16 Measurement and interpretation

It will have become evident to the reader that quantum mechanics is a most successful theory. This has been illustrated in many applications described in this book. In fact, quantum mechanics has never been shown to fail, although some specific applications may be beyond the reach of our calculational ability. Nevertheless quantum theory has features which seem strange compared with classical Newtonian mechanics and which some physicists have found difficult to accept. In this chapter, we shall discuss briefly some of the conceptual difficulties and experiments designed to resolve them, giving references for further study.

16.1 HIDDEN VARIABLES?

One of the most fundamental characteristics of quantum theory is its lack of determinism. When a single measurement of an observable A is made, the result is one of the eigenvalues a_n of A. However, unless the system is in an eigenstate of A it is absolutely impossible to predict in any particular measurement which of the eigenvalues a_n will be obtained. All that can be predicted is the frequency of obtaining the eigenvalue a_n when the measurement is repeated many times on a set of identically prepared systems. This lack of determinism is quite different from anything in classical physics. It is, of course, true that many situations arise in classical physics which can only be described statistically, for example the motion of the molecules in a gas. This classical indeterminism arises merely from our lack of detailed knowledge about the positions and velocities of each molecule. It is believed that although unobservable in practice, in fact each molecule at a given time has a well defined position and velocity, and the results of future measurements of the position and velocity of each molecule could, in principle, be determined [1]. Such considerations about classical systems have led to the supposition that quantum mechanics is an incomplete theory in that there are other variables, called '*hidden variables*', of which we are not directly aware, but which are required to determine the system completely. These hidden

[1] In practice, even the smallest errors made in the initial positions and velocities will cause the future positions and velocities to be indeterminable after a short interval of time, and the behaviour of the gas becomes 'chaotic'.

variables are postulated to behave in a classical deterministic manner, the apparent indeterminism exhibited by experiment arising from our lack of knowledge of the hidden sub-structure of the system studied. Thus apparently identical systems are perhaps characterised by different values of one or more hidden variables, which determine in some way which particular eigenvalues are obtained in a particular measurement.

It is of historical interest that de Broglie's original interpretation of the wave function falls into the class of hidden-variable theories. He supposed the wave function to be a physically real field propagating in space and coupled to an associated particle which has both a well defined position and momentum. The coupling between the particle and the 'pilot wave' gives rise to the observed diffraction phenomena. A deterministic theory of this type was elaborated in 1952 by D. Bohm who was able to account for the diffraction and interference shown in particle scattering, obtaining exactly the same results as those given by quantum mechanics. However, this model containing both waves and particles as separate, but connected, entities is extremely complex. To most people's minds it has even stranger features than those of quantum mechanics, and most physicists would reject it on the grounds of 'Occam's razor' [2]. Perhaps the least acceptable feature of Bohm's model, for those seeking an underlying classical mechanism to quantum theory is its non-locality. For instance, in the analysis of the two-slit experiment (see Chapter 2) using Bohm's model there is a force acting on a particle traversing one slit which is changed instantly if the second slit is opened, or closed. Such theories, in which an action at one place is transmitted instantaneously (or at least faster than the speed of light) to alter the situation at another, are called *non-local*. As we shall see, quantum mechanics is a non-local theory, but non-locality is generally not considered to be an acceptable feature of a classical theory. It might be thought that a sufficiently ingenious hidden variable theory might be constructed, which is both deterministic and local. However J. S. Bell in 1965 was able to lay down conditions that all deterministic local theories must satisfy. As we shall see in the next two sections these conditions are found to be violated by experiment.

16.2 THE EINSTEIN–PODOLSKY–ROSEN PARADOX

The most famous physicist to question the completeness of quantum theory was A. Einstein. In 1935, in collaboration with N. Rosen and B. Podolsky, he proposed the following criteria as the basis of any acceptable theory:

1. The quantities concerned in the theory should be 'physically real', physical reality being defined as follows: 'If, without in any way disturbing a system, we can predict with certainty the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.'

