1 PROBABILITY IN QUANTUM MECHANICS[1]

From about the beginning of the twentieth century experimental physics amassed an impressive array of strange phenomena which demonstrated the inadequacy of classical physics. The attempts to discover a theoretical structure for the new phenomena led at first to a confusion in which it appeared that light, and electrons, behaved sometimes like waves and sometimes like particles. This apparent inconsistency was completely resolved in 1926 and 1927 in the theory called quantum mechanics. The new theory asserts that there are experiments for which the exact outcome is fundamentally unpredictable and that in these cases one has to be satisfied with computing probabilities of various outcomes. But far more fundamental was the discovery that in nature the laws of combining probabilities were *not* those of the classical probability theory of Laplace. The quantum-mechanical laws of the physical world approach very closely the laws of Laplace as the size of the objects involved in the experiments increases. Therefore, the laws of probabilities which are conventionally applied are quite satisfactory in analyzing the behavior of the roulette wheel but not the behavior of a single electron or a photon of light.

A Conceptual Experiment. The concept of probability is not altered in quantum mechanics. When we say the probability of a certain outcome of an experiment is p, we mean the conventional thing, i.e., that if the experiment is repeated many times, one expects that the fraction of those which give the outcome in question is roughly p. We shall not be at all concerned with analyzing or defining this concept in more detail; for no departure from the concept used in classical statistics is required.

What is changed, and changed radically, is the method of calculating probabilities. The effect of this change is greatest when dealing with objects of atomic dimensions. For this reason we shall illustrate the laws of quantum mechanics by describing the results to be expected in some conceptual experiments dealing with a single electron.

Our imaginary experiment is illustrated in Fig. 1-1. At A we have a source of electrons S. The electrons at S all have the same energy but come out in all directions to impinge on a screen B. The screen

[1] Much of the material appearing in this chapter was originally presented as a lecture by R. P. Feynman and published as The Concept of Probability in Quantum Mechanics in the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Calif., 1951.

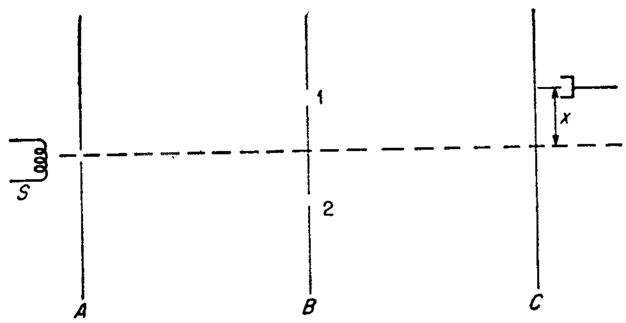

Fig. 1-1 The experimental arrangement. Electrons emitted at A make their way to the detector at screen C, but a screen B with two holes is interposed. The detector registers a count for each electron which arrives; the fraction which arrives when the detector is placed at a distance x from the center of the screen is measured and plotted against x, as in Fig. 1-2.

B has two holes, 1 and 2, through which the electrons may pass. Finally, behind the screen B at a plane C we have a detector of electrons which may be placed at various distances x from the center of the screen.

If the detector is extremely sensitive (as a Geiger counter is) it will be discovered that the current arriving at x is not continuous, but corresponds to a rain of particles. If the intensity of the source S is very low, the detector will record pulses representing the arrival of individual particles, separated by gaps in time during which nothing arrives. This is the reason we say electrons are particles. If we had detectors simultaneously all over the screen C, with a very weak source S, only one detector would respond, then after a little time, another would record the arrival of an electron, etc. There would never be a half response of the detector; either an entire electron would arrive or nothing would happen. And two detectors would never respond simultaneously (except for the coincidence that the source emitted two electrons within the resolving time of the detectors—a coincidence whose probability can be decreased by further decrease of the source intensity). In other words, the detector of Fig. 1-1 records the passage of a single corpuscular entity traveling from S through a hole in screen B to the point x.

This particular experiment has never been done in just this way. In the following description we are stating what the results would be according to the laws which fit every experiment of this type which has ever been performed. Some experiments which directly illustrate the conclusions we are reaching here have been done, but such experiments are usually more complicated. We prefer, for pedagogical reasons, to select experiments which are the simplest in principle and disregard the difficulties of actually doing them.

Incidentally, if one prefers, one could just as well use light instead of electrons in this experiment. The same points would be illustrated. The source S could be a source of monochromatic light and the sensitive detector a photoelectric cell or, better, a photomultiplier which would record pulses, each being the arrival of a single photon.

What we shall measure for various positions x of the detector is the mean number of pulses per second. In other words, we shall determine experimentally the (relative) probability P that the electron passes from S to x, as a function of x.

The graph of this probability as a function of x is the complicated curve illustrated qualitatively in Fig. 1-2a. It has several maxima and minima, and there are locations near the center of the screen at which electrons hardly ever arrive. It is the problem of physics to discover the laws governing the structure of this curve.

We might at first suppose (since the electrons behave as particles) that

I. Each electron which passes from S to x must go through either hole 1 or hole 2. As a consequence of I we expect that:

II. The chance of arrival at x should be the sum of two parts, P_1, the chance of arrival coming through hole 1, plus P_2, the chance of arrival coming through hole 2.

We may find out if this is true by direct experiment. Each of the component probabilities is easy to determine. We simply close hole

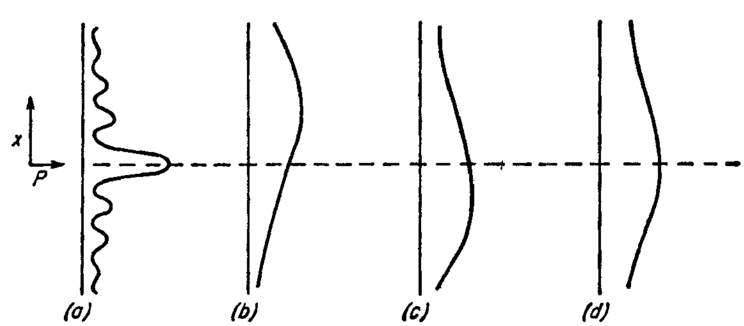

Fig. 1-2 Results of the experiment. Probability of arrival of electrons at x plotted against the position x of the detector. The result of the experiment of Fig. 1-1 is plotted here at (a). If only one hole is open, so the electrons can go through just hole 1, the result is (b). For just hole 2 open, it is (c). If we imagine each electron just goes through one hole or the other, we expect the curve (d) = (b) + (c) when both holes are open. This is considerably different from what we actually get, (a).

