Introduction to Quantum Mechanics

David J. Griffiths
Reed College

Prentice Hall
Upper Saddle River, New Jersey 07458

PREFACE

Unlike Newton's mechanics, or Maxwell's electrodynamics, or Einstein's relativity, quantum theory was not created—or even definitively packaged—by one individual, and it retains to this day some of the scars of its exhilirating but traumatic youth. There is no general consensus as to what its fundamental principles are, how it should be taught, or what it really "means." Every competent physicist can "do" quantum mechanics, but the stories we tell ourselves about what we are doing are as various as the tales of Scheherazade, and almost as implausible. Richard Feynman (one of its greatest practitioners) remarked, "I think I can safely say that nobody understands quantum mechanics."

The purpose of this book is to teach you how to *do* quantum mechanics. Apart from some essential background in Chapter 1, the deeper quasi-philosophical questions are saved for the end. I do not believe one can intelligently discuss what quantum mechanics *means* until one has a firm sense of what quantum mechanics *does*. But if you absolutely cannot wait, by all means read the Afterword immediately following Chapter 1.

Not only is quantum theory conceptually rich, it is also technically difficult, and exact solutions to all but the most artificial textbook examples are few and far between. It is therefore essential to develop special techniques for attacking more realistic problems. Accordingly, this book is divided into two parts[1]; Part I covers the basic theory, and Part II assembles an arsenal of approximation schemes, with illustrative applications. Although it is important to keep the two parts logically separate, it is not necessary to study the material in the order presented here. Some instructors, for example, may wish to treat time-independent perturbation theory immediately after Chapter 2.

[1]This structure was inspired by David Park's classic text *Introduction to the Quantum Theory*, 3rd ed., (New York: McGraw-Hill, 1992).

This book is intended for a one-semester or one-year course at the junior or senior level. A one-semester course will have to concentrate mainly on Part I; a full-year course should have room for supplementary material beyond Part II. The reader must be familiar with the rudiments of linear algebra, complex numbers, and calculus up through partial derivatives; some acquaintance with Fourier analysis and the Dirac delta function would help. Elementary classical mechanics is essential, of course, and a little electrodynamics would be useful in places. As always, the more physics and math you know the easier it will be, and the more you will get out of your study. But I would like to emphasize that quantum mechanics is not, in my view, something that flows smoothly and naturally from earlier theories. On the contrary, it represents an abrupt and revolutionary departure from classical ideas, calling forth a wholly new and radically counterintuitive way of thinking about the world. That, indeed, is what makes it such a fascinating subject.

At first glance, this book may strike you as forbiddingly mathematical. We encounter Legendre, Hermite, and Laguerre polynomials, spherical harmonics, Bessel, Neumann, and Hankel functions, Airy functions, and even the Riemann Zeta function —not to mention Fourier transforms, Hilbert spaces, Hermitian operators, Clebsch-Gordan coefficients, and Lagrange multipliers. Is all this baggage really necessary? Perhaps not, but physics is like carpentry: Using the right tool makes the job *easier*, not more difficult, and teaching quantum mechanics without the appropriate mathematical equipment is like asking the student to dig a foundation with a screwdriver. (On the other hand, it can be tedious and diverting if the instructor feels obliged to give elaborate lessons on the proper use of each tool. My own instinct is to hand the students shovels and tell them to start digging. They may develop blisters at first, but I still think this is the most efficient and exciting way to learn.) At any rate, I can assure you that there is no *deep* mathematics in this book, and if you run into something unfamiliar, and you don't find my explanation adequate, by all means *ask* someone about it, or look it up. There are many good books on mathematical methods—I particularly recommend Mary Boas, *Mathematical Methods in the Physical Sciences*, 2nd ed., Wiley, New York (1983), and George Arfken, *Mathematical Methods for Physicists*, 3rd ed., Academic Press, Orlando (1985). But whatever you do, don't let the mathematics—which, for us, is only a tool—interfere with the physics.

Several readers have noted that there are fewer worked examples in this book than is customary, and that some important material is relegated to the problems. This is no accident. I don't believe you can learn quantum mechanics without doing many exercises for yourself. Instructors should, of course, go over as many problems in class as time allows, but students should be warned that this is not a subject about which *any*one has natural intuitions—you're developing a whole new set of muscles here, and there is simply no substitute for calisthenics. Mark Semon suggested that I offer a "Michelin Guide" to the problems, with varying numbers of stars to indicate the level of difficulty and importance. This seemed like a good idea (though, like the quality of a restaurant, the significance of a problem is partly a matter of taste); I have adopted the following rating scheme:

* ∗ an essential problem that every reader should study;
* ∗∗ a somewhat more difficult or more peripheral problem;
* ∗ ∗ ∗ an unusually challenging problem that may take over an hour.

(No stars at all means fast food: OK if you're hungry, but not very nourishing.) Most of the one-star problems appear at the end of the relevant section; most of the three-star problems are at the end of the chapter. A solution manual is available (to instructors only) from the publisher.

I have benefited from the comments and advice of many colleagues, who suggested problems, read early drafts, or used a preliminary version in their courses. I would like to thank in particular Burt Brody (Bard College), Ash Carter (Drew University), Peter Collings (Swarthmore College), Jeff Dunham (Middlebury College), Greg Elliott (University of Puget Sound), Larry Hunter (Amherst College), Mark Semon (Bates College), Stavros Theodorakis (University of Cyprus), Dan Velleman (Amherst College), and all my colleagues at Reed College.

