

Chapter 2. The Probabilistic Interpretation of Measurements in Quantum Mechanics

§1 Quantum mechanics is a statistical theory

§2 Probability is intrinsic to quantum mechanics

§3 The equation for probability

§4 Irrelevant phase factors

§5 Pure states are orthonormal

§6 Extension to the continuous spectrum

§7 The connection between Dirac notation and Schrödinger notation

§8 The orthonormality of the continuous spectrum: the δ -function

§9 Coherent states and “quantum probability”

§10 Summary

©2014, Horia Metiu

§ 1 *Quantum mechanics is a statistical theory.* Let us analyze an experiment in which a stream of atoms moves through vacuum and crosses a laser beam. The experimentalist makes sure that before reaching the light beam, each atom in the stream has the lowest internal energy possible (denoted by E_0). The laser frequency Ω is set to satisfy

$$\hbar\Omega = E_1 - E_0$$

where E_1 is the first excited state of the atom. Each atom passing through the laser beam emerges from the illuminated region either in the state $|E_1\rangle$ (if it absorbs a photon) or in the state E_0 (if it does not absorb a photon).

In principle we can make the density of the stream of atoms so low that only one atom at a time passes through the laser beam. We can measure whether a given atom has absorbed a photon by monitoring whether it emits one.

We monitor every one of the N atoms that has passed through the laser beam and find that N_a have absorbed a photon and N_0 have not ($N_a + N_0 = N$). You are probably so used to hearing this statement that you no longer notice that it violates classical physics and common sense. Every atom in the stream has *exactly* the same history as every other atom. They all have the *same* energy when they approach the laser beam and they are all exposed to the *same* amount of radiation, for the *same* amount of time. Classical physics says that particles that have the same history have the same fate. They all take from the beam of light *the same amount of energy* and emerge from contact with the beam in the same state. The experiments say that they don't.

The behavior of the atoms would be even more puzzling if we had two students performing *the same* experiment in two different laboratories. Each student sends individual atoms, one by one, through the laser beam. They make sure that they have identical conditions in both laboratories. Student A performs the experiment with 10 atoms and obtains the following results for the states of the atom after it crosses the beam:

$$\{ |E_0\rangle, |E_0\rangle, |E_1\rangle, |E_0\rangle, |E_1\rangle, |E_0\rangle, |E_0\rangle, |E_0\rangle, |E_0\rangle, |E_1\rangle \}$$

In the first two experiments, the atom did not absorb light, in the third it did, etc.

The results obtained by Student B performing *the same experiment* are

$$\{ |E_0\rangle, |E_1\rangle, |E_0\rangle, |E_1\rangle, |E_0\rangle, |E_1\rangle, |E_0\rangle, |E_1\rangle, |E_0\rangle, |E_0\rangle \}$$

The experiments are identical but the results are different! The number of atoms observed by B is the same as the number observed by A. However, in A's experiments, three of the atoms were excited while B managed to excite four.

A fundamental assumption of classical science is that identical experiments must give identical results. The experiment described above forces us to abandon this assumption when we deal with quantum phenomena.

How are we going to do science if the experiments are not reproducible? In a subtle way they are. Let us ask students A and B to perform their experiments independently and identically, but send through the beam *a large number* N of atoms, one at a time. They count how many atoms absorb light. A finds that $N_a(A)$ do, and B finds that $N_a(B)$ do.

Here is what they will observe: the ratios

$$\nu_a(A, N) = \frac{N_a(A)}{N}$$

and

$$\nu_a(B, N) = \frac{N_a(B)}{N}$$

are close to one another. Moreover, they get closer and closer as the students increase N .

The limit

$$P_a = \lim_{N \rightarrow \infty} \frac{N_a(A)}{N}$$

is called the *probability* that an atom absorbs a photon. P_a does not depend on N as long as N is large enough. We do not need to label P_a with the name of the student since both students will measure the *same value* for P_a .

The ratio $N_a/N \equiv \nu_a(N)$ is called the frequency of photon absorption (there is no connection with the frequency of light; this ‘frequency’ tells us how frequently something takes place). If N is small, $\nu_a(N)$ depends on N , and the value of $\nu_a(N)$ obtained by A differs from that obtained by B. If N is sufficiently large, the two students obtain the same value for $\nu_a(N)$. There is determinism after all in the quantum world: if we perform two identical experiments, *the probability that a specific result is obtained is the same* in both experiments.

Since we are going to deal with probability throughout our study of quantum mechanics, here are a few practical observations about probability. When you flip a coin the probability that you get heads is $1/2$. This does not mean that if you flip twice you get heads once and tails once. It does not mean that if you flip ten times you get five heads and five tails. It does not mean that if you flip 100,000 times you get *exactly* 50,000 heads and 50,000 tails. The probability only tells us that the number of times you get heads divided by the total number of flips gets closer and closer to $1/2$, as the number of flips increases. The same interpretation holds for the probability that something happens in a quantum mechanical experiment. While the probabilities that something happens in the experiment are determined by the experimental conditions, we don’t know precisely the results of one experiment, on one atom, any more than we know the result of a coin flip in any individual flip.

