12

The Interpretation of Quantum Mechanics

Throughout this book, we have relied on the *Copenhagen interpretation* of quantum mechanics. This is the traditional interpretation of the quantum-mechanical formalism. Its main features were initially sketched by Heisenberg and by Bohr, and its details were later filled in by many collaborators and disciples of Bohr at the Institute for Theoretical Physics at Copenhagen. Because this interpretation provides us with only probabilistic information about the state of a quantum-mechanical system, and because this interpretation has some weird aspects that go counter to our intuition, its adequacy has often been questioned. Several other interpretations of quantum mechanics have been proposed, but none has been judged clearly superior to the Copenhagen interpretation, which still remains the only widely accepted of all the interpretations of quantum mechanics. Of course, scientific issues are not decided by popularity polls, but the wide acceptance of the Copenhagen interpretation means that a physicist who wants to communicate some result or discovery in quantum mechanics will feel compelled to couch the result in the language of the Copenhagen interpretation.

Critics of the Copenhagen interpretation do not challenge the accuracy of the numerical results calculated from quantum mechanics. At a pragmatic level, quantum mechanics works perfectly—the numerical results for, say, the eigenvalues of the angular momentum and the energy of the hydrogen atom are found to be in perfect agreement with experiment. But critics challenge whether the Copenhagen interpretation really gives us the most

complete, most exhaustive knowledge of a quantum system we can hope for. For instance, is it really impossible to say anything about the precise instantaneous position of the electron in the hydrogen atom and its motion as a function of time? Or is the inability of quantum mechanics to provide this information an indication of some deficiency of the theory? In the view of some critics, the probabilistic character of the predictions of quantum mechanics is held to reflect our ignorance of the details of the underlying dynamics. Theories that attempt to provide a more detailed knowledge than provided by the Copenhagen interpretation are said to contain *hidden variables*.

The discussion of the interpretation of quantum mechanics and of hidden variables has received a fresh stimulus in recent years, because it has become possible to perform an experiment originally conceived as a *Gedankenexperiment* by Einstein, Podolsky, and Rosen in 1935. A new theoretical analysis of this *Gedankenexperiment* by Bell in 1964 established that it could be used to discriminate between the Copenhagen interpretation and a wide class of theories with hidden variables, and this encouraged experimenters to attempt some actual versions of the experiment. The experimental results fully support the Copenhagen interpretation and contradict theories with hidden variables.

12.1 The Copenhagen Interpretation

The main features of the Copenhagen interpretation can be briefly summarized as follows:

1. The state vector $|\psi\rangle$ provides a complete characterization of the state of the system.
2. The state vector tells us the probability distribution for the result of the measurement of any observable quantity. This probability distribution applies to each *individual* quantum particle or quantum system.
3. The uncertainty relations indicate the intrinsic spreads in the values of complementary observables for the *individual* quantum particle or quantum system. These uncertainty relation deny the existence of sharp values of complementary observables.
4. Measurements produce unpredictable, discontinuous changes in the state vector, which do not obey the Schrö-

dinger equation. The outcome of a single measurement of an observable is unpredictable—the outcome can be any of the eigenvalues within the spread of the probability distribution. During the measurement, the state of the system collapses into an eigenstate of the observable.

This list of features overlaps, to some extent, with the axioms of quantum mechanics stated in Chapters 4 and 5. We could include all of these axioms in our list of features of the Copenhagen interpretation, but some of these axioms—for instance, the axiom for the time evolution of the state vector—do not pertain directly to the *interpretation* of quantum mechanics, and this is why we prefer not to include them here. We have used the features of the Copenhagen interpretation in the preceding chapters. Now we will discuss them critically.

The fundamental assumption of the Copenhagen interpretation is that the state vector $|\psi\rangle$ (or, in the position representation, the wavefunction ψ) provides a complete, exhaustive characterization of the state of the system. This means that the state vector encompasses all that can be said about the state of the system. The other assumptions and prescriptions of the Copenhagen interpretation are built upon this fundamental assumption.

In contrast to the classical characterization of the state of a system, where the instantaneous coordinates and momenta give us a detailed picture of the instantaneous configuration of the system, the quantum-mechanical characterization by means of the state vector gives us merely the probabilities for the outcome of measurements that we can perform on the system. For instance, if $|E_n\rangle$ is an energy eigenstate, then $|\langle E_n|\psi\rangle|^2$ gives us the probability that the outcome of an energy measurement is E_n. From the probability distribution for the different energy eigenvalues, we can calculate the expectation value of the energy; alternatively, we can calculate this expectation value, or average value, according to the concise formula

$$\langle E\rangle = \langle\psi|H|\psi\rangle \tag{1}$$

where H is the energy operator. Similar formulas give us the expectation values of all other physical observables. Because the state vector $|\psi\rangle$, or the wavefunction ψ, determines the expectation values of all observables, Schrödinger has called the wavefunction the "expectation-catalog."

We must resist the temptation to regard the wavefunction as some kind of snapshop of the instantaneous configuration of the system, in the way that, say, the classical wavefunction for a standing wave on a string is a snapshot of the instantaneous configuration of the string. The quantum-mechanical wavefunction of, say, an electron in an atom does not give us a picture of the shape of the instantaneous configuration of matter or of electric charge in the atom. It merely gives us the probability distribution of the electric charge; it merely provides us with the means of calculating expectation values. The quantum-mechanical wavefunction makes no assertions about the instantaneous position of the electron or about the instantaneous charge distribution in the atom. One of the advantages of the abstract state vector $|\psi\rangle$ over the wavefunction ψ is that as long as we deal with the abstract state vector we are unlikely to fall into the error of imagining the wavefunction as some kind of actual configuration of electric charge in space.

Quantum mechanics does not supply us with concrete mental pictures of the behavior of atoms and subatomic particles. Quantum mechanics does not tell us what atoms and subatomic particles are like; it merely tells us what happens when we perform measurements. As Heisenberg said: "The conception of objective reality . . . evaporated into the . . . mathematics that represents no longer the behavior of elementary particles but rather our knowledge of this behavior."[1]

The emphasis of the Copenhagen interpretation on measurements and on the procedures for measurements is in accord with the philosophical doctrines of positivism and operationalism. In brief, positivism asserts that the only meaningful statements we can make about a physical system are those that are verifiable by observation and experiment, and thus the only meaningful physical quantities are those that are measurable. And operationalism asserts that the definition of any physical quantity must spell out the experimental, or "operational," procedure for measuring the quantity. According to strict positivist doctrine, the aim of science is to describe and to predict, but not to explain; speculations about unobservable and unmeasurable properties are held to be irrelevant.

This emphasis on measurements is a strength and also a weakness of the Copenhagen interpretation: strength, since its lack of commitment to any detailed model of the atomic and subatomic

[1] W. Heisenberg, Daedalus, **87**, 99 (1958).

world makes it nearly impregnable; and weakness, since it fails to satisfy our craving for concrete mental images of atomic and subatomic processes. Of course, we can imagine the wavefunction, and this can help us to understand the mathematical properties of ψ; but when you imagine, say, the scattering of an incident proton wave on a nuclear target, you are not seeing the physical behavior of the proton, only the mathematical evolution of the wavefunction.

In the Copenhagen interpretation, the meaning of the quantum-mechanical probability distributions is quite different from that of the probability distributions familiar from classical statistical mechanics. When a classical physicist has recourse to a probability distribution to describe, say, the speed of a molecule in a gas, he does not mean to deny that the molecule has a perfectly well defined speed at each instant of time; but he does not know this speed—he only knows some macroscopic quantities of the gas, such as the average density, temperature, and pressure. Hence, in classical statistical mechanics, the probability distribution for molecular speeds reflects the ignorance of the observer of the precise microscopic conditions in the gas. This kind of probability distribution is called an *ensemble* distribution, since it describes the average conditions for a large number of molecules in a gas. In contrast, the quantum-mechanical probability distribution does not reflect our ignorance of the instantaneous position and momentum, but rather the non-existence of any well-defined position and momentum. The quantum-mechanical system does not consist of particles with well-defined albeit unknown positions and momenta, but of "particles" with intrinsically indeterminate positions and momenta. Thus, the quantum-mechanical probability distributions (and the quantum-mechanical uncertainties Δx and Δp_x; see below) refer to an individual particle, not to an ensemble of particles. An example due to Schrödinger brings this distinction into clear focus: Consider a particle in an energy eigenstate of the isotropic harmonic oscillator, say, a particle in the ground state, with $E = \frac{3}{2}\hbar\omega$. A classical probability distribution for this system with well-defined, but unknown, values of the position and momentum would necessarily require that the distance of the particle from the origin never exceed the distance at which the energy $\frac{3}{2}\hbar\omega$ equals the potential energy (this is the classical turning point of the motion); thus, if we were to assume that the particle had a well-defined instantaneous position and momentum, the probability distribution would have a sharp cutoff at the classical turning point.[2] But the quantum-mechanical probability distribution has no such sharp cutoff—it permits the particle to penetrate into the classically forbidden region beyond the turning point. (As we already discussed in Chapter 3, this penetration into a classically forbidden region leads to no inconsistencies, because, in consequence of the uncertainty relation, any attempt at detecting the presence of the particle in the forbidden region introduces a large uncertainty in the energy and blurs the distinction between the forbidden and the permitted region.) The important lesson we extract from this example is that the consistency of the Copenhagen interpretation demands that the quantum-mechanical probability distribution be associated with an individual particle.

