

2

Superposition

Paul Dirac began his classic *Principles of Quantum Mechanics* [1930] with a principle of superposition of states. The principle is mathematical but he introduced it by first describing some physical phenomena, and I'll follow his example. Some more or less familiar observations can be used to motivate a simple mathematical model that makes sense of them. These are all observations of interference of beams of matter. What does this mean?

2.1 Beams

It is easy to see how the Old English word ‘beam’ for a living tree came to refer to a timber post. But it is not so clear why we use the same word when advising drivers to turn off their high beams when they see an oncoming vehicle. The term ‘sunbeam’ may have entered the English language as a translation of the seventh-century Latin ‘*columna lucis*’ Bede used to refer to a column of fire like that which (according to Ecclesiastes) guided the Israelites by night on their flight from Egypt. After the scientific revolution people didn’t think of a beam of light as a rigid rod but as something emitted by a source that seemed to instantly illuminate distant objects. Whatever it was, this “something” counted as material since it pushed distant objects around and heated them up, however slightly. Newton thought light consisted of particles because it traveled in straight lines without bending around corners. Despite the subsequent successes of the rival wave theory, Feynman [1985] could still maintain that Newton was right (though for the wrong reasons).

Atoms and molecules certainly count as matter, and we can produce a beam of them by heating a substance in an oven and letting some of it escape through a hole and pass through another hole. If atoms are material then presumably so also are their constituent parts, including electrons and neutrons. An old-fashioned television tube produced a picture through the impact of a guided beam of electrons on a phosphorescent screen: a nuclear reactor acts as an “oven” from which a beam of neutrons may be extracted.

When a beam of matter encounters an object, the beam as well as the object may be affected. Newton studied many ways of modifying a sunbeam, including shining it through a prism to produce a spectrum of light and passing it through a glass lens resting on a flat piece of glass. In each case this produces colors—not of an object but

apparently of the light itself. Newton concluded that sunlight is a mixture of different types of “pure” light, where each type has some physical property other than color that makes it bend more or less when passing through a prism: we see light as colored only because when each different mixture of types of light enters our eye and affects our brain it produces a distinctive color sensation. But Newton had trouble accounting for details of how light is modified when passed through a glass lens resting on a glass plate. These are most clearly revealed when the light is of just one type, forming what we persist in calling a monochromatic beam.

Sunlight and most forms of artificial light are not monochromatic. I live in Tucson, a center for astronomy because of its clear night skies and nearby mountain-top sites for telescopes. Many street lamps here do produce very nearly monochromatic light by passing electricity through sodium vapor at low pressure. The astronomers like this because it makes it easy to remove stray street light from what they collect from distant stars. The local police are not so pleased because it makes it hard to see the natural colors of clothing, cars, hair, and skin in the streets at night, since everything appears some shade of the same yellow!

A concentric series of colored rings appears when a beam of light is shone from above onto a convex lens resting on a flat glass surface. These are called Newton’s rings after their famous but puzzled discoverer. They exhibit in a more controlled environment the same phenomenon that may be observed in the natural iridescence of some insect wings, and the colors that can be seen in soap bubbles, thin films of oil on water, and the peacock’s tail. All these phenomena are examples of interference of light.

Interference occurs when different beams or parts of the same beam take different paths before uniting in a single beam. When a beam of light encounters a thin layer of a different transparent medium (a soap bubble, or the air gap between a glass lens and the glass plate on which it rests) part is reflected and part is transmitted at its first surface. Some of the part that was transmitted is then reflected by the second surface, returns through the medium, and is then (mostly) emitted parallel to the other part. These two parts of the beam then merge to form a single beam which has been modified in the process. Importantly, how it has been modified depends on the difference between the lengths of the paths taken by the two parts. Newton’s rings appear because the difference in path lengths varies with the distance from the center of the lens, since this governs the size of the air gap between the lens surface and the glass plate.

