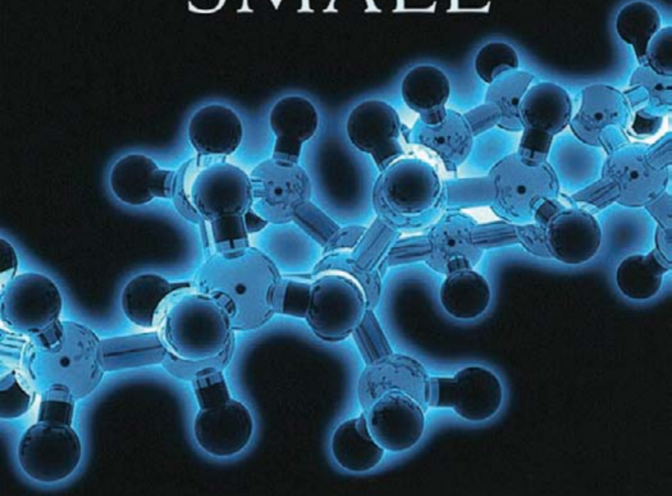
ABSOLUTELY SMALL



How Quantum Theory Explains Our Everyday World

MICHAEL D. FAYER, Ph.D.

Contents

Preface

If you are reading this, you probably fall into one of two broad categories of people. You may be one of my colleagues who is steeped in the mysteries of quantum theory and wants to see how someone writes a serious book on quantum theory with no math. Or, you may be one of the vast majority of people who look at the world around them without a clear view of why many things in everyday life are the way they are. These many things are not insignificant aspects of our environment that might be overlooked. Rather, they are important features of the world that are never explicated because they are seemingly beyond comprehension. What gives materials color, why does copper wire conduct electricity but glass doesn't, what is a trans fat anyway, and why is carbon dioxide a greenhouse gas while oxygen and nitrogen aren't? This lack of a picture of how things work arises from a seemingly insurmountable barrier to understanding. Usually that barrier is mathematics. To answer the questions posed above—and many more—requires an understanding of quantum theory, but it actually doesn't require mathematics.

This book will develop your quantum mechanics intuition, which will fundamentally change the way you view the world. You have an intuition for mechanics, but the mechanics you know is what we refer to as classical mechanics. When someone hits a long drive baseball, you know it goes up for a while, then the path turns over and the ball falls back to Earth. You know if the ball is hit harder, it takes off faster and will go farther before it hits the ground. Why does the ball behave this way? Because gravity is pulling it back to Earth. When you see the moon, you know it is orbiting the Earth. Why? Because gravity attracts the moon to the Earth. You don't sit down and start solving Newton's equations to calculate what is going on. You know from everyday experience that apples fall down not up and that if your car is going faster it will take longer to stop. However, you don't know from everyday experience why cherries are red and blueberries are blue. Color is intrinsically dependent on the quantum mechanical description of molecules. Everyday experience does not prepare you to understand the nature of things around you that depend on quantum phenomena. As mentioned here and detailed in the book, understanding features of everyday life, such as color or electricity, requires a quantum theory view of nature

Why no math? Imagine if this book contained discussions of a topic that started in English, jumped into Latin, then turned back to English. Then imagine that this jumping happened every time the details of an explanation were introduced. The language jumping is what occurs in books on quantum theory, except that instead of jumping from English to Latin, it jumps from English to math. In a hard core quantum mechanics book (for example, my own text, *Elements of Quantum Mechanics* [Oxford University Press, 2001]), you will find things like, "the interactions are described by the following set of coupled differential equations." After the equations, the text reads, "the solutions are," and more equations appear. In contrast, the presentation in this book is descriptive. Diagrams replace the many equations, with the exception of some small algebraic equations—and these simple equations are explained in detail.

Even without the usual overflow of math, the fundamental philosophical and conceptual basis for and applications of quantum theory are thoroughly developed. Therefore, anyone can come away with an understanding of quantum theory and a deeper understanding of the world around us. If you know a good deal of math, this book is still appropriate. You will acquire the conceptual understanding necessary to move on to a mathematical presentation of quantum theory. If you are willing to do some mental gymnastics, but no math, this book will provide you with the fundamentals of quantum theory, with applications to atomic and molecular matter.

1

Schrödinger's Cat

Why are cherries red and blueberries blue? What is the meaning of size? These two questions seem to be totally unrelated. But, in fact, the second question doesn't seem to be a question at all. Don't we all know the meaning of size? Some things are big, and some things are small. But, the development of quantum theory showed that the first two questions are intimately related and that we had a completely false concept of size until a couple of decades into the twentieth century. Our ideas about size, if we thought about size at all, worked just fine in our everyday lives. But beginning in approximately 1900, the physics that was used to describe all of nature, and the physics that still works remarkably well for landing a spacecraft on Mars, began to fall apart. In the end, a fundamentally new understanding of size was required not only to explain why cherries are red and blueberries are blue, but also to understand the molecules that make up our bodies, the microelectronics that run our computers, why carbon dioxide is a greenhouse gas, and how electricity can move through metals. Our everyday experiences teach us to think in terms of classical physics, the physics that was greatly advanced and formalized by Sir Isaac Newton (1642–1727). Everything we know from early childhood prepares us to view nature in a manner that is fundamentally wrong. This book is about the concept of absolute size and its consequence, quantum theory, which requires us to fundamentally change our way of thinking about nature. The first half of the book describes the basic concepts of quantum theory. The second half applies quantum theory to many aspects of the world around us through an examination of the properties of atoms and molecules and their roles in everyday life.

This book began with a simple question. Does quantum mechanics make sense? I was asked to address this question at "Wonderfest 2005, the Bay Area Festival of Science," sponsored by the University of California at Berkeley Department of Physics and the Stanford University Department of Chemistry. Wonderfest is a yearly event that presents a variety of lectures on "the latest findings" in a number of fields to an audience of nonspecialists. However, I was not asked to discuss the latest findings in my own research, but the topic, "does quantum mechanics make sense," which has been argued about by scientists and laypeople alike since the inception of quantum theory in 1900. In addition, I had only one-half hour to present my affirmative answer to the question. This was a tall order, so I spent several months thinking about the subject and a great deal of time preparing the lecture. After the event, I thought I had failed—not because it is impossible to make plain the important issues for nonspecialists, but because the time constraint was so severe. To get to the crux of the matter, certain concepts must be introduced so that contrasts between classical mechanics and quantum mechanics can be drawn. This book is my opportunity to address the quantum theory description of nature with sufficient time to do the subject justice. The book uses very simple math involving at most small equations. The idea is to make quantum theory completely accessible to the nonscientist. However, the fact that the book requires essentially no math does not mean that the material is simple. Reading Kierkegaard requires no math but is not simple. However, unlike Kierkegaard, the meaning of the material presented below should be evident to the reader who is willing to do a little mental exercise.

Classical mechanics describes the motion of a baseball, the spinning of a top, and the flight of an airplane. Quantum mechanics describes the motion of electrons and the shapes of molecules such as trans fats, as well as electrical conductivity and superconductivity. Classical mechanics is a limiting case of quantum mechanics. Quantum mechanics contains classical mechanics but not vice versa. In that respect, classical mechanics is wrong. However, we use classical mechanics to design bridges, cars, airplanes, and dams. We never worry about the fact that the designs were not done using the more general description of nature embodied in quantum theory. The use of classical mechanics will not cause the bridges to collapse, the cars to crash, the airplanes to fall from the sky, or the dams to burst. In its own realm, the realm of mechanics that we encounter in everyday life, classical mechanics works perfectly. Our intuitive feel of how the world works is built up from everyday experiences, and those experiences are, by and large, classical. Nonetheless, even in everyday life classical mechanics cannot explain why the molecules in a blueberry make it blue and the molecules in a cherry make it red. The instincts we have built up over a lifetime of observing certain aspects of nature leave us unprepared to intuitively understand other aspects of nature, even though such aspects of nature also pervade everyday life.