Finally, I wish to thank David Park and John Rasmussen (and their publishers) for permission to reproduce Figure 8.6, which is taken from Park's *Introduction to the Quantum Theory* (footnote 1), adapted from I. Perlman and J. O. Rasmussen, "Alpha Radioactivity," in *Encyclopedia of Physics*, vol. **42**, Springer-Verlag, 1957.

PART I

THEORY

CHAPTER 1

THE WAVE FUNCTION

1.1 THE SCHRÖDINGER EQUATION

Imagine a particle of mass m, constrained to move along the x-axis, subject to some specified force $F(x, t)$ (Figure 1.1). The program of *classical* mechanics is to determine the position of the particle at any given time: $x(t)$. Once we know that, we can figure out the velocity ($v = dx/dt$), the momentum ($p = mv$), the kinetic energy ($T = (1/2)mv^2$), or any other dynamical variable of interest. And how do we go about determining $x(t)$? We apply Newton's second law: $F = ma$. (For conservative systems—the only kind we shall consider, and, fortunately, the only kind that *occur* at the microscopic level—the force can be expressed as the derivative of a potential energy function,[1] $F = -\partial V/\partial x$, and Newton's law reads $m\, d^2x/dt^2 = -\partial V/\partial x$.) This, together with appropriate initial conditions (typically the position and velocity at $t = 0$), determines $x(t)$.

Quantum mechanics approaches this same problem quite differently. In this case what we're looking for is the **wave function**, $\Psi(x, t)$, of the particle, and we get it by solving the **Schrödinger equation**:

$$\boxed{i\hbar\frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2 \Psi}{\partial x^2} + V\Psi.} \qquad [1.1]$$

[1]Magnetic forces are an exception, but let's not worry about them just yet. By the way, we shall assume throughout this book that the motion is nonrelativistic ($v \ll c$).

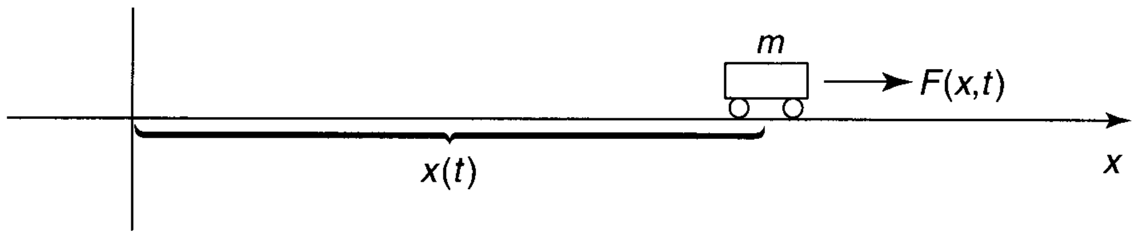

Figure 1.1: A "particle" constrained to move in one dimension under the influence of a specified force.

Here i is the square root of -1, and $\hbar$ is Planck's constant—or rather, his original constant (h) divided by 2π:

$$\hbar = \frac{h}{2\pi} = 1.054573 \times 10^{-34}\text{J s}. \qquad [1.2]$$

The Schrödinger equation plays a role logically analogous to Newton's second law: Given suitable initial conditions [typically, $\Psi(x, 0)$], the Schrödinger equation determines $\Psi(x, t)$ for all future time, just as, in classical mechanics, Newton's law determines $x(t)$ for all future time.

1.2 THE STATISTICAL INTERPRETATION

But what exactly *is* this "wave function", and what does it do for you once you've *got* it? After all, a particle, by its nature, is localized at a point, whereas the wave function (as its name suggests) is spread out in space (it's a function of x, for any given time t). How can such an object be said to describe the state of a *particle*? The answer is provided by Born's **statistical interpretation** of the wave function, which says that $|\Psi(x, t)|^2$ gives the *probability* of finding the particle at point x, at time t—or, more precisely,[2]

$$|\Psi(x,t)|^2\,dx = \left\{\begin{array}{l}\text{probability of finding the particle}\\ \text{between } x \text{ and } (x+dx),\ \text{at time } t.\end{array}\right\} \qquad [1.3]$$

For the wave function in Figure 1.2, you would be quite likely to find the particle in the vicinity of point A, and relatively unlikely to find it near point B.

The statistical interpretation introduces a kind of **indeterminacy** into quantum mechanics, for even if you know everything the theory has to tell you about the

[2]The wave function itself is complex, but $|\Psi|^2 = \Psi^*\Psi$ (where Ψ^* is the complex conjugate of Ψ) is real and nonnegative—as a probability, of course, must be.

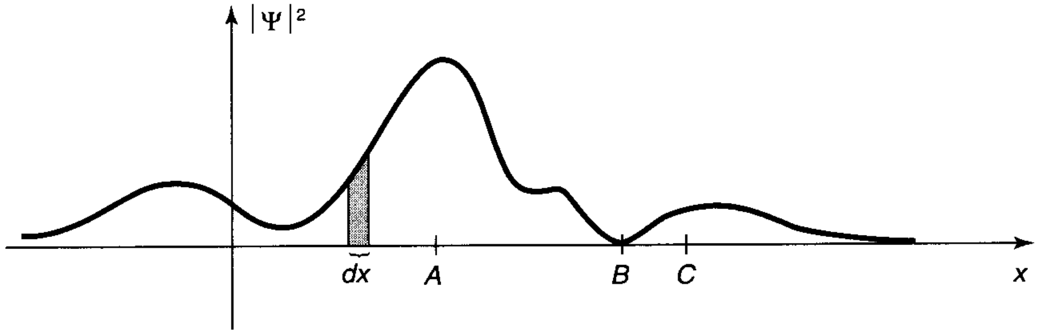

Figure 1.2: A typical wave function. The particle would be relatively likely to be found near A, and unlikely to be found near B. The shaded area represents the probability of finding the particle in the range dx.