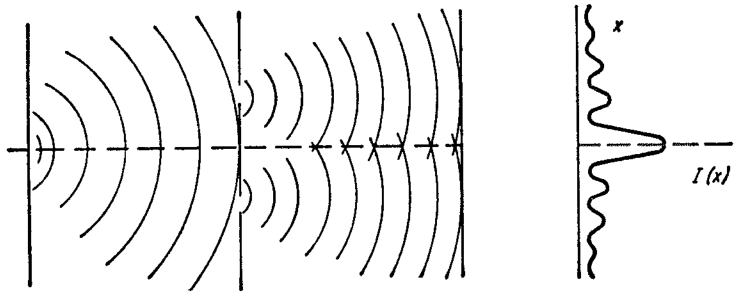

Fig. 1-3 An analogous experiment in wave interference. The complicated curve $P(x)$ in Fig. 1-2a is the same as the intensity $I(x)$ of waves which would arrive at x starting from S and coming through the holes. At some points x the wavelets from holes 1 and 2 interfere destructively (e.g., a crest from hole 1 arrives at the same time as a trough from hole 2); at others, constructively. This produces the complicated minima and maxima of the curve $I(x)$.

2 and measure the chance of arrival at x with only hole 1 open. This gives the chance P_1 of arrival at x for electrons coming through 1. The result is given in Fig. 1-2b. Similarly, by closing 1 we find the chance P_2 of arrival through hole 2 (Fig. 1-2c).

The sum of these (Fig. 1-2d) clearly does not agree with the curve (a). Hence, experiment tells us definitely that $P \neq P_1 + P_2$, or that II is false.

The Probability Amplitude. The chance of arrival at x with both holes open is *not* the sum of the chance with just hole 1 open plus that with just hole 2 open.

Actually, the complicated curve $P(x)$ is familiar, inasmuch as it is exactly the intensity of distribution in the interference pattern to be expected if waves starting from S pass through the two holes and impinge on the screen C (Fig. 1-3). The easiest way to represent wave amplitudes is by complex numbers. We can state the correct law for $P(x)$ mathematically by saying that $P(x)$ is the absolute square of a certain complex quantity (if electron spin is taken into account, it is a hypercomplex quantity) $\phi(x)$ which we call the *probability amplitude* of arrival at x. Furthermore, $\phi(x)$ is the sum of two contributions: ϕ_1, the amplitude of arrival through hole 1, plus ϕ_2, the amplitude of arrival through hole 2. In other words,

III. There are complex numbers ϕ_1 and ϕ_2 such that

$$P = |\phi|^2 \tag{1-1}$$

$$\phi = \phi_1 + \phi_2 \tag{1-2}$$

and

$$P_1 = |\phi_1|^2 \qquad P_2 = |\phi_2|^2 \tag{1-3}$$

In later chapters we shall discuss in detail the actual calculation of ϕ_1 and ϕ_2. Here we say only that ϕ_1, for example, may be calculated as a solution of a wave equation representing waves spreading from the source to 1 and from 1 to x. This reflects the wave properties of electrons (or in the case of light, photons).

To summarize: We *compute* the intensity (i.e., the absolute square of the amplitude) of waves which would arrive in the apparatus at x and then *interpret* this intensity as the probability that a particle will arrive at x.

Logical Difficulties. What is remarkable is that this dual use of wave and particle ideas does not lead to contradictions. This is so only if great care is taken as to what kind of statements one is permitted to make about the experimental situation.

To discuss this point in more detail, consider first the situation which arises from the observation that our new law III of composition of probabilities implies, in general, that it is not true that $P = P_1 + P_2$. We must conclude that when both holes are open, it is *not true* that the particle goes through one hole or the other. For if it had to go through one or the other, we could classify all the arrivals at x into two disjoint classes, namely, those arriving via hole 1 and those arriving through hole 2; and the frequency P of arrival at x would surely be the sum of the frequency P_1 of particles coming through hole 1 and the frequency P_2 of those coming through hole 2.

To extricate ourselves from the logical difficulties introduced by this startling conclusion, we might try various artifices. We might say, for example, that perhaps the electron travels in a complex trajectory going through hole 1, then back through hole 2 and finally out through 1 in some complicated manner. Or perhaps the electron spreads out somehow and passes partly through both holes so as to eventually produce the interference result III. Or perhaps the chance P_1 that the electron passes through hole 1 has not been determined correctly inasmuch as closing hole 2 might have influenced the motion near hole 1. Many such classical mechanisms have been tried to explain the result. When light photons are used (in which case the same law III applies), the two interfering paths 1 and 2 can be made to be many centimeters apart (in space), so that the two alternative trajectories must almost certainly be independent. That the actual situation is more profound than might at first be supposed is shown by the following experiment.

The Effect of Observation. We have concluded on logical grounds that since $P \neq P_1 + P_2$, it is not true that we can expect to analyze the electron's motion by the simple assumption that the electron passes through either hole 1 or hole 2. But it is easy to design an experiment to test our conclusion directly. We have merely to have a source of light behind the holes and watch to see through which hole the electron passes (see Fig. 1-4). For electrons scatter light, so that if light is scattered behind hole 1, we may conclude that an electron passed through hole 1; and if it is scattered in the neighborhood of hole 2, the electron has passed through hole 2.

The result of this experiment is to show unequivocally that the electron *does* pass through either hole 1 or hole 2! That is, for every electron which arrives at the screen C (assuming the light is strong enough that we do not miss seeing it) light is scattered either behind hole 1 or behind hole 2, and never (if the source S is very weak) at both places. (A more delicate experiment could even show that the charge passing through the holes passes through either one or the other and is in all cases the complete charge of one electron and not a fraction of it.)