The dependence on path difference is clearest when the original beam is monochromatic. The rings do not appear multicolored in this case, but they exhibit a notable variation in intensity: bright rings regularly alternate with dark rings where hardly any light can be seen. This is a good example of interference: one part of the beam interferes with another part. Where a bright ring appears the interference is constructive: together the partial beams form an emitted beam whose intensity is greater than the sum of their individual intensities. Where a dark ring appears the

interference is destructive: the intensity of the total beam emitted here is less than that of either individual intensity—the partial beams have canceled each other out, at least to some extent.

Changing the color of the incident beam changes the diameters of the rings, showing that they arise from some kind of interaction between the type of light and the difference in path lengths of the two parts of the beam. The multicolored rings seen when the beam is a mixture of types are a natural consequence of overlapping differently colored bright rings of different diameters.

Newton had trouble accounting for interference of light because he thought of a beam as composed of particles. Young later developed a wave model that could easily explain the interference of light by analogy to interference of waves in water—in one place a peak of a wave from some source may cancel a trough of a wave from a different source, while in another place their peaks and troughs coincide, as in this image of water waves made by the synchronized flapping of a floating bee's wings.¹ The wave theory of light became a cornerstone of classical physics after its mathematical development by Fresnel and Maxwell's theoretical identification of light as an electromagnetic field radiating from its source at an enormous speed of almost 300 million meters per second. But as we shall see, light behaves in ways that can't be encompassed by a classical physics of waves or of particles. So does atomic matter. But fortunately light and atomic matter behave in very similar ways!

If you stand in a garden in bright sunlight and view the window of a house from an angle, you will see not only inside the house but also a reflected image of flowers and trees. This is because besides letting light through from inside, the window also acts as a mirror by reflecting some light from outside. It is not difficult to construct a half-silvered mirror that reflects about as much of a beam of light as it transmits. Since the 1970s it has become possible to make something like a mirror for a beam of neutrons emerging from a nuclear reactor.

Such "mirrors" are the basic components of a neutron interferometer, depicted in Figure 2.1.

Cut from a single crystal of silicon, this looks a bit like an electric toaster for very thick slices of bread. I'll have more to say about the uses of this device later, but for now just consider what happens when the neutrons reach the first slab of silicon. A beam of neutrons all with the same speed is analogous to a monochromatic beam of light. If this beam hits the slab at just the right angle, part of it goes straight through while another part also goes through but is bent (as shown in the figure). By splitting the beam, the slab has essentially the same effect on the neutron beam that a half-silvered mirror has on a beam of light.

For beams of many other kinds of matter (atoms, molecules, electrons, etc.) analogs of mirrors have also been constructed. For each kind of matter there are detectors capable of measuring the intensity of the beam. Since passage from a source through

¹ See the image at http://upload.wikimedia.org/wikipedia/commons/c/c5/Ripples_waves_bee.jpg.

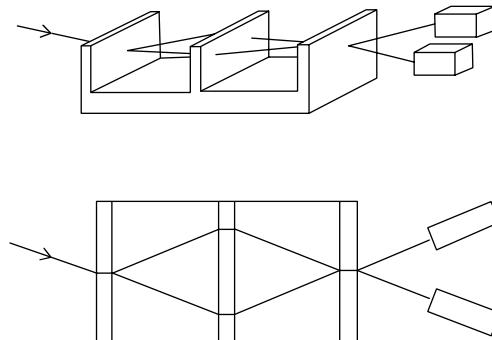


Figure 2.1 Neutron interferometer

a system of mirrors affects the beam intensity in basically the same way in each case, we can describe this without giving details of the kind of matter, its source, the mirrors, or the detectors. A generic analog of a half-silvered mirror is called a *beam splitter*.

We have now assembled all the necessary tools to exhibit the kind of peculiar observations that prompt the formulation of a principle of superposition of states. In sections 2.2 and 2.4 I'll deploy these tools by describing what is observed to happen in two series of schematic experiments, in each case building up to some clear but puzzling examples of interference phenomena.

