed to this orthodoxy – and indeed most physics professors who have accepted it and even taken part in teaching it – will probably find some aspects of the philosophy, as explained by Bohr and Heisenberg, a little surprising. Does the Copenhagen interpretation really insist, for example, that there simply is no real physical world at the microscopic level? This seems bizarre if not downright incomprehensible.

One can certainly find contemporary proponents of the Copenhagen philosophy who embrace such radical philosophical positions. The eminent Austrian experimentalist Anton Zeilinger, for example, summarized "The message of the quantum" in Copenhagen terms. He stresses the failure of classical notions of causality as follows:

> The discovery that individual events are irreducibly random is probably one of the most significant findings of the twentieth century. Before this, one could find comfort in the assumption that random events only seem random because of our ignorance. For example, although the brownian motion of a particle appears random, it can still be causally described if we know enough about the motions of the particles surrounding it.... But for the individual event in quantum physics, not only do we not know the cause, there is no cause. The instant when a radioactive atom decays, or the path taken by a photon behind a half-silvered beam-splitter are objectively random. There is nothing in the Universe that determines the way an individual event will happen. Since individual events may very well have macroscopic consequences ... the Universe is fundamentally unpredictable and open, not causally closed [17].

Zeilinger then insists that "the concept of reality itself is at stake" in certain experiments that we will discuss further in Chap. 8. As he elaborates:

> A criticism of realism also emerges from the notion of complementarity. It is not just that we are unable to measure two complementary quantities of a particle, such as its position and momentum, at the same time. Rather, the assumption that a particle possesses both position and momentum, before the measurement is made, is wrong. Our choice of measurement apparatus decides which of these quantities can become reality in the experiment.
>
> So, what is the message of the quantum? I suggest we look at the situation from a new angle. We have learned in the history of physics that it is important not to make distinctions that have no basis – such as the pre-newtonian distinction between the laws on Earth and those that govern the motion of heavenly bodies. I suggest that in a similar way, the distinction between reality and our knowledge of reality, between reality and information, cannot be made. There is no way to refer to reality without using the information we have about it [17].

Undoubtedly this almost idealistic (in the sense of Berkeley) anti-realism is part of "the Copenhagen interpretation" for many contemporary physicists.

But I think most physicists would find themselves slightly embarrassed by this kind of openly philosophical, anti-realist speculation. The more mainstream understanding of "the Copenhagen interpretation" is thus, I think, a little more restrained and pragmatic. This attitude is nicely captured in the widely used quantum mechanics text by David Griffiths, who explains that Born's statistical interpretation of the wave function

> ...introduces a kind of **indeterminacy** into quantum mechanics, for even if you know everything the theory has to tell you about the particle (to wit: its wave function), you cannot predict with certainty the outcome of a simple experiment to measure its position – all quantum mechanics has to offer is *statistical* information about the *possible* results. This indeterminacy has been profoundly disturbing to physicists and philosophers alike. Is it a peculiarity of nature, a deficiency in the theory, a fault in the measuring apparatus, or *what*?
>
> Suppose I *do* measure the position of the particle, and I find it to be at [some particular] point C. Question: Where was the particle just *before* I made the measurement? There are three plausible answers to this question, and they serve to characterize the main schools of thought regarding quantum indeterminacy:
>
> **1.** The **realist** position: *The particle was at C*. This certainly seems like a sensible response, and it is the one Einstein advocated. Note, however, that if this is true then quantum mechanics is an **incomplete** theory, since the particle *really was* at C, and yet quantum mechanics was unable to tell us so. To the realist, indeterminacy is not a fact of nature, but a reflection of our ignorance.... Evidently Ψ is not the whole story – some additional information (known as a **hidden variable**) is needed to provide a complete description of the particle.
>
> **2.** The **orthodox** position: *The particle wasn't really anywhere.* It was the act of measurement that forced the particle to 'take a stand' (though how and why it decided on the point C we dare not ask). Jordan said it most starkly: 'Observations not only *disturb* what is to be measured, they *produce* it. ... We *compel* [the particle] to assume a definite position.' This view (the so-called **Copenhagen interpretation**) is associated with Bohr and his followers. Among physicists it has always been the most widely accepted position. Note, however, that if it is correct there is something very peculiar about the act of measurement – something that over half a century of debate has done precious little to illuminate.
>
> **3.** The **agnostic** position: *Refuse to answer.* This is not quite as silly as it sounds – after all, what sense can there be in making assertions about the status of a particle *before* a measurement, when the only way of knowing whether you were right is precisely to conduct a measurement, in which case what you get is no longer 'before the measurement'? It is metaphysics (in the pejorative sense of the word) to worry about something that cannot, by its nature, be tested. Pauli said, 'One should no more rack one's brain about the problem of whether something one cannot know anything about exists all the same, than about the ancient question of how many angels are able to sit on the point of a needle.' For decades this was the 'fall-back' position of most physicists: They'd try to sell you answer 2, but if you were persistent they'd switch to 3 and terminate the conversation [18].

Incidentally, Griffiths goes on to suggest (just like Zeilinger) that certain experiments (pertaining to something called Bell's Theorem that is the subject of our Chap. 8) have recently "eliminated agnosticism as a viable option" and have "confirmed decisively the orthodox interpretation." About this, Griffiths is (like Zeilinger) simply wrong (in part because of an unnecessarily restrictive conception of the "realist" alternative); this will become clearer in the following two chapters.

In his overall characterization of the three options, however, Griffiths is admirably open and reasonable; many textbooks don't even acknowledge something like option

1 but instead just insist dogmatically that some superposition of **2** and **3** is the final truth, handed down from on high by Bohr, and not to be questioned. The final quoted sentence above is also, in my experience, perfectly accurate about the attitude of most physicists: they know they are supposed to believe **2** and so will do some minimal amount of due diligence trying to propagandize on behalf of the Copenhagen interpretation; but at the end of the day they don't take it too seriously and frankly don't really care and are perfectly content to just stop talking about it.

This pragmatic attitude was brilliantly captured by N. David Mermin, who wrote in a 1989 essay in *Physics Today*:

> If I were forced to sum up in one sentence what the Copenhagen interpretation says to me, it would be 'Shut up and calculate! [19].

This is probably the best – certainly the briefest – summary of how most physicists today understand "the Copenhagen interpretation". It captures perfectly the typical physicist's impatience for idle philosophical speculation and desire to get on with obviously practical things like using the theory to calculate predictions for how measurements should come out, and then testing those predictions with actual experiments.

Of course, this attitude is rather contrary to the point of view adopted in the present book. One should not, however, regard this as an endorsement of "idle philosophical speculation". Just the opposite, in fact. There is, I think, a deep irony in the fact that "Shut up and calculate!" is almost always deployed against people who want to *criticize* the orthodox, Copenhagen interpretation and construct an alternative theory that, for example, resolves the measurement problem. Such alternative theories typically postulate new sorts of microscopic objects, obeying new dynamical equations, in terms of which a uniform and coherent description of microscopic processes might be shown (through calculations!) to become possible. It is very surprising that physicists who value precisely-formulated theories and the calculations these make possible would prefer Bohr's philosophical speeches rather than the more hard-headed alternative theories we will cover in the remainder of the book. In a rational world, that is, "Shut up and calculate!" is what the *critics* should say to the Copenhagenists, whose dogmatic (yet simultaneously unserious) attachment to Bohr's philosophy prevents them from even asking the kinds of questions that might lead to real practical advances.

Clearly there are some deep issues – philosophical issues about the proper goals of science and sociological issues about how the physics community deals with disagreements over what questions are legitimate – that we will not be able to answer here. But one thing is for sure: to whatever extent the Copenhagen philosophy insists that it is not merely wrong, but *impossible*, to provide a uniform, coherent, realistic description of the world, which is nevertheless consistent with all known experimental facts, the Copenhagen philosophy is in that regard simply *wrong*. Several such candidate theories exist. Exploring them will occupy us for most of the rest of the book, starting, in the next chapter, with the pilot-wave theory of de Broglie and Bohm, which provides a stark, eye-opening contrast to the Copenhagen interpretation.

Projects:

6.1 Read the published version of Bohr's Como lecture [1] and report back on any aspects that you find surprising, interesting, novel, or illuminating.

6.2 In one of the passages quoted in Sect. 6.2, Heisenberg writes, about the quantum mechanical wave function:

> This probability function represents a mixture of two things, partly a fact and partly our knowledge of a fact. It represents a fact in so far as it assigns at the initial time the probability unity (i.e., complete certainty) to the initial situation: the electron moving with the observed velocity at the observed position; 'observed' means observed within the accuracy of the experiment. It represents our knowledge in so far as another observer could perhaps know the position of the electron more accurately. The error in the experiment does – at least to some extent – not represent a property of the electron but a deficiency in our knowledge of the electron. Also this deficiency of knowledge is expressed in the probability function [5].

Here Heisenberg wants to draw a distinction between the fundamental, irreducible type of uncertainty (described by his famous uncertainty relations) and the ordinary type of uncertainty that arises from, for example, imperfect measurements. Does Heisenberg's position here leave him open to the criticism that, if the same physical situation can be described by two different quantum mechanical wave functions (based on different amounts of uncertainty in at least the second sense), the quantum mechanical descriptions of physical states cannot be complete? Explain.

6.3 Do you think Bohr would agree with Heisenberg's suggestion that "another observer could perhaps know the position of the electron more accurately"? Explain.

6.4 What, according to Heisenberg, is the difference between the use of probabilities in classical physics, and their use in quantum mechanics?

6.5 Read Heisenberg's essay on "The Copenhagen Interpretation" [5] and report back on anything you find surprising, interesting, novel, or illuminating.

6.6 Read Bohr's "Discussion with Einstein..." [6] paper and report back on anything you find surprising, interesting, novel, or illuminating.

6.7 Show that, as claimed in Bohr's analysis of the two-slit experiment discussed in Sect. 6.3, the spacing Δx between adjacent interference maxima (see Fig. 6.2) is $\lambda L/d$.

6.8 In a 1979 paper [20], Wootters and Zurek provide a more detailed and quantitative analysis of the 2-slit experiment discussed in Sect. 6.3. Read their paper and summarize their arguments and conclusions.

6.9 Richard Feynman discusses the 2-slit experiment and its interpretation in Ref. [21]. How does Feynman's philosophical attitude toward quantum mechanics relate to the Copenhagen interpretation?

6.10 In the text it was suggested that (in his 1949 reminiscence) Bohr had misunderstood or misrepresented Einstein's 1930 photon box argument. Do you think Bohr might also have misunderstood/misrepresented Einstein's arguments regarding the diffraction and interference experiments we discussed in

Section 6.3? If so, explain how those arguments could be reformulated in a way that makes it clearer how they anticipate the 1935 EPR argument. If not, explain why the diffraction/interference examples are importantly different.

6.11 Work through the mathematical details of Bohr's analysis of the photon box experiment. (You might need or want to do a little independent research to understand the general relativistic formula for gravitational time-dilation.)

6.12 In Ref. [22], Dieks and Lam present a detailed analysis of Einstein's "photon box" thought experiment. Read their paper and summarize their arguments and conclusions. (Note that this requires a familiarity with operators and their commutators that is slightly beyond the level required elsewhere in this book.)

6.13 Commentators on the Einstein–Bohr debates often characterize Einstein as a kind of stubborn old conservative who simply couldn't get with the new quantum program. For example, Heisenberg wrote:

> Most scientists are willing to accept new empirical data and to recognize new results, provided they fit into their philosophical framework. But in the course of scientific progress it can happen that a new range of empirical data can be completely understood only when the enormous effort is made to enlarge this framework and to change the very structure of the thought process. In the case of quantum mechanics, Einstein was apparently no longer willing to take this step, or perhaps no longer able to do so [23].

And Max Born wrote:

> At first there were quite a number of serious scientists who did not want to know anything about the theory of relativity; conservative individuals, who were unable to free their minds from the prevailing philosophical principles.... Einstein himself belonged to this group in later years; he could no longer take in certain new ideas in physics which contradicted his own firmly held philosophical convictions [23].

What do you think of this? Of the two interlocutors, Einstein and Bohr, which one was really open to new theoretical concepts, and which one insisted on preserving old ideas, come what may?

6.14 Read Bohr's reply [9] to the EPR paper and report on anything you find surprising, interesting, novel, or illuminating.

6.15 In his reply [9] to EPR, Bohr provides a concrete kind of setup which would give rise to something like the entangled EPR state, in which "a subsequent single measurement of either of the position or of the momentum of one of the particles will automatically determine the position or momentum, respectively, of the other particle with any desired accuracy." What does he say about this and how would Einstein reply?

6.16 Interview some physicists about the Copenhagen interpretation. Ask them whether they basically agree with it. Then ask them to summarize what it says. You might also ask them specifically about whether Bohr successfully refuted the EPR argument and, if so, how the refutation works exactly. Summarize and share your findings.

6.17 The *Wikipedia* page on the Copenhagen interpretation provides (as of this writing) the following supposedly Copenhagen response to Schrödinger's cat: "The wave function reflects our knowledge of the system. The wave function

$[\frac{1}{\sqrt{2}}(\psi_{\text{alive}} + \psi_{\text{dead}})]$ means that, once the cat is observed, there is a 50% chance it will be dead, and 50% chance it will be alive." Do you think this accurately captures what Bohr would have said about Schrödinger's cat?

6.17 The *Wikipedia* page on the Copenhagen interpretation provides (as of this writing) the following supposedly Copenhagen response to the EPR argument: "Assuming wave functions are not real, wave-function collapse is interpreted subjectively. The moment one observer measures the spin of one particle, he knows the spin of the other. However, another observer cannot benefit until the results of that measurement have been relayed to him, at less than or equal to the speed of light." Do you think this provides a fair summary of Bohr's response to EPR?

6.18 What do you think Bohr would have thought about the slogan "Shut up and calculate!"? It might be helpful to do some research here regarding Bohr's thoughts about the applicability of "complementarity" outside of physics; see, for example, Mara Beller's "The Sokal Hoax: At Whom are we Laughing?" [24].

References

1. N. Bohr, The quantum postulate and the recent development of atomic theory. Nature **121**, 580–590 (14 April 1928)
2. H. Krips, Measurement in Quantum Theory, Stanford Encyclopedia of Philosophy (2008), http://stanford.library.sydney.edu.au/archives/fall2008/entries/qt-measurement/
3. M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974)
4. W. Heisenberg, *The History of Quantum Theory in Physics and Philosophy* (Harper & Row, New York, 1958)
5. W. Heisenberg, The Copenhagen Interpretation of Quantum Theory, in *Physics and Philosophy* (Harper & Row, New York, 1958)
6. N. Bohr, Discussion with Einstein on Epistemological Problems in Atomic Physics, in *Albert Einstein: Philosopher-Scientist*, ed. by P.A. Schilpp, vol. 1 (1949)
7. D. Howard, Revisiting the Einstein–Bohr dialogues. Iyyun: Jerus. Philos. Quarterly **56**, 57–90 (2007)
8. L. Rosenfeld, 1967, from Wheeler and Zurek, *Quantum Theory and Measurement* (Princeton University Press, Princeton 1983), p. 142
9. N. Bohr, Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **48**, 696–702 (15 Oct 1935)
10. A. Fine, *The Shaky Game*, 2nd edn. (University of Chicago Press, Chicago, 1996), p. 35
11. J.S. Bell, *Bertlmann's Socks and the Nature of Reality*, 2nd edn., Speakable and Unspeakable in Quantum Mechanics (Cambridge University Press, Cambridge, 2004)
12. A. Einstein, Reply to criticisms, in *Albert Einstein: Philosopher-Scientist*, ed. by P.A. Schilpp, vol. 2 (1949)
13. W. Heisenberg, Criticism and counterproposals to the copenhagen interpretation of quantum theory, in *Physics and Philosophy* (Harper and Row, New York, 1958), p. 129
14. N. Bohr, *Atomic Theory and the Description of Nature* (Cambridge University Press, Cambridge, 1961), p. 18
15. N. Bohr, On the notion of causality and complementarity. Dialectica **2**, 312–319 (1948)
16. A. Petersen, The philosophy of Niels Bohr, in *Niels Bohr: A Centenary Volume*, ed. by A.P. French, P.J. Kennedy, (Harvard University Press, Massachusetts, 1985), p. 305
17. A. Zeilinger, The message of the quantum. Nature **438**, 8 (2005)

18. D. Griffiths, *Introduction to Quantum Mechanics* (Prentice Hall, New Jersey, 1995)
19. N.D. Mermin, What's wrong with this pillow? Phys. Today (1989); see also Could Feynman have said this? Phys. Today (2004)
20. W.K. Wootters, W.H. Zurek, Complementarity in the double-slit experiment: quantum nonseparability and a quantitative statement of Bohr's principle. Phys. Rev. D **19**(2), 473–84 (1979)
21. R. Feynman, Probability and uncertainty – the quantum mechanical view of nature, in *The Character of Physical Law* (1965)
22. D. Dieks, S. Lam, Complementarity in the Bohr–Einstein Photon Box. Am. J. Phys. **76**, 838 (2006)
23. I. Born, *The Born-Einstein Letters* (Walker and Company, New York, 1971)
24. M. Beller, The Sokal Hoax: at whom are we laughing? Phys. Today 51, 29–34 (1998)

Chapter 7
The Pilot-Wave Theory

According to the orthodox interpretation of quantum mechanics, the theory provides – already, with wave functions alone – complete descriptions of physical states. In Chaps. 3–5 we reviewed three distinct but inter-related problems that afflict this view, at least according to people like Einstein, Schrödinger, and Bell. The alleged problems could be summarized by saying that there seems to be something deficient about quantum mechanical wave functions as descriptions of physical reality – either (at best) the wave functions provide only an *incomplete* description of what is actually going on physically, or (at worst) they fail to provide any comprehensible description of physically real processes at all.

Bohr and Heisenberg, of course, did not accept the criticisms and built a rather elaborate philosophical edifice in support of the claim that the theory is not only perfectly rational and comprehensible, but indeed complete. Their arguments, however, were never very convincing to the critics. For example, in 1949 (that is, well after the debates that led up to and followed the Schrödinger's Cat and EPR episodes of 1935), Einstein wrote that "the statistical quantum theory does not pretend to describe the individual system (and its development in time) completely". And so, Einstein said, "it appears unavoidable to look elsewhere for a complete description of the individual system...." From the point of view of a theory that *did* "accomplish a complete physical description", "the statistical quantum theory would ... take an approximately analogous position to the statistical mechanics within the framework of classical mechanics [1]."

The current chapter presents a concrete example of a theory of this sort – one which purports to *complete* the usual quantum mechanical description of physical states (by adding something new). The theory was first proposed, but then prematurely abandoned, by de Broglie in the mid 1920s [2]. The theory was then independently rediscovered and further developed by David Bohm in 1952 (and is therefore sometimes called "Bohmian Mechanics") [3]. Bell, who championed the theory until his untimely death in 1990, gave a very nice overview of its basic idea when he wrote:

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_7

> While the founding fathers agonized over the question
>
> ‘particle’ *or* ‘wave’
>
> de Broglie in 1925 proposed the obvious answer
>
> ‘particle’ *and* ‘wave’.
>
> Is it not clear from the smallness of the scintillation on the screen that we have to do with a particle? And is it not clear, from the diffraction and interference patterns, that the motion of the particle is directed by a wave? De Broglie showed in detail how the motion of a particle, passing through just one of two holes in [a] screen, could be influenced by waves propagating through both holes. And so influenced that the particle does not go where the waves cancel out, but is attracted to where they cooperate. This idea seems to me so natural and simple, to resolve the wave-particle dilemma in such a clear and ordinary way, that it is a great mystery to me that it was so generally ignored. Of the founding fathers, only Einstein thought that de Broglie was on the right lines. Discouraged, de Broglie abandoned his picture for many years. He took it up again only when it was rediscovered, and more systematically presented, in 1952, by David Bohm. .... There is no need in this picture to divide the world into ‘quantum’ and ‘classical’ parts. For the necessary ‘classical terms’ are available already for individual particles (their actual positions) and so also for macroscopic assemblies of particles [4].

Let’s try to understand in more detail what this theory says and how it works.

7.1 Overview

According to the de Broglie - Bohm pilot-wave theory, most of the puzzles and paradoxes of orthodox quantum mechanics arise from its using *incomplete* state descriptions. It is not, for example, that electrons are wave-like when not being observed, but then magically “collapse” to sharp positions when looked at. Instead, according to the pilot-wave theory, the electron is always a particle with a sharp position following a definite trajectory through space; the statistical wave-like phenomena (such as the build-up of the interference pattern in the two-slit experiment) arise because the motion of the particle is influenced by an associated wave. This is sometimes hard for people to understand because they are so accustomed, in ordinary quantum mechanics, to describing “particles” (like electrons) in terms of wave functions. So let me repeat it for emphasis: a single electron, according to the pilot-wave theory, is not one thing, but *two* – a wave *and* a (literal, pointlike) particle whose motion is controlled by the wave.

To formulate the theory in a rigorous way, we need to know the dynamical laws obeyed by both the wave and the particle. For the wave this is easy, because the wave is nothing but the ordinary quantum mechanical wave function Ψ obeying Schrödinger’s equation:

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}\Psi. \tag{7.1}$$

So that part is familiar and straightforward.

But what about the motion of the particle? Here we can take a clue from the de Broglie formula

$$p = \frac{h}{\lambda} = \hbar k \tag{7.2}$$

which relates the momentum p of the particle with the wavelength λ (or wave number k) of the associated wave. (Note that this equation is very difficult to understand unless there genuinely exist two things: a wave *and* a particle!) This suggests that when the wave function is a plane-wave $\Psi \sim e^{ikx}$ with a definite wave number k, the particle should move with velocity

$$v = \frac{p}{m} = \frac{\hbar}{m}k. \tag{7.3}$$

But what should the velocity be in the general case, where the wave function is not of this very special plane-wave type, and hence has no single well-defined wave number k?

The following seems like the simplest way of generalizing the last equation. Write the wave function in "polar form" $\Psi(x,t) = R(x,t)e^{iS(x,t)}$ (so that R is the modulus and S is the *phase* of the wave function) and then let

$$v = \frac{\hbar}{m}\frac{\partial S}{\partial x}. \tag{7.4}$$

For the plane-wave type solution, $S(x,t) = kx$ and so Eq. (7.4) reduces to Eq. (7.3). But Eq. (7.4) makes sense in general, for any $\Psi(x,t)$. Well, except for one thing: for a generic wave function $\Psi(x,t)$, the gradient of the phase $(\partial S/\partial x)$ will be a function of x and t. So where, exactly, should we evaluate the function to give the velocity of the particle? The obvious answer is: evaluate it at the actual location $X(t)$ of the particle! We will thus consider the following as the simplest possible candidate law describing how the particle moves under the influence of the wave:

$$\frac{dX(t)}{dt} = \frac{\hbar}{m}\left.\frac{\partial S(x,t)}{\partial x}\right|_{x=X(t)} \tag{7.5}$$

where $S(x,t)$ is the complex phase of the wave function. Note that this can be equivalently re-written (in terms of the wave function itself) as follows:

$$\frac{dX(t)}{dt} = \frac{\hbar}{m}\,\mathrm{Im}\left[\frac{\left(\frac{\partial\Psi}{\partial x}\right)}{\Psi}\right]\Bigg|_{x=X(t)}. \tag{7.6}$$

where "Im" means "the imaginary part".

That is basically all there is to the theory: a single electron (for example) is a wave *and* a particle, with the wave just obeying Schrödinger's equation and the particle moving, under the influence of the wave, according to Eq. (7.6). There is one other

aspect of the theory, though, that we will develop and explain here even though, in some sense, it is less fundamental. This has to do with how probabilities arise and are understood and explained in the theory.

Let's begin by reviewing/recalling an important fact about Schrödinger's equation, here, for simplicity, for a single particle moving in one dimension:

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V(x,t)\Psi. \tag{7.7}$$

The complex conjugate of Schrödinger's equation reads:

$$-i\hbar\frac{\partial\Psi^*}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi^*}{\partial x^2} + V(x,t)\Psi^*. \tag{7.8}$$

(Note that we assume here that the potential energy $V(x,t)$ is real.) Now multiply Eq. (7.7) by Ψ^*, multiply Eq. (7.8) by Ψ, and subtract the second from the first. The result is

$$i\hbar\left[\Psi^*\frac{\partial}{\partial t}\Psi + \Psi\frac{\partial}{\partial t}\Psi^*\right] = -\frac{\hbar^2}{2m}\left[\Psi^*\frac{\partial^2}{\partial x^2}\Psi - \Psi\frac{\partial^2}{\partial x^2}\Psi^*\right] \tag{7.9}$$

which can be simplified to

$$\frac{\partial}{\partial t}|\Psi|^2 = -\frac{\partial}{\partial x}\left[\frac{i\hbar}{2m}\left(\Psi\frac{\partial}{\partial x}\Psi^* - \Psi^*\frac{\partial}{\partial x}\Psi\right)\right]. \tag{7.10}$$

This has the form of the so-called "continuity equation"

$$\frac{\partial\rho}{\partial t} = -\frac{\partial}{\partial x}j \tag{7.11}$$

or, in three dimensions,

$$\frac{\partial\rho}{\partial t} = -\vec{\nabla}\cdot\vec{j}. \tag{7.12}$$

This continuity equation is satisfied, for example, in electrodynamics by the electric charge density ρ and the electric current density $\vec{j}$. In this context, we think of the continuity equation as expressing the local conservation of charge: a positive divergence of $\vec{j}$ at some point, which implies a net outward flow of electric charge away from that point, corresponds to a negative $\frac{\partial\rho}{\partial t}$, i.e., a decreasing charge density at the point.

So in the same way, Eq. (7.10) can be understood as expressing the local conservation of *probability* because we recognize $|\Psi|^2$ as the standard expression for the probability density (of finding the particle, if we look for it) in quantum mechanics. We thus identify

$$j = \frac{i\hbar}{2m}\left(\Psi\frac{\partial}{\partial x}\Psi^* - \Psi^*\frac{\partial}{\partial x}\Psi\right) \tag{7.13}$$

or, in three dimensions,

$$\vec{j} = \frac{i\hbar}{2m}\left(\Psi\vec{\nabla}\Psi^* - \Psi^*\vec{\nabla}\Psi\right) \tag{7.14}$$

as the "quantum probability current".

The fact that Schrödinger's equation implies that $|\Psi|^2$ obeys the continuity equation (with the $\vec{j}$ just given) is a completely standard (if slightly advanced) principle of orthodox quantum mechanics which has nothing in particular to do with the pilot-wave theory. But it relates to the pilot-wave theory in two (related!) ways.

First, recall that in electrodynamics the electrical current density (associated with, say, a single charged particle) is just the charge density multiplied by the particle's velocity: $\vec{j} = \rho\vec{v}$. And so one can write the velocity as the ratio of the current and charge densities like this:

$$\vec{v} = \frac{\vec{j}}{\rho}. \tag{7.15}$$

Now, in orthodox quantum mechanics, we have a probability density $\rho = |\Psi|^2$ and probability current $\vec{j}$. Of course, in *orthodox* quantum mechanics, there are no (literal) particles, but only wave functions. But – given orthodox quantum mechanics – if you wanted to propose that, in addition to the wave function, there is also a (literal) particle, it would be very natural and obvious – based on the analogy with electrodynamics – to guess that the velocity might be given by

$$\vec{v} = \frac{\vec{j}}{\rho} = \frac{i\hbar}{2m}\frac{\Psi\vec{\nabla}\Psi^* - \Psi^*\vec{\nabla}\Psi}{\Psi^*\Psi} \tag{7.16}$$

or, switching back to one dimension,

$$v = \frac{j}{\rho} = \frac{i\hbar}{2m}\frac{\Psi\frac{\partial}{\partial x}\Psi^* - \Psi^*\frac{\partial}{\partial x}\Psi}{\Psi^*\Psi}. \tag{7.17}$$

But it is now easy to see that this is yet another way of re-writing the equation, that we guessed above, for the velocity of the particle in the pilot-wave theory: since

$$\frac{\left(\Psi\frac{\partial}{\partial x}\Psi^* - \Psi^*\frac{\partial}{\partial x}\Psi\right)}{-2i} = \mathrm{Im}\left(\Psi^*\frac{\partial}{\partial x}\Psi\right) \tag{7.18}$$

Equation (7.17) becomes

$$v = \frac{\hbar}{m}\frac{\mathrm{Im}\left(\Psi^*\frac{\partial}{\partial x}\Psi\right)}{\Psi^*\Psi} = \frac{\hbar}{m}\mathrm{Im}\left[\frac{\left(\frac{\partial\Psi}{\partial x}\right)}{\Psi}\right] \tag{7.19}$$

which is the same as Eq. (7.6).

So that is the first reason for going into the quantum continuity equation here: it gives another illuminating perspective on the pilot-wave theory's new dynamical postulate (for how the particles should move).

The second reason is that it makes it possible to understand something about probability in the theory. In general, according to the pilot-wave theory, every particle is always definitely somewhere. But we do not usually know the exact location! Indeed, if we experimentally prepare an electron to have wave function $\Psi(x, 0)$, we cannot pick or control the exact initial particle position $X(0)$ – this will therefore be *random*, and one might expect that there should be some associated probability distribution $P(x, 0)$ to characterize this. But since we have already committed to a specific formula for the velocity that particles at various positions x and times t would have, it is clear that the initial probability distribution $P(x, 0)$ will change in time. It is possible to show that the probability distribution $P(x, t)$ should satisfy

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[\, v(x,t)\, P(x,t)\,\right] \tag{7.20}$$

(see the Projects). But then one can see that $P(x, t) = |\Psi(x, t)|^2$ is a special, equilibrium probability distribution for the pilot-wave theory, in the following sense: if $P(x, 0) = |\Psi(x, 0)|^2$ at the initial time ($t = 0$), then we will have $P(x, t) = |\Psi(x, t)|^2$ for all times. This is often described in the literature by saying that the distribution $P = |\Psi|^2$ is "equivariant". The proof of this is just the following: if $P = |\Psi|^2$ and $v = j/|\Psi|^2$, then Eq. (7.20) reduces to the continuity equation, Eq. (7.10), which we already showed is satisfied as a consequence of Schrödinger's equation.

There is a lot more that can be said about how to understand the quantum probabilities in the pilot-wave theory. One of the theory's main virtues is that something like the Born rule can be genuinely *derived* (from the basic dynamical postulates of the theory) rather than merely posited as an additional axiom. What we have just been explaining is a part (but only a part) of that derivation, but it would be too big a distraction to go any deeper into this. So for our purposes here it will have to suffice to think about the theory in something like the following way: at some cosmological initial time $t = 0$, the wave function was Ψ_0 and the particle positions were selected, randomly, according to the $|\Psi|^2$ distribution at that initial time. It then follows, from the two dynamical postulates of the theory, that we – today, inside the universe – should see particle positions that are distributed according to the Born rule: $P(x, t) = |\Psi(x, t)|^2$.

We will illustrate these ideas with a concrete example in the following section.

7.2 Particle in a Box

Let's try to see more clearly how the pilot-wave theory works by considering the simple example of a (one-dimensional) particle-in-a-box (PIB). Suppose to begin with that the system is in the ground state so that

$$\Psi(x,t) = \psi_1(x)e^{-iE_1t/\hbar}. \tag{7.21}$$

Since $\psi_1(x) = \sqrt{\frac{2}{L}}\sin(\pi x/L)$ is purely real, the complex phase associated with Ψ is just

$$S(x,t) = -iE_1t/\hbar. \tag{7.22}$$

This doesn't depend on x at all, so the particle velocity, according to Eq. (7.5), is *zero*. The particle, that is, just sits there at rest. This, as it turns out, is characteristic of so-called stationary states, which are indeed aptly named according to this theory.

Note that this applies also to some more interesting and realistic situations: for example, the electron in a Hydrogen atom in its ground state is not, according to the pilot-wave theory, orbiting the proton, but is instead just sitting there, at some fixed point near the proton. If that bothers you or seems physically impossible, you are probably tacitly expecting that if the electron is literally a particle, it should obey Newton's equations of motion, and should therefore accelerate toward the proton due to the electrostatic force. But the pilot-wave theory is not classical mechanics! The motion of the particle, according to this theory, is not determined by classical forces acting on it, but is instead determined by the structure of the wave function which guides it.

To see some non-trivial dynamics in the particle-in-a-box system, we need only let the quantum state be a superposition of energy eigenstates. For example, suppose the wave function is given by

$$\begin{aligned}\Psi(x,t) &= \frac{1}{\sqrt{2}}\left[\psi_1(x)e^{-iE_1t/\hbar} + \psi_2(x)e^{-iE_2t/\hbar}\right] \\ &= \frac{1}{\sqrt{L}}\left[\sin(\pi x/L)e^{-i\omega_1 t} + \sin(2\pi x/L)e^{-i\omega_2 t}\right].\end{aligned} \tag{7.23}$$

It is slightly cumbersome here to put this in polar form, but straightforward to use Eq. (7.6) to express the particle velocity as a function of its position:

$$\frac{dX(t)}{dt} = \frac{\hbar}{m}\mathrm{Im}\left[\frac{\frac{\pi}{L}\cos\left(\frac{\pi x}{L}\right)e^{-i\omega_1 t} + \frac{2\pi}{L}\cos\left(\frac{2\pi x}{L}\right)e^{-i\omega_2 t}}{\sin\left(\frac{\pi x}{L}\right)e^{-i\omega_1 t} + \sin\left(\frac{2\pi x}{L}\right)e^{-i\omega_2 t}}\right]_{x=X(t)}. \tag{7.24}$$

This is a bit of a messy first-order differential equation, but it's easy enough to let Mathematica solve it numerically. See Fig. 7.1 for some example world lines.

Basically, what happens is that – as the wave intensity sloshes back and forth within the box (as we saw in Chap. 2) – the particle is pushed back and forth with it. The ensemble of trajectories in Fig. 7.1 is, however, a little uneven and funny-looking because we have chosen an ensemble in which the initial positions $X(0)$ are equally spaced, i.e., the initial distribution $P(x)$ is constant. One can see in the figure that the distribution then becomes very non-constant (with several trajectories bunching closely together) after a short period of time.

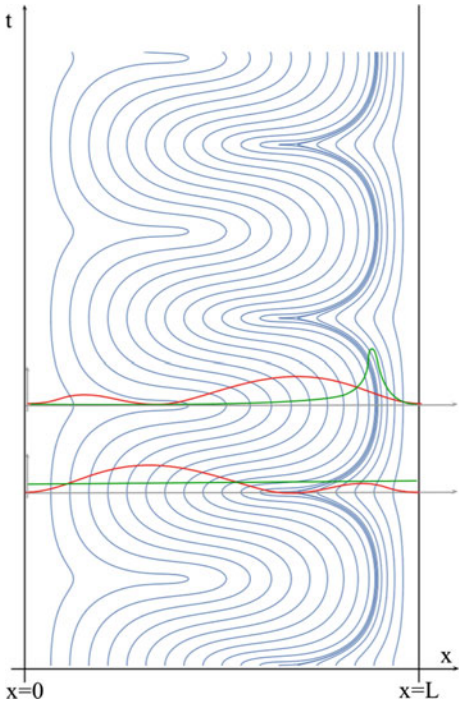


Fig. 7.1 The *blue curves* are a set of possible worldlines for a particle-in-a-box with wave function an equally-weighted superposition of the $n = 1$ and $n = 2$ energy eigenstates. (This is a space-time diagram, with the horizontal axis being the position x within the box, and the vertical axis representing the time t.) Note that at $t = 0$ the example trajectories are equally-spaced across the box. If we think of this as an ensemble of trajectories, we would say that the initial distribution $P(x)$ is constant. But then the distribution at later times is not constant, as illustrated by the rather extreme "clumping up" of the world lines. The distribution P is graphed, as a function of x, at two different times: see the *green curves* that live on the grey axes, whose vertical location is meant to indicate the time. The associated *red curves* show what $|\Psi|^2$ looks like at these same times

As described in the last section, however, there is a special distribution whose functional form is preserved in time. This is the distribution $P = |\Psi|^2$ in which the number of trajectories in the ensemble is proportional to the "intensity" $|\Psi|^2$ of (i.e., what is in orthodox QM thought of as the "probability density" associated with) the wave function Ψ. The claim, then, is that if we have an ensemble of particles (all moving under the influence of the same wave function $\Psi(x, t)$) with, at $t = 0$, the distribution $P(x, 0) = |\Psi(x, 0)|^2$, then it follows from Eqs. (7.1) and (7.6) that $P(x, t) = |\Psi(x, t)|^2$ for all t. As mentioned before, this property is sometimes called the "equivariance" of the $|\Psi|^2$ distribution.

The equivariance property is illustrated in Fig. 7.2, which is the same as Fig. 7.1 except that now the distribution of initial positions $X(0)$ is given by $|\Psi(x, 0)|^2$. One can see that, indeed, the distribution continues to be given by $|\Psi(x, t)|^2$ for later times.

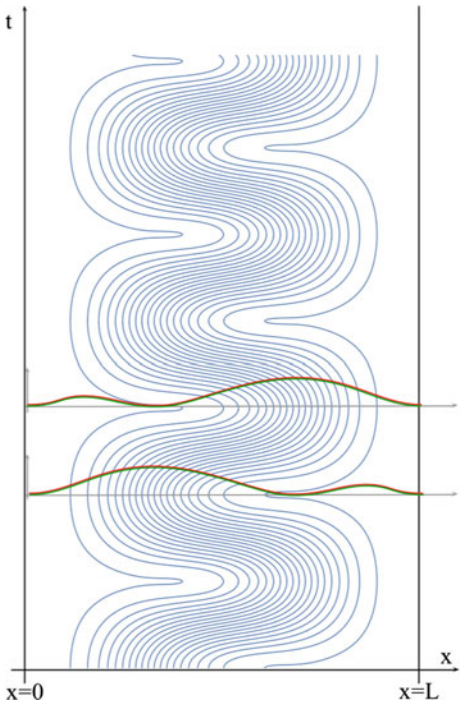


Fig. 7.2 Same as Fig. 7.1 but for an ensemble of initial positions $X(0)$ that are distributed with $P(x,0) = |\Psi(x,0)|^2$. This illustrates the "equivariance" property discussed in the previous section: if the positions of particles in the ensemble are $|\Psi|^2$-distributed at $t = 0$, then they will remain $|\Psi|^2$-distributed for all time. So the *green curves*(indicating P) and the *red curves*(indicating $|\Psi|^2$) coincide at all times here, unlike the situation depicted in the previous figure

So, this example illustrates all of the main ideas of the pilot-wave theory: a quantum system is a hybrid of particle-and-wave, with the wave being simply the usual wave function obeying Schrödinger's equation. The particle has a random initial position within the wave, and this position then evolves in time according to the guidance equation (which we have written in several mathematically equivalent forms). The motion of the particle is indeed well-captured by Bell's statement that the particle "is attracted to where [the contributions to Ψ] cooperate", i.e., where there is constructive interference, i.e., where $|\Psi|^2$ is large. For example, here, at $t = 0$ $|\Psi|^2$ is large on the left side of the box and small on the right, but after a short period of time $|\Psi|^2$ becomes small on the left and large on the right; the particles thus move from left to right to "follow" $|\Psi|^2$.

7.3 Other Single Particle Examples

Let's review a couple of other examples to get a sense of how the theory works. Consider, to start, the spreading Gaussian wave packet from Chap. 2. We saw that if, at $t = 0$, the wave function is given by

$$\Psi(x,0) = Ne^{-x^2/4\sigma^2} \tag{7.25}$$

then

$$\Psi(x,t) = N(t)\, e^{-x^2/4(\sigma^2 + i\hbar t/2m)}. \tag{7.26}$$

where $N(t)$ is a time- (but not position-) dependent complex normalization constant. We can put this in "polar form" by multiplying (inside the argument of the exponential) by the complex conjugate of $(\sigma^2 + i\hbar t/2m)$ divided by itself. This gives

$$\Psi(x,t) = N(t)\, \exp\left[\frac{-x^2\sigma^2}{4(\sigma^4 + \hbar^2 t^2/4m^2)}\right] \exp\left[\frac{ix^2\hbar t}{8m(\sigma^4 + \hbar^2 t^2/4m^2)}\right]. \tag{7.27}$$

So we can identify the complex phase $S(x,t)$ of the wave function as[1]

$$S(x,t) = \frac{x^2\hbar t}{8m(\sigma^4 + \hbar^2 t^2/4m^2)}. \tag{7.28}$$

Plugging this into Eq. (7.5) gives the following first-order differential equation for the position $X(t)$ of a particle being guided by this spreading Gaussian packet:

$$\frac{dX(t)}{dt} = X(t)\, \frac{t}{t^2 + 4m^2\sigma^4/\hbar^2}. \tag{7.29}$$

It is not hard to show that this differential equation is solved by

$$X(t) = X_0 \left(1 + \frac{t^2}{4m^2\sigma^4/\hbar^2}\right)^{1/2} \tag{7.30}$$

which can be re-written as

$$\left(\frac{X(t)}{X_0}\right)^2 - \left(\frac{t}{2m\sigma^2/\hbar}\right)^2 = 1. \tag{7.31}$$

This is the equation for a *hyperbola*, and so it turns out that the spreading Gaussian wave packet has the nice feature that the world lines of particles are hyperbolae. Some example trajectories are shown in Fig. 7.3.

To summarize this first example, when an initially narrow wave packet *spreads*, as of course occurs for example in the phenomenon we call *diffraction*, according to the pilot-wave theory the possible particle trajectories also spread out from one another, as one would expect on the basis of the equivariance property.

---

[1] Technically, there is also a contribution to the complex phase from what I called $N(t)$, but since that only depends on time, and we ultimately only care about the derivative of $S(x,t)$ with respect to x, I am just ignoring that other contribution. What's written here, then, is really just the x-dependent part of $S(x,t)$.

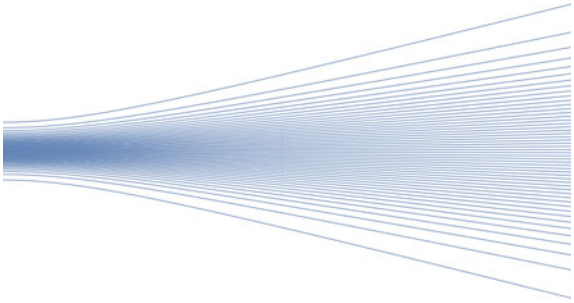


Fig. 7.3 Representative sample of particle trajectories for a spreading Gaussian wave packet. Here time runs to the *right* and x runs vertically (so it is a space-time diagram turned sideways). Or one can, in the spirit of Fig. 2.6, consider replacing t with a second spatial coordinate, and hence think of the lines as showing the trajectories through space that particles would follow downstream of a single Gaussian slit. That is, the figure can be understood as showing the trajectories that particles would follow when being guided by a diffracting wave function. Note that the distribution of initial particle positions $X(0)$ here is given by $|\Psi(x,0)^2|$, so (by "equivariance") the trajectories spread out from one another so as to keep $P = |\Psi|^2$ for subsequent times

We can use a similar technique to visualize the possible particle trajectories in the case of two-slit interference. Beginning with a superposition of two Gaussian wave packets (centered at $x = a$ and $x = -a$) we showed already in Chap. 2, Eq. (2.57), that

$$\Psi(x,t) = N(t)\left[e^{-\frac{(x-a)^2}{4(\sigma^2+i\hbar t/2m)}} + e^{-\frac{(x+a)^2}{4(\sigma^2+i\hbar t/2m)}}\right]. \tag{7.32}$$

This is a little harder to write explicitly in "polar" form (although there is a reasonably simple way of writing $S(x,t)$ explicitly). And the differential equation one gets for $X(t)$ has nothing as simple as hyperbolae as solutions. The only hope, really, is to solve the differential equation for $X(t)$ numerically using a computer. See Fig. 7.4 for the beautiful results!

There are a number of other simple example scenarios for which it is illuminating to consider the particle trajectories as posited by the pilot-wave theory. See, for example, the Projects for references to two papers which analyze (i) a simple case of reflection and transmission at a potential step and the case of quantum tunneling through a classically-forbidden region and (ii) the pilot-wave theory's account of *spin* and its measurement in for example a Stern–Gerlach type apparatus. But hopefully the examples discussed above already give you a fairly clear sense of how the theory works in simple situations. So let us then turn to exploring some other important aspects of the theory.