Quantum mechanics postulates that the situation found in the experiment performed by A and B is general. If we measure the value of an observable (the energy of the atom is one example) whose spectrum is a_1, a_2, a_3, \dots and if the conditions of the experiment are such that the measured value could be a_1 , or a_2, \dots or a_n , then we can only know the probabilities p_1, p_2, \dots, p_n that the measurement yields the values a_1, a_2, \dots, a_n , respectively. This means that we cannot know for certain the result of *one experiment*, performed on *one system*. The only exceptions to this rule are experiments that place the system in a pure state $|a_n\rangle$ of an observable A before we measure A. In such a case the result of the measurement is guaranteed to be a_n .

Scientists experimented with light absorption before quantum mechanics was discovered. Why did they believe that such experiments were reproducible? Why did they not observe the probabilistic nature of the atomic phenomena? In all their experiments they exposed a *very large number of atoms* (in a gas or a liquid) to a *very large number of photons*. They measured the intensity of light I when it exited from the gas and the intensity I_0 when it entered it. They found that the ratio I/I_0 was *the same* whenever they repeated the experiment. However, this is not precisely true. Let us denote by N_0 the number of photons entering the vessel per unit time and by N the number of photons exiting it per unit time. Let us imagine that we can measure N with *absolute precision*, and we perform the experiment twice with the same conditions and the same value of N_0 . Denote by N_1 the value of N obtained in the first experiment and by N_2 the number obtained in the second (identical) experiment. Measurements having absolute precision will find that $N_1 \neq N_2$. However, in the laboratory we cannot make experiments

with absolute precision and, if N_0 is large, the difference $N_1 - N_2$ is too small for us to detect. This fooled us into believing that the experiment is reproducible and the world is deterministic.

§ 2 *Probability is intrinsic to quantum mechanics.* You have encountered probabilities in situations where classical mechanics was valid (for example, in classical statistical mechanics). What is the difference between such classical situations and the probability forced on us by quantum mechanics?

Let us look at the example of Brownian motion. Mr. Brown observed in a microscope a liquid droplet containing a pollen particle and saw that the particle had a chaotic motion. When he observed the particle in several experiments, *performed under identical conditions*, he found that the distance traveled by the particle, in a fixed amount of time, differed from one experiment to another. The best he could do was to determine the probability that the particle will be at point \vec{r} at time t if it was at \vec{r}_0 at time t_0 . Sounds like quantum mechanics, doesn't it? Definitely not. In this experiment the molecules in the liquid and the pollen particle are sufficiently large that their motion is described by classical mechanics. Why then are we forced to use probability to describe it? I will argue in what follows that we must use probability to describe our results, for this classical system, because the experiments *were not performed under identical conditions*.

Imagine that we perform a computer experiment in which we follow the motion of 100,000 molecules and of a sphere of molecular dimensions placed among them. Since classical mechanics is valid, there is no law of nature (no Heisenberg uncertainty) to prevent us from knowing the position and the momentum of the sphere and of every molecule in the system, at any given time.

There is no law of physics to prevent us from knowing the forces with which the molecules act on each other and on the sphere. Given this knowledge of the initial conditions and of the forces, we can solve Newton's equation to calculate *precisely* the position and the momentum of each molecule and of the sphere, at any later time. No probability is involved in this calculation. If we repeat the calculations *with the same initial conditions*, we get the same result.

It appears that there is a conflict between this calculation and Mr. Brown's experiment, but there is none. If we plot the motion of the sphere and do not show the molecules forming the liquid (to simulate the observations made with a microscope), we would see the same disorganized motion that Mr. Brown saw in his microscope. This happens because the molecules in the liquid collide very frequently with the sphere. When they hit the sphere and bounce off it, they give the sphere a push. The net force acting on the sphere keeps changing as different molecules, having different velocities, collide with it at different times. When the sphere is small and only a small number of molecules collide with it in any small time interval, the magnitude and the direction of the net force acting on the sphere changes irregularly from one time interval to another. The sphere moves like a drunken sailor and this is what Mr. Brown saw.

Why does a person following the sphere in the microscope think that he must use probabilities to describe the motion and the one using a computer simulation claims that she can predict the position and the velocity of the sphere accurately? Who is wrong? Neither. The essential difference is that the person doing the computer simulation *knows the initial position and mo-*

mentum of each particle. She can control these quantities in her computer program. The experimentalist can only control the temperature and the pressure. If you give a hundred “simulators” the positions and the velocities at one time they will all find identical motion for their spheres. However, if you give them only the temperature and the pressure, many different positions and velocities are compatible with this information. They will pick different values for the initial velocities and positions (all of them compatible with the given temperature and pressure) and will find that the motion of the sphere differs from simulation to simulation. To compare their results, they will have to use probabilities.

In the case of Brownian motion, the experimentalist is forced to use probability out of ignorance: he does not control the initial conditions in his experiment. He can fix temperature and pressure but not the initial positions and momenta of all particles. The initial positions and momenta in one experiment differ from those in another experiment. This is why two experiments are not identical even though the pressure and the temperature are. This is not a limitation imposed by the laws of nature, but by the fact that molecules are small and we and our instruments are large. The simulator circumvents this limitation.