This raises the question of how we can perform an experimental measurement of probabilities when we are dealing with a single particle or a single system. A single trial, say, a single measurement of the position of a particle with some given wavefunction, cannot confirm the quantum-mechanical prediction for the probability distribution of the position. At the most, the single trial could prove quantum mechanics wrong—if the result of the single trial is a position that according to quantum mechanics has zero probability. For a comprehensive examination of the probability distribution, we must repeat the trial again and again, each time starting with the system prepared in the same way, so it has the same initial wavefunction for each trial. In practice, it is more convenient to prepare a large number of identical copies of the system, and measure the distribution of positions across this ensemble of copies. For instance, the measurement of the probability distribution in the diffraction pattern produced by an electron incident on a crystal is routinely performed by means of a beam of many electrons incident on the crystal. The diffracted electrons emerge from the crystal and strike a fluorescent screen, where they generate small flashes of light. Each flash of light amounts to a repeated trial of the experiment. However, under typical experimental conditions, the electrons arrive at the screen in quick succession, and we do not notice the individual flashes of light. What we see on the screen is a more or less steady pattern of

[2] Some hidden-variable theories bypass this requirement by modifying the potential energy. Thus, a hidden-variable theory contrived by Bohm adds to the ostensible potential energy $\frac{1}{2}m\omega^2x^2$ of the harmonic oscillator an extra term depending on the wavefunction $\psi(x)$ (see Problem 3). The classical turning point is then at infinity, and the probability distribution has no sharp cutoff.

bright and dark zones, which give us a direct picture of the probability distribution (see Fig. 1.2). According to the Copenhagen interpretation, the probability distribution for such an ensemble of repeated trials of the diffraction experiment is equal to the probability distribution for each individual electron, and the width of this probability distribution across the screen (or, more precisely, the rms deviation from the mean) is equal to the uncertainty in the position of each individual electron upon arrival at the screen. Of course, after the electron interacts with some atom in the screen and triggers the emission of a flash of light, the uncertainty in its position will be much smaller (this final uncertainty depends on the details of the interaction between the electron and the atom).

The state vector $|\psi\rangle$ presents us with a probability distribution for the possible values for every observable quantity. In general, this probability distribution spans several, or many, values of the observable, and therefore the outcome of a measurement of the observable is afflicted with uncertainties. Only in the exceptional case that $|\psi\rangle$ is an eigenvector of the observable does this uncertainty disappear—the outcome of the measurement is then certain to be the eigenvalue. However, the commutation relations of quantum mechanics place severe restrictions on what observables can simultaneously be free of uncertainties, that is, what observables can have simultaneous eigenvectors. For complementary observables, such as the position x and the momentum p_x, whose commutator has the canonical form $[x, p_x] = i\hbar$, there are no simultaneous eigenvectors, and the certainty in one of these observables implies total uncertainty in the other, in accord with the Heisenberg uncertainty relation

$$\Delta x \, \Delta p_x \geq \frac{\hbar}{2} \tag{2}$$

The uncertainties Δx, Δp_x, and other such quantum-mechanical uncertainties refer to an individual particle, not to an ensemble of particles. These quantum-mechanical uncertainties do not arise from our ignorance of some underlying details of the state of the particle or from some inadequacy of our measuring devices. Instead, the uncertainties reflect the nonexistence of such details; they reflect an intrinsic spread in the position and the momentum of the particle. The position and the momentum are not sharply defined; they are *indeterminate.*

The uncertainty relations are often called *indeterminacy relations,* to distinguish the quantum-mechanical uncertainties from ordinary experimental uncertainties arising from imperfections in the apparatus used in a measurement. To illustrate the distinction between these two kinds of uncertainties, consider the position measurement of an electron by means of a fluorescent screen. In this case, there are two different uncertainties: there is the initial intrinsic uncertainty of the position of the electron, associated with its initial wavefunction; and there is the ordinary experimental uncertainty of the measurement associated with the resolution attainable with the fluorescent screen. We can check the ordinary experimental uncertainty in the measurement of the position of an individual electron by an immediate repetition of this measurement, say, by means of a second fluorescent screen placed adjacent to the first (we will assume that the electron has enough energy to pass through both screens). The electron will then trigger the emission of a flash of light in the second screen, and the difference between the average positions of the flashes triggered in the first and the second screens tells us the ordinary experimental uncertainty. Note that this experimental uncertainty is never smaller than the quantum-mechanical uncertainty associated with the final wavefunction of the electron, after the measurement (in the best conceivable measurement, the experimental uncertainty attains the level of the final quantum-mechanical uncertainties). Whenever we perform a measurement, we must always make a careful distinction between these different kinds of uncertainty.

The Heisenberg uncertainty relation for the position and the momentum of a particle implies that classical determinism fails—the initial values of the position and the momentum (or velocity) of a particle cannot be used to predict the position and momentum at a later time. To make such a prediction for the motion of a particle, the classical physicist would need precise initial values of position and momentum; but Eq. (2) forbids simultaneous precise values of these two observables.

Although quantum mechanics lacks the simple determinism of classical physics, it retains a form of determinism in the state vector $|\psi\rangle$, which evolves in time according to the (general) Schrödinger equation,

$$\frac{\hbar}{i}\frac{d}{dt}|\psi\rangle = H|\psi\rangle \tag{3}$$

This equation expresses determinism and causality, since it permits us to predict the state vector at any later time from a given state vector at the initial time. Thus, in the words of Born: "The motion of particles is subject [only] to probabilistic laws, but the probability itself evolves in accord with causal laws."[3]

12.2 *Measurement and the Collapse of the Wavefunction*

The weirdest feature of the Copenhagen interpretation is that it requires that the wavefunction suffer a discontinuous, unpredictable change during the measurement. Consider, for instance, the impact of an electron on the fluorescent screen in the electron-diffraction experiment. This impact and the flash of light released in it constitute an (approximate) measurement of the position of the electron. Just before this measurement, the wavefunction was spread out all over the screen; immediately after the measurement, the electron position is known to lie within some small spot on the screen, and the wavefunction must therefore have an extent no larger than this spot. Thus, during the measurement, the wavefunction suffers an unpredictable *collapse,* or *reduction.* The collapse is unpredictable, since we have no way of knowing onto what part of the screen the wavefunction will collapse—we know only the probability distribution of the spots on which the wavefunction collapses, that is, the probability distribution of positions for the electron on the screen.

Note that a single measurement tells us very little about the wavefunction *before* the measurement. If a measurement finds an electron at some spot, this tells us only that the wavefunction before the measurement was different from zero at that spot. But the measurement tells us much about the wavefunction just *after* the measurement. In general, a precise measurement of an observable collapses the wavefunction into an eigenstate of that observable. Thus, the wavefunction after such an ideal measurement is precisely known. For instance, an ideal measurement of the position of an electron collapses the wavefunction into a delta function (the practical measurement with a fluorescent screen has a limited precision, given by the experimental uncertainty; and the wavefunction after the measurement is not a delta function, but a wave packet of a width of a few Å). A measurement of the energy of an electron collapses the wavefunction into an eigenstate of energy.

[3] M. Born, Z. Phys., **38**, 803 (1926).

A measurement of the spin collapses the wavefunction into an eigenstate of spin, and so on. The apparatus plays a crucial role in selecting the kind of eigenstate into which the wavefunction collapses. The apparatus dictates whether the system will collapse into some eigenstate of position, or of momentum, or of spin, and so on. But, of course, the apparatus does not dictate which specific eigenstate of position, or of energy, or of spin, and so on, the system will collapse into; this aspect of the collapse is unpredictable.

Bohr has emphasized that quantum mechanics does not describe quantum systems per se; instead, it describes a whole phenomenon, which includes, in an inextricable way, both the quantum system and the apparatus used to measure it: ". . . an independent reality in the ordinary physical sense can neither be ascribed to the phenomenon nor to the agencies of observation."[4] According to the Copenhagen interpretation, quantum systems in themselves do not have sharply defined attributes, only diffuse potentialities, which are capable of becoming sharply defined when we perform suitable measurements. The attributes of a quantum system depend on the apparatus used to measure them, and they exist only in relation to this apparatus. Thus, the attributes are a joint property of the system and the apparatus. This intimate symbiotic relationship between system and apparatus implies a break with naive realism, according to which the attributes of a physical system belong to the system in itself, and they are supposed to exist independently of the environment surrounding the system. However, in the view of most physicists, the antirealism of the Copenhagen interpretation extends only to the attributes of physical systems, not to the physical systems themselves. The Copenhagen interpretation denies the realism of contingent attributes, but it does not deny the realism of physical systems or of the material world.

In an ideal measurement, the collapse is instantaneous—at one instant the wavefunction has one configuration, at the next instant it has collapsed to a new, drastically different configuration. Such an instantaneous collapse would seem to conflict with the theory of Special Relativity, according to which signals can never exceed the speed of light (if signals can be sent with a speed exceeding the speed of light, then you can send messages into your own past, in blatant violation of causality). But it is easy to

[4] N. Bohr, *Quantum Theory and the Description of Nature,* Chapter II.

see that the collapse process cannot be used to transmit messages from one observer to another. For instance, consider an electron wave that has spread out over a very large volume, say several light years, and suppose that an observer at one end of this electron wave detects the electron on her fluorescent screen and brings about the collapse of the wavefunction. This means that it will thereafter be impossible for another, distant observer to detect the electron on his fluorescent screen; but this does not give this other observer a message of any sort, since he has no way of knowing that his attempts at detecting the electron have been condemned to failure. The other observer is now on a fool's errand—he can continue to grope around searching for the electron, but he is unable to conclude that the electron wave has collapsed until he has explored *every* volume element in space, including the vicinity of the first observer, where, of course, he will finally get the message.

The change of the wavefunction during the collapse is not governed by the Schrödinger equation. As we will see in Section 12.3, the Copenhagen interpretation brazenly postulates that this collapse is merely a mathematical procedure, not a physical process. We might be tempted to suppose that the collapse is produced by the dynamics of the interaction of the measured system with the measuring apparatus. But such an interaction, if treated according to the time evolution specified by the Schrödinger equation, is not by itself enough to bring about the collapse of the wavefunction. For instance, when the diffracted electron wave strikes the fluorescent screen, it interacts with *all* the atoms in the screen and scatters off them with some loss of energy (inelastic scattering). This interaction of the electron with the atoms in the screen tells us the probabilities for the emission of flashes of light by the atoms, but it does not tell us which of the many atoms on the screen will actually emit the light, and thereby signal the collapse of the electron wavefunction into its vicinity. Thus, although the interaction of the system and the measuring apparatus is required for the measurement to be possible, the collapse is not produced by this interaction.

It might be argued that single atoms, or groups of atoms, in the fluorescent screen do not constitute a macroscopic measuring device, and that therefore the collapse is to be expected to occur only at the next stage of the measurement process, when the flash of light from the screen triggers a macroscopic measurement device, such as a photomultiplier tube or the human eye. The interaction

of an electron with the atoms of a fluorescent screen is complicated, and the ensuing complete chain of events is difficult to analyze in detail. Instead, let us deal with a different example of measurement, in which the interactions are more obvious, and the evolution of the joint system–apparatus state vector can be examined in some detail.