SCHRÖDINGER'S CAT

Schrödinger's Cat is frequently used to illustrate the paradoxes that seem to permeate the quantum mechanical description of nature. Erwin Schrödinger (1887–1961) and Paul A.M. Dirac (1902–1984) shared the Nobel Prize in Physics in 1933 for their contributions to the development of quantum theory, specifically "for the discovery of new productive forms of atomic theory." Schrödinger never liked the fundamental interpretation of the mathematics that underpins quantum theory. The ideas that bothered Schrödinger are the exact topics that will be discussed in this book. He used what has come to be known as "Schrödinger's Cat" to illustrate some of the issues that troubled him. Here, Schrödinger's Cat will be reprised in a modified version that provides a simple illustration of the fact that quantum mechanics doesn't seem to make sense when discussed in terms of everyday life. The cats offered here are to drive the issues home and are not in Schrödinger's original form, which was more esoteric. The scenario presented will be returned to later. It will be discussed as an analogy to real experiments explained by quantum theory, but not as an actual physical example of quantum mechanics in action.

Imagine that you are presented with 1000 boxes and that you are going to participate in an experiment by opening them all. You are told that there is a half-dead cat in each box. Thus, if you opened one of the boxes, you might expect to find a very sick cat. Actually, the statement needs to be clarified. The correct statement is that each of the cats is not half dead, but rather each cat is in a state that is simultaneously completely dead and perfectly healthy. It is a 50-50 mixture of dead and healthy. In other words, there is a 50% chance that it is dead and a 50% chance that it is alive. Each of the thousand cats in the thousand boxes is in the exact same state. The quantum experimentalist who prepared the boxes did not place 500 dead cats in 500 boxes and 500 live cats in the other 500 boxes. Rather, he placed identical cats that are in some sense 50-50 mixtures of dead and perfectly healthy in each box. While the cats are in the closed boxes, they do not change; they remain in the live-dead mixed state. Furthermore, you are told that when you open a box and look in, you will determine the cat's fate. The act of looking to see if the cat is alive will determine if the cat is dead or alive.

You open the first box, and you find a perfectly healthy cat. You open the next three boxes and find three dead cats. You open another box and find a live cat. When you are finished opening the 1000 boxes, you have found 500 live cats and 500 dead cats. Perhaps, more astonishing, would be if you start again with a new set of one 1000 boxes, each containing again a 50-50 mixture of live-dead cats. If you open the boxes in the same order as in the first trial, you will not necessarily get the same result for any one box. Say box 10 in the first run produced a live cat on inspection. In the second run, you may find it produced a dead cat. The first experimental run gives you no information on what any one box will contain the second time. However, after opening all 1000 boxes on the second run, you again find 500 live cats and 500 dead cats.

I have to admit to simplifying a little bit here. In two runs of the Schrödinger's Cats experiment, you probably would not get exactly 500 live and 500 dead cats on each run. This is somewhat like flipping an honest coin 1000 times. Because the probability of getting heads is one half and the probability of getting tails is one half, after 1000 flips you will get approximately 500 heads. However, you might also get 496 heads or 512 heads. The probability of getting exactly 500 heads or 500 live cats out of 1000 trials is 0.025 (2.5%). The probability of getting 496 heads is 0.024 (2.4%) and 512 heads is 0.019 (1.9%). The probability of getting only 400 heads or 400 live cats out of 1000 trials is $4.6 \times 10^{-11} = 0.000000000046$. So the probable outcomes are clustered around 500 out of a 1000 or 50%. Knowing that you have 1000 Schrödinger's Cat boxes with 50-50 mixtures of live-dead cats or 1000 flips of an honest coin, you can't say what will happen when you open one box or flip the coin one time. In fact, you can't even say exactly what will happen when you open all 1000 boxes or flip the coin 1000 times. You can say what the probability of getting a particular result is for one event and what the likely cumulative results will be for many events.

NOT LIKE FLIPPING COINS

A fundamental difference exists between Schrödinger's Cats, or more correctly real quantum experiments, and flipping pennies. Before I flip a penny, it is either heads or tails. When I flip it, I may not know what the outcome will be, but the penny starts in a well-defined state, either heads or tails, and ends in a well-defined state, either heads or tails. It is possible to construct a machine that flips a penny so precisely that it always lands with the same result. Nothing inherent in nature prevents the construction of such a machine. If a penny with heads up is inserted into the machine, a switch could determine whether the penny lands heads or tails. In flipping a coin by hand, the nonreproducibility of the flip is what randomizes the outcome. However, a box containing Schrödinger's Cat is completely different. The cat is a 50-50 mixture of live and dead. It is the act of opening the box and observing the state of the cat that causes it to change from a "mixed state" into a "pure state" of either alive or dead. It doesn't matter how precisely the boxes are opened. Unlike flipping pennies, a machine constructed to open each of the 1000 boxes exactly the same way will not make the results come out the same. The only thing that can be known about opening any one box is that there is a 50% chance of finding a live cat.

REAL PHENOMENA CAN BEHAVE LIKE SCHRÖDINGER'S CATS

As described, the Schrödinger's Cat problem cannot be actualized. However, in nature many particles and situations do behave in a manner analogous to opening Schrödinger's Cat boxes. Particles such as photons (particles of light), electrons, atoms, and molecules have "mixed states" that become "pure states" upon observation, in a manner like that described for Schrödinger's Cats. The things that make up everyday matter, processes, and phenomena behave at a fundamental level in a way that, at first, is as counterintuitive as Schrödinger's Cats. However, the problem does not lie with the behavior of electrons and atoms, but rather with our intuition of how things should behave. Our intuition is based on our everyday experiences. We take in information with our senses, which are only capable of observing phenomena that involve the behavior of matter governed by the laws of classical mechanics. It is necessary to develop a new understanding of nature and a new intuition to understand and accept the quantum mechanical world that is all around us but not intuitively understandable from our sensory perceptions.

2

Size Is Absolute

The fundamental nature of size is central to understanding the differences between the aspects of the everyday world that fit into our intuitive view of nature and the world of quantum phenomena, which is also manifested all around us. We have a good feel for the motion of baseballs, but we mainly gloss over our lack of knowledge of what gives things different colors or why the heating element in an electric stove gets hot and glows red. The motion of baseballs can be described with classical mechanics, but color and electrical heating are quantum phenomena. The differences between classical and quantum phenomena depend on the definition of size.

The quantum mechanical concept of size is the correct view, and it is different from our familiar notion of size. Our common concept of size is central to classical mechanics. The failure to treat size properly, and all of the associated consequences of that failure, is ultimately responsible for the inability of classical mechanics to properly describe and explain the behavior of the basic constituents of matter. A quantum mechanical description of matter is at the heart of technological fields as diverse as microelectronics and the computer design of pharmaceuticals.

SIZE IS RELATIVE IN EVERYDAY LIFE

In classical mechanics, size is relative. In quantum mechanics, size is absolute. What does relative versus absolute size mean, and why does it matter?

In classical mechanics and in everyday life, we determine whether something is big or small by comparing it to something else. Figure 2.1 shows two rocks. Looking at them, we would say that the rock on the left is bigger than the rock on the right. However, because there is nothing else to compare them to, we can't tell if they are what we might commonly call a big rock and a small rock. Figure 2.2 shows the rock on the left again, but this time there is something to compare it to. The size of the rock is clear because we have the size of a human hand as a reference. Because we know how big a typical hand is, we get a good feel for how big the rock is relative to the hand. Once we have the something against which to

Figure 2.1. *Two rocks.*

make a size comparison, we can say that the rock is relatively small, but not tiny. If I were to describe the rock over the phone, I could say it is somewhat bigger than the palm of your hand, and the person I am talking to would have a good idea of how big the rock is. In the absence of something of known size for comparison, there is no way to make a size determination.