[2] William of Occam (1290?–1349). His maxim states that 'It is vain to do with

2. The theory should be local, i.e. there is no action at a distance in nature.

Einstein, Podolsky and Rosen were able to give an example of a quantum mechanical system which did not satisfy these conditions and concluded that the quantum description of nature was incomplete. Following Bohm, we shall investigate a simpler situation than that proposed by Einstein and his collaborators, but which exhibits similar features. Consider a system with total spin $S=0$ which splits into two identical particles, 1 and 2, each of spin $1/2$. When the particles are well separated let us measure the component of the spin of particle 1 parallel to some direction, which we shall define as the z-axis, Since the particle has spin $1/2$, either the result $+\hbar/2$ or the result $-\hbar/2$ is obtained. Suppose that in a particular experiment the result $+\hbar/2$ is found, then since the total spin of the two-particle system is zero, a measurement of the z-component of the spin of particle 2 has to produce the result $-\hbar/2$. Subsequent measurements of the component of the spin of particle 2 parallel to the x-axis produce the results $+\hbar/2$ and $-\hbar/2$ on a fifty–fifty basis. Similarly, if the x-component of the spin of particle 1 is measured and found to be $\hbar/2$, the x-component of the spin of particle 2 must be $-\hbar/2$ and the z-component of the spin of particle 2 will be found to be $+\hbar/2$ or $-\hbar/2$ on a fifty–fifty basis. We see that the act of measuring a component of the spin of particle 1 has altered the result obtained in measuring a component of the spin of the other particle. This alteration takes place instantly no matter how far apart particles 1 and 2 may be. Thus the quantum description does not obey the conditions 1 and 2. This fact is often known as the Einstein–Podolsky–Rosen paradox. However, N. Bohr rejected the idea that the result is paradoxical, taking the view that in condition 1 physical reality can only refer to situations in which the experimental arrangement is completely specified and suggesting that this is not the case because the system is disturbed from the outset by the experimenters' decision to measure the x-component rather than the z-component of the spin of particle 1.

The quantum situation can be contrasted with what would be observed if spin were a classical variable. It would remain true that the spin components of particles 1 and 2 would be found to be equal and opposite, because the total spin is zero. However, this would be the case because the spin vectors have definite values and directions right from the beginning when the state was created, and the act of measurement on particle 1 would not change the state of particle 2 in any way. Now it might be that the quantum results could be explained by this common-cause argument. For example, there might be a classical hidden variable (or variables), the value of which was determined when the spin-zero system was created and which subsequently determined the experimental results. However, the experimental violation of Bell's theorem, which we will now discuss, shows that this explanation is in fact incorrect.

16.3 BELL'S THEOREM

Consider again the spin-zero system of the previous paragraph, composed of two spin-1/2 particles. Choosing a particular direction z, the spin wave function is (see [6.301])

$$\chi(1,2)=\frac{1}{\sqrt{2}}[\alpha(1)\beta(2)-\beta(1)\alpha(2)]. \qquad [16.1]$$

The component of the spin of particle 1 along a certain direction specified by the unit vector $\hat{\mathbf{a}}$ is $S_a(1)$ where

$$S_a(1)=\mathbf{S}(1).\hat{\mathbf{a}}=\boldsymbol{\sigma}(1).\hat{\mathbf{a}}(\hbar/2) \qquad [16.2]$$

where $\boldsymbol{\sigma}(i)$ are the Pauli spin matrices for particle i. The result of a single measurement of $S_a(1)$ is either $\hbar/2$ or $-\hbar/2$. These values occur in a series of measurements on a fifty–fifty basis, so that the expectation value of $S_a(1)$ is zero. This result also follows by calculating $\langle S_a(1)\rangle$ using the wave function [16.1].

If both the components of the spin of particle 1 along $\hat{\mathbf{a}}$ and that of particle 2 along $\hat{\mathbf{b}}$ are measured jointly, the corresponding observable is $(\boldsymbol{\sigma}(1).\hat{\mathbf{a}}\boldsymbol{\sigma}(2).\hat{\mathbf{b}})\hbar^2/4$. The average result of a series of joint measurements of $S_a(1)$ and $S_b(2)$ is the expectation value of this operator. Denoting the expectation value of $(\boldsymbol{\sigma}(1).\hat{\mathbf{a}}\boldsymbol{\sigma}(2).\hat{\mathbf{b}})$ by $E(\hat{\mathbf{a}},\hat{\mathbf{b}})$ we find using [16.1] that

$$E(\hat{\mathbf{a}},\hat{\mathbf{b}})=\langle\chi|\boldsymbol{\sigma}(1).\hat{\mathbf{a}}\boldsymbol{\sigma}(2).\hat{\mathbf{b}}|\chi\rangle=-\hat{\mathbf{a}}.\hat{\mathbf{b}}$$
$$=-\cos\phi \qquad [16.3]$$

where ϕ is the angle between $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$. If $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ are in the same direction, the value of $E(\hat{\mathbf{a}}.\hat{\mathbf{b}})$ is -1, which just tells us that the two spin components are both of magnitude $\hbar/2$ and of opposite sign.

Now suppose there is a hidden variable [3] λ which specifies the state of the system completely and which determines the values of the quantum variables obtained in a given experiment. The dynamical evolution of the hidden variable is considered to be subject to conditions 1 and 2 of the previous section. Each spin-zero system has a definite value of λ. When a large number of such systems are prepared identically let a fraction $p(\lambda)$ have values of the hidden variable between λ and $\lambda+\mathrm{d}\lambda$, normalised so that

$$\int p(\lambda)\,\mathrm{d}\lambda=1, \qquad p(\lambda)\geqslant 0. \qquad [16.4]$$

Let us denote the result of a measurement of the spin component $S_a(1)$ of particle 1 along a direction $\hat{\mathbf{a}}$ by $A(\hat{\mathbf{a}},\lambda)\hbar/2$ and of a measurement of the spin component $S_b(2)$ of particle 2 along a direction $\hat{\mathbf{b}}$ by $B(\hat{\mathbf{b}},\lambda)\hbar/2$, where $A(\hat{\mathbf{a}},\lambda)$ and $B(\hat{\mathbf{b}},\lambda)$ can only take the values ± 1. Since the overall spin is zero, we must