2.2 Some Interference Phenomena

Suppose a source produces a beam of matter of intensity I , as measured by a suitable detector. For a beam of neutrons or other particles, I may indicate how many neutrons are detected every second: for a beam of light, I may indicate the power of the beam in watts (or fractions of a watt). If a beam of measured intensity I is directed at a 50–50 beam splitter that absorbs none of it, detectors placed to collect the reflected and transmitted beams as in Figure 2.2 will each measure intensity $1/2 I$. On average, half of the particles of a particle beam are detected at D_1 , the other half at D_2 .

Now suppose that these detectors are removed and two more 50–50 beam splitters are placed, one to intercept the transmitted beam, the other to intercept the reflected beam (see Figure 2.3). What would you expect to be the intensity measured by detectors placed to intercept beams transmitted and reflected by these additional beam splitters?

If a beam splitter acts as some kind of filter, transmitting one sort of particle or light but reflecting another sort, then one might expect detectors marked TT and RR each to register intensity $1/2 I$ while those marked TR and RT register nothing. (A prism does act much like a filter for light, as Newton showed. It affects a sunbeam passing through it—a continuous spectrum of different monochromatic beams emerge, each

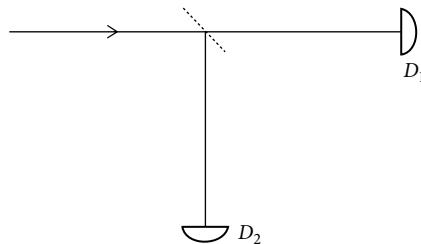


Figure 2.2 Beam splitter

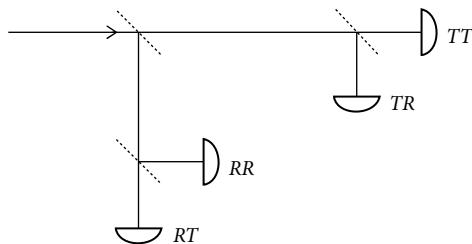


Figure 2.3 Three beam splitters

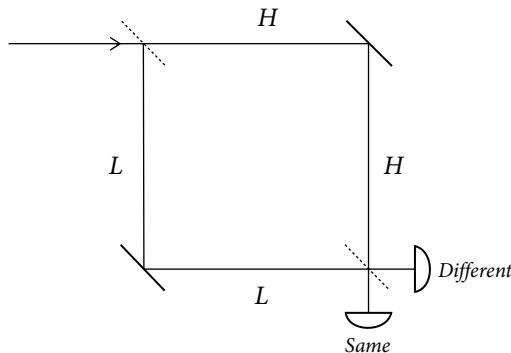


Figure 2.4 Balanced Mach-Zehnder interferometer

at a slightly different angle: these angles don't change when each monochromatic beam is bent by a second prism.) But if a beam splitter just divides up a beam without filtering it, then one would expect each detector (TT , RR , TR , RT) to register the same intensity, $1/4 I$: and indeed that is what happens.

Now consider the arrangement in Figure 2.4, where a dotted line represents a 50–50 beam splitter and a bold line represents a fully reflecting mirror for this type of beam. When paths H, L have the same length between the two beam splitters this is said to constitute a *balanced* interferometer.

I've labeled one detector *Same* to remind you that it would be expected to detect that part of the beam either transmitted at both beam splitters (*TT*) or reflected at both beam splitters (*RR*): the detector called *Different* would be expected to detect that part of the beam transmitted at one beam splitter but reflected at the other. What happened in the arrangement of Figure 2.3 seemed to show that a beam splitter does not act as a filter. So one would expect that each partial beam following path *H,L* would be split equally by the second beam splitter, and that each detector would therefore measure intensity $1/2 I$. But that is *not* what happens. Instead, detector *Different* measures intensity I while detector *Same* measures intensity 0. Partial beams following paths *H,L* appear to have interfered with each other so their intensities completely cancel each other at *Same* while wholly reinforcing each other at *Different*.