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 the 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. As d'Espagnat put it, "the position of the particle was never indeterminate, but was merely unknown to the experimenter."[3] 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."[4] This view (the so-called **Copenhagen interpretation**) is associated with Bohr and his followers. Among physicists it has always been the

---

[3]Bernard d'Espagnat, *The Quantum Theory and Reality*, Scientific American, Nov. 1979 (Vol. 241), p. 165.

[4]Quoted in a lovely article by N. David Mermin, *Is the moon there when nobody looks?*, Physics Today, April 1985, p. 38.

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 perjorative 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."[5] 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.

Until fairly recently, all three positions (realist, orthodox, and agnostic) had their partisans. But in 1964 John Bell astonished the physics community by showing that it makes an *observable* difference if the particle had a precise (though unknown) position prior to the measurement. Bell's discovery effectively eliminated agnosticism as a viable option, and made it an *experimental* question whether 1 or 2 is the correct choice. I'll return to this story at the end of the book, when you will be in a better position to appreciate Bell's theorem; for now, suffice it to say that the experiments have confirmed decisively the orthodox interpretation[6]: A particle simply does not have a precise position prior to measurement, any more than the ripples on a pond do; it is the measurement process that insists on one particular number, and thereby in a sense *creates* the specific result, limited only by the statistical weighting imposed by the wave function.

But what if I made a *second* measurement, immediately after the first? Would I get C again, or does the act of measurement cough up some completely new number each time? On this question everyone is in agreement: A repeated measurement (on the same particle) must return the same value. Indeed, it would be tough to prove that the particle was really found at C in the first instance if this could not be confirmed by immediate repetition of the measurement. How does the orthodox interpretation account for the fact that the second measurement is bound to give the value C? Evidently the first measurement radically alters the wave function, so that it is now sharply peaked about C (Figure 1.3). We say that the wave function **collapses** upon measurement, to a spike at the point C (Ψ soon spreads out again, in accordance with the Schrödinger equation, so the second measurement must be made quickly). There

---

[5]Quoted by Mermin (previous footnote), p. 40.

[6]This statement is a little too strong: There remain a few theoretical and experimental loopholes, some of which I shall discuss in the Afterword. And there exist other formulations (such as the **many worlds** interpretation) that do not fit cleanly into any of my three categories. But I think it is wise, at least from a pedagogical point of view, to adopt a clear and coherent platform at this stage, and worry about the alternatives later.

Figure 1.3: Collapse of the wave function: graph of $|\Psi|^2$ immediately after a measurement has found the particle at point C.

are, then, two entirely distinct kinds of physical processes: "ordinary" ones, in which the wave function evolves in a leisurely fashion under the Schrödinger equation, and "measurements", in which Ψ suddenly and discontinuously collapses.[7]

1.3 PROBABILITY

Because of the statistical interpretation, **probability** plays a central role in quantum mechanics, so I digress now for a brief discussion of the theory of probability. It is mainly a question of introducing some notation and terminology, and I shall do it in the context of a simple example.

Imagine a room containing 14 people, whose ages are as follows:

one person aged 14
one person aged 15
three people aged 16
two people aged 22
two people aged 24
five people aged 25.

If we let $N(j)$ represent the number of people of age j, then

[7]The role of measurement in quantum mechanics is so critical and so bizarre that you may well be wondering what precisely *constitutes* a measurement. Does it have to do with the interaction between a microscopic (quantum) system and a macroscopic (classical) measuring apparatus (as Bohr insisted), or is it characterized by the leaving of a permanent "record" (as Heisenberg claimed), or does it involve the intervention of a conscious "observer" (as Wigner proposed)? I'll return to this thorny issue in the Afterword; for the moment let's take the naive view: A measurement is the kind of thing that a scientist does in the laboratory, with rulers, stopwatches, Geiger counters, and so on.

AFTERWORD

Now that you have (I hope) a sound understanding of what quantum mechanics *says*, I should like to return to the question of what it *means*—continuing the story begun in Section 1.2. The source of the problem is the indeterminacy associated with the statistical interpretation of the wave function. For Ψ (or, more generally, the *quantum state*—it could be a spinor, for example) does not uniquely determine the outcome of a measurement; all it provides is the statistical distribution of all possible results. This raises a profound question: Did the physical system "actually have" the attribute in question *prior* to the measurement (the so-called **realist** viewpoint), or did the act of measurement itself "create" the property, limited only by the statistical constraint imposed by the wave function (the **orthodox** position)—or can we duck the question entirely, on the grounds that it is "metaphysical" (the **agnostic** response)?

According to the realist, quantum mechanics is an *incomplete* theory, for even if you know *everything quantum mechanics has to tell you* about the system (to wit, its wave function), you still cannot determine all of its features. Evidently there is some *other* information, external to quantum mechanics, which (together with Ψ) is required for a complete description of physical reality.

The orthodox position raises even more disturbing problems, for if the act of measurement forces the system to "take a stand," helping to *create* an attribute that was not there previously,[1] then there is something very peculiar about the measurement process. Moreover, to account for the fact that an immediately repeated measurement yields the same result, we are forced to assume that the act of measurement **collapses**

[1]This may be *strange*, but it is not *mystical*, as some popularizers would like to suggest. The so-called **wave-particle duality**, which Niels Bohr elevated into a cosmic principle (**complementarity**), makes electrons sound like unpredictable adolescents, who sometimes behave like adults, and sometimes, for no particular reason, like children. I prefer to avoid such language. When I say that a particle does not have a particular attribute until a measurement intervenes, I have in mind, for example, an electron in the spin state $\chi = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; a measurement of the x-component of its angular momentum could return the value $\hbar/2$, or (with equal probability) the value $-\hbar/2$, but until the measurement is made it simply *does not have* a well-defined value of S_x.

the wave function, in a manner that is difficult, at best, to reconcile with the normal evolution prescribed by the Schrödinger equation.