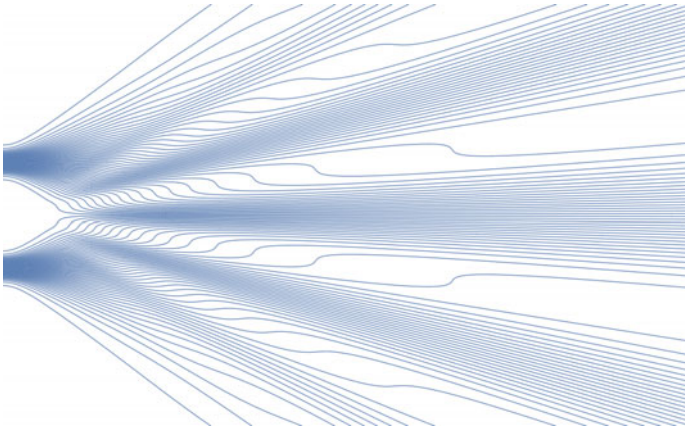


Fig. 7.4 Representative sample of particle trajectories for the case of two initially-separated Gaussian wave-packets. As in the previous figure, this is technically a space-time diagram turned sideways – but one may also legitimately think of it as showing the trajectories, through space, of particles which have just emerged, moving to the *right*, through a double- (Gaussian) slit screen. This type of image, of the particle trajectories for the double-slit experiment according to the pilot-wave theory, was first presented in Ref. [5] and has become iconic for the pilot-wave theory because it captures so clearly, in a picture, how the theory explains the (otherwise) puzzling wave-particle-duality. The discrete flashes on the detection screen correspond to places where (literal, pointlike) particles collide with the screen; but the highly non-classical motion of the particles is influenced by the accompanying pilot-wave such that the particle trajectories bunch up around points of constructive interference. An ensemble of such trajectories (with suitably random initial conditions) will therefore perfectly account (in, to use Bell's phrase, "a clear and ordinary way" [4]) for the type of statistical interference pattern we saw in Fig. 2.8

7.4 Measurement

We assumed, in our discussion of the one-particle examples above, that if we make a position measurement at some time t when the wave function is $\Psi(x, t)$ and the actual particle position is $X(t)$, we will see the particle where it *is*. For example, if a particular particle in the double slit experiment is following one of the trajectories shown in Fig. 7.4, we will see a "flash" on the detection screen right where the particle runs into it, i.e., at the location where the trajectory it's following intersects the screen (on, for example, the right of the figure).

But probably the most important virtue of the pilot-wave theory is that we do not need to divide up the world into "quantum system" (which we describe using the theory) and "classical environment" (which we take for granted and make uncontrolled assumptions about) in order to understand measurements and their outcomes. Instead, we are free (indeed, required!) to enlarge the "quantum system" (which we describe using the theory) until it includes literally everything – the entire universe. This is of course in contrast to ordinary quantum mechanics which, as we discussed in detail in Chap. 3, seems to require one to introduce what Bell called a "shifty

split" (i.e., an artificial division of the world into distinct "quantum" and "classical" realms, with special ad hoc exceptions to the usual dynamical rules when the two realms interact). The claim, then, is that unlike orthodox quantum mechanics, the pilot-wave theory is not afflicted with a "measurement problem."

Let us discuss this in terms of the simple example, from Chap. 3, in which the energy of a particle-in-a-box (with degree of freedom x) is measured, and the outcome displayed in the position of a "pointer" (with degree of freedom y). As discussed back in that chapter, a schematic interaction Hamiltonian

$$\hat{H}_{int} = \lambda \hat{H}_x \hat{p}_y \tag{7.33}$$

can be shown to generate the expected kind of behavior, namely, that if the initial wave function is given by

$$\Psi(x, y, 0) = \psi_n(x)\phi(y) \tag{7.34}$$

(where $\psi_n(x)$ is the nth energy eigenstate of the particle-in-a-box and $\phi(y)$ is a gaussian wave packet centered at $y = 0$, the "ready" position of the pointer), Schrödinger's equation will give the time-evolved wave function

$$\Psi(x, y, t) = \psi_n(x)\phi(y - \lambda E_n t) \tag{7.35}$$

in which the wave packet for the pointer has *moved* a distance proportional to the energy E_n of the particle in the box. In short, the post-interaction position of the pointer registers the actual energy E_n of the particle-in-the-box... just as it should if the process is going to be described as a measurement of that energy!

The trouble arose when we considered what happens if the particle-in-a-box starts out in a superposition of different energy eigenstates. From the linearity of Schrödinger's equation, it is clear that if

$$\Psi(x, y, 0) = \left[\sum_i c_i \psi_i(x)\right] \phi(y) \tag{7.36}$$

then we will have

$$\Psi(x, y, t) = \sum_i c_i \psi_i(x)\phi(y - \lambda E_i t). \tag{7.37}$$

That is, instead of having a well-defined post-interaction position which we can interpret as registering the (single, well-defined) outcome of the measurement, the pointer itself becomes "infected" with the quantum superposition. That is, the final state is a superposition of terms like: "the particle is in the ground state and the pointer indicates $E = E_1$", but also "the particle is in the first excited state and the pointer indicates $E = E_2$", and so on. The wave function alone fails to pick out a particular result; instead it contains, so to speak, all possible results in parallel. But since in an

actual measurement of this kind we always *observe* a single, definite result, it seems that the wave function alone is inadequate to account for our observations. That, in a nutshell, was the measurement problem.

How does the pilot-wave theory resolve the problem, given that, as we have said, the theory *also* says that there is a wave function which obeys Schrödinger's equation? It is true that, according to the pilot-wave theory, the PIB-pointer system has a wave function that ends up in the state described by Eq. (7.37). But the crucial idea is that, according to the pilot-wave theory, *the wave function alone does not provide a complete description of the physical situation*. There is, in addition, the actual position $X(t)$ of the PIB and – crucially here – the actual position $Y(t)$ of the pointer.

Let's think a little bit about what the theory says these actual particle positions do. The details are a little bit complicated (mostly because of the somewhat unusual form of the interaction Hamiltonian) so I'll save those for the Projects at the end of the chapter. But, in principle, the idea is simple: the position $X(t)$ of the particle-in-the-box evolves according to

$$\frac{dX(t)}{dt} = \frac{j_x}{|\Psi|^2} \tag{7.38}$$

where j_x is the x-component of the quantum probability current associated with this system. Similarly, the position $Y(t)$ of the pointer evolves according to

$$\frac{dY(t)}{dt} = \frac{j_y}{|\Psi|^2} \tag{7.39}$$

where j_y is the y-component of the probability current. And at some level you don't really need to worry about what, exactly, the trajectories look like, because you know – from the equivariance property – that if the initial values $X(0)$ and $Y(0)$ start out random (and suitably distributed) the positions will remain $|\Psi|^2$-distributed for all times. And so the positions $X(t)$ and $Y(t)$ later on will be random and they will in particular lie somewhere in the support of $\Psi(x, y, t)$. Let's think qualitatively about what that means.

Figure 7.5 shows a "cartoon" representation of how the wave function $\Psi(x, y, t)$ evolves in time as the measurement interaction proceeds. At $t = 0$, the wave function has support between $x = 0$ and $x = L$ and for $y \approx 0$. But as time goes on, the superposed terms in Ψ move, in the y-direction, by different amounts, so that after a while the wave function has support in a set of discrete, non-overlapping "islands" of the configuration space. But the *particles* have definite positions $X(t)$ and $Y(t)$ which, together, can be understood as defining an "actual configuration point" which traces out some kind of trajectory through configuration space. (I've indicated a possible beginning and end to this trajectory in the figure by putting a dot at $\{X(0), Y(0)\}$ on the left and $\{X(t), Y(t)\}$ on the right.) The exact trajectory will be, in general, rather twisted and complicated: as the several wave function "islands" slide across one another (in the process of going from their initial, stacked configuration, to their final, separated, configuration) there is a complicated interference pattern, somewhat

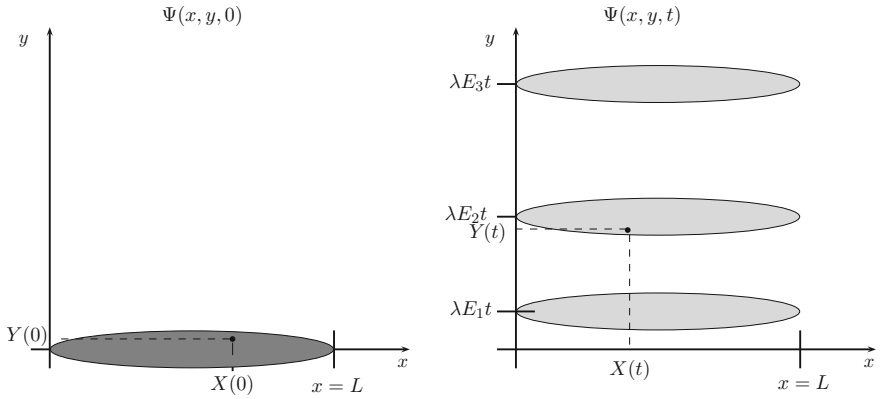


Fig. 7.5 The graph on the *left highlights* (in *dark gray*) the region of the two-dimensional configuration space where $\Psi(x, y, 0)$ has support. Later, at time t, the wave function has split apart into several non-overlapping "islands". This is depicted in the graph on the *right*. The simultaneous presence of all these islands constitutes, for orthodox quantum mechanics, the measurement problem. But for the pilot-wave theory, the actually-realized outcome of the measurement is not to be found in the wave function, but rather in the final position of the pointer. And this, in the pilot-wave theory, will be some one (random but perfectly definite) value, indicated here by the vertical position $Y(t)$ of the *dot* which represents the actual configuration point (X, Y). The indicated $Y(t)$ is in the support of the $n = 2$ branch of the wave function – i.e., $Y(t)$ is approximately $\lambda E_2 t$ – so we would say in this case that the energy measurement had the outcome $E = E_2$. Note that the outcome might have been different had the (random) initial positions $X(0)$ and $Y(0)$ been different

like the one that makes the particle trajectories bend this way and that in the double slit setup. But as the "islands" cease to overlap, things calm down, and the actual configuration point $\{X(t), Y(t)\}$ finds itself in one or the other of the islands. And so, in particular, the pointer has some specific post-interaction position, $Y(t)$ which is either approximately $\lambda E_1 t$, or $\lambda E_2 t$, or $\lambda E_3 t$, etc. So, at the end of the experiment, the position of the pointer is in no way "blurry" or "indefinite" or "superposed". The pointer has a definite position which corresponds to exactly *one* of the energy values we regard as possible outcomes of the experiment.

Indeed, note that, by the equivariance property (and, again, assuming the initial positions $X(0)$ and $Y(0)$ are suitably random), the probability for the final configuration point $\{X(t), Y(t)\}$ to lie in the nth "island" is equal to the integral of $|\Psi|^2$ across that island. But this is simply $|c_n|^2$ – the very thing we would have identified, according to the textbook quantum rules, as the probability for the measurement to have the n^{th} outcome. So the pilot-wave theory reproduces the statistical predictions of ordinary QM. But it does this while treating the system-being-measured and the measuring apparatus itself, on an equal footing, both as part of a big system that is described in a uniform way by the theory.

It is perhaps also worth stressing here that the pilot-wave theory generates the same *statistical* predictions as ordinary QM, even though the theory is completely *deterministic*. Recall that in ordinary QM, the Schrödinger equation part of the dynamics is

completely deterministic; it is the collapse postulate (i.e., the dynamics that momentarily pre-empts the usual Schrödinger evolution when a "measurement" occurs) that introduces the randomness which results in the theory making statistical predictions. In the pilot-wave theory, there is no collapse postulate: the wave function obeys the Schrödinger equation *always*. And the dynamics describing the motion of the particles is also completely deterministic. Randomness enters – like in classical statistical mechanics – only through the initial conditions. Basically, we can never be sure (in advance) how a quantum measurement will come out, because we can never know (in sufficient microscopic detail) what the initial positions of all the particles were.

There is much more that can be said about this issue. It turns out, for example, that the uncertainty (associated with initial particle positions) is *unavoidable*. If, for example, you prepare a system to have wave function $\psi(x)$, then the *best possible knowledge* that you could in principle have, about the position X of the associated particle, is (according to the theory) that $P(x) = |\psi|^2$. Thus, although lots of detailed microscopic structure *exists*, according to the pilot-wave theory – for example, every particle in the universe has at every moment a precisely defined position and velocity – not all of this structure is *accessible* to us. The theory thus allows us to understand Heisenberg's uncertainty principle as just that – uncertainty about facts which exist. This is in contrast to the usual interpretation, in which, to be precise about it, it is not so much that we are *uncertain* about these things, but that the things themselves are objectively *indefinite*.

Let us develop one more point which ties some of these ideas together. We have been stressing that in the pilot-wave theory one can – and indeed must – treat *everything* (ultimately, the whole universe) quantum mechanically. So for example when a measuring apparatus interacts with some particle (one of whose properties is being "measured") we must solve Schrödinger's equation for the entire system comprising both the particle and the apparatus. And as we saw, no additional ad hoc postulates (about "collapse", etc.) are needed to explain how the measurement has a single, definite outcome: although the wave function (for the whole big system) ends up in an entangled superposition, the particles (being, after all, literal particles) have definite positions all the time.

That's great if one just wants to understand why there is no "measurement problem" in the pilot-wave theory. But the broader claim is that the pilot-wave theory reproduces all of the statistical implications of ordinary textbook quantum theory. And ordinary textbook quantum theory generally does not include measuring devices in the systems being described quantum mechanically, but instead just talks about (for example) the PIB whose energy is being measured. And one of the things it says about the PIB is that the PIB has its own wave function (both before and after the measurement) and that, during the measurement, this wave function *collapses*. So this naturally raises the question: even though there is nothing like a collapse *postulate* in the pilot-wave theory, can the theory nevertheless shed some light on the success and utility of the textbook quantum rules?

The answer is yes, and this is one of the most interesting (and also least appreciated) aspects of the theory. To begin with, note that even though there is in some sense really only one big wave function in the pilot-wave theory (namely, the wave

function of the whole universe) the theory allows one to define the wave function for a sub-system. The way to do this is as follows: just evaluate the universal wave function at the actual locations of all the particles *outside* the subsystem. For example, take the case of the PIB-pointer system we've been discussing here. The wave function for the whole PIB-pointer system is

$$\Psi(x, y, t) = \sum_i c_i \psi_i(x) \phi(y - \lambda E_i t). \tag{7.40}$$

By simply evaluating this at $y = Y(t)$ – the actual location of the pointer particle – one thus has a function which depends only on x and t and can be understood as the wave function, call it $\chi(x, t)$ of the PIB sub-system:

$$\chi(x, t) \sim \sum_i c_i \psi_i(x) \phi(Y(t) - \lambda E_i t). \tag{7.41}$$

The "$\sim$" is there, instead of an "$=$" sign, because it would probably be sensible to define the sub-system wave function in such a way that it is properly normalized. The RHS, however, is not. But this is a minor technical detail that we simply leave aside for now.

Here is the important thing. At $t = 0$ (or, in general, before the interaction with the measuring apparatus is turned on), the PIB sub-system wave function is

$$\chi(x, 0) \sim \sum_i c_i \psi_i(x)\, \phi\,(Y(0)) \sim \sum_i c_i \psi_i(x) \tag{7.42}$$

since $\phi\,(Y(0))$ is just a constant that doesn't depend on x or i. This is just what we would ordinarily have said the PIB's pre-measurement wave function is, according to the textbook theory. So that is not too interesting or surprising. But consider what happens for large t (after, say, the measurement has gone to completion). Recall in particular that the actual pointer position $Y(t)$ ends up (at random, depending on the uncontrollable initial conditions) *either* near $\lambda E_1 t$ *or* $\lambda E_2 t$ *or* $\lambda E_3 t$, etc. That is, $Y(t) \approx \lambda E_n t$ for some particular n which we describe as the actual outcome of the experiment. But since ϕ is something like a narrow Gaussian wave packet, this means that $\phi(Y(t) - \lambda E_i t)$ will be approximately zero for all values of i except $i = n$, the one corresponding to the realized outcome. And so this means that, for large t, the PIB sub-system wave function can be written

$$\chi(x, t) \sim \sum_i c_i \psi_i(x) \phi(Y(t) - \lambda E_i t) \approx c_n \psi_n(x) \phi(Y(t) - \lambda E_n t). \tag{7.43}$$

But this is equivalent to saying

$$\chi(x, t) = \psi_n(x) \tag{7.44}$$

since c_n and $\phi\,(Y(t) - \lambda E_n t)$ are again just constants that don't depend on x.

Thus, the wave function of the PIB sub-system evolves, during the course of the interaction with the measuring apparatus, from a superposition of several energy eigenstates, into the one particular eigenstate that corresponds to the realized outcome of the experiment. In fact this evolution is perfectly smooth and continuous; but if the interaction is strong, the evolution will occur rapidly, and one might be forgiven for describing it as apparently discontinuous. The point is, of course, that here the pilot-wave theory is providing an *explanation* for the process that is described in ordinary quantum theory as the collapse of the wave function. But whereas in ordinary quantum mechanics the collapse is an implausible, ad hoc exception to the usual dynamical rules, the transition of sub-system wave functions to appropriate eigenstates during suitable interactions is, in the pilot-wave theory, a *consequence* of the standard dynamical rules that apply all of the time.

7.5 Contextuality

In the earlier section, when we were thinking about the pilot-wave theory's account of the 2-slit experiment, we assumed that the visible "flash" on the detection screen occurs where the particle in fact hits the screen. We assumed, that is, that position measurements simply reveal the pre-existing positions of the particles. But the analogous thing does not appear to hold in the case of the energy measurement we discussed subsequently. The "measurement of the PIB's energy" had, to be sure, a definite outcome – E_n – but this in no way implied that the PIB somehow secretly had this particular amount of energy prior to the measurement interaction. Indeed, it's not even really clear what that would mean according to the pilot-wave theory: prior to the interaction with the measuring device, the PIB's wave function was a superposition of several energy eigenstates, and the particle had some definite position X within that wave; but there is simply nothing there that would allow us (or should make us feel the urge to) attribute some definite pre-measurement energy to the particle. It seems, instead, more reasonable to summarize the situation by saying that the PIB doesn't really have any particular energy prior to the measurement, although it does have one *after* the measurement.

Here is another example of how, in the pilot-wave theory, measurements do not necessarily just passively reveal pre-existing values. We mentioned at the beginning of this chapter that, for an electron in the ground state of Hydrogen or a particle-in-a-box potential, the (literal, pointlike) particle will be motionless. The same will be true for an electron in the ground state of a simple harmonic oscillator potential; let us analyze this case in some detail using some bits of mathematics that have already been worked out for other purposes.

Thus, consider an electron moving in one dimension which experiences the potential energy

$$V(x) = \frac{1}{2} m\omega^2 x^2. \tag{7.45}$$

The lowest-energy solution of the time-independent Schrödinger equation is a Gaussian wave function

$$\psi(x) = Ne^{-x^2/4\sigma^2} \tag{7.46}$$

where the width σ of the wave packet is related to the (classical angular) frequency ω of the oscillator and the mass m of the particle through

$$\sigma^2 = \frac{\hbar}{2m\omega}. \tag{7.47}$$

The energy eigenvalue of this state is $E = \frac{1}{2}\hbar\omega$.

If the electron is in this ground state, its wave function will be

$$\psi(x, t) = Ne^{-x^2/4\sigma^2} e^{-iEt/\hbar}. \tag{7.48}$$

The complex phase $S(x, t)$ depends on t only and so, like the earlier examples, Eq. (7.4) implies that the velocity of the (literal, pointlike) particle is zero, regardless of its precise location X within the Gaussian wave packet.

And this of course means that the *momentum* of the particle is zero as well, assuming that by "the momentum of the particle" we just mean its mass multiplied by its instantaneous velocity: $p = m\frac{dX}{dt}$. But this should be slightly troubling since, as discussed in Chap. 2, the generalized Born rule implies that a measurement of the momentum of the electron in this situation is exceedingly *unlikely* to yield the value $p = 0$. Recall in particular that the Gaussian wave function $\psi(x) = Ne^{-x^2/4\sigma^2}$ can be written as a linear combination of momentum eigenstates $\psi(x) = \int \phi(k)\frac{e^{ikx}}{\sqrt{2\pi}}dk$ with

$$\phi(k) = \sqrt{2}N\sigma\, e^{-k^2\sigma^2}. \tag{7.49}$$

This, according to the generalized Born rule, implies that the probability for a momentum measurement to yield a value between p and $p + dp$ is

$$\begin{aligned} P(p)\,dp &= P(k)\,dk \\ &= |\phi(k)|^2 dk \\ &= 2N^2\sigma^2\, e^{-2k^2\sigma^2}\, dk \\ &= \frac{2N^2\sigma^2}{\hbar}\, e^{-2p^2\sigma^2/\hbar^2}\, dp \end{aligned} \tag{7.50}$$

where we have used $p = \hbar k$ to relate the wave number k to the momentum p.

It is thus clear that, if the pilot-wave theory is going to be able to reproduce the usual quantum statistical predictions, it cannot be that momentum measurements simply reveal the pre-existing momentum! And, of course, it turns out that they do not. To understand in detail what the theory does say, about how such measurements will come out, we just need to analyze the measurement procedure in detail, using the theory.

This particular example lends itself well to imagining a so-called "time-of-flight" procedure for measuring the momentum. The idea here is that, to determine the momentum of the electron, one could "turn off" the potential energy $V(x)$ which confines the electron to the vicinity of the origin, let the particle fly freely away from the origin for a long time, observe its *position*, and then infer what the momentum must have been to allow it to arrive at that position.

We have already worked through all of the mathematics required to analyze this type of momentum measurement. For example, we saw in Sect. 7.3 that, for a free particle whose wave function is, at $t = 0$, $\psi(x, 0) = N\, e^{-x^2/4\sigma^2}$, the particle trajectories are given by Eq. (7.30). Since our momentum measurement procedure involves letting the particles fly freely for a very long time, it is sufficient to take the large t limit in which

$$X(t) \approx X_0 \frac{\hbar t}{2m\sigma^2}. \tag{7.51}$$

If the particle is observed at $X(t)$ at time t, we will infer that its velocity has been $v = X(t)/t$. Thus, according to the pilot-wave theory, a particle whose initial ($t = 0$) position was X_0 will produce a measured momentum value

$$p = mv = \frac{X_0 \hbar}{2\sigma^2}. \tag{7.52}$$

Qualitatively, one sees how this particular method of measuring the momentum of the electron yields non-zero values even though the electron's momentum was, just prior to the initiation of the measurement procedure, zero: turning off the confining potential energy changes the subsequent time-evolution of the electron's wave function, which in turn causes the particle to acquire a non-zero momentum! (Or, at least, this is what happens if, as is overwhelmingly probable, the particle's initial position does not happen to be precisely $X_0 = 0$.) The outcome of the measurement does indeed in some sense come into existence as a result of the measurement intervention. But the process by which this occurs is clear and comprehensible and governed by the same quantum laws that (according to the pilot-wave theory) always apply.

It is easy to check also that the quantitative statistics work out correctly. Using Eq. (7.52) to relate the measured momentum value p to the initial position X_0 of the particle within the wave packet, we may assert that the probability for the momentum measurement to yield a value between p and $p + dp$ is equal to the probability that the initial position of the particle was in the range that would lead to those outcomes. But then we know how to express that probability in terms of the initial wave function. Putting these pieces together, we find that

$$\begin{aligned} P(p)\, dp &= P(X_0)\, dX_0 \\ &= |\psi(X_0, 0)|^2\, dX_0 \\ &= N^2\, e^{-X_0^2/2\sigma^2}\, dX_0 \\ &= \frac{2N^2\sigma^2}{\hbar}\, e^{-2p^2\sigma^2/\hbar^2}\, dp. \end{aligned} \tag{7.53}$$

This is precisely the same Gaussian distribution of p-values that we found, in Eq. (7.50), was predicted by the generalized Born rule of ordinary quantum mechanics. So the pilot-wave theory not only explains qualitatively how a particle whose pre-measurement momentum is zero can nevertheless be measured to have a non-zero momentum, but it precisely agrees with ordinary quantum mechanics about the precise statistical distribution of those measured values.

This example nicely illustrates the point that, for measurable quantities other than position, measurements according to the pilot-wave theory do not just passively real the pre-existing value of the quantity in question. This is part of what is meant by saying that, for the pilot-wave theory, properties like momentum (and energy and spin) are "contextual".

But this notion of "contextuality" goes a little bit deeper. It is not just that the result of a measurement of a certain property can in general be different from the pre-measurement value of that property. Rather, there may be no such meaningful thing as "the pre-measurement value of that property" at all! We have already suggested something along these lines in the case of the measurement of the energy of the PIB (whose wave function is initially a superposition of several energy eigenstates). A complete description of the pre-measurement state of the PIB consists, according to the pilot-wave theory, of the PIB wave function (a superposition of several energy eigenstates) and the position X of the associated particle. It is simply not clear how, from these ingredients, one would construct some specific energy value to attribute to the PIB as a "pre-measurement value".

But, you might object, the pilot-wave theory is deterministic! So surely the outcome of the energy measurement (i.e., the final position $Y(t)$ of the energy-measuring-apparatus pointer) is determined by, i.e., is some complicated function of, the initial states of the PIB and the pointer and the details (captured by the interaction Hamiltonian H_{int}) of their interaction. That is true, but does not affect the overall point. The heart of the matter is that the measurement outcome depends not only on the initial state of the measured system (and the initial state of the measuring apparatus) but also on details pertaining to the specific *way* the measurement is carried out.

Concretely, in our example of the measurement of the energy of the PIB, the outcome of the measurement will depend not only on the initial conditions – $\Psi(x, y, 0)$, $X(0)$, and $Y(0)$ – but also on the value of λ, which controls the "strength" of the PIB-Pointer interaction and so determines, for example, the amount of time it takes for the configuration space "islands" described around Fig. 7.5 to separate. From a purely dynamical point of view, it is hardly surprising that the configuration point $\{X(t), Y(t)\}$, which after all moves in some complicated and erratic way while the "islands" are still in the process of separating, can end up in a different "island" depending on the amount of time it takes for them to separate. But this means that different – and perfectly, equally legitimate – methods of "measuring the energy of the PIB" will yield different measurement outcomes, even if they are implemented on perfectly identical systems. Surely this demonstrates the complete pointlessness of trying to imagine that there is, according to the pilot-wave theory, some pre-existing

energy value which is revealed by (or even somehow affected by and then revealed by) the measurement procedure.

Bell has pointed out that any residual feeling of queasiness – about the fact that "measurements" do not, in general, merely reveal some pre-existing value for the quantity being measured – is almost certainly just a result of the connotations of the word "measurement". If, to you, the word "measurement" *means* "simply finding out something that was already definite" then it turns out that, according to the pilot-wave theory, (the procedures that are conventionally called) "position measurements" are indeed genuine measurements, whereas (the procedures that are conventionally called) "energy measurements" (and "momentum measurements" and "spin measurements"...) are not actually measurements at all. Perhaps, as Bell suggested, using a different word (like "experiment" instead of "measurement") would help us avoid inappropriate expectations. But there is nothing here that is actually problematic:

> the word ['measurement'] comes loaded with meaning from everyday life, meaning which is entirely inappropriate in the quantum context. When it is said that something is 'measured' it is difficult not to think of the result as referring to some pre-existing property of the object in question. [But t]his is to disregard Bohr's insistence that in quantum phenomena the apparatus as well as the system is essentially involved. If it were not so, how could we understand, for example, that measurement of a component of 'angular momentum' – in an arbitrarily chosen direction – yields one of a discrete set of values? When one forgets the role of the apparatus, as the word measurement makes all too likely, one despairs of ordinary logic – hence 'quantum logic'. When one remembers the role of the apparatus, ordinary logic is just fine [6].

For our purposes, all of this is important because it allows us to understand how, exactly, it is possible for the pilot-wave theory to exist, and work, in the face of the "no hidden variables" theorems that were mentioned back in Chap. 3. It seems that the (largely unacknowledged) linguistic connotations of the word "measurement" have contributed significantly to generations of physicists abandoning the hidden variables program and succumbing to the Copenhagen philosophy (or entertaining even more radical proposals such as the abandoning of the laws of logic). In particular, the conventional use of the word "measurement" (to describe experiments which output a value for the position, momentum, energy, spin, etc., of a particle) has led people to believe that it is reasonable to insist that any "hidden variable" account of such processes would have to attribute definite pre-measurement values that are simply revealed by the measurement, for all such quantities. But the "no hidden variables" theorems prove that "hidden variable" theories *of that sort* are impossible.

The "no hidden variables" theorems, that is, invariably apply only to "non-contextual" hidden variables theories. That is why those theorems do not in an any sense rule out the pilot-wave theory. But more importantly, the pilot-wave theory shows rather clearly that "contextuality" is in no way contrived or problematic, but is instead a completely straightforward consequence of the theory's very natural dynamical postulates. One just needs to take seriously the idea (which is inherent in understanding the measurement problem as a problem) that what a theory says about "measurements" should be extracted from (rather than awkwardly appended to) the theory's fundamental dynamical postulates.

7.6 The Many-Particle Theory and Nonlocality

In the opening of a paper he wrote in 1982, Bell describes his own first-person perspective on the "no hidden variables" theorem of von Neumann and its relation to the pilot-wave theory:

> When I was a student I had much difficulty with quantum mechanics. It was comforting to find that even Einstein had such difficulties for a long time. Indeed they had led him to the heretical conclusion that something was missing in the theory: 'I am, in fact, firmly convinced that the essentially statistical character of contemporary quantum theory is solely to be ascribed to the fact that this (theory) operates with an incomplete description of physical systems.'
>
> More explicitly, in 'a complete physical description, the statistical quantum theory would ... take an approximately analogous position to the statistical mechanics within the framework of classical mechanics...'.
>
> Einstein did not seem to know that this possibility, of peaceful coexistence between quantum statistical predictions and a more complete theoretical description, had been disposed of with great rigour by J. von Neumann. I myself did not know von Neumann's demonstration at first hand, for at that time it was available only in German, which I could not read. However I knew of it from the beautiful book by Born, *Natural Philosophy of Cause and Chance*, which was in fact one of the highlights of my physics education. Discussing how physics might develop Born wrote: 'I expect ... that we shall have to sacrifice some current ideas and to use still more abstract methods. However these are only opinions. A more concrete contribution to this question has been made by J.v. Neumann in his brilliant book, *Mathematische Grundlagen der Quantenmechanik*. He puts the theory on an axiomatic basis by deriving it from a few postulates of a very plausible and general character, about the properties of 'expectation values' (averages) and their representation by mathematical symbols. The result is that the formalism of quantum mechanics is uniquely determined by these axioms; in particular, no concealed parameters can be introduced with the help of which the indeterministic description could be transformed into a deterministic one. Hence if a future theory should be deterministic, it cannot be a modification of the present one but must be essentially different. How this could be possible without sacrificing a whole treasure of well established results I leave to the determinists to worry about.'
>
> Having read this, I relegated the question to the back of my mind and got on with more practical things.
>
> But in 1952 I saw the impossible done. It was in papers by David Bohm. Bohm showed explicitly how parameters could indeed be introduced, into nonrelativistic wave mechanics, with the help of which the indeterministic description could be transformed into a deterministic one. More importantly, in my opinion, the subjectivity of the orthodox version, the necessary reference to the 'observer', could be eliminated.
>
> Moreover, the essential idea was one that had been advanced already by de Broglie in 1927, in his 'pilot wave' picture.
>
> But why then had Born not told me of this 'pilot wave'? If only to point out what was wrong with it? Why did von Neumann not consider it? More extraordinarily, why did people go on producing 'impossibility' proofs, after 1952, and as recently as 1978? When even Pauli, Rosenfeld, and Heisenberg, could produce no more devastating criticism of Bohm's version than to brand it as 'metaphysical' and 'ideological'? Why is the pilot wave picture ignored in text books? Should it not be taught, not as the only way, but as an antidote to the prevailing complacency? To show that vagueness, subjectivity, and indeterminism, are not forced on us by experimental facts, but by deliberate theoretical choice? [7]

These are all very interesting and good questions that deserve answers. But we will not try to answer them here. I just wanted to give you a reference point for what you should probably be thinking at this point in the chapter: if this pilot-wave theory is as wonderful as it seems, why haven't I heard of it before? The theory seems to completely eliminate the "measurement problem" and, although the (puzzling) wave function of the universe still plays a role in the theory, there is no serious "ontology problem" since the everyday world of material objects is not supposed to be made of the wave function but is instead composed of the particles – whose existence in three-dimensional physical space is obviously unproblematic.

Overall, the reaction to the theory – by people like Pauli, Rosenfeld, and Heisenberg – is indeed puzzling. There is some kind of deep philosophical bias against what seems on the surface to be a far more scientific approach than the openly philosophical Copenhagen interpretation. But there is one feature of the pilot-wave theory, a technical physics feature, not at all "philosophical", which explains at least some of the physics community's near-unanimous dismissal of the theory: it not only fails to solve "the locality problem" that we discussed extensively in Chap. 4, but indeed makes the non-locality (which, at least according to Einstein, already afflicted ordinary QM) more blatant, explicit, and problematic.

The pilot-wave theory is manifestly non-local in the following sense: the velocity of each particle, at a given instant, depends on the instantaneous positions of all other particles (at least when there is entanglement). For example, consider a two-particle system with wave function $\Psi(x_1, x_2, t)$. The velocity of particle 1 at time t is given by

$$v_1(t) = \frac{dX_1(t)}{dt} = \frac{\hbar}{m_1}\,\mathrm{Im}\left[\frac{\left(\frac{\partial\Psi(x_1, X_2(t), t)}{\partial x_1}\right)}{\Psi(x_1, X_2(t), t)}\right]\Bigg|_{x_1 = X_1(t)}. \tag{7.54}$$

The point is that the right hand side depends on $X_2(t)$, the position of the other particle – even though this could be a million miles away. How particle 1 moves will depend, according to the theory, on what's happening with particle 2, and the dependence is immediate (with nothing like a speed-of-light time delay) and independent of the distance between the particles.

To make the non-locality even more explicit and dramatic, let's consider a situation in which "what's happening with particle 2" can be influenced in some way, say by some human agent who decides whether to make a certain kind of measurement on particle 2. As we have seen, such an intervention will influence the evolution of the wave function and hence affect the subsequent trajectories of the particles. The dramatic and shocking thing is that the subsequent trajectory of particle 1 can be affected by an experimental intervention that is localized in the vicinity of particle 2.

So, consider in particular two "particle-in-a-box" sub-systems that are well-separated in space (so the origins of the x_1 and x_2 coordinate systems are far apart from each other) in the entangled state:

$$\Psi(x_1, x_2, t) = \frac{1}{\sqrt{2}}\left[\psi_1(x_1)\psi_2(x_2) + i\psi_2(x_1)\psi_1(x_2)\right] e^{-i(E_1+E_2)t/\hbar} \tag{7.55}$$

where, as usual,

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right). \tag{7.56}$$

Remember, though, that $x_1 = 0$ is something like a million miles to the left and $x_2 = 0$ is a million miles to the right.

The wave function Ψ in Eq. (7.55) is an energy eigenstate of the two-PIB system, but turns out to have nonzero probability currents j_{x_1} and j_{x_2} (because of the relative phase between the two terms). That is, it turns out that the two particles, according to the pilot-wave theory, move – and move in tandem – as long as their joint wave function remains in this state. A density plot of $|\Psi|^2$ is shown in Fig. 7.6, and some plots of the associated motion of the two particles are shown in Fig. 7.7.

As long as the two-PIB system remains, undisturbed, in the quantum state Eq. (7.55), the two particles each just continue oscillating back and forth in their respective boxes. But now suppose that somebody decides to measure (say) the energy of particle 2. In the pilot-wave theory, we can analyze this measurement in the same schematic way we've done before: consider an energy measuring device with a moveable pointer whose final position will indicate the outcome of the energy measurement. We can suppose that the wave function of the pointer begins in a "ready" state $\phi(y)$ that is a Gaussian packet centered on $y = 0$. (The actual pointer

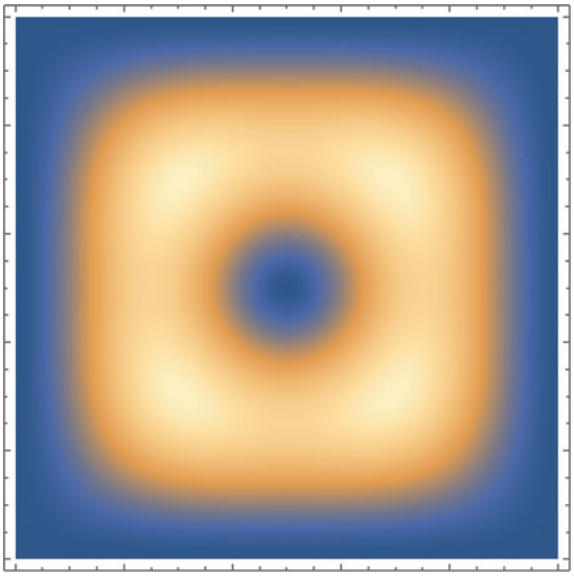


Fig. 7.6 Density plot of $|\Psi|^2$ in configuration space, with Ψ given by Eq. (7.55). The horizontal axis is x_1 and the vertical axis is x_2; there is a node in the *center* (where $\Psi = 0$) and then a "ring" where $|\Psi|^2$ is *large*. Note, though, that even though $|\Psi|^2$ is independent of time, the probability $|\Psi|^2$ is not stationary, but is instead flowing, clockwise, around the ring, like in a whirlpool

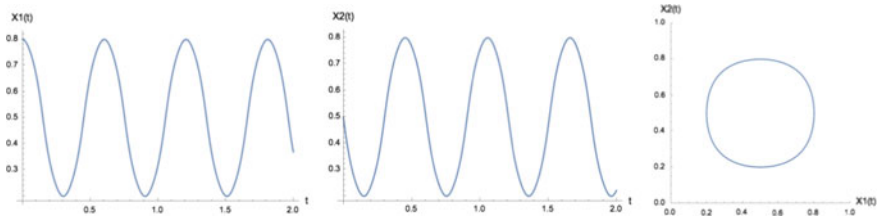


Fig. 7.7 The *left* and *center panels* show how the positions of the two particles ($X_1(t)$ and $X_2(t)$) vary with time: each particle essentially oscillates back and forth inside its box. The *right panel* shows the trajectory of the configuration point $\{X_1(t), X_2(t)\}$ through configuration space. (The trajectory is a closed clockwise loop.)

particle will have some random initial position $Y(0)$ in the support of this packet, i.e., near $Y(0) = 0$.) The initial wave function of the (now three-particle!) system will thus be

$$\Psi(x_1, x_2, y, t) = \frac{1}{\sqrt{2}} [\psi_1(x_1)\psi_2(x_2) + i\psi_2(x_1)\psi_1(x_2)] \ \phi(y). \tag{7.57}$$

If and when the measurement is actually carried out, an interaction Hamiltonian such as

$$\hat{H}_{int} = \lambda \hat{H}_{x_2} \hat{p}_y \tag{7.58}$$

will then couple the position of the pointer to the energy of particle 2 and hence the three-particle wave function will evolve into:

$$\Psi(x_1, x_2, t) = \frac{1}{\sqrt{2}} [\psi_1(x_1)\psi_2(x_2)\phi(y - \lambda E_2 t/\hbar) + i\psi_2(x_1)\psi_1(x_2)\phi(y - \lambda E_1 t/\hbar)] . \tag{7.59}$$

Let's try to visualize this in the (now three-dimensional!) configuration space. Prior to the measurement interaction, the initial wave function – Eq. (7.57) – has support only around $y = 0$ and then looks, in the x_1-x_2-plane, like what we talked about before: it is a superposition of the "particle 1 in the ground state and particle 2 in the first excited state" and "particle 1 in the first excited state and particle 2 in the ground state" states which has a node ($\Psi = 0$) at $(x_1, x_2) = (L/2, L/2)$ and a "ring" of large $|\Psi|^2$ around the node. As the measurement interaction proceeds, though, the two superposed PIB states get "lifted" – by different amounts! – in the y-direction in configuration space and so cease to overlap. This is pictured in Fig. 7.8.

The actual configuration point $\{X_1, X_2, Y\}$ of course ends up randomly (depending on the detailed initial particle positions) in one or the other of the two now-separated "islands" of wave function support in configuration space. In particular, as we discussed before, the pointer ends up with some definite position, either $Y \approx \lambda E_1 t$ (indicating "particle 2 has energy E_1") or $Y \approx \lambda E_2 t$ (indicating "particle 2 has energy E_2"). But then consider the implications of this for the subsequent motion of particle 1.

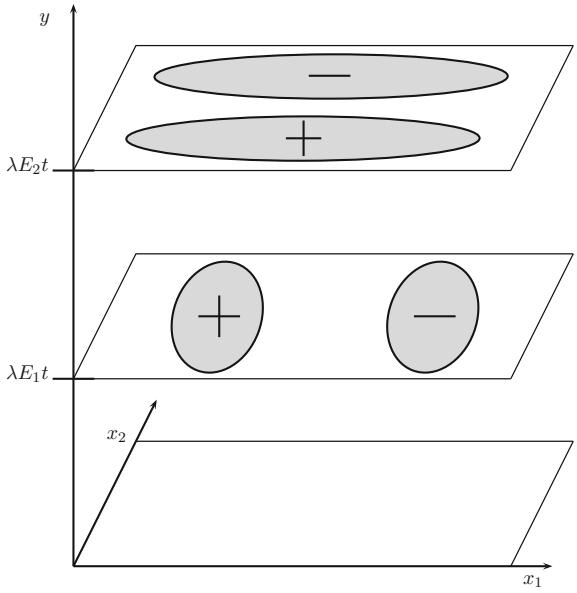


Fig. 7.8 The wave function $\Psi(x_1, x_2, y)$ for the example discussed in the text, in which the energy of one of two entangled particles-in-boxes is measured and the outcome displayed in a pointer with coordinate y. Prior to the measurement, the two superposed terms overlap, in the $y = 0$ plane, giving rise to the wave function with the structure depicted earlier, in Fig. 7.6. But the coupling to the measuring device's pointer causes the two terms to separate as shown here: the $\psi_1(x_1)\psi_2(x_2)$ term ("particle 1 is in the ground state and particle 2 is in the first excited state") is displaced in the y-direction by a distance $\lambda E_2 t$, while the $\psi_2(x_1)\psi_1(x_2)$ term ("particle 1 is in the first excited state and particle 2 is in the ground state") is displaced in the y-direction by a smaller distance, $\lambda E_1 t$. In the pilot-wave theory, the actual configuration point $\{X_1, X_2, Y\}$ will end up in one or the other of these "islands" in configuration space, depending on the initial positions of the three particles. But then since the particle velocities depend only on the structure of the wave function near the actual configuration point, the subsequent evolution of all three particles will be dictated exclusively by just one of the two terms in the wave function. And this implies (among other things) that after the measurement on particle 2, particle 1 will stop moving

There are two equivalent ways to put the point.

First, since the velocity of particle 1 depends only on the structure of Ψ around the actual configuration point, the relevant part of Ψ is now *either* $\psi_1(x_1)\psi_2(x_2)$ (if $Y \approx \lambda E_2 t$) *or* $\psi_2(x_1)\psi_1(x_2)$ (if $Y \approx \lambda E_1 t$). It is one term or the other, rather than their superposition, which will now determine the subsequent motion of particle 1. But for both possibilities it turns out that the velocity of particle 1 will be zero!

The second way to express the same point is to say that while, prior to the measurement of particle 2's energy, the "conditional wave function" (CWF) of particle 1 is a superposition of the ground state and first excited state wave functions (which gives rise to the oscillatory motion we saw at the beginning of this chapter), after the measurement of particle 2's energy, the CWF of particle 1 "collapses" to *either* the

ground state *or* the first excited state. And, as we talked about before, both of these possibilities imply that particle 1 will be at rest.

However you think about it, though, the crucial point is that, by measuring the energy of (i.e., by intervening in the affairs of) particle 2 (which remember is, say, a million miles to the right!), we have caused a sudden and dramatic change in the behavior of particle 1 (which, remember, is a million miles to the left): it went from oscillating back and forth, to just sitting there at rest. We made it stop moving, instantaneously, from two million miles away! It is a truly blatant case of "spooky action at a distance" which seems impossible to reconcile with the relativistic concept of locality according to which all causal influences propagate at or below the speed of light.

It is, however, worth noting that, although there is a blatant violation of relativistic locality here, the instantaneous action-at-a-distance cannot be used to transmit signals or information. This is implied by the fact that the pilot-wave theory makes the same statistical predictions as ordinary QM, but it is worth saying a little more here. One has to remember that, although, according to the pilot-wave theory, particle 1 is initially oscillating back and forth, one cannot *observe* this motion (or its subsequent cessation). Or rather, if you *did* try to observe it, this observation would require interacting with particle 1 with some physical observation equipment, which would disrupt its subsequent evolution and "break" the entanglement with particle 2 (just like the energy measurement in the above example does), and hence imply that a later measurement on particle 2 has no effect whatsoever on particle 1!