This is particularly difficult to understand if the source emits atoms or other particles. Consider a beam of neutrons emerging from a nuclear reactor. You can control their speed and intensity so they are typically separated from each other by 300 meters in the beam. The neutron interferometer pictured in Figure 2.1 effectively realizes the scheme of Figure 2.4. It is much less than a meter long. So it is very unlikely that more than one neutron is present in the interferometer at once. It is natural to assume that each neutron detected travels through the interferometer either along path *H* or along path *L*. If you have any doubts, you can check by blocking either path: this cuts the total detected intensity by half, while blocking both paths reduces the total detected intensity to 0. Moreover, it seems that no neutron splits in two, with half following path *H* and the other half following path *L*. For the two detectors almost never respond simultaneously, and nor would detectors placed along both paths inside the interferometer.

Suppose a particular particle travels along path *H*. If path *L* is blocked, then it is just as likely to be detected at *Same* as at *Different*, since each detector measures intensity $1/2 I$ when path *L* is blocked (as the observations of Figure 2.3 already lead one to expect). But if path *L* is now reopened, it is never detected at *Same* but always at *Different*! It appears that altering a path along which a particle never travels somehow affects its behavior. Similarly, altering path *H* seems to affect the behavior of a particle assumed to travel along path *L*.

Instead of blocking a path, you can change its length. Make path *L* very slightly longer (or shorter) and you will not detect the full intensity I at *Different*—you will detect some of the original beam intensity at *Same*. Gradually changing the difference in path lengths by changing either length produces a continuous variation in the relative intensities measured at the two detectors, until for a particular path difference it is detector *Same* that measures the full intensity I while nothing is detected at *Different*. If a particle does take either path *H* or path *L* through the interferometer, then very slight variations of the length of the path it does not take are reliably correlated with variations in the relative intensities detected.

It is not always easy to gradually vary the length of a path. But there are other ways of altering a path through an interferometer to exhibit the continuous variation in

relative intensity associated with interference between paths. Inserting a thin sample of the right kind of material in a path through a neutron interferometer will not significantly reduce the intensity of neutrons detected following that path: you can check this after first blocking any other path through the device. But it can temporarily slow down any neutrons that go through it, much as light slows down while it goes through glass or water. If the sample is wedge-shaped, then you can vary the thickness neutrons pass through by gradually inserting or withdrawing the wedge. So you can continuously vary how long neutrons taking that path are delayed by their passage. This is another way of continuously varying the interference between the two paths through the interferometer, as you can observe by comparing how much intensity each detector records.

Yet another way of altering a path through a neutron interferometer reveals new and important features of single-particle interference. But since this requires devices that are not exactly familiar household objects, it is best to introduce the idea by describing its analog for beams of light.

2.3 Polarization

Most people today are familiar with polarized sunglasses and most photographers are familiar with clear polarizing filters. These both reduce glare from reflected light by absorbing light that has been polarized by that reflection. A beam of direct sunlight is not polarized. But if it is passed through a clear polarizing filter its intensity is reduced while the rest of the beam goes straight through with no color change. The emerging beam is polarized along an axis at right angles to its direction of travel determined by what is called the polarization axis of the polarizer. So far ‘polarized’ is just a word. Its cash value is revealed by what happens when the beam that emerges from one polarizer encounters a second, similar polarizer.

If their polarization axes are perfectly aligned, the second polarizer reduces the beam intensity hardly at all. But if they are at right angles to one another (and also to the direction of travel) then the second polarizer blocks essentially all of the beam. By rotating the second polarizer in its plane one can vary the intensity of the beam it transmits continuously between these two extremes. Newton studied this phenomenon and speculated that it might occur if the particles have “sides”. Fresnel developed a rival wave theory that waves of light are rapidly propagating periodic disturbances of some otherwise stationary transparent medium. He adapted this theory to the phenomena of polarization by taking these disturbances to occur in a direction at right angles to the direction in which they propagate—so-called transverse waves. But we can now observe phenomena associated with the polarization of light for which neither Newton’s particle theory nor Fresnel’s wave theory could account.