In light of this, it is no wonder that generations of physicists retreated to the agnostic position, and advised their students not to waste time worrying about the conceptual foundations of the theory.

A.1 The EPR Paradox

In 1935, Einstein, Podolsky, and Rosen[2] published the famous **EPR paradox**, which was designed to prove (on purely theoretical grounds) that the realist position is the only sustainable one. I'll describe a simplified version of the EPR paradox, due to David Bohm. Consider the decay of the neutral pi meson into an electron and a positron:

$$\pi^0 \rightarrow e^- + e^+.$$

Assuming the pion was at rest, the electron and positron fly off in opposite directions (Figure A.1). Now, the pion has spin zero, so conservation of angular momentum requires that the electron and positron are in the singlet configuration:

$$\frac{1}{\sqrt{2}}(\uparrow_-\downarrow_+ - \downarrow_-\uparrow_+). \qquad [A.1]$$

If the electron is found to have spin up, the positron must have spin down, and vice versa. Quantum mechanics can't tell you *which* combination you'll get, in any particular pion decay, but it does say that the measurements will be *correlated*, and you'll get each combination half the time (on average). Now suppose we let the electron and positron fly *way* off—10 meters, in a practical experiment, or, in principle, 10 light years—and then you measure the spin of the electron. Say you get spin up. Immediately you know that someone 20 meters (or 20 light years) away will get spin down, if that person examines the positron.

To the realist, there's nothing surprising in this—the electron *really had* spin up (and the positron spin down) from the moment they were created—it's just that quantum mechanics didn't know about it. But the "orthodox" view holds that neither particle had either spin up *or* spin down until the act of measurement intervened: Your measurement of the electron collapsed the wave function, and instantaneously "produced" the spin of the positron 20 meters (or 20 light years) away. Einstein, Podolsky, and Rosen considered any such "spooky action-at-a-distance" (Einstein's words) preposterous. They concluded that the orthodox position is untenable; the

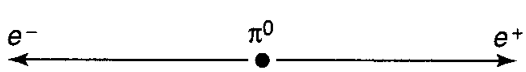

Figure A.1: Bohm's version of the EPR experiment: π^0 at rest decays into electron-positron pair.

[2]A. Einstein, B. Podolsky, and N. Rosen, *Phys. Rev.* **47**, 777 (1935).

electron and positron must have had well-defined spins all along, whether quantum mechanics can calculate them or not.

The fundamental assumption on which the EPR argument rests is that no influence can propagate faster than the speed of light. We call this the principle of **locality**. You might be tempted to propose that the collapse of the wave function is *not* instantaneous, but somehow "travels" out at some finite velocity. However, this would lead to violations of angular momentum conservation, for if we measured the spin of the positron before the news of the collapse had reached it, there would be a 50-50 probability of finding *both* particles with spin up. Whatever one might think of such a theory in the abstract, the experiments are unambiguous: No such violation occurs—the correlation of the spins is perfect.

A.2 Bell's Theorem

Einstein, Podolsky, and Rosen did not doubt that quantum mechanics is *correct*, as far as it goes; they only claimed that it is an *incomplete* discription of physical reality: The wave function is not the whole story—some *other* quantity, λ, is needed, in addition to Ψ, to characterize the state of a system fully. We call λ the "hidden variable" because, at this stage, we have no idea how to calculate or measure it.[3] Over the years, a number of hidden variable theories have been proposed, to supplement quantum mechanics; they tend to be cumbersome and implausible, but never mind—until 1964 the program seemed eminently worth pursuing. But in that year J. S. Bell proved that *any* local hidden variable theory is *incompatible* with quantum mechanics.[4]

Bell suggested a generalization of the EPR/Bohm experiment: Instead of orienting the electron and positron detectors along the *same* direction, he allowed them to be rotated independently. The first measures the component of the electron spin in the direction of a unit vector **a**, and the second measures the spin of the positron along the direction **b** (Figure A.2). For simplicity, let's record the spins in units of $\hbar/2$; then each detector registers the value $+1$ (for spin up) or -1 (spin down), along the direction in question. A table of results, for many π^0 decays, might look like this:

electron	positron	product
$+1$	-1	-1
$+1$	$+1$	$+1$
-1	$+1$	-1
$+1$	-1	-1
-1	-1	$+1$
$\vdots$	$\vdots$	$\vdots$

[3]The hidden variable could be a single number, or it could be a whole *collection* of numbers; perhaps λ is to be calculated in some future theory, or maybe it is for some reason of principle incalculable. It hardly matters. All I am asserting is that there must be *something*—if only a *list* of the outcomes of every possible experiment—associated with the system prior to a measurement.

[4]Bell's original paper [*Physics* **1**, 195 (1964)] is a gem: brief, accessible, and beautifully written.

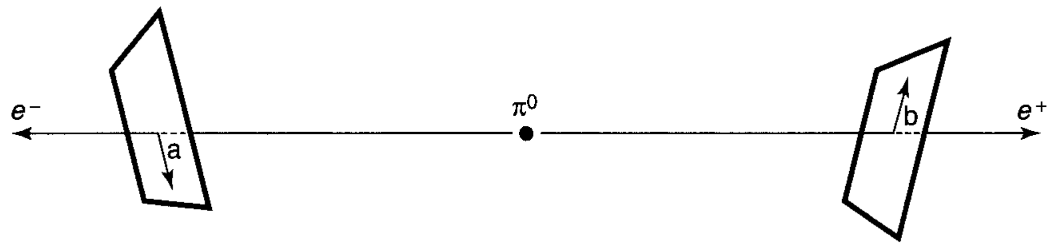

Figure A.2: Bell's version of the EPR-Bohm experiment: detectors independently oriented in directions **a** and **b**.