Or one can think of it this way: although particle 1 has, at every moment, a definite position $X_1(t)$, that position is not known to anybody; all that is known is a probability distribution for where particle 1 might be found if looked for. But this probability distribution, as it turns out, is independent of time and in particular doesn't change even when the distant intervention (which causes particle 1 to stop moving) occurs. You could imagine, for example, setting up thousands of copies of this system (with identical wave functions but of course random and different particle positions), with Alice stationed a million miles to the left with all the particle 1s, and Bob stationed a million miles to the right with all the particle 2 s and an arsenal of energy-measuring devices. Suppose, by prior arrangement, Bob will pick some random time within a few seconds of $t = 0$ to measure the energies of all of his particle 2s. What could Alice do to monitor her particles and try to observe the effect of Bob's intervention on them?

She could, for example, pick a few hundred of the particles and measure their positions at $t = 0$, then do the same thing at $t = 1$ s for a different set of a few hundred particles, then do the same thing again at $t = 2$ s, and so on. And the point is, she would never be able to tell in this way when Bob had performed the intervention which, in fact, according to the theory, influences the motion of her collection of particles. She would just see the same exact random distribution in the sets of particles inspected before Bob's intervention, as she sees in the sets inspected after Bob's intervention.

Similarly, if Alice instead measures the energies of a few hundred of her particles, she will get the same statistical distribution of outcomes (namely: about half E_1 and about half E_2) whether Bob has made his measurements yet, or not. Each individual

one of Bob's measurements affects the motion of one of Alice's particles, according to the pilot-wave theory, and indeed the outcome of Bob's measurement allows him to know, in advance and with absolute certainty, how a subsequent measurement, by Alice, of the energy of the entangled partner, will come out. But because Bob cannot control the outcome of his own experiment, the non-local causal influence which produces the perfect correlation, is useless for purposes of communication.

So although, according to the theory, there are these blatant violations of relativistic causality, they are in some sense "behind the scenes" – hidden away where we can't in practice see them or use them to (say) send signals faster than light. This, in some sense, saves the theory from the worst kind of conflict with relativity: it would be bad, for example, if the theory implied that you could relay a message into your own past and arrange, say, to have your parents killed before you were born. (That is a classic example of the kind of paradoxical situation that could arise if faster-than-light signalling – which remember implies signalling into the past in some reference frames if relativity is true – were possible.) But this is not much comfort for anybody who takes relativity theory seriously, as telling us something about the fundamental structure of space and time, rather than just prohibiting certain types of communication among humans. If relativity theory is taken to prohibit faster-than-light causal influences, then the pilot-wave theory is just inconsistent with relativity, end of discussion.

7.7 Reactions

As mentioned before, the basic idea of the pilot-wave theory (that is, the idea of resolving the wave-particle duality dilemma by having *both* waves and particles) was proposed already in the 1920 s by de Broglie. When he presented his ideas, they were roundly rejected by nearly everyone and de Broglie himself abandoned the idea shortly thereafter. It was basically only Einstein who had a favorable reaction to de Broglie's ideas: recall his (Einstein's) comment from 1927 that

> one can remove [the "boxes" type objection, against nonlocality] only in the following way, that one does not describe the process solely by the Schrödinger wave, but that at the same time one localises the particle during the propagation. I think Mr de Broglie is right to search in this direction [8].

Twenty five years later – during which time de Broglie had completely abandoned and forgotten the pilot-wave idea, and Einstein had gone off on his own to try to develop his "unified field theory" program – David Bohm independently rediscovered and developed and published the pilot-wave idea. Prior to this publication, Bohm wrote: "I can't believe that I should have been the one to see this" and expressed an optimistic expectation "that the physics community would react with enthusiasm [9]." But instead the community reacted very negatively. Oppenheimer dismissed Bohm's ideas as "juvenile deviationism" and said that "if we cannot disprove Bohm, then we must agree to ignore him." Rosenfeld called the theory "very ingenious, but

basically wrong". Wolfgang Pauli called it "foolish simplicity" which "is of course beyond all help [9, 10]".

None of this is particularly surprising, in the sense that all of these people were (by then) proponents of the (by then) orthodox Copenhagen interpretation which (as we have seen) is quite antagonistic toward the very idea of trying to give a precise and realistic description of microscopic processes. It is somewhat more puzzling, then, that even Einstein – the greatest critic of the Copenhagen interpretation and the greatest champion of the pilot-wave idea back in the 1920s – did not seem to think highly of the theory, even though it seems to be exactly the kind of thing that Einstein had sought and even though Einstein had directly and personally influenced and encouraged Bohm into the line of thinking that led to his (re-) discovery of the pilot-wave ideas [10]. Einstein wrote, in a letter to Max Born:

> Have you noticed that Bohm believes (as de Broglie did, 25 years ago) that he is able to interpret the quantum theory in deterministic terms? That way seems too cheap to me [11].

It is not clear exactly what Einstein meant by "too cheap", but it seems likely that the theory did not strike him as a step in the right direction since it failed to eliminate (but instead in some ways exacerbated) the one feature that Einstein found most unacceptable in orthodox quantum theory: non-locality.

Heisenberg (less surprisingly) also didn't much like the pilot-wave theory, and gave his reasons in some detail in an essay called "Criticisms and Counterproposals to the Copenhagen Interpretation of Quantum Theory":

> When one analyzes the papers of the first group [of criticisms/counterproposals, namely, those who do not "want to change the Copenhagen interpretation so far as predictions of experimental results are concerned", but try "to change the language of this interpretation in order to get a closer resemblance to classical physics"] it is important to realize from the beginning that their interpretations cannot be refuted by experiment, since they only repeat the Copenhagen interpretation in a different language. From a strictly positivistic standpoint one may even say that we are here concerned not with counterproposals to the Copenhagen interpretation but with its exact repetition in a different language. Therefore, one can only dispute the suitability of this language. One group of counterproposals works with the idea of 'hidden parameters'. Since the quantum-theoretical laws determine in general the results of an experiment only statistically, one would from the classical standpoint be inclined to think that there exist some hidden parameters which escape observation in any ordinary experiment but which determine the outcome of the experiment in the normal causal way. Therefore, some papers try to construct such parameters within the framework of quantum mechanics.
>
> Along this line, for instance, Bohm has made a counter-proposal to the Copenhagen interpretation, which has recently been taken up to some extent also by de Broglie. Bohm's interpretation has been worked out in detail. It may therefore serve here as a basis for the discussions. Bohm considers the particles as 'objectively real' structures, like the point masses in Newtonian mechanics. The waves in configuration space are in his interpretation 'objectively real' too, like electric fields. Configuration space is a space of many dimensions referring to the different co-ordinates of all the particles belonging to the system. Here we meet a first difficulty: what does it mean to call waves in configuration space 'real'? This space is a very abstract space.... but things are in the ordinary three-dimensional space, not in an abstract configuration space. One may call the waves in configuration space 'objective' when one wants to say that these waves do not depend on any observer; but one can scarcely call them 'real' unless one is willing to change the meaning of the word.

> ....One consequence of this interpretation is, as Pauli has emphasized, that the electrons in the ground states of many atoms should be at rest, not performing any orbital motion around the atomic nucleus..... [Bohm responds that] when the quantum theory for the measuring equipment is taken into account – especially some strange quantum potentials introduced ad hoc by Bohm – then the statement is admissible that the electrons 'really' always are at rest. In measurements of the position of the particle, Bohm takes the ordinary interpretation of the experiments as correct; in measurements of the velocity he rejects it. At this price Bohm considers himself able to assert: 'We do not need to abandon the precise, rational and objective description of individual systems in the realm of quantum theory.' This objective description, however, reveals itself as a kind of 'ideological superstructure', which has little to do with immediate physical reality....
>
> ....Bohm's language, as we have already pointed out, says nothing about physics that is different from what the Copenhagen interpretation says [12].

Readers of Chap. 5 will sympathize with Heisenberg's point that the wave function on configuration space is hard to take seriously as a physically real field (even if the pilot-wave theory arguably suffers less from an "ontology problem" than theories for which the wave function is supposed to be the *only* physical reality). And those who followed the discussion in Sect. 7.5 will recognize Heisenberg's complaint about electrons in atoms being (he thinks, implausibly) at rest as well as his complaint (albeit using different terminology) that position, for Bohm, is non-contextual while velocity is contextual. Beneath all of this, though, is the more controversial philosophical question about what it means for a theory to "say something about physics".

In general, then, it is fair to say that the pilot-wave theory never got a particularly favorable reception, either when it was proposed initially in 1927 by de Broglie or when it was proposed later in 1952 by Bohm. And it continues to be regarded, by most physicists, as (at best) not worthy of serious attention. But – as we've seen – it *was* viewed very positively by Bell, who recognized it immediately as an explicit counter-example to the supposed proofs (by von Neumann and others) that no "hidden variable" completion of quantum mechanics was mathematically possible. This realization led Bell to wonder how the "impossibility proofs" had gone wrong, exactly – how they had wrongly convinced so many people that something which clearly *is* possible, isn't possible. Bell answered this question (in effect, by explicitly identifying the tacit assumption of "non-contextuality" in the proofs) in an important 1964 paper which (due to an unfortunate editorial accident) remained unpublished until 1966 [12].

At the end of that paper, he explains that, while Bohm's pilot-wave theory is a clear-cut counter-example to any assertion that deterministic hidden variable theories are impossible, the theory does have the unappealing feature we noted in the last section – what Bell refers to as a "grossly nonlocal character":

> in this theory an explicit causal mechanism exists whereby the disposition of one piece of apparatus affects the results obtained with a distant piece. In fact the Einstein–Podolsky–Rosen paradox is resolved in the way which Einstein would have liked least [13].

Bell then ends the paper with the following paragraph:

> Bohm of course was well aware of these features of his scheme, and has given them much attention. However, it must be stressed that, to the present writer's knowledge, there is no

> *proof* that *any* hidden variable account of quantum mechanics *must* have this extraordinary character.[*] It would therefore be interesting, perhaps, to pursue some further 'impossibility proofs', replacing the arbitrary axioms objected to above [namely, "non-contextuality"] by some condition of locality, or of separability of distant systems [13].

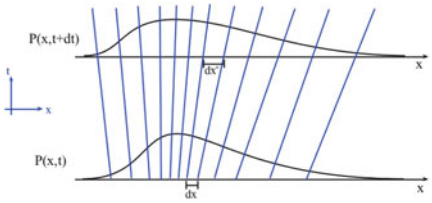
The "[∗]" points to a footnote which was added before the delayed publication of the paper: "Since the completion of this paper such a proof has been found: J.S. Bell, Physics 1, 195, [1964]". That is, between the completion of this first paper in 1964, and its publication in 1966, Bell had already discovered and published the answer to his own question: would it be possible to construct a hidden variable completion of QM, with all of the virtues of the pilot-wave theory, but without the troubling non-local character?

His answer is the subject of Chap. 8.

Projects:

7.1 Show that Eq. (7.6) really is equivalent to Eq. (7.5).

7.2 Show that the probability distribution $P(x,t)$ for an ensemble of particles moving with velocities $v(x,t)$ should satisfy Eq. (7.20). Hint: argue, based on this picture



that all the trajectories (shown in the figure as blue lines on a space-time diagram) in dx at time t will be in dx' at time $t+dt$, i.e., $P(x+v(x,t)dt, t+dt)dx' = P(x,t)dx$. This can (with some additional work) then be shown to be equivalent to

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[P(x,t)v(x,t)\right]. \tag{7.60}$$

7.3 Work through the derivation of Eq. (7.12) – the quantum continuity equation for a particle in three dimensions – from the time-dependent Schrödinger equation, and thereby confirm the expression in Eq. (7.14) for the quantum probability current.

7.4 Show that, indeed, Eq. (7.17) is equivalent to the earlier expressions for the particle velocity in the pilot-wave theory.

7.5 Confirm that Eq. (7.30) really solves Eq. (7.29).

7.6 Massage Eq. (7.32) into polar form. Let Mathematica numerically solve the differential equation $\frac{dX}{dt} = \frac{\hbar}{m}\frac{\partial S}{\partial x}$ to recreate trajectories like the ones shown in Fig. 7.4.

7.7 For the toy model of a measurement discussed in Sect. 7.4, Schrödinger's equation reads

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}\Psi \tag{7.61}$$

with $\hat{H} = \lambda\hat{H}_x\hat{p}_y$, where, in turn, $\hat{H}_x = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)$ and $\hat{p}_y = -i\hbar\frac{\partial}{\partial y}$. Show that the x- and y-components of the quantum probability current can be written

$$j_x = -\frac{\lambda\hbar^2}{m}\mathrm{Re}\left[\Psi^*\frac{\partial}{\partial x}\frac{\partial}{\partial y}\Psi\right] \tag{7.62}$$

and

$$j_y = \frac{\lambda\hbar^2}{m}\frac{\partial\Psi^*}{\partial x}\frac{\partial\Psi}{\partial x} \tag{7.63}$$

respectively. That is, show that the Schrödinger equation implies the continuity equation

$$\frac{\partial|\Psi|^2}{\partial t} = -\frac{\partial j_x}{\partial x} - \frac{\partial j_y}{\partial y} \tag{7.64}$$

with j_x and j_y as given above.

7.8 Use the results of the previous Project to argue that, in the pilot-wave theory, the velocities of the two particles involved in the toy model of measurement are given by

$$\frac{dX}{dt} = -\frac{\lambda\hbar^2}{m}\mathrm{Re}\left[\frac{\frac{\partial}{\partial x}\frac{\partial}{\partial y}\Psi}{\Psi}\right] \tag{7.65}$$

and

$$\frac{dY}{dt} = \frac{\lambda\hbar^2}{m}\left|\frac{\partial\Psi/\partial x}{\Psi}\right|^2. \tag{7.66}$$

Let Mathematica numerically solve these differential equations to find some example trajectories $X(t)$, $Y(t)$. Use the known solution of Schrödinger's equation for this problem:

$$\Psi(x, y, t) = \sum_i c_i\psi_i(x)\phi(y - \lambda E_i t). \tag{7.67}$$

7.9 Use your Mathematica program from the previous Project to demonstrate the "contextuality" of energy measurements in the pilot-wave theory. In particular, find specific initial conditions that lead to *different* outcomes for the energy measurement for different values of λ. (You can do this by trial and error: just pick some random values for $X(0)$ and $Y(0)$ and then fiddle with the value of λ. You probably won't have to try too many different values of λ before you find a couple of values that produce distinct values of $E_n \approx \frac{Y(t)}{\lambda t}$.)

7.10 Show that Eq. (7.48) really solves the (time-dependent) Schrödinger equation, with $V(x)$ given by Eq. (7.45), as long as the packet width σ and the frequency ω are related as in Eq. (7.47).

7.11 Calculate j_{x_1} and j_{x_2} for the wave function in Eq. (7.55).

7.12 In the passage quoted in Sect. 7.7, Heisenberg refers to "some strange quantum potentials introduced ad hoc by Bohm". This is a reference to a slightly different formulation of the pilot-wave theory, in terms of which Bohm presents the theory in his 1952 papers. To see how this works, take a time derivative of Eq. (7.6) to derive an expression for the *acceleration* of the particle. It is important here that the right hand side of Eq. (7.6) depends on time in two different ways, so one must use the "convective derivative" $\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{dx}{dt}\frac{\partial}{\partial x}$. If all goes well you should be able to write the equation describing the motion of the particle in the somewhat more Newtonian-mechanical-like form,

$$ma = -\frac{\partial}{\partial x}(V + Q) \tag{7.68}$$

where V is the regular ("classical") potential energy function (which appears in Schrödinger's equation) and then Q is a new, so-called "quantum potential" which depends on the structure of the wave function ψ. Find the expression for Q. (It can be expressed most simply in terms of the R in $\psi = Re^{iS}$.) Think about how to understand, from this more classical perspective on the motion of the particle, how (for example) the electron particle in a ground-state Hydrogen atom remains at rest.

7.13 The previous Project might suggest that, in addition to the kinetic energy $K = \frac{1}{2}mv^2$ and classical potential energy $V(x)$, a particle in the pilot-wave theory also possesses some "quantum potential energy", Q. Would the inclusion of this "quantum potential energy" make it possible to regard the measurement of the energy of a particle as revealing a pre-existing energy value? In other words, does the possibility of re-formulating the theory in this more Newtonian-mechanical-like way undermine our conclusion that energy is, in the pilot-wave theory, contextual?

7.14 Suppose we define the total energy of a particle in the pilot-wave theory as $E = K + V + Q$ as suggested in the previous Project. Is the total energy E of a particle conserved according to the theory?

7.15 One of Heisenberg's criticisms of Bohm's theory is that "[i]n measurements of the position of the particle, Bohm takes the ordinary interpretation of the experiments as correct; in measurements of the velocity he rejects it." Heisenberg here means that, in the pilot-wave picture, position is non-contextual whereas velocity is contextual. Heisenberg seems to think that this is the result of a choice and is therefore arbitrary and unbelievable. But we have shown in the Chapter that the contextuality of (for example) momentum and energy is not a choice at all, but simply a consequence of the basic dynamical postulates of the theory. Complete the rebuttal of Heisenberg's criticism by showing that position measurements just do, according again to the dynamical postulates,

reveal pre-existing position values. (Hint: consider a particle with wave function $\psi_0(x)$ whose position is to be measured using an apparatus whose pointer has an initial wave function $\phi(y)$ and interacts with the particle according to $\hat{H}_{int} = \lambda \hat{x}\,\hat{p}_y$. Show that the x- and y-components of the quantum probability current can be taken to be $j_x = 0$ and $j_y = \lambda x \Psi^* \Psi$ so that $dX/dt = 0$ and $dY/dt = \lambda X$. This implies that the final displacement of the pointer is proportional to X, the actual position of the particle whose position is being measured.)

7.16 Read through Bohm's 1952 papers [3] and report on anything you find interesting or surprising.

7.17 Read through Ref. [14], "The pilot-wave perspective on quantum scattering and tunneling," and summarize its main points.

7.18 Read through Ref. [15], "The pilot-wave perspective on spin," and summarize its main points. In particular, explain in detail how the pilot-wave theory accounts for the EPR-Bohm correlations. Bell says that the theory resolves the EPR paradox "in the way which Einstein would have liked least". What exactly does he mean?

7.19 Can the pilot-wave theory be diagnosed as "nonlocal" using Bell's formulation of locality (or the slightly modified formulation) from Chap. 1? How about using the related necessary condition for locality that we developed in Chap. 5?

7.20 In the text, the non-locality of the pilot-wave theory is explained in terms of an entangled two-particle state (with a measurement of one of the particles non-locally affecting the motion of the other, distant particle). But we saw in Chap. 4 that textbook quantum theory is already apparently non-local in the simpler, single-particle "Einstein's boxes" scenario. Is any non-locality involved in the pilot-wave theory's account of "Einstein's boxes"? Explain.

7.21 Recall the passage quoted in Sect. 6.6, in which (the textbook author) David Griffiths explains three frequently-encountered attitudes toward quantum mechanics: the "realist" position, the "orthodox" position, and the "agnostic" position. Would Griffiths classify the pilot-wave theory as "realist"? Note that your answer will depend on whether or not you think he intends what he says about a position measurement, by way of defining what he means by "realist", to apply *just* to position measurements, or instead to apply more generally to measurements of any property. Why do you think Griffiths isn't more explicit about this issue, namely, whether, to count as "realist", a theory should merely say that position measurements reveal pre-existing position values, or instead must say that a measurement of *any* quantity must reveal its pre-existing value?

7.22 In the passage quoted in Sect. 6.6, Griffiths seems to define "realism" as meaning that a theory posits specifically non-contextual hidden variables (at least for position). The pilot-wave theory would count as "realist" in this sense (if we interpret this notion of "realism" as requiring non-contextual hidden variables *only* for positions... obviously the pilot-wave theory would *not* be "realist" if that is taken to require non-contextual hidden variables for not only position, but also momentum, energy, etc.). But there would seem to be a more basic

notion of "realism" that the pilot-wave theory embodies – a notion that doesn't pertain to anything so obscure and specific as non-contextual hidden variables, but which instead means something along the lines of "there's a real world out there, independent of us, and it's the job of physics theories to describe it." Try your hand at describing/formulating this more basic notion of "realism" (more carefully than I did in the previous sentence). Does the Copenhagen interpretation count as "realist" in this more basic sense? Would a hidden variable theory according to which *all* properties are contextual, count as "realist" in this more basic sense?

7.23 We have argued that the pilot-wave theory completely resolves the measurement problem (of Chap. 3), but suffers acutely from the locality problem (of Chap. 4). What about the ontology problem (of Chap. 5), about which only a few brief remarks have been made in passing? Summarize your thinking about this subtle question.

7.24 Use the method introduced in Projects 2.9 and 2.10 to find an asymptotic ($t \to \infty$) approximation to $\psi(x, t)$ for a free particle whose $\psi(x, 0) = \sqrt{\frac{2}{L}} \sin(\pi x/L)$ for $0 < x < L$ (and zero otherwise). Use the result to determine the possible values that one might find for a "time-of-flight" type measurement of the momentum of a particle which is initially in the ground state of a PIB potential. Explain how these values relate to the values one would naively expect for a classical particle with the same energy.

References

1. A. Einstein, Reply to criticisms, in *Albert Einstein: Philosopher-Scientist*, vol. 2, ed. by P.A. Schilpp (1949)
2. G. Bacciogallupi, A. Valentini, Quantum theory at the crossroads, http://arxiv.org/pdf/quant-ph/0609184.pdf
3. D. Bohm, A suggested interpretation of the quantum theory in terms of 'Hidden' variables (parts I and II). Phys. Rev. **85**(2), 166–179 and 180–193 (1952)
4. J.S. Bell, Six possible worlds of quantum mechanics, *Speakable and Unspeakable in Quantum Mechanics*, 2nd edn. (Cambridge, 2004)
5. C. Philippidis, C. Dewdney, B.J. Hiley, Quantum interference and the quantum potential. Nuovo Cimento, **52 B**, 15–28 (1979)
6. J.S. Bell, Against 'Measurement', *Speakable and Unspeakable in Quantum Mechanics*, 2nd edn. (Cambridge, 2004)
7. J.S. Bell, On the impossible pilot wave, *Speakable and Unspeakable in Quantum Mechanics*, 2nd edn. (Cambridge, 2004)
8. Einstein's remarks from Solvay 1927, translated in Bacciogallupi and Valentini, Quantum theory at the crossroads, pp. 485–487, http://arxiv.org/pdf/quant-ph/0609184.pdf
9. S. Goldstein, A theorist ignored (review of F. David Peat's biography of David Bohm, Infinite Potential). Science **275**(28), 1893 (1997)
10. J. Bricmont, *Making Sense of Quantum Mechanics* (Springer, New York, 2016)
11. Einstein, letter of May 12, 1952, to Max Born, in Irene Born, trans., *The Born-Einstein Letters* (Walker and Company, New York, 1971), p. 192
12. W. Heisenberg, Criticism and counterproposals to the Copenhagen interpretation of quantum theory. *Physics and Philosophy* (Harper & Row, New York, 1958)

13. J.S. Bell, On the problem of hidden variables in quantum mechanics. Rev. Mod. Phys. **38**(3), 447–452 (1966). (Reprinted in *Speakable and Unspeakable in Quantum Mechanics*, 2nd edn. (Cambridge, 2004).)
14. T. Norsen, The pilot-wave perspective on quantum scattering and tunneling. Am. J. Phys. **81**(4), 258–266 (2013)
15. T. Norsen, The pilot-wave perspective on spin. Am. J. Phys. **82**(4), 337–348 (2014)

Chapter 8
Bell's Theorem

The result which has come to be known as "Bell's Theorem" – but which Bell himself instead referred to as the "locality inequality theorem" [1] – first appeared in Bell's 1964 paper, "On the Einstein–Podolsky–Rosen paradox" [2]. Following Bell's own presentation, we begin here by recalling (from Chap. 4) the EPR argument, in the updated form introduced by Bohm in 1951.

8.1 EPRB Revisited

In Bohm's re-formulation, we consider a pair of spatially separated spin 1/2 particles in the spin "singlet" state

$$\psi_s = \frac{1}{\sqrt{2}}\left[\psi^1_{+z}\psi^2_{-z} - \psi^1_{-z}\psi^2_{+z}\right]. \tag{8.1}$$

With the state written in this form, it is apparent that, from a measurement of the z-component of particle 1's spin, we can immediately infer the z-component of particle 2's spin. The two particles' z-spins are perfectly anti-correlated: a $+1$ outcome on one side implies a -1 outcome on the other side, and vice versa. But according to the locality assumption, measuring the z-spin of particle 1 (say, nearby) should not disturb the physical state of the (say, distant) particle 2. And so, according to the reasoning introduced by EPR in 1935, the distant particle must *already possess* a definite z-spin value (which is then simply revealed when its z-spin is measured). The only alternative is that the distant particle's z-spin somehow comes into existence (crystallizing out of a prior fog, so to speak) as a result of our measurement on the nearby particle; but that would constitute a violation of local causality. The EPR claim is that the only way to avoid non-locality is to attribute a pre-determined z-spin value to the distant particle.

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_8

The singlet state ψ_s can also be re-written in terms of the single particle eigenstates for the x-component of the spin:

$$\psi^1_{\pm x} = \frac{1}{\sqrt{2}}\left[\psi^1_{+z} \pm \psi^1_{-z}\right] \tag{8.2}$$

and identically for particle 2. Solving for $\psi_{\pm z}$ in terms of $\psi_{\pm x}$ and plugging in gives

$$\psi_s = \frac{1}{\sqrt{2}}\left[\psi^1_{+x}\psi^2_{-x} - \psi^1_{-x}\psi^2_{+x}\right]. \tag{8.3}$$

It is a special property of the singlet state that it takes exactly the same form, written in terms of the states of definite x-spin, as it does written in terms of the states of definite z-spin. This allows the same EPR reasoning to be run again, this time about x-spin values: the distant particle must *also* possess a definite pre-measurement value for its x-spin since we could determine this value, with certainty, and without disturbing the physical state of the distant particle, by measuring the x-spin of the nearby particle.

And note – contra Bohr's reply to EPR – that the fact that we could only measure *either* the nearby particle's x-spin *or* its z-spin, does not block the inference to the pre-determinateness of *both* properties of the distant particle. It is true that we could only *learn* about one or the other of the two properties on a given particle pair. But the assumption is that we have a genuinely free choice about which property (if either!) to measure on the nearby particle. *If* we choose to measure the z-spin of the nearby particle, we would learn the value of the z-spin of the distant particle (without disturbing its physical state in any way) and hence be in a position to infer its existence. Whether we in fact *do* so choose is immaterial: for the existence of the distant particle's z-spin to depend upon whether or not we choose to measure the nearby particle's z-spin, would be a violation of locality. So, again, the claim is that to avoid non-locality ("spooky action-at-a-distance") we must attribute definite, pre-measurement values to both the x- and z-spin of the distant particle.

And, finally, the argument is symmetric with respect to the two particles: if the *distant* particle must already possess definite values of both x-spin and z-spin (because we, here, could determine those values indirectly, without in any way influencing the physical state of the particle), so must the *nearby* particle (because someone over there could determine the values indirectly, without in any way influencing the physical state of the nearby particle).

So the upshot of the EPR-Bohm argument is that when a pair of particles is prepared in the spin state ψ_s, each particle in the pair must already possess a definite value for both x- and z-spin, and these values (for the two particles) must apparently be correlated to guarantee that any of the possible subsequent measurements agree with the quantum mechanical statistics. Locality thus requires a "hidden variable" theory of the sort summarized in the following table:

Pair type	Particle 1	Particle 2	Frequency
1	$\{+1, +1\}$	$\{-1, -1\}$	25%
2	$\{+1, -1\}$	$\{-1, +1\}$	25%
3	$\{-1, +1\}$	$\{+1, -1\}$	25%
4	$\{-1, -1\}$	$\{+1, +1\}$	25%

According to this model, the description of the particle pair in terms of the quantum state ψ_s is decidedly *incomplete*. There are facts about the x- and z-spins of both particles that are not contained in ψ_s! In particular, each pair of particles (prepared in the quantum state ψ_s) is in fact of one of the four sub-types described in the four rows of the table. Which of the four types a given pair of particles ends up is somehow just random, with equal 25% probabilities for each of the four types.

The exact nature of the four types is described in the "Particle 1" and "Particle 2" columns. For example, $\{+1, -1\}$ means that the particle in question is spin-up ("+1") along the x-direction and spin-down ("−1") along the z-direction. Notice that both the x-spin and z-spin values are perfectly anti-correlated within each pair type. For example, if Particle 1 is spin-up along z, then Particle 2 is spin-down along z. This ensures that, if the same property (x-spin or z-spin) is measured on both particles, the results will always be opposite (as predicted by QM). (This also explains why there are exactly four allowed "pair types". The other logical possibilities, for example $\{+1, +1\}$ for particle 1 and $\{+1, -1\}$ for particle 2, would violate the perfect correlation property for at least one possible set of measurements – here, if the x-spin is measured on both particles. Such pair types, if included, would need to be assigned frequencies of exactly zero in order for the model to reproduce the quantum predictions, so we might as well just exclude them entirely from the discussion.)

And notice also that the equal 25% frequencies for all four pair types are required to match the rest of the quantum predictions. For example, what happens if the x-spin of particle 1 is measured and the z-spin of particle 2 is measured? The quantum statistics can be read off from ψ_s re-written in this form:

$$\psi_s = \frac{1}{2}\left[\psi^1_{+x}\psi^2_{-z} + \psi^1_{-x}\psi^2_{-z} - \psi^1_{+x}\psi^2_{+z} + \psi^1_{-x}\psi^2_{+z}\right]. \tag{8.4}$$

The four possible joint outcomes (up-up, down-up, up-down, and down-down) thus have equal, 25%, probabilities. We can thus reproduce the complete slate of quantum mechanical statistical predictions (for any set of measurements along the x- and z-axes) by letting each of the four pair types in our hidden variable model occur with 25% frequency.

It is clear that this sort of "hidden variable" model, in which particles carry predetermined values for possible spin measurements along the x- and z-directions, can reproduce the quantum predictions but without the non-locality associated with ordinary quantum theory's collapse postulate (combined with the claim that the wave function provides a complete state description). The matter effectively stood there for several decades, with EPR having shown that such a model is needed

to account for the quantum correlations in a local way, but with most physicists believing that Bohr had somehow refuted the EPR argument and therefore ignoring the issue entirely. As we will see, though, Bell moved the issue forward in the 1960s by asking: could this same local hidden variable model continue to reproduce the quantum mechanical predictions in a more general setting, where spin measurement along more and different axes are also allowed?

8.2 A Preliminary Bell Inequality

As we just showed, it is rather straightforward to reproduce the quantum mechanical predictions, for all possible spin measurements along the x- and z-directions, on a pair of entangled spin 1/2 particles, with a hidden variable model in which each particle's x-spin and z-spin are pre-determined. But let us broaden the discussion. Suppose that instead of restricting ourselves to measuring the spins of the particles along the x- and z-directions, we allow spin measurements in arbitrary (not necessarily orthogonal!) directions; and suppose that instead of considering only two possible directions (along which to measure the particles' spins) we allow the experimenter on each side to choose from among *three* possible axes. Let's call the three axes $\hat{a}$, $\hat{b}$, and $\hat{c}$. Notice that, since the singlet state ψ_s can be written

$$\psi_s = \frac{1}{\sqrt{2}}\left[\psi^1_{+a}\psi^2_{-a} - \psi^1_{-a}\psi^2_{+a}\right] \tag{8.5}$$

– and identically for $\hat{b}$ and $\hat{c}$ – it is clear that, according to quantum mechanics, the outcomes should be perfectly anti-correlated (either "up-down" or "down-up") whenever the two experimenters happen to measure their particles' spins along the same axis. Agreement with this aspect of the quantum mechanical predictions – and continuing to insist on locality – again requires a "hidden variable" theory of the sort we considered before, and requires in particular that the values of (for example) a-spin (and then identically for b-spin and c-spin) should be perfectly anti-correlated. A little contemplation reveals that there are now *eight* types of particle pairs which might be produced (with nonzero frequency) when particle pairs are prepared in the quantum state ψ_s. The types are described in the following table:

Pair type	Particle 1	Particle 2	Frequency
1	$\{+1,+1,+1\}$	$\{-1,-1,-1\}$	F_1
2	$\{+1,+1,-1\}$	$\{-1,-1,+1\}$	F_2
3	$\{+1,-1,+1\}$	$\{-1,+1,-1\}$	F_3
4	$\{-1,+1,+1\}$	$\{+1,-1,-1\}$	F_4
5	$\{+1,-1,-1\}$	$\{-1,+1,+1\}$	F_5
6	$\{-1,+1,-1\}$	$\{+1,-1,+1\}$	F_6
7	$\{-1,-1,+1\}$	$\{+1,+1,-1\}$	F_7
8	$\{-1,-1,-1\}$	$\{+1,+1,+1\}$	F_8

As before, the lists in the "Particle 1" and "Particle 2" columns tell us how a particle which is a member of the indicated pair type will respond to all three possible questions that might be put to it. So, for example, "$\{+1, -1, +1\}$" means that the particle will be found spin-up along $\hat{a}$ (if so measured!), spin-down along $\hat{b}$ (if so measured!), and spin-up along $\hat{c}$ (if so measured!).

Notice that the frequencies F_i of the 8 different pair types are left unspecified. The hope, of course, will be to pick values (as we were able to do in the previous section) so that the statistical predictions of the local hidden variable theory will agree with those of quantum mechanics.

But, as it turns out, this is impossible. The proof that it is impossible is, of course, "Bell's theorem", which involves showing that the predictions of the local hidden variable theory we are considering are constrained by a certain inequality ("Bell's inequality") that the quantum mechanical predictions do not respect. In short, there will be situations where – no matter exactly how the frequencies F_i are selected – the local hidden variable theory cannot reproduce the quantum mechanical statistics. We develop the proof in the remainder of this section in a way that is a little simpler than what Bell did in his original 1964 paper; in subsequent sections we return to consider Bell's own way of presenting things.

Notice first that we can express probabilities for specific possible outcomes in terms of the frequencies F_i that appear in the table. For example, suppose that particle 1 is measured along the $\hat{a}$ direction and particle 2 is measured along the $\hat{b}$ direction. What, for example, is the probability $P_{ab}(++)$ that both measurements have outcome "spin-up"? To answer, we can simply scan down the table on the previous page and look for the pair types for which this will occur. In particular, here, we need a "$+1$" as the first entry in the Particle 1 column (indicating that Particle 1 will be measured "spin-up" in the $\hat{a}$ direction) and a "$+1$" as the second entry in the Particle 2 column (indicating that Particle 2 will be measured "spin-up" in the $\hat{b}$ direction).

I find the appropriate entries in row 3 and row 5. This means that pairs of type 3 and type 5 will yield the outcomes "particle 1 is spin-up along $\hat{a}$" and "particle 2 is spin-up along $\hat{b}$". (Pairs of all the other types will yield at least one different outcome if the particles' spins are measured along $\hat{a}$ and $\hat{b}$ respectively.) And so the probability of seeing that particular outcome ("$++$") is just the probability that a given particle pair is of type 3 or type 5. That is:

$$P_{ab}(++) = F_3 + F_5. \tag{8.6}$$

Let's practice with a couple of other examples. What is the probability $P_{bc}(++)$ of seeing both particles "spin-up" given that particle 1 is measured along the $\hat{b}$ direction and particle 2 is measured along the $\hat{c}$ direction? I find:

$$P_{bc}(++) = F_2 + F_6. \tag{8.7}$$

And similarly

$$P_{ac}(++) = F_2 + F_5. \tag{8.8}$$

Make sure you see where these equations are coming from (and make sure you agree with what I wrote!).

Now, amazingly, we are already in a position to write down a (preliminary example of a) Bell inequality. Since the F_i's represent the frequencies with which pairs of different types are supposed to be produced when we create a particle pair in the singlet state, they must all be positive and they should add to 1. And so it must be the case that

$$F_2 + F_5 \leq F_3 + F_5 + F_2 + F_6 \tag{8.9}$$

since the right hand side is the same as the left hand side plus two additional terms which cannot be smaller than zero! But this means that, for a local hidden variable theory of the sort being considered here, it must be the case that

$$P_{ac}(++) \leq P_{ab}(++) + P_{bc}(++). \tag{8.10}$$

That is, no matter how we pick the frequencies F_i, a theory in which spin measurements simply reveal pre-existing values will have to make statistical predictions that obey Equation (or actually, Inequality) (8.10).

Now the incredible thing is that it is possible to choose directions $\hat{a}$, $\hat{b}$, and $\hat{c}$ such that this same inequality is *violated* by the quantum mechanical predictions. So let us recall in more detail how the relevant quantum predictions come about. In Chap. 2, we saw that the one-particle spin states $\psi_{\pm n}$ (corresponding to particles being definitely spin up or definitely spin down along the $\hat{n}$ axis, an angle θ down from the z-axis in the x-z-plane) were given by

$$\psi_{+n} = \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} = \cos(\theta/2)\psi_{+z} + \sin(\theta/2)\psi_{-z} \tag{8.11}$$

and

$$\psi_{-n} = \begin{pmatrix} \sin(\theta/2) \\ -\cos(\theta/2) \end{pmatrix} = \sin(\theta/2)\psi_{+z} - \cos(\theta/2)\psi_{-z}. \tag{8.12}$$

It is fairly straightforward to invert this relationship (solving for $\psi_{\pm z}$ in terms of $\psi_{\pm n}$). The result is that

$$\psi_{+z} = \cos(\theta/2)\psi_{+n} + \sin(\theta/2)\psi_{-n} \tag{8.13}$$

and

$$\psi_{-z} = \sin(\theta/2)\psi_{+n} - \cos(\theta/2)\psi_{-n}. \tag{8.14}$$

This allows us to rewrite the singlet state as follows:

$$\psi_s = \frac{1}{\sqrt{2}}\left[\sin\left(\frac{\theta}{2}\right)\psi^1_{+z}\psi^2_{+n} - \cos\left(\frac{\theta}{2}\right)\psi^1_{+z}\psi^2_{-n} - \cos\left(\frac{\theta}{2}\right)\psi^1_{-z}\psi^2_{+n} - \sin\left(\frac{\theta}{2}\right)\psi^1_{-z}\psi^2_{-n}\right]. \tag{8.15}$$

From this, we can read off (as the square of the coefficient in front of the $\psi^1_{+z}\psi^2_{+n}$ term) the probability of seeing two "spin-up" outcomes when we measure particle 1 along the z-axis and particle 2 along a direction that is an angle θ away from the z-axis. This is simply:

$$P_{z,\theta}(++) = \frac{1}{2}\sin^2(\theta/2). \tag{8.16}$$

Since (as we have seen) the singlet state is symmetrical – and since what direction we choose to call the z-direction is ultimately arbitrary – this formula turns out to give the quantum mechanical probability for a "++" outcome whenever the two measurement directions have an angle θ between them (whether one of them is the "z-axis" or not). So we can use this general formula now to compute the quantum mechanical prediction for all three of the probabilities that appeared in Eq. (8.10), our baby Bell inequality.

Suppose we pick the three directions $\hat{a}$, $\hat{b}$, and $\hat{c}$ as follows: pick $\hat{b} = \hat{z}$, and then pick $\hat{a}$ and $\hat{c}$ to be tilted at angle θ away from the z-axis, in opposite directions. Then we have that, according to QM,

$$P^{QM}_{ab}(++) = \frac{1}{2}\sin^2(\theta/2) \tag{8.17}$$

and similarly

$$P^{QM}_{bc}(++) = \frac{1}{2}\sin^2(\theta/2). \tag{8.18}$$

What about $P^{QM}_{ac}(++)$? The angle between $\hat{a}$ and $\hat{c}$ is 2θ so the general formula gives

$$P^{QM}_{ac}(++) = \frac{1}{2}\sin^2(\theta). \tag{8.19}$$

Now it is a plain and simple mathematical fact that

$$\frac{1}{2}\sin^2(\theta) \leq \frac{1}{2}\sin^2(\theta/2) + \frac{1}{2}\sin^2(\theta/2) \tag{8.20}$$

is *false* for $\theta \leq \pi/2$. The biggest violation of the inequality occurs for $\theta = \pi/3 = 60°$. In that case we have

$$P^{QM}_{ab}(++) = \frac{1}{8}, \tag{8.21}$$

$$P^{QM}_{bc}(++) = \frac{1}{8}, \tag{8.22}$$

and

$$P^{QM}_{ac}(++) = \frac{3}{8}. \tag{8.23}$$

And, as plain as day, it is not the case that

$$\frac{3}{8} \leq \frac{1}{8} + \frac{1}{8}. \tag{8.24}$$

Bell's inequality is violated by the quantum mechanical predictions.

And so there is, in principle, a certain kind of experiment that we could do to test whether quantum mechanics is right, or the type of local hidden variable theory suggested by EPR is right. The two theories make empirically distinguishable predictions. What is the experiment, exactly? Well, we would need to produce a bunch of spin 1/2 particle pairs, have them fly off in opposite directions toward Stern–Gerlach measuring devices which could be oriented along one of the possible three directions ($\hat{a}$, $\hat{b}$, or $\hat{c}$). For reasons that we will discuss further in the next section, it should ideally be the case that the measurement direction on each side be chosen randomly and at the last possible second before the particles arrive.

Then we would simply keep track of how often, when the $\hat{a}$-spin of particle 1 was measured and the $\hat{b}$-spin of particle 2 was measured, both measurements yielded the "spin-up" outcome. That is, we would empirically measure $P_{ab}(++)$. And similarly for $P_{bc}(++)$ and $P_{ac}(++)$. And then at the end of the day we would compare these probabilities. If Bell's inequality

$$P_{ac}(++) \leq P_{ab}(++) + P_{bc}(++) \tag{8.25}$$

was respected by the experimental data, it would constitute a refutation of Quantum Mechanics and a vindication of the local hidden variables theory; whereas if the measured probabilities agreed with the QM predictions it would constitute an experimental refutation of the local hidden variable theory.

8.3 The Real Bell (and the CHSH) Inequality

Although the experiment just sketched would indeed be possible and gives one the flavor of how a Bell inequality can be experimentally tested, the inequality derived in the previous section is really just a kind of preliminary toy example of a Bell-type inequality. In this section we show how the real Bell inequality (that is, the actual inequality first derived by Bell in his 1964 paper), as well as the closely-related "CHSH Inequality," can be developed. These, as it turns out, suggest far superior types of experimental tests which we will discuss in the following section.

To motivate this discussion, perhaps it is worth thinking about why the experimental test proposed at the end of the last section is somehow less than ideal. Part of the answer is simply that it is very inefficient: if the directions (along which to measure the spins of the two particles) are selected randomly for each particle pair, then only about $1/3$ of the time will we happen to make one of the three types of measurements (namely, ab, bc, or ac) that are relevant to Eq. (8.10). So $2/3$ of the data – $2/3$ of the particle pairs produced – are simply wasted.

You are probably thinking that one could eliminate the waste by just fixing the Stern–Gerlach device on the particle 1 side in the $\hat{a}$ direction, and similarly fixing the Stern–Gerlach device on the particle 2 side in the $\hat{b}$ direction, collecting data for (say) a million particle pairs, then switching the detectors to the *bc* orientations, collecting more data, and then finally switching to the *ac* orientations and collecting a last set of data. And that is true. You could do that (and some of the early experiments were along these lines). But (as hinted at previously) it is important that the orientations of the measuring devices be set randomly and at the last possible second before the particles arrive – ideally, so late that the measurement on particle 1 cannot be influenced (by any signal propagating at the speed of light or slower) by the orientation of particle 2's measuring device (or vice versa).

To understand this, imagine that both detectors are just fixed in place – say, in the *ac* orientation – for a run of many particle pairs. Then each particle pair could "know", already when it is created, and without any violation of local causality, that particle 1 will be measured along the $\hat{a}$ direction and particle 2 will be measured along the $\hat{c}$ direction. But one could imagine that, in such a circumstance, the particle source would be free to emit particles not only of the 8 types captured in our earlier table, but also "rogue" types in which, for example, particle 1 has properties $\{+1, -1, -1\}$ and particle 2 has properties $\{+1, -1, +1\}$. After all, if the particles "know" in advance that they will each be measured along particular, pre-determined directions, there is no reason the pre-existing spin components along all three directions should have to be perfectly correlated. Similarly, if the particle pairs "know", in advance of being emitted, the directions along which their spins will later be measured, it might be possible for the source to adjust the frequencies F_i in response: for example, perhaps when the detectors are in the *ab* configuration, F_1 is big and F_2 is small, whereas when the detectors are in the *bc* configuration, F_1 is small and F_2 is big.

It should be clear that in either of these scenarios ("rogue" particle types, or measurement-axis-dependent pair frequencies), the straightforward type of hidden variable model we've described is no longer required and our derivation of the inequality would not go through. Turning this around, then, it should make sense that, in order for the straightforward type of hidden variable model we've described to genuinely be required by locality, we must have in mind an experimental setup in which it is impossible for the particles to know in advance along which axes their spins will be measured. So if at the end of the day we want a clean experimental discrimination between Quantum Mechanics and the sort of local theory implied by the EPR-Bohm argument, the experimental test must implement strict "Einstein locality" conditions in which the device setting on each side is only determined while the particles are in flight, and, indeed, determined sufficiently late that information about it is locally inaccessible to the measurement on the other side.