It now appears that we have come to a paradox. For suppose that we combine the two experiments. We watch to see through which hole the electron passes and at the same time measure the chance that the electron arrives at x. Then for each electron which arrives at x we can say experimentally whether it came through hole 1 or hole 2. First we may verify that P_1 is given by curve (b), because if we select, of the electrons which arrive at x, only those which appear to come through hole 1 (by scattering light there), we find they are,

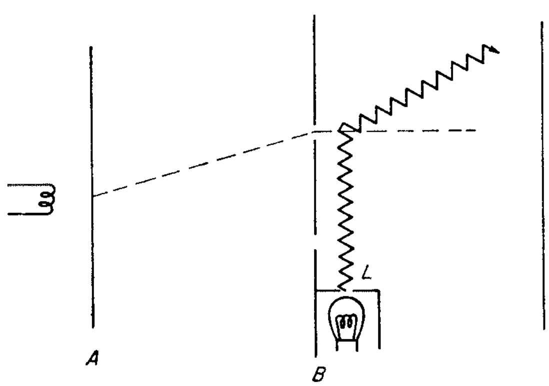

Fig. 1-4 A modification of the experiment of Fig. 1-1. Here we place a light source L behind the screen B and look for light scattered by the electrons passing through hole 1 or hole 2. With a strong light source every electron is indeed found to pass by one or the other hole. But now the probability of arrival at x is no longer given by the curve of Fig. 1-2a, but is instead given by Fig. 1-2d.

indeed, very nearly distributed as in curve (*b*). (This result is obtained whether hole 2 is open or closed, so we have verified that there is no subtle influence of closing 2 on the motion near hole 1.) If we select the electrons scattering light at 2, we get (very nearly) P_2 of Fig. 1-2*c*. But now each electron appears at either 1 or 2 and we can separate our electrons into disjoint classes. So, if we take both together, we *must* get the distribution $P = P_1 + P_2$ illustrated in Fig. 1-2*d*. And experimentally we do! Somehow now the distribution does *not* show the interference effects III of curve (*a*)!

What has been changed? When we watch the electrons to see through which hole they pass, we obtain the result $P = P_1 + P_2$. When we do not watch, we get a different result,

$$P = |\phi_1 + \phi_2|^2 \neq P_1 + P_2$$

Just by watching the electrons, we have changed the chance that they arrive at x. How is this possible? The answer is that, to watch them, we used light and the light in collision with the electron may be expected to alter its motion, or, more exactly, to alter its chance of arrival at x.

On the other hand, can we not use weaker light and thus expect a weaker effect? A negligible disturbance certainly cannot be presumed to produce the finite change in distribution from (*a*) to (*d*). But weak light does not mean a weaker disturbance. Light comes in photons of energy $h\nu$, where ν is the frequency, or of momentum h/λ, where λ is the wavelength. Weakening the light just means using fewer photons, so that we may miss seeing an electron. But when we do see one, it means a complete photon was scattered and a finite momentum of order h/λ is given to the electron.

The electrons that we miss seeing are distributed according to the interference law (*a*), while those we do see and which therefore have scattered a photon arrive at x with the probability $P = P_1 + P_2$ in (*d*). The net distribution in this case is therefore the weighted mean of (*a*) and (*d*). In strong light when nearly all electrons scatter light, it is nearly (*d*); and in very weak light, when very few scatter, it becomes more like (*a*).

It might still be suggested that since the momentum carried by the light is h/λ, weaker effects could be produced by using light of longer wavelength λ. But there is a limit to this. If light of too long a wavelength is used, we shall not be able to tell whether it was scattered from behind hole 1 or hole 2; for the source of light of wavelength λ cannot be located in space with precision greater than order λ.

We thus see that any physical agency designed to determine through which hole the electron passes must produce, lest we have a paradox, enough disturbance to alter the distribution from (*a*) to (*d*).

It was first noticed by Heisenberg, and stated in his uncertainty principle, that the consistency of the then-new mechanics required a limitation to the subtlety to which experiments could be performed. In our case the principle says that an attempt to design apparatus to determine through what hole the electron passed, and delicate enough so as not to deflect the electron sufficiently to destroy the interference pattern, must fail. It is clear that the consistency of quantum mechanics requires that it must be a general statement involving all the agencies of the physical world which might be used to determine through which hole an electron passes. The world cannot be half quantum-mechanical, half classical. No exception to the uncertainty principle has been discovered.

2 THE UNCERTAINTY PRINCIPLE

We shall state the uncertainty principle as follows: Any determination of the alternative taken by a process capable of following more than one alternative destroys the interference between alternatives. Heisenberg's original statement of the uncertainty principle was not given in the form we have used here. We shall interrupt our argument for a few paragraphs to discuss Heisenberg's original statement.

In classical physics a particle can be described as moving along a definite trajectory and having, for example, a precise position and velocity at any particular time. Such a picture would not lead to the odd results that we have seen are characteristic of quantum mechanics. Heisenberg's uncertainty principle gives the limits of accuracy of such classical ideas. For example, the idea that a particle has both a definite position and a definite momentum has its limitations. A real system (i.e., one obeying quantum mechanics) looked upon from a classical view appears to be one in which the position or momentum is not definite, but is uncertain. The uncertainty in position can be reduced by careful measurement, and other measurements may make the momentum definite. But, as Heisenberg stated in his principle, both cannot be accurately known simultaneously; the product of the uncertainties of momentum and position involved in any experiment cannot be smaller than a number with the order of magnitude of $\hbar$.†
That such a result is required by physical consistency in the situation

† $\hbar = h/2\pi = 1.054 \times 10^{-27}$ erg-sec, where h = Planck's constant.

we have been discussing can be shown by considering still another way of trying to determine through which hole the electron passes.

Example. Notice that if an electron is deflected in passing through one of the holes, its vertical component of momentum is changed. Furthermore, an electron arriving at the detector x after passing through hole 1 is deflected by a different amount, and thus suffers a different change in momentum, than an electron arriving at x via hole 2. Suppose that the screen at B is not rigidly supported, but is free to move up and down (Fig. 1-5). Any change in the vertical component of the momentum of an electron upon passing through a hole will be accompanied by an equal and opposite change in the momentum of the screen. This change in momentum can be measured by measuring the velocity of the screen before and after the passage of an electron. Call δp the difference in momentum change between electrons passing through hole 1 or hole 2. Then an unambiguous determination of the hole used by a particular electron requires a momentum determination of the screen to an accuracy of better than δp.