[3] There may be many such variables but this does not alter the essential nature of the discussion.

have

$$A(\hat{\mathbf{a}},\lambda) = -B(\hat{\mathbf{a}},\lambda). \qquad [16.5]$$

In a 'complete' theory $A(\hat{\mathbf{a}},\lambda)$ and $B(\hat{\mathbf{b}},\lambda)$ are entirely specified by the value of λ. In a local theory, if the measurements are made at points which are well separated, $A(\hat{\mathbf{a}},\lambda)$ can only depend on λ and $\hat{\mathbf{a}}$ but not on $\hat{\mathbf{b}}$. Similarly, $B(\hat{\mathbf{b}},\lambda)$ is independent of $\hat{\mathbf{a}}$, so that the result of a measurement of S_a and S_b jointly is the uncorrelated product of $A(\hat{\mathbf{a}},\lambda)\hbar/2$ and $B(\hat{\mathbf{b}},\lambda)\hbar/2$. The average of a series of joint measurements is $(\hbar^2/4)\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{b}})$, where

$$\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{b}}) = \int p(\lambda)A(\hat{\mathbf{a}},\lambda)B(\hat{\mathbf{b}},\lambda)\,\mathrm{d}\lambda. \qquad [16.6]$$

This is in general non-factorisable and is correlated despite the fact that the individual product $A(\hat{\mathbf{a}},\lambda)B(\hat{\mathbf{b}},\lambda)$ is not. However, the correlation is through the common cause, the theory remaining a local one. Note that to distinguish the averages calculated by the local hidden variable theory from the quantum result [16.3], we have employed a script $\mathscr{E}$. We notice that if $\hat{\mathbf{a}} = \hat{\mathbf{b}}$, then

$$\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{a}}) = E(\hat{\mathbf{a}},\hat{\mathbf{a}}) = -1 \qquad [16.7]$$

because of the conditions [16.4] and [16.5].

Now consider a joint measurement of the component of the spin of 1 along $\hat{\mathbf{a}}$ and of 2 along $\hat{\mathbf{c}}$, where $\hat{\mathbf{c}}$ is a different direction from $\hat{\mathbf{b}}$. We find that

$$\begin{aligned}\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{b}}) - \mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{c}}) &= \int p(\lambda)\{A(\hat{\mathbf{a}},\lambda)B(\hat{\mathbf{b}},\lambda) - A(\hat{\mathbf{a}},\lambda)B(\hat{\mathbf{c}},\lambda)\}\,\mathrm{d}\lambda \\ &= -\int p(\lambda)A(\hat{\mathbf{a}},\lambda)A(\hat{\mathbf{b}},\lambda)\{1 + A(\hat{\mathbf{b}},\lambda)B(\hat{\mathbf{c}},\lambda)\}\,\mathrm{d}\lambda \end{aligned} \qquad [16.8]$$

where we have used [16.5], together with the fact that $\{A(\hat{\mathbf{b}},\lambda)\}^2 = 1$.

The inequality

$$|\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{b}}) - \mathscr{E}(\hat{\mathbf{a}}.\hat{\mathbf{c}})| < \int p(\lambda)\{1 + A(\hat{\mathbf{b}},\lambda)B(\hat{\mathbf{c}},\lambda)\}\,\mathrm{d}\lambda \qquad [16.9]$$

or

$$|\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{b}}) - \mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{c}})| < 1 + \mathscr{E}(\hat{\mathbf{b}},\hat{\mathbf{c}}) \qquad [16.10]$$

then follows because A only takes the values $+1$ or -1, and $p(\lambda)$ is always non-negative. This is one of a family of inequalities deduced by Bell, which must be satisfied by any real local theory. It is easy to find directions $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$ and $\hat{\mathbf{c}}$ for which the quantum expectation value violates [16.10], so we can conclude that *quantum mechanics is inherently a non-local theory*. For example, let the angle between $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}$ be $2\pi/3$ and let $\hat{\mathbf{b}}$ lie in the same plane as $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}$ making an angle $\pi/3$ with each of them. Then from [16.3] we see that the quantum expectation values satisfy the relations

$$|E(\hat{\mathbf{a}},\hat{\mathbf{b}}) - E(\hat{\mathbf{a}},\hat{\mathbf{c}})| = |-\cos(2\pi/3) + \cos(\pi/3)| = 1 \qquad [16.11]$$

and

$$1 + E(\hat{\mathbf{b}},\hat{\mathbf{c}}) = [1 - \cos(\pi/3)] = \frac{1}{2} \qquad [16.12]$$

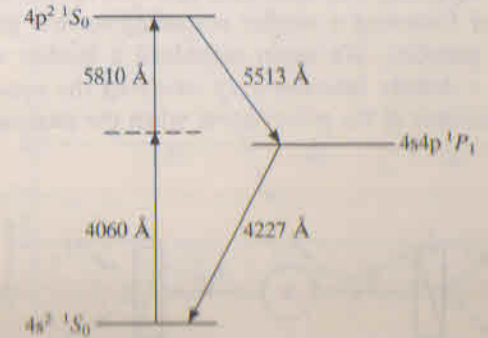
with the consequence that

$$|E(\hat{\mathbf{a}},\hat{\mathbf{b}}) - E(\hat{\mathbf{a}},\hat{\mathbf{c}})| > 1 + E(\hat{\mathbf{b}},\hat{\mathbf{c}}) \qquad [16.13]$$

which violates the inequality [16.10].