Consider the measurement of the vertical component of the spin of an atom with the Stern–Gerlach apparatus shown in Fig. 12.1. In this measurement, we send the atom through the inhomogeneous magnetic field of a magnet, which displaces the trajectory of the atom vertically upward if the spin is up, and vertically downward if the spin is down. Two detectors serve to discover whether the atom has taken the high road or the low road. In the original Stern–Gerlach experiment, the detector was simply a photographic plate. But such a device absorbs the atom and precludes any further measurements. For our purposes, it will prove more instructive to use some detector that permits the atom to proceed on its trajectory. A suitable detector might consist of a laser beam that intersects the trajectory and a photomultiplier tube aimed at the intersection. If the atom crosses the laser beam, it scatters a

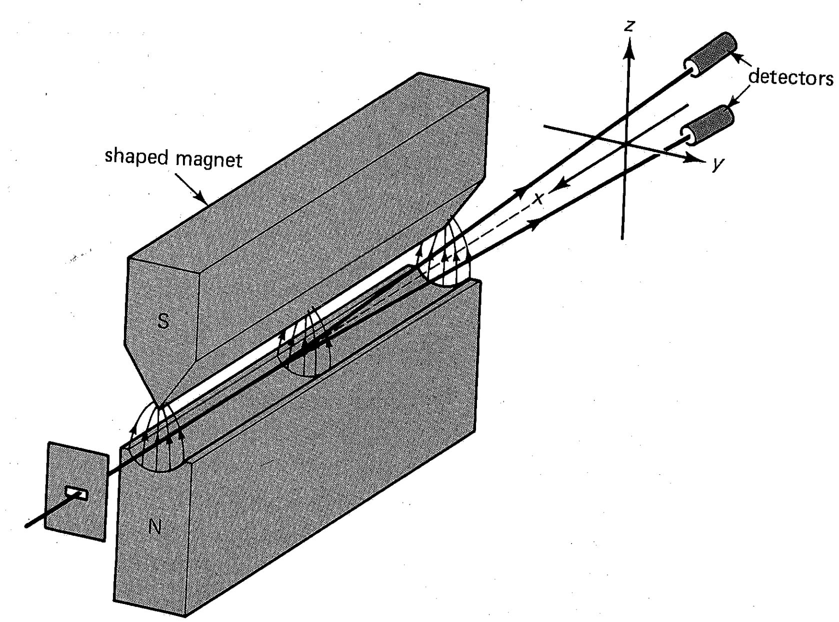

Fig. 12.1 Stern–Gerlach apparatus for the measurement of the vertical component of the spin.

few photons, which reveal the presence of the atom when they trigger the photomultiplier tube. Since the resolution of this simple optical detector is of no concern to us, we can use photons of long wavelength, which hardly disturb the atom at all. Note that this apparatus contains two basic elements: a discriminating device (the inhomogeneous magnetic field) that sends the atom one way or another according to the vertical component of its spin, and an amplifying device (the photomultiplier) that produces a macroscopic pulse of current when triggered by an incident atom. These two basic elements are quite typical of most measuring devices used for measurements on quantum systems.

We can easily see that as long as the atom, the apparatus, and their interactions are governed by the Schrödinger equation, a collapse of the state vector is not possible. We assume that the atom has spin $\frac{1}{2}$ and that the initial state of the atom is some superposition of the spin-up and spin-down states, say, the superposition

$$\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle) \qquad (4)$$

which corresponds to an initial state of horizontal spin, in the $+x$ direction [see Eq. (9.33)]. According to the usual rules for calculating probabilities for the outcome of a measurement, this initial state of the atom has a probability of $\frac{1}{2}$ for spin up, and $\frac{1}{2}$ for spin down. The initial state of the detectors, before the measurement, is that neither of them has been triggered; we designate this state by $|\text{none}\rangle$. The initial state vector for the joint atom–detector system is then

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)\,|\text{none}\rangle = \frac{1}{\sqrt{2}}|+\rangle\,|\text{none}\rangle + \frac{1}{\sqrt{2}}|-\rangle\,|\text{none}\rangle \qquad (5)$$

Note that in this state vector we have not bothered to indicate explicitly the state of translational motion of the atom. Within the approximations of the Stern–Gerlach experiment, the translational motion of the atom proceeds according to the classical mechanics, and the upward or downward displacement of the trajectory for the spin-up or the spin-down state is completely determined by the (classical) parameters of the apparatus. Thus, there is a unique correspondence between spin states and trajectories, and we can pretend that the spin states $|\pm\rangle$ include an implicit specification of translational states.

The atom passes through the apparatus and interacts with the detectors. During this interaction, which is described by some suitable interaction Hamiltonian, the spin-up state triggers the upper detector, but not the lower; we designate the corresponding state vector of the detectors by $|\text{upper}\rangle$. The spin-down state triggers the lower detector, but not the upper; we designate the corresponding state vector by $|\text{lower}\rangle$. Thus, if the initial state were one or the other of the states of definite spin, the interaction during the measurement would produce the following transition to a definite final state:

$$|+\rangle\,|\text{none}\rangle \rightarrow |+\rangle\,|\text{upper}\rangle \quad \text{or} \quad |-\rangle\,|\text{none}\rangle \rightarrow |-\rangle\,|\text{lower}\rangle \qquad (6)$$

Since the Schrödinger equation is linear, the superposition (5) of initial states will therefore produce a corresponding superposition of final states:

$$|\psi\rangle = \frac{1}{\sqrt{2}}|+\rangle\,|\text{upper}\rangle + \frac{1}{\sqrt{2}}|-\rangle\,|\text{lower}\rangle \qquad (7)$$

Thus, the result of the interaction is not a collapse into one or another of the states of definite spin up or down and a definite response from the detector, but a superposition in which the spin-up and spin-down states are correlated with the detector states. But such a superposition cannot be regarded as a completed measurement, since it fails to make a definite choice between the spin-up and spin-down states. In fact, the state vector (7) is merely the initial state vector (5) translated in time. With a slight modification of the apparatus, it is even possible to restore the initial horizontal spin state of the atom by a further translation in time. If we add a second magnet to the apparatus in tandem with the first and with a reversed magnetic field, then the trajectory of the atom will suffer a reversed displacement in the second magnet, and the initial spin state of the atom will be restored when the upper and the lower trajectories again merge into one.[5] Such a restoration of the horizontal spin state demonstrates that our Stern–Gerlach experiment cannot be regarded as a completed measurement of the vertical component of the spin.

[5] In order to achieve a complete restoration of the initial state vector, we also have to reset the detectors, so their state vector is restored to $|\text{none}\rangle$. Since the detectors are macroscopic devices, a restoration to the exact initial quantum state is difficult, perhaps impossible. Some versions of the Copenhagen interpretation use lack of reversibility as a criterion for what constitutes a completed measurement (see the next section).

Although the state vector (7) does not, in itself, provide an adequate description of the outcome of a measurement, we might hope that we can bring about its collapse into one or another of the two states of this correlated superposition by performing a measurement on the joint atom–apparatus system. For this purpose, we might employ a secondary apparatus that observes the primary apparatus and checks which detector has triggered. The usual rule for calculating probabilities tells us that such a measurement on the state vector (7) has a probability of $\frac{1}{2}$ for the result $|+\rangle$ $|\text{upper}\rangle$ and a probability of $\frac{1}{2}$ for the result $|-\rangle$ $|\text{lower}\rangle$; thus, the probabilities for the outcomes of the secondary measurement are consistent with the probabilities for the primary measurement that we attempted with the Stern–Gerlach apparatus. However, if the secondary apparatus and its interaction with the primary apparatus are, again, governed by the Schrödinger equation, then this attempt at a measurement yields, again, a superposition:

$$|\psi\rangle = \frac{1}{\sqrt{2}}|+\rangle\,|\text{upper}\rangle\,|\text{upper confirmed}\rangle + \frac{1}{\sqrt{2}}|-\rangle\,|\text{lower}\rangle\,|\text{lower confirmed}\rangle \qquad (8)$$

From this example we see that stacking one apparatus on top of another does not bring about the desired collapse of the state vector.

The absence of collapse in any system governed by the Schrödinger equation—and the concomitant impossibility of bringing a measurement to completion, no matter how many apparata are stacked one on top of another—is called *von Neumann's catastrophe of infinite regression*. This absence of collapse was established by von Neumann, who made the first rigorous examination of the mathematical foundations of quantum mechanics. Von Neumann decided that the collapse of the state vector during measurement must be inserted into quantum mechanics as a separate axiom. If we arrange any number of apparata in a sequential stack (with a primary apparatus, a secondary apparatus, a tertiary apparatus, etc.), in which each apparatus checks on the apparatus ranking below it and is, in turn, checked by the apparatus ranking above it, we must postulate that the collapse of the state vector occurs somewhere in this stack. As in the cases of one or two apparata discussed in connection with Eqs. (7) and (8), the probabilities for the different outcomes of measurement are unaffected by whether we postulate that the collapse occurs in the primary apparatus, or the secondary apparatus, or the tertiary apparatus, etc.

Apart from its inadequacy for describing the outcome of a measurement, the state vector (7) has some weird properties. This state vector represents an ambivalent state, in which the detectors are in a schizoid superposition of having triggered and not having triggered. Such superpositions are a familiar feature of the atomic or subatomic world, and since we have no direct experience with that world, our intuition is willing to accept such superpositions in that world. But in Eq. (7) we encounter such a superposition in the macroscopic world, where it directly clashes with our intuition.

The weirdness of such superpositions of macroscopic states is brought to an extreme in a celebrated *Gedankenexperiment* contrived by Schrödinger:[6]

> A cat is locked into a steel chamber, along with the following diabolical device (which one must secure against direct intervention by the cat): In a Geiger counter there is a minuscule amount of radioactive substance, so small, that in the course of one hour *perhaps* 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 with 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 phrase) mixed or smeared out in equal parts.

Although such a schizoid superposition of a live-cat state vector and a dead-cat state vector does violence to our intuition, we cannot disprove it by any experiment. As soon as we open the chamber, or use any measuring device to detect the life signs of the cat, the state vector collapses into either the live-cat configuration or the dead-cat configuration, with equal probabilities. Thus, we can never "see" the cat in the superposed state. Instead, the act of observation or of measurement does something very drastic to the state of the cat—it flips the cat into either the live state or the dead state.