Quantum mechanics is different. It claims that, in general, it is impossible to control the initial conditions so that the outcome of an experiment on a single system is predetermined. The only exception occurs when an experimentalist manages to prepare a pure state of an observable A , before measuring A . In quantum mechanics probabilities are needed not because we do not completely control the conditions of the experiment, but because that

is the way the world of the very small particle behaves. There is however a form of determinism: identical experiments lead to identical probabilities for the results of the measurements.

§ 3 *The equation for probability.* Let us assume that we expose a system to some external agent (a pulse of light, an electron beam, etc.) and this interaction forces the system to have the state $|\psi\rangle$. After that, we perform a measurement that determines the value of an observable A, whose spectrum is $\{a_1, a_2, a_3, \dots\}$. What is the probability that the result of a measurement of A is a_n ?

If you want to use quantum mechanics to calculate the probability, you will have to solve the time-dependent Schrödinger equation that describes the interaction of the system with the external agent. This will tell you what the state $|\psi\rangle$ is at the time when A is measured. The equation is such that if $|\psi\rangle$ is a solution then $\alpha|\psi\rangle$ is also a solution, where α is any complex number. The only physically meaningful solution $|\psi\rangle$ is the one that satisfies the equation

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

A ket that satisfies this condition is said to be *normalized*.

We want to calculate the probability $P_\psi(a_n)$ that a measurement of A when the system is in state $|\psi\rangle$, will yield the value a_n . To perform this calculation we need to know the ket $|a_n\rangle$. We can calculate it by solving a differential equation (you'll see later what the equation is and how it is solved). This equation has an infinite number of solutions because if $|a_n\rangle$ is a solution, then $\alpha|a_n\rangle$, where α is a complex number, is also a solution. The

only physically meaningful solution is the one satisfying

$$\langle a_n | a_n \rangle = 1 \quad (2)$$

We can now formulate one of the fundamental assumptions of quantum mechanics: if we prepare a system in a state described by the *normalized* state $|\psi\rangle$, the probability that when we measure the magnitude of an observable A , we obtain the value a_n is

$$P_\psi(a_n) = |\langle a_n | \psi \rangle|^2 \quad (3)$$

where $|a_n\rangle$ is the normalized pure state in which A has the value a_n . This equation is correct only if both $|\psi\rangle$ and $|a_n\rangle$ are *normalized*. The vertical bars in the equation mean that we calculate the scalar product $\langle a_n | \psi \rangle$ and if this is a complex number we obtain the probability by calculating the square of its absolute value.

This prescription was proposed by Max Born in 1927, after the work of Heisenberg, Schrödinger, and Dirac. People solved the Schrödinger equation but did not know how to interpret the wave function. It was Born who proposed that the wave function is not a real wave (as Schrödinger believed) and that its only usefulness is to calculate the probability of the result of a measurement (a collision in the case of Born). In the paper Born mistakenly proposed that the wave function itself gives a probability and the paper has a footnote saying that while correcting the proofs of his paper Professor Born realized that the wave function can be a complex number and therefore it cannot be a probability. In the footnote he proposes that the probability is given by the absolute value squared of the wave function. This footnote is the main reason for awarding him the Nobel prize (so the citation says), but his

other contributions to quantum theory were very fundamental and deserved a Nobel prize (in my opinion). The definition given here is a generalization of Born's and I don't know who extended it (it is likely that it was Dirac).

Exercise 1 Use the properties of the scalar product to show that

$$|\langle \psi | a_n \rangle|^2 = |\langle a_n | \psi \rangle|^2$$

Does this mean that $\langle \psi | a_n \rangle = \langle a_n | \psi \rangle$?

Exercise 2 Show that if $|a\rangle$ is an arbitrary ket then the ket $|b\rangle$ defined by

$$|b\rangle = \frac{|a\rangle}{\sqrt{\langle a | a \rangle}}$$

is normalized.

Exercise 3 Show that

$$P_\psi(a_n) = \frac{|\langle a_n | \psi \rangle|^2}{\langle a_n | a_n \rangle \langle \psi | \psi \rangle}$$

whether or not $|\psi\rangle$ and $|a_n\rangle$ are normalized.

§ 4 *Irrelevant phase factors.* Here is how we measure the probability in Eq. 3. I consider first the case when the experimentalist is skillful enough to perform the experiment with a single molecule. He uses his apparatus to

force the molecule to have the state $|\psi\rangle$. Then, he measures A and the result is one of the values a_n from the spectrum of A. This is one experiment that must be repeated N times, under identical conditions for the parameters that he can control. He records the number of times $N(a_n)$ that the measurement of A gave the result a_n . The ratio

$$\frac{N(a_n)}{N} \quad (4)$$

will equal

$$|\langle a_n | \psi \rangle|^2 \quad (5)$$

if N is large enough. How does he know that N is large enough? He keeps increasing N until the ratio $N(a_n)/N$ no longer changes.