If you search the phrase “single photon detector” on the Web, you will find links to many companies offering to sell you such a device: you could be forgiven for assuming that what it detects are the single photons we are now convinced compose a beam of

light. There are observations that support that conviction, including some I'll describe in the next chapter. Einstein revealed the power of this idea in a paper he published in 1905. In two other papers published that year he first formulated his theory of relativity and showed the equivalence of mass and energy. But the one paper Einstein published in 1905 whose ideas he himself described as "very revolutionary" and which were later mentioned by the Nobel committee when awarding him his prize, adopted the viewpoint that:

the energy of a light ray spreading out from a point source is not continuously distributed over an increasing space but consists of a finite number of energy quanta which are localized at points in space, which move without dividing, and which can only be produced and absorbed as complete units. [1905, 1965]

Although someone else introduced the term later (to refer to something else!) these quanta came to be known as photons. I defer until later (Chapters 12 and 13) a careful attempt to answer the questions of what it means to say a light beam is composed of photons and whether we should believe it is. But for now let's just bypass the first question and assume a positive answer to the second.

Suppose that a single photon detector is placed behind a polarizer. A severely attenuated beam of light is directed toward this polarizer, but first encounters another similar polarizer. The beam is composed of so few photons that each separately encounters this first polarizer (see Figure 2.5).

While absorbing some photons from the beam, the first polarizer may pass others: these then encounter the second polarizer. Experience with more intense beams of light leads one to expect that, on average, the fraction passed by the second polarizer will depend on the relative orientations of the two polarizers' polarization axes. If these are aligned, essentially the same number detected passing the first polarizer with the second polarizer removed will still be detected when it is replaced: but few if any will be detected if the polarizers' axes are then crossed, that is, at right angles to each other. Observations confirm this expectation, in line with the assumption that photons in a beam behave independently in this setup. You may wonder whether photons that pass a polarizer differ somehow from those it absorbs: maybe a polarizer works like a grating of bars parallel to its axis, so photons with "sides" closely aligned with these bars slip through more often than those whose "sides" are badly misaligned with its axis? Observations in the setup of Figure 2.5 don't address this suggestion, but insertion

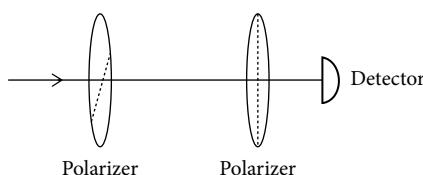


Figure 2.5 Polarization analysis

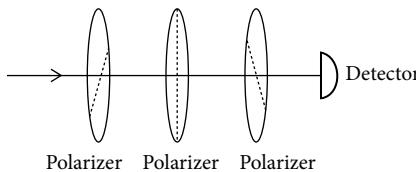


Figure 2.6 Analysis with a third polarizer

of a third similar polarizer between crossed polarizers (as in Figure 2.6) at least casts doubt on it.

If the intermediate polarizer's axis is aligned either with that of the initial polarizer or with that of the final polarizer, no photons are detected, in accord with the suggested model of photons and their interaction with polarizers. But if the axis of this intermediate polarizer is rotated around the direction of the beam, some photons will be detected. The average rate at which they are detected depends smoothly on how far it is rotated. It is greatest when the axis of the intermediate polarizer is arranged to make equal 45 degree angles with the other two polarizers' axes: in this arrangement a quarter as many photons are detected on average as when the intermediate polarizer is removed and the other polarizers' axes are aligned.