Wigner has added an extra wrinkle to this *Gedankenexperiment* by proposing that we omit the cyanide capsule and that we

[6] E. Schrödinger, Naturwissenschaften, **23**, 807 (1935); a translation of this paper appeared in Proc. Am. Philos. Soc., **124**, 323 (1980).

replace the cat by a human volunteer, Wigner's friend. We let Wigner's friend watch the Geiger counter for a while, and then open the chamber, and ask her to tell us what has happened. If Wigner's friend tells us that the Geiger counter clicked some time ago, we would presumably have to conclude that her presence in the chamber was enough to collapse the state vector, and that our opening of the chamber had no further effect on the state vector.

12.3 *Alternative Interpretations of the Collapse*

Physicists have made a variety of attempts at resolving the conundrum posed by the collapse of the wavefunction during measurement. Most of these attempts accept the main features of the Copenhagen interpretation (listed in Section 12.1), but propose different ways of dealing with the collapse. Here we will briefly discuss four such attempts: the orthodox Copenhagen picture, the popular picture, the subjective picture, and the many-worlds picture.[7]

Orthodox Copenhagen Picture. This is the picture conceived by Bohr and by Heisenberg.[8] An essential feature of this picture is that the results obtained in any experiment are to be described in classical terms. Bohr argued that such a classical view of experimental results is imperative to enable physicists to communicate these results to each other: "However far the phenomena transcend the scope of classical physical explanation, the account of all evidence must be expressed in classical terms. The argument is simply that by the word 'experiment' we refer to a situation where we can tell others what we have done and what we have learned and that, therefore, the account of the experimental arrangement and of the results of the observations must be expressed in unambiguous language with suitable application of the terminology of classical physics."[9] Thus, the apparatus is supposed to indicate the result of a measurement with a well-defined pointer position

[7] There is no general agreement on names for these and other different pictures, or even on the number of different pictures.

[8] The following discussion is based mainly on the exposition of the Copenhagen interpretation given by Heisenberg in his *Physics and Philosophy,* Chapter III. Bohr never gave such a systematic exposition of his views; but the isolated statements that he made are consistent with Heisenberg's views.

[9] N. Bohr, *Atomic Physics and Human Knowledge,* p. 39.

on a scale or a well-defined digital readout, without any significant uncertainty.

The orthodox Copenhagen picture does not claim that the laws of quantum physics are inapplicable to the apparatus; on the contrary, Bohr was quite aware that quantum physics is ultimately responsible for all the properties of the materials out of which the apparatus is constructed, and, in his refutation of the *Gedankenexperimente* of Einstein (see Section 12.4), he did not hesitate to apply the uncertainty relations to macroscopic pieces of equipment. But a "good" apparatus is supposed to be designed in such a way that quantum uncertainties in the readout are insignificant. According to Bohr's criterion for a "good" apparatus, if a superposition of different macroscopic apparatus states—such as Eq. (7)—where to occur, it would demonstrate an inadequacy in the design of the apparatus, and the inability of this apparatus to bring the measurement to completion.

In the orthodox Copenhagen picture, the collapse of the wavefunction is not a physical process, merely a mathematical procedure, or a bookkeeping procedure. The wavefunction is not a material entity, merely a mathematical construct. The wavefunction is the expectation catalog that characterizes the quantum system. This catalog lists all the possible outcomes for all the possible experiments we might perform on the system; it tells us that if we perform some experiment, then the outcome will be this or that, with this or that predicted probability. As long as we do not perform any experiment on the system, the expectation catalog evolves continuously in time, according to the Schrödinger equation. But if we perform an experiment on the system and measure some observable, the expectation catalog changes discontinuously. During the measurement, one of the possible outcomes listed in the expectation catalog becomes the actual outcome, and all the other outcomes are rejected. This means that the expectation catalog must suddenly be altered—all the entries in the catalog must be deleted, except, of course, the one entry that became actual. But this sudden alteration, or collapse, of the expectation catalog is merely a reflection of our sudden change of knowledge. The only definite *physical* change during the measurement is the change that occurs in the apparatus, which switches from one well-defined state to another.

Both Bohr and Heisenberg have emphasized that the wavefunction does not tell us what actually happens in the quantum system between one measurement and the next; it does not provide a history of events in the quantum system. The wavefunction, in conjunction with the Schrödinger equation, merely tells us that if we perform a given experiment on the system at one time (say, a measurement of the x component of the spin of an atom), and some other given experiment at a later time (say, a measurement of the z component of the spin), then the outcome of the second experiment is probabilistically related to the outcome of the first. Thus, the mathematical machinery of quantum mechanics provides us with probabilistic connections between one experiment and the next, but it does not provide us with a mental picture of what happens in between (as already mentioned in Section 12.1, we must not confuse a mental picture of the time evolution of the wavefunction with a mental picture of the quantum system itself). Quantum mechanics tells us nothing about the quantum system itself, only about what happens in measurements. In Bohr's words: "The formalism of quantum mechanics is to be considered as a tool for deriving predictions of a . . . statistical character as regards information obtainable under experimental conditions described in classical terms."[10]

The orthodox Copenhagen interpretation insists on a sharp dichotomy between quantum system and apparatus. We must begin any description of an experiment by specifying the system to be measured, the apparatus with which it is to interact, and the dividing line, or the *Heisenberg cut,* between the system and the apparatus. The state of the quantum system is characterized by its wavefunction, but the state of the apparatus (or, at least, the state of the readout end of the apparatus) is characterized by well-defined classical parameters.

Although we must draw a sharp line between the quantum system and the apparatus, we have considerable freedom in just where we draw this sharp line. As is clear from the discussion of Eqs. (7) and (8), the probabilities for outcomes of measurements are not altered when we extend our quantum system so as to include some part of the apparatus with which it interacts during the measurement. For instance, if we attempt to detect a photon with a photomultiplier tube, we can regard the photon as the system and the photomultiplier tube as the apparatus. Alternatively, we can regard the photon and some portion of the photomultiplier tube as the system and the remainder as the apparatus. In the photomultiplier tube, the incident photon ejects a photoelectron from the faceplate, this electron strikes the first dynode and ejects

[10] N. Bohr; see Jammer, The *Philosophy of Quantum Mechanics*, p. 204.

several electrons; each of these strikes the second dynode and ejects more electrons, and so on. We can draw the dividing line between system and apparatus at the faceplate, or at the first dynode, or at the second, and so on. Depending on our choice of dividing line, the measured system will consist of a photon or a photon and one or several electrons; accordingly, the wavefunction of the system will have to include the wavefunction of these electrons. However, the Copenhagen picture does not permit us to shift the dividing line all the way to the output end of the photomultiplier, where a *classical* pulse of current emerges. According to Bohr, the classical mode of description becomes compulsory by the occurrence of an irreversible process of amplification; this brings the measurement to completion. This criterion for the completion of a measurement has been enthusiastically advocated by Wheeler who declared: "A phenomenon is not yet a phenomenon until it has been brought to a close by an irreversible act of amplification, such as the blackening of a grain of silver bromide emulsion or the triggering of a photodetector."[11] Wheeler has emphasized that a decisive test for the completion of a measurement is the registration of the information acquired in the measurement, in the form of a permanent, indelible record.

In the *Gedankenexperiment* of Schrödinger's cat, the orthodox Copenhagen interpretation claims that the quantum-mechanical wavefunction collapses when the Geiger counter makes a measurement on the radioactive substance, and therefore the state of the Geiger counter (and the state of the cat) never forms a superposition of two macroscopically different states. At each instant, the Geiger counter either performs an irreversible act of amplification or does not perform such an act, that is, the Geiger counter adopts either a definite state of discharge or a definite state of no discharge. This means that the Geiger counter acquires information about the radioactive decay, and brings about the collapse of the wavefunction of the radioactive substance. The human observer is not required to bring about the collapse. When the observer opens the chamber, he receives the information about the collapsed wavefunction; but since this information was already available in the output of the Geiger counter, he produces no further collapse of the wavefunction.

Although the criterion of irreversible amplification for the

[11] J. A. Wheeler in *Quantum Theory and Measurement*, edited by J. A. Wheeler and W. H. Zurek, p. 185.

completion of a measurement and the collapse of the wavefunction seems quite plausible, it is fraught with ambiguities. In a photomultiplier tube the amplification increases by stages. At what stage of this process of successive amplifications with successively increasing irreversibility will we attain sufficient amplification and sufficient irreversibility for the completion of the measurement? Furthermore, the amplification process spans some time, and this raises the question of whether perhaps the collapse of the wavefunction is also spans some time, which would mean that the collapse is not truly discontinuous.

Popular Picture. Physicists have a deep predilection for continuity in nature (*Natura non facit saltus*), and they tend to be uncomfortable with the discontinuous collapse and with the somewhat capricious dichotomy between measured system and apparatus demanded by the orthodox Copenhagen picture. The popular picture is an alternative to the orthodox Copenhagen picture; it is favored by many, perhaps by most, of the physicists of today. In the popular picture, there is no collapse. The state vector evolves continuously at all times, according to the Schrödinger equation. Both the system and the apparatus are treated quantum-mechanically, and they are described by a joint state vector. A measurement is regarded as an interaction between the system and the apparatus, as in our example of the Stern–Gerlach experiment of Section 12.2. During such an interaction, the state vectors of the system and the apparatus become correlated, and the joint state vector forms a superposition of these correlated state vectors. Thus, in our example of the Stern–Gerlach experiment, the result of the measurement is the joint state vector

$$|\psi\rangle = \frac{1}{\sqrt{2}}\,|+\rangle|\text{upper}\rangle + \frac{1}{\sqrt{2}}\,|-\rangle\,|\text{lower}\rangle \tag{9}$$

Consider, now, the expectation value of any operator R that acts on the spin states (but not on the apparatus states). According to the usual prescription for the calculation of an expectation value,

$$\begin{aligned}\langle\psi|R|\psi\rangle = \tfrac{1}{2}(\langle+|R|+\rangle\,\langle\text{upper}|\text{upper}\rangle + \langle-|R|-\rangle\,\langle\text{lower}|\text{lower}\rangle \\ + \langle+|R|-\rangle\,\langle\text{upper}|\text{lower}\rangle + \langle-|R|+\rangle\,\langle\text{lower}|\text{upper}\rangle)\end{aligned} \tag{10}$$

The two apparatus states are, of course, normalized, with $\langle\text{upper}|\text{upper}\rangle = 1$ and $\langle\text{lower}|\text{lower}\rangle = 1$. But the apparatus states

are also orthogonal, since the triggering of one detector has no associated probability for triggering of the other detector (if the triggering of one detector tends to produce a triggering of the other detector, there is a defect in the design of the electric connections in the apparatus, and the experimenter must repair the apparatus). Thus, the last two terms in Eq. (10) vanish, and the first two terms reduce to

$$\langle\psi|R|\psi\rangle = \tfrac{1}{2}(\langle+|R|+\rangle + \langle-|R|-\rangle) \tag{11}$$

This result is the same as what we obtain if we take the expectation value of R for the collapsed state vector $|+\rangle|\text{upper}\rangle$ and for the collapsed state vector $|-\rangle|\text{lower}\rangle$, and we average these two possible expectation values. We therefore reach the conclusion that, on the average, for repeated measurements, the expectation value of R is the same as in the Copenhagen picture. Thus, although for each individual measurement, the collapsed state vectors $|+\rangle|\text{upper}\rangle$ and $|-\rangle|\text{lower}\rangle$ differ from the superposed state vector (9), we cannot detect this difference experimentally, because the average expectation values for repeated measurements are indistinguishable.