Figure 2.1 demonstrates how much we rely on comparing one thing to another to determine size. In Figure 2.1, the two rocks are on a white background, with no other features for reference. Their proximity immediately leads us to compare them and to decide that the rock on the left is larger than the rock on the right. Figure 2.3 shows the rock on the right in its natural setting. Now we can see that it is actually a very large rock. The hand on the rock gives a very good reference from which to judge its size. Like the rock in the hand, the rock with the hand on top provides us with a scale that permits a relative determination of size. It is clear from these simple illustrations that under normal circumstances, we take size to be relative. We know how big something is by comparing it to something else.

FIGURE 2.2. *The rock from Figure 2.1 in a hand.*

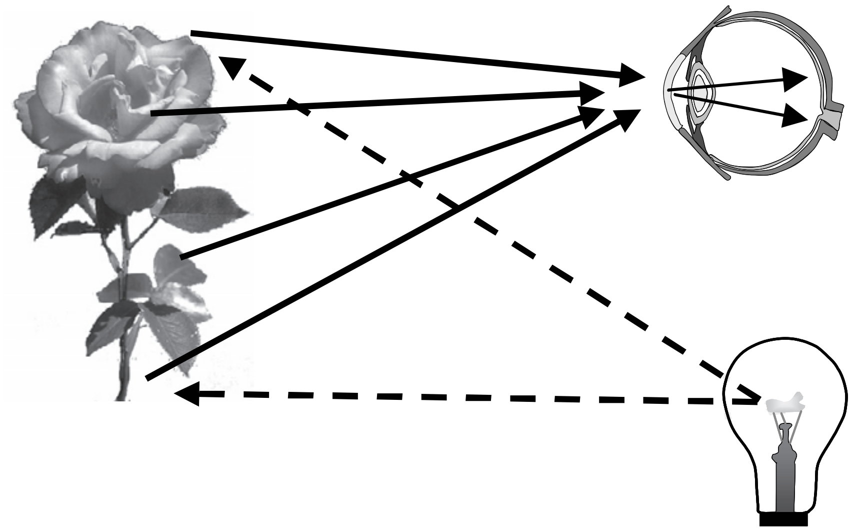
FIGURE 2.3. *The other rock from Figure 2.1, but now in a context from which its size can be judged.*

OBSERVATION METHOD CAN MATTER

Why does the definition of size, relative versus absolute, matter? To observe something, we must interact with it. This is true in both classical and quantum mechanics.

Figure 2.4 illustrates the observation of a rose. In a totally dark room, we cannot see the rose. In Figure 2.4, however, light emanating from the light bulb falls on the rose. Some of the light is absorbed, and some of it bounces off. (Which colors are absorbed, and therefore, which colors bounce off to make the leaves look green and the petals look red, is a strictly quantum mechanical phenomenon that will be discussed in Chapter 8.) A portion of the light that

FIGURE 2.4. *The light bulb illuminates the rose. The light that bounces off the rose enters the eye, enabling us to see the rose.*



bounces off is detected by the eye and processed by the brain to observe the rose. The observer is interacting with the rose through the light that bounces off of it.

Once we recognize that we must interact with an object to observe it, we are in a position to define big and small. The definition of what is big and what is small is identical in classical mechanics and quantum mechanics. If the disturbance to an object caused by an observation, which is another way of saying a measurement, is negligible, then the object is big. If the disturbance is nonnegligible, the object is small. In classical mechanics, we make the following assumption.

> *Assume: When making an observation, it is always possible to find a way to make a negligible disturbance.*

If you perform the correct experiment, then the disturbance that accompanies the measurement is negligible. Therefore, you can observe a system without changing it. However, if you do the wrong experiment in trying to study a system, you make a nonnegligible disturbance, and the object is small. A nonnegligible disturbance changes the system in some way and, it is desirable, if possible, to make a measurement that doesn't change the thing you are trying to measure. Classical theory assumes that you can reduce the size of the disturbance to be as small as desired. No matter what is under observation, it is possible to find an experimental method that will cause a negligible disturbance. This assumed ability to find an experiment that produces a negligible disturbance implies that *size is only relative*. The size of an object depends on the object and on your experimental technique. There is nothing inherent. Any object can be considered to be big by observing it with the correct method, a method that causes a negligible disturbance.

Suppose you decide to examine the wall of a room in which you are sitting by throwing many billiard balls at it. In your experiment, you will observe where the balls land after they bounce off of the wall. You start throwing balls and, pretty soon, plaster is flying all over the place. Holes appear in the wall, and the balls you throw later on don't seem to bounce off the same way the earlier balls did. This may not be surprising because of the gaping holes that your measurement method is making in the wall. You decide that this is not a very good experiment for observing the wall. You start over again after having a good painter restore the wall to its original state. This time you decide to shine light on the wall and observe the light that bounces off of the wall. You find that this method works very well. You can get a very detailed look at the wall. You observe the wall with light for an extended period of time, and the properties you observe do not change.

BIG OR SMALL—IT'S THE SIZE OF THE DISTURBANCE

When the wall was observed with billiard balls, it was small because the observation made a nonnegligible disturbance. When the wall was observed with light, it was big. The observation made a negligible disturbance. In these experiments, which can be well described with classical mechanics, the wall's size was relative. Do the poor experiment (observation with billiard balls), and the wall is small. Do a good experiment (observation with light), and the wall is big.

In classical mechanics, there is nothing intrinsic about size. Find the right experiment, and any object is big. It is up to the experimental scientist to design or develop the right experiment. Nothing intrinsic in classical mechanics theory prevents a good experiment from being performed. A good experiment is one that produces a negligible disturbance during the measurement. In other words, a good experiment does not change the object that is being observed, and, therefore, the observation is made on a big object.

CAUSALITY FOR BIG OBJECTS

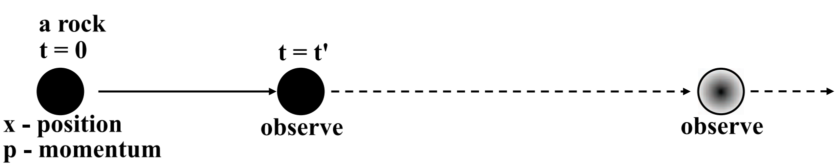
The importance of being able to make any object big is that it can be observed without changing it. Observing an object without changing it is intimately related to the concept of causality in classical mechanics. Causality can be defined and applied in many ways. One statement of causality is that equal causes have equal effects. This implies that the characteristics of a system are caused by earlier events according to the laws of physics. In other words, if you know in complete detail the previous history of a system, you will know its current state and how it will progress. This idea of causality led Pierre-Simon, Marquis de Laplace (1749–1827), one of the most renowned physicists and mathematicians, to declare that if the current state of the world were known with complete precision, the state of the world could be computed for any time in the future. Of course, we cannot know the state of the world with total accuracy, but for many systems, classical mechanics permits a very accurate prediction of future events based on accurate knowledge of the current state of a system. The prediction of the trajectory of a shell in precision artillery and the prediction of solar eclipses are examples of how well causality in classical mechanics works.

As a simple but very important example, consider the trajectory of a free particle, such as a rock flying through space. A free particle is an object that has no forces acting on it, that is, no air resistance, no gravity, etc. Physicists love discussing free particles because they are the simplest of all systems. However, it is necessary to point out that a free particle never really exists in nature. Even a rock in intergalactic space has weak gravity influencing it, weak light shining on it, and occasionally bumps into a hydrogen atom out there among the galaxies. Nonetheless, a free particle is useful to discuss and can almost be realized in a laboratory. So our free particle is a hypothetical true free particle despite its impossibility.

The free particle was set in motion some time ago with a momentum p, and at the time we will call zero, t = 0, it is at location x. x is the particle's position along the horizontal axis. The trajectory of the rock is shown in Figure 2.5 beginning at t = 0. The momentum is p = mV where m is the mass of the object and V is its velocity. The mass on earth is just its normal weight. If the rock is on the moon, it has the same mass, but it would have one-sixth the weight because of the weaker pull of gravity on the moon.