If the experiment is set up in such a way that the momentum of screen B can be measured to the required accuracy, then, since we can

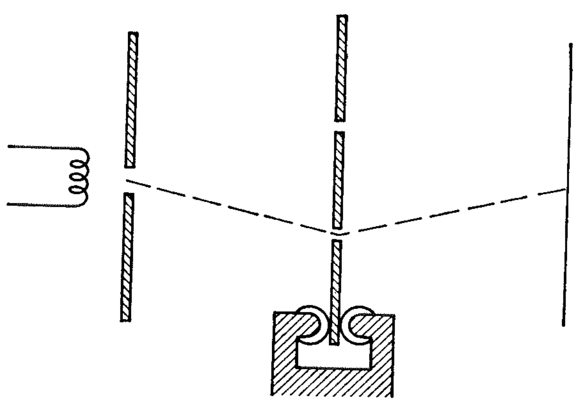

Fig. 1-5 Another modification of the experiment of Fig. 1-1. The screen B is left free to move vertically. If the electron passes hole 2 and arrives at the detector (at $x = 0$, for example), it is deflected upward and the screen B should recoil downward. The hole through which the electron passes can be determined for each passage by starting with the screen at rest and measuring whether it is recoiling up or down afterward. According to Heisenberg's uncertainty principle, however, such precise momentum measurements on screen B would be inconsistent with accurate knowledge of its vertical position, so we could not be sure that the center line of the holes is correctly set. Instead of $P(x)$ of Fig. 1-2a, we get this smeared a little in the vertical direction, so it looks like Fig. 1-2d.

determine the hole passed through, we must find that the resulting distribution of electrons is that of curve (d) of Fig. 1-2. The interference pattern of curve (a) must be lost. How can this happen? To understand, note that the construction of a distribution curve in the plane C requires an accurate knowledge of the vertical position of the two holes in screen B. Thus we must measure not only the momentum of screen B but also its position. If the interference pattern of curve (a) is to be established, the vertical position of B must be known to an accuracy of better than $d/2$, where d is the spacing between maxima of the curve (a). For suppose the vertical position of B is not known to this accuracy; then the vertical position of every point in Fig. 1-2a cannot be specified with an accuracy greater than $d/2$, since the zero point of the vertical scale must be lined up with some nominal zero point on B. Then the value of P at any particular height x must be obtained by averaging over all values within a distance $d/2$ of x. Clearly, the interference pattern will be smeared out by this averaging process. The resulting curve will look like Fig. 1-2d.

The interference pattern in the original experiment is the sign of a wave-like behavior of the electrons. The pattern is the same for any wave motion, so we may use the well-known result from the theory of light diffraction that the relation between the separation a of the holes, the distance l between screen B and the plane C, the wavelength λ of the light, and d is

$$\frac{a}{l} = \frac{\lambda}{d} \tag{1-4}$$

as shown in Fig. 1-6. In Chap. 3 we shall find that the wavelength of the electron waves is intimately connected with the momentum of the electron by the relation

$$p = \frac{h}{\lambda} \tag{1-5}$$

If p is the total momentum of an electron (and we assume all the electrons have the same total momentum), then for $l \gg a$

$$\frac{\delta p}{p} \simeq \frac{a}{l} \tag{1-6}$$

as shown in Fig. 1-7. It follows that

$$d = \frac{h}{\delta p} \tag{1-7}$$

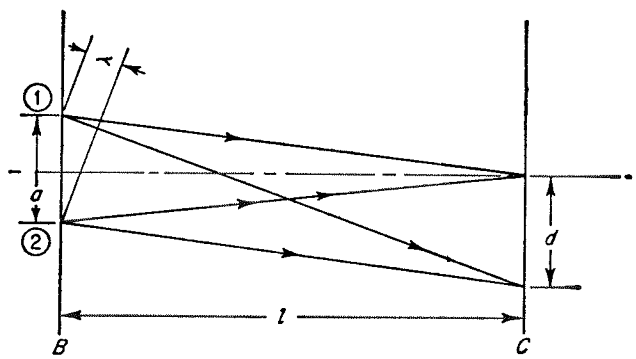

Fig. 1-6 Two beams of light, starting in phase at holes 1 and 2, will interfere constructively when they reach the screen C if they take the same time to travel from B to C. This means that a maximum in the diffraction pattern for light beams passing through two holes will occur at the center of the screen. As we move down the screen, the next maximum will occur at a distance d, which is far enough from the center that, in traveling to this point, the beam from hole 1 will have traveled exactly one wavelength λ farther than the beam from hole 2.

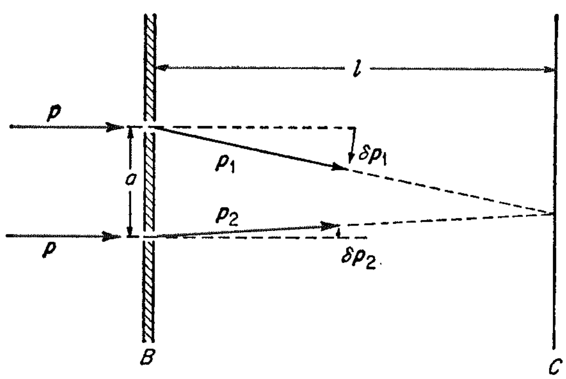

Fig. 1-7 The deflection of an electron in passing through a hole in the screen B is actually a change in momentum $\delta\mathbf{p}$. This change amounts to the addition of a small component of momentum in a direction approximately perpendicular to the original momentum vector. The change in energy is completely negligible. For small deflection angles, the total momentum vector keeps the same magnitude (approximately). Then the deflection angle is represented to a very good approximation by $|\delta\mathbf{p}|/|\mathbf{p}|$. If two electrons, one starting from hole 1 with momentum $\mathbf{p}_1$ and the other starting from hole 2 with momentum $\mathbf{p}_2$, reach the same point on the screen C, then the angles through which they were deflected must differ by approximately a/l. Since we cannot say through which hole an electron has come, the uncertainty in the vertical component of momentum which the electron receives on passing through the screen B must be equivalent to this uncertainty in deflection angle. This gives relation $|\mathbf{p}_1 - \mathbf{p}_2|/|\mathbf{p}| = |\delta\mathbf{p}|/|\mathbf{p}| = a/l$.