This last observation strongly suggests that polarizers don't simply act as filters, letting one type of light through unchanged while blocking light of another type. A polarizer can apparently also modify some photons while letting them pass: modification by an intermediate polarizer is sometimes enough to permit a photon to pass through each of two crossed polarizers. If a photon did have "sides" or other properties whose relation to a polarizer determined whether or not it would pass that polarizer, then its passage would have to completely "reset" these properties—for how likely it is to be detected with a third polarizer depends only on the angle between the axes of the second and third polarizers. Chapter 4 will present a famous argument against even this possibility. If cogent, this shows that photons have no properties that make some pass a polarizer while others are absorbed: qualitatively identical photons behave differently in just the same circumstances, individually at random but with statistical regularity.

Newton was not the first to observe and comment on polarization phenomena: some have speculated that the "sunstone" Vikings used to navigate the Atlantic ocean centuries earlier was a polarizing crystal of calcium carbonate still known as Iceland spar. A beam of light shone on such a crystal at the correct angle is split into two beams that separate inside the crystal to emerge parallel: one (called the ordinary ray) goes straight through, but the other (the extraordinary ray) is displaced, as shown in Figure 2.7.

The important thing is that each of these beams is polarized, with their axes of polarization at right angles to one another and to their common direction of travel on exiting the crystal. This is easily shown by placing a polarizer to intercept each

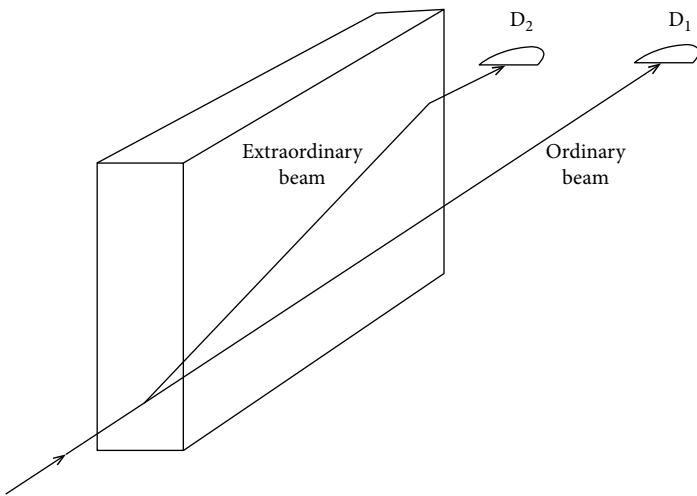


Figure 2.7 Action of calcite crystal

beam separately and noting how its intensity is observed to vary (using your eye or other detector) as that polarizer is rotated around the beam axis.

Iceland spar is a two-channel polarizer because (unlike a polarizing “filter”) it yields two oppositely polarized beams: there are others. They all act as polarizing beam splitters for light. Using mirrors if necessary, it is easy to arrange for the two beams to leave a polarizing beam splitter in perpendicular directions. There are also clear “filters” capable of rotating the polarization axis of polarized monochromatic light they pass by any desired angle: for historical reasons these are called half-wave plates.²

2.4 More Interference Phenomena

After this brief introduction to polarization of light beams it is time to turn to a more abstract level of description of a series of phenomena that can also be displayed by observations on beams of other types of matter such as electrons, atoms, and neutrons. I already noted that a beam of neutrons all with the same speed displays interference analogous to that displayed by a monochromatic beam of light. So does a beam of electrons, atoms, molecules, etc. Some beams of each of these kinds of matter also manifest behavior closely analogous to that displayed by polarized light: indeed, physicists often talk about polarized beams of neutrons, and so will I.

The analog for neutrons of photon polarization is called spin. The behavior of electrons, atoms, molecules, etc. can be modeled by assigning them a spin state, just as the behavior of photons can be modeled by assigning them a polarization state. I will explain later how this modeling works. But now I must issue a warning. I mentioned

² See the demonstration at https://www.youtube.com/watch?v=_sUVXHfUVsY.

the problems created by the assumption that a photon has some definite polarization properties that determine what will happen when it encounters a polarizer. The same kind of problems recur if one thinks of a particle's spin as a property that determines what will happen when it encounters a device that is sensitive to its spin state. I will later address the question of how the modeling techniques introduced in this chapter are able to resolve these problems, beginning in Chapter 4.