Does this determine the value of $\langle a_n | \psi \rangle$? Almost, but not quite. To see why I say this, I need to remind you of a few results of algebra. In general, the scalar product $\langle a_n | \psi \rangle$ in Eq. 3 is a complex number. Therefore we can write it in polar form

$$\langle a_n | \psi \rangle = \rho \exp(-i\phi) \quad (6)$$

Here ρ is a real number called the absolute value of $\langle a_n | \psi \rangle$, and ϕ is a real number called the *phase* of $\langle a_n | \psi \rangle$. In quantum mechanics the term $\exp(-i\phi)$, where ϕ is a real number, multiplying a ket or a scalar product, is called a ‘phase factor’.

Let us use this expression to calculate the probability (which we can measure)

$$|\langle a_n | \psi \rangle|^2 = \langle a_n | \psi \rangle^* \langle a_n | \psi \rangle = \rho \exp(i\phi) \rho \exp(-i\phi) = \rho^2 \quad (7)$$

This tells me that if I measure $P_\psi(a_n) \equiv |\langle a_n | \psi \rangle|^2 = \rho^2$, I cannot determine the phase ϕ of $\langle a_n | \psi \rangle$. This has two consequences. (1) If I measure the

probability $P_\psi(a_n)$, I can determine the absolute value of the scalar product $\langle a_n | \psi \rangle$ but not its phase. (2) If I calculate the scalar product (by using quantum mechanics) then $\langle a_n | \psi \rangle$ and $\exp(-i\lambda)\langle a_n | \psi \rangle$, where λ is a real number, are equivalent, as far as physics is concerned, since they give the same probability $P_\psi(a_n)$.

We can push this result a bit further: states represented by the kets $|\psi\rangle$ and $\exp(-i\phi)|\psi\rangle$ are physically equivalent because they give the same result when used in Eq. 3 to calculate the probability that a measurement of A gives the result a_n .

Exercise 4 Prove the statement made above.

It is important to note that the quantity $\exp(i\phi)(|\psi\rangle + |\eta\rangle)$ is physically equivalent to $|\psi\rangle + |\eta\rangle$ but $|\psi\rangle + \exp(i\phi)|\eta\rangle$ is *not* equivalent to $|\psi\rangle + |\eta\rangle$. If you change the phase of one ket in a sum of kets, you change the physical properties of the system. But you can multiply the sum by any phase factor and you cause no change in the behavior of the system.

Exercise 5 An experimentalist prepares a system in the state $|\psi\rangle + |\eta\rangle$ and another one prepares the system in the state $|\psi\rangle + \exp(-i\phi)|\eta\rangle$, where ϕ is a real number. Calculate the probability that when they measure A they obtain the result a_n . Will this probability be the same in both cases? How about the states $|\psi\rangle + |\eta\rangle$ and $\exp(-i\phi)(|\psi\rangle + |\eta\rangle)$, or $|\eta\rangle + |\psi\rangle$ and $|\eta\rangle - |\psi\rangle$?

Exercise 6 An experimentalist prepares a system in a state $|\psi\rangle$ which is described as an element in the space ℓ^2 given by:

$$|\psi\rangle = \{\psi_1, \psi_2, \dots\} \quad (8)$$

where ψ_j , $j = 1, 2, \dots$ are complex numbers. He measures the value of an observable A whose spectrum is $\{a_1, a_2, \dots\}$. The pure state $|a_n\rangle$ is given in ℓ^2 by

$$|a_n\rangle = \{\alpha_1, \alpha_2, \dots\} \quad (9)$$

where α_j , $j = 1, 2, \dots$ are complex numbers. Calculate the probability $P_\psi(a_n)$ that the measurement yields a_n .

Exercise 7 An experimentalist prepares a system in a state $|\psi\rangle$ which is described as an element $\psi(x)$ in the space L^2 . He measures the value of an observable A whose spectrum is $\{a_1, a_2, \dots\}$. The pure state $|a_n\rangle$ is given in L^2 by $\alpha(x)$. Calculate the probability $P_\psi(a_n)$ that the measurement yields a_n .

Exercise 8 One experimentalist prepares a system in the state $|\psi\rangle$, another one in $-|\psi\rangle$, and a third one in $0.2|\psi\rangle$. They all measure the value of a quantity A. Will the probabilities that A takes the value a_n be the same in these experiments? (Hint: before using a ket in the formula for probability you must normalize it.)

§ 5 *Pure states are orthonormal.* Let us examine what happens if an experimentalist measures the magnitude of an observable A , having the spectrum a_0, a_1, a_2, \dots , in a system that *has been prepared to be in a pure state* $|\psi\rangle \equiv |a_n\rangle$. The probability that the result of such a measurement is a_m is (use Eq. 3)

$$P_\psi(a_m) = |\langle a_m | \psi \rangle|^2 = |\langle a_m | a_n \rangle|^2 = |\langle a_n | a_m \rangle|^2 \quad (10)$$

By the definition of a pure state we are *certain* that if we measure A , when the system is in the state $|\psi\rangle = |a_n\rangle$, the result is a_n . This means that $P_\psi(a_m) = 1$ if $m = n$ and $P_\psi(a_m) = 0$ if $m \neq n$. These statements are expressed concisely by writing

$$|\langle a_n | a_m \rangle|^2 = \delta_{n,m} \quad (11)$$

where

$$\delta_{nm} = \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m \end{cases} \quad (12)$$

The symbol $\delta_{n,m}$ is called a Kronecker delta. From Eq. 11 we derive

$$\langle a_n | a_m \rangle = \delta_{n,m} \quad (13)$$

In deriving Eq. 13 from Eq. 11 we set a physically irrelevant phase factor to be equal to 1.