A very qualitative way to think about momentum is that it is a measure of the force that an object could exert on another object if

FIGURE 2.5. *A free particle in the form of a rock is shown moving along a trajectory.*



they collided. Imagine that a small boy weighing 50 pounds runs into you going 20 miles per hour. He will probably knock you down. Now imagine that a 200-pound man runs into you going 5 miles per hour. He will probably also knock you down. The small boy is light and moving fast. The man is heavy and moving slow. Both have the same momentum, 1000 lb−miles/hour. (lb is the unit for pound.) In some sense, both would have the same impact when they collide with you. Of course, this example should not be taken too literally. The boy might hit you in the legs while the man would hit you in the chest. But in a situation where these types of differences did not occur, either would have essentially the same effect when running into you.

Momentum is a vector because the velocity is a vector. A vector has a magnitude and a direction. The velocity is the speed and the direction. Driving north at 60 mph is not the same as driving south at 60 mph. The speed is the same, but the direction is different. The momentum has a magnitude mV and a direction because the velocity has a direction. In Figure 2.5, the motion is from left to right across the page.

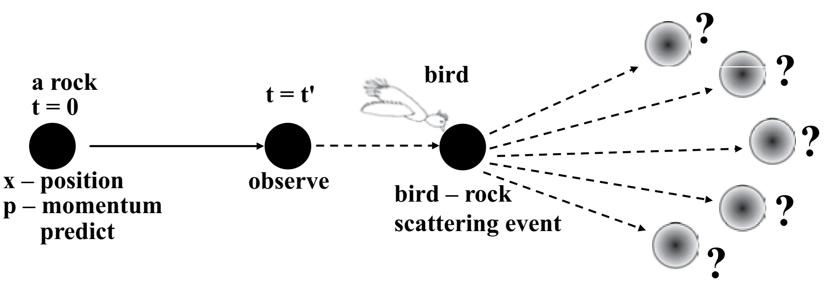
At $t = 0$ we observe (make measurements of) the rock's position and momentum. Once we know x and p at $t = 0$, we can predict the trajectory of the rock at all later times. For a free particle, predicting the trajectory is very simple. Because there are no forces acting on the particle, no air resistance to slow it down or gravity to pull it down to earth, the particle will continue in a straight line indefinitely. At some later time called t' (t prime), $t = t'$, the rock will have moved a distance $d = Vt'$. The distance is the velocity multiplied by how long the particle has been traveling. Since we started at time equal to zero, $t = 0$, then t' is how long the particle has been moving—for example, one second. So at time t' we know exactly where to look for the rock. We can make an observation to see if the particle is where we think it should be and, sure enough, there it is, as shown in Figure 2.5. We can predict where it will be at a later time, and observe that it is in fact there. This is shown on the right side of Figure 2.5. We have predicted where the particle will be, and when we make an observation, it is there. The rock is traveling with a well-defined trajectory, and the principle of causality is obeyed.

NONNEGLIGIBLE DISTURBANCES MATTER

Now consider Figure 2.6. The rock is prepared identically to the situation shown in Figure 2.5. At $t = 0$, it has position x and momentum p. Again it is observed at $t = t'$.

Its position is as predicted from the values of x and p at $t = 0$. However, some time after $t = t'$, a bird flies into a rock. (You will have to forgive my drawing of the bird. This is the best I can do on a computer with a mouse.) In the jargon of physics, we might refer to this as a bird-rock scattering event. The bird hitting the rock makes a nonnegligible disturbance. Therefore, it is not surprising that a measurement of the position and momentum made some

FIGURE 2.6. *A free particle in the form of a rock is moving along a trajectory. At time $t = 0$, it has position x and momentum p. At a later time, $t = t'$, it has moved to a new position where it is observed, and its future position is predicted. However, some time later, a bird flies into the rock. The prediction made at t' is no longer valid.*



time after the scattering event will not coincide with the predictions made based on the trajectory determined at $t = 0$. According to the precepts of classical mechanics, if we knew everything about the bird, the rock, and how they interact (collide with each other), we could make a prediction of what would happen after the bird-rock scattering event. We could test our predictions by observation. Observation is possible in classical mechanics because we can find a method for observation that makes a negligible disturbance of the system. That is, we can always find a way to make the system big. But the important point is that following a nonnegligible disturbance, it is not surprising that predictions are not fulfilled, as they were based on the known trajectory that existed prior to the disturbance.

THERE IS ALWAYS A DISTURBANCE

Quantum theory is fundamentally different from classical mechanics in the way it treats size and experimental observation; the difference makes size absolute. Dirac succinctly put forward the assumption that makes size absolute.

> *Assume: There is a limit to the fineness of our powers of observation and the smallness of the accompanying disturbance, a limit that is inherent in the nature of things and can never be surpassed by improved technique or increased skill on the part of the observer.*

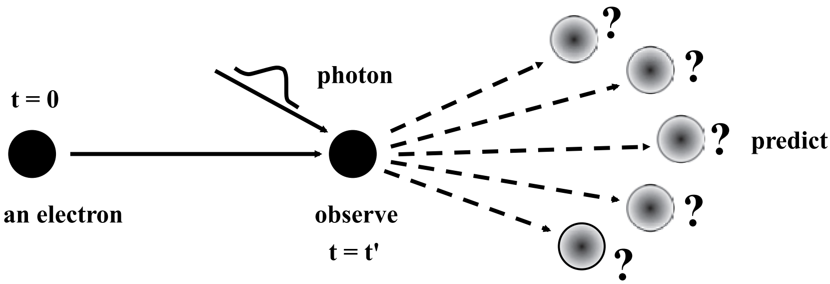
This statement is a wild departure from classical thinking. It says that whenever you observe a system (make a measurement), there is always a disturbance; it may be small, but it is always there. The size of this disturbance is part of nature. No improvements in instrumentation or new methods of observation can make this minimum disturbance vanish or become smaller.

SIZE IS ABSOLUTE

Dirac's statement has ramifications that are part of all formulations of quantum theory. His assumption immediately makes size absolute. An object is big in the absolute sense if the minimum disturbance that accompanies a measurement is negligible. An object is small in the absolute sense if the inherent minimum disturbance is not negligible. At the most fundamental level, classical mechanics is not set up to describe objects that are small in the absolute sense. In classical mechanics, any object can be made "big" by finding the right experiment to use in making an observation. In the development of classical mechanics, it was never envisioned that because of the inherent properties of nature, it was impossible to improve methodology to the point where an observation did not change a system. Therefore, classical mechanics is not set up to deal with objects that are small in an absolute sense. Its inability to treat objects that are absolutely small, such as electrons or atoms, is the reason that classical mechanics fails when it is applied to the description of such objects.

Figure 2.7 illustrates the nature of the problem. An electron is a particle that is small in the absolute sense. (Later we will discuss in detail the meaning of the word particle, which is not the same as the classical concept of particle.) At $t = 0$, it is moving along a trajectory. As with the rock, we want to see if it is actually doing what we think it is doing so that we can make subsequent predictions. We use the least invasive method to observe the electron; we let it interact with a single particle of light, a photon. (Below is a detailed discussion of the nature of light and what it means to have a particle of light.) Here is what makes this problem completely different from that illustrated in Figure 2.5. Because an electron is absolutely small, even observing it with a single particle of light causes a nonnegligible disturbance. The electron is changed by the observation. We cannot make subsequent predictions of what it will

FIGURE 2.7. *At time, $t = 0$, an electron is moving along some trajectory. At time, $t = t'$, we observe it in a minimally invasive manner by letting it interact with a single particle of light, a photon. (Photons are discussed in detail later.) The electron-photon interaction causes a nonnegligible disturbance. It is not possible to make a causal prediction of what happens after the observation.*



do once we observe it to see if it is doing what we think it is doing. Causality applies to undisturbed systems. The act of observing the electron disturbs it. You can predict what a system is doing as long as you don't look to see if it is actually doing what you think it should be doing. Therefore, causality does not apply to systems that are absolutely small. The act of observation destroys causality. Indeterminacy, that is a certain type of indefiniteness, comes into the calculation of observables for absolutely small systems. A system is absolutely small if the minimum disturbance that accompanies a measurement is not negligible. An absolutely small system can't be observed without changing it.