Experimental verification

Now the question arises as to whether the correlations observed experimentally agree with the quantum result and violate the locality condition, or whether Bell's inequality holding for real local theories is satisfied. Since 1972 a number of experiments of increasing accuracy have been carried out and analysed using an elaboration of the preceding theory which allows for the less than perfect efficiencies of the detectors. Rather than using particles the experiments have measured the states of polarisation of a two-photon system in which the photons propagate in opposite directions. In the experiments of A. Aspect and co-workers carried out in 1982, the $4p^2\,{}^1S_0$ level of calcium is populated by two-photon excitation, using two lasers – a krypton laser at 4060 Å and a tunable dye laser at 5810 Å. This level decays back into the $4s^2\,{}^1S_0$ level by cascading through the $4s4p\,{}^1P_1$ level (see Fig. 16.1) with the emission of two photons of wavelengths 5513 Å and 4227 Å. Since the total angular momentum is zero in both the initial $4p^2\,{}^1S_0$ and final $4s^2\,{}^1S_0$ levels, the overall angular momentum of the two-photon system is also zero. As a consequence, if the photons propagate in



16.1 Study of two-photon correlations. The $4p^2\,{}^1S_0$ level of calcium is populated by two-photon excitation. The correlations between the polarisation of two photons emitted in the decay back to the $4s^2\,{}^1S_0$ level, via the $4s4p\,{}^1P$ level, are studied in the experiment of Aspect and co-workers.

opposite directions parallel to the z-axis, the polarisation part of the wave function is of the form

$$\psi(1,2)=\frac{1}{\sqrt{2}}[u(1)u(2)+v(1)v(2)] \qquad [16.14]$$

where $u(i)$ represents a photon linearly polarised along the x-axis and $v(i)$ a photon linearly polarised along the y-axis. The polarisations are thus completely correlated and similar arguments apply to those we outlined for the spin one-half particles. The principle of the experiment is illustrated in Fig. 16.2. Two photons, $h\nu_1$ and $h\nu_2$, emerge in opposite directions parallel to the z-axis from the source S, and are detected by the linear polarisation analysers I and II. The analysers I and II are orientated parallel to the unit vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$, respectively, which lie in xy planes normal to the direction of propagation. These analysers record the result $+1$ if a photon is found to be linearly polarised parallel to $\hat{\mathbf{a}}$ (or $\hat{\mathbf{b}}$) and the result -1 when the polarisation is normal to $\hat{\mathbf{a}}$ (or $\hat{\mathbf{b}}$). Four coincidence rates, $N_{++}(\hat{\mathbf{a}},\hat{\mathbf{b}})$, $N_{--}(\hat{\mathbf{a}},\hat{\mathbf{b}})$, $N_{+-}(\hat{\mathbf{a}},\hat{\mathbf{b}})$ and $N_{-+}(\hat{\mathbf{a}},\hat{\mathbf{b}})$ are measured, where $N_{++}(\hat{\mathbf{a}},\hat{\mathbf{b}})$ is the rate at which results $+1$ are obtained in I oriented parallel to $\hat{\mathbf{a}}$ and II orientated parallel to $\hat{\mathbf{b}}$ simultaneously. The rates N_{--}, N_{+-} and N_{-+} are defined similarly.

A correlation coefficient $E(\hat{\mathbf{a}},\hat{\mathbf{b}})$ can be defined in terms of the four coincidence rates as

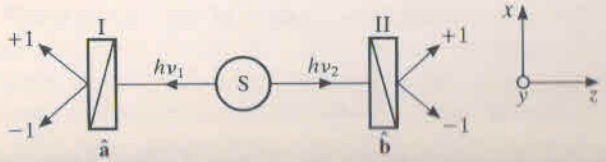
$$E(\hat{\mathbf{a}},\hat{\mathbf{b}})=\frac{N_{++}(\hat{\mathbf{a}},\hat{\mathbf{b}})+N_{--}(\hat{\mathbf{a}},\hat{\mathbf{b}})-N_{+-}(\hat{\mathbf{a}},\hat{\mathbf{b}})-N_{-+}(\hat{\mathbf{a}},\hat{\mathbf{b}})}{N_{++}(\hat{\mathbf{a}},\hat{\mathbf{b}})+N_{--}(\hat{\mathbf{a}},\hat{\mathbf{b}})+N_{+-}(\hat{\mathbf{a}},\hat{\mathbf{b}})+N_{-+}(\hat{\mathbf{a}},\hat{\mathbf{b}})} \qquad [16.15]$$

The coefficient $E(\hat{\mathbf{a}},\hat{\mathbf{b}})$ can be calculated using quantum mechanics by an argument rather similar to that used in obtaining the spin-1/2 result [16.3]. In the present case it is found that

$$E(\hat{\mathbf{a}},\hat{\mathbf{b}})=\cos 2\phi \qquad [16.16]$$

where ϕ is the angle between $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$.