As I told you in Chapter 1, two kets $|a\rangle$ and $|b\rangle$ that satisfy

$$\langle a | b \rangle = 0 \quad (14)$$

are said to be *orthogonal*, and a ket $|a\rangle$ that satisfies

$$\langle a | a \rangle = 1 \quad (15)$$

is said to be *normalized*. Furthermore, a set of kets $\{|a_1\rangle, |a_2\rangle, |a_3\rangle, \dots\}$ is said to be *orthonormal* if they satisfy

$$\langle a_n | a_m \rangle = \delta_{n,m} \quad (16)$$

for all values of n and m .

We can now state our main conclusion: the definition (Eq. 10) of the probability in quantum mechanics implies that if $\{a_1, a_2, a_3, \dots\}$ is the *discrete spectrum* of an observable, then any pair of kets from the set of pure states $\{|a_1\rangle, |a_2\rangle, |a_3\rangle, \dots\}$ must be orthonormal. This must be true for the pure states of any observable as long as they correspond to the discrete spectrum.

§ 6 *Extension to the continuous spectrum.* We know that the coordinate x and the momentum p have continuous spectra. The energy spectrum of a hydrogen atom has both discrete values (when the electron and the proton are bound) and continuous values (when the electron and the proton have so much energy that are no longer bound to each other). How do we calculate the probability that a measurement gives a certain value in the continuous spectrum?

Consider a system in which a particle is in a state $|\psi\rangle$. Since the coordinate X of the particle is an observable, we can introduce the pure state $|x\rangle$ in which we know for certain that the coordinate has the value x . A simple generalization of Eq. 10 seems to imply that the probability that a measurement of position gives the value x , when the system is in a state $|\psi\rangle$, is

$$P_\psi(x) \equiv |\langle x | \psi \rangle|^2 \quad (17)$$

This definition is impossible to use in practice. What do I mean when

I say that the position of particle is $x = 3$? I cannot measure the position with infinite precision. When I measure X and find it to be equal to 2.9999999999999999, is this equal to 3? How about 3.0000000000000001? The statement that $x = 3$ is operationally meaningless when we talk about physical measurements of a *continuous* variable.

A better way is to define the probability $P_\psi(x) dx$ that the coordinate X takes values between x and $x + dx$, by

$$P_\psi(x) dx \equiv |\langle x | \psi \rangle|^2 dx \quad (18)$$

Here dx is an infinitesimal spatial interval. I can calculate the probability that X takes values in a finite interval $[x_0 - \Delta/2, x_0 + \Delta/2]$ from

$$\int_{x_0 - \Delta/2}^{x_0 + \Delta/2} |\langle x | \psi \rangle|^2 dx \quad (19)$$

It is quite possible to measure this probability.

There is of course nothing special about the coordinate and we can apply this equation to any measurement of a quantity having a continuous spectrum. If we are interested in momentum, we can define the pure state $|p\rangle$, in which the particle has the momentum p and the probability that a system in state $|\psi\rangle$ has a momentum between p and $p + dp$ is

$$P_\psi(p) dp = |\langle p | \psi \rangle|^2 dp \quad (20)$$

§ 7 *The connection between Dirac notation and Schrödinger notation.* In the introductory quantum mechanics you used what I will call the Schrödinger notation. In this notation the state of the system is described by a *wave function* $\psi(x)$ and the probability $P_\psi(x) dx$ that the particle position

X is between x and $x + dx$ is

$$P_\psi(x) dx = |\psi(x)|^2 dx = \psi(x)^* \psi(x) dx \quad (21)$$

If Dirac notation describes the same physical reality as the Schrödinger notation this expression must be identical to $|\langle x | \psi \rangle|^2 dx$ (see Eq. 19). This identification gives us a chance to find out how $\psi(x)$ is represented in Dirac's notation.

We can write Eq. 19 as

$$P_\psi(x) = |\langle x | \psi \rangle|^2 = \langle x | \psi \rangle^* \langle x | \psi \rangle = \langle \psi | x \rangle \langle x | \psi \rangle \quad (22)$$

I used here the property of the scalar product: $\langle x | \psi \rangle^* = \langle \psi | x \rangle$. Since $P_\psi(x)$ is given by Eq. 22 and by Eq. 21, the right-hand sides of these equations must be equal:

$$\psi(x)^* \psi(x) = \langle \psi | x \rangle \langle x | \psi \rangle \quad (23)$$

Eq. 23 is satisfied if we take

$$\psi(x) = e^{i\alpha} \langle x | \psi \rangle \quad (24)$$

where α is some real number, and so

$$\psi(x)^* = e^{-i\alpha} \langle \psi | x \rangle \quad (25)$$

There is no physical condition that we can use to pin down the value of α . This is not a problem, since $\psi(x)$ is not a measurable quantity. The probability distribution $\psi(x)^* \psi(x)$, which is measurable, is independent of α ; we can take any value for α without affecting the distribution, and we choose $\alpha = 0$. This makes $e^{i\alpha} = 1$.