CAN'T CALCULATE THE FUTURE—ONLY PROBABILITIES

Unlike in classical mechanics, once an observation is made for a quantum system, it is not possible to say exactly what another observation will yield. This lack of exactitude is not like the bird hitting the rock in Figure 2.6. In the bird-rock case, it is possible, if difficult, in principle to predict the result of the next observation. We would need to know all of the properties of the bird and the rock, as well as the exact details of how the bird hit the rock (e.g., the velocities and masses of the bird and the rock and the angle at which they hit). In the electron–photon case, it is impossible to predict exactly what the results of the next observation will be. What quantum theory can do is predict the probability of obtaining a particular result. In the Schrödinger's Cats example, when a box was opened, either a dead cat or a live cat was found. There was no way to predict which it would be. Opening the box (observing the cat) changed the cat from being in a type of mixed live-dead state into either a pure live state or a pure dead state. If many boxes were opened, the probability of finding a cat alive was 50%, but there was no way to predict what would happen when a particular box was opened (a single measurement). The cat problem is not physically realizable and, therefore, it is not a true quantum mechanical problem. A physically real problem that is akin to the cat problem is discussed in a later chapter. The cat problem was intended to introduce the idea that an observation can change a system and that only a probability could be ascertained from a series of experiments. For real systems that are absolutely small, quantum mechanics is the theory that permits the calculation and understanding of the distribution of probabilities that are obtained when measurements are made on many identically prepared systems. How quantum mechanical probability distributions arise and how to think about the nature of the disturbance that accompanies measurements on absolutely small systems are discussed in the following chapters.

3

Some Things About Waves

To address the nature of the inherent disturbance that accompanies a measurement and to understand what can and cannot be measured about an absolutely small quantum mechanical system, first it is necessary to spend some time discussing classical waves and the classical description of light. At the beginning of the twentieth century, a variety of experiments produced results that could not be explained with classical mechanics. The earliest of these involved light. Therefore, we will first discuss an experiment that seemed to show that classical ideas work perfectly. Then, in Chapter 4, we will present one of the experiments that demonstrated that the classical mechanics description could not be correct and, furthermore, that a classical reanalysis of the experiment seemed to work, but actually didn't. Finally, the correct analysis of the experiment involving light will be given using quantum ideas, which will bring us back to Schrödinger's Cat.

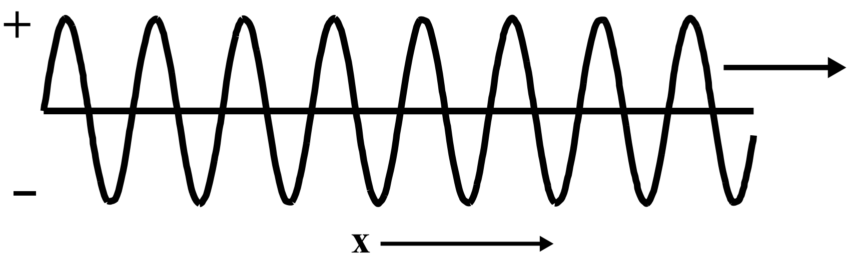
WHAT ARE WAVES?

There are many types of classical waves, water waves, sound waves, and light waves (electromagnetic waves). All waves have certain common properties, including amplitude, wavelength, speed, and direction of propagation (the direction in which a wave is traveling). Figure 3.1 shows a wave traveling in the x direction. The amplitude of the wave is the "distance" between its positive and negative peaks, the up-to-down distance. The wavelength is the distance along the direction of propagation between two positive or negative peaks. This is the distance over which the wave repeats itself. If you are riding on the wave and you move any integer number of wavelengths forward or backward along the wave, everything looks the same. The wave is traveling with some velocity, V.

WAVES HAVE VELOCITIES AND FREQUENCIES

The velocity depends on the type of wave, and the velocity of a wave needs a little discussion. Imagine you are standing beside the wave in Figure 3.1, but the wave is so long that you cannot see its beginning or end. Still, you can determine its velocity using a timing device. Start timing when a positive peak just reaches you and stop timing when the next positive peak reaches you. You now have enough information to determine the wave's velocity.

Figure 3.1. *A wave traveling in the x direction. The black line represents zero amplitude of the wave. The wave undergoes positive and negative oscillations about zero. The distance between the peaks is the wavelength. The wave is traveling along x with a velocity V.*



The wave has traveled d, a distance, of one wavelength, in time t. The distance equals the velocity multiplied by the time, $d = Vt$. (If you are in a car going at velocity, $V = 60$ miles per hour, and you travel for a time, $t = 1$ hour, then you have traveled a distance, $d = 60$ miles.) If we take the distance of one wavelength and divide it by the time it took to travel one wavelength, then we know the velocity, $V = d/t$. Watching the wave go by is like watching a very long train go by. You see boxcar after boxcar pass you. If you know the length of one boxcar and how long it takes that one boxcar to pass by, then you can determine the velocity of the train.

Another important property of waves that is related to their velocity and wavelength is the frequency. Scientists love using Greek letters to represent things because we tend to use up all of the Roman letters early on. There is no reason the velocity has to be V or distance d or time, t, but these are usually used, and many of the letters of the Roman alphabet have common usages. Therefore, we turn to the Greek alphabet. It is common to call a wave's wavelength λ (lambda) and a wave's frequency ν (nu). To see what the frequency is, again consider the train of box cars passing by. If you count how many boxcars go by in a certain amount of time, you have found the box car frequency. If 10 boxcars go by in a minute, the frequency is 10 per minute, which would usually be written as 10/minute. The frequency of a wave is determined by how many cycles (peaks) go by a point in a second. If 1000 cycles pass by a point in a second, the frequency is $\nu = 1000/s = 1000$ Hz. Lowercase s is used for the units of seconds. Per second has its own unit, Hz, for Hertz, which is in honor of Gustav Ludwig Hertz (1887–1975), who shared the Nobel Prize in Physics in 1925 with James Franck "for their discovery of the laws governing the impact of an electron upon an atom." The wavelength, velocity, and frequency of a wave are related through the equation, $\lambda\nu = V$.

OCEAN WAVES

Waves in the deep ocean travel with the crest above the average sea level and the troughs below sea level. A typical ocean wave has a wavelength $\lambda = 160$ m (520 ft) and travels with a velocity of 60 km/hr (60 kilometers per hour, or 38 miles per hour). The period, which is the time between wave crests, is 10 s, so the frequency $\nu = 0.1$ Hz. The amplitude is just the distance between a crest and a trough. Therefore, it is relatively straightforward to visualize the amplitude. (Waves break at the beach because the troughs drag on the ocean bottom in shallow water, which slows them down. The crests move faster than the troughs and fold over to produce the breaking waves we see at the beach. Waves traveling in the ocean do not break.)

SOUND WAVES

Sound waves are density waves in air. A standard tuning fork A above middle C is 440 Hz. When you strike the tuning fork, the tines vibrate at 440 Hz. The vibration produces sound waves. The tines moving back and forth "push" the air back and forth at 440 Hz, producing a wave with frequency, $\nu = 440$ Hz. At 70°F, the speed of sound is $V = 770$ miles per hour, which is 345 m/s. Because $\lambda\nu = V$, the wavelength of the 440 Hz sound wave is $\lambda = 0.78$ m (2.55 ft). The sound wave consists of air density going above the average density and then below the average density, more air and then less air. The density is the weight of air in a unit of volume, for example the number of grams in a cubic centimeter (g/cm^3). Increased density can be associated with increased pressure. So you could also think of the sound wave as a pressure wave in which the air pressure goes up and down at 440 Hz. When the sound wave enters your ear, the up-and-down oscillation of the pressure causes your eardrum to move in and out at the frequency of the sound wave, in this case, 440 Hz. The motion of the eardrum transfers the sound into the interior of the ear and tiny hairs are wiggled depending on the frequency of the sound. The motion of these hairs stimulates nerves, and the brain decodes the nerve impulses into what we perceive as sounds.