Recall the generic balanced Mach-Zehnder interferometer depicted in Figure 2.4. Polarizing a beam of matter before it enters the interferometer reduces its intensity. But it has no other effect on what is observed: the remaining intensity is still all detected at *Different*, none at *Same*. But if you also place a half-wave plate along one of the paths *H* or *L* inside the interferometer then you will observe something new (see Figure 2.8).

Rotating the half-wave plate around the direction of that path varies the relative amounts of the beam that are detected at *Different* and *Same*. At one angle *Different* detects all the intensity, at another angle *Same* detects the same intensity as *Different* (even if one then varies the path lengths), while rotating through intermediate angles continuously varies the relative detected intensities between these two extremes. So rotating its polarization axis is another way of altering a beam of matter that affects how it interferes. It can do so without changing the length of the path this beam travels.

Suppose you fix the polarization state of the incoming beam as *P* by passing it through a polarizer. By adjusting the angle of the half-wave plate you can make *Different* and *Same* detect equal intensities. At this angle the plate rotates the polarization of its beam so its polarization state is opposite to *P*. This means the two beams are oppositely polarized when they encounter the second beam splitter. The detectors then respond as if each beam behaved just the way it would if the other beam were not there at all: they respond as if there were no interference. One could then plausibly maintain that each particle either took path *H* or took path *L* between the beam splitters. Indeed one could claim to tell which of these paths a particle took by replacing the second

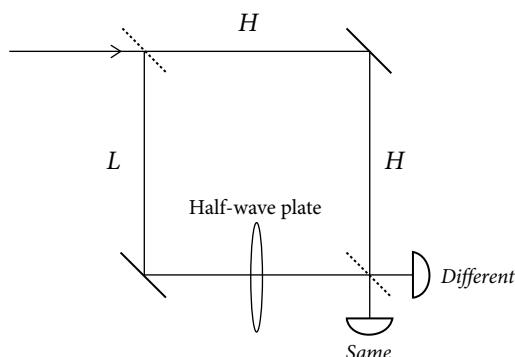


Figure 2.8 Half-wave plate in balanced interferometer

beam splitter with a polarizing beam splitter (PBS)—a device that acts as a polarizing filter as well as a beam splitter. If its axis is opposite to P , it seems that any particle detected must have followed a path taking it through the half-wave plate: if its axis is P , then any particle detected must presumably have followed the other path through the interferometer.

Instead, suppose you set the axis of the PBS half way between P and its opposite. A particle encountering the PBS is now just as likely to pass through whether its polarization state is P or its opposite. Even if such a particle had followed path H or path L through the detector, detecting it would not entitle you to say which path it had taken. You might still maintain that it must have taken one path or the other, even though you can no longer tell which. But the interesting thing is that with the axis of the PBS set like this *Same* no longer detects anything—only *Different* detects any particles! Set this way, the PBS has restored interference between the paths. If you continue to maintain that each detected particle either took path H or path L then it seems you must admit that you can detect interference only when you have removed any possibility of determining which path a detected particle took.

I have not yet mentioned a famous example of interference that Feynman [1963] once claimed contains the only mystery of quantum behavior—the two-slit experiment with electrons, as depicted in Figure 2.9.