We will discuss this point a bit more in the following sections; for now, suffice it to say that for a variety of technical and practical reasons, it would be nice to develop a Bell type inequality that doesn't focus so narrowly on one specific outcome for each of just three possible measurements, but instead embraces all possible measurements and outcomes in a more democratic way.

To begin to develop such an inequality, let us consider the "correlation coefficient" $C(\hat{n}_1, \hat{n}_2)$, defined as the expected value of the product of the two measurement outcomes when measurements are made along direction $\hat{n}_1$ and $\hat{n}_2$ on the two particles. Since the product of the outcomes is either $+1$ (if the outcomes are "++" or "−−") or -1 (if the outcomes are "+−" or "−+"), the correlation coefficient is thus the probability for "++" plus the probability for "−−" but then minus the probability for "+−" and minus the probability for "−+":

$$C(\hat{n}_1, \hat{n}_2) = P_{n_1 n_2}(++) + P_{n_1 n_2}(--) - P_{n_1 n_2}(+-) - P_{n_1 n_2}(-+). \tag{8.26}$$

Using the same technique we used in the previous section, i.e., reading off the probabilities as the squares of the coefficients of the four terms in Eq. (8.15), it is easy to work out that the Quantum Mechanical prediction for the correlation coefficient is

$$C^{QM}(\hat{n}_1, \hat{n}_2) = \frac{1}{2}\sin^2\left(\frac{\theta}{2}\right) + \frac{1}{2}\sin^2\left(\frac{\theta}{2}\right) - \frac{1}{2}\cos^2\left(\frac{\theta}{2}\right) - \frac{1}{2}\cos^2\left(\frac{\theta}{2}\right) = -\cos(\theta) \tag{8.27}$$

where θ is the angle between $\hat{n}_1$ and $\hat{n}_2$. Note that, for $\theta = 0$, $C = -1$, meaning that the two outcomes are perfectly anti-correlated (always opposite). Whereas for $\theta = 90°$, $C = 0$, meaning that there is no correlation at all between the outcomes. This is all consistent with what we have seen before, namely, that if the two particles' spins are measured along the same direction, the individual outcomes are necessarily opposite, whereas if they are measured along orthogonal axes, a "+1" outcome on one side is equally likely to be accompanied by a "+1" or a "−1" on the other side, and so on.

What about the local hidden variable theory described in the previous section? Writing

$$C(\hat{n}_1, \hat{n}_2) = P_{n_1 n_2}(++) + P_{n_1 n_2}(--) - P_{n_1 n_2}(+-) - P_{n_1 n_2}(-+) \tag{8.28}$$

we see that each of the four probabilities on the right hand side can be expressed in terms of the frequencies F_i from our table. For given measurement directions $\hat{n}_1$ and $\hat{n}_2$, each frequency will appear, either with a plus sign or a minus sign. For example:

$$C(\hat{a}, \hat{b}) = F_3 + F_5 + F_4 + F_6 - F_1 - F_2 - F_7 - F_8. \tag{8.29}$$

(Take a minute and make sure you understand exactly how I got this!) Similarly:

$$C(\hat{a}, \hat{c}) = F_2 + F_5 + F_4 + F_7 - F_1 - F_3 - F_6 - F_8 \tag{8.30}$$

and

$$C(\hat{b}, \hat{c}) = F_2 + F_3 + F_6 + F_7 - F_1 - F_4 - F_5 - F_8. \tag{8.31}$$

But then, as before, obviously-true inequalities involving the F_i can be seen to be equivalent to inequalities involving the correlation coefficients for different measurement settings. For example, the trivial inequality

$$|F_3 + F_6 - F_7 - F_2| \leq F_2 + F_3 + F_6 + F_7 \tag{8.32}$$

turns out to be equivalent to

$$\left|C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{c})\right| \leq 1 + C(\hat{b}, \hat{c}). \tag{8.33}$$

This is actually the original "Bell inequality" that Bell derived (using a somewhat different method) in his original 1964 paper, and it is easy to see that it is violated by the Quantum Mechanical predictions. For $\hat{a}$, $\hat{b}$, and $\hat{c}$ as before (with $\hat{b}$ in the middle and then $\hat{a}$ and $\hat{c}$ θ away in opposite directions) the greatest violation is again achieved for $\theta = \pi/3 = 60°$. So then $C^{QM}(\hat{a}, \hat{b}) = -\cos(60°) = -1/2$ and $C^{QM}(\hat{a}, \hat{c}) = -\cos(120°) = +1/2$ and $C^{QM}(\hat{b}, \hat{c}) = -1/2$ so we have

$$\left|-\frac{1}{2} - \frac{1}{2}\right| \leq 1 - \frac{1}{2} \tag{8.34}$$

which reduces to

$$1 \leq \frac{1}{2} \tag{8.35}$$

which is definitely not true! So the QM predictions violate Bell's Inequality, which is a constraint on the correlations (between the outcomes on the two sides) for the kind of theory implied by the EPR-Bohm argument.

One can develop a second, closely-related Bell-type inequality by starting with the trivial inequality

$$|F_5 + F_4 - F_1 - F_8| \leq F_1 + F_4 + F_5 + F_8 \tag{8.36}$$

which turns out to be equivalent to

$$\left|C(\hat{a}', \hat{b}) + C(\hat{a}', \hat{c})\right| \leq 1 - C(\hat{b}, \hat{c}) \tag{8.37}$$

where $\hat{a}'$ may be (but need not be!) the same direction as the previous $\hat{a}$.

Note, finally, that by adding Equations (actually, inequalities) (8.33) and (8.37) we arrive at

$$\left|C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{c})\right| + \left|C(\hat{a}', \hat{b}) + C(\hat{a}', \hat{c})\right| \leq 2 \tag{8.38}$$

which is a particularly important Bell-type inequality called the "CHSH inequality" (after Clauser, Horne, Shimony, and Holt who first derived it in 1969) [3].

The CHSH inequality is particularly well-suited to experimental test, because only two measurement angles appear in each "wing" of the experiment. That is, suppose that for each particle pair that is created, Particle 1 is sent toward Alice, who will (randomly and at the last possible second) choose to measure the spin of Particle 1 along either the direction $\hat{a}$ or the direction $\hat{a}'$. And similarly, Particle 2 is sent toward Bob, who will (randomly and at the last possible second) choose to measure the spin of Particle 2 along either the direction $\hat{b}$ or the direction $\hat{c}$.

Alice and Bob then record their respective outcomes for that particular particle pair, and get ready for the particles from the next pair to arrive. After collecting data for some time, Alice and Bob meet somewhere and compare notes. Importantly, every single outcome of every single measurement on every single particle pair gets used to determine an empirical value for one of the four correlation coefficients appearing in the CHSH inequality. So none of the data is wasted.

What does QM predict for the CHSH parameter? Well, suppose we pick the directions $\hat{a}$, $\hat{a}'$, $\hat{b}$, and $\hat{c}$ as follows: $\hat{a}$ will be the z-axis, and $\hat{a}'$ will be the x-axis. Then $\hat{b}$ will be halfway between the x- and z-axes, i.e., 45° down from the z-axis toward the x-axis. And $\hat{c}$ will be 45° away from the x-axis in the other direction, i.e., 135° down from the z-axis toward the x-axis, i.e., halfway between the x-axis and the negative z-axis. See Fig. 8.1.

These turn out to be the directions that give the greatest possible discrepancy between the local hidden variable and the quantum mechanical predictions. To see exactly what QM predicts, note that the angle between $\hat{a}$ and $\hat{b}$ is 45°, so, according to Eq. (8.27)

$$C^{QM}(\hat{a}, \hat{b}) = -\cos(45°) = -1/\sqrt{2}. \tag{8.39}$$

Similarly, we have that

$$C^{QM}(\hat{a}, \hat{c}) = +1/\sqrt{2}, \tag{8.40}$$

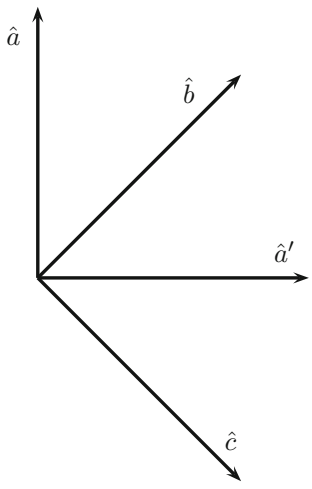


Fig. 8.1 The four directions that give the greatest discrepancy between the local hidden variable and quantum mechanical predictions for the CHSH inequality

$$C^{QM}(\hat{a}', \hat{b}) = -1/\sqrt{2}, \tag{8.41}$$

and

$$C^{QM}(\hat{a}', \hat{c}) = -1/\sqrt{2}. \tag{8.42}$$

Putting these together, we have that

$$\left|C^{QM}(\hat{a}, \hat{b}) - C^{QM}(\hat{a}, \hat{c})\right| + \left|C^{QM}(\hat{a}', \hat{b}) + C^{QM}(\hat{a}', \hat{c})\right| = 2\sqrt{2}. \tag{8.43}$$

This is a factor of $\sqrt{2}$ – i.e., about 40% – bigger than should be allowed if a local hidden variable theory is true.

8.4 Experiments

Let us then review some of the actual experiments of this sort.

The first really systematic test of the CHSH inequality was done by Alain Aspect and collaborators in 1982 [4]. Instead of using pairs of spin 1/2 particles in the spin singlet state, they used pairs of photons (emitted from excited Calcium atoms) whose *polarizations* are entangled in a way that is perfectly analogous to the singlet state we've been discussing. Note, however, that there is a factor-of-2 difference between the angles involved in the spin 1/2 case and the photon polarization case. Whereas, for example, the two possible spin directions of a spin 1/2 particle are *opposite* ("up" and "down", different by 180°) the two possible polarizations of a photon are *orthogonal* (e.g., "horizontal" and "vertical", different by 90°). In addition, whereas spin-1/2 particles in the singlet state display perfect anti-correlation (when their spins are measured along the same axes), the photon pairs instead display perfect (positive) correlation (when their polarizations are measured along the same axes). So the quantum prediction for the polarization correlation coefficient $C^{QM}(\hat{n}_1, \hat{n}_2)$ in the case of photons is $\cos(2\theta)$ rather than the $-\cos(\theta)$ we saw previously for the case of spin-1/2 particles. But otherwise everything is just as we've been discussing.

A schematic diagram of the experiment and a graph of their data are reproduced (from the 1982 paper) in Fig. 8.2. There is essentially perfect agreement between the experimental results and the QM predictions, and the CHSH parameter (that is, the combination of correlation coefficients that can be no greater than 2 for local hidden variable theories) was

$$S_{expt} = 2.697 \pm 0.015, \tag{8.44}$$

i.e., well above the maximum possible value (namely, 2) allowed for the local hidden variable theories.

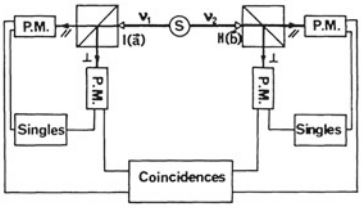


FIG. 2. Experimental setup. Two polarimeters I and II, in orientations $\vec{a}$ and $\vec{b}$, perform true dichotomic measurements of linear polarization on photons ν_1 and ν_2. Each polarimeter is rotatable around the axis of the incident beam. The counting electronics monitors the singles and the coincidences.

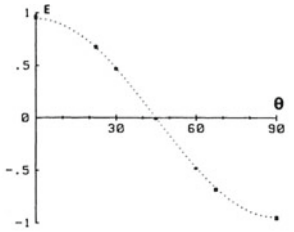


FIG. 3. Correlation of polarizations as a function of the relative angle of the polarimeters. The indicated errors are ±2 standard deviations. The dotted curve is not a fit to the data, but quantum mechanical predictions for the actual experiment. For ideal polarizers, the curve would reach the values ±1.

Fig. 8.2 The *left panel* shows a schematic diagram of the setup for the Aspect et al. experiment, Ref. [4]; the *right panel* shows their data. The essentially perfect agreement with $C^{QM} = \cos(2\theta)$ is evident. In this experiment they found the CHSH parameter to be $S_{expt} = 2.697 \pm 0.015$, i.e., well above the maximum value of 2 allowed for local hidden variable theories and in excellent agreement with the QM prediction (which, for detector efficiency and alignment reasons) for their experiment was $S_{QM} = 2.70 \pm 0.05$ (i.e., slightly less than $2\sqrt{2}$). (Reprinted with permission from A. Aspect, P. Grangier, G. Roger, *Physical Review Letters*, 49, 91–94, 1982, by the American Physical Society.)

Here is how Aspect et al. summarize their conclusions:

> ...our experiment yields the strongest violation of Bell's inequalities ever achieved, and excellent agreement with quantum mechanics. Since it is a straightforward transposition of the ideal Einstein–Podolsky–Rosen-Bohm scheme, the experimental procedure is very simple, and needs no auxiliary measurements as in previous experiments with single-channel polarizers. We are thus led to the rejection of realistic [i.e., "hidden variable"] local theories if we accept the assumption that there is no bias in the detected samples: Experiments support this natural assumption.
>
> Only two loopholes remain open for advocates of realistic theories without action at a distance. The first one, exploiting the low efficiencies of detectors, could be ruled out by a feasible experiment. The second one, exploiting the static character of all previous experiments, could also be ruled out by a 'timing experiment' with variable analyzers now in progress [4].

What is this "assumption that there is no bias in the detected samples"? The idea here is that the committed advocate of a local hidden variable theory could claim that the CHSH inequality wasn't *really* shown to be violated because, actually, only a small fraction of all emitted photon pairs were successfully detected. (This has to do with the fact that the photon source in this experiment emits photon pairs isotropically, so only the occasional pairs which just happen to be aimed right at the detectors, will actually be detected.) "If only" (says the conspiracy theorist here) "*all* of the pairs had been detected, we might have found a CHSH parameter less than 2." And, indeed, that is in principle possible, although it seems exceedingly unlikely. In order for the experimental results to be biased in this way, there would need to be some reason why the polarization correlations between pairs which happen to be aimed at the detectors, are significantly different than those between pairs going in other directions. But neither quantum mechanics nor any other available

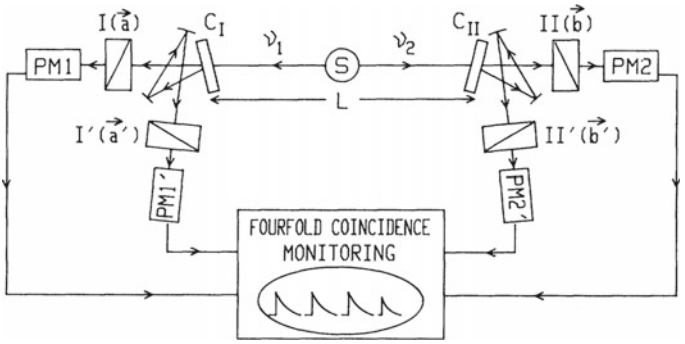


Fig. 8.3 Aspect et al.'s schematic diagram of their "Experimental test of Bell's inequalities using time-varying analyzers" from Ref. [5]. They write: "timing experiment with optical switches. Each switching device (C_I, C_{II}) is followed by two polarizers in two different orientations. Each combination is equivalent to a polarizer switched fast between two orientations." (Reprinted with permission from A. Aspect, J. Dalibard, G. Roger, *Physical Review Letters*, 49, 1804–1807, 1982, by the American Physical Society.)

idea provides any basis for suspecting such a thing. So this way of trying to elude the apparent implications of Aspect's first experiment seems like grasping at straws. Nevertheless, increasing the fraction of particle pairs detected (i.e., working to "close the detector efficiency loophole") is something that experimenters worked on in subsequent experiments.

The second "loophole" that Aspect mentions, however – having to do with the "static character" of this and previous experiments – is a bit of a more serious issue. Indeed, this is precisely the issue I already raised in the previous section: in order to really discriminate quantum mechanics from local hidden variable theories of the sort implied by the EPR-Bohm argument, the measurement devices should not remain static; instead, the settings (that is, for Alice, the choice between measuring along $\hat{a}$ and $\hat{a}'$, and, for Bob, the choice between measuring along $\hat{b}$ and $\hat{c}$) should be made randomly and at the last possible second.

In a follow-up experiment (also published in 1982!) Aspect and his collaborators devised an ingenious mechanism for switching between different measurement settings. (See Fig. 8.3) Instead of physically rotating a polarization measuring device (which would simply not be feasible on the required time scales), each photon was shunted to one or the other of two static measuring devices by an "acousto-optical switch" consisting of a cell of water in which a (roughly) 50 MHz acoustic standing wave is set up. During the part of the cycle in which the amplitude of the standing wave is small, incident photons simply pass straight through. But during the part of the cycle in which the amplitude is large, the water (with now a spatially-varying density) acts like a diffraction grating and photons passing through are deflected at some angle. The 50 MHz standing wave frequency corresponds to about a 10 nanosecond period during which any incident photons are shunted in the one direction, before switching to the other direction. This timescale can be compared to the

time – $L/c \approx 40$ nanoseconds – it would take a signal, propagating at the speed of light, to get from Alice's measuring device to Bob's (or vice versa). If the fast back-and-forth switching between the two measurement directions on either side is considered "effectively random" (and note here that the frequencies were deliberately made incommensurate on the two sides) one can thus say that the choice of measurement settings on the two sides are spacelike separated. As Aspect et al. write:

> The new feature of this experiment is that we change the settings of the polarizers, at a rate greater than c/L. The ideal scheme has not been completed since the change is not truly random, but rather quasiperiodic. Nevertheless, the two switches on the two sides are driven by different generators at different frequencies. It is then very natural to assume that they function in an uncorrelated way.
>
> A more ideal experiment with random and complete switching would be necessary for a fully conclusive argument against the whole class of supplementary-parameter [i.e., hidden variable] theories obeying Einstein's [local] causality. However, our observed violation of Bell's inequalities indicates that the experimental accuracy was good enough for pointing out a hypothetical discrepancy with the predictions of quantum mechanics. No such effect was observed [5].

Thus, again in this improved version of the experiment, results consistent with QM – and inconsistent with a Bell-type inequality – were observed.

Some of the possible improvements described by Aspect et al. were implemented by a new experiment performed in 1998 by Gregor Weihs, Anton Zeilinger, and collaborators [6]. They used a new (more controllable and more efficient) source of polarization-entangled photon pairs (called "type-II parametric down-conversion"). The photons were carried from the central source to the polarization measuring stations (at opposite ends of the campus of the University of Innsbruck in Austria) along fibre optic cables. And, crucially, the experiment used "high speed physical random number generators and fast electro-optic modulators" to arrange for the choice of measurement axis (along which each particle's polarization was measured) to be made, for the first time, genuinely randomly and at unambiguously space-like separation from the similar choice occuring in the other wing of the experiment.[1] Weihs et al. report typical measured CHSH parameter values of $S = 2.73 \pm 0.02$ – in excellent agreement with the quantum mechanical predictions (as applicable to their particular experimental setup) and in clear violation of the CHSH inequality.

The experiment by Weihs et al. remained essentially the state-of-the-art until quite recently, when *three* different groups published results showing that the QM predictions remain correct even when the various "loopholes" described originally by Aspect are simultaneously closed [7–9].

[1] Technical detail: actually, instead of using the output of the random number generator to physically rotate the polarization measuring device (which could never be done quickly enough), the output was fed into an electro-optic modulator which rotated (by one of two possible amounts) the polarization of the incoming photon.

8.5 What Does It Mean?

I have been describing Bell's inequalities as constraints on the predictions of local hidden variable theories of the sort implied by the EPR-Bohm argument we reviewed at the beginning of the chapter. From this point of view, the experiments we surveyed in the last section prove (to me, at least, quite convincingly) that the predictions of QM are *correct* and the predictions of the local hidden variable theories are simply *wrong*. And leaving aside the conspiracy theorists, everybody agrees about this.

There is, however, a surprising amount of controversy about what, exactly, should be inferred from the empirical violations of Bell inequalities.

In particular, many people have taken Bell's theorem as a proof that hidden variables theories are not viable.[2] Eugene Wigner, for example, wrote (about the possibility of a hidden variables "completion" of ordinary QM) that the proof of von Neumann "uses assumptions which, in my opinion, can quite reasonably be questioned." (Here Wigner was in total agreement with Bell, who recall showed in his 1966 paper that von Neumann's assumptions were in fact totally arbitrary and unwarranted.) But Wigner goes on to state:

> In my opinion, the most convincing argument against the theory of hidden variables was presented by J.S. Bell [10].

A similar remark has been made by the eminent theoretician Rudolf Peierls:

> If people are obstinate in opposing the accepted view they can think of many new possibilities, but there is no sensible view of hidden variables which doesn't conflict with these experimental results [i.e., Aspect's experiments]. That was proved by John Bell, who has great merit in establishing this. Prior to that there was a proof due to the mathematician von Neumann, but he made an assumption which is not really necessary [11].

More recently, Stephn Hawking summarized the situation as follows:

> Einstein's view was what would now be called a hidden variable theory. Hidden variable theories might seem to be the most obvious way to incorporate the Uncertainty Principle into physics. They form the basis of the mental picture of the universe, held by many scientists, and almost all philosophers of science. But these hidden variable theories are wrong. The British physicist, John Bell ... devised an experimental test that would distinguish hidden variable theories. When the experiment was carried out carefully, the results were inconsistent with hidden variables. Thus it seems that even God is bound by the Uncertainty Principle.... God does play dice with the universe [12].

It is easy to multiply examples. In a review article on "One Hundred Years of Quantum Physics", Daniel Kleppner and Roman Jackiw of MIT wrote, about the experiments reviewed in the last section, that "[t]heir collective data came down decisively against the possibility of hidden variables. For most scientists this resolved any doubt about the validity of quantum mechanics" [13]. And in a similar article celebrating the 100 year anniversary of QM, Max Tegmark and John Wheeler wrote the following:

---

[2]Most of the quotes in the following paragraph were collected by Jean Bricmont in Sect. 7.5 of *Making Sense of Quantum Mechanics*.

> Could the apparent quantum randomness be replaced by some kind of unknown quantity carried out inside particles, so-called 'hidden variables'? CERN theorist John Bell showed that in this case, quantities that could be measured in certain difficult experiments would inevitably disagree with standard quantum predictions. After many years, technology allowed researchers to conduct these experiments and eliminate hidden variables as a possibility [14].

In their preface to the published proceedings of a conference honoring Bell 10 years after his death, Reinhold Bertlmann (a long-time colleague, collaborator, and friend of Bell's... recall his mis-matched socks from Chap. 4!) and Anton Zeilinger (one of the authors of the Innsbruck experiment paper discussed in the last section and one of the most prominent experimental quantum physicists) explained how, although Bell had seemingly opened the door to hidden variables by refuting von Neumann's supposed impossibility proof,

> he immediately dealt them [i.e., hidden variables] a major blow. In 1964 ... he showed that any hidden variables theory, which obeys Einstein's requirement of locality, i.e., no influence travelling faster than the speed of light, would automatically be in conflict with quantum mechanics. [...] While a very tiny [experimental] loophole in principle remains for local realism, it is a very safe position to assume that quantum mechanics has definitely been shown to be the right theory. Thus, a very deep philosophical question, namely, whether or not events observed in the quantum world can be described by an underlying deterministic theory, has been answered by experiment, thanks to the momentous achievement of John Bell [15].

What is going on here? How can all these people claim that the experimental violation of Bell's inequality somehow refutes the possibility of an underlying deterministic or hidden variable completion of quantum mechanics, when such a theory (namely, the de Broglie - Bohm pilot-wave theory) already actually exists and is demonstrably consistent with these experiments?

Part of the answer, to be sure, is that most physicists are simply not as aware as they should be about the existence of the pilot-wave theory. They've heard of it but never looked into it and hence don't actually understand how it works and, as we've seen, they dismiss the broad category of hidden variable theories (of which the pilot-wave theory is just one concrete example) on the grounds that they have been ruled out, experimentally, as shown by Bell. There is a kind of rich and tragic irony here, in citing Bell as having supposedly refuted hidden variables theories (and hence entrenching the unjustified belief that the pilot-wave theory cannot be right and must not be worth looking into) when, as we have seen, Bell's theorem was actually inspired by Bohm's 1952 pilot-wave theory papers, and indeed Bell remained far and away the pilot-wave theory's greatest champion until his death in 1990.

But there is more going on, in the citation of Bell's theorem as refuting the hidden variables program, than mere ignorance of the pilot-wave theory. Some of the people who make this kind of argument do know about the pilot-wave theory, and reject it on the grounds that it is non-local and hence in apparent conflict with relativity. This point of view was perhaps best encapsulated by David Mermin's remark:

> To those for whom nonlocality is anathema, Bell's theorem finally spells the death of the hidden-variables program [16].

The idea here, apparently, is that Bell's inequality – which we now know from experiment is *false* – follows from the conjunction of two premises: locality and hidden variables. (Or sometimes, instead, of "hidden variables" people will say "determinism" or "realism" – or something essentially equivalent but a little more cryptic called "counter-factual definiteness".) But if these two premises, together, imply something that is false, at least one of the premises must be wrong. According to this viewpoint, then, we have to *choose* between the following two options:

1. Uphold locality and reject hidden variables, i.e., retain consistency with relativistic causality and admit (as everyone has told us we should have done anyway) that Einstein was wrong and Bohr was right regarding the question of whether the quantum mechanical description of reality can be considered complete.
2. Uphold hidden variables and reject locality, i.e., side with the somewhat senile Einstein in his stubborn, arbitrary, and philosophical demand that "God does not play dice" and insist that, despite being one of the most successful and highly-confirmed theories in the history of science, relativity is somehow wrong.

If those were the two available options, it would indeed be a no-brainer. Obviously we should choose option 1. Selecting option 2 would be crazy.

In this way of looking at the matter, non-locality is the price one has to pay for attempting to restore determinism (and/or "realism") to quantum theory... and the price, obviously, is simply too high. Maintaining consistency with relativistic causality (i.e., locality) is mandatory, and if that means we need to abandon the quest for a more complete underlying model of quantum phenomena, so be it; indeed, most would say, good riddance.

That, I think, captures the viewpoint of the vast majority of physicists today. But, I believe, it is completely and utterly and hopelessly wrong. We do not face anything like the choice between options 1 and 2 above, and indeed, at the end of the day, Bell's theorem tells us absolutely nothing about "hidden variables" or determinism or counter-factual definiteness or "realism" or whether the moon is there or not when you aren't looking [17] or *any* of these sorts of things that people so frequently say it is fundamentally about. Everybody is just simply wrong here, because they have forgotten (or, more commonly, because they never understood in the first place) a crucial part of the broader context of Bell's theorem.

In particular, they have forgotten the EPR argument – which, remember, is supposed to be a proof that deterministic hidden variables are *required, in the first place, precisely in order to avoid non-locality*. The sort of hidden variable theory that Bell's theorem ends up ruling out, that is, is not something that Bell – or for that matter Einstein – just dreamed up. It's not something they just liked or randomly felt like considering. It's something they considered specifically because they recognized it as the only possible hope for maintaining locality in the face of the perfect EPR correlations.

Bell's theorem, then – taken here to mean the proof that local hidden variable theories are wrong – must be understood as the second part of an overall two-part argument, the first part of which is the EPR argument. Schematically, the two-part argument goes like this:

EPR: locality → X

Bell: X → conflict with experiment

Here "X" stands for something like "local deterministic hidden variables", but somehow the logic is easier to grasp by suppressing this. Obviously, if locality → X, and X in turn implies a conflict with experiment, then we cannot maintain X, which means we cannot maintain locality (because locality entails X!).

So according to this view, what should be concluded from the experimentally observed violations of Bell-type inequalities is not that we cannot have hidden variables (we can!), and not even that we must choose between hidden variables and locality. It only appears that we face such a choice if we look *only* at the second, Bell-part of the two-part argument. But if we remember also the first, EPR-part of the argument, we remember that the choice is highly constrained: keeping locality but abandoning hidden variables is not an available option at all. We must, that is, simply conclude that locality – that the prohibition on faster-than-light causation that seems somehow to be implied by relativity theory – is false. Relativistic local causality is wrong, is in conflict with experimental data. Faster-than-light causal influences really exist in Nature!

That is, to be sure, a shocking conclusion and raises all kinds of pressing questions that proliferate in all directions. But we will not be able to pursue them in detail here. I will instead close this section by sharing that this "alternative" view – according to which the upshot of Bell's theorem is that locality is false – is not only my view (and that of some other contemporary physicists and philosophers of science), but was also the view of the person in the best possible position to understand Bell's reasoning: Bell himself.

In his introductory remarks at a 1984 conference, for example, Bell said that "the real problem with quantum theory" is the "essential conflict between any sharp formulation and fundamental relativity" and went on to speak of the "incompatibility, at the deepest level, between the two fundamental pillars of contemporary theory" (meaning quantum theory and relativity theory) [18].

Indeed, Bell even went so far as to suggest, in response to his theorem and the relevant experimental data, the rejection of "fundamental relativity" and the return to a Lorentzian view in which there is a dynamically privileged (though probably empirically undetectable) reference frame:

> "It may well be that a relativistic version of [quantum] theory, while Lorentz invariant and local at the observational level, may be necessarily non-local and with a preferred frame (or aether) at the fundamental level" [19].

And elsewhere:

> "... I would say the cheapest resolution is something like going back to relativity as it was before Einstein, when people like Lorentz and Poincaré thought that there was an aether – a preferred frame of reference – but that our measuring instruments were distorted by motion in such a way that we could not detect motion through the aether. Now, in that way you can imagine that there is a preferred frame of reference, and in this preferred frame of reference things do go faster than light. .... Behind the apparent Lorentz invariance of the phenomena, there is a deeper level which is not Lorentz invariant... [This] pre-Einstein

> position of Lorentz and Poincaré, Larmor and Fitzgerald, was perfectly coherent, and is not inconsistent with relativity theory. The idea that there is an aether, and these Fitzgerald contractions and Larmor dilations occur, and that as a result the instruments do not detect motion through the aether – that is a perfectly coherent point of view" [20].

Why did Bell take so seriously these sorts of ideas, which everybody else today regards as completely outmoded and wrong? Because he thinks his theorem (and the associated experimental evidence) proves that nonlocality is a fact of Nature, rather than merely a defect of a type of theory we shouldn't believe in.

And, as I have tried to explain, he thinks that because he sees his theorem as building from where the EPR argument left off. He makes his reasoning particularly clear in his classic 1981 paper, "Bertlmann's Socks and the Nature of Reality", in which he reacts against the confusion described above (namely, forgetting about the EPR argument and hence inferring a completely wrong conclusion from the theorem) by laying out the two-part argument:

> "Let me summarize once again the logic that leads to the impasse. The EPRB correlations are such that the result of the experiment on one side immediately foretells that on the other, whenever the analyzers happen to be parallel. If we do not accept the intervention on one side as a causal influence on the other, we seem obliged to admit that the results on both sides are determined in advance anyway, independently of the intervention on the other side, by signals from the source and by the local magnet settings. [That's the first, EPR-part of the argument.] But this has implications for non-parallel settings which conflict with those of quantum mechanics. [That's the second part, what is usually (alone) called "Bell's theorem".] So we *cannot* dismiss intervention on one side as a causal influence on the other" [21].

The last sentence expresses the overall conclusion of the two-part argument, that (something about) the measurement on one side *does* influence, faster than light, the results on the other side.

Earlier in the same paper, Bell rehearses the EPR argument and then underscores its logical structure as follows:

> "It is important to note that to the limited degree to which *determinism* plays a role in the EPR argument, it is not assumed but *inferred*. What is held sacred is the principle of 'local causality' - or 'no action at a distance'. Of course, mere *correlation* between distant events does not by itself imply action at a distance, but only correlation between the signals reaching the two places. These signals, in the idealized example of Bohm, must be sufficient to *determine* whether the particles go up or down. For any residual undeterminism could only spoil the perfect correlation.
>
> "It is remarkably difficult to get this point across, that determinism is not a *presupposition* of the analysis. There is a widespread and erroneous conviction that for Einstein[*] determinism was always *the* sacred principle. The quotability of his famous 'God does not play dice' has not helped in this respect" [21].

The footnote referred to after the mention of Einstein reads:

> "And his followers. My own first paper on this subject [i.e., Bell's 1964 paper presenting "Bell's theorem"] starts with a summary of the EPR argument *from locality to* deterministic hidden variables. But the commentators have almost universally reported that it begins with deterministic hidden variables" [21].

I personally find this footnote remarkable and extremely revealing. Bell describes himself as a follower of Einstein (meaning, presumably, that for him, like for Einstein, it is 'local causality' rather than determinism which is *the* sacred principle) and then says explicitly that "the commentators have almost universally" misunderstood Bell's theorem (as presented in his original 1964 paper) because they have failed to appreciate the relevance of "the EPR argument *from locality to* deterministic hidden variables."

8.6 (Bell's) Locality Inequality Theorem

As we saw, Bell claimed rather unambiguously in 1981 that his original 1964 paper "starts with a summary of the EPR argument *from locality to* deterministic hidden variables". Whether or not Bell had actually had this two-part argument in mind from the beginning, however, has been the focus of some discussion and debate during the recent 50th anniversary celebration of Bell's paper [22]. Suffice it to say that, on the one hand, it is clear that the introductory sections of Bell's 1964 paper begin by reminding the reader of what Einstein et al. had already established, several decades earlier. But, on the other hand, Bell's summary of the EPR argument is indeed somewhat unfortunately brief and informal.

Happily, though, Bell continued to write and give talks about "Bell's theorem" throughout the period between 1964 and 1990, and in these talks and papers we see a systematic attempt to clarify, sharpen, and make more explicit several aspects of the reasoning, and thereby to pre-empt the sort of misunderstanding discussed above. There are a couple of threads to this development. One is the thing we have just been focusing on: making the EPR argument and its relationship to his new discovery more explicit and clear.

But another thread, the one I want to focus on here, involved eliminating the middle-man, so to speak – that is, constructing a simpler and more direct demonstration of the incompatibility between local causality and experiment (via an empirically testable inequality). That is, whereas in Bell's earlier presentations, he tends to use the "two-part argument" described in the last section (locality implies deterministic local hidden variables, and then deterministic local hidden variables imply a Bell-type inequality), in his later presentations Bell instead lays out much more explicitly what exactly he means by "locality" and then shows *directly* how locality entails (for example) the CHSH inequality. So I thought it would be good to end the chapter by rehearsing this more direct presentation that represents, I think, Bell's mature sense of what he proved and why it's important.

We have already discussed Bell's formulation of "locality" – way back in Chap. 1 and then again in the context of the EPR argument in Chap. 4 (and again in Chap. 5). See Fig. 8.4 for a brief recap of the idea that, in a theory respecting relativistic local causality, certain information (at appropriate space-like separation from a given event) must necessarily be irrelevant for making predictions about the given event, once what happens in the backward light cone of that event is sufficiently specified.

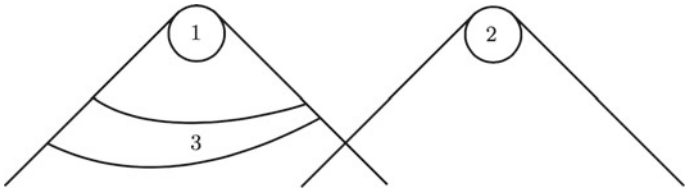


Fig. 8.4 Space-time regions relevant to Bell's formulation of local causality. Bell writes: "Full specification of what happens in 3 makes events in 2 irrelevant for predictions about 1 in a locally causal theory" [23]

Let us then sketch how the empirically-testable CHSH inequality can be derived – *directly* – from Bell's formulation of local causality. Consider the now-familiar sort of experimental setup in which a source creates particle pairs in a state that QM would describe as the singlet state. Of course, we want to be completely general here and allow that perhaps some other (hidden variable type) theory turns out to provide the correct description. So we will say that a given particle pair has state λ. This *might* be just the QM singlet state ψ_s, or it might be the singlet state wave function plus some additional "hidden variables", or it might be something entirely distinct from the QM wave function. We will keep it completely unspecified, completely general.

So then, the particles fly off in opposite directions toward measuring stations womanned, and manned, respectively, by Alice and Bob. Alice uses some kind of random number generator to pick, at the last possible second, an axis $\hat{n}_1$ along which to measure the spin of her particle, and Bob similarly uses an independent random number generator to pick, also at the last possible second, an axis $\hat{n}_2$ along which to measure the spin of his particle. Let's call Alice's outcome A (with, as usual, $A = +1$ meaning "spin up" and $A = -1$ meaning "spin down") and similarly for Bob's outcome, B. All of these things are depicted on the space-time diagram in Fig. 8.5.

Note that we utilize the necessary condition for locality introduced in Chap. 5: the (large!) region 3 in Fig. 8.5 not only shields off the past light cone of region 1 from region 2 (as in the previous figure), but also vice versa. Thus a complete specification of the physical state of things in region 3 should render everything about region 2 (in particular, both Bob's randomly-selected setting $\hat{n}_2$ and his outcome B) irrelevant for predictions about region 1 – and it should also render everything about region 1 (in particular, both Alice's randomly-selected setting $\hat{n}_1$ and her outcome A) irrelevant for predictions about region 2.

This is the key idea behind the *factorization* of the joint probability $P(A, B|\hat{n}_1, \hat{n}_2, \lambda)$ for outcomes A and B (conditioned on all the things these outcomes might depend upon). The definition of conditional probability implies that this joint probability can be written as

$$P(A, B|\hat{n}_1, \hat{n}_2, \lambda) = P(A|\hat{n}_1, \hat{n}_2, B, \lambda) \cdot P(B|\hat{n}_1, \hat{n}_2, \lambda). \tag{8.45}$$

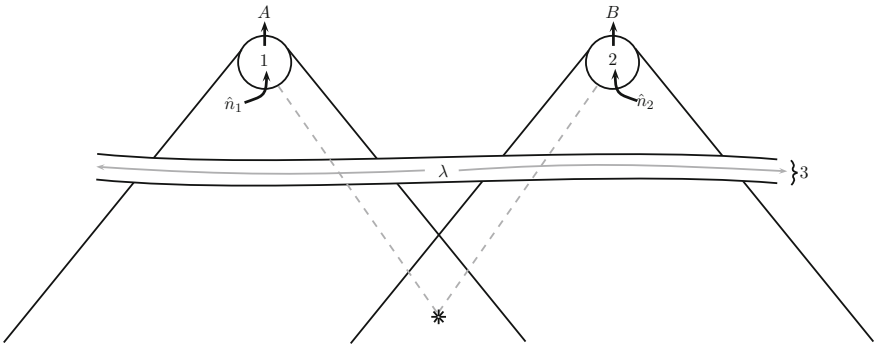


Fig. 8.5 Space-time diagram for the Bell experiment. The particle pair is emitted at the "flash" at the *bottom* of the diagram; world-lines for the two individual particles flying apart in opposite directions are represented by the *gray dashed lines*. The (large!) region 3 encompasses both particles at some intermediate time and shields the two measurement regions, 1 and 2, from their overlapping past light cones in the way that is required in Bell's formulation of locality. (Note that the complete description of the particle pair, λ, in region 3 need not assign specific facts to specific points in space; this allows λ to be something like a two-particle quantum-mechanical wave function which, as discussed in Chap. 5, does not imply a clear ontology in 3D space.) The apparatus settings $\hat{n}_1$ and $\hat{n}_2$ are shown as "inputs" to the measurements occuring in regions 1 and 2, whereas the individual outcomes A and B are shown as "outputs"

But then, in each of the two factors on the right hand side, Bell's definition of local causality implies that the entries relating to the space-like separated region are redundant. In particular, local causality requires that

$$P(A|\hat{n}_1, \hat{n}_2, B, \lambda) = P(A|\hat{n}_1, \lambda) \tag{8.46}$$

and

$$P(B|\hat{n}_1, \hat{n}_2, \lambda) = P(B|\hat{n}_2, \lambda). \tag{8.47}$$

In words: the probability for Alice's experiment to have outcome A should only depend on the state λ of the particle pair (for that run of the experiment) and the setting $\hat{n}_1$ of her apparatus; it should *not* depend on the setting of Bob's apparatus or on the outcome B of his experiment. And similarly, the probability for Bob's experiment to have outcome B should not depend on the setting of Alice's apparatus.

Plugging in we see that, for a locally causal theory, the joint probability factorizes as follows:

$$P(A, B|\hat{n}_1, \hat{n}_2, \lambda) = P(A|\hat{n}_1, \lambda)\, P(B|\hat{n}_2, \lambda). \tag{8.48}$$

From here, it turns out to be a straightforward mathematical exercise to derive the CHSH inequality. Recall first that the correlation coefficient $C(\hat{n}_1, \hat{n}_2)$ is defined as the expected value of the product of the outcomes A and B. This is simply the sum (over all four possible joint outcomes for A and B) of the product of the outcomes, weighted by the probability of that particular joint outcome. Notice also that, when we

speak of the "expected value", we mean: over a run of many particle pairs in which, for all we know, the exact state λ of the particle pair may vary from run to run. So we should also, for each possible joint outcome, average over the possible states λ that might have been produced by the source. (Let's assume, for definiteness but without loss of generality, that a continuously infinite spectrum of different possible λs are possible, with probability density $\rho(\lambda)$.) Our expression for the correlation coefficient then looks like:

$$\begin{aligned} C(\hat{n}_1, \hat{n}_2) &= \int \sum_{A,B} A \cdot B \cdot P(A, B|\hat{n}_1, \hat{n}_2, \lambda)\, \rho(\lambda)\, d\lambda \\ &= \int \sum_{A,B} A \cdot B \cdot P(A|\hat{n}_1, \lambda) \cdot P(B|\hat{n}_2, \lambda)\, \rho(\lambda)\, d\lambda \\ &= \int \left[\sum_{A} A \cdot P(A|\hat{n}_1\, \lambda) \right] \left[\sum_{B} B \cdot P(B|\hat{n}_2, \lambda) \right] \rho(\lambda)\, d\lambda \\ &= \int \bar{A}(\hat{n}_1, \lambda)\, \bar{B}(\hat{n}_2, \lambda)\, \rho(\lambda)\, d\lambda \end{aligned} \tag{8.49}$$

where, in the first step, we have used the factorized expression for the joint probability (which follows from local causality) and in the last step we have defined

$$\bar{A}(\hat{n}_1, \lambda) = \sum_{A} A \cdot P(A|\hat{n}_1, \lambda) \tag{8.50}$$

as the average value of A (and similarly for B). Since the only two possible outcomes for A are $+1$ and -1, it is obvious that this average value must be between -1 and $+1$, i.e.,

$$\left|\bar{A}(\hat{n}_1, \lambda)\right| \leq 1 \tag{8.51}$$

and similarly

$$\left|\bar{B}(\hat{n}_2, \lambda)\right| \leq 1. \tag{8.52}$$

Now let us consider the combinations of correlation coefficients that appear in the CHSH inequality. To begin with,

$$C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{c}) = \int \bar{A}(\hat{a}, \lambda) \left[\bar{B}(\hat{b}, \lambda) - \bar{B}(\hat{c}, \lambda) \right] \rho(\lambda)\, d\lambda \tag{8.53}$$

so that

$$\left| C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{c}) \right| \leq \int \left| \bar{B}(\hat{b}, \lambda) - \bar{B}(\hat{c}, \lambda) \right| \rho(\lambda)\, d\lambda \tag{8.54}$$

since $\left|\bar{A}(\hat{a}, \lambda)\right| \leq 1$.

In a similar way, we have that

$$\left|C(\hat{a}', \hat{b}) + C(\hat{a}', \hat{c})\right| \leq \int \left|\bar{B}(\hat{b}, \lambda) + \bar{B}(\hat{c}, \lambda)\right| \rho(\lambda)\, d\lambda. \tag{8.55}$$

Adding the last two equations, noting that $|x - y| + |x + y|$ is either $2x$, or $-2x$, or $2y$, or $-2y$, and is hence definitely less than or equal to 2 as long as $|x| \leq 1$ and $|y| \leq 1$, and using the fact that $\int \rho(\lambda) d\lambda = 1$, we arrive at the CHSH inequality:

$$\left|C(\hat{a}, \hat{b}) - C(\hat{a}, \hat{c})\right| + \left|C(\hat{a}', \hat{b}) + C(\hat{a}', \hat{c})\right| \leq 2. \tag{8.56}$$

And so, any theory that respects Bell's "local causality" condition must make predictions for the correlations in this kind of experiment which respect the inequality. But since the actual experimental data shows a clear violation of the inequality, it follows that all theories which respect Bell's "local causality" condition are *wrong*. The true theory, whatever that is exactly, must violate "local causality". But that is just a complicated way of saying that Nature itself violates local causality, i.e., the faster-than-light causal influences (which "local causality" prohibits) really exist in the world.

That, as we have already acknowledged, is profound and deeply troubling. And this way of arriving at the conclusion should make much clearer that one cannot escape it simply, for example, by upholding the orthodox completeness doctrine and rejecting hidden variables (or determinism or "realism"). This should help you appreciate why, when Bell referred to his own theorem, he (modestly) called it the "locality inequality theorem" [1].