The amplitude of a sound wave is the difference between the maximum and minimum density (maximum and minimum pressure). In contrast to an ocean wave, you cannot see the amplitude of a sound wave, but you can certainly hear the differences in the amplitudes of sound waves. It is relatively simple to obtain electrical signals from sound waves, which is what a microphone does. Once an electrical signal is produced from a sound wave, its amplitude can be measured by measuring the size of the electrical signal. Like all classical waves, sound waves propagate in a direction and have an amplitude, a wavelength, and a velocity.

CLASSICAL LIGHT WAVES

The discussion of ocean waves and sound waves sets the stage for the classical description of light as light waves. In the classical description of light, explicated in great detail with Maxwell's Equations (James Clerk Maxwell, 1831–1879), light is described as an electromagnetic wave. The wave has an electric field and a magnetic field, both of which oscillate at the same frequency. You have experienced electric and magnetic fields. If you have seen a magnet pull a small object to it, then you have seen the effect of a magnetic field. The magnetic field from a magnet is static, not oscillatory as in light. You may have also seen the effects of electric fields. If you have combed your hair on a very dry day with a plastic comb, you may have noticed that your hair is attracted to it. After combing, very small bits of paper may jump to the comb as the comb is brought close to them. These effects are caused by a static electric field. An electromagnetic wave has both electric and magnetic fields that oscillate.

Unlike ocean waves, which travel in water, and sound waves, which travel in air, light waves can travel in a vacuum. In a vacuum, the velocity of light is given the symbol c, and $c = 3 \times 10^8$ m/s. The speed of light is about a million times faster than the speed of sound. This is the reason why you see distant lightning long before you hear it. Sound takes about 5 seconds to travel a mile. Light takes about 0.000005 s or 5 μs (microseconds) to travel a mile. The velocity of light is slower when it is not traveling in a vacuum. In air it is almost the same as in a vacuum, but in glass it travels at about two-thirds of c.

What is an electromagnetic wave, which is the classical description of light? In a water wave, we have the height of the water above and below sea level oscillating. In a sound wave, the air density or pressure oscillates above and below the normal values. If you take a small volume, the amount of air (number of molecules that make up air, mostly oxygen and nitrogen) goes above and below the average amount of air in the volume. In an electromagnetic wave, two things actually oscillate, an electric field and a magnetic field. We usually talk about the electric field because it is easier to measure than the magnetic field. The oscillating electric field is an electric wave. When you listen to the radio, the radio antenna is a piece of wire that detects the radio waves. Radio waves are just low frequency electromagnetic waves. They are the same as light waves, but much lower in frequency. The electric field in an electromagnetic wave oscillates positive and negative from a maximum positive amplitude value to the same negative value. The metal in a radio antenna has many electrons that can be moved by an electric field. (Electrons will be discussed in detail further on, and electrical conduction will be discussed in Chapter 19.) The oscillating electric field of a radio wave causes the electrons in the antenna to oscillate back and forth. The electronics in the radio amplify the oscillations of the electrons in the antenna and convert these oscillations into an electrical signal that drives the speakers to make the sound waves that you hear. So we can think of light classically as an oscillating electric field and an oscillating magnetic field. Both oscillate at the same frequency and travel together at the same speed in the same direction. This is why they are called electromagnetic waves.

VISIBLE LIGHT

For light in a vacuum, $\lambda\nu = c$. The visible wavelengths, that is, the wavelengths we can see with our eyes, range from 700 nm (red) to 400 nm (blue). (A nm is a nanometer, which is 10^{-9} meters or 0.000000001 meters.) The visible wavelengths of light are very small; the velocity of light is very high. Therefore, the frequencies of visible light waves are very high. Red light has $\nu = 4.3 \times 10^{14}$ Hz, and blue light has $\nu = 7.5 \times 10^{14}$ Hz. 10^{14} is 100 trillion. Contrast light frequencies to a sound wave frequency (440 Hz) or an ocean wave frequency (0.1 Hz). Unlike an ocean wave or a sound wave, there is a complication in measuring the amplitude of a light wave. The frequency of light is so high that even the most modern electronics cannot see the oscillations. Rather than measuring the amplitude of the wave, defined as the amplitude of the oscillating electric field, the intensity of light is measured. The intensity, I, is proportional to the absolute valued squared of the electric field E, which is written as $I \propto |E|^2$. The absolute value, the two vertical lines $|\ |$, just means, for example, if there is a sign, positive or negative, we ignore it and just make everything positive. A photodetector, like the CCD in a digital camera (a CCD, or charge coupled device, makes an electrical signal when light strikes it), measures the amount of light, the intensity, rather than the amplitude of a light wave. Your eye does not directly measure the frequency of light waves in contrast to your ear, which measures the frequency of sound waves.

ADDING WAVES TOGETHER—INTERFERENCE

Waves of any kind, including light waves, can be added together to give new waves. Figure 3.2 shows on the left two identical waves (same wavelength, same amplitude, propagating in the same direction) that are in phase. (The waves are actually on top of each other, but they have been displaced so that we can see them individually.) "In phase" means that the positive peaks of one wave line up exactly with the positive peaks of the other wave, and therefore, the negative peaks also line up. The vertical dashed line in Figure 3.2 shows that the peaks line up. When waves are in phase, we say that the phase difference is 0° (zero degrees). One cycle of a wave spans a phase of 360°. Starting at any point on a wave, if you go along the wave for 360°, you are in an equivalent position, like going 360° around a circle. When two identical waves are added in phase, the resultant wave has twice the amplitude. This is called constructive interference, as shown on the right side of Figure 3.2.

Waves that are 180° out of phase can also be added together. As shown on the left side of Figure 3.3, waves that are 180° out of phase have the positive peaks of the top wave exactly lined up with the negative peaks of the bottom wave and vice versa. (Again, for interference to occur the waves need to actually be on top of each other, but they have been displaced so that we can see them clearly.) The dashed vertical line in Figure 3.3 shows that the positive peak of one wave is exactly lined up with the negative peak of the other

FIGURE 3.2. *Two identical waves that are in phase. The waves undergo positive and negative oscillations about zero (horizontal line). The positive peaks line up, and the negative peaks line up. They undergo constructive interference (are added together) to form a wave with twice the amplitude.*

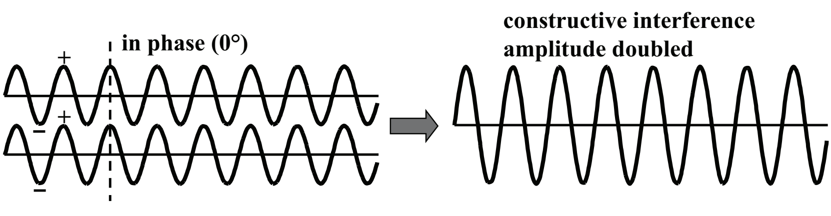
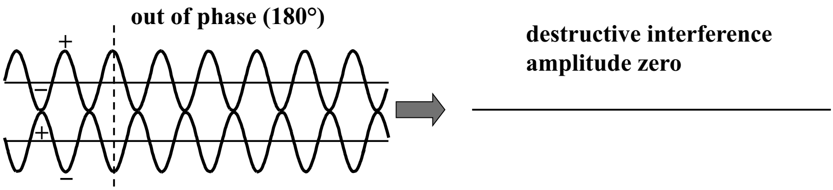


FIGURE 3.3. *Two identical waves that are 180° out of phase. The waves undergo positive and negative oscillations about zero (horizontal line). The positive peaks of the top wave line up exactly with the negative peaks of the bottom wave, and the negative peaks of the top wave line up exactly with the positive peaks of the bottom wave. The two waves undergo destructive interference when they are added together to produce zero amplitude.*



wave. When two identical waves that are 180° out of phase are added, the positive peaks and the negative peaks exactly cancel. For example, take the maximum positive value to be +1 and the maximum negative value to be −1. Adding +1 and −1 gives zero. In Figure 3.3 each point on the top wave that is positive lines up perfectly with a point on the bottom wave that is the same amount negative, and each point of the top wave that is negative lines up with an equivalent point on the bottom wave that is the same amount positive. Therefore, the waves exactly cancel to give zero amplitude as shown on the right side of the figure. This cancellation is called destructive interference.