We have found that in Dirac's notation, the wave function $\psi(x)$ used by Schrödinger is given by

$$\psi(x) = \langle x | \psi \rangle \quad (26)$$

Dirac's notation emphasizes that the wave function, which he writes as $\langle x | \psi \rangle$, is a composite of two objects: $|\psi\rangle$ which tells us the state of the system and $|x\rangle$ which appears because we want to know the position of the particle.

§ 8 *The orthonormality of the continuous spectrum: the δ -function.* We showed in §5 that the definition of probability implies that the pure states of an observable with a discrete spectrum must form an orthonormal set. How do we generalize this for the pure states of an observable with a continuous spectrum?

We start from the following mathematical identity: for the discrete pure states $|a_n\rangle$ of an observable we have

$$\sum_{n=1}^{\infty} f(a_n) \langle a_n | a_m \rangle = \sum_{n=1}^{\infty} f(a_n) \delta_{nm} = f(a_m) \quad (27)$$

This is true for an arbitrary function $f(a)$ and follows from the definition of δ_{nm} (Eq. 12).

To generalize this to states $|x\rangle$ and $|y\rangle$ describing observables having a *continuous spectrum*, we replace the sum with an integral and require that

$$\int_{-\infty}^{+\infty} f(x) \langle x | y \rangle dx = f(y) \quad (28)$$

which is the analog of Eq. 27. This imposes a stringent condition on the symbol $\langle x | y \rangle$.

Dirac introduced the notation

$$\langle x | y \rangle = \delta(x - y) \quad (29)$$

and defined its meaning through the equations

$$\int_{-\infty}^{+\infty} f(x)\delta(x - y) dx \equiv f(y) \quad (30)$$

and

$$\delta(x - y) \equiv 0 \text{ if } x \neq y \quad (31)$$

The symbol $\delta(z)$ defined by Eq. 30 is called a δ -function or *Dirac's delta function*; it is a generalization of the Kronecker delta.

This notation baffled and irritated mathematicians. von Neumann wrote a whole chapter in his famous book on the mathematics of quantum mechanics to prove that no function having the properties required by Dirac existed. A few decades later, mathematicians managed to develop a rigorous theory of the δ -function and of various generalizations of it. They called it distribution theory to emphasize that the objects they dealt with *were not functions*.

For our purposes, you can think of a δ -function in two ways.

(1) $\delta(x)$ is the limit of a sequence of functions $\phi(x; \varepsilon)$ having the property that

$$\lim_{\varepsilon \rightarrow 0} \int_{-\infty}^{+\infty} \phi(x - y; \varepsilon) f(x) dx = f(y) \quad (32)$$

There are many functions that satisfy this definition; here are a few examples:

$$\delta(x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\sqrt{\pi\varepsilon}} \exp\left[-\frac{x^2}{\varepsilon}\right] \quad (33)$$

or

$$\delta(x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\pi} \frac{\varepsilon}{x^2 + \varepsilon^2} \quad (34)$$

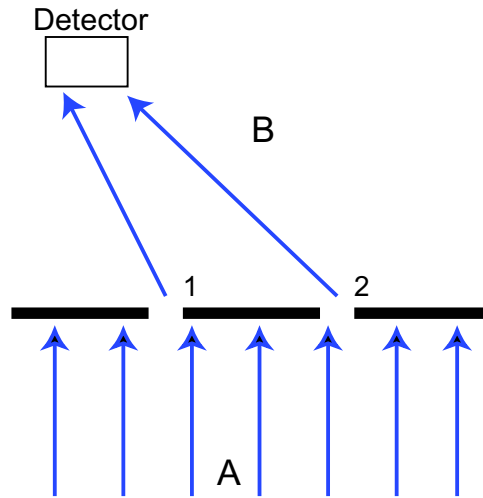


Figure 1: A two-slit experiment

or

$$\delta(x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\pi} \frac{\sin(x/\varepsilon)}{x} \quad (35)$$

One can show that any one of the functions defined by Eqs. 33–35 satisfies Eq. 32.

(2) Or you could think of $\delta(x - y)$ as a symbol that satisfies Eqs. 30 and 31 and *is meaningful only when it appears under an integral*.

The δ -function has many interesting and surprising properties and must be handled with care. You will learn more about it when needed.

§ 9 Coherent states and “quantum probability”. The probability defined by Eq. 3 (for the discrete values in the spectrum) or by Eq. 22 (for the values in the continuous spectrum) has peculiar properties that are not encountered in the ordinary probability theory. I will explain what this means by using an example.