The corresponding correlation coefficient $\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{b}})$ for a real local hidden variable theory is found by following a similar reasoning to that given above for the case of spin-1/2 particles. We again introduce a hidden variable (or set of variables) λ, and a density function $p(\lambda)$ satisfying the condition [16.4]. The result of a measurement of the polarisation when the analyser I is parallel to $\hat{\mathbf{a}}$



16.2 Schematic diagram of an experiment to measure the correlation of linear polarisation when two photons $h\nu_1$ and $h\nu_2$ are emitted from a source S. (*By courtesy of A. Aspect.*)

is denoted by $A(\hat{\mathbf{a}},\lambda)$ where $A(\hat{\mathbf{a}},\lambda)$ can take the values $+1$ or -1, while $B(\hat{\mathbf{b}},\lambda)$ is the corresponding quantity for measurements by the analyser II. The correlation function $\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{b}})$ is then of exactly the same form as [16.6], that is

$$\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{b}})=\int p(\lambda)A(\hat{\mathbf{a}},\lambda)B(\hat{\mathbf{b}},\lambda)\,\mathrm{d}\lambda. \qquad [16.17]$$

Out of the several Bell's inequalities which can be deduced, the one chosen in the experiment of Aspect and co-workers is based on a quantity S defined as

$$S=\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{b}})-\mathscr{E}(\hat{\mathbf{a}},\hat{\mathbf{b}}')+\mathscr{E}(\hat{\mathbf{a}}',\hat{\mathbf{b}})+\mathscr{E}(\hat{\mathbf{a}}',\hat{\mathbf{b}}') \qquad [16.18]$$

where $\hat{\mathbf{a}},\hat{\mathbf{a}}'$ and $\hat{\mathbf{b}},\hat{\mathbf{b}}'$ represent two different orientations of the analysers I and II, respectively. Provided A and B can only take the values $+1$ or -1 and that [16.4] and [16.17] hold, it is not difficult to obtain the Bell inequality

$$-2\leqslant S\leqslant 2. \qquad [16.19]$$

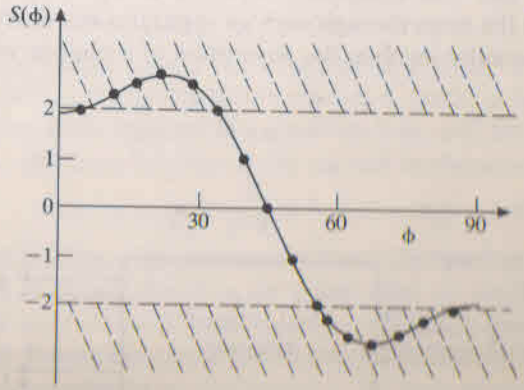
On the other hand if the correlations $\mathscr{E}$ in [16.18] are replaced by the corresponding quantum correlations E, defined by [16.16], $|S|$ can exceed 2 for various orientations of the four vectors $\hat{\mathbf{a}}$, $\hat{\mathbf{a}}'$, $\hat{\mathbf{b}}$ and $\hat{\mathbf{b}}'$. In the 1982 experiments of A. Aspect and co-workers, the orientations were chosen such that

$$\hat{\mathbf{a}}.\hat{\mathbf{b}}=\hat{\mathbf{b}}.\hat{\mathbf{a}}'=\hat{\mathbf{a}}'.\hat{\mathbf{b}}'=\cos\phi \qquad [16.20a]$$

and

$$\hat{\mathbf{a}}.\hat{\mathbf{b}}'=\cos 3\phi. \qquad [16.20b]$$

The experimental values of S were then obtained for various angles ϕ between $0°$ and $90°$. The results are shown in Fig. 16.3 together with the predictions of quantum mechanics (the solid curve). The experimental results violate Bell's

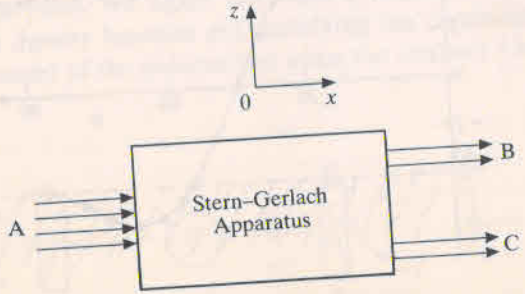


16.3 Experimental data obtained for the correlation quantity $S(\phi)$ by A. Aspect, P. Grangier and G. Roger. The solid curve represents the prediction of quantum mechanics and the shaded areas are those in which Bell's inequality [16.19] is violated. (*By courtesy of A. Aspect.*)

Further experiments, notably by Perrie and collaborators [illegible] violation of Bell's inequalities. Hence it can be concluded that no local hidden variable theory underlies quantum mechanics. As we pointed out earlier, it is possible to construct non-local hidden variable theories which lead to the same experimental consequences as quantum mechanics, but there is no compelling reason to introduce such complexities.