Bell proposed to calculate the *average* value of the *product* of the spins, for a given set of detector orientations. Call this average $P(\mathbf{a}, \mathbf{b})$. If the detectors are parallel ($\mathbf{b} = \mathbf{a}$), we recover the original EPRB configuration; in this case one is spin up and the other spin down, so the product is always -1, and hence so too is the average:

$$P(\mathbf{a}, \mathbf{a}) = -1. \qquad [A.2]$$

By the same token, if they are *anti*-parallel ($\mathbf{b} = -\mathbf{a}$), then every product is $+1$, so

$$P(\mathbf{a}, -\mathbf{a}) = +1. \qquad [A.3]$$

For arbitrary orientations, quantum mechanics predicts

$$\boxed{P(\mathbf{a}, \mathbf{b}) = -\mathbf{a} \cdot \mathbf{b}} \qquad [A.4]$$

(see Problem 4.44). What Bell discovered is that this result is impossible in any local hidden variable theory.

The argument is stunningly simple. Suppose that the "complete" state of the electron/positron system is characterized by the hidden variable(s) λ; λ varies, in some way that we neither understand nor control, from one pion decay to the next. Suppose further that the outcome of the *electron* measurement is independent of the orientation ($\mathbf{b}$) of the *positron* detector—which may, after all, be chosen by the experimenter at the positron end just before the electron measurement is made, and hence far too late for any subluminal message to get back to the electron detector. (This is the locality assumption.) Then there exists some function $A(\mathbf{a}, \lambda)$ which gives the result of an electron measurement, and some other function $B(\mathbf{b}, \lambda)$ for the positron measurement. These functions can only[5] take on the values ± 1:

$$A(\mathbf{a}, \lambda) = \pm 1; \quad B(\mathbf{b}, \lambda) = \pm 1. \qquad [A.5]$$

[5]This already concedes far more than a *classical* determinist would be prepared to allow, for it abandons any notion that the particles could have well-defined angular momentum vectors with simultaneously determinate components. But never mind—the point of Bell's argument is to demonstrate that quantum mechanics is incompatible with *any* local deterministic theory—even one that bends over backward to be accommodating.

When the detectors are aligned, the results are perfectly (anti)correlated:

$$A(\mathbf{a}, \lambda) = -B(\mathbf{a}, \lambda), \quad [A.6]$$

for all λ.

Now, the average of the product of the measurements is

$$P(\mathbf{a}, \mathbf{b}) = \int \rho(\lambda) A(\mathbf{a}, \lambda) B(\mathbf{b}, \lambda)\, d\lambda, \quad [A.7]$$

where $\rho(\lambda)$ is the probability density for the hidden variable. [Like any probability density, it is nonnegative, and satisfies the normalization condition $\int \rho(\lambda)\, d\lambda = 1$, but beyond this we make no assumptions about $\rho(\lambda)$; different hidden variable theories would presumably deliver quite different expressions for ρ.] In view of Equation A.6, we can eliminate B:

$$P(\mathbf{a}, \mathbf{b}) = -\int \rho(\lambda) A(\mathbf{a}, \lambda) A(\mathbf{b}, \lambda)\, d\lambda. \quad [A.8]$$

If **c** is any *other* unit vector,

$$P(\mathbf{a}, \mathbf{b}) - P(\mathbf{a}, \mathbf{c}) = -\int \rho(\lambda)\big[A(\mathbf{a}, \lambda)A(\mathbf{b}, \lambda) - A(\mathbf{a}, \lambda)A(\mathbf{c}, \lambda)\big]\, d\lambda. \quad [A.9]$$

Or, since $[A(\mathbf{b}, \lambda)]^2 = 1$:

$$P(\mathbf{a}, \mathbf{b}) - P(\mathbf{a}, \mathbf{c}) = -\int \rho(\lambda)\big[1 - A(\mathbf{b}, \lambda)A(\mathbf{c}, \lambda)\big]A(\mathbf{a}, \lambda)A(\mathbf{b}, \lambda)\, d\lambda. \quad [A.10]$$

But it follows from Equation A.5 that $-1 \leq [A(\mathbf{a}, \lambda)A(\mathbf{b}, \lambda)] \leq +1$, and $\rho(\lambda)[1 - A(\mathbf{b}, \lambda)A(\mathbf{c}, \lambda)] \geq 0$, so

$$|P(\mathbf{a}, \mathbf{b}) - P(\mathbf{a}, \mathbf{c})| \leq \int \rho(\lambda)\big[1 - A(\mathbf{b}, \lambda)A(\mathbf{c}, \lambda)\big]\, d\lambda, \quad [A.11]$$

or, more simply,

$$\boxed{\Big|P(\mathbf{a}, \mathbf{b}) - P(\mathbf{a}, \mathbf{c})\Big| \leq 1 + P(\mathbf{b}, \mathbf{c}).} \quad [A.12]$$

This is the famous **Bell inequality**. It holds for *any* local hidden variable theory (subject only to the minimal requirements of Equations A.5 and A.6), for we have made no assumptions whatever as to the nature or number of the hidden variables or their distribution (ρ).