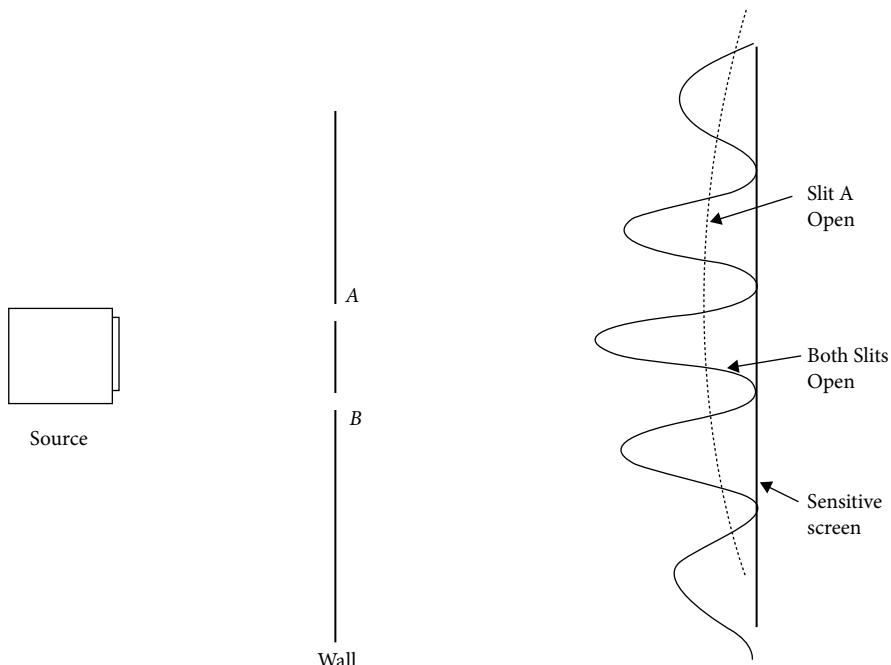


Figure 2.9 Two-slit interference

I left it till now for two reasons. First, when Feynman described it before 1963, it had not actually been done—it was merely a thought experiment: the first real experiment with individual electrons was performed (with the expected results) much later, and no experiment successfully followed Feynman's precise methodology until 2012.³ Second, if you think in terms of direct paths from source through a slit to a point on the screen, there is only a tiny spatial separation (less than a thousandth of a millimeter) between a path through one slit and a path through the other: Figure 2.9 is certainly not drawn to scale! Analogous paths through a Mach–Zehnder interferometer can be much more widely separated: paths through a neutron interferometer are separated by centimeters, making it particularly hard to imagine how a tiny subatomic particle could follow both at once.

Popular discussions of the two-slit experiment with electrons often heighten the sense of mystery by claiming that any attempt to observe through which slit an electron went will inevitably destroy the interference pattern shown in 2.9, replacing it by a smooth pattern—a simple sum of the pattern you get when only slit A is open and the pattern you get when only slit B is open. Those single-slit patterns are smooth and show none of the rapid variations of intensity we saw characterize interference fringes (remember Newton's rings). One way to observe is to shine light on the electrons and see which slit they go through by collecting the reflected light in a microscope. Heisenberg analyzed essentially this proposal in his influential [1930] book and used it to illustrate his so-called uncertainty principle. His idea was that observing the path of an electron by shining light on it would inevitably disturb its motion: the disturbance produced by an observation accurate enough to determine through which slit it went would so change its subsequent motion as to destroy the interference pattern. Heisenberg took this as an instance of the general principle that there is a strict limit on how precisely one can observe both the position and the momentum of an object.

But as we've seen, interference between two beams may be affected, or destroyed entirely, without in any way affecting the motion of a particle in either beam—it suffices to alter the polarization state of particles in one beam. Moreover, merely altering the beam polarization to destroy interference is not a way of observing which path a particle takes through an interferometer. Instead, what this does is to make possible a subsequent observational procedure that some may interpret as an indirect determination of that path. A different subsequent procedure may on the other hand remove even that possibility, thereby restoring detectable interference. Whether interference is detectable between beams of matter is not simply a function of the inevitable clumsiness involved in our attempts to observe which path a particle takes. Quantum theory itself will help us to say when such interference will be detectable, as we shall see in Chapter 5.

³ See Frabboni *et al.* [2008], Bach *et al.* [2013].