16.4 THE PROBLEM OF MEASUREMENT

The process of measurement differs in quantum mechanics in several respects from that in classical mechanics. For example, we have already seen that *complementary* pairs of variables exist, such as momentum and position which can only be measured simultaneously up to limits imposed by the uncertainty relations. The word measurement itself is used in two rather different senses in quantum mechanics, which it is important to distinguish. First, there are those measurements, often called *ideal measurements*, which can be repeated immediately. For example, if a beam of atoms, each of spin one-half, is passed through a Stern–Gerlach apparatus (see Chapter 1), orientated to split the beam in the z-direction, one of the two emerging beams (see Fig. 16.4) contains only atoms with $S_z = +\hbar/2$ and the other only atoms with $S_z = -\hbar/2$. This measurement, which is better called *state preparation*, can be immediately repeated, by passing, for instance, the $S_z = +\hbar/2$ beam through a second Stern–Gerlach apparatus, orientated in the same way as the first. If this is done, all the particles will emerge deflected in the direction corresponding to $S_z = \hbar/2$ and none in that corresponding to $S_z = -\hbar/2$. The passing of a beam of atoms through a Stern–Gerlach apparatus prepares an ensemble of atoms in a known state and is not really a measurement in the usual sense. To measure whether or not an atom has spin component $S_z = \hbar/2$ or $S_z = -\hbar/2$, it is not only necessary to pass the atom through such an apparatus but also to detect it. The act of detection usually involves the absorption of a photon, or the capture of



16.4 An example of state preparation. A Stern–Gerlach apparatus is orientated to split an unpolarised beam of spin one-half atoms A, in the z-direction. The upper energy beam B contains only atoms with $S_z = +\hbar/2$ and the lower beam C contains only atoms with $S_z = -\hbar/2$.

an ensemble of objects is prepared in a reproducible state, and [illegible] which involve the detection of the object together with the determination of one or more of its properties, within certain limits – for example, energy, position, momentum, and so on. The former process allows one to say something about the future behaviour of the system, while the latter gives information about its immediate past.

The relationship between the wave function Ψ, or the corresponding state vector $|\Psi\rangle$, and the results of measurements have been discussed in Chapter 5. It is all-important to recognise that a knowledge of $|\Psi\rangle$ does not allow us to predict the results of a measurement on a single object of an observable A, except in the special case for which $|\Psi\rangle$ is an eigenstate of A belonging to some eigenvalue a_n, in which case the value a_n will be obtained. Apart from this case, $|\Psi\rangle$ only determines the frequency of obtaining a particular eigenvalue when the measurement is repeated a large number of times on a set of identically prepared objects or systems (called an ensemble of systems). It is therefore incorrect to associate a wave function with a single object (with the exception noted). On the contrary, $|\Psi\rangle$ provides us with the maximum possible information about an ensemble of identically prepared systems. It is important to be clear that the individual systems do not possess precise (but unknown) values of the variable A before a measurement is made. This would amount to the kind of hidden variable theory that we have seen to be untenable; rather a member of the ensemble cannot be said to possess a value for A (unless $|\Psi\rangle$ is an eigenvector of A) until a measurement has been performed. This view, although apparently correct, gives rise to some difficulties in interpretation, as Schrödinger pointed out in his famous cat paradox.

Schrödinger's cat

In this thought experiment, a cat is placed in a closed box together with a radioactive atom and a device which releases a deadly poison on the decay of the atom. If the wave function of a dead cat is ψ_d and that of a live cat is ψ_l, after a time t, the wave function of the cat will be the superposition

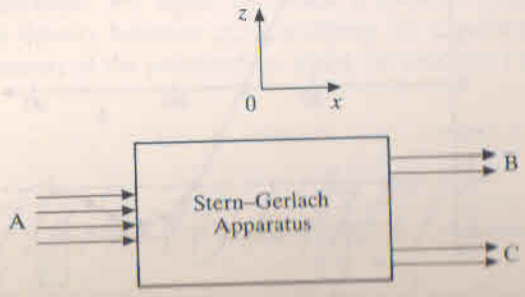
$$\Psi(t) = \psi_l \exp(-t/2\tau) + \psi_d(1 - \exp(-t/2\tau)) \qquad [16.21]$$

where τ is the half-life of the radioactive atom. On opening the box at a certain time, the cat is clearly found to be either alive or dead. This is an act of measurement which forces the system into one of the eigenstates of being alive or being dead. It is an example of the reduction or collapse of the wave function on measurement, referred to in Chapter 2. All sorts of curious questions now arise. Before the box is opened the cat is apparently neither alive or dead, but in a superposition; what does that mean? If the cat is found to be dead, did it die on opening the box, or before? All these paradoxical questions arise in the

inequalities and are in excellent agreement with the quantum predictions. Further experiments, notably by Perrie and collaborators in 1985 also show a violation of Bell's inequalities. Hence it can be concluded that no local hidden variable theory underlies quantum mechanics. As we pointed out earlier, it is possible to construct non-local hidden variable theories which lead to the same experimental consequences as quantum mechanics, but there is no compelling reason to introduce such complexities.