Exercise 1. The calculation leading to Eq. (11) was based on the simple initial state vector (5), with equal coefficients for the spin-up and spin-down states. Repeat the calculation for a general initial state, consisting of a superposition of spin-up and spin-down states with arbitrary coefficients c_1 and c_2. Show that the expectation values $\langle\psi|R|\psi\rangle$ is, again, an average of the two terms given in Eq. (11), but with weights $|c_1|^2$ and $|c_2|^2$. Thus, the general result is in agreement with the Copenhagen interpretation.

This means that in the popular picture there is collapse without collapse: the state vector does not really collapse, but the results for expectation values are the same as though it had collapsed. When we want to calculate an expectation value, we can therefore use collapse as a shortcut. As Eq. (11) shows, we can omit the state vector of the apparatus in our calculation, and we can *pretend* that the state vector of the system has collapsed to $|+\rangle$ or to $|-\rangle$. This shortcut always yields the same result as the honest calculation with the true, uncollapsed joint state vector given in Eq. (9).

From Eq. (10) we see that what permits the uncollapsed state vector to mimic an average calculated from the collapsed state vectors is the cancellation of the last two terms in this equation, that is, the off-diagonal terms. A somewhat different version of the popular picture attempts to achieve such a cancellation by exploiting unpredictable, random phase differences that are supposedly introduced into the state vector for the system when it interacts with the apparatus during measurement.[12] This version of the popular picture argues that the microscopic quantum state of the apparatus is not known, and is not reproducible from one repetition of the measurement to the next; even if we "reset" the apparatus for each repetition of the measurement, there will be uncontrollable and unpredictable fluctuations in its microscopic quantum state. When the measured system interacts with this apparatus, the different superposed parts of its state vector acquire different, random phase factors, which make the different parts in the superposition incoherent. But an incoherent superposition of several state vectors is equivalent, on the average, to an ensemble of collapsed state vectors. We can understand this equivalence between an incoherent superposition and an ensemble of collapsed state vectors by means of our simple example of measurement of the spin of an atom in a Stern–Gerlach experiment. The initial state vector of the atom [see Eq. (4)] is a coherent superposition of the spin-up and spin-down states. If the interaction with the apparatus inserts extra, random phase factors into this superposition, we obtain a final state vector

$$|\psi\rangle = \frac{e^{i\alpha_1}}{\sqrt{2}}|+\rangle + \frac{e^{i\alpha_2}}{\sqrt{2}}|-\rangle \tag{12}$$

Such a superposition with random phase factors is called a *mixture*. According to Eq. (12), the expectation value of any arbitrary operator R is then

$$\begin{aligned}\langle\psi|R|\psi\rangle &= \frac{1}{2}\langle+|R|+\rangle + \frac{1}{2}\langle-|R|-\rangle \\ &\quad + \frac{e^{i(\alpha_1-\alpha_2)}}{2}\langle-|R|+\rangle + \frac{e^{i(\alpha_2-\alpha_1)}}{2}\langle+|R|-\rangle \end{aligned}\tag{13}$$

For random phases α_1 and α_2, the factor $e^{i(\alpha_1-\alpha_2)}$ averages to zero when we perform the measurement repeatedly, and therefore the average value of (13) over repeated measurements is simply

$$\langle\psi|R|\psi\rangle = \frac{1}{2}(\langle+|R|+\rangle + \langle-|R|-\rangle) \tag{14}$$

[12] D. Bohm, *Quantum Theory*, Chapter 22, Sections 6–12.

Note that here, as in Eq. (10), the off-diagonal terms have canceled. We therefore, again, reach the conclusion that, on the average, for repeated measurements, the expectation value of R is the same as for the Copenhagen picture.

The cancellation of off-diagonal terms by random phases [Eq. (13)] seems simpler and more straightforward than the cancellation by orthogonality of apparatus states [Eq. (10)]. However, the random-phase scheme suffers from a fatal defect. The phases must ultimately arise from the interactions between the system and the apparatus. If we want to calculate these phases, we must begin with an initial joint system–apparatus state vector, such as in Eq. (5), and we must investigate its time evolution during the interaction. The result will then be a correlated joint system–apparatus state vector, such as in Eq. (7), with extra, random phases. But this state vector *cannot* be factored into a product of a system state vector of the form (12) and some apparatus state vector. Thus, interactions cannot lead to a final state vector of the form of Eq. (12) for the system after the measurement.

Although random phases by themselves do not provide a consistent picture of the collapse, random phases could possibly play an ancillary role in suppressing interference effects in the correlated joint system–apparatus state vector given in Eq. (9). One difficulty with this state vector is that, to the extent that the system–apparatus interaction is reversible, the state vector (9) could possibly evolve back into the initial state vector (5). But if the two terms in the state vector (9) acquire extra, random phase factors $e^{i\alpha_1}$ and $e^{i\alpha_2}$, then the measurement becomes irreversible. Once the system has acquired random phases, we have lost essential information about the state vector, and we cannot reverse the evolution of the state vector in time and return to the initial state. Thus, the picture of random phases provides us with an explicit model of how irreversibility might enter the measurement process.

Subjective Picture. Another proposal for the collapse is that it is produced in the mind of the observer, by the intervention of the observer's consciousness. This notion was first proposed by von Neumann. As we saw in the preceding section, no apparatus governed by the Schrödinger equation can bring about the collapse, and neither can a stack of apparata arranged to check on each other. However, experience tells us that if a human observer is looking at one of the apparata in the stack, he always perceives the apparatus in a definite state. This forces us to accept that the collapse occurs, somehow, no later than in this observed apparatus or, at the most, no later than within the human observer. Since any apparatus is built of atomic or subatomic pieces, it presumably obeys the Schrödinger equation, and is free of collapse. As a last resort, von Neumann therefore suggested that the collapse occurs when the signals from the apparatus register in the observer's consciousness. This picture of the collapse process was adopted by London and Bauer[13] and by Wigner,[14] who saw in it the resolution of the conundrum posed by the *Gedankenexperiment* of Wigner's friend (see Section 12.2). Wigner proposed that the collapse is brought about by some (unknown) nonlinear process whenever the quantum system interacts with the consciousness of an observer.

But this proposal raises some awkward questions. Exactly what is meant by "consciousness"? What level of consciousness is sufficient to bring about collapse? Is human consciousness required, or is that of a cat or of a mosquito sufficient? Some of these questions can be bypassed by postulating that there is only one consciousness (my own) in the entire universe. This is the philosophical doctrine of solipsism. It is logically unexceptionable, but it is viewed with distaste by most physicists, whose scientific training tells them to be cautious about accepting claims made by one observer alone.

Many-Worlds Picture. Another, radically different treatment of the collapse problem is the many-worlds picture of Everett.[15] In this picture, as in the popular picture, there is no collapse, and the state vector evolves according to the Schrödinger equation at all times. But the many-worlds picture differs from the popular picture in that it includes the observer as part of the quantum-mechanical system. Thus, the many-worlds picture eliminates the dividing line (Heisenberg cut) between the observer and the apparatus, whereas the popular picture implicitly retains this dividing line. The interaction between measured system, apparatus, and observer produces a joint state vector consisting of a superposition of correlated joint state vectors. For instance, in the example of the Stern–Gerlach experiment in Section 12.2, the state vector for

[13] F. London and E. Bauer, *La théorie de l'observation en mécanique quantique* (Hermann, Paris, 1939). Translated in *Quantum Theory and Measurement,* edited by J. A. Wheeler and W. H. Zurek.

[14] E Wigner, *Symmetries and Reflections,* p. 183.

[15] H. Everett, Rev. Mod. Phys., **29,** 454 (1957).

the joint atom–apparatus–observer system after the measurement is a correlated joint state vector of the form given in Eq. (8), where we now regard the states $|\text{upper confirmed}\rangle$ and $|\text{lower confirmed}\rangle$ as states of the observer. The many-worlds interpretation insists that such a schizoid superposition with two or more terms, or "branches," with different observer states, is the correct description of the outcome of the measurement. The two terms in the superposition (8) are interpreted as one branch in which the apparatus has detected spin up and the observer has seen the apparatus detect spin up, and one branch in which the apparatus has detected spin down and the observer has seen it detect spin down. Thus, in each branch the state of the observer is consistent with the state of the apparatus, and in each branch the observer is unaware that something different has happened in the other branch, or even that there is another branch.

Although all of the branches exist simultaneously, the cloned observers in the individual branches do not interact,[16] and they remain forever unaware of each other. The cloned observers effectively inhabit separate worlds. Whenever there is a measurement, the history of the world splits into two or more branches, corresponding to the different outcomes of this measurement. Note that in the many-worlds picture, any measurement-like interaction occurring anywhere gives rise to new branches; thus the universe is continually splitting into a myriads of branches, and each of us is continually splitting into myriads of clones, even when the measurements are not being performed in our immediate vicinity.

In the many-worlds picture, we cannot directly interpret the coefficient $1/\sqrt{2}$ in Eq. (8) as probability amplitudes, since there is no external (outside-the-universe) observer who can measure the state of the system. When the observer is in the state $|\text{upper confirmed}\rangle$, he is not aware of the other state, or of the coefficients $1/\sqrt{2}$. So how can he obtain probabilities in measurements? To answer this question, the many-worlds picture examines what happens if the observer performs a sequence of repeated measurements and records (or remembers) the results. Each measurement generates a new branch of the world, and after all the repeated

[16] The matrix element of the Hamiltonian is zero between any two distinct macroscopic states; if this were not so, then the Hamiltonian could produce transitions from one state of the apparatus or observer to the other, that is, it could change what the apparatus has detected or what the observer has seen.

measurements are completed, there are many branches. At the end of each branch sits a clone of the observer with a sequence of results in his memory, for instance a sequence + + − + + − − − + − + + . . . or a sequence + − − + − + − + + + − + ... Everett proved that for almost every one of these many branches, the sequences of +'s and −'s are random.[17] Thus, almost all the cloned observers will decide that they have verified the prediction of quantum mechanics for repeated measurements of the spin. Everett takes this to mean that in a "typical" branch of the world, the predictions of quantum mechanics will be verified; and he assumes that our branch—that is, the branch at the end of which we sit—is a typical world.