We send a particle (an electron, a He atom, a neutron, or any particle with small mass) towards an impenetrable screen that has two slits (Fig. 1). It is essential for what follows that the state of this particle, before reaching the slits (Region A in Fig. 1), is such that we are *uncertain* of its position: as far as we can tell the particle could go through either slit. I denote by $|\psi_\alpha\rangle$ the state that the particle has if it passes through slit α ($\alpha = 1$ or 2). We don't know through which slit the particle passes, and the state in Region B is (according to the superposition principle, Eq. 1.9)

$$|\psi\rangle = |\psi_1\rangle + |\psi_2\rangle \quad (36)$$

I have invoked the principle of linear superposition but this is not necessary. If we solve the Schrödinger equation to calculate the state of the particle behind the screen, the result has the form given by Eq. 36.

The probability that the particle reaches the detector located at x is

$$P_\psi(x) = |\langle x | \psi \rangle|^2 \quad (37)$$

To be more realistic, I should integrate over the area of the detector, but the point I want to make here does not depend on this detail. Using Eq. 36 in Eq. 37, together with the properties of the scalar product, leads to

$$\begin{aligned} P_\psi(x) &= |\langle x | \psi \rangle|^2 = \langle x | \psi \rangle^* \langle x | \psi \rangle \\ &= (\langle x | \psi_1 \rangle + \langle x | \psi_2 \rangle)^* (\langle x | \psi_1 \rangle + \langle x | \psi_2 \rangle) \\ &= \langle x | \psi_1 \rangle^* \langle x | \psi_1 \rangle + \langle x | \psi_2 \rangle^* \langle x | \psi_2 \rangle + \langle x | \psi_1 \rangle^* \langle x | \psi_2 \rangle + \langle x | \psi_2 \rangle^* \langle x | \psi_1 \rangle \\ &= P_{\psi_1}(x) + P_{\psi_2}(x) + \langle x | \psi_1 \rangle^* \langle x | \psi_2 \rangle + (\langle x | \psi_1 \rangle^* \langle x | \psi_2 \rangle)^* \\ &= P_{\psi_1}(x) + P_{\psi_2}(x) + 2\Re(\langle x | \psi_1 \rangle^* \langle x | \psi_2 \rangle) \end{aligned} \quad (38)$$

I used the following properties of complex numbers: $(z^*)^* = z$ and $z^* + z = 2\Re(z)$ where $\Re(x + iy) = x$ when x and y are real numbers; $\Re(z)$ is the real part of the complex number z . I also used the fact that $\langle x | \psi_1 \rangle^* \langle x | \psi_1 \rangle = P_{\psi_1}(x)$, which is the probability that the electron reached the detector after passing through slit 1. Similarly, $\langle x | \psi_2 \rangle^* \langle x | \psi_2 \rangle = P_{\psi_2}(x)$.

The ordinary theory of probability, based on our classical experience, postulates that the probability that the particle reaches the detector is the probability that it passed through slit 1 and made it to the detector, *plus* the probability that it passed through slit 2 and made it to the detector. This would give $P_\psi(x) = P_{\psi_1}(x) + P_{\psi_2}(x)$ and *not* Eq. 38. We might think of reconciling the two results by setting $\Re(\langle x | \psi_1 \rangle^* \langle x | \psi_2 \rangle)$ to zero. But this is possible only if either $\langle x | \psi_1 \rangle$ or $\langle x | \psi_2 \rangle$ is zero, which means that either $P_{\psi_1}(x)$ or $P_{\psi_2}(x)$ is zero. If $P_{\psi_1}(x) = 0$ then we know for sure that the particle did not pass through slit 1. But this violates our premise that the uncertainty in the position of the particle approaching the slits is such that we cannot know that the particle goes only through one of the slits.

We must live with the term $\Re(\langle x | \psi_1 \rangle^* \langle x | \psi_2 \rangle)$ in Eq. 38 and accept that the properties of the probability in quantum mechanics are different from those in classical probability. This additional term $\Re(\langle x | \psi_1 \rangle^* \langle x | \psi_2 \rangle)$ is called an *interference term* and such terms are frequently present in many quantum processes. Whether they appear or not depends on how a state is prepared and what quantity is measured.

In developing this argument, I glossed over some technical details regarding the normalization of the ket describing the electron in the space behind the slits. A more precise analysis leads to the same conclusion as the one we

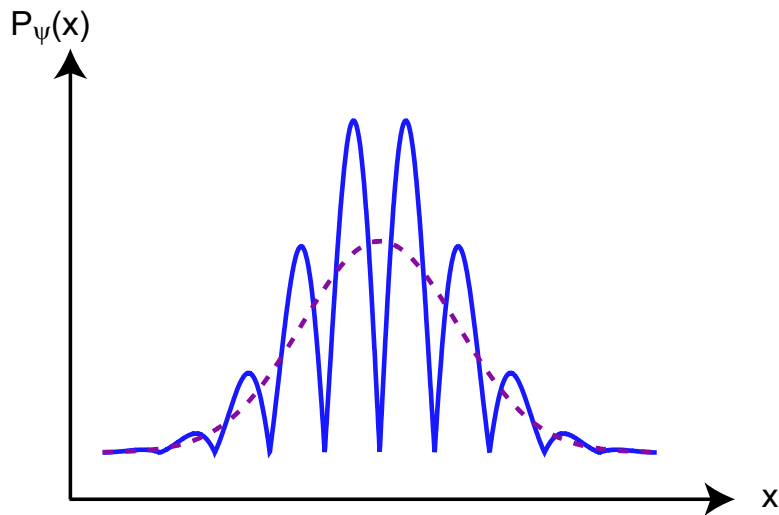


Figure 2: Probability with (solid blue line) and without (dashed purple line) interference

have reached.