Projects:

8.1 Create your own (preliminary, toy) Bell inequality like the one discussed in Sect. 8.2.

8.2 Use the table from Sect. 8.2 to show that, indeed, Eqs. (8.32) and (8.33) are equivalent.

8.3 Work through and understand all the detailed steps in the derivation, from Sect. 8.6, of the CHSH inequality (many of which are glossed over hastily in the text).

8.4 In Sect. 8.3, we discussed the need to assume that the numbers F_i (characterizing the fraction of particle pairs that are of each possible type) are independent of the axes along which the particle spins will be measured. (This assumption in the derivation of the inequality is then rendered applicable in an ideal experimental test of the inequality by letting the measurement axes be chosen randomly and only after the particle pairs have been emitted.) A similar assumption is made in the derivation of the CHSH inequality in Sect. 8.6, but the terminology is a little different and the assumption was not highlighted in the text. Explain, in the terminology of Sect. 8.6, what this assumption is, and point out the first equation in the derivation which would be invalid without this assumption.

8.5 Read Bell's 1964 paper, Ref. [2]. Summarize his method of deriving an inequality and comment on whether you think he is presenting a proof that the empirical predictions of quantum mechanics are inconsistent with locality, or inconsistent with the joint assumptions of locality and "realism" (i.e., deterministic hidden variables).

8.6 Read the (first) 1982 paper of Aspect et al., Ref. [4]. Summarize their experimental setup and procedure.

8.7 Read the (other) 1982 paper of Aspect et al., Ref. [5]. Summarize the relationship of this experiment to their earlier one, and comment on any other features you find interesting or surprising.

8.8 Read the 1998 paper of Weihs et al., Ref. [6]. Describe what was novel about their experiment (relative to Aspect's 1982 experiments) and summarize their results.

8.9 Read Bell's "Bertlmann's Socks..." paper, Ref. [21]. Summarize Bell's amusing derivation of a locality inequality (in terms of socks and washing machines).

8.10 Your friend is a sociologist doing her senior thesis on the political opinions of twins. She invites pairs of twins to show up and earn $20 by participating in her study. When a pair arrives, she sends the older twin into the room on the left with her assistant, Alice, and sends the younger twin into the room on the right with her other assistant, Bob. The rooms are extremely well soundproofed and the doors are tightly locked after each subject enters his/her room. After the doors are locked, the assistant (Alice or Bob) rolls a 3-sided die to randomly choose one of three pre-determined yes/no questions to ask the subjects. (What the questions are don't matter here, but you could imagine they are something like: Q1 is "Should we raise the minimum wage?", Q2 is "Should the fed raise interest rates?", and Q3 is "Should Roe-vs-Wade be overturned?") The assistants record the subjects' answers, and the whole process is repeated for several hundred pairs of subjects. Afterwards, your friend collects and analyzes all of the data and notices the following:

- When both twins happen to be asked the *same* question, they always answer the same way (either both "yes" or both "no").
- When the older twin is asked Q1 and the younger twin is asked Q2, the answers are (respectively) "yes" and "no" 20% of the time.
- When the older twin is asked Q2 and the younger twin is asked Q3, the answers are (respectively) "yes" and "no" 15% of the time.
- When the older twin is asked Q1 and the younger twin is asked Q3, the answers are (respectively) "yes" and "no" 40% of the time.

What should you advise your friend to conclude?

8.11 Interview some physicists to find out what they think Bell's Theorem is and proves. If they say that Bell's theorem proves you can't have a deterministic/hidden variables theory, you might consider following up by asking them how they reconcile this with the existence of the de Broglie - Bohm pilot-wave theory. You might also consider asking them if they think that ordinary QM (without any hidden variables) is a local theory and, if so, how they recon-

cile this with the perfect EPR correlations (i.e., ask them exactly how, in their understanding, ordinary QM explains these correlations in a local way).

8.12 Recall the assumption, made in the derivation of Bell inequalities, that, in the notation of Sect. 8.6, the probability distribution $\rho(\lambda)$ is independent of the settings $\hat{n}_1$ and $\hat{n}_2$. We have discussed how an ideal experimental test involves randomly choosing the settings "at the last second" in order to ensure that this condition is satisfied, the idea being that then there is no way for the particle source to have "known" about the settings – they didn't even exist yet when the source emitted the particles! But this is a little too quick. Sometimes people imagine that the settings could actually be determined, directly, by some kind of "free will choices" made at the last second by Alice and Bob. And depending on whether one believes in, and/or how one understands, "free will", that might indeed ensure that $\rho(\lambda)$ is independent of the settings. But nobody has ever performed an experiment like that; the real experiments use various sorts of random-number generators to determine the settings. But (again, depending on what sort of random number generator is used, exactly, and perhaps depending on whether one believes in "hidden variables") the outputs of random number generators are not actually random. In principle, the outputs are determined by something, which was in turn determined by something else, and so on into the past. Sketch a space-time diagram to make it clear how, at least in principle, both the settings $\hat{n}_1$ and $\hat{n}_2$ and the state λ of the particle pair could all be influenced/determined by something in their overlapping past light cones, and could therefore be correlated (such that $\rho(\lambda)$ is different for different values of $\hat{n}_1$ and $\hat{n}_2$) without any funny-business like non-locality or backwards-in-time causation. How seriously do you think this possibility should be taken? It may be helpful to think about the extent to which a similar independence assumption is needed in other scientific experiments that have nothing to do with quantum mechanics, e.g., a controlled drug trial in which patients are randomly assigned to receive either the drug or a placebo.[3]

8.13 Read one (or more!) of the three recent experimental papers reporting improved tests of Bell's inequalities, Refs. [7–9]. There are at least two interesting things to pay attention to in these papers. First, how do their experiments work and how do they represent improvements over the earlier experiments of Aspect (et al.) and Weihs (et al.)? And second, how do the authors talk about what their experiments show? That is, do they regard them as refuting hidden variables, or proving non-locality, or what exactly?

[3] Just to give you a sense of the spectrum of views which exist on this issue, the assumption – that $\rho(\lambda)$ is independent of the settings $\hat{n}_1$ and $\hat{n}_2$ – has been called the "no conspiracies" assumption, with the implication that you'd have to be a crazy conspiracy theorist to take it seriously; on the other hand, the Nobel Prize winning particle physicist Gerard 't Hooft, among others, thinks that relativity and quantum theory can and should be reconciled by denying that this assumption applies to the real experiments.

References

1. Bell, Preface to the first edition, in *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed. (Cambridge, 2004)
2. Bell, On the Einstein Podolsky Rosen paradox. Physics **1**, 195–200 (1964). (Reprinted in *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed. (Cambridge, 2004))
3. H. Clauser, Shimony, Holt, Proposed experiment to test local hidden-variable theories. Phys. Rev. Lett. **23**(15), 880–884 (1969)
4. Aspect, Grangier, Roger, Experimental realization of Einstein-Podolsky-Rosen-Bohm *Gedankenexperiment*: a new violation of Bell's inequalities. Phys. Rev. Lett. **49**(2), 91–94 (1982)
5. Aspect, Dalibard, Roger, Experimental test of Bell's inequalities using time-varying analyzers. Phys. Rev. Lett. **49**(25), 1804–1807 (1982)
6. Weihs, Jennewein, Simon, Weinfurter, Zeilinger, Violation of Bell's inequality under strict Einstein locality conditions. Phys. Rev. Lett. **81**, 5039 (1998)
7. Hensen et al., Loophole-free Bell inequality violation using electron spins separated by 1.3 km. Nature **526**, 682–686 (2015)
8. Giustina et al., Significant-loophole-free test of Bell's theorem with entangled photons. Phys. Rev. Lett. **115**, 250401 (2015)
9. Shalm et al., A strong loophole-free test of local realism. Phys. Rev. Lett. **115**, 250402 (2015)
10. Wigner, Interpretation of quantum mechanics, in *Quantum Theory and Measurement*, ed. by Wheeler and Zurek (Princeton University Press, Princeton, 1983)
11. Interview with R. Peierls, in Davies and Brown, in *The Ghost in the Atom* (Cambridge, Cambridge University Press, 1993)
12. Hawking, Does god play dice? (1999), http://www.hawking.org.uk/does-god-play-dice.html
13. Kleppner and Jackiw, One hundred years of quantum physics, in *Science*, August 11, 2000
14. Tegmark and Wheeler, 100 years of the quantum, in *Scientific American*, February 2001
15. Bertlmann and Zeilinger, *Preface to Quantum [Un]speakables: From Bell to QUantum Information* (Springer, Berlin, 2002)
16. Mermin, Hidden variables and the two theorems of John Bell. Rev. Modern Phys. **65**, 803–815 (1993)
17. Mermin, Is the moon there when nobody looks? Reality and the quantum theory. Phys. Today (1985)
18. Bell, Speakable and unspeakable in quantum mechanics, in *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed. (Cambridge, Cambridge University Press, 2004)
19. Bell, Quantum mechanics for cosmologists, in *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed. (Cambridge, Cambridge University Press, 2004)
20. Interview with J.S. Bell, in Davies and Brown, *The Ghost in the Atom* (Cambridge, Cambridge University Press, 1993)
21. Bell, Bertlmann's socks and the nature of reality, in *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed. (Cambridge, Cambridge University Press, 2004)
22. See, for example, H. Wiseman, The two Bell's theorems of John Bell. J. Phys. A Math. Theor. **47**, 424001 (2014), and T. Norsen, Are there really two different Bell's theorems? http://www.ijqf.org/archives/1646
23. Bell, La Nouvelle Cuisine, in *Speakable and Unspeakable in Quantum Mehanics*, 2nd ed. (Cambridge, Cambridge University Press, 2004)

Chapter 9
The Spontaneous Collapse Theory

Commenting on the quantum measurement problem as illustrated by Schrödinger's infamous cat, Bell remarked: "Either the wavefunction, as given by the Schrödinger equation, is not everything, or it is not right [1]." We have seen, in Chap. 7, how the measurement problem can be avoided if the wave function is not everything: by *supplementing* the wave function with additional objects (like the always-definite positions of particles in the pilot-wave theory) we can have a theory which actually predicts that definite things should happen, without anything like *ad hoc* and ill-defined exceptions to the usual dynamical rules.

In this chapter, we explore the other possibility mentioned by Bell – that the wavefunction, as given by the Schrödinger equation, isn't right. Recall that the essence of the measurement problem was that, according to Schrödinger's equation, interactions between (for example) a particle and a measuring apparatus do not typically result in the measuring apparatus pointer having a definite post-interaction position. Instead, the measuring apparatus gets infected with whatever quantum superposition was present in the initial state of the particle being measured. In ordinary quantum mechanics, this seemingly problematic result is already avoided in the way suggested by Bell – the "collapse postulate" is precisely a claim that the wave function *as given by Schrödinger's equation* isn't right. In particular, when a *measurement* occurs, the wave function ceases to evolve according to Schrödinger's equation, and (momentarily) does something entirely different instead. This avoids the seemingly problematic idea of superpositions of macroscopically distinct situations, but at a heavy price: it seems unbelievable that there is a fundamental distinction between "measurement" and "non-measurement" processes. Somehow, the true fundamental theory should treat all processes in a consistent, uniform fashion.

The "spontaneous collapse" theory is, at root, an attempt to remove this troubling dualism by positing, for the wave function, a single, universally-applicable dynamical evolution law which will somehow accomplish, in a single stroke, the two jobs done respectively by the Schrödinger equation and the collapse postulate in ordinary QM. The idea, more specifically, is to *modify* Schrödinger's equation with stochastic

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_9

non-linear terms which will have the effect of preserving the Schrödinger evolution for microscopic systems (where we know it is correct) but also ensuring that macroscopic things like pointers (and cats!) end up in the sorts of definite, non-superposed states we observe them to always end up in.

9.1 Ghirardi, Rimini, and Weber

The main idea of the spontaneous collapse theory is sometimes traced back to a 1966 paper by David Bohm and Jeffrey Bub, which explores a type of hidden variable theory rather unlike Bohm's 1952 pilot-wave theory. In the pilot-wave theory, of course, the additional variables (namely, the positions of the particles) are controlled by the wave function, which thus plays a somewhat mysterious background role. In the 1966 paper, by contrast, it is the hidden variables – here something like a background field – which influence the evolution of the wave function and give rise to deviations from the normal Schrödinger-equation evolution thereof.

This motivated Philip Pearle and also, somewhat later, Nicolas Gisin – both of whom were very concerned by the measurement problem of ordinary QM – to begin exploring stochastic modifications of the usual Schrödinger equation. Some progress was made toward the goal of reconciling wave function dynamics with the appearance of definite outcomes, but no systematic method of achieving the desired ends was identified, and several difficulties (including for example the apparent inevitability of conflicts with relativity when deviations from the Schrödinger evolution were contemplated) were brought into sharper focus.

A breakthrough appeared in 1986, when three Italian physicists (Ghirardi, Rimini, and Weber – hereafter "GRW") took fuller advantage of the fundamentality of *position*: if you can get the *positions* of macroscopic things right, then you will automatically get other properties right as well, since the outcomes of measurements of other properties (such as energy, momentum, spin, etc.) are always registered in the position of some macroscopic object (like our ubiquitous pointer) [2]. So whereas the earlier proposals had struggled with the problem of deciding which basis to use in narrowing the wave function (does one narrow in momentum space when a momentum measurement is happening?), GRW proposed the simple and elegant idea that wave functions should occasionally (randomly, spontaneously) localize exclusively in position space.[1]

In the theory, it is as if, at randomly selected moments, some outside observer makes a (somewhat rough) position measurement and thus collapses the particle's wave function (but, due to the roughness, to a finite-width Gaussian wave packet

[1] Note that there is an interesting parallel here to the pilot-wave theory, which eludes the "no hidden variables" theorems by letting non-position properties (such as momentum, energy, and spin) be "contextual". This difference, between the way position and other properties are treated by the theory, does not prevent the theory from generating correct empirical predictions for measurements of non-position properties since even the outcomes of momentum/energy/spin measurements are registered, at the end of the day, in the position of some pointer.

rather than a perfectly sharp delta function). But of course the whole point of the theory is to avoid the idea of some mysterious "outside observer" whose interventions imply exceptions to the usual dynamical behavior... hence "as if". According to GRW, the occasional collapses or "localizations" of the wave function should be considered as purely natural – part of the ordinary, universal way that wave functions evolve in time.

In just a moment, we'll talk through the technical details of these spontaneous localizations, starting first, in the present section, with the simple case of a single particle (in 1-D for simplicity). Then in the following sections we will explain how the theory describes multi-particle systems, including those that we would commonly describe as involving "measurements".

But first, let me just acknowledge that the theory, as it will be explained, maybe doesn't seem to do a very good job of truly *unifying* the two different types of wave-function time-evolution posited by ordinary QM. The GRW evolution will amount to: wave functions just evolve according to Schrödinger's equation most of the time, except for these occasional random moments when they instead suffer a spontaneous localization. The supposed unification here perhaps feels a bit like taking these two allegedly incompatible dynamical evolution laws, wrapping a bow around both of them together, and saying "Voila!" Putting this point another way, it may feel like there is somehow not much difference between the GRW theory and standard textbook QM: whereas orthodox QM says "Wave-functions evolve according to Schrödinger's equation, except during measurements, when they instead collapse" GRW says "Wave-functions evolve according to Schrödinger's equation, except at certain random moments, when they instead collapse." Other than a minor change in the words, is there really any difference?

It's a fair question, in response to which two things might be said.

One is that while the GRW process does indeed have a somewhat implausibly dualistic character, this can to at least some degree be eliminated. For example, models have been developed (especially by Pearle) in which, instead of abrupt intermittent wave function collapses, one has gentler localizations that are occuring continuously in time. The net effect (that is, something like the total amount of localization that happens per unit time) is roughly the same, but – because the Schrödinger-type and collapse-type evolutions are simultaneous and omnipresent – the dynamics feels a little more natural, coherent, and plausible. These so-called "continuous spontaneous localization" (CSL) models also solve some other technical issues with the GRW process as we will explain it. So, to some degree, one can take our discussion of the GRW process merely as a kind of pedagogical simplification of an overall concept which can perhaps be implemented somewhat more elegantly.

A second point, though, is that to some extent, no matter how you slice them up, "spontaneous collapse" theories just are somewhat dualistic. They are, after all, literally designed to unify the two dynamical postulates of ordinary QM. The duality, then, is somehow a problem only if one is expecting something that is somehow radically different from ordinary QM. But perhaps we should not expect that, and we should instead view the spontaneous collapse idea simply as an attempt to replace the "loose talk" (about "measurements" and "observers") in ordinary QM, with sharp

mathematics. From this point of view, we should not regard the GRW theory so much as an alternative to ordinary, textbook QM, but rather as something like "ordinary QM v2.0". In this vein, Bell said about the GRW theory: "I do think [the spontaneous collapse theories have] a certain kind of goodness... in the sense that they are honest attempts to replace the wooly words by real mathematical equations – equations which you don't have to talk away – equations which you simply calculate with and take the results seriously [3]."

All right. With all of that as preamble, let's finally jump into exploring in mathematical detail how the GRW theory works.

So, as has been said, the theory posits that the wave function of a single particle evolves according to Schrödinger's equation

$$i\hbar\frac{\partial\psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi(x,t)}{\partial x^2} + V(x,t)\psi(x,t) \tag{9.1}$$

most of the time. But the Schrödinger evolution is interrupted by occasional localizations. The Schrödinger-equation part of the evolution is already well-understood, so we will focus our exposition on the localizations.

First, when do they happen? For a single particle, there is supposed to be a constant probability per unit time, $\frac{dP}{dt} = \lambda$, for a spontaneous localization to occur. This will give rise to a (Poisson-distributed) sequence of times $t_1, t_2, t_3, \ldots$, with an average "waiting time" $\tau = t_{n+1} - t_n$ between the localizations given by $\tau = 1/\lambda$. For reasons that will be discussed as we proceed, GRW suggest that the constant λ should have a value in the neighborhood of

$$\lambda \approx 10^{-16}\, s^{-1} \tag{9.2}$$

so that the average time between localizations is

$$\tau = \frac{1}{\lambda} \approx 10^{16}\ \text{s} = 3 \times 10^8\ \text{years}. \tag{9.3}$$

That's three hundred million years – a very long time! So, for a single particle, the spontaneous localizations are quite rare. It may even appear that localizations occuring at such a slow rate would be totally negligible. But, as we will see in the next section, they will become quite important in the evolution of macroscopic systems containing a large number of particles. Before turning to that, though, let's understand in more precise detail what exactly happens at one of these intermittent localization events.

Consider the Gaussian function

$$g_r(x) = \frac{1}{(2\pi\sigma^2)^{1/4}} e^{-(x-r)^2/4\sigma^2} \tag{9.4}$$

which has a half-width of about σ, is centered at the point $x = r$, and is normalized in the following sense:

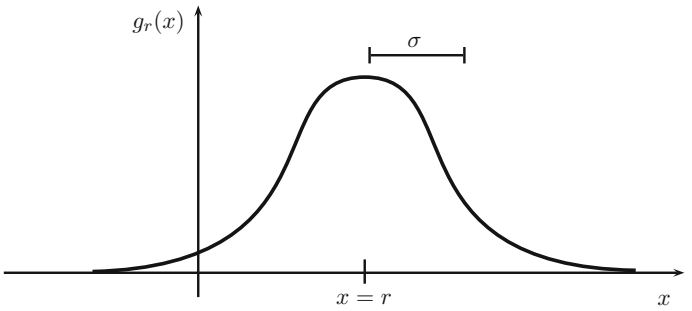


Fig. 9.1 The localization of a wave function in the GRW theory basically consists of its being multiplied by the Gaussian function $g_r(x)$ shown here

$$\int_{-\infty}^{\infty} |g_r(x)|^2 \, dx \; = \; 1. \tag{9.5}$$

(Since $g_r(x)$ is real-valued, the absolute value bars here are unnecessary but, of course, totally harmless.) The function $g_r(x)$ is shown in Fig. 9.1. Note that, again for reasons that will emerge as our presentation proceeds, the value of the constant σ is postulated by GRW to have a value in the neighborhood of

$$\sigma \approx 10^{-7} \, \text{m}. \tag{9.6}$$

It is important that this is fairly small on the macroscopic scale, but fairly large compared to, for example, the size of an atom.

The basic idea is then that, during an episode of "spontaneous localization", the wave function gets suddenly multiplied by $g_r(x)$. Suppose one of the localizations happens at time t. Then the wave function $\psi(x, t^+)$ just after time t is given by

$$\psi(x, t^+) \sim g_r(x)\psi(x, t^-) \tag{9.7}$$

where $\psi(x, t^-)$ is what the wave function was right *before* time t. That is the basic idea, but there are a couple of mathematical details to iron out.

First of all, I wrote "$\sim$" rather than "$=$" just above because the product on the right hand side will not generally be a properly normalized wave function. This is easy enough to fix by writing instead

$$\psi(x, t^+) = \frac{g_r(x)\psi(x, t^-)}{N(r)} \tag{9.8}$$

where the re-normalization factor $N(r)$ given by

$$N(r)^2 = \int \left| g_r(x)\psi(x, t^-) \right|^2 \, dx \tag{9.9}$$

ensures that $\psi(x, t^+)$ is properly normalized:

$$\int \left|\psi(x, t^+)\right|^2 dx = \frac{1}{N(r)^2} \int \left|g_r(x)\psi(x, t^-)\right|^2 dx = 1. \tag{9.10}$$

The second mathematical detail addresses the question: what is the value of r, i.e., what point does the wave function get localized *around*? The answer is that r is *random*, with a probability distribution

$$P(r) = N(r)^2 = \int \left|g_r(x)\psi(x, t^-)\right|^2 dx. \tag{9.11}$$

This says, basically, that the wave function is most likely to localize around some point $x = r$ where the wave function modulus is large to begin with. In a little more detail, it says that the probability for localization at the point $x = r$ is proportional to what would usually be regarded as the total probability associated with the new (but not yet normalized) localized state, if the localization did occur at $x = r$. (You are invited to prove that $P(r)$ as defined here really is a valid probability distribution in the Projects.)

Let's work through a couple of simple examples to clarify the idea.

To begin with, suppose the wave function is initially extremely spread out so that it has (say, over some region of width $L \gg \sigma$) a *constant* value:

$$\psi(x, t^-) = \frac{1}{\sqrt{L}}. \tag{9.12}$$

(The actual value here doesn't matter much for our purposes, but we might as well take the wave function to be properly normalized.) Now, at time t, let's say a spontaneous localization happens to occur. It is (approximately) equally likely to occur at any point r where the wave function $\psi(x, t^-)$ has support, since

$$P(r) = N(r)^2 = \int \left|g_r(x)\psi(x, t^-)\right|^2 dx \approx \begin{cases} 1/L & \text{where } \psi(x, t^-) = 1/\sqrt{L} \\ 0 & \text{where } \psi(x, t^-) = 0 \end{cases}. \tag{9.13}$$

The reason for the "≈" is that technically, at the edges of the width-L region where the wave function is initially non-zero, P(r) will be a little smaller than $1/L$, and similarly it'll be a little bigger than zero just outside that region where $\psi(x, t^-) = 0$. That is, $P(r)$ will have a smooth transition at the edges, as shown in the lower graph of Fig. 9.2. But still, leaving aside the edge effects, we can say that the localization is equally likely to occur at any point in the width-L region where the wave function was nonzero.

And of course, after the localization, the multiplication of the initially-constant wave function by the Gaussian $g_r(x)$ produces a Gaussian wave function, centered at the randomly-selected point r. The transition from $\psi(x, t^-)$ to $\psi(x, t^+)$ is sketched in the upper graph of Fig. 9.2.

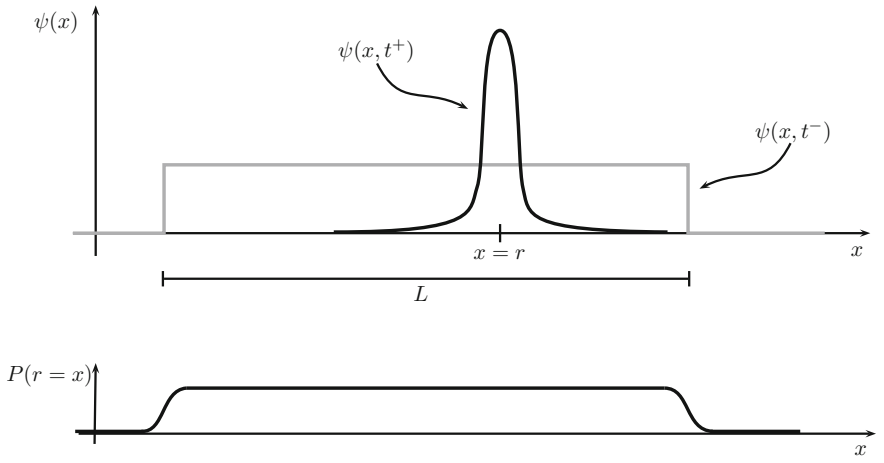


Fig. 9.2 If the wave function $\psi(x, t^-)$ is roughly constant over some region, a spontaneous localization will narrow it down to a Gaussian, of width σ, centered at some point r which is (approximately) equally likely to be anywhere in the region where $\psi(x, t^-)$ was nonzero

OK, so, if the wave function is initially very spread out compared to the length scale σ, a spontaneous localization does exactly what is advertised – it *localizes* the wave function around some new, randomly selected point where the wave function was originally big.

As a second example, let's take the opposite limit, where the wave function is already, initially, very narrowly peaked. It is convenient to take the extreme limiting case of a position eigenstate, i.e., a wave function which is a Dirac delta function:

$$\psi(x, t^-) \sim \delta(x - a). \tag{9.14}$$

Of course, this is not a properly normalized state, and (while not really making any difference at the end of the day), that will be slightly annoying as we try to figure out what our general formulas imply for things like the probability distribution $P(r)$. With apologies to any mathematicians who are reading, we can elude these problems in a simple way by writing

$$\psi(x, t^-) = \frac{1}{\sqrt{\delta(0)}}\,\delta(x - a). \tag{9.15}$$

Now suppose a spontaneous localization happens. This means the wave function will be multiplied by a Gaussian centered at some random point r. What is the probability distribution for this point? Well,

$$P(r) = N(r)^2 = \int \left| \frac{g_r(x)\delta(x - a)}{\sqrt{\delta(0)}} \right|^2 dx = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(r-a)^2/2\sigma^2} = g_a(r)^2. \tag{9.16}$$

That is, $P(r)$ is a Gaussian function, of width σ, centered at the same point $x = a$ where the wave function is initially concentrated. That makes sense.

So some point r (within about σ either way from $x = a$) is randomly chosen. What does the wave function look like after multiplication by $g_r(x)$ and re-normalization? Well,

$$\psi(x, t^+) = \frac{g_r(x)\psi(x, t^-)}{N(r)} = \frac{g_r(x)}{N(r)} \frac{\delta(x-a)}{\sqrt{\delta(0)}} = \frac{g_r(a)}{N(r)} \frac{\delta(x-a)}{\sqrt{\delta(0)}} = \psi(x, t^-) \tag{9.17}$$

since, as we just showed, $N(r) = g_r(a)$. In this case as shown in Fig. 9.3, the spontaneous localization actually doesn't change the wave function at all! (And note in particular that the wave function stays the same no matter which value of r was selected.) This actually makes sense: a δ-function wave function is already as localized as it is possible for a wave function to be, so the spontaneous localization doesn't change it at all.

Another interesting case to consider is an initially Gaussian wave function of width w_0. It turns out that, for $w_0 \gg \sigma$, the width after the localization decreases to about σ, whereas if $w_0 \ll \sigma$, the width is basically unaffected by the "localization". That is, this case smoothly interpolates between the two extremes embodied in our two examples. Rather than pursue that here, though, I'll let you work it out in the Projects.

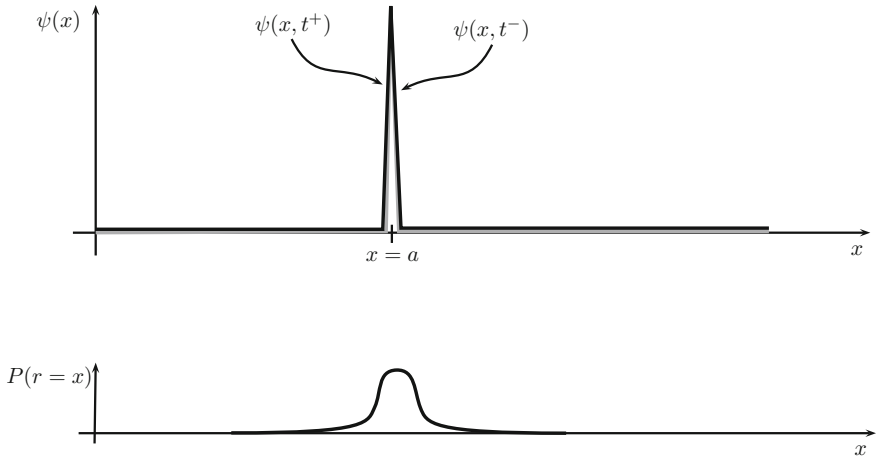


Fig. 9.3 If the wave function $\psi(x, t^-)$ is a δ-function (centered at $x = a$), then the probability distribution $P(r)$ is a width-σ Gaussian centered at $x = a$. But that turns out to be irrelevant because, no matter what value of r is chosen, the wave function is unaffected by the spontaneous localization: $\psi(x, t^-) = \psi(x, t^+)$

For a third and final example here, let's consider an "Einstein's boxes" kind of situation in which a particle is in a superposition of two relatively-sharply-defined positions. Concretely, suppose that

$$\psi(x, t^-) = \frac{1}{\sqrt{2}}\left[\frac{\delta(x+a)}{\sqrt{\delta(0)}} + \frac{\delta(x-a)}{\sqrt{\delta(0)}}\right] \tag{9.18}$$

so that the particle is in a 50/50 superposition of "being at $x = -a$" and "being at $x = +a$". And let us assume that the two possible positions here are very distant, i.e., $a \gg \sigma$.

The spontaneous localization process in this example is illustrated in Fig. 9.4. The probability distribution $P(r = x)$ for where the localization will be centered consists of two symmetric Gaussian functions centered at $x = +a$ and $x = -a$ respectively. Suppose that, by chance, $r \approx +a$. Then we have that

$$N(r)^2 = \frac{1}{2}\left[g_a(-a)^2 + g_a(+a)^2\right] = \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{-2a^2/\sigma^2} + 1\right]. \tag{9.19}$$

For $a \gg \sigma$, the first term is extremely small compared to 1 and we may thus take

$$N \approx \frac{1}{\sqrt{2}}\frac{1}{(2\pi\sigma^2)^{1/4}}. \tag{9.20}$$

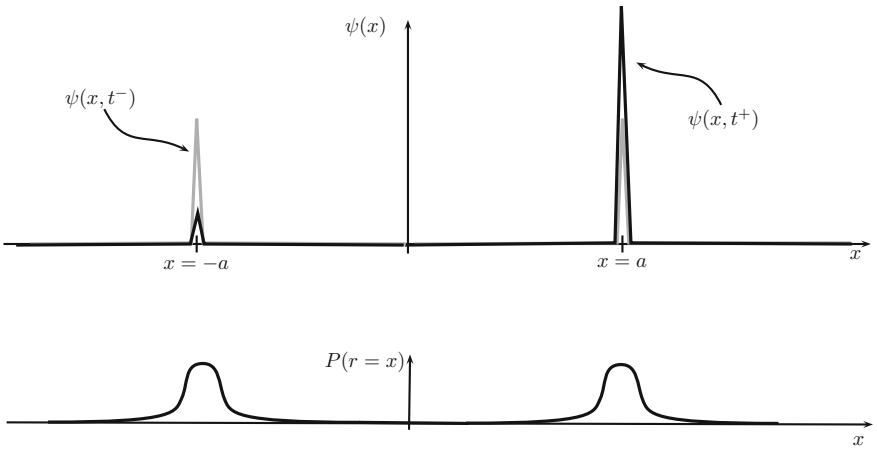


Fig. 9.4 If the wave function $\psi(x, t^-)$ is a superposition of two δ-functions, separated by a distance much larger than σ, the localization promotes one of the δ-functions while greatly suppressing the size of the other. For all practical purposes, the post-localization wave function is just one or the other of the previously-superposed spikes, so the localization has the effect of erasing spatial superpositions over a length scale greater than σ. Note that, as shown in the $P(r = x)$ graph below, the localization is equally likely to promote the $x = a$ or the $x = -a$ term. What is shown above is, obviously, the case in which $r \approx +a$ so that $\psi(x, t^+)$ is basically $\delta(x - a)$ – but with just a tiny bit of $\delta(x + a)$ remaining as well

The post-collapse wave function is then given by

$$\begin{aligned}\psi(x,t^+) &= \frac{g_{+a}(x)\psi(x,t^-)}{N} \\ &= \frac{1}{\sqrt{2}}\frac{1}{\sqrt{\delta(0)}}\frac{1}{N}\left[g_{+a}(x)\delta(x+a)+g_{+a}(x)\delta(x-a)\right] \\ &\approx \frac{\frac{1}{\sqrt{2}}g_a(a)}{N}\frac{\delta(x-a)}{\sqrt{\delta(0)}} \\ &= \frac{\delta(x-a)}{\sqrt{\delta(0)}} \end{aligned} \tag{9.21}$$

where we have again thrown out a term that is small by a factor like e^{-2a^2/σ^2} which is extremely small if $a \gg \sigma$.

So basically, if $r \approx +a$, the localization completely annihilates the delta function spike at $x = -a$ and leaves only a (re-normalized) spike at $x = +a$. (Of course, it was equally probable that instead we would have had $r = -a$ in which case the fates of the two spikes would have been reversed.) A particle which is in a superposition of two distinct locations (separated by a distance greater than σ) will not remain in that superposition forever; instead, according to GRW, the particle will eventually be located *definitely on the left* or *definitely on the right* – and this transition will happen spontaneously, without the need of anything like an external intervention or observation.

It might occur to you to worry that this spontaneous localization could destroy the interference that is observed in, for example, the two-slit experiment: if the two slits are separated by a distance greater than $\sigma \approx 10^{-7}$ m – and in typical demonstrations of interference, they are! – then the wave function of a particle which happens to suffer a spontaneous localization while it is traversing the 2-slit apparatus would *not* form an interference pattern at the screen, but would instead form something like a single-slit diffraction pattern. Does this mean that the GRW theory contradicts the observation of interference? No, for recall that, according to the theory, an individual particle only suffers a spontaneous localization every 300 million years or so. So unless you have a *lot* of time on your hands and send particles through the apparatus *very* slowly, you would never expect to see deviations from the usual quantum mechanical predictions in this kind of situation. Virtually all of the particles sent through would remain uncollapsed during the entire duration of their journey from source to screen.

9.2 Multiple Particle Systems and Measurement

As hinted at before, the incredible slowness/rarity of the GRW localizations might make one think that the localizations can just be completely ignored and will play no role whatever in the theory's predictions. But that is only true as long as we are thinking of individual particles. To understand the role of the localizations in the

GRW theory's solution to the measurement problem, we therefore need to see how the theory describes multi-particle systems.

The generalization of the theory to many-particle systems is pretty straightforward. In a nutshell, the idea is just that each individual particle suffers spontaneous localizations in the same way that we described in the previous section. How things play out then depends importantly on whether or not there is *entanglement*. Let's begin by discussing the simpler case in which there is no entanglement.

Consider, then, a two-particle system which, at the moment t^- just before a spontaneous localization occurs, is in the (non-entangled, i.e., factorizable) quantum state

$$\Psi(x_1, x_2, t^-) = \psi(x_1, t^-)\phi(x_2, t^-). \tag{9.22}$$

In a multi-particle situation like this, it is supposed to be irreducibly random *which* particle suffers the first localization (in addition to being irreducibly random exactly when and where that localization occurs). But, for definiteness, suppose that particle 2 suffers a localization at time t. Then, just as in the previous section, we have

$$\Psi(x_1, x_2, t^+) = \frac{g_r(x_2)\Psi(x_1, x_2, t^-)}{N(r)} \tag{9.23}$$

where, in the obvious generalization of what we saw previously,

$$N(r)^2 = \int \left|g_r(x_2)\Psi(x_1, x_2, t^-)\right|^2 \, dx_1 \, dx_2. \tag{9.24}$$

And note that, also just as before, the probability density for the localization to be centered at the point r is $P(r) = N(r)^2$.

Plugging in the non-entangled two-particle state, Eq. (9.22), we see that

$$\Psi(x_1, x_2, t^+) = \psi(x_1, t^-)\frac{g_r(x_2)\phi(x_2, t^-)}{N(r)}. \tag{9.25}$$

The important point here is that when the overall quantum state involves no entanglement, an individual spontaneous localization only affects the particular particle that is "hit" by it. The overall state remains an unentangled product state, and the factors representing the wave functions of the *other* particles (in our example here, particle 1) are in no way affected by the localization.

But let us now turn to the more interesting case where there *is* entanglement between the two particles. Take, for simplicity, the same sort of example we considered in the previous section, in which a particle is located either at $x = +a$ or at $x = -a$, but suppose now there are *two* particles in a superposition of "both particles

are at $x = +a$" and "both particles are at $x = -a$". Suppose in particular that, at a time t^- just before one of the particles happens to suffer a spontaneous localization, the two-particle wave function is

$$\Psi(x_1, x_2, t^-) \sim \frac{1}{\sqrt{2}} [\delta(x_1 - a)\delta(x_2 - a) + \delta(x_1 + a)\delta(x_2 + a)] . \tag{9.26}$$

Now, what happens to this wave function if one of the particles suffers a spontaneous localization? We cannot, as before, just say that "the wave function of one of the particles gets localized, while that of the other is unaffected"... For an entangled state like this the particles cannot even be said to possess their own individual wave functions! So let's just let the math tell us what happens, supposing, again arbitrarily, that it is particle 2 which nominally suffers the localization:

$$\Psi(x_1, x_2, t^+) = \frac{g_r(x_2)\Psi(x_1, x_2, t^-)}{N(r)} \tag{9.27}$$

where r is random, with probability distribution $N(r)^2$. Here, just as before, the definition of $N(r)$ is

$$N(r)^2 = \int \left|g_r(x_2)\Psi(x_1, x_2, t^-)\right|^2 \, dx_1 \, dx_2. \tag{9.28}$$

This will be large in a small (size-σ) neighborhood around $r = +a$ as well as a small neighborhood around $r = -a$. That is, the spatial probability distribution for the center of the localization will look exactly like it did in the last example of the previous section.

Suppose that, for this particular localization, it happens that $r \approx +a$. Then (leaving out the uninteresting re-normalization factor) the two-particle wave function after the localization will look like

$$\begin{aligned}
\Psi(x_1, x_2, t^+) &\sim g_a(x_2)\Psi(x_1, x_2, t^-) \\
&= g_a(x_2)\frac{1}{\sqrt{2}} [\delta(x_1 - a)\delta(x_2 - a) + \delta(x_2 + a)\delta(x_2 + a)] \\
&= \frac{1}{\sqrt{2}} [\delta(x_1 - a)\delta(x_2 - a)g_a(x_2) + \delta(x_1 + a)\delta(x_2 + a)g_a(x_2)] \\
&= \frac{1}{\sqrt{2}} [\delta(x_1 - a)\delta(x_2 - a)g_a(a) + \delta(x_1 + a)\delta(x_2 + a)g_a(-a)]
\end{aligned} \tag{9.29}$$

Now here is the crucial point. The factor $g_a(a)$ (in the first term in the square brackets) is "big" – it is just the value of g at exactly the place where g peaks. But the factor $g_a(-a)$ (in the second term in the square brackets) is vanishingly small, if the separation $(2a)$ between the two places the particles might have been is large

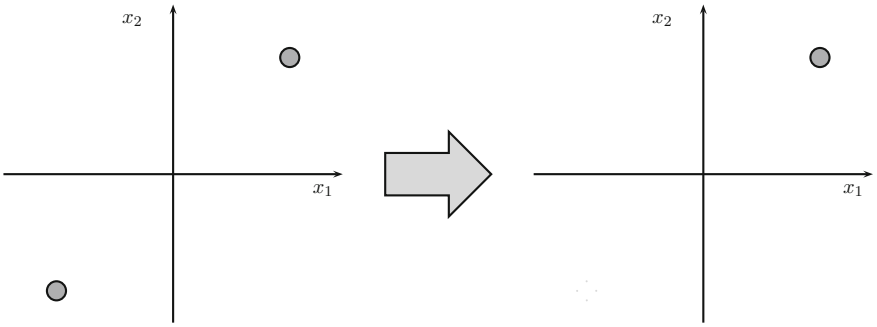


Fig. 9.5 The *left* graph is a configuration space map of the two-particle wave function $\Psi(x_1, x_2)$ for two particles in a superposition of "both particles on the *left*" and "both particles on the *right*". The *right* graph is the same map, after a single spontaneous localization. Despite nominally acting on just one of the two particles, the spontaneous localization gives rise to a wave function which has both particles localized together

compared to the width σ of the localization function g. So, to an excellent approximation, we have that, after the spontaneously localization (which, remember, was nominally associated with just one of the two entangled particles)

$$\Psi(x_1, x_2, t^+) = \delta(x_1 - a)\delta(x_2 - a) \tag{9.30}$$

which is of course a state in which *both* particles are definitely located at $x = a$. This process is illustrated in Fig. 9.5.

That was of course just one of several possibilities. We assumed arbitrarily that particle 2 happens to suffer the first spontaneous localization, and that this localization happens to be centered around $r = +a$. If you think through it, though, it should be clear that the final state would have been exactly the same had it been instead particle 1 that suffered a collapse centered near $r = +a$. So it doesn't actually matter which particle gets "hit" – *either* one getting localized localizes *both* because their positions started out in the special entangled state. And it should also be clear that, if either particle instead suffered a localization centered near $x = -a$, then both particles would have ended up definitely localized at $x = -a$.

Now we are finally in a position to understand how this spontaneous collapse theory solves the measurement problem. Suppose that, instead of just two particles being in an entangled state that binds their positions together (even as they remain in a superposition of, say, "all being on the left" or "all being on the right") it is instead some macroscopically-large number, like $N \approx 10^{23}$ particles whose positions are so bound. That is, suppose the initial state is something like

$$\begin{aligned}&\Psi(x_1, x_2, \ldots, x_N, t^-)\\ &\sim \frac{1}{\sqrt{2}}\left[\delta(x_1 - a)\delta(x_2 - a)\cdots\delta(x_N - a) + \delta(x_1 + a)\delta(x_2 + a)\cdots\delta(x_N + a)\right].\end{aligned} \tag{9.31}$$

Then we can see that, as soon as *any one* of the N particles suffers a spontaneous localization, the *entire set* of particles will localize along the following lines:

$$\Psi(x_1, x_2, \ldots, x_N, t^+) \sim \delta(x_1 - r)\delta(x_2 - r) \cdots \delta(x_N - r) \tag{9.32}$$

with $r = +a$ or $r = -a$ with 50/50 probability. Everything is just the same as before, with one important exception. With just one particle, we would typically need to wait around $\tau = 300$ million years for the particle to spontaneously localize. For two particles, we would typically need to wait around $\tau/2 \approx 150$ million years. But for $N \approx 10^{23}$ particles, we would typically need to wait $\tau/N \approx 30$ nanoseconds. That is, because of the enormous number of individual particles comprising anything remotely macroscopic, a macroscopic object (like, say, a pointer or a cat) will suffer a constant barrage of spontaneous localizations (millions or billions or trillions of them per second), which will for all practical purposes prevent it from ever getting into the kind of macroscopic superposition state which so worried Einstein and Schrödinger. As Bell expressed this point: "Quite generally any embarrassing macroscopic ambiguity in the usual theory is only momentary in the GRW theory. The cat is not both dead and alive for more than a split second [1]."

Let us illustrate this one last time with our standard example of a quantum measurement process: a single "particle in a box" which begins in a superposition of several different energy eigenstates, but which is then coupled to an energy measuring device, represented schematically as a pointer whose position moves by an amount proportional to the energy of the particle. As should be familiar from earlier treatments, if the coupling begins at $t = 0$, the Schrödinger equation dictates that the wave function at time t will be given by

$$\Psi(x, y, t) = \sum_i c_i \psi_i(x, t)\phi_0(y - \lambda E_i t) \tag{9.33}$$

where the ψ_i are the energy eigenfunctions (with corresponding energy eigenvalues E_i) for the particle-in-a-box and ϕ_0 is a narrow Gaussian wave packet representing the position of the center of mass of the (roughly 10^{23}) particles composing the pointer.