But it is easy to show that the quantum mechanical prediction (Equation A.4) is incompatible with Bell's inequality. For example, suppose all three vectors lie in

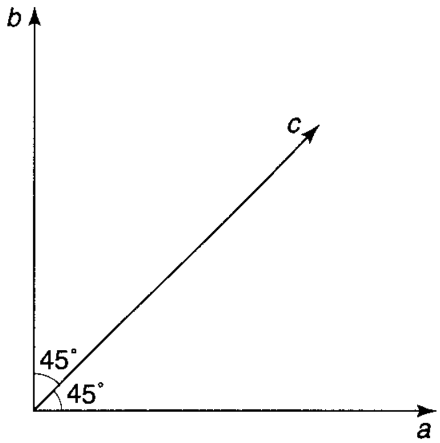

Figure A.3: An orientation of the detectors that demonstrates quantum violations of Bell's inequality.

a plane, and **c** makes a 45° angle with **a** and **b** (Figure A.3); in this case quantum mechanics says

$$P(\mathbf{a}, \mathbf{b}) = 0, \quad P(\mathbf{a}, \mathbf{c}) = P(\mathbf{b}, \mathbf{c}) = -0.707,$$

which is patently inconsistent with Bell's inequality:

$$0.707 \not\leq 1 - 0.707 = 0.293.$$

With Bell's modification, then, the EPR paradox proves something far more radical than its authors imagined: If they are right, then not only is quantum mechanics *incomplete*, it is downright *wrong*. On the other hand, if quantum mechanics is right, then *no* hidden variable theory is going to rescue us from the nonlocality Einstein considered so preposterous. Moreover, we are provided with a very simple experiment to settle the issue once and for all.

Many experiments to test Bell's inequality were performed in the 1960s and 1970's, culminating in the work of Aspect, Grangier, and Roger.[6] The details do not concern us here (they actually used two-photon atomic transitions, not pion decays). To exclude the remote possibility that the positron detector might somehow "sense" the orientation of the electron detector, both orientations were set quasi-randomly *after* the photons were already in flight. The results were in excellent agreement with the predictions of quantum mechanics and clearly incompatible with Bell's inequality.[7]

Ironically, the experimental confirmation of quantum mechanics came as something of a shock to the scientific community. But not because it spelled the demise of "realism"—most physicists had long since adjusted to this (and for those who could

---

[6]A. Aspect, P. Grangier, and G. Roger, *Phys. Rev. Lett.* **49**, 91 (1982).

[7]Bell's theorem involves *averages*, and it is conceivable that an apparatus such as Aspect's contains some secret bias which selects out a nonrepresentative sample, thus distorting the average. Recently, an improved version of Bell's theorem has been proposed in which a *single measurement* suffices to distinguish between the quantum prediction and that of any local hidden variable theory. See D. Greenberger, M. Horne, A. Shimony, and A. Zeilinger, *Am. J. Phys.* **58**, 1131, (1990) and N. David Mermin, *Am. J. Phys.* **58**, 731, (1990).

Figure A.4: The shadow of the bug moves across the screen at a velocity v' greater than c, provided that the screen is far enough away.

not, there remained the possibility of *nonlocal* hidden variable theories, to which Bell's theorem does not apply[8]). The real shock was the proof that *nature itself is fundamentally nonlocal.* Nonlocality, in the form of the instantaneous collapse of the wave function (and for that matter also in the symmetrization requirement for identical particles) had always been a feature of the orthodox interpretation, but before Aspect's experiment it was possible to hope that quantum nonlocality was somehow a nonphysical artifact of the formalism, with no detectable consequences. That hope can no longer be sustained, and we are obliged to reexamine our objection to instantaneous action at a distance.

Why *are* physicists so alarmed at the idea of superluminal influences? After all, there are many things that travel faster than light. If a bug flies across the beam of a movie projector, the speed of its shadow is proportional to the distance to the screen; in principle, that distance can be as large as you like, and hence the *shadow* can move at arbitrarily high velocity (Figure A.4). However, the shadow does not carry any *energy*; nor can it transmit any *message* from one point to another on the screen. A person at point X cannot *cause anything to happen* at point Y by manipulating the passing shadow.

On the other hand, a *causal* influence that propagated faster than light would carry unacceptable implications. For according to special relativity there exist inertial frames in which such a signal propagates *backward in time*—the effect preceding the cause—and this leads to inescapable logical anomalies. (You could, for example, arrange to kill your infant grandfather.) The question is, are the superluminal influences predicted by quantum mechanics and detected by Aspect *causal*, in this sense,

[8]It is a curious twist of fate that the EPR paradox, which *assumed* locality to *prove* realism, led finally to the repudiation of locality and left the issue of realism undecided—the outcome (as Mermin put it) Einstein would have liked *least.* Most physicists today consider that if they can't have *local* realism, there's not much point in realism at *all*, and for this reason nonlocal hidden variable theories occupy a rather peripheral place. Still, some authors—notably Bell himself, in *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 1987)—argue that such theories offer the best hope of bridging the gap between the measured system and the measuring apparatus, and for supplying an intelligible mechanism for the collapse of the wave function.

or are they somehow ethereal enough (like the motion of the shadow) to escape the philosophical objection?