Since experimentally we find that the interference pattern has been lost, it must be that the uncertainty δx in the measurement of the position of B is larger than $d/2$. Thus

$$\delta p \ \delta x \geq \frac{h}{2} \tag{1-8}$$

which agrees (in order of magnitude) with the usual statement of the uncertainty principle.

A similar analysis can be applied to the previous measuring device where the scattering of light was used to determine through which hole the electron passed. Such an analysis produces the same lower limit for the uncertainties of measurement.

The uncertainty principle is not "proved" by considering a few such experiments. It is only illustrated. The evidence for it is of two kinds. First, no one has yet found any experimental way to defeat the limitations in measurements which it implies. Second, the laws of quantum mechanics seem to require it if their consistency is to be maintained, and the prediction of these laws has been confirmed again and again with great precision.

INTERFERING ALTERNATIVES

Two Kinds of Alternatives. From a physical standpoint the two routes are independent alternatives, yet the implication that the probability is the sum $P_1 + P_2$ is false. This means that either the premise or the reasoning which leads to such a conclusion must be false. Since our habits of thought are very strong, many physicists find that it is much more convenient to deny the premise than to deny the reasoning. To avoid the logical inconsistencies into which it is so easy to stumble, they take the following view: When no attempt is made to determine through which hole the electron passes, one cannot say it must pass through one hole or the other. Only in a situation where apparatus is operating to determine through which hole the electron goes is it permissible to say that it passes through one or the other. When you watch, you find that it goes through either one hole or the other hole; but if you are not looking, you cannot say that it goes either one way or the other! Nature demands that we walk a logical tightrope if we wish to describe her.

Contrary to that way of thinking, we shall in this book follow the suggestion made in the first part of this chapter and deny the reasoning; i.e., we shall not compute probabilities by adding probabilities for all alternatives. In order to make definite the new rules for combining probabilities, it will be convenient to define two meanings for the word "alternative." The first of these meanings carries with it the concept of exclusion. Thus holes 1 and 2 are *exclusive alternatives* if one of them is closed or if some apparatus that can unambiguously determine which hole is used is operating. The other meaning of the word "alternative" carries with it a concept of combination or interference. (The term *interference* has the same meaning here as it has in optics, i.e., either constructive or destructive interference.) Thus we shall say, holes 1 and 2 present *interfering alternatives* to the electron when (1) both holes are open and (2) no attempt is made to determine through which hole the electron passes. When the alternatives are of this interfering type, the laws of probability must be changed to the form given in Eqs. (1-1) and (1-2).

The concept of interfering alternatives is fundamental to all of quantum mechanics. In some situations we may have both kinds of alternatives present. Suppose we ask, in the two-hole experiment, for the probability that the electron arrives at some point, say, within 1 cm of the center of the screen. We may mean by that the probability that if there were counters arranged all over the screen (so one or another would go off when the electron arrived), the counter which went off was within 1 cm of $x = 0$. Here the various possibilities are that the electron arrives at some counter via some hole. The holes represent interfering alternatives, but the counters are exclusive alternatives. Thus we first add $\phi_1 + \phi_2$ for a fixed x, square that, and then sum these resultant probabilities over x from -1 to $+1$ cm.

It is not hard, with a little experience, to tell which kind of alternatives is involved. For example, suppose that information about the alternatives is available (or could be made available without altering the result), but this information is not used. Nevertheless, in this case a sum of probabilities (in the ordinary sense) must be carried out over *exclusive* alternatives. These exclusive alternatives are those which *could* have been separately identified by the information.

Some Illustrations. When alternatives cannot possibly be resolved by any experiment, they always interfere. A striking illustration of this is the scattering of two nuclei at 90°, say, in the center-of-gravity system, as illustrated in Fig. 1-8. Suppose A represents an α particle and B some other nucleus. Ask for the probability that A is scattered to position 1 and B to 2. The amplitude is, say, $\phi_{AB}(1,2)$. The probability of this is $p = |\phi_{AB}(1,2)|^2$. Suppose we do not distinguish what kind of nucleus arrives at 1, that is, whether it

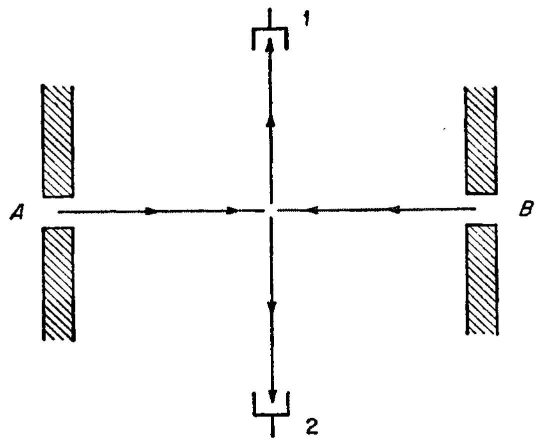

Fig. 1-8 Scattering of one nucleus by another in the center-of-gravity system. The scattering of two identical nuclei shows striking interference effects. There are two interfering alternatives here. The particle which arrives at 1, say, can have been that which started either from A or from B. If the original nuclei were not identical, tests of identity at 1 could determine which alternative had actually been taken, so they are exclusive alternatives and the special interference effects do not arise in this case.

is B or A. If it is B, the amplitude is $\phi_{AB}(2,1)$ [which equals $\phi_{AB}(1,2)$, because we have taken a 90° angle]. The chance that some nucleus arrives at 1 and the other at 2 is

$$|\phi_{AB}(1,2)|^2 + |\phi_{AB}(2,1)|^2 = 2p \tag{1-9}$$

We have added the probabilities. The cases A arrives at 1 and B arrives at 1 are exclusive alternatives because we could, if we wished, determine the character of the nucleus at 1 without disturbing the previous scattering process.

But if A is an α particle, what happens if B is also an α particle? Then no experiment can distinguish which is which, and we cannot know when one arrives at 1 whether it is A or B. We have interfering alternatives, and the probability is

$$|\phi_{AB}(1,2) + \phi_{AB}(2,1)|^2 = 4p \tag{1-10}$$

This interesting result is readily verified experimentally.