INTERFERENCE PATTERNS AND THE OPTICAL INTERFEROMETER

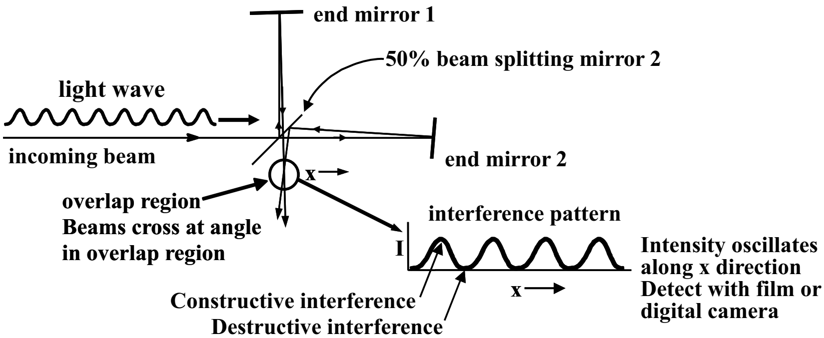
Waves do not have to be right on top of each other and going in the same direction to interfere. They just have to overlap in some region of space, and interference can occur in that region. When Davies

Symphony Hall in San Francisco was opened in 1980, it had acoustic problems. While the problems were very complicated, it is easy to see how they developed. Imagine that you are sitting in the audience pretty far back from the orchestra. When a 440 Hz A is played, the acoustic wave comes directly at you but it also bounces off of the walls on either side of you. If there is a reflection from the wall to your right and a reflection from the wall to you left so that the reflected acoustic waves (sound waves) from each wall comes to your row of seats at, for example, a 30° angle, an interference pattern will be produced along your row of seats. There will be places where reflected waves constructively interfere and make the sound louder and places where the waves destructively interfere and make the sound softer. The spacing between a peak and a null of the interference pattern is 2.4 ft (see below for the spacing formula). So depending on your seat, the 440 Hz A will be louder or softer. Of course, there are many frequency acoustic waves coming at you from many directions. The combined interference effects distorted the sound that should have been coming straight at you from the orchestra. The problem in Davies Hall was fixed in 1992 by the installation of 88 carefully designed panels hanging from the ceiling along the two side walls. No two panels are identical. They are filled with sand and weigh as much as 8500 pounds. These panels prevented the reflections from the walls from going into the audience.

Light can also undergo interference phenomena. The classical view of optical interference patterns can reproduce experimental results, as we are about to see. However, as discussed in Chapters 4 and 5, ultimately the classical description fails when other experiments are considered. The correct description will introduce the quantum mechanical superposition principle and bring us back to Schrödinger's Cats.

Figure 3.4 shows a diagram of an interferometer used by Michelson (Albert Abraham Michelson, 1853–1931) in his studies of

FIGURE 3.4. *The incoming light wave hits a 50% reflecting mirror. Half of the light goes through the mirror and half reflects from it. The light in each leg of the interferometer reflects from the end mirrors. Part of each beam crosses in the overlap region at a small angle. To the right of the circled overlap region is a blowup of what is seen along the x direction when two beams cross. An interference pattern is formed in which the intensity varies along x from a maximum value to zero periodically.*



the nature of light waves. Michelson won the Nobel Prize in Physics in 1907 "for his optical precision instruments and the spectroscopic and metrological investigations carried out with their aid." Michelson and Morley, who was a coworker of Michelson, used an interferometer to attempt to determine the nature of the medium in which light waves propagated. Water waves propagate in water. Sound waves propagate in air. The Michelson-Morley experiment showed that light waves do not have an underlying medium, which had been called the aether. Light can propagate in a vacuum. There is no aether that pervades space. Light waves traveling to us from the stars are not traveling in a medium the way ocean waves and sound waves travel in water and air, respectively. This was an important step in recognizing that light waves are not waves in the same sense as sound waves. Here we only want to understand the classical description of what is observed with an interferometer.

In Figure 3.4, a beam of light, taken to be a light wave, enters the apparatus from the left. The light hits a partially reflecting "beam-splitting" mirror that reflects 50% of the light intensity and transmits 50% of the light intensity. In the wave description of light, there is no problem having part of the wave go one way and part the other way. The reflected light goes vertically up the page, reflects from end mirror 1, which is at a small angle so the reflected beam does not quite go right back along the same path. The reflected beam goes down the page and part of it goes right through the beam-splitting mirror. (Part of this beam reflects from the beam splitter, but we are not concerned with this portion.) This path is leg 1 of the interferometer. The 50% of the original beam that goes through the beam splitter hits end mirror 2, which is also at a small angle. This reflected beam travels back to the left, almost retracing its original path. It reflects from the beam splitter. (The portion that goes through the beam splitter is unimportant for our considerations.) The reflected portion heads down the page. This path is leg 2 of the interferometer. The result is that the two beams, one that traveled leg 1 and one that traveled leg 2, come together after traveling the same distance and cross at a small angle in the "overlap region" shown by the circle in Figure 3.4. This crossing of the light waves is like the crossing of the sound waves in Davies Symphony Hall that caused the interference problems.

In Figure 3.4 the light beams are drawn as lines, but in any real experiment the beams have a width. The x direction shown in the figure is perpendicular to the bisector of the angle (the line that splits the angle) made by the crossing beams. Since the angle is small, the x direction is basically perpendicular to the propagation direction of the beams, and in the figure it is the horizontal direction. A blowup of what is seen along the x direction in the overlap region is shown in the lower right portion of the figure. In the graph the vertical axis is the intensity of the light, I, and the horizontal axis is the position along x. Because the beams cross at a small angle, the phase relationship between them varies along the x direction, and there are alternating regions of constructive and destructive interference. The intensity of the light varies from a maximum value to zero back to the maximum, and again to zero, and so on. The crossed light waves form regions of constructive and destructive interference. At the intensity maxima, the light waves are in phase (0°—see Figure 3.2), and they add constructively to give increased amplitude. At the zeros of intensity, the light waves are 180° out of phase (see Figure 3.3), and they add destructively, to exactly cancel. This pattern can be observed by placing a piece of photographic film or a digital camera in the overlap region to measure the intensity at the different points along the x direction.

For a small angle, the fringe spacing, that is, the spacing, d, between a pair of intensity peaks or nulls is given by $d = \lambda/\theta$, where λ is the wavelength of light, and θ is the angle between the beams in radians (1 radian = 57.3 degrees). If 700 nm red light is used, and the angle between the beams is 1°, the fringe spacing is 40 μm or 1.6 thousandths of an inch. These fringes can be seen with film or a good digital camera. If the angle is 0.1°, the fringe spacing is 0.4 mm, which you can see by eye. If the angle is 0.01° (an exceedingly small angle), the fringe spacing is 4 mm (about a sixth of an inch), which you can easily see by eye. To have 4 mm fringes, the beams that cross must be much larger in diameter than 4 mm.