The particle-in-a-box is just a single particle, so the probability that it will happen to suffer a spontaneous localization during the time of the experiment is negligible. The pointer, on the other hand, being macroscopic, will suffer repeated localizations. If, however, the width of the wave packet ϕ_0 describing its center-of-mass location is small compared to σ, these localizations will have essentially no effect on the overall wave function for early times during which the individual wave packets – corresponding to terms with different values of i in Eq. (9.33) – remain overlapping in configuration space.

However, as soon as the different terms begin to fail to overlap, such that the spacing between them is of order σ, the situation will be just like that discussed earlier in this section: because all 10^{23} particles composing the pointer are bound together (by the usual sorts of intra- and inter-atomic forces) a single spontaneous

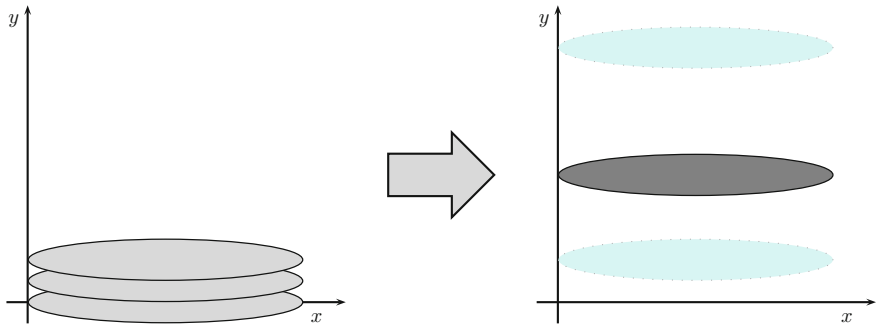


Fig. 9.6 Evolution of the wave-function (in the schematic, two-dimensional configuration space whose axes are the position x of the particle-in-a-box and the position y of the center-of-mass of the pointer) for our toy measurement example. As soon as the individual terms in the superposition (which are separating along the y-direction of configuration space) have separation of order σ, the superposition collapses to just one term, randomly selected from all the possibilities, with probability $|c_i|^2$. Thus, before there would be time for anybody to notice or worry about a troubling macroscopic superposition, the overall wave function describes the pointer as having a well-defined center-of-mass position (here, arbitrarily, $y \approx \lambda E_2 t$) and the particle-in-a-box as having the correct, associated energy E_2

localization of any of the particles will localize all of them, i.e., will localize the entire macroscopic pointer, to just one of the terms. The others will, for all practical purposes, disappear. This is illustrated in Fig. 9.6.

Notice, in particular, that although the spontaneous localizations exclusively localize the particles in *position* space, the particle-in-a-box (whose energy is being measured in this example) ends up in a state of definite energy, and, indeed, the particular state corresponding to the final position of the pointer on the energy-measuring device. Thus, no special/additional/contradictory prescription is required to cause sub-system wave functions to collapse (in the way that ordinary QM says they do) when some arbitrary measurement is performed on them. The measurement *outcome* being displayed in the macroscopic spatial configuration of some aspect of the measuring device (here, the pointer, but one could just as well think of the distribution of ink droplets on a computer printout, or the distribution of photons emitted from a computer screen, for example) is perfectly sufficient in general. This should help clarify the earlier comments about the importance of recognizing the fundamentality of position.

9.3 Ontology, Locality, and Relativity

As we have explained it so far, the GRW theory describes the world in terms of a wave function. Because the collapse/localization mechanism is built into the dynamical evolution law for the wave function, the theory has no need to follow orthodox QM

in postulating a separately-existing macroscopic world and associated exceptions to the usual dynamical laws. That is, by providing a uniform description of the world that avoids (noticeable) macroscopic superpositions, the GRW theory avoids the measurement problem that plagues ordinary QM. But what about the other two problems associated with standard QM that we reviewed in earlier chapters?

We begin with the ontology problem. The wave function for an N-particle system (where, for GRW, ultimately N is the total number of particles in the entire universe) is something like a field on $3N$-dimensional configuration space. This does not, in any obvious or straightforward way, attribute definite properties to particular locations in regular, 3-dimensional physical space. Since the ultimate goal must be to provide a coherent description and explanation of the observable 3-dimensional physical world, it is clear that more needs to be said about what, according to the theory, the physical world is made of, and how it relates to the universal wave function whose evolution we have already discussed.

Two possibilities have gained traction in the literature. The first is, interestingly, just the early idea of Schrödinger that we discussed in Chap. 4. Recall that Schrödinger's idea was that the wave function (on configuration space) could be used to define (for example) a *mass density* associated with each individual particle, according to

$$\rho_i(x,t) = m_i \int |\Psi(x_1, x_2, \ldots, x_N, t)|^2 \, \delta(x_i - x) \, dx_1 \, dx_2 \cdots dx_N. \tag{9.34}$$

The total mass density could then be written

$$\rho(x,t) = \sum_i \rho_i(x,t) \tag{9.35}$$

and the original hope was that this field $\rho(x,t)$ would contain, at least at an appropriately coarse-grained level, an image of the familiar macroscopic world of everyday perception, including things like pointers with definite positions and unambiguously alive or dead cats.

Schrödinger himself gave up this interpretation of the wave function (as representing a continuous matter density in physical space) because it simply did not work the way he had hoped for. If the wave function obeys Schrödinger's equation all of the time, then the mass field $\rho(\vec{x},t)$ inherits (or one might say, makes ontologically clear) whatever problematic superpositions arose in the wave function itself. For example, in the Schrödinger's cat kind of situation, the mass field would not contain just a living cat or just a dead cat, but both – superimposed on top of one another, so to speak. If one imagines extrapolating to a description of the entire world, with frequent splittings of the universal wave function into different "branches" each of which corresponds to some more or less definite macroscopic situation, *all* of these different "possibilities" would be superimposed in Ψ and hence $\rho(\vec{x},t)$ would be, for lack of a better term, a complete and utter mess. The ρ generated by the theory,

that is, simply wouldn't look anything like what we know the world is supposed to look like. So the theory seems clearly wrong.

But by altering the rules of wave-function evolution, the GRW dynamics avoids precisely this sort of trouble. That is, if the wave function of the universe evolves, not according to Schrödinger's equation, but instead according to the GRW process, the mass density field $\rho(\vec{x}, t)$ associated with the wave function will correspond to just one of the sensible macroscopic possibilities. (Or at least, any appreciable non-sensible macroscopic blurriness will not last for more than a split second.) The three dimensional world, consisting of a mass field $\rho(\vec{x}, t)$ produced by a universal wave function obeying the GRW dynamics, that is, *will look right*. There will be tables and chairs and trees and planets with adequately-sharp shapes, structures, and trajectories; cats will be unambiguously alive or dead; and so on.

So that is one possible way of understanding the ontology of the physical world according to this theory. In the literature, this has come to be called "GRWm", meaning: the universal wave function evolves according to the GRW dynamics, and the ontology is understood as a *m*ass density field.

The other possible way was suggested by Bell:

> There is nothing in this theory but the wavefunction. It is in the wavefunction that we must find an image of the physical world, and in particular of the arrangement of things in ordinary three-dimensional space. But the wavefunction as a whole lives in a much bigger space, of $3N$-dimensions. It makes no sense to ask for the amplitude or phase or whatever of the wavefunction at a point in ordinary space. It has neither amplitude nor phase nor anything else until a multitude of points in ordinary three-dimensional space are specified. However, the GRW jumps (which are part of the wavefunction, not something else) are well localized in ordinary space. Indeed each is centered on a particular spacetime point $[\vec{x}, t]$. So we can propose these events as the basis of the 'local beables' [Bell's term for the physical space ontology] of the theory. These are the mathematical counterparts in the theory to real events at definite places and times in the real world.... A piece of matter then is a galaxy of such events [1].

The idea, then, is that each spontaneous localization, which happens at a particular time and is centered at a particular location in 3D space, produces a kind of "matter point" at that location in space-time. These "matter points" have, in the subsequent literature, come to be called "flashes", and so this version of GRW has come to be called "GRWf".

It is helpful to visualize the two options here, so in Fig. 9.7. I have sketched, on spacetime diagrams, the story of what is going on with the pointer in our toy measurement example, according to GRWm and GRWf.

With two definite proposals for understanding the ontology of the GRW theory, we are in a position to ask: does the theory (in either version) respect the idea of relativistic locality, i.e., no (spooky, faster-than-light) action at a distance? The answer, simply and unambiguously, is: no. GRW (with either of the proposed ontologies) is a non-local theory. This, of course, is not surprising given that we know, from Bell's theorem (Chap. 8), that *any* theory which agrees with the quantum mechanical predictions will have to be non-local. (But see Chap. 10!) Still, it is worthwhile to understand in more detail how the non-locality appears.

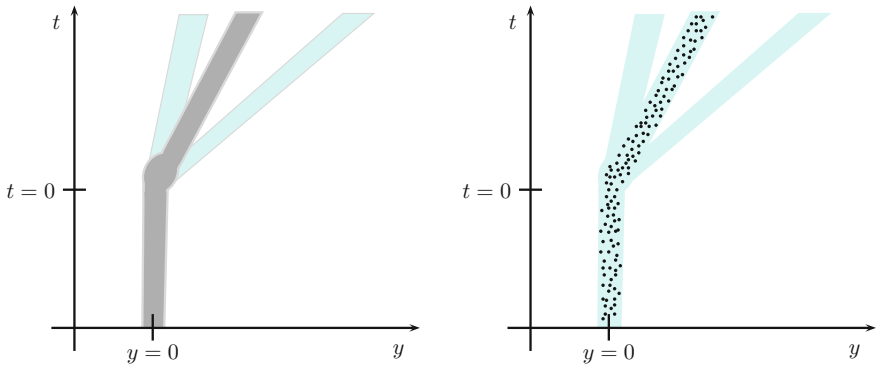


Fig. 9.7 The panel on the *left* shows the mass density field $\rho(y, t)$ associated with all the particles composing the pointer, for the familiar toy measurement example in which the energy of a particle-in-a-box is measured and the outcome, E_2, is indicated by the position of a pointer. Here "the pointer" consists of a lump of nonzero mass density that begins near $y = 0$ and then starts moving to the *right* at a certain well-defined rate just after the measurement interaction begins at $t = 0$. The slight "bulge" around $t = 0$ is meant to suggest that, as the individual terms in the wave function begin to separate, there is a brief period of time in which $\rho(y, t)$ includes several superimposed possibilities. But after a tiny fraction of a second, a spontaneous localization picks just one of the possibilities, the rest disappear, and $\rho(y, t)$ contains just the one realized possibility. The *right* panel shows the same situation, but for the flash ontology. The *black dots* represent the discrete, space-time point flashes and something like the overall motion of the pointer to the *right*, at a basically well-defined rate, can indeed be understood as "a galaxy of such events". It is interesting to contemplate, however, the fact that (in the same way that a real galaxy is mostly empty space), most of the time the pointer is, according to GRWf, literally nothing. That is, for the overwhelming majority of horizontal slices you could draw through the diagram (corresponding to particular moments), the slice would intersect precisely zero of the *dots/flashes*. The physical world, according to GRW, is curiously sparse and pointillistic at the micro-scale... though it coarse-grains to produce a sensible image of the familiar world at the macro-scale

It is easiest to see and understand in the case of GRWm, so let us begin there. Consider a kind of double Einstein's boxes situation, in which two particles are each split between two pairs of half-boxes. In particular, suppose that Alice, a million miles to the left, has a particle which is split between the half box in her left hand (state ψ_L^A) and the half box in her right hand (state ψ_R^A). And suppose that Bob, a million miles to the right, has a second particle which is similarly split between the half box in his left hand (state ψ_L^B) and the half box in his right hand (state ψ_R^B). And suppose that, by some prior careful arrangement, the two particles are in the following entangled state:

$$\psi(x_1, x_2) = \frac{1}{\sqrt{2}}\left[\psi_L^A(x_1)\psi_L^B(x_2) + \psi_R^A(x_1)\psi_R^B(x_2)\right]. \tag{9.36}$$

Suppose also that Bob is prepared with a position measuring device (the center-of-mass position of whose macroscopic pointer we denote y) which can interact with the two half-boxes he's holding and determine whether particle 2 is in the half box

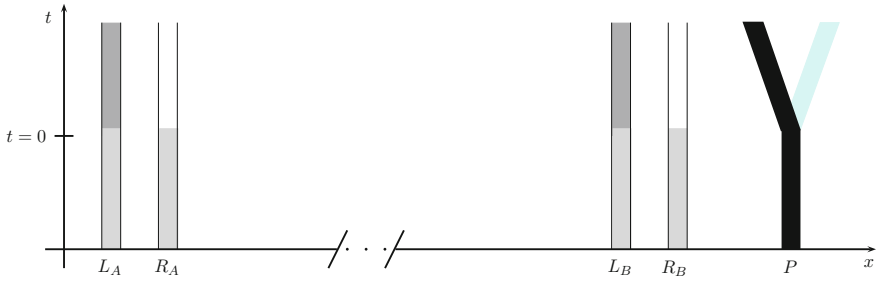


Fig. 9.8 Space-time diagram showing the mass densities of the various objects described in the text: the contents of the half-boxes in Alice's left and right hands (L_A and R_A), the contents of the half boxes in Bob's *left* and *right* hands (L_B and R_B) and the pointer P on Bob's position measuring device. At $t = 0$ Bob initiates the measurement of the position of his particle; very shortly after $t = 0$, a spontaneous localization in one of the (many!) pointer particles collapses the wave function in such a way that subsequently, say: (*i*) the entire mass density associated with the pointer moves unambiguously to the left, (*ii*) the mass density associated with Bob's particle coalesces entirely into L_B (i.e., the density there doubles while the density in R_B suddenly goes to zero), and (*iii*) the mass density associated with Alice's particle (millions of miles away!) also coalesces entirely into L_A. The change in the mass density distribution associated with Alice's particle, as a consequence of Bob's measurement on his particle, is a clear-cut case of non-local action-at-a-distance. Note, though, that as in the analogous case in the pilot-wave theory, even though what's happening in Alice's boxes is affected by Bob's distant actions, Alice has no way to observe this change. She could open her boxes and see where the particle is, and she would of course find it somewhere. But she would have no way to know whether it was her own observation that triggered her particle to randomly coalesce either in her right hand or her left hand, or whether, instead, the particle had already coalesced in one place or the other as a result of Bob's distant actions. So although there is nonlocal action-at-a-distance, according to the theory, the nonlocality cannot be used to transmit messages faster than light, and so avoids the most blatant sort of conflict with relativity theory

in his left hand, or instead the one in his right hand. The measuring device is initially in its ready state, with the pointer at $y = 0$, and we assume the pointer moves to the right/left if particle 2 is found in the right/left-hand box.

Now suppose that at $t = 0$ Bob decides to proceed with the measurement, i.e., to let the measuring device begin interacting with his half-boxes. A space-time diagram showing the mass densities associated with the two particles and the pointer is shown in Fig. 9.8. The important point is as follows. Prior to $t = 0$, the mass density associated with Alice's particle is genuinely split 50/50 between her two half-boxes. And it would (with extremely high probability) have remained so split (for millions of years!) had Bob not initiated the position measurement on his own particle. But when he does initiate this position measurement, it has the effect, by the mechanism we discussed in the previous section, of causing (in some very short period of time) a collapse to one or the other of the definite, initially superposed states. Thus not only Bob's particle, but also Alice's distant one, will switch from being "evenly smeared" between the two half boxes, to being definitely in one or the other of the two half boxes, as a direct result of Bob's decision to initiate his measurement procedure. Bob's decision – a million miles to the right – thus (almost) instantaneously influences

the distribution of mass (associated with Alice's particle) even though Alice's particle is a million miles to the left. It is a clear-cut case of nonlocal action-at-a-distance. (You are invited, in the Projects, to render this diagnosis in a more formal way by applying Bell's locality condition or one of our modifications of it.)

The case of GRWf is basically the same, although it is slightly harder to draw a nice picture to capture the nonlocality since, as mentioned, for individual particles, the space-time diagram would be almost completely empty. Still, the same kind of analysis applies. Suppose, for example, that the two particles are prepared, just as before, in the state

$$\psi(x_1, x_2) = \frac{1}{\sqrt{2}}\left[\psi_L^A(x_1)\psi_L^B(x_2) + \psi_R^A(x_1)\psi_R^B(x_2)\right]. \tag{9.37}$$

Now, there is a certain (very small, but nonzero) probability that, according to the theory, there will be a "flash" inside the box in Alice's left hand in, say, the next one minute. However, if Bob measures the position of his particle in the same way we described before, this probability (for a flash to appear in L_A) will either double (if Bob's measurement "finds" his particle on the left) or will go to zero (if Bob instead "finds" his particle on the right). Thus, the probability for a certain event over where Alice is, a million miles to the left, will be different depending on what happens over where Bob is, a million miles to the right, even when we are conditionalizing those probabilities on a complete specification of events (including, here, in particular, the fact that there have been no prior flashes associated with Alice's particle!) in the past light cone of the event in question.

So, with either the "m" or "f" ontology, the GRW theory is nonlocal, just like the pilot-wave theory, and just like we should have expected on the grounds of Bell's theorem.

However, as first pointed out by Bell, there is a sense in which the spontaneous collapse theories seem to be more compatible with relativity – or at least a little more promising in that respect – than the pilot-wave theory [1]. This has to do with the fact that the spontaneous collapse theories are irreducibly stochastic (unlike the pilot-wave theory, which is deterministic). The technical details are somewhat beyond the level of this book, but it should be noted that spontaneous collapse theories with both "mass field" and "flash" ontologies have been constructed which, despite being non-local, appear to be more plausibly consistent with a more serious notion of fundamental Lorentz invariance than appears to be possible with pilot-wave type theories [4, 5]. That is, the spontaneous collapse theories warrant a somewhat hopeful attitude toward the project – which you may not until this very moment even have conceived of as a possibility – of *reconciling* non-locality (which we know, from Bell's theorem, must be present) with some satisfying notion of fundamental relativity (which, needless to say, there is strong reason to demand).[2]

[2]The idea that it might be possible, after all, to reconcile relativity with non-locality may perhaps suggest that earlier chapters have over-stated the extent to which Bell's formulation of locality successfully captures the idea of "no faster-than-light causal influences" that we ordinarily take to be an implication of relativity theory. Let me assure you that this is not the case. The quantum

The technical details involved in these issues definitely render them beyond the scope of the present book. But one can nevertheless appreciate that certain seemingly simple questions – for example, "What precisely does it *mean* for a theory to be fundamentally relativistic?" – turn out to be surprisingly difficult to answer in a context in which they are entangled with the possibility of irreducibly stochastic (non-deterministic) laws, unclarity about ontology (how quantum wave functions relate to goings-on in 3+1-dimensional space-time), and other issues we have grappled with in this book. Suffice it to say here that this remains an area of continuing controversy and ongoing research, but that the spontaneous collapse theories have put on the table, for further analysis and contemplation, the previously-unrecognized possibility that Bell's theorem (and the associated experiments) could live in harmony with fundamental relativity.

9.4 Empirical Tests of GRW

So far we have presented the spontaneous collapse theory as a way of reformulating quantum mechanics so that it (i) posits a clear set of unambiguous and universal dynamical rules and (ii) provides a coherent ontology in terms of which directly observable macroscopic features of the real world can be recognized. The goal has been basically to clean up the foundational problems that plague ordinary quantum theory, while maintaining (as closely as possible) quantum theory's seemingly accurate empirical predictions. But, as hinted at the beginning of this chapter, because they predict that spontaneous collapses will occur with specific length- and time-scales, the spontaneous collapse theories do, in principle, make slightly different empirical predictions from ordinary quantum mechanics. And of course this is nice, because it means that spontaneous collapse theories can be tested, experimentally, against other versions of quantum mechanics.

The easiest type of test to understand involves something like two-slit interference. Recall that, for a single particle, spontaneous collapses, which localize the particle's wave function to a distance scale of order σ, happen with frequency λ, i.e., on a timescale $\tau = 1/\lambda$. It should thus be clear that, in an interference experiment with individual particles, spontaneous collapse theories will predict that the interference should start to disappear if the spatial separation between the individual components of the wave function exceeds σ for a time period greater than τ. Thus, the successful observation of interference places experimental constraints on the values of σ and τ, or equivalently, σ and λ.

(Footnote 2 continued)
non-locality really does mean that there are causal linkages between space-like separated events, of a sort normally thought to be prohibited by relativity. The possibility of perhaps reconciling non-locality with relativity does not mean that we previously misunderstood or misformulated locality. Rather, it means that there may have been something rather deep and subtle wrong with the way we were thinking about relativity (or causality or both) that fooled us into thinking that relativity was incompatible with space-like separated events being causally linked.

Of course, for interference experiments involving single particles, the particles are typically in a state of spatial superposition for only some small fraction of a second. So the experimental constraint is something like $\tau \gg 1$ s, i.e., $\lambda \ll 1\ \mathrm{s}^{-1}$. The value of λ proposed by GRW, recall, was $\lambda \approx 10^{-16}\ \mathrm{s}^{-1}$. So the experimental constraint coming from, say, single-neutron interferometry, is almost completely useless: it tells us only that, if the spontaneous collapse theories are right, the frequency of collapses must be much smaller than something that is already 16 orders of magnitude bigger than what we guessed the frequency might be!

However, as we saw previously, the effective collapse rate for an object consisting of N particles is $N\lambda$. So, by performing interference experiments with atoms, molecules, and even larger objects, we can start to get experimental constraints that are at least in the neighborhood of the hypothesized values of the collapse parameters. For example, in 1999, a group led by Markus Arndt and Anton Zeilinger in Vienna demonstrated interference using "buckyballs", which are C_{60} molecules [6]. Sixty carbon atoms, each containing 12 nucleons (6 protons and 6 neutrons) and 6 electrons, is roughly a thousand elementary particles. So the spontaneous collapse rate for buckyballs should be about a thousand times faster than the fundamental (per particle) collapse rate, and so the experimental constraint on the GRW parameters is about three orders of magnitude closer to relevancy.

In subsequent years, interference with even bigger molecules has been demonstrated, and there are plans for pushing this particular envelope even further [7]. In addition, experimental limits on the spontaneous collapse parameters can also be extracted from other kinds of observations. For example, spontaneous localizations add high-momentum Fourier components to wave functions and thereby add energy to systems that would not otherwise be present. Such additions of energy might be observed as anomalous warming of otherwise-thermally-isolated systems, or perhaps anomalous emission of high-energy particles such as X-rays. Some of these processes can be explored in more detail in the Projects.

In a very nice recent paper, Tumulka and Feldmann have organized the various known experimental constraints on the spontaneous collapse parameters, and produced a "parameter diagram" showing ranges of values that are excluded by the different types of observations [8]. We reproduce one of their diagrams here as Fig. 9.9. As might have been anticipated from our previous discussion, the most stringent constraints actually do not come from interference experiments, but arise instead from observations of systems (like the intergalactic medium!) with considerably more particles. Note also that these kinds of observational constraints tend to exclude the "upper left" portion of the parameter space, i.e., large values of λ and small values of σ.

A rather different kind of constraint tends to exclude the opposite region of the parameter space, i.e., very small values of λ and very large values of σ. The idea here goes back to one of the original motivating goals of the spontaneous collapse theories, which is to avoid the embarrassing sort of macroscopic superposition that is illustrated by Schrödinger's cat. More specifically, the idea is that we *know*, just from ordinary direct perceptual experience of the physical world around us, that macroscopic things do not appear "blurry". So their positions must be sharply defined at length scales

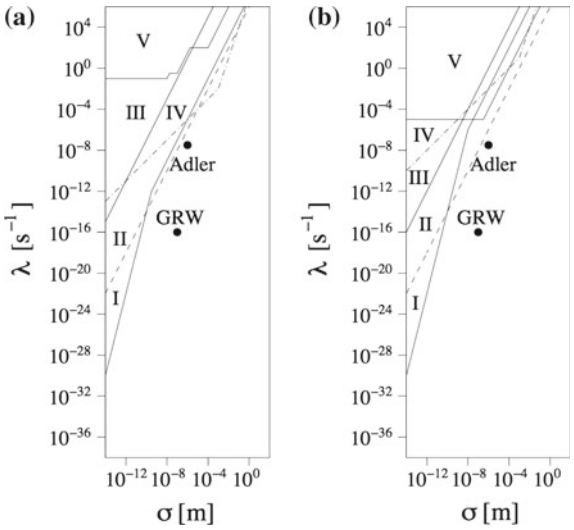


Fig. 9.9 A map of the parameter space, for both (**a**) the GRW theory we've discussed in detail as well as (**b**) the related "continuous spontaneous localization" [CSL] theory that was alluded to earlier, showing the values of σ (the spatial width of the collapse function $g_r(x)$) and λ (the collapse frequency) that are excluded by various sorts of observations, from Ref. [8]. The five numbered categories of experimental/observational constraints are "I = spontaneous x-ray emission, II = spontaneous warming of the intergalactic medium (dashed line), III = spontaneous warming of air, IV = decay of supercurrents (dashed-and-dotted line), V = diffraction experiments [8]." Note that, on each panel, the two dots represent the parameter values suggested originally by GRW and another slightly different suggestion by Stephen Adler. https://doi.org/10.1088/1751-8113/45/6/065304

where any blurriness would be perceptually evident – say, something of order a millimeter. Or at least, visible macroscopic things should not remain blurry at a distance scale much larger than a millimeter, for a time long enough for us to notice the blurriness! One can see in principle here how small values of λ and large values of σ can be excluded as "perceptually unsatisfactory". See Fig. 9.10 for Tumulka and Feldmann's nice diagram showing both the "Empirically Refuted Region" and (what they call the "Philosophically Unsatisfactory Region" but I would prefer to call) the "Perceptually Unsatisfactory Region" of parameter space for both GRW and CSL.

We close this section and this Chapter with one final Figure from the paper by Tumulka and Feldmann. In Fig. 9.11 we reproduce their diagram showing the progression of experimental constraints, coming from interference experiments, over time. The visual implication is that we are perhaps only two or three decades away from the ability to experimentally probe the parameter values originally suggested by GRW. Thus, the "open window" – between the ERR and PUR in Fig. 9.10 – may close in the near future, and we will know, once and for all, whether or not the spontaneous collapse models are right.

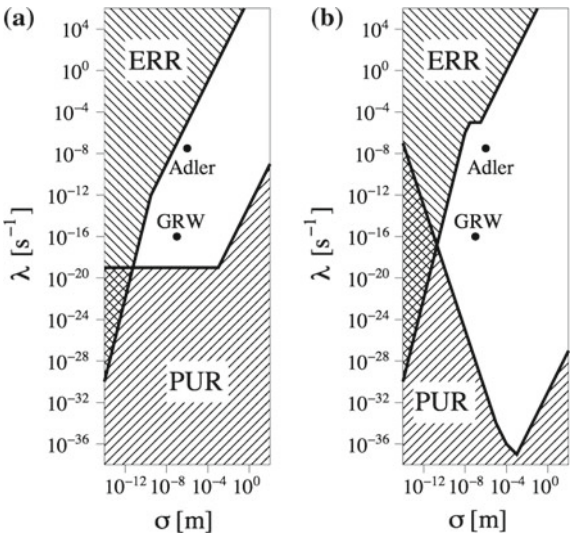


Fig. 9.10 Map of parameter space, again for both GRW and CSL theories, showing now both the "Empirically Refuted Region" (ERR) and the "Perceptually Unsatisfactory Region" (PUR) as discussed in the text. From Ref. [8]. Figure © IOP Publishing. Reproduced with permission. All rights reserved. https://doi.org/10.1088/1751-8113/45/6/065304

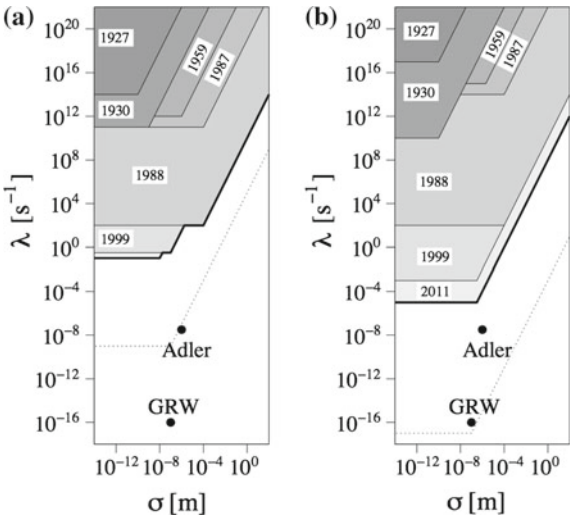


Fig. 9.11 The "Empirically Refuted Region" (ERR) of the GRW and CSL parameter spaces has steadily advanced, in recent decades, leaving an ever-narrowing window of parameter values which are compatible both with experimental and perceptual evidence. This suggests that, within perhaps a couple of decades, we will either have direct experimental evidence in support of the spontaneous collapse models, or the models will have been ruled out as either empirically or perceptually unacceptable. From Ref. [8]. Figure © IOP Publishing. Reproduced with permission. All rights reserved. https://doi.org/10.1088/1751-8113/45/6/065304

Projects

9.1 Work carefully through all the steps to convince yourself that Eqs. (9.16) and (9.17) are correct.

9.2 Suppose a particle with a Gaussian wave function $\psi(x) \sim e^{-x^2/4w_0^2}$ suffers a spontaneous collapse centered at $x = r$. Show that the post-collapse wave function remains Gaussian, and find a formula for its width w in terms of w_0 and σ. (Confirm that your expression for w implies that $w \approx w_0$ if $\sigma \gg w_0$, and implies that $w \approx \sigma$ if $w_0 \gg \sigma$.)

9.3 Suppose a particle has a Gaussian wave function $\psi(x) \sim e^{-x^2/4w_0^2}$ at the moment just before it suffers a spontaneous collapse. What is the probability density $P(r)$ for the collapse to be centered at $x = r$?

9.4 Argue that the probability distribution $P(r)$ defined in Eq.(9.11) is indeed a legitimate probability distribution since $P(r) > 0$ and $\int P(r)\, dr = 1$.

9.5 The discussion in Sect. 9.2 suggests that whereas for a single particle the localization rate is λ, for a collection of N particles the localization rate is $N\lambda$. This is basically equivalent to saying that collections of particles should have an overall localization rate that is proportional to the total mass of the collection – an idea that can and probably should be instituted as part of the formulation of the theory at the fundamental level: different particle species (electrons and protons, for example) may have different fundamental localization rates, with the rates being proportional to the mass of the particle. Assuming such a modification of the theory, is it nucleons (neutrons and protons) or electrons that suffer most of the localizations associated with ordinary matter?

9.6 It may appear puzzling that the spatial probability distribution for the point r at which a localization is centered, is given by $P(r) = N(r)^2$ rather than the seemingly simpler and approximately equivalent alternative $P(r) = |\psi(r, t^-)|^2$. The reason for this has to do with the requirement that non-local signaling (i.e., instantaneous communication across arbitrary distances) should be impossible. Consider a situation involving two entangled and spatially-separated particles, 1 and 2, and suppose that particle 1 suffers a spontaneous collapse centered at $x_1 = r$ at time t. Show that the pre-collapse marginal distribution for particle 2 to be observed at position x_2, namely

$$P(x_2, t^-) = \int \left|\psi(x_1, x_2, t^-)\right|^2 dx_1 \tag{9.38}$$

is the same as the post-collapse marginal distribution (averaged over all the points r at which the collapse might have been centered)

$$P(x_2, t^+) = \int \int \left|\psi(x_1, x_2, t^+)\right|^2 \, P(r)\, dx_1\, dr \tag{9.39}$$

provided $P(r) = N(r)^2$. (This means, for example, that Alice cannot tell, by measurements made on her particle, whether a distant entangled particle has suffered a collapse. This in turn prevents Bob from sending her a message, by for example choosing whether or not to allow his particle – entangled with her distant one – to interact with a macroscopic object such as a measuring device and thereby trigger a collapse.)

9.7 Consider a one gram pointer. In GRW with the "flash" ontology, approximately how many flashes occur, associated with the pointer, per second, if the flash rate f is as given in the text? (Assume for simplicity that only the nucleons are hit by spontaneous localizations.)

9.8 Approximately what fraction of the particles composing your body will pop briefly into existence (in a "flash") at least once during your lifetime, according to GRWf?

9.9 Consider the conduction electrons in a macroscopic piece of metal. These can be thought of as having wave functions that spread out over the entire, macroscopic extent of the metal. For such an electron with essentially zero momentum, its kinetic energy will also be approximately zero. However, if it happens to suffer a spontaneous localization its wave function will subsequently be a width-σ Gaussian. Estimate the increase in the particle's kinetic energy that results from this spontaneous localization, and use this to estimate the rate at which the temperature of a thermally isolated piece of metal should increase according to the spontaneous collapse theory. Would this "anomalous heating" be easy to detect, experimentally?

9.10 Use Bell's formulation of locality (and/or one of our modified versions from Chap. 1 or Chap. 5) to more formally diagnose GRWm as a non-local theory, using the example displayed in and discussed around Fig. 9.8. (Note: you will need to think carefully about which formulation of locality it is possible and appropriate to use here.)

9.11 In Sect. 9.3, we discussed the non-local character of both GRWm and GRWf in terms of a "double Einstein's boxes situation, in which two particles are each split between two pairs of half-boxes." This example is nice because it provides another opportunity to think about how the spontaneous collapses function in the presence of entanglement. But it is really more complicated than is minimally necessary to establish the non-locality of the theory. Show and explain how a "[single] Einstein's boxes" situation, like that discussed in Sect. 4.1 of Chap. 4, can already be used to diagnose the spontaneous collapse theories as non-local. (Note that this means, interestingly, that there are situations whose explanation involves non-locality in GRW, but is local in the pilot-wave theory.)

9.12 A single character printed in ink contains something of order 10^{17} carbon atoms or roughly 10^{18} nucleons. In GRWf, how many flashes per second (associated with that small amount of ink) do you think are sufficient to say that the ink drop is really there, with the particular shape we see? (Hint: human visual perception can be modeled as something like a digital camera which captures roughly 30 frames per second. Consistency with perceptual experience would seem to require that typical frames contain enough flashes to construct the shape of the appropriate letter unambiguously.) Use your estimate to calculate the minimum localization rate λ compatible with "perceptual acceptability", and compare your calculated value to Fig. 9.10.

9.13 In Ref. [8], Tumulka and Feldmann raise an interesting question: what if some future experiment demonstrates violations of ordinary QM and confirms the empirical predictions of GRW/CSL, but for parameter values which lie in the "perceptually unsatisfactory region" (PUR) of Fig. 9.10. What would you say/conclude in such a situation?

9.14 True of false: according to the spontaneous collapse theories, matter is made of particles. Explain.

9.15 There are a lot of things to like about the spontaneous collapse theories: they sharpen, with precise mathematics, the "loose talk" of Copenhagen QM; they provide comprehensible (if unanticipated) ontologies; and they make empirically testable predictions that differ from other versions of QM. But it is also possible to find the spontaneous collapse theories somewhat contrived and *ad hoc*. Explain why, by listing and discussing some of the details about the theory's formulation which seem arbitrary and/or which could easily be changed without dramatically affecting the theory's structure or predictions.

References

1. J.S. Bell, Are there quantum jumps? in *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed. (Cambridge University Press, Cambridge, 2004)
2. G.C. Ghirardi, A. Rimini, T. Weber, Unified dynamics for microscopic and macroscopic systems. Phys. Rev. D **34**, 470 (1986)
3. J.S. Bell, quoted by G.C. Ghirardi in *Sneaking a Look at God's Cards*, Gerald Malsbary, trans., Revised Edition (Princeton University Press, Princeton, 2005), p. 415
4. R. Tumulka, A relativistic version of the Ghirardi-Rimini-Weber model. J. Stat. Phys. **125**, 821–840 (2006)
5. D. Dürr, D.J. Bedingham, G. Ghirardi, S. Goldstein, R. Tumulka, N. Zanghi, Matter density and relativistic models of wave function collapse. J. Stat. Phys. **154**, 623–631 (2014)
6. M. Arndt, O. Nairz, J. Vos-Andreae, C. Keller, G. van der Zouw, A. Zeilinger, Waveparticle duality of C60 molecules. Nature **401**, 680 (1999)
7. M. Arndt, K. Hornberger, Testing the limits of quantum mechanical superpositions. Nat. Phys. **10**, 271–277 (2014)
8. W. Feldmann, R. Tumulka, Parameter diagrams of the GRW and CSL theories of wave function collapse. J. Phys. A: Math. Theor. **45**, 065304 (2012)

Chapter 10
The Many-Worlds Theory

The last version of quantum mechanics that we will explore in detail was developed by Hugh Everett III while he was a graduate student under John Wheeler in the 1950s. Everett's basic idea is at once beautifully elegant and uncomfortably radical. Max Jammer rightly described it as "one of the most daring and most ambitious theories ever constructed in the history of science" [1].

Some idea about the nature of Everett's proposal can be gleaned by the different titles used for various draft versions of his PhD thesis: "Quantum Mechanics by the Method of the Universal Wave Function", "Wave Mechanics Without Probability", and "On the Foundations of Quantum Mechanics [2]." Everett's thesis advisor, John Wheeler, was a strong proponent of Bohr's Copenhagen interpretation and was thus sensitive not only about the radical nature of Everett's proposal, but also about Everett's sharp criticisms of the Copenhagen philosophy. Wheeler thus demanded that Everett produce a significantly toned-down presentation of his ideas; this was ultimately published in 1957 with the somewhat cryptic title "[The] 'Relative State' Formulation of Quantum Mechanics [3]." The somewhat muted nature of the presentation in this published version probably contributed to Everett's ideas not being widely understood or appreciated for several subsequent decades, and his near-complete departure from the world of theoretical physics. But Everett did inspire a few early followers such as Bryce deWitt who, along with his own graduate student Neill Graham, published the original, full-length version of Wheeler's thesis, as well as some other commentary, as "The Many-Worlds Interpretation of Quantum Mechanics [4]". This title is probably the most accurately descriptive of Everett's ideas, and is the one by which the theory has largely come to be described today.

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_10

10.1 The Basic Idea

As with the Spontaneous Collapse theory of the last chapter, Everett's theory is principally motivated by the Measurement Problem that we studied in Chap. 3. In Everett's description, the usual quantum formalism contains two incompatible rules, "Process 1" and "Process 2", for how the states of quantum systems evolve. Process 1 is the discontinuous and random change that is postulated to occur when an observer or measuring instrument from outside the quantum system interacts with it in an appropriate way, whereas Process 2 is the continuous and deterministic state evolution described by the Schrödinger equation. Everett bemoans the fact that, since measuring instruments (and ultimately observers) are just physical systems like any other, it is simply not clear when the two very different Processes are supposed to apply. Discussing an isolated system that includes an observer or measuring instrument, Everett writes:

> Can the change with time of the state of the *total* system be described by Process 2? If so, then it would appear that no discontinuous probabilistic process like Process 1 can take place. If not, we are forced to admit that systems which contain observers are not subject to the same kind of quantum-mechanical description as we admit for all other physical systems. [And note that when we speak of an "observer", we really mean things like] photoelectric cells, photographic plates, and similar devices where a mechanistic attitude can hardly be contested [3].

Moreover, if one wants to apply quantum mechanics to the universe as a whole (which is natural in cosmology and in particular in the quest to unify quantum theory and gravitation) the idea of an "outside observer" becomes obviously incoherent:

> No way is evident to apply the conventional formulation of quantum mechanics to a system that is not subject to *external* observation. The whole interpretive scheme of that formalism rests upon the notion of external observation. The probabilities of the various possible outcomes of the observation are prescribed exclusively by Process 1. Without that part of the formalism there is no means whatever to ascribe a physical interpretation to the conventional machinery. But Process 1 is out of the question for systems not subject to external observation [3].

Everett's central idea, therefore, is to simply abandon Process 1:

> This paper proposes to regard pure wave mechanics (Process 2 only) as a complete theory. It postulates that a wave function that obeys a linear wave equation everywhere and at all times supplies a complete mathematical model for every isolated physical system without exception. It further postulates that every system that is subject to external observation can be regarded as part of a larger isolated system. The wave function is taken as the basic physical entity.... [3]

Let us think through what that means, in the context of our standard example: the measurement of the energy of a particle-in-a-box (whose spatial coordinate we call x) using an energy-measuring-device (whose pointer has center of mass coordinate y). We would describe the measuring device as faithfully and accurately measuring the energy of the particle if, when the particle is initially in an energy eigenstate $\psi_k(x)$ (with eigenvalue E_k), the interaction causes the apparatus pointer to move by

a distance proportional to E_k. That is, we assume that Process 2 – Schrödinger's equation – evolves the joint quantum state of the particle and pointer as follows:

$$\psi_k(x)\phi_0(y) \rightarrow \psi_k(x)\phi_0(y - \lambda E_k t) \tag{10.1}$$

where t is the duration of the interaction. It then immediately follows from the linearity of Schrödinger's equation that, if the particle-in-a-box is initially in a superposition of several different energy eigenstates, the system will evolve into an entangled superposition as follows:

$$\left[\sum_i c_i\, \psi_i(x)\right] \phi_0(y) \rightarrow \sum_i c_i\, \psi_i(x)\, \phi_0(y - \lambda E_i t). \tag{10.2}$$

All three versions of quantum theory that we have studied so far have regarded this last formula as problematic, and have thus proposed some way of resolving the problem. The orthodox/Copenhagen view, for example, would say that it was inappropriate to try to describe the measurement interaction in terms of a quantum mechanical wave function; measuring devices are classical objects, so we should have treated the interaction instead using Process 1, according to which the post-interaction state involves a collapsed wave function for the particle-in-a-box and a pointer with a definite, classical position. The pilot-wave theory accepts that there is a wave function associated with the particle-pointer system, and that the wave function indeed evolves into a state like that in Eq. (10.2), but insists that the real physical state of the pointer is not to be found in this wave function but instead in the actual positions of the associated particles, which will be unambiguous and unproblematic. Finally, the spontaneous collapse theory insists that the wave function for a system including a macroscopic pointer will simply not obey Schrödinger's equation, and so the troubling macroscopic superposition state, Eq. (10.2), simply will not arise (or at least, will not arise for long enough to notice!).

In contrast to all three of these views, Everett wants to say, about Eq. (10.2), that it is fine; there is no problem. To understand this, though, it will help to briefly review what the problem with Eq. (10.2) was supposed to be. In short, the problem was that Eq. (10.2) involves a superposition of different positions for the (macroscopic, directly observable) pointer. It's frankly not even exactly clear what this means, or what it would look like, but apparently it is some kind of state in which the pointer somehow has several different positions at the same time. It seems it should appear in some sense "blurred" among the several different positions. But, of course, nobody has ever seen a pointer in a state like that. Real pointers always point this way or that. And so Eq. (10.2) simply cannot provide a faithful, direct, literal, complete description of the physical state of the pointer. Or at least that is what we have been taking for granted until now.

Everett, though, invites us to consider in more detail what – according to quantum mechanics – the pointer in a state like Eq. (10.2) would look like. To analyze this, we should consider the different possible states that an observer might get into upon

interacting with a pointer. Suppose, to begin with, that the observer – whose (many!) degrees of freedom we call z – begins in a "ready" state, $\chi_0(z)$. Suppose he interacts with a pointer with a reasonably well-defined position $y = y_k$. Then the observer should get into a state $\chi_k(z)$ in which he has seen (and, for example, remembers seeing, and will, if asked, report having seen) that the pointer has position y_k. That is, during the time of interaction between the pointer and the observer, Schrödinger's equation should evolve the quantum state as follows:

$$\phi_0(y - y_k)\,\chi_0(z) \rightarrow \phi_0(y - y_k)\,\chi_k(z). \tag{10.3}$$

But then, just as before, it immediately follows from the linearity of Schrödinger's equation that if the observer observes the position of the pointer in the particle-pointer system, described by Eq. (10.2), the quantum state of the particle-pointer-observer system will evolve as follows:

$$\left[\sum_i c_i\,\psi_i(x)\,\phi_0(y - y_i)\right]\chi_0(z) \rightarrow \sum_i c_i\,\psi_i(x)\,\phi_0(y - y_i)\,\chi_i(z). \tag{10.4}$$

So, in the same way that the pointer failed to pick out some one particular outcome from the set of superposed "possibilities", but instead got tangled up in the superposition, so now the observer (of the pointer) does not end up in a state that corresponds to seeing some one particular location for the pointer. Instead he, too, joins the entangled superposition. It is, of course, unclear exactly what to make of this. But notice right away one thing that this definitely does *not* say: it does not say (or at any rate, does not seem to say) that the observer will be in a state in which he definitely experiences (and remembers experiencing and will report, if asked, having experienced) the pointer as "looking blurry" or "being smeared out among several different positions". Everett explained this as follows in his thesis:

> Why doesn't our observer see a smeared out needle? The answer is quite simple. He behaves just like the apparatus did. When he looks at the needle (interacts) he himself becomes smeared out, but at the same time correlated to the apparatus, and hence to the system.... [T]he observer himself has split into a number of observers, each of which sees a definite result of the measurement.... As an analogy one can imagine an intelligent amoeba with a good memory. As time progresses the amoeba is constantly splitting [2].