If electrons scatter electrons, the result is different in two ways. First, the electron has a quality we call *spin*, and a given electron may be in one of two states called *spin up* and *spin down*. The spin is not changed to first approximation for scattering at low energy. The spin carries a magnetic moment. At low velocities the main forces are electrical, owing to charge, and the magnetic influences make only a small correction, which we neglect. So if electron A has spin up and B has spin down, we could later tell which arrived at 1 by measuring its spin. If up, it is A; if down, it is B. The scattering probability is then

$$|\phi(1,2)|^2 + |\phi(2,1)|^2 = 2p \tag{1-11}$$

in this case.

If, however, both A and B start with spin up, we cannot later tell which is which and we would expect

$$|\phi_{AB}(1,2) + \phi_{AB}(2,1)|^2 = 4p \tag{1-12}$$

Actually this is wrong and, remarkably, electrons obey a different rule. The amplitude for an event in which the identity of a pair of electrons is reversed contributes 180° out of phase. That is, the case of both spin up gives

$$|\phi_{AB}(1,2) - \phi_{AB}(2,1)|^2 \tag{1-13}$$

In our case of 90° scattering $\phi_{AB}(1,2) = \phi_{AB}(2,1)$, so this is zero.

Fermions and Bosons. This rule of the 180° phase shift for alternatives involving exchange in identity of electrons is very odd, and its ultimate reason in nature is still only imperfectly understood. Other particles besides electrons obey it. Such particles are called fermions, and are said to obey Fermi, or antisymmetric, statistics. Electrons, protons, neutrons, neutrinos, and μ mesons are fermions. So are compounds of an odd number of these such as a nitrogen atom, which contains seven electrons, seven protons, and seven neutrons. This 180° rule was first stated by Pauli and is the full quantum-mechanical basis of his exclusion principle, which controls the character of the chemists' periodic table.

Particles for which interchange does not alter the phase are called bosons and are said to obey Bose, or symmetrical, statistics. Examples of bosons are photons, π mesons, and systems containing an even number of Fermi particles such as an α particle, which is two protons and two neutrons. All particles are either one or the other, bosons or fermions. These interference properties can have profound and mysterious effects. For example, helium liquid made of atoms of atomic mass 4 (bosons) at temperatures of one or two degrees Kelvin can flow without any resistance through small tubes, whereas the liquid made of atoms of mass 3 (fermions) does not have this property.

The concept of identity of particles is far more complete and definite in quantum mechanics than it is in classical mechanics. Classically, two particles which seem identical could be nearly identical, or identical for all practical purposes, in the sense that they may be so closely equal that present experimental techniques cannot detect any difference. However, the door is left open for some future technique to establish the difference. In quantum mechanics, however, the situation is different. We can give a direct test to determine whether or not particles are completely indistinguishable.

If the particles in the experiment diagramed in Fig. 1-8, starting at A and B, were only approximately identical, then improvements in experimental techniques would enable us to determine by close scrutiny of the particle arriving at 1, for example, whether it came from A or B. In this situation the alternatives of the two initial positions must be exclusive, and there must be no interference between the amplitudes describing these alternatives. Now the important point is that this act of scrutiny would take place *after* the scattering had taken place. This means that the observation could not possibly affect the scattering process, and this in turn implies that we would expect no interference between the amplitudes describing the alternatives (that it is either the particle from A or the particle from B which arrives at 1). In this case we must conclude from the uncertainty principle that there is no way, even in principle, to ever distinguish between these possibilities. That is, when a particle arrives at 1, it is completely impossible by any test whatsoever, now or in the future, to determine whether the particle started from A or B. In this more rigorous sense of identity, all electrons are identical, as are all protons, etc.

As a second example we consider the scattering of neutrons from a crystal. When neutrons of wavelength somewhat shorter than the atomic spacing are scattered from the atoms in a crystal, we get very strong interference effects. The neutrons emerge only in certain discrete directions determined by the Bragg law of reflection, just as for X rays. The interfering alternatives which enter this example are the alternative possibilities that it is this, or that, atom which does the scattering of a particular neutron. (The amplitude to scatter neutrons from any atom is so small that we need not consider alternatives in which a neutron is scattered by more than one atom.) The waves of amplitude describing the motion of a neutron which start from these atoms interfere constructively only in certain definite directions.

Now there is an interesting complication which enters this apparently simple picture. Neutrons, like electrons, carry a spin, which can be analyzed in two states, spin up and spin down. Suppose the scattering material is composed of an atomic species which has a similar spin property, such as carbon-13. In this case an experiment will reveal two apparently different types of scattering. It is found that besides the scattering in discrete directions, as described in the preceding paragraph, there is a diffused scattering in all directions. Why should this be?

A clue to the source of these two types of scattering is provided by the following observation. Suppose all the neutrons which enter the experiment are prepared in such a manner that their spin direction is up. If the spin direction of the emerging neutrons is analyzed, it will be found that there are some up and some down; those which still have spin up are scattered only at the discrete Bragg angles, while those whose spin has been changed to down come out scattered diffusely in all directions!

Now in order that a neutron flip its spin from up to down, the law of conservation of angular momentum requires that the spin of the scattering nucleus be changed from down to up. Therefore, in principle, the particular nucleus which was responsible for scattering that particular neutron could be determined. We could, in principle, note down before the experiment the spin state of all the scattering nuclei in the crystal. Then, after the neutron is scattered, we could reinvestigate the crystal and see which nucleus had changed its spin from down to up. If no crystal nucleus underwent such a change in spin, then neither did the neutron, and we cannot tell from which nucleus the neutron was actually scattered. In this case the alternatives interfere and the Bragg law of scattering results.

If, on the other hand, one crystal nucleus is found to have changed spin, then we know that this nucleus did the scattering. There are no interfering alternatives. The spherical waves of amplitude which emerge from this particular nucleus describe the motion of the scattered neutron, and only the waves emerging from this nucleus enter into that description. In this case there is equal likelihood to find the scattered neutron coming out in any direction.