16.4 THE PROBLEM OF MEASUREMENT

The process of measurement differs in quantum mechanics in several respects from that in classical mechanics. For example, we have already seen that *complementary* pairs of variables exist, such as momentum and position which can only be measured simultaneously up to limits imposed by the uncertainty relations. The word measurement itself is used in two rather different senses in quantum mechanics, which it is important to distinguish. First, there are those measurements, often called *ideal measurements*, which can be repeated immediately. For example, if a beam of atoms, each of spin one-half, is passed through a Stern–Gerlach apparatus (see Chapter 1), orientated to split the beam in the z-direction, one of the two emerging beams (see Fig. 16.4) contains only atoms with $S_z = +\hbar/2$ and the other only atoms with $S_z = -\hbar/2$. This measurement, which is better called *state preparation*, can be immediately repeated, by passing, for instance, the $S_z = +\hbar/2$ beam through a second Stern–Gerlach apparatus, orientated in the same way as the first. If this is done, all the particles will emerge deflected in the direction corresponding to $S_z = \hbar/2$ and none in that corresponding to $S_z = -\hbar/2$. The passing of a beam of atoms through a Stern–Gerlach apparatus prepares an ensemble of atoms in a known state and is not really a measurement in the usual sense. To measure whether or not an atom has spin component $S_z = \hbar/2$ or $S_z = -\hbar/2$, it is not only necessary to pass the atom through such an apparatus but also to detect it. The act of detection usually involves the absorption of a photon, or the capture of



16.4 An example of state preparation. A Stern–Gerlach apparatus is orientated to split an unpolarised beam of spin one-half atoms A, in the z-direction. The upper energy beam B contains only atoms with $S_z = +\hbar/2$ and the lower beam C contains only atoms with $S_z = -\hbar/2$.

a particle, which are not repeatable processes. Thus measurements which can be immediately repeated on the same system are the exception rather than the rule. To repeat, we will distinguish between those 'ideal measurements' in which an ensemble of objects is prepared in a reproducible state, and true measurements which involve the detection of the object together with the determination of one or more of its properties, within certain limits – for example, energy, position, momentum, and so on. The former process allows one to say something about the future behaviour of the system, while the latter gives information about its immediate past.

The relationship between the wave function Ψ, or the corresponding state vector $|\Psi\rangle$, and the results of measurements have been discussed in Chapter 5. It is all-important to recognise that a knowledge of $|\Psi\rangle$ does not allow us to predict the results of a measurement on a single object of an observable A, except in the special case for which $|\Psi\rangle$ is an eigenstate of A belonging to some eigenvalue a_n, in which case the value a_n will be obtained. Apart from this case, $|\Psi\rangle$ only determines the frequency of obtaining a particular eigenvalue when the measurement is repeated a large number of times on a set of identically prepared objects or systems (called an ensemble of systems). It is therefore incorrect to associate a wave function with a single object (with the exception noted). On the contrary, $|\Psi\rangle$ provides us with the maximum possible information about an ensemble of identically prepared systems. It is important to be clear that the individual systems do not possess precise (but unknown) values of the variable A before a measurement is made. This would amount to the kind of hidden variable theory that we have seen to be untenable; rather a member of the ensemble cannot be said to possess a value for A (unless $|\Psi\rangle$ is an eigenvector of A) until a measurement has been performed. This view, although apparently correct, gives rise to some difficulties in interpretation, as Schrödinger pointed out in his famous cat paradox.

Schrödinger's cat

In this thought experiment, a cat is placed in a closed box together with a radioactive atom and a device which releases a deadly poison on the decay of the atom. If the wave function of a dead cat is ψ_d and that of a live cat is ψ_l, after a time t, the wave function of the cat will be the superposition

$$\Psi(t) = \psi_l \exp(-t/2\tau) + \psi_d(1 - \exp(-t/2\tau)) \qquad [16.21]$$

where τ is the half-life of the radioactive atom. On opening the box at a certain time, the cat is clearly found to be either alive or dead. This is an act of measurement which forces the system into one of the eigenstates of being alive or being dead. It is an example of the reduction or collapse of the wave function on measurement, referred to in Chapter 2. All sorts of curious questions now arise. Before the box is opened the cat is apparently neither alive or dead, but in a superposition; what does that mean? If the cat is found to be dead, did it die on opening the box, or before? All these paradoxical questions arise in the