Whatever else one wants to say, there is a suggestion here that our assumption that there was some kind of fatal problem with Eq. (10.2) – that the particle-pointer system just obviously wouldn't look right if this were the correct and complete state description – was perhaps too hasty.

It will be helpful here to follow Everett in introducing the concept of a "relative state". As he points out, in a system described, for example, by Eq. (10.2), neither of the components – the particle or the pointer – can be attributed a definite state of their own. That's essentially what "entanglement" means. But Everett points out that we can define a "relative state" for each component, relative, in particular, to the

other component's being in a particular state. For example, the state of the pointer relative to the particle-in-a-box being in state $\psi_k(x)$ is defined to be

$$\phi^{\text{rel. to } \psi_k(x)}(y) \sim \int \psi_k^*(x)\Psi(x, y)dx \tag{10.5}$$

where $\Psi(x, y)$ is just the joint particle-pointer state given in Eq. (10.2). (The "$\sim$" is because the right hand side here is not properly normalized.) Plugging in, we find

$$\begin{aligned}\phi^{\text{rel. to } \psi_k(x)}(y) &\sim \int \psi_k^*(x) \sum_i c_i\psi_i(x)\phi_0(y - \lambda E_i t) \\ &= c_k\phi_0(y - \lambda E_k t)\end{aligned} \tag{10.6}$$

since the different $\psi_i(x)$'s are orthonormal: $\int \psi_k^*(x)\psi_i(x)dx = \delta_{i,k}$. So the (properly normalized) relative state is just

$$\phi^{\text{rel. to } \psi_k(x)}(y) = \phi_0(y - \lambda E_k t). \tag{10.7}$$

In words: relative to the PIB being in a particular energy eigenstate, the pointer ends up in a perfectly definite and appropriate state, namely, one in which it indicates the energy E_k of the PIB.

The converse also holds: relative to the pointer indicating outcome E_k, the PIB is in the state $\psi_k(x)$. And we can generalize this concept to bring in the observer as well: when the overall particle-pointer-observer wave function is given by the right hand side of Eq. (10.4), relative to the PIB being in the state $\psi_k(x)$, not only does the pointer indicate that its energy is E_k, but the observer *sees* (and remembers seeing and will report having seen) the pointer indicating that its energy is E_k.

This idea of "relative state" provides a way of capturing the fact that, although it remains puzzling, a state like the right hand side of Eq. (10.4) is not just utter chaos. It is not just some kind of incomprehensible blur in which everything is happening in a completely mixed-up way. Rather, there are definite *correlations* built into the state: it is an orderly mixture, in some sense, of several individually perfectly reasonable situations, in which the PIB has a definite energy, and the pointer indicates correctly what its energy is, and the observer sees the pointer indicating its energy and correctly and validly infers what its energy is.

It is clear that subsequent interactions will work exactly the same way, and just bring more and more of the world into the mixture. For example, the air molecules in the vicinity of the pointer get jostled around in slightly different ways depending on how fast the pointer moves during its journey from its "ready" position to its final position, and where exactly that final position is. Of course, there are several distinct final positions in the mix, and so, just like the observer, the air molecules join the entangled superposition. They cannot be said to have any particular definite state of their own, but relative to the pointer being in a particular final position, their configuration is clear and definite and sensible. Similarly, if the observer's mom

calls him on the phone to ask how his energy measurement turned out, she will now also join the entangled superposition, as will the ink molecules in the physics journal where he publishes his results. There will not be any one particular fact of the matter about what his mom hears or what is printed in the journal; but "relative to" the observer seeing the pointer at $y = \lambda E_k t$ (and relative to the pointer being at $y = \lambda E_k t$ and relative to the energy of the PIB being E_k) mom will hear, and the journal will report in print, that the energy measurement came out E_k. And so on.

It is as if the big (and ever-expanding) entangled superposition, which we previously took as just somehow obviously wrong, is actually describing all of these perfectly coherent stories playing out in parallel. Except that, for Everett, it is not merely "as if" this were the case. Everett's idea is precisely that this is literally the case. By eliminating "Process 1" and letting the universe be described by a single wave function, evolving always exclusively according to the linear Schrödinger equation (Process 2), we arrive at the following picture: whenever we would have said (according to one of the previously considered formulations of QM) that there were several distinct possibilities, only one of which is in fact realized, instead in Everett's theory *all* of the possibilities are realized; the world splits into several branches, each of which realizes one of the possibilities. Further interactions then produce further branchings in each original branch, and so on. The overall pattern of iterative branching is indicated schematically in Fig. 10.1.

A few words of clarification are in order. First, although Everett's theory is often called the "many worlds" theory and the different branches are sometimes referred to as (for example) "parallel universes", these turns of phrase can also cause confusion

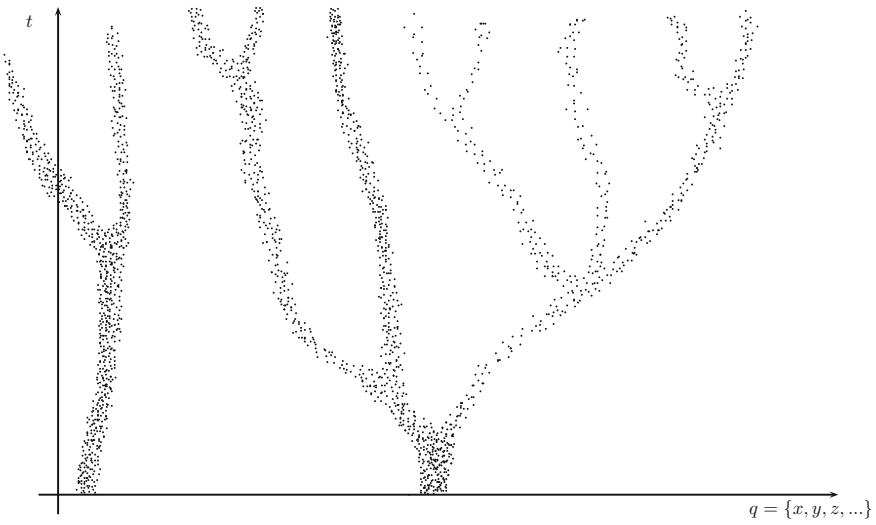


Fig. 10.1 Schematic depiction of the wave function of the universe, evolving in time, with an iterative branching structure

and suggest misleading pictures. Really, according to Everett's theory, there is only one universe, only one world. It's not, for example, that every time a quantum event (triggering a branching) occurs, a whole new copy of the physical universe is created, ex nihilo, "next to" the old one, such that, over time, more and more and more universes are all existing, in some sense separately. (People who misunderstand Everett's idea in this way often complain, for example, that the multiplication of worlds flagrantly violates the idea of mass or energy conservation.)

Instead, it is supposed to be the case that the matter in the one, only-existing universe just has these different patterns going on in it, all, so to speak, on top of one another. Perhaps a good analogy here would be to light waves: if you're driving your car during the daytime and turn on the headlights, the region in front of the lights has (let's say) some light waves, propagating east, emitted by the headlights – and also some light waves, headed down, emitted by the sun. These two things are happening in the same place and are associated with the same one underlying field. They are distinct structural patterns in that field. But the dynamics of the field is such that the two patterns do not affect each other. The light waves from the sun just do their thing, passing downward, in the same way they would if the light from the car headlights were not there, and vice versa. The non-interaction of these two light waves explains why it is appropriate to think of what's going on in terms of these two overlapping but distinct patterns.

One should remember, though, that unlike electromagnetic waves which propagate in 3-dimensional physical space, the quantum mechanical wave function exists in a very high-dimensional space. So one should for example recognize that the horizontal, "spatial" axis in Fig. 10.1 is a very schematic simplified way of representing what is in fact a space of enormous dimension. This is also relevant to understanding why the different branches, once formed, do not interact. In principle, packets can interact, by *interfering* with each other, if they overlap. But here "overlap" means "overlap in configuration space" – because that is where the wave function lives. If one is thinking about a single particle moving in one dimension, it may seem very probable that, for example, if the wave packet splits into two "branches", one of which moves off to the left and the other to the right, it might occur (for example if one of the packets reflects off something and moves back the other way) that the two packets might come again to overlap, producing some interference effects. *In princple* this can happen, but due to a phenomenon called "decoherence" this basically never happens in practice once the difference between two branches becomes macroscopic (which by the way is when you'd first be justified in thinking of them as distinct branches).

You can think of it this way: configuration space is *really high-dimensional*. So there's just a lot of room there. If a branching event occurs, like when our energy measuring device interacts with the PIB, it's not just – as our schematic treatment in terms of the center of mass coordinate y might suggest – that the two wave function packets separate by a small macroscopic distance d (say, a centimeter). In fact, there are some enormous number – 10^{23} or something – of particles in the pointer. So the two wave packets in configuration space are not just separated by distance d. Rather, they are separated by distance d *in each of some* 10^{23} *distinct coordinates*. So, by

a high-dimensional analog of the Pythagorean theorem, the packets are actually separated by something like a centimeter times $\sqrt{10^{23}}$, i.e., about *two million miles*. The packets are, for all practical purposes, permanently and irreparably separated by a vast distance in configuration space, never to interact again. (And note that the separation and its irreparability only continue to increase as the pointer interacts with air molecules in its vicinity, which then in turn interact with further degrees of freedom that they are in contact with, and so on.)

So that is a nice way to think about the process, decoherence, that makes these individual branches in the wave function very stable, separate, non-interacting. What might seem like a very small difference between two branches actually (when we remember the enormous and ever-increasing number of particles that are involved) implies that the branches are extremely well-separated in the vast open wilderness of configuration space and will hence never see each other again.

Everett summarizes the overall idea as follows:

> We thus arrive at the following picture: Throughout all of a sequence of observation processes, there is only one physical system representing the observer, yet there is no single unique *state* of the observer (which follows from the representations of interacting systems). Nevertheless, there is a representation in terms of a *superposition*, each element of which contains a definite observer state and a corresponding system state. Thus with each succeeding observation (or interaction), the observer state 'branches' into a number of different states. Each branch represents a different outcome of the measurement and the *corresponding* eigenstate for the object-system state. All branches exist simultaneously in the superposition after any given sequence of observations.
>
> [In a footnote he adds:] From the viewpoint of the theory *all* elements of a superposition (all 'branches') are 'actual', none any more 'real' than the rest. It is unnecessary to suppose that all but one are somehow destroyed, since all the separate elements of a superposition individually obey the wave equation with complete indifference to the presence or absence ('actuality' or not) of any other elements. This total lack of effect of one branch on another also implies that no observer will ever be aware of any 'splitting' process [3].

10.2 Probability

In the last section, we started to come to grips with Everett's central idea of simply omitting, from the axioms of quantum theory, the measurement postulates (such as the Born rule) which seem difficult to reconcile, at the fundamental level, with Schrödinger's equation. However, in the conventional interpretation, these measurement postulates provide practically the entire *testable* content of the theory – they tell us, in particular, about the probabilities for various possible measurement outcomes. And it is precisely the fact that these probabilities match up with the empirically observed outcome frequencies, that we believe in the quantum formalism in the first place. So if Everett's "many worlds" theory is to be worth taking seriously at all, it will need to be able to account for these conventional probabilistic claims.

From his very first presentation of the many worlds idea, Everett recognized the importance of being able to somehow derive the Born rule (in the context of his new

theory which adamantly does not just posit it as an axiom). Indeed, Everett claimed to provide such a derivation/explanation already in 1957:

> The new theory is not based on any radical departure from the conventional one. The special postulates in the old theory which deal with observation are omitted in the new theory. The altered theory thereby acquires a new character. It has to be analyzed in and for itself before any identification becomes possible between the quantities of the theory and the properties of the world of experience. The identification, when made, leads back to the omitted postulates of the conventional theory that deal with observation, but in a manner which clarifies their role and logical position [3].

However, Everett's claim has been met with skepticism and in general this issue has remained a highly controversial one ever since Everett's original proposal.

Let's try to understand what's at issue here, starting with a simple example. Suppose an experimenter prepares a spin 1/2 particle in the "spin-up along the x-direction" state,

$$\psi_{+x} = \frac{1}{\sqrt{2}} \left(\psi_{+z} + \psi_{-z} \right), \tag{10.8}$$

and then performs a measurement of the z-component of the particle's spin. According to conventional quantum mechanics, we'd say that there is a 50% probability that the measurement comes out spin-up (let's call that "heads" for simplicity here) and a 50% probability that it comes out spin-down ("tails"). But of course, in Everett's view, that's not right. Instead, according to Everett, both things happen: the act of measuring the z-spin (i.e., setting up a coupling between the z-component of the particle's spin and some eventually-macroscopic degrees of freedom that include those belonging to the observer himself) triggers a branching of the universal wave function, and each outcome occurs in one of the two branches. As it is sometimes put, the observer has two "descendants" – one who sees the experiment come out "H" and one who sees it come out "T".

Now suppose the experimenter does this N times – that is, suppose he prepares a *bunch* of spin 1/2 particles in the state ψ_{+x} and then measures their z-spins, one at a time. The branching structure that will be produced is illustrated, for the case $N = 4$, in Fig. 10.2. At the end, there are $2^4 = 16$ different branches, and the experimenter observed a different sequence in each one: $HHHH$ for the branch on the left, $HHHT$ for the next one over, and so on, all the way over to $TTTT$ on the right. All 16 of these branches appear, in the expression for the total wave function, with the same amplitude, so aside from each involving a distinct sequence of outcomes, they all seem to be on an equal footing.

However, for purely combinatoric reasons, certain statistical patterns of outcomes occur in more branches. For example, there is only the one branch in which the observer saw 4 Hs, and similarly there is just the one branch in which the observer saw 4 Ts. But there are *four* branches in which the observer saw 3 Hs and 1 T. (These four branches have the following sequences: $HHHT$, $HHTH$, $HTHH$, and $THHH$.) Similarly, there are four branches in which the observer saw 1 H and 3 Ts. And finally there are *six* branches in which the observer sees 2 Hs and 2 Ts. One begins to see the overall pattern: although every possible sequence of outcomes

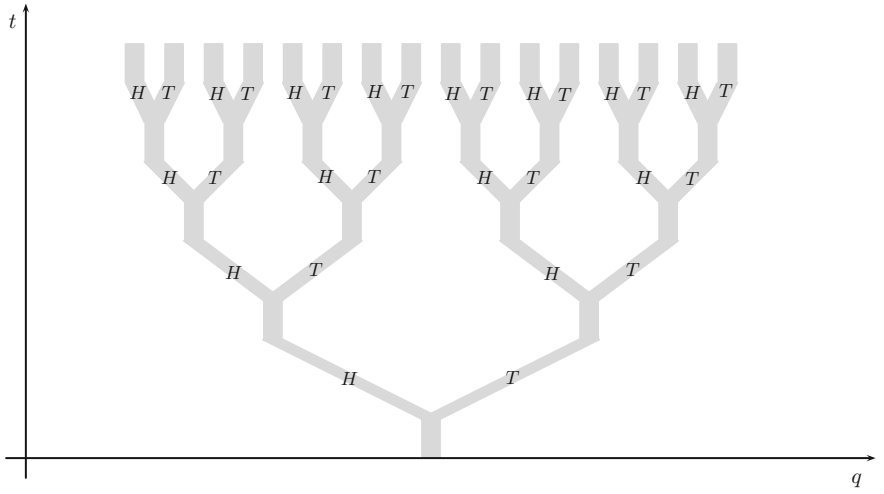


Fig. 10.2 The branching structure created by an experiment in which a "quantum coin" is flipped 4 times

occurs in precisely one world, *most* of the worlds will exhibit statistics that are close to those associated with the Born rule (here, equal numbers of Hs and Ts).

In the general case of N binary quantum measurements (which we'll continue to think of as coin flips for simplicity), the number $g_N(n)$ of worlds in which exactly n Hs are observed will be "N choose n":

$$g_N(n) = \binom{N}{n} = \frac{N!}{n!\,(N-n)!}. \tag{10.9}$$

And so (there being 2^N worlds at the end of the sequence of experiments) the *fraction* $f_N(n)$ of worlds in which exactly n Hs are observed will be

$$f_N(n) = \binom{N}{n}\left(\frac{1}{2}\right)^N = \frac{N!}{n!\,(N-n)!}\left(\frac{1}{2}\right)^N. \tag{10.10}$$

For large N, this function of n is well-approximated by a normalized Gaussian whose center point is at $n = N/2$ and whose half-width σ is $\sqrt{N}/2$. That is,

$$f_N(n) \sim e^{-(n-N/2)^2/(N/2)}. \tag{10.11}$$

See Fig. 10.3 for a sketch.

So again, although each possible sequence is realized in exactly one branch, and no sequence is in any way preferred over any other, it is the case that, at the end of the experiment, the overwhelming majority of observers – that is, the overwhelming majority of descendants of the original observer – will observe a sequence in which

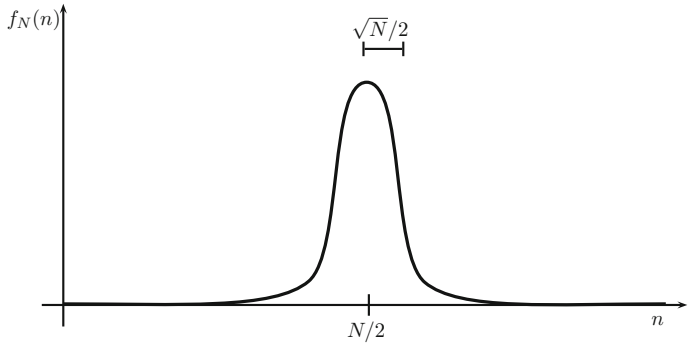


Fig. 10.3 In an experiment involving N binary measurements (in which, according to conventional quantum theory, the probability for each possible outcome is 50%), the fraction $f_N(n)$ of Everettian worlds in which n Hs are observed will be sharply peaked around $n = N/2$. That is, the overwhelming majority of observers in the different worlds (i.e., the overwhelming majority of descendants of the original experimenter) will see approximately Born rule statistics

there are roughly equal numbers of Hs and Ts. That is, we can reproduce the Born rule by looking at the statistical patterns that are *typical* for universes, i.e., present in *most* of the universes. Of course, there will some universes in which very non-Born-rule statistics (e.g., a string of N consecutive Hs!) will be observed. But so long as such rogue universes represent a vanishingly small fraction of the total number of universes it perhaps seems somewhat reasonable to ignore them and claim that, according to Everett's theory, observers should typically expect to see Born-rule statistics.

But there is a serious problem with this line of thinking: it only works for the special case that, in the conventional way of describing the situation, the outcome probabilities are 50/50. Or, to put the same point in Everettian terms, it only works for the special case in which the two branches created by each individual measurement event appear (in the overall expression for the wave function) with equal amplitudes. To see this, let's consider the more general case in which, say, the initial preparation of each spin 1/2 particle has it being spin-up along a direction $\hat{n}$ such that

$$\psi_{+n} = \sqrt{p}\psi_{+z} + \sqrt{q}\psi_{-z} \tag{10.12}$$

where $p + q = 1$, i.e., p is what would ordinarily be called the probability of H (i.e., the particle coming out spin-up along z), but which is in the context of Everett's theory instead called the *branch weight* of the "spin-up along z" branch that the measurement creates.

It is easy to see that, for the $N = 4$ case, the "tree of outcome sequences" is exactly the same as what was already displayed in Fig. 10.2. The only difference is that now the branch weights are not all equal. For example, the $HHHH$ branch has a branch weight p^4; the four branches with three Hs and one T each have branch weight p^3q; the six branches with two Hs and two Ts have branch weight p^2q^2; and so on. In general, the weight of a branch with n Hs and $(N - n)$ Ts will be

$$w_N(n) = p^n q^{(N-n)} = p^n (1-p)^{(N-n)}. \tag{10.13}$$

Now, the Born rule tells us that (in conventional terms) the probability of a H for each flip is p. So in a sequence of N flips, the expected number of Hs will be Np. For example, if $N = 100$ and $p = 90\%$ we should expect to see about 90 Hs. But if we just naively count worlds the way we did before, it remains true that the overwhelming number of worlds have approximately 50 Hs and approximately 50 Ts.

Therefore, in order to continue accounting for the usual Born rule statistics in the Everettian model, it is necessary to *weight* the worlds differently – and in particular to weight each branch by, what else, its branch weight – when we compute the world-fraction which displays a certain characteristic. We thus define the weighted world fraction as follows:

$$f_N^w(n) = g_N(n) w_N(n). \tag{10.14}$$

(Note that what we called $f_N(n)$ before is just this same formula but for the special case $p = q = 1/2$ in which the weight function $w_N(n)$ is equal to $1/2^N$ independent of n.) One can show that this weighted world fraction function is, for large N, sharply peaked around $n = Np$. (See the Projects.) That is, when we include the non-equal weightings, we can still say that the overwhelming majority of worlds (in the weighted-by-their-branch-weights sense) will exhibit approximately Born rule statistics. The idea here is visualized in Fig. 10.4.

That sounds good, but also raises a number of questions. For example: what, exactly, are these "branch weights" that we've been talking about? Well, they are nothing but the (absolute) squares of the amplitudes of the different branches, i.e., the different terms in the universal wave function. If, that is, after some sequence of measurements, the universal wave function has the structure

$$\Psi = \sum_i c_i \psi_i \tag{10.15}$$

(where the index i is labeling particular sequences of measurement outcomes, or whatever) the branch weight associated with the ith branch is just

$$w_i = |c_i|^2. \tag{10.16}$$

That is, the formula for the branch weights – the equation telling us how much to "care" about each individual branch in the tree – is really just the Born rule. So the overall argument has a strong air of circularity about it: if you weight the branches using the Born rule, then (the Born-rule-weighted-sense-of) "most" of the branches will display Born rule statistics. It seems that we get the Born rule out (as a description of the statistics that will be observed in typical branches) only because we put the

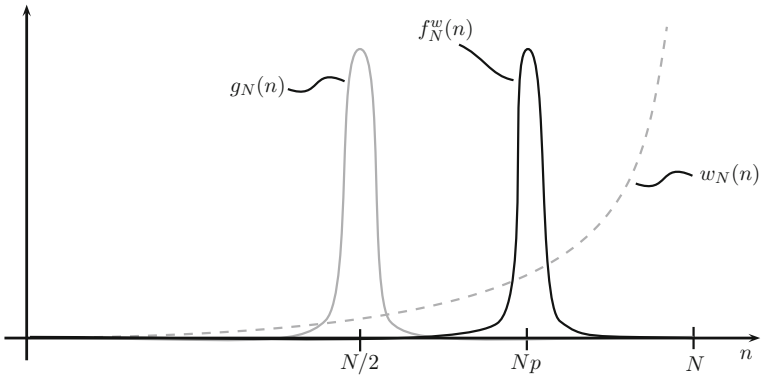


Fig. 10.4 Graphical illustration of that fact that, although the (raw, unweighted) number of worlds in which n "Heads" results appear in a sequence of N quantum coin flips is strongly peaked around $N/2$, we can nevertheless say that (in a weighted sense) "most" worlds will display Born rule statistics, i.e., $n \approx Np$. Here the raw world-counting function $g_N(n)$ is shown as the solid gray curve; the weighting function $w_N(n)$, according to which worlds with larger amplitude or "branch weight" are more strongly emphasized in the accounting, is shown as the dashed gray curve; and the weighted world fraction $f_N^w(n) = g_N(n)w_N(n)$ is shown as the solid black curve. The case $p = 3/4$ is shown

Born rule in (as a measure of how much each branch should count in our assessment of what is typical).

There is a long history of proponents of the many worlds interpretation trying to give further arguments to prove that Eq. (10.16) is somehow the only mathematically reasonable way to weight the different branches. If this could be convincingly established, it would significantly reduce the feeling of circularity. You are invited, in the Projects, to analyze and assess the argument that Everett presented already in 1957.

But there are some deeper concerns as well. For example, the very idea that we should use this un-equal weighting seems somewhat in conflict with Everett's overall idea. Recall, for example, Everett's statement that "none [of the branches is] any more 'real' than the rest". But what is this non-trivial weighting function, other than some kind of measure of how real, exactly, each (supposedly equally real) branch is? Suppose there are just two branches, one – say, in which a red light flashes – with a branch weight of $1/100$, and the other – in which, say, a blue light flashes instead – with a branch weight of $99/100$. Everett would have us say that the vast majority of worlds – namely, 99% of them – include a flashing blue light. This may make some kind of sense from the point of view of an external God-like observer, who can somehow "see" that the blue-light world is *brighter* (more intense? heavier?? more real???) than the red-light world. But in some sense Everett's whole program is to abandon, as non-existent and meaningless, the idea of such external God-like observers, and instead to exclusively consider what the world is like "from the inside", i.e., according to observers who are part of the world and governed by its fundamental laws.

There has indeed been a trend in the recent literature on this issue, away from treating the "branch weights" as somehow objective facts that we (the theorists analyzing the merits of Everett's theory) must acknowledge, and toward treating them instead as measures of how much individual observers within an Everettian world should care about their various descendants. As Simon Saunders has summarized this point,

> In recent years, with the development of decision-theory methods for quantifying subjective probability in quantum mechanics, the link between probability in the subjective sense and an objective counterpart has been greatly clarified. Specifically, it can be shown that agents who are rational, in order to achieve their ends, have no option but to use the modulus squared branch amplitudes in weighting their utilities. In this sense the Born rule has been *derived* [5].

This claim, though, remains controversial, with questions proliferating about: the necessity of defining (or, indeed, expecting) the "rationality" of agents in the way required in the derivation; the appropriateness and relevance of focusing on how agents who believe in Everett's theory should behave as opposed to explaining why we should interpret our empirical observations in Everettian terms; and even whether the concept of "probability" can possibly mean *anything* in a picture where, with certainty, everything that can happen *will* happen.

We will not be able to resolve these issues here. What should be clear, though, is that the unusual, many-worlds character of Everett's proposal forces us to reimagine certain concepts – like "probability" – that play an important role in quantum theory. As we will see, this uncomfortable "stretching" of concepts previously thought to be well-understood is a theme that will re-appear as we continue our exploration of Everett's proposal.

10.3 Ontology

So far we have followed Everett (and his followers) in essentially taking for granted that each branch of the universal wave function can be understood as describing a sensible physical world, with stars and planets and trees and cats and measuring equipment and observers with brains (whose physical states give rise to appropriate conscious experiences) and so on. That is, we have just been assuming that (at least at an appropriately macroscopic coarse-grained level) each branch of the wave function corresponds to a physical world basically of the sort we take ourselves to experience.

But we should remember, from Chap. 5, that the wave function is a funny and abstract kind of mathematical object. There is no obvious and straightforward sense in which the wave function can be understood to directly describe physical goings-on in ordinary three-dimensional space, because the wave function is something like a field on an abstract, high-dimensional configuration space. So we should not simply take for granted that the wave function (or any individual branch of the wave function) describes a three-dimensional physical world of the sort we are accustomed to

imagining exists. Instead, we should ask: if it does, how, exactly, does the description work?

One possibility is Schrödinger's original idea that the wave function can be used to compute a mass density field (on physical, 3D space). Recall that, in this scheme, the mass density of the ith particle would be given by

$$\rho_i(x,t) = m_i \int |\Psi(x_1, x_2, \ldots, x_N, t)|^2 \, \delta(x - x_i) \, dx_1 dx_2 \cdots dx_N \tag{10.17}$$

and the total mass density would then be

$$\rho(x,t) = \sum_i \rho_i(x,t). \tag{10.18}$$

In the context of the GRW theory we discussed in Chap. 9, in which only one branch of the universal wave function survives the spontaneous collapses, we were able to recognize this mass density field as corresponding to a world that "looks right" at the macroscopic scale. But in the context of Everett's proposal – in which all branches of the universal wave function survive – the mass density field becomes a big incoherent mess. In an illuminating discussion of this idea [6] an analogy has been given to an old TV set which is badly tuned and is therefore receiving and displaying the programs from several different channels all at once. Indeed, this was the primary reason that Schrödinger himself abandoned this idea as a possible way of understanding the ontology associated with the quantum wave function.

But is the mess really so incoherent? Just like, in the TV set analogy, the different programs (being displayed on top of one another) do not *interact* with each other, so the contributions to the mass density field from different branches of the wave function remain dynamically independent. That is, just like two characters from one of the TV programs will interact with each other (but neither can in any sense interact with the characters from one of the other simultaneously-displayed programs), so with the different contributions to the mass density field associated with different branches of the wave function. We should recognize, that is, that the total mass density field, Eq. (10.18), only looks like an incoherent mess to some God-like external observer (and only then, perhaps implausibly, if She is unable to disentangle the overlapping programs). In keeping with the Everettian philosophy, though, we recognize this as irrelevant. To an observer living in that universe, himself made out of some portion of $\rho(x,t)$ which only interacts with other portions of $\rho(x,t)$ arising from the same branch of the universal wave function, the world looks entirely coherent. Such an observer would, in effect, be happily oblivious to the fact that there were countless alternative programs playing out literally right on top of him, but the (limited part of the) world he actually experiences would indeed "look right" – i.e., have the same kind of overall macroscopic coherence we are familiar with from our actual experiences.

Let us explain, more formally, how and why the different contributions to $\rho(x,t)$ arising from different branches of the wave function can be thought of

as non-interacting and causally independent. Suppose the wave function can be written as a linear combination of macroscopically-distinct packets

$$\Psi(x_1, x_2, \ldots, x_N, t) = \sum_\alpha c_\alpha \Psi_\alpha(x_1, x_2, \ldots, x_N, t) \tag{10.19}$$

where, as discussed in Sect. 10.1, the individual packets are well-separated in configuration space so that

$$\int \Psi^*_\beta(x_1, x_2, \ldots, x_N, t)\Psi_\alpha(x_1, x_2, \ldots, x_N, t)\, dx_1,\ dx_2 \cdots dx_N = 0 \tag{10.20}$$

if $\alpha \neq \beta$. (The requirement that the different terms be *macroscopically* distinct effectively ensures that two terms which are orthogonal in this sense at one time will remain orthogonal in the future.) It then follows from Eq. (10.17) that the mass density associated with the ith particle can be written as

$$\rho_i(x, t) = \sum_\alpha |c_\alpha|^2 \rho_i^\alpha(x, t) \tag{10.21}$$

where

$$\rho_i^\alpha(x, t) = m_i \int |\Psi_\alpha(x_1, x_2, \ldots, x_N, t)|^2\, \delta(x - x_i) dx_1 dx_2 \cdots dx_N \tag{10.22}$$

is the mass density of particle i arising specifically from the α branch of the wave function.

The total mass density can then similarly be written as

$$\rho(x, t) = \sum_\alpha \rho_\alpha(x, t) \tag{10.23}$$

where

$$\rho_\alpha(x, t) = \sum_i \rho_i^\alpha(x, t) \tag{10.24}$$

is the total mass density associated with the α branch of the wave function.

The important point here is captured by Eq. (10.23), which says that the total mass density can be broken apart into distinct pieces that (like the programs from different TV channels that are being displayed simultaneously) each play out independently of the others. In a sense, there is nothing new here compared to the way we were thinking about Everett's many-worlds proposal previously. The point is just that the mass density ontology provides a definite, viable way of extracting, from the evolving universal wave function Ψ, a coherent (many worlds!) story about physical goings-on in ordinary three-dimensional space, i.e., this is a way to give a precise meaning to the way we were already talking about Everett's idea in earlier sections.

A few contemporary proponents of the Everettian picture (for example, Lev Vaidman) seem to basically understand the theory in this way. But for the most part, Everett's contemporary followers resist the idea that some special, explicit postulate about the ontology of the theory is required.

The reason for this resistance is the idea that one of the main virtues of Everett's approach is its elegance, its parsimony: there is just the wave function, obeying Schrödinger's equation, full stop. This is supposed to be in contrast, for example, to the pilot-wave picture, which followers of Everett would regard (because it posits not only the wave function obeying Schrödinger's equation, but in addition particles moving in accordance with some further dynamical law) as ontologically cluttered and cumbersome. That is, "wave function monism" – the idea that the wave function is all there is – plays a very important role, for Everettians, in explaining and justifying their preference for the Everettian theory.

One can indeed appreciate how an explicit endorsement of something like Schrödinger's mass density ontology – Eqs. (10.17) and (10.18) – would feel dangerously and suspiciously similar to the pilot-wave theory's explicit postulation of additional ontology. But, of course, the problem is that it is very difficult to understand what to make of Everett's theory if one just says "the wave function is everything" and leaves it at that. Independent of whatever worries one might have about the many worlds idea, such a position would mean that the theory suffers rather acutely from the ontology problem we discussed in Chap. 5.

So the problem faced by proponents of Everett's theory is to, on the one hand, avoid the ontology problem by finding some way of extracting, from the theory, an explanation for our experience of material objects moving and interacting in three-dimensional space, while at the same time avoiding the need to postulate additional things, distinct from and additional to the wave function itself. One approach to this problem has been to argue that familiar macroscopic structures in 3D can be understood to emerge from the structure in the wave function, in the same way that complicated macroscopic objects like, say, tigers can be understood as complex macro-patterns of more basic ontological posits. As David Wallace elaborates,

> It is simply untrue that any entity not directly represented in the basic axioms of our theory is an illusion. Rather, science is replete with perfectly respectable entities which are nowhere to be found in the underlying microphysics.... Tigers [for example] are (I take it!) unquestionably real, objective physical objects, but the Standard Model [of particle physics] contains quarks, electrons and the like, but no tigers. Instead, tigers should be understood as patterns, or structures, *within* the states of that microphysical theory.... The moral of the story is: there are structural facts about many microphysical systems which, although perfectly real and objective (try telling a deer that a nearby tiger is not objectively real) simply cannot be seen if we persist in describing those systems in purely microphysical language. Talk of zoology is of course grounded in cell biology, and cell biology in molecular physics, but the entities of zoology cannot be discarded in favour of the austere ontology of molecular physics alone. Rather, those entities are structures instantiated within the molecular physics, and the task of almost all science is to study structures of this kind [7].

The idea is that, in something like that same way, the ordinary world of macroscopic objects (including tables and chairs and planets and trees and cats and

human observers) is already there, instantiated within the complicated ripplings in the structure of the universal wave function. In particular:

> Structurally speaking, the dynamical behaviour of each wavepacket [i.e., each decoherent branch of the wave function] is the same as the behaviour of a macroscopic classical system. And if there are multiple wavepackets, the system is dynamically isomorphic to a collection of independent classical systems [7].

That, I think, is exactly correct, but seems also to miss the point of the ontology problem.

It is true that a relatively narrow and well-isolated wave packet propagating through $3N$-dimensional configuration space is equivalent to an (approximate) *trajectory* through $3N$-dimensional configuration space and hence, in turn, isomorphic to (i.e., mathematically interchangeable with) a description of N particle trajectories in 3-dimensional space, i.e., a classical system. But surely this mathematical isomorphism does not imply that the real physical existence of a propagating wave packet in $3N$-dimensional space somehow brings about the additional real physical existence of N particles moving and interacting in 3D. A single billiard ball, bouncing around on a square two-dimensional billiards table, for example, is mathematically isomorphic to two beads (one small enough to pass through the hole of the other so they don't interact with each other) bouncing back and forth from the ends of a wire. Does each really-existing billiard ball on a table thus somehow call into existence a pair of beads on a wire somewhere? Nobody believes this, yet it seems like a perfectly fair analogy to what would be required for a wavepacket in configuration space to genuinely give rise to a classical system of particles in three-dimensional, physical space.

The sticking point is really the trans-dimensional character of the required sort of emergence. If, for example, the fundamental quantum mechanical description were in terms of N single-particle wave functions propagating in 3D space, there would be *no difficulty at all* in understanding how a rough macroscopic description in terms of atoms, molecules, and ultimately tigers, could be appropriate and entirely consistent with that fundamental ontology. There is no problem, that is, in understanding how something like a tiger can be understood as emerging from a fundamental ontology involving waves. The problem is in understanding specifically how something like a tiger (which is a certain complex pattern of microscopic goings-on *in three-dimensional space*) could be understood as emerging from a fundamental ontology involving only waves that live in an entirely different, much higher-dimensional space.

Perhaps some ultimately-satisfying account of the needed sort of trans-dimensional emergence could be given. Or perhaps this is the wrong way to think about it. Wallace, for example, seems to have suggested that the very appearance that we live in a three-dimensional world could itself be emergent:

> Note firstly that the very assumption that a certain entity [namely, a certain branch of the universal wave function] which is structurally like our world is not *our world* is manifestly question-begging. How do we know that space is three-dimensional? We look around us. How do we know that we are seeing something fundamental rather than emergent? We

> don't; all of our observations ... are structural observations, and only the sort of a prioristic knowledge now fundamentally discredited in philosophy could tell us more [7].

He means here that the idea that we live in a three-dimensional world should not be taken as some kind of a priori dogma which has to appear, in stone, at the most basic level. This, too, could be emergent from some very different more elementary processes, as they appear to rough creatures like us. But (although Wallace would certainly deny that this is an appropriate way to express it) there is a suggestion here that the three-dimensionality of the world is then something like an illusion. And if we can be deceived, through our ordinary direct perceptual experience of the world, about something so basic as that, one worries that it might become difficult to hold off concerns about what else we might have been deluded about, and so why we should believe the quantum formalism in the first place.

Again, the goal here is not to resolve these issues, but to help make you aware of their existence. Suffice it, then, to note that (just as with "probability"), controversial questions about ontology persist for Everett's theory. Can the theory explain the existence (or, at least, the appearance) of familiar three-dimensional material worlds of the sort we ordinarily take ourselves to inhabit and of which, according to the theory, there are actually many? And in particular, can it account for such worlds exclusively on the basis of the universal wave function? The needed structures are unquestionably present there – as shown by the possibility of understanding the ontology in terms of Schrödinger's mass density field. But does the singling-out of, for example, that particular bit of structure – as the thing we should look at to understand what the theory says about goings-on in three-dimensional space – constitute the postulation of additional ontology, beyond the wave function, as is done unapologetically in the pilot-wave theory? If so, it is hard to understand why one would not then just prefer to adopt the pilot-wave theory and skip the difficulties (pertaining, for example, to "probability") that arise due to the many-worlds character of Everett's theory. But if not, why only that particular structure? And are certain things we took as basic facts about our world (like its three-dimensionality) then rendered merely illusory, and, if so, is that even a problem?

These are some of the questions that would, I think, need to be addressed before an Everettian theory could be considered to be as ontologically satisfactory as the pilot-wave theory, or GRWm, or GRWf.

10.4 Locality

The question of whether Everett's theory respects relativistic local causality is yet another subtle and controversial one. Among the theory's supporters, it is widely believed that the theory is – uniquely among available options – locally causal. And so this claim, that Everett's theory is somehow uniquely compatible with relativity, is a big part of the reason why the theory's supporters support it.

The claim is generally based on two different lines of reasoning. The first is that, in ordinary quantum mechanics, the non-locality originated from the wave-function collapse postulate. So, by retaining only (an appropriately relativistic generalization of) the Schrödinger equation – i.e., by simply eliminating the collapse postulate from the dynamics – Everett's theory supposedly retains the local part, and abandons the nonlocal part, of ordinary QM, and is therefore itself perfectly local.

The second line of reasoning addresses the question of how Everett's theory supposedly eludes Bell's proof (discussed in Chap. 8) that *any* empirically viable theory must include nonlocal dynamics. The claim here is that Bell's arguments involve a previously-unacknowledged assumption which does not apply to Everett's theory – namely, Bell assumes that the spin measurements (made by Alice and Bob at opposite ends of the experimental setup) *have definite outcomes*. That is, Bell assumes that, for each individual spin measurement, there is a particular unambiguous result – the particle is either found spin up along the axis in question, or spin down. Bell's inequality is then a constraint on the statistical correlations that are possible, if locality is respected, within this set of unambiguous particular measurement outcomes. But, the Everettians point out, it is simply not true in Everett's theory that each individual spin measurement has a single particular outcome; instead, there is a branching point in the structure of the world and both "possible" outcomes are realized, one in each branch.

It is certainly true that Bell wasn't anticipating that kind of possibility and that his theorem does indeed tacitly assume that experiments have particular, definite, single realized outcomes. And there is some value in pointing out precisely how the many-worlds theory manages to elude Bell's general argument. But in a way it is unnecessary to ask, of any extant candidate theory, whether and how Bell's theorem applies to it. (The use of Bell's theorem is that it allows us to diagnose, as either non-local or in conflict with the experimental facts, all of the *non-extant* theories – the theories that nobody has managed or bothered to think of yet.) We can instead assess the theory's status vis-a-vis locality directly, by just seeing whether or not the theory respects our explicitly formulated notion of locality from Chap. 1.

When we attempt to do this for Everett's theory, however, we immediately realize that the difficulties we reviewed in the last two sections – pertaining to the ontology of the theory and the role and meaning of probability within it – preclude anything like a straightforward diagnosis. If, for example, we understand the theory as positing the existence of nothing but the wave function – thought of as a kind of field in a 3N-dimensional abstract space, and with the appearance of three-dimensionality being some kind of emergent delusion within our conscious experiences – then it will be completely impossible to say anything meaningful about whether the theory does or does not respect locality. Locality, remember, is the idea that causal influences between physically real objects in ordinary 3-dimensional space never propagate faster than the speed of light. If, according to a theory, there *are* no physically real objects in ordinary 3-dimensional space, then concepts like "local" and "non-local" are simply, radically, fatally, inapplicable. The theory, so understood, would be "not even non-local" in precisely the sense introduced back in Chap. 5.

Of course, this is just an extreme example, intended to make a pedagogical point. Probably no actual proponent of Everett's theory would endorse the perspective described in the previous paragraph. Still, it is a crucial and under-appreciated point that a theory has to clearly articulate an ontology of physical objects in three-dimensional space (and, if that ontology is not openly posited like the particles of the pilot-wave theory, must explain clearly how the ontology of physical objects in three-dimensional space relates to and emerges from whatever *is* openly posited) before the theory's status vis-a-vis local causality can be meaningfully assessed.

Ambiguities surrounding the concept of "probability" also prevent a straightforward application, to Everett's theory, of Bell's formulation of locality. Recall that, in Bell's formulation, "locality" was the requirement that the probability assigned to each event in space-time, conditioned on a complete description of events in a slice across the past light cone, should be independent of events with suitable space-like separation. In earlier chapters, we have always been concerned in particular with events that correspond to definite, observable occurrences such as a certain experiment having a particular outcome. Such events can of course still be said to occur in Everett's theory, but (what would previously have been described as) the different possibilities are not related to one another in the familiar way in Everett's theory, and this undermines and obscures the applicability of certain probabilistic ideas.

For example, it would normally be assumed that, since a given spin measurement on a spin-1/2 particle has two possible outcomes, two probabilities like $P(\text{up}|\lambda)$ and $P(\text{down}|\lambda)$ – where "up" and "down" mean, respectively, that the "spin-up" and "spin-down" outcomes are manifested in the macroscopic ontology in the appropriate space-time region – should sum to 100%:

$$P(\text{up}|\lambda) + P(\text{down}|\lambda) = 1. \tag{10.25}$$

But, in the context of Everett's theory, each of these probabilities is (for generic λ) already 100%: *both* outcomes will, with certainty, be instantiated, right on top of one another, in the appropriate space-time region. Thus, for Everett:

$$P(\text{up}|\lambda) + P(\text{down}|\lambda) = 2. \tag{10.26}$$

This illustrates the sense in which certain basic assumptions about how probabilities work – having to do with the mutual exclusivity of the possibilities to which we conventionally assign probabilities – take a very different form in the context of Everett's theory and thus prevent certain probability assignments from working in the familiar ways.

In any case, you can perhaps begin to see why the question of whether the many worlds theory is local, is a subtle and controversial one. Not only is the many-worlds character of the theory *weird* in a profound and radical way (so that normal everyday assumptions, like that experiments always have definite specific outcomes, as well as more technical assumptions like that the probabilities associated with what we normally think of as distinct possible outcomes should sum to one, fail to apply), but it also remains very obscure how/whether the theory's postulates relate to and/or

account for and/or give rise to physical objects and process in ordinary 3-dimensional physical space.

Still, let us attempt to set these abstract worries aside, and get a concrete feeling for how the many worlds theory talks about one of the example situations we've used to discuss the non-locality of other theories.

To make things as definite as possible, we'll consider a version of the theory in which a mass density field $\rho(\vec{x}, t)$ is explicitly postulated as the way to understand what 3-space ontology the wave function is describing. And let's analyze the Einstein's Boxes scenario from the point of view of this version of the theory. Suppose, then, that there is a single particle, split between two "half boxes" located at widely separate locations. Suppose further that Alice is stationed near the half-box on the left and decides to implement, at $t = 0$, a measurement to see whether or not the particle is contained in her half box; the measurement outcome is registered on a pointer which swings to the right by some distance d if the particle is detected, and stays stationary if the particle is not detected. Meanwhile, Bob is stationed near the half-box on the right and also decides to implement, at $t = 0$, a similar measurement. Thus, prior to $t = 0$, the wave function for the particle-and-two-pointers system can be written

$$\Psi(x, y, z) = \frac{1}{\sqrt{2}} [\psi_L(x) + \psi_R(x)] \phi_0(y) \chi_0(z) \tag{10.27}$$

where $\psi_{L/R}(x)$ are wave functions with support exclusively in the left/right half-boxes, and $\phi_0(y)$ and $\chi_0(z)$ are narrow wave packets centered at $y = 0$ and $z = 0$, the undeflected "ready" positions of the two pointers. The mass density ρ associated with this wave function will have contributions at the undeflected positions of the two pointers and will also involve half of the split particle's worth of mass density in each of the half boxes.