Well, let's consider Bell's experiment. Does the measurement of the electron *influence* the outcome of the positron measurement? Assuredly it *does*—otherwise we cannot account for the correlation of the data. But does the measurement of the electron *cause* a particular outcome for the positron? Not in any ordinary sense of the word. There is no way the person monitoring the electron detector could use his measurement to send a signal to the person at the positron detector, since he does not control the outcome of his own measurement (he cannot *make* a given electron come out spin up, any more than the person at X can affect the passing shadow of the bug). It is true that he can decide *whether to make a measurement at all*, but the positron monitor, having immediate access only to data at his end of the line, cannot tell whether the electron was measured or not. For the lists of data compiled at the two ends, considered separately, are completely random. It is only when we *compare* the two lists later that we discover the remarkable correlations. In another reference frame, the positron measurements occur *before* the electron measurements, and yet this leads to no logical paradox—the observed correlation is entirely symmetrical in its treatment, and it is a matter of indifference whether we say the observation of the electron influenced the measurement of the positron, or the other way around. This is a wonderfully delicate kind of influence, whose only manifestation is a subtle correlation between two lists of otherwise random data.

We are led, then, to distinguish two types of influence: the "causal" variety, which produce actual changes in some physical property of the receiver, detectable by measurements on that subsystem alone, and an "ethereal" kind, which do not transmit energy or information, and for which the only evidence is a correlation in the data taken on the two separate subsystems—a correlation which by its nature cannot be detected by examining either list alone. Causal infiuences *cannot* propagate faster than light, but there is no compelling reason why ethereal ones should not. The influences associated with the collapse of the wave function are of the latter type, and the fact that they "travel" faster than light may be surprising, but it is not, after all, catastrophic.[9]

A.3 What is a Measurement?

The measurement process plays a mischievous role in quantum mechanics: It is here that indeterminacy, nonlocality, the collapse of the wave function, and all the attendant conceptual difficulties arise. Absent measurement, the wave function evolves in a leisurely and deterministic way, according to the Schrödinger equation, and quantum mechanics looks like a rather ordinary field theory [much simpler than classical

[9] An enormous amount has been written about Bell's theorem. My favorite is an inspired essay by David Mermin in *Physics Today* (April 1985, page 38). An extensive bibliography will be found in L. E. Ballentine, *Am. J. Phys.* **55**, 785 (1987).

electrodynamics, for example, since there is only *one* field (Ψ), instead of *two* (**E** and **B**), and it's a *scalar*]. It is the bizarre role of the measurement process that gives quantum mechanics its extraordinary richness and subtlety. But what, exactly, *is* a measurement? What makes it so different from other physical processes?[10] And how can we tell when a measurement has occurred?

Schrödinger posed the essential question most starkly, in his famous **cat paradox**:[11]

> A cat is placed in a steel chamber, together with the following hellish contraption .... In a Geiger counter there is a tiny amount of radioactive substance, so tiny that maybe within an hour one of the atoms decays, but equally probably none of them decays. If one decays then the counter triggers and via a relay activates a little hammer which breaks a container of cyanide. If one has left this entire system for an hour, then one would say the cat is living if no atom has decayed. The first decay would have poisoned it. The wave function of the entire system would express this by containing equal parts of the living and dead cat.

At the end of the hour, the wave function of the cat has the schematic form

$$\psi = \frac{1}{\sqrt{2}}(\psi_{\text{alive}} + \psi_{\text{dead}}). \qquad \text{[A.13]}$$

The cat is neither alive nor dead, but rather a linear combination of the two, until a measurement occurs—until, say, you peek in the window to check. At that moment your observation forces the cat to "take a stand": dead or alive. And if you find it to be dead, then it's really *you* who killed it, by looking in the window.

Schrödinger regarded this as patent nonsense, and I think most physicists would agree with him. There is something absurd about the very idea of a *macroscopic* object being in a linear combination of two palpably different states. An electron can be in a linear combination of spin up and spin down, but a cat simply cannot *be* in a linear combination of alive and dead. How are we to reconcile this with the orthodox interpretation of quantum mechanics?

The most widely accepted answer is that the triggering of the Geiger counter constitutes the "measurement," in the sense of the statistical interpretation, not the intervention of a human observer. It is the essence of a measurement that some *macroscopic* system is affected (the Geiger counter, in this instance). The measurement occurs at the moment when the microscopic system (described by the laws of

---

[10]There is a school of thought that rejects this distinction, holding that the system and the measuring apparatus should be described by one great big wave function which itself evolves according to the Schrödinger equation. In such theories there is no collapse of the wave function, but one must typically abandon any hope of describing individual events—quantum mechanics (in this view) applies only to *ensembles* of identically prepared systems. See, for example, Philip Pearle *Am. J. Phys.* **35**, 742 (1967), or, more recently, Leslie E. Ballentine, *Quantum Mechanics*, (Prentice Hall, Englewood Cliffs, NJ, 1990).

[11]E. Schrödinger, *Naturwiss.* **48**, 52 (1935); translation by Josef M. Jauch, *Foundations of Quantum Mechanics*, (Reading, MA: Addison-Wesley, 1968), p. 185.

quantum mechanics) interacts with the macroscopic system (described by the laws of classical mechanics) in such a way as to leave a permanent record. The macroscopic system itself is not permitted to occupy a linear combination of distinct states.[12]

I would not pretend that this is an entirely satisfactory resolution, but at least it avoids the stultifying solipsism of Wigner and others, who persuaded themselves that it is the intervention of human consciousness that constitutes a measurement in quantum mechanics. Part of the problem is the word "measurement" itself, which certainly carries an suggestion of human involvement. Heisenberg proposed the word "event", which might be preferable. But I'm afraid "measurement" is so ingrained by now that we're stuck with it. And, in the end, no manipulation of the terminology can completely exorcise this mysterious ghost.