The presence of the interference term has interesting consequences. The probability that the particle is located at x (the position of the detector) looks similar to the solid line in Fig. 2. If I remove the interference term $2\Re(\langle c | \psi_1 \rangle^* \langle x | \psi_2 \rangle)$ from Eq. 38 and plot the result, I obtain the dotted line. The wiggles in the solid line come from the interference term. They are very sensitive to the distance between the slits and the energy of the particle going through the slits. X-ray and neutron scattering experiments, used to determine the structure of crystals, are based on a similar interference.

You have probably learned that a similar experiment can be performed with acoustic waves passing through two slits. Does it mean that the electrons or neutrons going through two slits are waves? Not quite. If we experiment

with an acoustic wave, the whole interference pattern is built in one experiment. This is not true in our quantum two-slit experiment. Imagine that we fill the space behind the screen with tiny detectors. When we send an acoustic wave through the slits, all detectors are triggered. We get a large signal from detectors located where the wave is intense and a weak signal from those located where the intensity of the wave is low. If we send an electron, *only one detector is triggered*. When we send a second electron, having the same properties as the first, another detector is triggered. A single electron does not behave like a wave. It behaves like a particle. However, if I send N electrons independent of each other, but prepared identically, and if I denote by $\frac{N(x)}{N}\Delta x$ the number of electrons arriving at the detector located in the range $[x, x + \Delta x]$, then this quantity converges (as N increases) toward the value of $P_\psi(x)$ given by Eq. 38. *The electron is a particle but the probability that it reaches a certain position (in this particular experiment) behaves as if a wave had passed through the slits.*

We can generalize this observation as follows. If we act on a particle and create a coherent state

$$|\psi\rangle = \sum_{\alpha=1}^N |\psi_\alpha\rangle \quad (39)$$

then the probability that a measurement of an observable A yields the result a_n is

$$P_\psi(a_n) = |\langle a_n | \psi \rangle|^2 = \sum_{\alpha=1}^N P_{\psi_\alpha}(a_n) + 2\Re \left(\sum_{\alpha=1}^N \sum_{\beta=1}^N \langle a_n | \psi_\alpha \rangle^* \langle a_n | \psi_\beta \rangle \right) \quad (40)$$

The double sum is the interference term. Note the similarity with the two-slit experiment, in which we measured x instead of A and had two alternatives instead of N .

The additional term is called an interference term because of its similarity to the results of measurements on waves. If we sent a sound wave through two slits the compression ρ caused by an acoustic wave passing through the slits and measured behind the screen is

$$\rho = \sum_{\alpha=1}^2 \rho_{\alpha} \quad (41)$$

where ρ_{α} is the compression caused by the wave coming from slit α . Sound detectors work by measuring the *energy* of the wave, which is proportional to ρ^2 . Therefore the signal is proportional to

$$\rho^2 = \left(\sum_{\alpha=1}^2 \rho_{\alpha} \right)^2 = \sum_{\alpha=1}^2 \rho_{\alpha}^2 + 2\rho_1\rho_2 \quad (42)$$

ρ_{α}^2 is the signal obtained if only source α emitted. The term $\rho_1\rho_2$ is an interference term between the waves that passed through different slits.

The similarity between Eqs. 40 and 42 was used to point out that a particle has wave-like behavior. This, among other observations, led to the name ‘wave mechanics’ in the early literature on quantum phenomena. Formally the similarity appears because when we calculate the probability or the strength of the signal, we *add some quantities and then we square the sum*. When the square is performed we get “cross terms”, which are the cause of interference. In classical probability theory, we add probabilities and there are no cross terms.

§ 10 Summary. Numerous experiments force us to accept that, in general, measurements performed on a single quantum system are not reproducible. As a result we need to describe the results of a measurement by probabilities. Reproducible results are obtained only if the system is in a pure state of an

observable A and we perform a measurement of A . If we know the state of the system and the pure states of the observable, then quantum mechanics provides a recipe for calculating the probability that the measurement will yield a given value. A consequence of this equation is that the pure states of any observable must be orthonormal. If the state of the system is a coherent superposition of states the probability defined in quantum mechanics differs from that defined in classical probability theory: it has an additional interference term.

At this point the theory is rather abstract. The probability is defined in terms of the scalar products $\langle a_n | \psi \rangle$ and we have not learned how to evaluate them. We will have to learn how to write and solve the equation that gives us the pure states $|a_n\rangle$ of an observable A . We also need to learn how to provide a theoretical description of the experimental process that brings the system into the state $|\psi\rangle$ on which we perform the measurement of A . Then we must use this description to calculate $|\psi\rangle$. Only then can we evaluate $\langle a_n | \psi \rangle$ and the probability $P_\psi(a_n)$.