The concept of searching through all the nuclei in a crystal to find which one has changed its spin state is surely a needle-in-the-haystack type of activity, but nature is not concerned with the practical difficulties of experimentation. The important fact is that in principle it is possible without producing any disturbance of the scattered neutron to determine (in this latter case where the spin states change) which crystal nucleus actually did the scattering. The existence of this possibility means that even if we do not actually carry out this determination, we are nevertheless dealing with exclusive (and thus noninterfering) alternatives.

On the other hand, the fact that we get interference between alternatives in the situation where the spin states of the neutrons were not changed means that it is impossible, even in principle, to ever discover which particular crystal nucleus did the scattering—impossible, at least, without disturbing the situation during or before the scattering.

4 SUMMARY OF PROBABILITY CONCEPTS

Alternatives and the Uncertainty Principle. The purpose of this introductory chapter has been to explain the meaning of a probability amplitude and its importance in quantum mechanics and to discuss the rules for manipulation of these amplitudes. Thus we have stated that there is a quantity called a *probability amplitude* associated with every method whereby an event in nature can take place. For example, an electron going from a source at S (Fig. 1-1) to a detector at x has one amplitude for completing this course while passing through hole 1 of the screen at B and another amplitude in passing through hole 2. Further, we can associate an amplitude with the overall event by adding together the amplitudes of each alternative method. Thus, for example, the overall amplitude for arrival at x is given in Eq. (1-2) as

$$\phi = \phi_1 + \phi_2 \tag{1-14}$$

Next, we interpret the absolute square of the overall amplitude as the probability that the event will happen. For example, the probability that an electron reaches the detector is

$$P = |\phi_1 + \phi_2|^2 \tag{1-15}$$

If we interrupt the course of the event before its conclusion with an observation on the state of the particles involved in the event, we disturb the construction of the overall amplitude. Thus if we observe the system of particles to be in one particular state, we exclude the possibility that it can be in any other state, and the amplitudes associated with the excluded states can no longer be added in as alternatives in computing the overall amplitude. For example, if we determine with the help of some sort of measuring equipment that the electron passes through hole 1, the amplitude for arrival at the detector is just ϕ_1. Further, it does not matter if we actually observe and record the outcome of the measurement or not, so long as the measuring equipment is working. Obviously, we could observe the outcome at any time we wished. The operation of the measuring equipment is sufficient to disturb the system and its probability amplitude.

This latter fact is the basis of the Heisenberg uncertainty principle, which states that there is a natural limit to the subtlety of any experiment or the refinement of any measurement.

The Structure of the Amplitude. The amplitude for an event is the sum of the amplitudes for the various alternative ways that the event can occur. This permits the amplitude to be analyzed in many different ways depending on the different classes into which the alternatives can be divided. The most detailed analysis results from considering that a particle going from A to B, for example, in a given time interval, can be considered to have done this by going in a certain motion (position vs. time) or path in space and time. We shall therefore associate an amplitude with each possible motion. The total amplitude will be the sum of a contribution from each of the paths.

This idea can be made more clear by a further consideration of our experiment with the two holes. Suppose we put a couple of extra screens between the source and the hole. Call these screens D and E. In each of them we drill a few holes which we number D_1, D_2, . . . and E_1, E_2, . . . (Fig. 1-9). For simplicity, we shall assume the electrons are constrained to move in the xy plane. Then there are several alternative paths which an electron may take in going from the source to the hole in screen B. It could go from source to D_2, and then E_3, and then the hole 1; or it could go from the source to D_3, then E_1, and finally to the hole 1, etc. Each of these paths has its own amplitude. The complete amplitude is the sum of all of them.

Next, suppose we continue to drill holes in the screens D and E until there is nothing left of the screens. The path of an electron must now be specified by the height x_D at which the electron passes

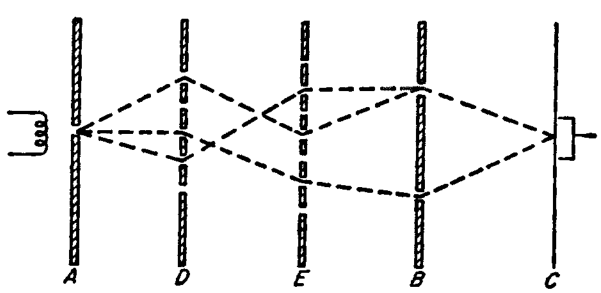

Fig. 1-9 When several holes are drilled in the screens D and E placed between the source in screen A and the final position in screen C, several alternative routes are available for each electron. For each of these routes there is an amplitude. The result of any experiment in which all of the holes are open requires the addition of all of these amplitudes, one for each possible path.

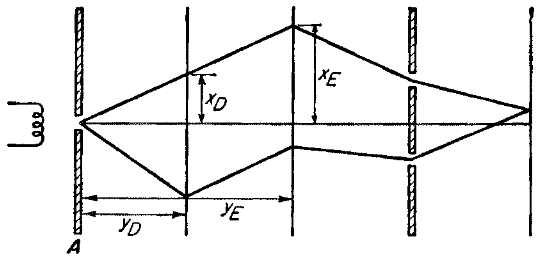

Fig. 1-10 More and more holes are cut in the screens at y_D and y_E. Eventually, the screens are completely riddled with holes, and the electron has a continuous range of positions, up and down along each screen, at which it can pass through the position of the screen. In this case the sum of alternatives becomes a double integral over the continuous parameters x_D and x_E describing the alternative heights at which the electron passes the position of the screens at y_D and y_E.

the position y_D at the nonexistent screen D, together with the height x_E, at the position y_E, as in Fig. 1-10. To each pair of heights there corresponds an amplitude. The principle of superposition still applies, and we must take the sum (or by now, the integral) of these amplitudes over all possible values of x_D and x_E.

Clearly, the next thing to do is to place more and more screens between the source and the hole 1 and in each screen drill so many holes that there is nothing left. Throughout this process we continue to refine the definition of the path of the electron, until finally we arrive at the sensible idea that a path is merely height as a particular function of distance, or $x(y)$. We also continue to apply the principle of superposition, until we arrive at the integral over all paths of the amplitude for each path.