12.4 *The Einstein–Podolsky–Rosen Paradox*

The fundamental assumption of the Copenhagen interpretation is that the physical state of an individual system is completely specified by the wavefunction ψ. This fundamental assumption leads to the uncertainty relations, which tell us that the coordinates and momenta are indeterminate, and that causality, in the sense of classical mechanics, is impossible. Physicists brought up in the traditions of classical mechanics found it hard to accept these features of quantum mechanics. The most illustrious and most severe critic of the Copenhagen interpretation was Einstein, who insisted that even in the realm of the atom there must exist precisely defined dynamical variables and strict deterministic behavior. In view of the practical success of quantum theory, Einstein was willing to accept that $|\psi|^2$ gives a probability distribution, but he insisted that this probability distribution must be interpreted as an ensemble distribution, which does not arise from an intrinsic indeterminacy of the dynamical variables, but only from our ignorance of their values.

Over the years, Einstein challenged the completeness assumption of the Copenhagen interpretation by a variety of clever *Gedankenexperimente*. At first, the thrust of these was directed at the uncertainty relation. Einstein wanted to find a counterexample to these uncertainty relations, by contriving some measurement procedure that would simultaneously determine the coordinate and the momentum of a particle. One such *Gedankenexperiment*, proposed by Einstein in a discussion with Bohr at the 1928 Solvay Meeting, was based on the momentum exchange between the incident particle and a slotted plate, such as might be used to demonstrate diffraction effects. Figure 12.2 shows the experimental arrangement (in the arrangement actually examined by Einstein and Bohr, another plate with two slots was placed in tandem with the single-slot plate, but this is an unessential complication). The particle is incident on the plate from the left, suffers diffraction while passing through the slot, and lands at some (unpredictable) position on the screen at the right. The passage through the slot amounts to a measurement of the vertical position, with an uncertainty $\Delta y = a$, the width of the slot. In the usual analysis of this *Gedankenexperiment*, the vertical momentum is calculated from the diffraction angle, which is known from the observed impact point on the screen; obviously,

$$p_y = p \sin\theta \tag{15}$$

The uncertainty in the angle θ is roughly given by the width of the central diffraction peak, $\Delta(\sin\theta) = \lambda/a = h/ap$, which leads to an uncertainty Δp_y:

$$\Delta p_y = p\,\Delta(\sin\theta) \simeq p\frac{h}{ap} = \frac{h}{a} \tag{16}$$

The product of the uncertainties in y and p_y is therefore

$$\Delta y\,\Delta p_y \simeq a\frac{h}{a} = h \tag{17}$$

which is consistent with the uncertainty principle.

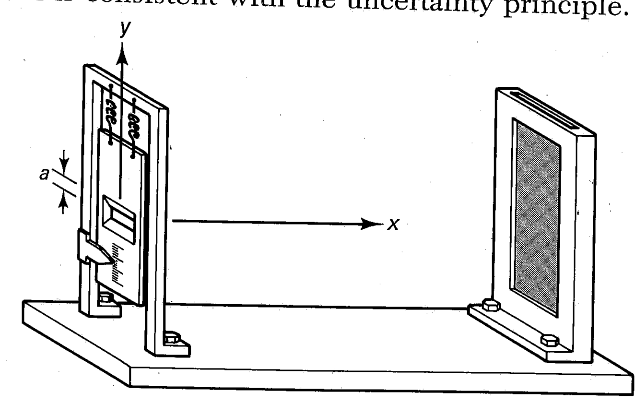

Fig. 12.2 A slotted plate used f...

[17] More generally, if the terms in the superposition have different coefficients c_1 and c_2, the numbers of +'s and −'s in the sequences are weighted in proportion to $|c_1|^2$ and $|c_2|^2$.

However, Einstein proposed to modify this *Gedankenexperiment;* he proposed to measure the momentum p_y, not by the impact point on the screen, but by the recoil suffered by the plate. For this purpose, he suggested that the plate be loosely suspended (by the springs in Fig. 12.2), so it can move and its recoil motion can be determined. Since the recoil momentum of the plate, which is a large macroscopic body for which the laws of classical mechanics ought to hold, can presumably be measured with arbitrary precision, it should be possible to violate the uncertainty relation. But Bohr was quick to notice that the plate is itself subject to the uncertainty principle, and if its momentum is measured with an uncertainty of Δp_y smaller than h/a, then its position will become uncertain by an amount in excess of $h/\Delta p_y = a$, and this means that the momentum and the position of the particle deduced from the position and the momentum of the plate will, again, obey the uncertainty relation.

In fact, it is easy to see that the uncertainty relations are self-consistent: if all bodies obey the uncertainty relations, then a measurement of one body by another can never lead to a result that violates the uncertainty relation. But it is crucial for this consistency that *all* bodies in the universe obey the uncertainty relation; if there were some purely classical body somewhere, with perfectly well-defined position and momentum, then by examining the collision of this body with another body, we could violate the uncertainty relations for the position and momentum of this other body.

Blocked in his direct attacks on the uncertainty relations, Einstein, in a joint venture with Podolsky and Rosen,[18] launched a more subtle attack on the completeness assumption on which the uncertainty relations are based. Their argument, which became known as the Einstein–Podolsky–Rosen (EPR) paradox, begins with the hypothesis that the quantum-mechanical predictions for the results of measurements are correct and tries to show, by means of a *Gedankenexperiment,* that the quantum-mechanical description of the state of the system is incomplete, that is, the system is endowed with physical properties that go beyond those permitted by quantum mechanics.

The EPR paradox hinges on the examination of the joint quantum-mechanical state of two particles that are initially correlated in such perfect way that a measurement performed on one of the

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

particles immediately tells us the state of the other particle, without any need to measure or disturb this other particle. Einstein, Podolsky, and Rosen contemplated two particles of opposite momenta, released initially at one point. But, as remarked by Bohm, the EPR paradox can equally well be stated in terms of two particles of spin $\frac{1}{2}$ in a state of net spin zero, that is, in a state in which their spins are opposite. We will discuss the EPR paradox in terms of such a spin state, because this eases the mathematics and because actual laboratory trials of the experiment have made use of spin states.

Consider two particles of spin $\frac{1}{2}$, such as two protons or two neutrons, in a state of net spin zero. Suppose that the particles are initially close together, but then they move apart to a large distance, while they remain in the original state of net spin zero. Once they are widely separated, we measure the spin of one of these particles. Since the net spin is zero, the measurement of the spin of the first particle immediately allows us to infer the spin of the other particle—it must always be opposite to the spin of the first particle. For instance, if we measure the z component of the spin of the first particle and find $\hbar/2$, then we immediately know that the z component of the spin of the second particle must be $-\hbar/2$. The crucial step in the argument of Einstein, Podolsky, and Rosen is now this: Since our measurement did not touch this second particle, its state before the measurement ought to be the same as after, and therefore this particle ought to have had a well-defined z component of spin even *before* we performed the measurement. But we can now apply the same argument to a measurement of the x component of the spin; if we repeat the experiment and measure the x component instead of the z component, then our argument leads us the conclusion that the second particle ought to have had a well-defined x component of the spin before the measurement. And we can apply the same argument to the y component, and conclude that the second particle ought to have had a well-defined y component of the spin before the measurement. Thus, *all* of the components of the spin of the second particle ought to be well defined, in contradiction to quantum mechanics, which asserts that if one component is well defined, then the others are indeterminate. Accordingly, Einstein, Podolsky, and Rosen claimed that the quantum-mechanical description provided by the state vector cannot be complete. In their view, the state vector must be supplemented or replaced by some extra "hidden variables," and the spin components must be expressed as functions of these hidden variables, so all the spin components are simultaneously well defined.

Note that the crucial step in the EPR argument hinges on the reality of the attributes of the particles and on the locality of the measurement procedure. The spin of the second particle is supposed to exist, in itself, even if we do not measure it; and the measurement performed on the first particle is supposed to produce no effect on the second, distant, particle. Quantum mechanics refutes this paradox by denying both of these suppositions. The Copenhagen interpretation tells us that the particles do not have attributes in themselves, but only in relation to a measurement procedure. Furthermore, it tells us that a measurement performed on one portion of a wavefunction, at one place, affects the entire wavefunction, even its very distant portions.

According to quantum mechanics, the state vectors of the two particles are so intimately intertwined that it makes no sense to speak of the state vector of each individual particle, and it makes no sense to speak of a real value of spin of each. We can see this from the expression for the eigenstate of net spin zero ($s = 0$, $m_s = 0$) formed from the two states of spin one-half:

$$|0,0\rangle = \frac{1}{\sqrt{2}}(|+\rangle|-\rangle - |-\rangle\,|+\rangle) \tag{18}$$

Here, the first bra in each term indicates the spin state of the first particle, and the second bra that of the second [see Eq. (9.95)]. For each individual particle, this state $|0,0\rangle$ is neither an eigenstate of the individual z component of spin, nor even a simple superposition of the eigenstates $|+\rangle$ and $|-\rangle$. There is no definite state vector for the individual particles—only a joint state vector for the system. Thus, it is not surprising that a measurement of the spin of one particle affects the other particle. The measurement of the spin of one particle changes the *whole* state.

We cannot measure one portion of the quantum-mechanical wavefunction and leave the rest undisturbed. When we measure any portion of the wavefunction, the *whole* wavefunction collapses. The strange simultaneous collapse of the spin states of both particles in the EPR *Gedankenexperiment* is no more remarkable than the simultaneous collapse of all parts of the wavefunction of a single particle. When we place a detector in one part of such a single-particle wavefunction, and we find (or do not find) the particle in the volume of the detector, this affects the *entire* wavefunction throughout space. The spin state vectors of the two particles are just as intertwined as the portions of the wavefunction of a single particle. The system of the two spinning particles has a single wavefunction, which happens to depend on two variables. The wavefunction cannot be regarded as consisting of separate, disjoint pieces. Our intuitive expectation that we can measure one portion of a wavefunction without disturbing its distant portions is brought about by an overemphasis on the position representation. Excessive reliance on this representation misleads us into expecting analogies between the behavior of the quantum-mechanical wavefunctions and classical wavefunctions. If we use the abstract state vector, we find it easier to resist this temptation.