After $t = 0$, when both measurements have gone to completion, the wave function will evolve into

$$\Psi(x, y, z) = \frac{1}{\sqrt{2}} [\psi_L(x)\, \phi_0(y - d)\, \chi_0(z) + \psi_R(x)\, \phi_0(y)\, \chi_0(z - d)] \tag{10.28}$$

which is a superposition of two terms: one in which the particle is on the left, the pointer on the left has deflected (indicating that the particle was detected there), and the pointer on the right remains undeflected (indicating that the particle was not detected there) – and another in which the particle is on the right, the pointer on the left has not deflected (indicating that the particle was not detected there) and the pointer on the right has deflected (indicating that the particle was detected there). These two terms are well-separated in configuration space (especially when one remembers that our schematic degrees of freedom y and z are really proxies for some huge macroscopic number of individual particle positions) and so the two terms can be understood as describing distinct, no-longer-interacting worlds.

The mass density is then simply the sum of the individually reasonable mass densities associated with each individual world. That is, $\rho = \rho_1 + \rho_2$ where ρ_1 has

the mass density associated with the particle being contained exclusively in the left box, Alice's pointer having swung to the right indicating that the particle is there on the left, and Bob's pointer remaining in its undeflected position... and where ρ_2 instead has the mass density associated with the particle being contained exclusively in the *right* box, Alice's pointer remaining in its undeflected position, and Bob's pointer having swung to the right indicating that the particle is there on the right. See Fig. 10.5 for a sketch of how the mass density evolves during the process.

As shown in the Figure, from this "God's eye" perspective, nothing particularly dramatic happens here and there isn't much of a suggestion of nonlocality. The mass density associated with the particle-in-the-two-half-boxes is initially split between the two half-boxes and the pointers are both sitting in their ready positions. Then, as the interactions proceed around $t = 0$, the mass density associated with the two pointers splits in half so that both pointers now have "split" positions in the same way that the particle did initially. It perhaps seems plausible to say that Alice's pointer splits into these two different positions in response to the (purely local) fact that the particle is only half-contained in her half-box, and similarly for Bob's pointer. And so it may seem plausible to say that (weird though the many-worlds character here may be, with each pointer pointing to two different positions!) there is not really any suggestion of nonlocality here.

But this appearance is somewhat misleading since there are *relational facts* about the various pieces of mass density which are not captured in Fig. 10.5. In particular,

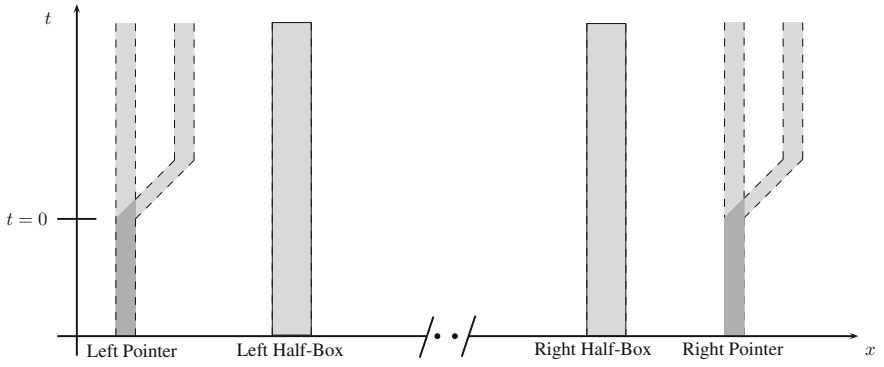


Fig. 10.5 Alice and Bob perform simultaneous measurements to detect the presence of a particle which is initially "split" between their two locations. Alice's and Bob's pointers are initially in their undeflected, "ready" positions, and the mass density associated with the particle is split between the two half-boxes. After $t = 0$, when both Alice and Bob each initiate an interaction which causes their pointer to deflect if the particle is present in their half-box, the two pointers "split", with half of each pointer's mass density remaining in the undeflected position (indicating the non-detection of the particle) and the other half of each pointer's mass density deflecting to the right (indicating the successful detection of the particle). Overall, it perhaps appears that there is no hint of nonlocality here: Alice's choice to initiate a measurement procedure causes her pointer (and shortly thereafter, her self!) to split, and similarly for Bob and his pointer, and the contents of the two half-boxes never change and hence appear unaffected by any of the measurements, distant or otherwise

there are facts, implied by Eq. (10.28), about which pieces of mass density are in the same world – the same branch – as each other. Remember here that the decomposition of the total mass density field ρ into the sum, $\rho_1 + \rho_2$, is robustly implied by the fact that the wave function is itself the sum of two disjoint terms, i.e., the two terms which are extremely well-separated in configuration space. So the very fundamental concept of Everett's theory not just allows, but requires, us to consider the separate branch-identities of these terms. See Fig. 10.6 for an attempt to visualize the same process again, but now including these relational facts about which pieces of mass density are "in the same universe as" which other pieces.

The point of this further elaboration is to stress the following: it is *not* the case that Alice's measurement merely causes a "local splitting" of her pointer and her self, with Bob's measurement also causing a second, *independent*, "local splitting" of his pointer and his self. If the two splittings were independent in this sense, you would expect that, if for example Alice and Bob get together later in the day to compare notes on the outcomes of their experiments, the interaction between the two Alices and the two Bobs would generate *four* branches: (i) one in which "yes-Alice" (i.e., the Alice who detected that the particle was present at her location) meets "yes-Bob"; (ii) one in which "yes-Alice" meets "no-Bob"; (iii) one in which "no-Alice" meets

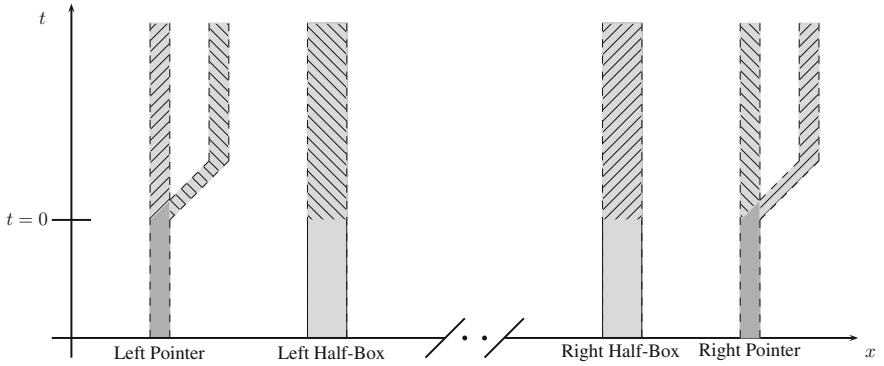


Fig. 10.6 Same as Fig. 10.5 but with the various contributions to the mass density ρ now marked to indicate the splitting into two distinct worlds or branches: The down-to-the-right striping indicates the world in which the particle is found on the left, by Alice, while the up-to-the-right striping indicates the world in which the particle is found on the right, by Bob. (Note that, prior to the measurements, the mass density associated with the particle is not identified with one or the other of these worlds, even though there would be a rather obvious way of doing this and it wouldn't be terribly misleading to do it. The reason is that, as long as the splitting of the wave function into two terms remains based on a purely microscopic difference, we don't really have separate worlds at all, in Everett's sense. (This is immaterial as long as Alice and Bob are inevitably going to carry out their measurements as we have been describing. But in principle, prior to their doing this, they could decide instead to bring the two half-boxes back together in the middle and perform some kind of interference experiment instead, and we would expect that they would indeed be able to see interference. This remains a possibility precisely because no actual branching in Everett's sense occurs until a more macroscopic number of degrees of freedom is involved in the decomposition of the wave function into disjoint terms

"yes-Bob"; and (iv) one in which "no-Alice" meets "no-Bob". But this is not right. There will not be a branch in which "yes-Alice" meets "yes-Bob" and there will not be a branch in which "no-Alice" meets "no-Bob". Only branches (ii) and (iii) will actually exist later, if Alice and Bob get together to chat, because *already*, after they have each completed their measurements but not yet gotten together to chat, there are just two distinct worlds.

This is inherent in the mathematical structure of the wave function even though it does not appear in the mass density field ρ. In general, it should be clear that, in using Ψ to compute ρ, we lose (by integrating) a lot of information. (Remember the examples, from Chap. 5, of very different wave functions which all produce the same mass density fields.) So while ρ is supposed, on this understanding of Everett's theory, to tell us what is going on in 3D physical space, there is in some sense much more that is true about goings on in physical space than is contained in the mass density field. In particular, as we are seeing here, there are relational facts – about which contributions to the mass field are in the same world as which other contributions – that really exist and have dynamical implications for how events will play out in the future as different sub-systems continue to interact with one another.

For our example here, the situation seems to be as follows. Alice's measurement of the contents of her half-box induce a splitting; she has two descendants, one of whom ("yes-Alice") sees the particle in her half-box and the other of whom ("no-Alice") fails to see the particle. Simultaneously, but at a distant location, Bob's measurement induces another splitting, and he too has two descendants, one of whom ("yes-Bob") sees and one of whom ("no-Bob") fails to see the particle in his half-box. But the two splittings are correlated despite their spatial separation: "yes-Alice" and "no-Bob" are, so to speak, born into the same post-measurement world, and "no-Alice" and "yes-Bob" are also born into the same post-measurement world. These spatially-separated birthings are correlated in a way that seems impossible to understand in any purely local way.

That said, I do not think it is possible to really argue cleanly, the way we have done for both the pilot-wave and spontaneous collapse theories, that there is an unambiguous violation of Bell's formulation of local causality. This is partly because that formulation is built around the concept of probability (it demands, remember, that a certain probability not change when distant events are specified) and the question about how to understand the usual quantum mechanical probabilities in the context of Everett's many worlds theory remains rather murky. It is also in part a result of the murkiness of the ontology posited by the theory. Thinking in terms of the mass density ontology at least gives us something reasonably clear that we can draw pictures of and think about (even though it is perhaps somewhat contrary to the Everettian spirit of insisting that the wave function *alone*, evolving in accordance with Schrödinger's equation *always*, is sufficient). But things still remain murky, with the intuitive non-locality somehow being associated with the relational facts which are not captured by the mass field.

And so the conclusion of this discussion will, unfortunately, but like our discussions of Probability and Ontology, be somewhat anti-climactic. It simply is not clear, in the context of Everett's version of quantum theory, how we should understand and

formulate the concept of "local causality", whether we should think of the theory as local or non-local, or, indeed, whether we should care about whether the theory is in some sense local or not. (Note that the closely related – but, as we saw in Chap. 9, not identical – question of the theory's compatibility with fundamental relativity also remains, I think, an open question.)

By way of bringing our discussion of the many-worlds theory to a close, I think it is helpful to acknowledge the truly shocking nature of the idea, in Bryce de Witt's description, that the

> ...universe is constantly splitting into a stupendous number of branches, all resulting from the measurementlike interactions between its myriad components. Moreover, every quantum transition taking place on every star, in every galaxy, in every remote corner of the universe is splitting our local world on earth into myriads of copies of itself [8].

As de Witt continues:

> 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. Here is schizophrenia with a vengeance [8].

I think it should be admitted, however, that although our intuitions recoil at this suggestion, the picture is compelling and elegant as an approach to addressing the measurement problem of ordinary quantum mechanics, and should be regarded (despite its initially shocking character!) as seemingly compatible with experience.

On the other hand, I think it must also be admitted that the theory does not yet provide sufficient clarity regarding the several issues we focused on in this chapter – probability, ontology, and locality – and that it therefore remains impossible to assess in anything like a final or conclusive way. It remains, to a greater extent than the pilot-wave theory or the spontaneous collapse theories, a work-in-progress.

Projects:

10.1 According to Everett's theory, if your friend measures the z-component of the spin of a spin-1/2 particle that is initially in the state ψ_{+x}, he gets into an entangled superposition (with the spin-1/2 particle and the measuring equipment) in which he experiences, in some sense, *both* outcomes: spin-up and spin-down. So, how will your friend respond if you ask him which outcome he experienced? Explain.

10.2 True or false: according to Everett's theory, matter is made of particles. Explain.

10.3 Suppose a measurement of the energy of a particle-in-a-box produces the joint PIB-pointer state

$$\psi(x, y) = \frac{1}{\sqrt{2}} [\psi_1(x)\phi_1(y) + \psi_2(x)\phi_2(y)] \tag{10.29}$$

where $\psi_n(x)$ is the nth energy eigenstate for the PIB and $\phi_n(y)$ is a pointer state that is sharply peaked around $y = Y_n$, the position that indicates the nth

outcome for the energy measurement. What is the state of the pointer *relative to* the PIB having state $\psi_1(x)$? What is the state of the PIB *relative to* the pointer having state $\phi_2(y)$? What is the state of the pointer *relative to* the PIB having state $\frac{1}{\sqrt{2}}[\psi_1(x) + \psi_2(x)]$?

10.4 It was claimed in Sect. 2 that, if the worlds are appropriately weighted in the counting, we can still say that the overwhelming number of worlds display outcome statistics that are compatible with the Born rule. Show in particular that, if Eqs. (10.9) and (10.13) are plugged into Eq. (10.14), the resulting function $f_N^w(n)$ does indeed peak at $n = Np$. (Hint: set d/dn of $\ln[f]$ to zero, and use the Stirling approximation, $\ln(m!) \approx m\ln(m)$.)

10.5 Consider the case of $N = 4$ "quantum coin flips", as discussed in Sect. 10.2, but with the branch weight for the H outcomes being $p = 3/4$. There is one world in which the sequence $HHHH$ is observed; its weighted world-fraction is therefore $f^w(4) = 1 \times (3/4)^4 \approx 0.316$. Calculate in a similar way $f^w(n)$ for $n = 0, 1, 2,$ and 3. Which value of n produces the largest weighted world-count? Is this what you would expect?

10.6 Suppose an observer measures some quantity (on a system that is not initially in an eigenstate for that quantity), and then subsequently re-measures the same quantity again, but using a different measuring apparatus. (Thus, there will be three degrees of freedom involved – say, "x" for the system whose property is being measured, "y" for the position of the pointer of the first measuring apparatus, and "z" for the position of the pointer of the second measuring apparatus.) Will the results of the two measurements agree in each branch of the wave function, or will there be some branches (i.e., some worlds) in which the measurements disagree? Explain how this relates to the collapse postulate of ordinary QM.

10.7 Suppose a spin-1/2 particle is in the state ψ_{+z}. First its z-spin is measured, then its x-spin is measured, and then its z-spin is measured again. Will the results of the two z-spin measurements agree in each branch of the wave function, or will there be some branches (i.e., some worlds) in which the measurements disagree? Explain how this relates to the collapse postulate of ordinary QM.

10.8 Imagine that you live in Everett's universe and are about to perform a biased quantum coin flip with $p = 3/4$. Explain what is problematic with each of the following statements you might consider making: (i) "The probability that I will see H is 3/4." (ii) "Of all my descendants, the probability that the one who is really *me* sees H is 3/4." (iii) "There will be descendants in branches with all possible outcomes, but the probability that I will end up *experiencing* a branch with outcome H is 3/4." Can you construct a similar statement, assigning a 3/4 probability to *something*, that would actually make sense in the Everettian point of view?

10.9 In Everett's 1957 paper, Ref. [3], he gives the following derivation of the branch weighting rule. The goal is to find a function w which assigns weights to the different terms in $\psi = \sum_i c_i \psi_i$. If the individual factors ψ_i are properly normalized, then the weight assigned to a given term can only depend on

the complex number c_i. But a pure phase can be absorbed into the states ψ_i, so that the weight function should only depend on the modulus of the expansion coefficients: $w(c_i) = w(|c_i|)$. Now, with further time-evolution, the nth branch will split into additional sub-branches: $c_n\psi_n = \sum_j a_j\phi_j$. Assuming again that all the states (ψ_n and the ϕ_j) are properly normalized, this implies $|c_n|^2 = \sum_j |a_j|^2$. Now, If we require that this further splitting preserves the total weight of the involved branches, we have

$$w(c_n) = \sum_j w(a_j). \tag{10.30}$$

Everett calls this the "additivity requirement". Using the above, it implies

$$w\left(\sqrt{\sum_j |a_j|^2}\right) = \sum_j w\left(|a_j|\right) \tag{10.31}$$

which implies that $w(x) = cx^2$ for some constant c that will be 1 if the total weight is normalized to 1. To see this, Everett suggests defining a new function $g(x) = w(\sqrt{x})$, in terms of which the previous equation reads:

$$g\left(\sum_j |a_j|^2\right) = \sum_j g\left(|a_j|^2\right). \tag{10.32}$$

One can see that this requires $g(x) = cx$. Fill in the gaps in the mathematics and reasoning here to make the argument fully clear. Then assess it. What, exactly, does it prove in the context of Everett's theory? Does the argument completely remove the circularity alluded to in the text?

10.10 In our preliminary discussion of the Einstein's Boxes scenario, depicted in Fig. 10.5, we said that "it perhaps appears that there is no hint of nonlocality here". Make this a little more formal by applying our modification of Bell's locality condition from Chap. 1, Eq. 1.28. Let χ_1 denote, say, the presence of a nonzero mass density, for the Left Pointer, at its undeflected position (after the measurement has gone to completion). And let χ_2 and χ_2' represent, respectively, nonzero mass densities, for the Right Pointer, at its undeflected and deflected positions. ($\mathcal{C}_\Sigma$ here just includes everything that was true prior to $t = 0$.) Show that the condition is formally respected.

10.11 The attitude of Everettians toward the issue of quantum non-locality is pretty-well captured by Everett's comments from 1957: "Consider the case where the states of two object systems are correlated, but where the two systems do not interact. Let one observer perform a specified observation on the first system, then let another observer perform an observation on the second system, and finally let the first observer repeat his observation. Then it is found that the first observer always [i.e., in each branch] gets the same result both times, and the

observation by the second observer has no effect whatsoever on the outcome of the first's observations. Fictitious paradoxes like that of Einstein, Podolsky, and Rosen which are concerned with such correlated, noninteracting systems are easily investigated and clarified in the present scheme." [3] Why do you think Everett calls the EPR paradox "fictitious"? Explain how you understand Everett to be thinking about this kind of situation. Do you agree that there is just clearly nothing non-local going on here, according to the Everett theory, such that it makes sense to call EPR's suggestion of non-locality "fictitious"? Explain.

10.12 Everettians, starting with David Deutsch, have accused the pilot-wave theory of being a "parallel universe theor[y] in a state of chronic denial [9]." The basis for this accusation is the fact that the pilot-wave theory also has the wave function of the universe obeying Schrödinger's equation all the time. So the many-worlds structure that Everettians find in that wave function is, they argue, just as present in that wave function in the pilot-wave picture, as it is in the wave-function monist Everettian picture. What do you make of this accusation? How do you think a proponent of the pilot-wave theory would or should respond?

10.13 Tim Maudlin pointed out in Ref. [10] that GRWm (but, interestingly, not GRWf) also has a kind of many-worlds character: since the GRW localizations involve multiplication by a Gaussian function which is small (but never quite zero) far from the Gaussian's center, the mass density field associated with the "un-selected" branches of the wave function is, while very small compared to the "selected" branch, not zero. What do you think? Is GRWm really a single-universe theory (because those other, "un-selected" worlds are so dim that it is reasonable to ignore them), or is it really a many-worlds theory in denial (because, dim or not, and anyway the dimness isn't visible from the inside, those "un-selected" worlds have all the right structure to count as real worlds)?

10.14 Proponents and critics of Everett's theory both sometimes appeal to Occam's razor in support of their position. The proponents say that, because the theory dispenses with the measurement axioms of ordinary QM (and because it doesn't replace those with anything like additional dynamical laws for "hidden variables"), Everettism is by far the simplest, most parsimonious version of quantum theory. On the other hand critics say that Everett's worldview, with the huge number of "parallel universes" that are totally unobservable to us, is ridiculously extravagant. Explain precisely how each side interprets and applies Occam's razor, i.e., explain what leads the two sides to these two opposite conclusions even though they are allegedly appealing to the same criterion. What do you think? Is Everett's theory clean and elegant, or ugly and complicated?

10.15 David Deutsch has argued that evidence for an Everettian multiplicity of universes is ubiquitous:

> The point that theorists tend to miss is that the multiplicity of reality is not only, or even primarily, a consequence of quantum *theory*. It is quite simply an observed fact. Any interference experiment (such as the two-slit experiment), when performed with individual particles one at a time, has no known interpretation in which the particle we see is the only physical entity passing through the apparatus. We know that the invisible entities passing through obey the same phenomenological equations of motion ... as the single particle we do see. And we know from [EPR] type experiments, such as that of Aspect, that these not-directly-perceptible particles are arranged in extended 'layers' each of which behaves internally like an approximately classical universe. Admittedly all these observations detect other universes only indirectly. But then we can detect pterodactyls and quarks only indirectly too. The evidence that other universes exist is at least as strong as the evidence for pterodactyls or quarks [9].

What do you think of this argument? Is there really no single-universe theory that can explain the results of the double-slit experiment?

10.16 In the last section of Chap. 9, we saw that certain regions of the $\{\lambda, \sigma\}$ parameter space for spontaneous collapse theories were empirically refuted, and certain other regions were considered "Perceptually/Philosophically Unsatisfactory". The many-worlds theory can be thought of as a spontaneous collapse theory, but with collapse rate $\lambda = 0$. So is the many-worlds theory "Perceptually/Philosophically Unsatisfactory"? Explain the assumption that is made in diagnosing small-λ versions of spontaneous collapse theory as unsatisfactory, and how Everett would challenge this assumption.

References

1. M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974)
2. B. Peter, Everett and wheeler, the untold story, in *Many Worlds? Everett, Quantum Theory, and Reality*, ed. by S. Saunders, J. Barrett, A. Kent, D. Wallace (Oxford University Press, Oxford, 2010)
3. H. Everett, [The] 'Relative State' formulation of quantum mechanics. Rev. Mod. Phys. **29**(3), 454–462 (1957)
4. B.S. de Witt, N. Graham, *The Many-Worlds Interpretation of Quantum Mechanics* (Princeton University Press, Princeton, 1973)
5. S. Saunders, Many worlds? An introduction in *Many Worlds? Everett, Quantum Theory, and Reality*, ed. by S. Saunders, J. Barrett, A. Kent, D. Wallace (Oxford University Press, Oxford, 2010)
6. S. Goldstein, V. Allori, R. Tumulka, N. Zanghi, Many-worlds and Schrödinger's first quantum theory. Br. J. Philos. Sci. **62**(1), 1–27 (2011), arXiv:0903.2211
7. D. Wallace, Decoherence and ontology in *Many Worlds? Everett, Quantum Theory, and Reality*, ed. by S. Saunders, J. Barrett, A. Kent, D. Wallace (Oxford University Press, Oxford, 2010)
8. B. de Witt, Quantum mechanics and reality. Phys. Today **23**(9), 30–35 (1970)
9. D. Deutsch, Comment on lockwood. Br. J. Philos. Sci. **47**, 222–8 (1996)
10. T. Maudlin, Can the world be only wavefunction? in *Many Worlds? Everett, Quantum Theory, and Reality*, ed. by S. Saunders, J. Barrett, A. Kent, D. Wallace (Oxford University Press, Oxford, 2010)

Afterword

In December of 1923, shortly after completing his PhD in the physics department at Harvard, the promising young Illinois native John Slater traveled to Copenhagen to conduct post-doctoral research with Niels Bohr and Bohr's close associate Hans Kramers. Slater had read Bohr's early papers on atomic models and been impressed by them: "I liked the way in which [Bohr] would go straight to the physical side of things instead of wrapping it up in a great deal of mathematics," Slater would explain much later in an archival interview conducted by Thomas Kuhn. "I felt that he must understand the physics of it quite well [1]."

Slater was particularly interested in the interaction of matter and radiation and was particularly struck by the mounting experimental evidence for Einstein's light quantum hypothesis, i.e., the existence of photons. Upon arriving in Copenhagen, Slater was contemplating a pilot-wave type model, in which real photon particles would be guided by a wave obeying Maxwell's equations: "I wanted to see how definite one could get in tying together the fields and the photons [1]."

But Bohr and Kramers turned out to be quite hostile to the majority of – and certainly the spirit of – Slater's ideas. The three ended up co-authoring a paper (which was influential despite turning out to be completely wrong) that jumped off from some of Slater's ideas about matter-radiation interaction, but contradicted, for example, Slater's picture of real, "definite" photons, and Slater's principle of strict energy conservation. As Slater would later admit, though, "those papers were dictated by Bohr and Kramers very much against my wishes. I fought with them so seriously that I've never had any respect for those people since. I had a horrible time in Copenhagen [1]."

As Slater elaborated:

> I went there. Bohr was very nice, he invited me to Christmas dinner, I told him about my ideas, he felt these were fine, 'But, you see, they're much too definite. Now we cannot have this exact conservation. We must not think too specifically about the photons. We don't have photons like that.' In other words, [Bohr] wanted to make the whole thing just as vague as he could [1].

Undoubtedly this was neither the first nor the last time that a relatively young scientist has felt pressured – even railroaded – by more senior colleagues. But what I find

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4

particularly interesting about this episode is not what felt to Slater like rude or even professionally inappropriate behavior on the part of Bohr and Kramers, but instead the philosophical disconnect that seemingly prevented them from agreeing about how best to proceed with the physics.

Slater, for his part, reported that he was not dogmatically attached to the idea of definite photons:

> ... I was willing to knock the photons.... but I wanted to put these ideas together and get something more definite about it. That was when I was working on [a paper which ended up with the title] 'A quantum theory of optical phenomena'. I wanted to start working on that. I always approach a problem in the sense of wanting to be able to make it definite and work out the details. I feel that if you can't work out the details you can't be sure it's right. I have a great distrust of the hand-waving approach to anything. I had supposed, when I went to Copenhagen, that although Bohr's papers looked like hand-waving, they were just covering up all the mathematics and careful thought that had gone on underneath.
>
> The thing I convinced myself of after a month, was that there was nothing underneath. It was all just hand-waving. I just said 'I'm not going to content myself with this. I'm going to go ahead and see if I can't work out a physical picture ... which at least would show that these ideas can be made to hang together logically.
>
> So I tried to see if any set of hypotheses could be hung together that would be somewhat logical. I think that the final result was that I could do this. Well, I was working on that. Bohr was contemptuous of it. He would have nothing whatever to do with it. He said he wasn't interested in looking at it or anything like it. Same way with Kramers. They just had no use whatever for this. So I decided I had no use whatever for being around them, [and] I went away. And I've never had any respect for Mr. Bohr since [1].

Finally, in a statement that I decided to open this book with because I regard it as perfectly and beautifully capturing the attitude that sensible physicists should have to the Copenhagen philosophy, Slater noted:

> Bohr always would go in for this remark, 'You cannot really explain it in the framework of space and time.' By God, I was determined I was going to explain it in the framework of space and time.
>
> In other words, that was Bohr's point of view on everything, and that was the fundamental difference of opinion between us.... Bohr was fundamentally of a mystical turn of mind and I'm fundamentally of a matter-of-fact turn of mind [1].

Despite – or perhaps in part because of – his frustrations with Bohr and his early departure from Copenhagen, Slater went on to have a long and distinguished career, making major contributions across many decades in atomic, molecular, and solid-state physics.

I find Slater's story at once tragic and refreshing – tragic in the sense that unquestionably progress in fundamental physics was stifled by the Copenhagen philosophy, which surely turned off, in addition to Slater, other promising researchers with "matter-of-fact turn[s] of mind" – but also refreshing to know that Schrödinger and Einstein were not the *only* voices of dissent against the rising tide of Copenhagen orthodoxy.

That tide did rise. As the Nobel prize-winner Murray Gell-Mann would note: "Niels Bohr brainwashed a whole generation of theorists into thinking that the job

(of finding an adequate presentation of quantum mechanics) was done fifty years ago [2]."

But happily that tide is now in the process of receding. Many physicists continue to pay superficial lip-service to the Copenhagen interpretation, but this is really just a result of inertia. Very few physicists (and even fewer philosophers of physics) who are actually interested in foundational questions in quantum mechanics – and, thankfully, this is an increasing number due to the recent explosion of interest in novel technological applications of fundamental quantum mechanical principles – take the Copenhagen approach seriously. It is increasingly recognized as, in essence, what Slater pegged it as from the very beginning: a lot of vague and philosophical hand-waving intended to paper over the failure to "explain [things] in the framework of space and time [1]."

I hope that this book has helped encourage readers to adopt Slater's attitude. In particular, I hope that an increased familiarity with the pilot-wave, spontaneous collapse, and many-worlds pictures will give today's promising young physicists a clearer sense of what is possible so that they feel confident in demanding more – more, that is, than Bohr thought possible – from microphysics going forward.

Regarding this menu of currently-available options for understanding quantum theory, I have to agree with Bell's assessment that "the pilot wave picture undoubtedly shows the best craftsmanship among the pictures we have considered [3]." Its apparent incompatibility with relativity is indeed troubling. However, Bell's theorem and the associated experiments seem to reveal that the pilot-wave theory's non-locality is a reflection of a genuine feature of the world, rather than a disqualifying flaw. On the other hand, there are hints that the required sort of non-locality could be more plausibly unified with fundamental relativity in the context of spontaneous collapse theories. And some still hold out hope that the many-worlds theory will show that even non-locality is not required. But still, despite the difficulty with relativity, it is hard not to view the pilot-wave theory as the most plausible and promising currently-available option, compared to the relatively contrived and *ad hoc* spontaneous collapse theories and the many-worlds theory with its sprawling difficulties with probability and ontology.

My first major hope for this book, though, is not to convert people into proponents of the pilot-wave theory; instead my hope is simply to help people understand that the common aspiration of the pilot-wave, spontaneous collapse, and many-worlds theories – namely, to give a uniform and coherent account of physics from the microscopic to the macroscopic scales – is attainable.

Bell's theorem – the conflict between any such coherent account and the relativistic notion of local causality – remains very troubling. My second major hope for this book is thus to help people appreciate that, while it is not terribly difficult to achieve a realistic picture of the quantum realm, it *is* rather difficult to unify these pictures with a similarly-realistic understanding of relativity theory. It is tragic that, more than 50 years after Bell's discovery, so many physicists remain unaware of this major challenge. One strongly suspects that several more widely-known theoretical challenges (such as unifying general relativity with relativistic quantum field theory)

might be related to this largely-unrecognized one. More attention and more creativity are needed on this and many other related issues.

I hope the book has established a fertile base from which to further explore foundational questions in quantum physics. To help you get started, I thought I could close the book by suggesting some possible directions and topics for further study and research.

One obvious possibility is to learn more about one of the theories that was introduced in this book:

- For the pilot-wave theory, there are interesting questions about how the "quantum equilibrium hypothesis" – which ensures consistency with the Born rule statistics of ordinary QM – should be undestood, derived, and explained. Important entries into the literature include Refs. [4–6]. There has also been interesting work on relativistic extensions of the pilot-wave picture; see, for example, Refs. [7–9]. Finally there are interesting questions, related to the discussions in our Chap. 5, about how to understand the wave function in the context of the pilot-wave theory. References [10, 11] provide two contrasting perspectives.
- Important recent work on the spontaneous collapse theories includes the relativistic formulations of GRWf [12] and GRWm [13], and a precise formalization of the phenomenological implications of GRW [14]. Reference [15] provides a recent and systematic review of collapse models and the ongoing attempts to constrain them experimentally. On the more philosophical side, David Albert's book – which argues that spontaneous collapse theories may shed novel light on problems in the foundations of statistical mechanics, thermodynamics, and cosmology – is also noteworthy [16].
- The recent resurgence of popularity in Everett's many-worlds theory, by both physicists (especially cosmologists) and philosophers of science, has produced a broad literature. Everettians' recent attempts to address the problem of understanding and explaining the Born rule quantum probabilities are especially important. David Wallace's essay "How to prove the Born rule" is particularly noteworthy in a collection of very high-quality papers about the physics and philosophy of Everett's theory, Ref. [17]. (References [18, 19] also provide useful overviews.) With the possible exception of Ref. [20], Ref. [21] seems to remain the state-of-the-art when it comes to the relationship between Everett's theory and the ontology problem. On the more pure physics side, Max Tegmark has reviewed the Everett interpretation and its relationship to other, especially cosmological, notions of many- or parallel- universes [22].

Debates in the foundations of quantum mechanics sometimes appear as endless partisan squabbling between dogmatically-committed proponents of different viewpoints, none of whose minds are ever changed. Still, in addition to learning how proponents of different theories describe and develop their favored perspectives, it can be valuable to understand the criticisms that proponents of one theory level against its rivals. Noteworthy examples here include an Everettian critique of the pilot-wave theory by Brown and Wallace [23], replies (defending the pilot-wave perspective) by Maudlin and Valentini (and Brown's reply to Valentini's reply) in Ref. [17]. J. Bricmont, a

defender of the pilot-wave theory, has made a number of sharp criticisms of the spontaneous collapse and many-worlds theories in Ref. [24].

There also exist, of course, many other "interpretations" of quantum mechanics that were (undoubtedly to the great annoyance of their defenders) not included in this book. That is, of course, because I view them as less worthwhile. But there may be value in exploring them nevertheless. Reference [25] provides an accessible introduction, by a critic, to the "consistent histories" interpretation (which can be understood as a more formal and modernized version of the Copenhagen interpretation) in relation to two of the theories included in this book; some accessible presentations by one of the theory's defenders are collected in Ref. [26]. A rather different attempt to modernize the Copenhagen interpretation is so-called "QBism" (which at least at one point stood for Quantum Bayesianism); see Ref. [27] for an accessible introduction. A completely different idea is that virtually everything that is puzzling or problematic about quantum mechanics disappears if only we allow for the possibility of backwards-in-time causal influences; see Ref. [28] for a recent overview and Ref. [29] for an older but more formally developed proposal. There are of course many others as well, but these tend to be closely related to one of the perspectives already mentioned, or to have a very dubious status, or both. The Wikipedia page on "Interpretations of quantum mechanics" seems to contain a fairly complete list.

Some important recent developments have been spurred by a viewpoint which is not exactly a candidate theory, but more an open-ended category of possible theories: the "ψ-epistemic" viewpoint according to which the quantum wave function should be understood as describing our incomplete knowledge of physical states, rather than the physical states themselves. (Something in this vicinity was, of course, Einstein's view and there is significant overlap with the idea of "hidden variables"; note, though, that, despite its status as a so-called hidden variable theory, the pilot-wave theory is "ψ-ontic" rather than "ψ-epistemic" since it claims that a complete physical state description includes ψ.) See Ref. [30] for a somewhat-dated but excellent review, and Ref. [31] for the more recent and highly thought-provoking paper that stimulated an important subsequent proof, in Ref. [32], that the ψ-epistemic viewpoint is actually in conflict with the statistical predictions of quantum mechanics so that, if those statistical predictions are right, the quantum state ψ must be (in a certain sense) part of the ontology of any viable theory. This so-called PBR theorem (the initials of its authors) has generated significant further debate and elaboration.

Topics that were raised in the present book, but in a relatively simplified or superficial way, are also good ones for further independent reading and research. One important example here is the "no hidden variables" theorems mentioned in Chap. 3. Crucial references here include Bell's classic paper [33] and Mermin's review article [34]. References [35, 36] are also of interest.

Another example is the concept of "decoherence" which, despite only being mentioned explicitly in Chap. 10, plays a crucial role in all of the quantum theories we've explored. Indeed, there are many physicists who believe that decoherence alone (i.e., without the need to bring in hidden variables, spontaneous collapses, or many-worlds) already solves the measurement problem. The recent article [37] and book [38] by

Max Schlosshauer would be good entry-points into the vast literature on this subject. (Continued study of the important background concept of "density matrices", in standard quantum mechanics texts, would also be helpful.) Ref. [39] is also of interest.

Finally, there is a near-infinite list of topics, not really touched on in this book but related to the foundations of quantum mechanics, on which there has been some interesting recent development. One important example is the concept of "weak measurement", which was initiated by the 1988 paper of Aharanov, Albert, and Vaidman with the intriguing title "How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100". [40] See References [41], (especially) [42, 43] to follow a particularly interesting thread that relates closely to the pilot-wave theory. Reference [44] is another example (of many that could be mentioned) of an interesting application of the idea of "weak measurement".

Another even-more sprawling area with connections to the material we've covered is the field of "quantum information", which includes such concepts as "quantum cryptography" and "quantum computing". The textbook by Nielsen and Chuang is an excellent place to start learning about this broad and highly-active area. [45]

References

1. Oral history interview transcript with John C. Slater, October 3, 1963, American Institute of Physics, Niels Bohr Library and Archives, https://www.aip.org/history-programs/niels-bohr-library/oral-histories/4892-1
2. M. Gell-Mann, in D. Huff, O. Prewett, *The Nature of the Physical Universe*, 1976 Nobel Conference, New York, (1979), p. 29. Also cited by K. R. Popper, Quantum Theory and the Schism in Physics, (1982), P.10
3. J.S. Bell, Six possible worlds of quantum mechanics. Found. Phys. **22**(10), 1201–1215 (1992)
4. D. Dürr, S. Goldstein, N. Zanghì, Quantum equilibrium and the origin of absolute uncertainty. J. Stat. Phys. **67**, 843–907 (1992), arxiv:aquant-ph/0308039
5. A. Valentini, Signal-locality, uncertainty, and the subquantum H-theorem, Parts I and II, Phys. Lett. A, **156**,1–2, (3 June 1991), 5–11,1–2, (19 August 1991), 1–8
6. A. Valentini, H. Westman, Dynamical origin of quantum probabilities. Proc. R. Soc. A **461**(8), (January, 2005), arxiv:quant-ph/0403034
7. D. Dürr, S. Goldstein, S. Teufel, N. Zanghì, Hypersurface bohm-dirac models. Phys. Rev. A **60**, 2729–2736 (1999), arxiv:quant-ph/9801070
8. D. Dürr, S. Goldstein, R. Tumulka, N. Zanghì, Bell-Type Quantum Field Theories. J. Phys. A: Math. Gen. **38**, R1–R43 (2005), arxiv:quant-ph/0407116
9. D. Dürr, S. Goldstein, T. Norsen, W. Sturyve, N. Zanghì, Can Bohmian mechanics be made relativistic? Proc. R. Soc. A **470**. doi:10.1098/rspa.2013.0699 (2013), (arXiv:1307.1714)
10. S. Goldstein, N. Zanghi, Reality and the role of the wavefunction in quantum theory, in *The Wave Function: Essays in the Metaphysics of Quantum Mechanics*, ed. by D. Albert, A. Ney (Oxford, 2012)
11. T. Norsen, D. Marian, X. Oriols, Can the wave function in configuration space be replaced by single-particle wave functions in physical space? Synthese **192**(10), 3125–3151, (arXiv:1410.3676)
12. R. Tumulka, A relativistic version of the Ghirardi-Rimini-Weber model. J. Stat. Phys. **125**, 821–840, arxiv:quant-ph/0406094

13. D. Bedingham et al., Matter density and relativistic models of wave function collapse. J. Stat. Phys. **154**, 623–631 (2014)
14. S. Goldstein, R. Tumulka, N. Zanghì, The quantum formalism and the GRW formalism. J. Stat. Phys. **149**, 142–201 (2012), (arXiv:0710-0885)
15. A. Bassi et al., Models of wave-function collapse, underlying theories, and experimental tests, Rev. Mod. Phys. **85**, 471 (2 April 2013)
16. D. Albert, *Time and Chance* (Harvard University Press, Cambridge, 2000)
17. S. Saunders, J. Barrett, A. Kent, D. Wallace (eds.), *Many Worlds? Everett, Quantum Theory, and Reality* (Oxford, New York, 2010)
18. L. Vaidman, Many-worlds interpretation of quantum mechanics, Stanf. Encycl. Philos., https://plato.stanford.edu/entries/qm-manyworlds/
19. D. Wallace, *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation* (Oxford, New York, 2012)
20. D. Wallace, C.J. Timpson, Quantum mechanics on spacetime I: spacetime state realism. Brit. J. Phil. Sci. **61**, 697–727 (2010)
21. V. Allori et al., Many worlds and schrödinger's first quantum theory. Brit. J. Philos. Sci. **62**, 1–27 (2011), (arXiv:0903.2211)
22. M. Tegmark, Parallel Universes in *Science and Ultimate Reality: From Quantum to Cosmos*, honoring John Wheeler's 90th birthday, J.D. Barrow, P.C.W. Daview, C.L. Harper, eds., Cambridge (2003) (astro-ph/0302131) See also arXiv:0905.1283 and arXiv:0905.2182 and arXiv:1008.1066
23. H.R. Brown, D. Wallace, Solving the measurement problem: de broglie - bohm loses out to everett. Found. Phys. **35**, 517 (2005). arxiv:quant-ph/0403094
24. J. Bricmont, *Making Sense of Quantum Mechanics* (Springer, Berlin, 2016)
25. S. Goldstein, Quantum theory without observers, Phys. Today, 42–46, (March 1998), 38–42 (April 1998)
26. R. Griffiths, *Consistent Quantum Theory* (Cambridge University Press, UK, 2002). See also arXiv:1501.04813, arXiv:1304.4425, arXiv:1105.3932, and arXiv:1007.4281
27. C. Fuchs, N.D. Mermin, R. Schack, An introduction to QBism with application to the locality of quantum mechanics. Am. J. Phys. **82**, 749 (2014), (arXiv:1311.5253)
28. H. Price, K. Wharton, Disentangling the quantum world. Entropy **17**, 7752–7767 (2015), (arXiv:1508.01140)
29. J. Cramer, The transactional interpretation of quantum mechanics. Rev. Mod. Phys. **58**, 647–688 (1986)
30. L. Ballentine, The statistical interpretation fo quantum mechanics. Rev. Mod. Phys. **42**, 358 (1970)
31. R. Spekkens, In defense of the epistemic view of quantum states: a toy theory. Phys. Rev. A **75**, 032110 (2007), arxiv:quant-ph/0401052
32. M. Pusey, J. Barrett, T. Rudolph, On the reality of the quantum state. Nature Phys. **8**, 475–478 (2012), (arXiv:1111.3328)
33. J.S. Bell, On the problem of hidden variables in quantum mechanics. Rev. Mod. Phys. **38**, 447 (1966)
34. N.D. Mermin, Hidden variables and the two theorems of John Bell. Rev. Mod. Phys. **65**, 803 (1993)
35. D. Greenberger, M. Horne, A. Shimony, A. Zeilinger, Bell's theorem without inequalities. Am. J. Phys. **58**, 1131 (1990)
36. M. Waegell, P.K. Aravind, Proofs of the Kochen-Specker theorem based on a system of three qubits. J. Phys. A: Math. Theor. **45**(40), (2012), (arXiv:1205.5015)
37. M. Schlosshauer, Decoherence, the measurement problem, and interpretations of quantum mechanics. Rev. Mod. Phys. **76**, 1267 (2005), arxiv:quant-ph/0312059
38. *Decoherence and the Quantum-to-Classical Transition* (Springer, Berlin, 2007)
39. S. Adler, Why decoherence has not solved the measurement problem: A response to P.W. Anderson. Stud. Hist. Phil. Sci. B: Stud. ist. Phil. Mod. Phys. **34**(1), 135–142 (March 2003), arxiv:quant-ph/0112095

40. Y. Aharanov, D.Z. Albert, L. Vaidman, How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. Phys. Rev. Lett. **60**, 1351 (1988)
41. H. Wiseman, Grounding Bohmian mechanics in weak values and bayesianism. New J. Phys. **9**, 165 (2007)
42. S. Kocsis, B. Braverman, S. Braverman, S. Ravets, M.J. Stevens, R. Mirin, L.K. Shalm, A. Steinberg, Observing the average trajectories of single photons in a two-slit interferometer. Science **332**(6034), 1170–1173 (03 Jun 2011)
43. D. Mahler, L. Rozema, K. Fisher, L. Vermeyden, K. Resch, H. Wiseman, A. Steinberg, Experimental nonlocal and surreal Bohmian trajectories. Sci. Adv. **2**(2), 19 (2016)
44. T. Denkmayr et al., Observation of a quantum cheshire cat in a matter-wave interferometer experiment. Nature commun. **5**, 4492 (2014)
45. M. Nielsen, I. Chuang, *Quantum Computation and Quantum Information*, 10 Anniversary edn. (Cambridge, UK, 2011)