most acute form if the wave function is held to determine the properties of a particular cat. Part of the difficulty is resolved by recognising that the wave function [16.21] refers to a large number of identically prepared cat experiments. After a time t, a certain fraction of the cats will have died and a certain fraction will be alive, as determined by $\Psi(t)$, but no prediction of the time of death of an individual cat can be made. The objection can now be made that this explanation supposes that each individual cat is in the definite state 'alive' or 'dead' before the measurement is made, in contradiction to the view expressed above that the assignment of values before a measurement is meaningless. A possible resolution of this difficulty must take account of the fact that a cat is a macroscopic object and by the correspondence principle we can expect the variable 'alive' or 'dead' to behave classically. Furthermore, the transition alive → dead is irreversible so that the cat itself acts as a measuring apparatus which records whether or not the radioactive atom has decayed. In the next paragraph, we will elaborate on this point and try to examine more closely the interaction between the apparatus and the system on which measurements are made.

An analysis of measurement

Let us now try to analyse the quantum mechanics of measurement a little more closely. We shall suppose we have an apparatus which is capable of measuring the value of an observable A. The apparatus will be taken to be macroscopic and capable of recording the result of the measurement on a piece of paper or on a magnetic tape or other medium. The result recorded in a single measurement must be one of the eigenvalues a_n of A, and for the purpose of discussion we will imagine that there is no experimental error. Let the eigenvector of A corresponding to the eigenvalue a_n be denoted by $|\phi_n\rangle$, and the set of eigenvectors $|\phi_n\rangle$, $n = 1, 2, \ldots$, be complete. The apparatus will be described by a very complicated wave function, but when presented with a system described by the eigenvector $|\phi_n\rangle$ it must record the value a_n for A, so that the apparatus must possess distinct eigenstates $|\Phi_n\rangle$, one for each n.

Before the measurement, the initial eigenvector describing the apparatus can be denoted by $|\Phi_0\rangle$, and if the system being measured is in the state $|\phi_n\rangle$, the initial state vector of the (system + apparatus) is

$$|\Psi(t)\rangle = |\phi_n\rangle \otimes |\Phi_0\rangle = |\phi_n; \Phi_0\rangle. \qquad [16.22]$$

This state vector will satisfy some time-dependent Schrödinger equation, and will evolve so that after the measurement it will become

$$|\Psi(t')\rangle = |\phi_n; \Phi_n\rangle. \qquad [16.23]$$

The apparatus will now be in the state with eigenvector $|\Phi_n\rangle$, and the value a_n will have been recorded, while the measured system will be described by the state vector $|\phi_n\rangle$. More generally, if the initial state of the system is described

by the state vector $|\psi(t)\rangle$ having the expansion

$$|\psi(t)\rangle = \sum_n b_n |\phi_n\rangle \qquad [16.24]$$

then the final state of the combined (system + apparatus) will be described by $|\Psi(t')\rangle$ where

$$|\Psi(t')\rangle = \sum_n b_n |\phi_n; \Phi_n\rangle. \qquad [16.25]$$

According to our interpretation, the meaning of $|\Psi(t')\rangle$ is that in a large number of measurements on the ensemble of identically prepared systems, the fraction in which the value a_n is found is given by $|b_n|^2$, all state vectors being normalised to unity. We notice again that if we attempt to associate $|\Psi(t')\rangle$ with a single system the equation [16.25] would be difficult to interpret.

The reduction of the wave function

Instead of concentrating on the combined (system + apparatus), let us now see what happens if we work in terms of the state vector of the system only. We take the special case in which the measurements are repeatable. Before the measurement the system is described by $|\psi\rangle$ with the expansion [16.24]. After making a large number of measurements on members of the system, we know that a fraction $|b_n|^2$ of these will have given the result a_n. That fraction of the system will be described after the measurements by the corresponding eigenvector $|\phi_n\rangle$. Looking at the whole ensemble, we see that after the measurements, we have a *mixed state*, a fraction $|b_1|^2$ of the system being in the pure state $|\phi_1\rangle$, $|b_2|^2$ in $|\phi_2\rangle$, and so on. This situation can be described by a density matrix ρ_{nm} which is diagonal in the basis $|\phi_n\rangle$ and is given by

$$\rho_{nm} = |b_n|^2 \delta_{nm}. \qquad [16.26]$$

The change from the pure state $|\psi\rangle$ to the mixed state described by [16.26] is an example of the reduction of the wave function on measurement. This has caused confusion because there is no way in which a linear Schrödinger equation can convert a pure state into a mixed state. However, we see that this reduction is not present in the treatment of the complete physical system (measured system + apparatus) and arises because we choose to direct our attention to the measured system only – it does not correspond to any physical process in nature.

We now ask the question as to whether the two descriptions of the state of affairs after the measurement, by the complete wave function [16.25] or by the density matrix [16.26] are mutually consistent. To do this, consider some different observable C. For consistency we must obtain the same expectation value of C from either description. Starting from the mixed state we have