Note that the simultaneous collapse of the spin states of the two particles cannot be used to transmit signals from one location to the other. If the measurement of the spin of one of the particles reveals it to be in the state $|+\rangle$, then the other particle collapses into the state $|-\rangle$; but an observer who then measures the spin of this second particle and finds it to be $|-\rangle$ has no way of knowing that this result was enforced by the previous measurement at the other location—he will only know this after he has received a telegram or a letter from the other location informing him of the previous result.

Although quantum mechanics gives a perfectly logical answer to the EPR paradox, it does not give an answer that satisfies our intuition. The EPR paradox shows that the weirdness of quantum mechanics can be found even in systems involving macroscopic distances. The two spinning particles are separated by a large distance, and they do not interact—nevertheless, they form a single, mathematically inseparable system. Quantum mechanics asks us to ignore our intuition and to accept the weird intertwined, nonlocal behavior of the particles in this *Gedankenexperiment*.

12.5 *Bell's Theorem*

In hidden-variable theories, the unpredictable results for a sequence of repeated measurements are attributed to our lack of knowledge of the values of the hidden variables. The average value obtained in a sequence of measurements is taken to equal an average over the unknown (and unknowable?) values of the hidden variables; thus, the average value obtained in a measurement is taken to equal an ensemble average. Of course, the hidden variables and the ensemble used in the averaging are chosen so as to obtain agreement with the expectation values calculated from quantum mechanics. Einstein and other physicists who favored hidden variables took it for granted that the predictions of quantum mechanics can always be duplicated by adopting some sufficiently large set of hidden variables with a sufficiently complicated ensemble distribution. However, in 1964, Bell[19] demonstrated that not all of the subtleties of the probabilistic predictions of quantum mechanics can be duplicated by hidden variables. He demonstrated that the correlations among spin measurements on two particles of spin one-half in a state of zero net spin cannot be duplicated by local hidden variables.

Consider a sequence of measurements of the components of the spins of the two particles along two different directions. The component of the spin of particle 1 is measured along the direction of the unit vector **a**, and the component of the spin of particle 2 is measured along the direction of the unit vector **b**. The results of these measurements are S_{a1} and S_{b2}, respectively, where the spin components S_{a1} and S_{b2} take the usual values $\pm\frac{1}{2}\hbar$. If the directions **a** and **b** are the same, measurements on the quantum-mechanical spin state exhibit a perfect correlation, or rather, a perfect anticorrelation: whenever the measurement of the spin of one of the particles yields the value $S_{a1} = \frac{1}{2}\hbar$, the measurement of the spin of the other particle yields the opposite value $S_{a2} = -\frac{1}{2}\hbar$.

However, if the directions **a** and **b** are not the same, then the anticorrelation of the paired spin measurements will not be perfect. In general, we can characterize the amount of correlation observed in a sequence of a large number of repeated measurements by a *correlation coefficient,* defined as the average value of the product $(4/\hbar^2)S_{a1}S_{b2}$:

$$C(\mathbf{a}, \mathbf{b}) = \left[\frac{4}{\hbar^2} S_{a1} S_{b2}\right]_{av} \tag{19}$$

Note that for each paired spin measurement, the value of $(4/\hbar^2)S_{a1}S_{b2}$ is either +1 or −1; hence, the correlation coefficient is the average of a sequence of +1's and −1's, and necessarily falls within the range $-1 \leq C(\mathbf{a}, \mathbf{b}) \leq +1$. If for each paired spin measurement in our sequence, the observed values of S_{a1} and S_{b2} are exactly opposite, then the correlation coefficient will be $C(\mathbf{a}, \mathbf{b}) = -1$; this characterizes a perfect anticorrelation. If for

[19] J. S. Bell, Physics, **1**, 195 (1964).

each paired spin measurement, the observed values of S_{a1} and S_{b2} are equal, then the correlation coefficient will be $C(\mathbf{a}, \mathbf{b}) = +1$, a perfect correlation. If some pairs of measurements yield opposite spins and some pairs equal spins, then the correlation coefficient will fall between the extreme values $+1$ and -1.

The average in Eq. (19) has been indicated with a square bracket to emphasize that it is calculated directly from the experimental data. Thus, this definition of the correlation coefficient does not hinge on any particular theory of the spin. But, of course, any theory of spin will make a prediction for the value of the correlation coefficient. In quantum mechanics, the average $[(4/\hbar^2)S_{a1}S_{b2}]_{av}$ over the experimental data for a long sequence of repeated measurements is predicted to equal the expectation value $\langle(4/\hbar^2)S_{a1}S_{b2}\rangle$. By evaluating this expectation value in the quantum-mechanical state of net spin zero, we can show that the correlation coefficient is

$$C(\mathbf{a}, \mathbf{b}) = \left\langle \left(\frac{4}{\hbar^2}\right) S_{a1}S_{b2} \right\rangle = -\cos\theta \tag{20}$$

where θ is the angle between the directions of $\mathbf{a}$ and $\mathbf{b}$. Note that for $\theta = 0$, this yields $C(\mathbf{a}, \mathbf{b}) = -1$, as expected. And for $\theta = 90°$ it yields $C(\mathbf{a}, \mathbf{b}) = 0$. This is also expected, since the second spin is always opposite the first, and therefore has equal probabilities for the two possible eigenstates ($S_{b2} = \pm\frac{1}{2}\hbar$) of spin at right angles; consequently, there is no correlation between S_{b2} and S_{a1}.

For a derivation of the formula (20), let us assume that $\mathbf{a}$ is along the $+z$ direction and that $\mathbf{b}$ is in the z-x plane, at an angle θ with the z axis. The spinor for the zero-spin state is

$$\frac{1}{\sqrt{2}}\left(|+\rangle\,|-\rangle - |-\rangle\,|+\rangle\right) \tag{21}$$

The correlation coefficient is the expectation value of $(4/\hbar^2)S_{z1}S_{b2}$ in this state:

$$\begin{aligned} C(\mathbf{a}, \mathbf{b}) &= \frac{1}{2}\left(\langle+|\,\langle-| - \langle-|\,\langle+|\right)\left(\frac{4}{\hbar^2}\, S_{z1}S_{b2}\right)\left(|+\rangle\,|-\rangle - |-\rangle\,|+\rangle\right) \\ &= \frac{2}{\hbar^2}\left(\langle+|\, S_{z1}\,|+\rangle\,\langle-|\, S_{b2}\,|-\rangle - \langle+|\, S_{z1}\,|-\rangle\,\langle-|\, S_{b2}\,|+\rangle \right. \\ &\quad \left. - \langle-|\, S_{z1}\,|+\rangle\,\langle+|\, S_{b2}\,|-\rangle + \langle-|\, S_{z1}\,|-\rangle\,\langle+|\, S_{b2}\,|+\rangle\right) \end{aligned} \tag{22}$$

Here, the second and the third terms are zero, since $\langle+|\, S_{z1}\,|-\rangle = 0$. In the first and fourth terms, the expectation values of S_{z1} are easy to evaluate: $\langle + | S_{z1} | + \rangle = \hbar/2$ and $\langle - | S_{z1} | - \rangle = -\hbar/2$. However, the expectation values of S_{b2} are more difficult, since $|+\rangle$ and $|-\rangle$ are not eigenstates of S_{b2}, but of S_{z2}. To get around this difficulty, we use the vector properties of the spin operator **S** and express S_{b2} as a superposition of S_{z2} and S_{x2}:

$$S_{b2} = S_{z2} \cos \theta + S_{x2} \sin \theta \tag{23}$$

This equation is merely the standard formula for the transformation of the z component of a vector when the z axis is rotated by an angle θ toward the x axis. Thus, the expectation values of S_{b2} in the first and fourth terms of Eq. (22) are

$$\langle + | S_{b2} | + \rangle = \langle + | S_{z2} | + \rangle \cos \theta + \langle + | S_{x2} | + \rangle \sin \theta = \frac{\hbar}{2} \cos \theta + 0 \tag{24}$$

and

$$\langle - | S_{b2} | - \rangle = \langle - | S_{z2} | - \rangle \cos \theta + \langle - | S_{x2} | - \rangle \sin \theta = -\frac{\hbar}{2} \cos \theta + 0 \tag{25}$$

Combining these results, we find

$$C(\mathbf{a}, \mathbf{b}) = \frac{2}{\hbar^2} \left(-\frac{\hbar^2}{4} \cos \theta - \frac{\hbar^2}{4} \cos \theta \right) = -\cos \theta$$

This establishes the validity of the formula (20).

Bell examined the correlation coefficients for measurements of the spin components along three (or more) different directions. He proved that in any local hidden-variable theory the correlation coefficients are restricted by an inequality, and that this inequality is *not* satisfied by the correlation coefficient predicted by quantum mechanics.

Consider three different directions specified by the unit vectors **a**, **b**, and **c**, and suppose that we perform paired measurements of the spin components of the two particles along these directions, taking two directions at a time. We begin with a sequence of paired measurements along the directions **a** (for particle 1) and **b** (for particle 2); then a sequence of measurements along the directions **a** and **c**; and finally, a sequence of measurements along the directions **b** and **c**. The correlation coefficients for these sequences of measurements are $C(\mathbf{a}, \mathbf{b})$, $C(\mathbf{a}, \mathbf{c})$, and $C(\mathbf{b}, \mathbf{c})$, respectively. According to hidden-variable theory, the predicted values of these correlation coefficients are ensemble averages over the hidden variables, with some distribution function (weight function). The number and the kind of hidden variables and the shape of the distribution function depend on the details of the hidden-variable theory. But Bell proved that in any local hidden-variable theory, the correlation coefficients necessarily obey the inequality

$$|C(\mathbf{a}, \mathbf{b}) - C(\mathbf{a}, \mathbf{c})| - C(\mathbf{b}, \mathbf{c}) \leq 1 \tag{26}$$

This result, known as *Bell's theorem*, is independent of the details of hidden-variable theory; it makes no difference how many hidden variables the theory contains, and how their probability distributions are contrived.