A.4 The Quantum Zeno Paradox

The collapse of the wave function is undoubtedly the *most* peculiar feature of this whole story. It was introduced on purely theoretical grounds, to account for the fact that an immediately repeated measurement reproduces the same value. But surely such a radical postulate must carry directly observable consequences. In 1977 Misra and Sudarshan[13] proposed what they called the **quantum Zeno effect** as a dramatic experimental demonstration of the collapse of the wave function. Their idea was to take an unstable system (an atom in an excited state, say) and subject it to repeated measurements. Each observation collapses the wave function, resetting the clock, and it is possible by this means to delay indefinitely the expected transition to the lower state.[14]

Specifically, suppose a system starts out in the excited state ψ_2, which has a natural lifetime τ for transition to the ground state ψ_1. Ordinarily, for times substantially less than τ, the probability of a transition is proportional to t (see Equation 9.42); in fact, since the transition rate is $1/\tau$,

$$P_{2\to 1} = \frac{t}{\tau}. \qquad \text{[A.14]}$$

If we make a measurement after a time t, then, the probability that the system is still in the *upper* state is

$$P_1(t) = 1 - \frac{t}{\tau}. \qquad \text{[A.15]}$$

[12]Of course, in some ultimate sense the macroscopic system is *itself* described by the laws of quantum mechanics. But wave functions, in the first instance, describe individual elementary particles; the wave function of a macroscopic object would be a monstrously complicated composite, built out of all the wave functions of its 10^{23} constituent particles. Presumably somewhere in the statistics of large numbers macroscopic linear combinations become extremely improbable.

[13]B. Misra and E. C. G. Sudarshan, *J. Math. Phys.* **18**, 756 (1977).

[14]This phenomenon doesn't have much to do with Zeno, but it *is* reminiscent of the old adage "a watched pot never boils," so it is sometimes called the **watched pot effect**.

Suppose we *do* find it to be in the upper state. In that case the wave function collapses back to ψ_2, and the process starts all over again. If we make a *second* measurement, at $2t$, the probability that the system is *still* in the upper state is evidently

$$\left(1-\frac{t}{\tau}\right)^2 \approx 1-\frac{2t}{\tau}, \qquad \text{[A.16]}$$

which is the same as it would have been had we never made the measurement at t. This is certainly what one would naively expect; if it were the whole story there would be nothing gained by observing the system, and there would be no quantum Zeno effect.

However, for *extremely* short times, the probability of a transition is *not* proportional to t, but rather to t^2 (see Equation 9.39)[15]:

$$P_{2\to 1} = \alpha t^2. \qquad \text{[A.17]}$$

In this case the probability that the system is still in the upper state after the two measurements is

$$\left(1-\alpha t^2\right)^2 \approx 1-2\alpha t^2, \qquad \text{[A.18]}$$

whereas if we had never made the first measurement it would have been

$$1-\alpha(2t)^2 \approx 1-4\alpha t^2. \qquad \text{[A.19]}$$

Evidently our observation of the system after time t *decreased* the net probability of a transition to the lower state!

Indeed, if we examine the system at n regular intervals, from $t=0$ out to $t=T$ (that is, we make measurements at T/n, $2T/n$, $3T/n$, …, T), the probability that the system is still in the upper state at the end is

$$\left(1-\alpha(T/n)^2\right)^n \approx 1-\frac{\alpha}{n}T^2, \qquad \text{[A.20]}$$

which goes to 1 in the limit $n\to\infty$: A *continuously* observed unstable system never decays at all! Some authors regard this as an absurd conclusion, and a proof that the collapse of the wave function is fallacious. However, their argument hinges on a rather loose interpretation of what constitutes "observation." If the track of a particle in a bubble chamber amounts to "continuous observation," then the case is closed, for such particles certainly do decay (in fact, their lifetime is not measureably extended by the presence of the detector). But such a particle is only intermittently interacting with the atoms in the chamber, and for the quantum Zeno effect to occur the successive measurements must be made *extremely* rapidly to catch the system in the t^2 regime.

---

[15] In the argument leading to linear time dependence, we assumed that the function $\sin^2(\Omega t/2)/\Omega^2$ in Equation 9.39 was a sharp spike. However, the *width* of the "spike" is of order $\Delta\omega = 4\pi/t$, and for *extremely* short t this approximation fails, and the integral becomes $(t^2/4)\int \rho(\omega)d\omega$.

As it turns out, the experiment is impractical for spontaneous transitions, but it can be done using *induced* transitions, and the results are in excellent agreement with the theoretical predictions.[16] Unfortunately, this experiment is not as compelling a confirmation of the collapse of the wave function as its designers hoped; the observed effect can be accounted for in other ways.[17]

****

In this book I have tried to present a consistent and coherent story: The wave function (Ψ) represents the state of a particle (or system); particles do not in general possess specific dynamical properties (position, momentum, energy, angular momentum, etc.) until an act of measurement intervenes; the probability of getting a particular value in any given experiment is determined by the statistical interpretation of Ψ; upon measurement the wave function collapses, so that an immediately repeated measurement is certain to yield the same result. There are other possible interpretations—nonlocal hidden variable theories, the **many worlds** picture, ensemble models, and others—but I believe this one is conceptually the *simplest*, and certainly it is the one shared by most physicists today. It has stood the test of time, and emerged unscathed from every experimental challenge. But I cannot believe this is the end of the story; at the very least, we have much to learn about the nature of measurement and the mechanism of collapse. And it is entirely possible that future generations will look back, from the vantage point of a more sophisticated theory, and wonder how we could have been so gullible.

---

[16] W. M. Itano, D. J. Heinzen, J. J. Bollinger, and D. J. Wineland, *Phys. Rev. A* **41**, 2295 (1990).

[17] L. E. Ballentine, *Found. Phys.* **20**, 1329 (1990); T. Petrosky, S. Tasaki, and I. Prigogine, *Phys. Lett. A* **151**, 109 (1990).