Now we can make a still finer specification of the motion. Not only can we think of the particular path $x(y)$ in space, but we can specify the *time* at which it passes each point in space. That is, a path will (in our two-dimensional case) be given if the two functions $x(t)$, $y(t)$ are given. Thus we have the idea of an amplitude to take a certain path $x(t)$, $y(t)$. The total amplitude to arrive is the sum or integral of this amplitude over all possible paths. The problem of defining this concept of a sum or integral over all paths in a mathematically more precise way will be taken up in Chap. 2.

Chapter 2 also contains the formula for the amplitude for any given path. Once this is given, the laws of nonrelativistic quantum mechanics are completely stated, and all that remains is a demonstration of the application of these laws in a number of interesting special cases.

1-5 SOME REMAINING THOUGHTS

We shall find that in quantum mechanics, the amplitudes ϕ are solutions of a completely deterministic equation (the Schrödinger equation). Knowledge of ϕ at $t = 0$ implies its knowledge at all subsequent times. The interpretation of $|\phi|^2$ as the probability of an event is an indeterministic interpretation. It implies that the result of an experiment is not exactly predictable. It is very remarkable that this interpretation does not lead to any inconsistencies. That it is true has been amply demonstrated by analyses of many particular situations by Heisenberg, Bohr, Born, von Neumann, and many other physicists. In spite of all these analyses the fact that no inconsistency can arise is not thoroughly obvious. For this reason quantum mechanics appears as a difficult and somewhat mysterious subject to a beginner. The mystery gradually decreases as more examples are tried out, but one never quite loses the feeling that there is something peculiar about the subject.

There are a few interpretational problems on which work may still be done. They are very difficult to state until they are completely worked out. One is to show that the probability interpretation of ϕ is the *only* consistent interpretation of this quantity. We and our measuring instruments are part of nature and so are, in principle, described by an amplitude function satisfying a deterministic equation. Why can we only predict the probability that a given experiment will lead to a definite result? From what does the uncertainty arise? Almost without doubt it arises from the need to amplify the effects of single atomic events to such a level that they may be readily observed by large systems. The details of this have been analyzed only on the assumption that $|\phi|^2$ is a probability, and the consistency of this assumption has been shown. It would be an interesting problem to show that *no other* consistent interpretation can be made.

Other problems which may be further analyzed are those dealing with the theory of knowledge. For example, there seems to be a lack of symmetry in time in our knowledge. Our knowledge of the past is qualitatively different from that of the future. In what way is only the probability of a future event accessible to us, whereas the certainty of a past event can often apparently be asserted? These matters again have been analyzed to a great extent. Possibly a little more can be said to clarify the situation, however. Obviously, we are again involved in the consequences of the large size of ourselves and of our measuring equipment. The usual separation of observer and observed which is now needed in analyzing measurements in quantum mechanics should not really be necessary, or at least should be even more thoroughly analyzed. What seems to be needed is the statistical mechanics of amplifying apparatus.

The analyses of such problems are, of course, in the nature of philosophical questions. They are not necessary for the further development of physics. We know we have a consistent interpretation of ϕ and, almost without doubt, the only consistent one. The problem of today seems to be the discovery of the laws governing the behavior of ϕ for phenomena involving nuclei and mesons. The interpretation of ϕ is interesting. But the much more intriguing question is: What new modifications of our thinking will be required to permit us to analyze phenomena occurring within nuclear dimensions?

1-6 THE PURPOSE OF THIS BOOK

So far, we have given the form the quantum-mechanical laws must take, i.e., that a probability amplitude exists, and we have outlined one possible scheme for calculating this amplitude. There are other ways to formulate this. In a more usual approach to quantum mechanics the amplitude is calculated by solving a kind of wave equation. For particles of low velocity, it is called the Schrödinger equation. A more accurate equation valid for electrons of velocity arbitrarily close to the velocity of light is the Dirac equation. In this case the probability amplitude is a kind of hypercomplex number. We shall not discuss the Dirac equation in this book, nor shall we investigate the effects of spin. Instead, we limit our attention to low-velocity electrons, extending our horizon somewhat in the direction of quantum electrodynamics by investigating photons, particles whose behavior is determined by Maxwell's equation.

In this book we shall give the laws to compute the probability amplitude for nonrelativistic problems in a manner which is somewhat unconventional. In some ways, particularly in developing a conceptual understanding of quantum mechanics, it may be preferred, but in others, e.g., in making computations for the simpler problems and for understanding the literature, it is disadvantageous.

The more conventional view, via the Schrödinger equation, is already presented in many books, but the views to be presented here have appeared only in abbreviated form in papers in the journals.[1]

[1] R. P. Feynman, The Space-Time Approach to Non-relativistic Quantum Mechanics, *Rev. Mod. Phys.*, vol. 20, p. 367, 1948.

A central aim of this book is to collect this work into one volume where it may be expanded with sufficient clarity and detail to be of use to the interested student.

In order to keep the subject within bounds, we shall not make a complete development of quantum mechanics. Instead, whenever a topic has reached such a point that further elucidation would best be made by conventional arguments appearing in other books, we refer to those books. Because of this incompleteness, this book cannot serve as a complete textbook of quantum mechanics. It can serve as an introduction to the ideas of the subject if used in conjunction with another book that deals with the Schrödinger equation, matrix mechanics, and applications of quantum mechanics.

On the other hand, we shall use the space saved (by our not developing all of quantum mechanics in detail) to consider the application of the mathematical methods used in the formulation of quantum mechanics to other branches of physics.

It is a problem of the future to discover the exact manner of computing amplitudes for processes involving the apparently more complicated particles, namely, neutrons, protons, and mesons. Of course, one can doubt that, when the unknown laws are discovered, we shall find ourselves computing amplitudes at all. However, the situation today does not seem analogous to that preceding the discovery of quantum mechanics.

In the 1920's there were many indications that the fundamental theorems and concepts of classical mechanics were wrong, i.e., there were many paradoxes. General laws could be proved independently of the detailed forces involved. Some of these laws did not hold. For example, each spectral line showed a degree of freedom for an atom, and at temperature T each degree of freedom should have an energy kT, contributing R to the specific heat. Yet this very high specific heat expected from the enormous number of spectral lines did not appear.

Today, any general law that we have been able to deduce from the principle of superposition of amplitudes, such as the characteristics of angular momentum, seems to work. But the detailed interactions still elude us. This suggests that amplitudes will exist in a future theory, but their method of calculation may be strange to us.