For a proof of the theorem, it is convenient to start with an examination of the quantity

$$g = -\frac{4}{\hbar^2} S_{a1} S_{b1} \left(1 - \frac{4}{\hbar^2} S_{b1} S_{c1} \right) \tag{27}$$

which, as we will see, is closely related to the correlation coefficients. In a local hidden-variable theory, a measurement at one place does not affect what happens at another, distant, place; and it follows from this, by the EPR argument, that each particle has well-defined simultaneous spin components. Thus, in such a theory (but not in quantum mechanics) the components of the spin along all three directions **a**, **b**, and **c** are well defined, although they are not necessarily known to us, and they are likely to be different for each repetition of the measurement, because the values of the hidden variables are likely to be different. But for the purposes of Bell's theorem, it suffices that, for each particle 1 in our sequence of repeated measurements, the quantity g has some well-defined value.

Since $(S_{b1})^2 = \hbar^2/4$, we can rewrite g as

$$g = -\frac{4}{\hbar^2} (S_{a1} S_{b1} - S_{a1} S_{c1}) \tag{28}$$

We can relate this expression to the correlation coefficients by taking into account that, in the configuration of net spin zero for the two particles, their spins must be opposite, and hence their spin components along any direction must also be opposite:

$$\begin{aligned} S_{b1} &= -S_{b2} \\ S_{c1} &= -S_{c2} \end{aligned} \tag{29}$$

Substituting these equations into Eq. (28), we obtain

$$g = \frac{4}{\hbar^2}(S_{a1}S_{b2} - S_{a1}S_{c2}) \tag{30}$$

From this we see that the ensemble-average value of g is

$$[g]_{av} = C(\mathbf{a}, \mathbf{b}) - C(\mathbf{a}, \mathbf{c}) \tag{31}$$

Next, we examine the absolute value of $[g]_{av}$. Since the absolute value of the average of g is less than or equal to the average of the absolute value of g, and since $|S_{a1}S_{b2}| = \hbar^2/4$, Eq. (27) leads to

$$|[g]_{av}| \le [|g|]_{av} = \left[\left|\frac{4}{\hbar^2}S_{a1}S_{b1}\left(1 - \frac{4}{\hbar^2}S_{b1}S_{c1}\right)\right|\right]_{av}$$

$$= \left[\left(1 - \frac{4}{\hbar^2}S_{b1}S_{c1}\right)\right]_{av}$$

$$= 1 + \left[\frac{4}{\hbar^2}S_{b1}S_{c2}\right]_{av} \tag{32}$$

But the second term on the right side of the last equation is $C(\mathbf{b}, \mathbf{c})$, and thus

$$|[g]_{av}| \ge 1 + C(\mathbf{b}, \mathbf{c})$$

Combining this with Eq. (31), we immediately obtain the inequality (26) for the correlation coefficients.

The inequality (26), called *Bell's inequality,* is obeyed by every local hidden-variable theory. But this inequality is *not* obeyed by quantum mechanics. For the sake of simplicity, let us consider the special case with **a**, **b**, and **c** in the same plane, say the z-x plane, and with **a** along the $+z$ axis, **b** at an angle of θ with respect to the $+z$ axis, and **c** at an angle of 2θ with respect to the $+z$ axis. The quantum-mechanical correlation coefficients are then [see Eq. (20)]

$$C(\mathbf{a}, \mathbf{b}) = -\cos\theta$$

$$C(\mathbf{a}, \mathbf{c}) = -\cos 2\theta$$

$$C(\mathbf{b}, \mathbf{c}) = -\cos\theta$$

Thus, the quantum-mechanical expression for the left side of Bell's inequality is

$$|C(\mathbf{a}, \mathbf{b}) - C(\mathbf{a}, \mathbf{c})| - C(\mathbf{b}, \mathbf{c}) = |-\cos\theta + \cos 2\theta| + \cos\theta$$

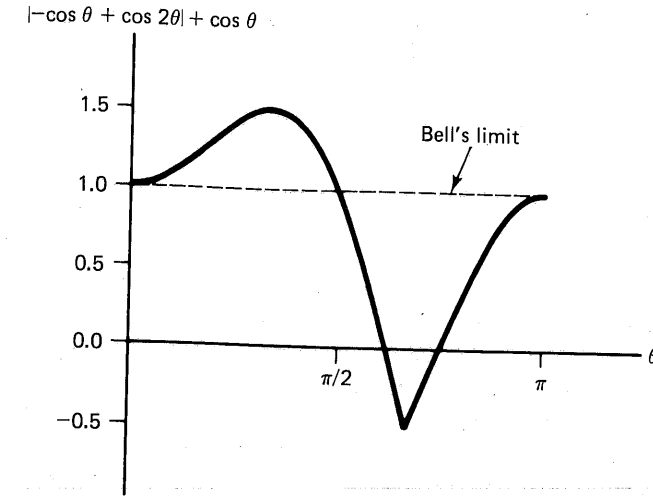

Fig. 12.3 Plot of $|-\cos\theta + \cos 2\theta| + \cos\theta$ vs. θ. The dashed line is the upper limit set by Bell's inequality.

Figure 12.3 shows a plot of this expression as a function of θ. We see that in the interval $0 < \theta < \pi/2$, the quantum-mechanical result exceeds the limit set by Bell's inequality. Thus, the quantum-mechanical result is inconsistent with all local hidden-variable theories.

Bell's inequality provides us with a way to discriminate experimentally between the predictions of quantum mechanics and those of local hidden-variable theories. Before Bell's theorem, such a discrimination was thought to be nearly impossible, since hidden-variable theories are designed to mimic the results of quantum mechanics as best they can. After Bell's theorem, several experiments were performed to test the quantum-mechanical prediction for correlation coefficients vs. the hidden-variable prediction. Most of these experiments studied the correlations of the polarizations of paired photons of net spin zero emitted by an atom. Inequalities similar to the Bell inequality (26) can be derived for the correlations of the polarizations of such photons. The inequalities tested in these experiments actually were generalized Bell inequalities, that make allowances for the less-than-perfect counting efficiencies of the photon polarizers and detectors, and for the experimental errors introduced by the apparatus. One weakness of the original Bell inequality is that, like the EPR argument, it relies on an ideal apparatus capable of measuring a spin component with absolute precision—we can use the relation $S_{a2} = -S_{a1}$ to find the spin of the second particle from a measurement performed on the first particle only if the first measurement is *absolutely precise*. The generalized Bell inequalities do not make such extreme demands of the apparatus.

The pairs photons used in experimental tests of Bell's inequality are emitted in a cascade process, in which an atom quickly makes two successive transitions from an upper state of angular momentum $j = 0$, to an intermediate state of $j = 1$, and finally to a lower state of $j = 0$. Since the initial and the final states have angular momenta zero, the net angular momentum carried away by the two photons emitted in these two transitions must be zero, and their polarizations are therefore perfectly correlated, like the spins of the two particles we considered above.[20] Experiments have also been performed with pairs of protons obtained by proton–proton scattering at low energies. From our discussion of partial waves in scattering (see Section 11.5), we know that when a low-energy proton is incident on a target proton, most of the scattering is contributed by the partial wave of zero orbital angular momentum. This means the protons are in a symmetric orbital state, and the Pauli exclusion principle demands that they must then be in an antisymmetric spin state, that is, a state of net spin zero.

With one exception, attributed to systematic experimental errors, all these experiments found correlations that agreed with the predictions of quantum mechanics and that exceeded the upper limit demanded by Bell's inequality. In the best of these experiments, by A. Aspect and his associates at the Institut d'Optique d'Orsay,[21] the experimental results exceeded the Bell inequality by more than 40 standard deviations. These experimental results conclusively rule out local hidden-variable theories. Thus, nature tells us that the weird nonlocal character of quantum mechanics brought out in the EPR paradox must be accepted. This does not necessarily mean that we have to accept quantum mechanics. We could still think of contriving a *nonlocal* hidden-variable theory, even though such a theory would be somewhat pointless, since, according to the EPR argument, the main rationale for a hidden-variable theory is the attainment of locality.

[20] The polarization of a photon is in direct correspondence to its spin state. The two states of circular polarization correspond to spin parallel and spin antiparallel to the momentum of the photon. The two states of plane polarization correspond to superpositions of these spin states.

[21] For a review of recent experimental results, see the article by A. Aspect and P. Grangier in *Quantum Concepts in Space and Time,* edited by R. Penrose and C. J. Isham. For a comprehensive review of earlier experiments, see J. F. Clauser and A. Shimony, Rep. Prog. Phys., **41,** 1881 (1978).

In a nonlocal theory, a measurement at one place can affect a measurement at another place. In such a theory, some hidden variables might generate a (nonlocal) influence between the detectors, so the orientation of the axis of polarization of detector 1 alters the behavior of detector 2, in such a way that the measured correlations match those predicted by quantum mechanics. However, a modification of the two-photon correlation experiment by Aspect established that if this influence exists, it must propagate from one detector to the other at a speed exceeding the speed of light. In the modified experiment, the polarizations of the detectors were independently switched from one direction to another in a time shorter than the light travel time between detectors. The experimental results were, again, in agreement with quantum mechanics, and in disagreement with Bell's inequality. This establishes that any nonlocal influence exerted by one detector on the other would have to proceed via superluminal action-at-a-distance. Furthermore, this action-at-a-distance would have to be contrived in such a way that an experimenter can never use it to send deliberate signals, which would violate causality. These features of a nonlocal hidden-variable theory would be even more weird than the features of quantum mechanics.

PROBLEMS

1. In one attempt at finding a counterexample to the energy–time uncertainty relation, Einstein proposed that a closed box full of radiation be equipped with a shutter operated by a clock (see Fig. 12.4). The box is first weighed precisely with a spring balance, then the clock opens the shutter for an interval Δt and releases a photon, and then the box is again weighed precisely. This would seem to permit a precise determination of the energy of the photon ($\Delta E = 0$), in contradiction with the uncertainty relation $\Delta E\, \Delta t \geq \hbar/2$. Bohr refuted this counterexample by noting that the vertical uncertainty in the position of the box during the weighing leads to an uncertainty in the rate of the clock, via the gravitational time-dilation effect of Einstein's theory of general relativity. A simpler refutation can be constructed by taking into account that the *clock and the shutter* are a quantum system, which is subject to the energy–time uncertainty relation.

   (a) Prove that the clock and shutter cannot be in an energy eigenstate. (Hint: Consider that the hands of the clock and the shutter have time-dependent positions.)