As discussed, in the classical description, light is an electromagnetic wave, and the intensity is proportional to the square of the electric field amplitude (size of the wave in Figure 3.1). In the following, we are not going to worry about units. By including a lot of constants, the units in the following all work out, but they are unimportant for our purposes here. Take the electric field in one of the beams in one leg of the interferometer to have an amplitude of 10. Then the intensity is 100 ($10^2 = 100 = 10 \times 10$). The other beam also has I = 100. These are the intensities when we are not observing in the beam overlap region. When the beams are separated, the sum of their intensities is 200. What happens in the overlap region? Waves constructively interfere in some places and destructively interfere in others (see Figure 3.4, lower right). Therefore, to determine the intensities in the overlap region, it is necessary to add the electric field amplitudes and then square the result to find the intensities. At an intensity maximum in the overlap region, the waves are perfectly in phase and add constructively. The electric field from beam 1 adds to the electric field from beam 2, that is, $E = 10 + 10 = 20$. Then the intensity in a peak in the interference pattern is $I = E^2 = 20^2 = 400$. The intensity is 400, twice as great as the intensity of just the sum of the intensities of the two beams by themselves when they are not constructively interfering. In a null of the interference pattern, the waves destructively interfere perfectly. An electric field of $+10$ adds to an electric field of -10, to give zero. The electric field equals zero, and $I = 0$. Therefore, the interference pattern is caused by alternating regions of constructive and destructive interference of electromagnetic waves. In some places the waves add, and we see a peak. In some places the waves subtract to give zero. Interference is a well-known property of waves, and the interference pattern produced by the interferometer seemed to be a perfect example of a wave phenomenon.

The interferometer and the interference pattern shown in Figure 3.4 can be described in complete detail using classical electromagnetic theory. The details of the interference pattern can be calculated with Maxwell's equations. This and many other experiments, including the transmission of radio waves, can be described with classical theory. Therefore, classical theory, which treats light as a wave, appeared to be correct up to the beginning of the twentieth century. However, Chapter 4 shows how Einstein's explanation of one phenomenon, the photoelectric effect, caused the beautiful and seemingly infallible edifice of classical electromagnetic theory to require fundamental rethinking.

4

The Photoelectric Effect and Einstein's Explanation

AT THE END OF THE NINETEENTH CENTURY, classical electromagnetic theory was one of the great triumphs of classical mechanics. It was capable of explaining a wide variety of experimental observations. But early in the twentieth century, new experiments were causing problems for the classical wave picture of light. One experiment in particular, along with its explanation, showed a fundamental problem with the seemingly indestructible wave theory of light.

THE PHOTOELECTRIC EFFECT

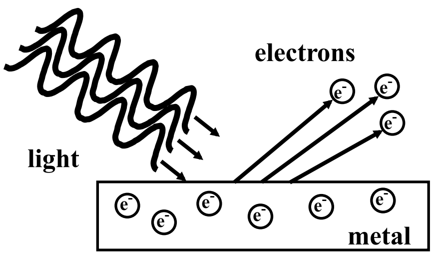
The experiment is the observation of the photoelectric effect. In the photoelectric effect, light shines on a metal surface and, under the right conditions, electrons fly out of the metal. For our purposes here, electrons are electrically charged particles. The electron charge is negative. (Later we will see that electrons are not strictly particles for the same reason that light is not a wave.) Because electrons are charged particles, they are easy to detect. They can produce electrical signals in detection equipment. Figure 4.1 shows a schematic of the photoelectric effect with the incoming light viewed as a wave.

It is possible to measure the number of electrons that come out of the metal and their speed. For a particular metal and a given color of light, say blue, it is found that the electrons come out with a well-defined speed, and that the number of electrons that come out depends on the intensity of the light. If the intensity of light is increased, more electrons come out, but each electron has the same speed, independent of the intensity of the light. If the color of light is changed to red, the electron speed is slower, and if the color is made redder and redder, the electrons' speed is slower and slower. For red enough light, electrons cease to come out of the metal.

THE WAVE PICTURE DOESN'T WORK

The problem for classical theory with these observations is that they are totally inconsistent with a wave picture of light. First, consider the intensity dependence. In the wave picture, a higher light intensity means that the amplitude of the wave is larger. Anyone who

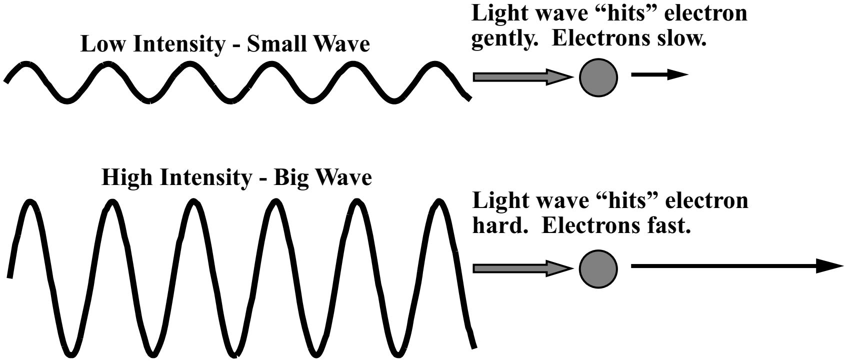
Figure 4.1. *The photoelectric effect. Light impinges on a metal, and electrons (negatively charged particles) are ejected. In the classical picture, light is a wave, and the interaction of the wave with the electrons in the metal causes them to fly out.*



has been in ocean waves knows that a small wave hits you gently and a big wave hits you hard. As illustrated in Figure 4.2, low-intensity light is an electromagnetic wave with small amplitude. Such a wave should "hit" the electrons rather gently. The electrons should emerge from the metal with a relatively slow speed. In contrast, high intensity light has associated with it a large amplitude wave. This large amplitude wave should "hit" the electrons hard, and electrons should fly away from the metal with a high speed.

To put this more clearly, the light wave has associated with it an oscillating electric field. The electric field swings from positive to negative to positive to negative at the frequency of the light. An electron in the metal will be pulled in one direction when the field is positive and pushed in the other direction when the electronic field is negative. Thus, the oscillating electric field throws the elec-

FIGURE 4.2. *A wave picture of the intensity dependence of the photoelectric effect. Low-intensity light has a small wave amplitude. Therefore, the wave should "hit" the electrons gently, and they will come out of the metal with a low speed. High-intensity light has a large wave amplitude. The large wave should hit the electrons hard, and the electrons will come out of the metal with a high speed.*



tron back and forth. According to classical theory, if the wave has large enough amplitude, it will throw the electron right out of the metal. If the amplitude of the wave is bigger (higher intensity), it will throw the electron out harder, and the electron should come out of the metal having a greater speed. However, this is not what is observed. When the intensity of light is increased, electrons come out of the metal with the same speed, but more electrons come out.

Furthermore, when the light color is shifted to the red (longer wavelength), the electrons come out of the metal with a lower speed no matter how high the intensity is. Even in the wave picture, longer wavelength light is less energetic, but it should be possible to turn up the intensity, making a bigger amplitude wave, and therefore increase the speed of the electrons that fly out of the metal. But, as with a bluer wavelength, turning up the intensity causes more electrons to emerge from the metal, but for a given color, they all come out moving with the same speed.

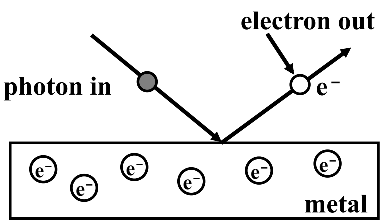
An additional problem is that when the color is shifted far enough to the red, the electrons stop coming out. The electrons have some binding energy to the metal, that is, the negatively charged electrons are attracted to the positively charged metal atom nuclei. (Atoms will be discussed in detail beginning in Chapter 9 and metals in Chapter 19.) This binding energy is what keeps the electrons from flying out of the metal in the absence of light. In the wave picture, it should always be possible to turn up the intensity high enough, and therefore make the amplitude of the oscillating electric field large enough, to overcome the binding energy. If you are standing in the ocean, a small wave may not knock you off of your feet, but if the waves get bigger and bigger, eventually they will be big enough to break the binding of your feet to the ocean floor and send you flying. But with light, for a red enough color, no matter how big the wave is, the binding of the electrons to the metal is not overcome.

EINSTEIN GIVES THE EXPLANATION

The upshot of these experimental observations is that the wave picture of light that describes the interference pattern of Figure 3.4 so well does not properly describe the photoelectric effect. The explanation for the photoelectric effect was given by Einstein in 1905 (Albert Einstein, 1879–1955). Einstein won the Nobel Prize in Physics in 1921 “for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect.” It may seem surprising that Einstein, known for his Theory of Relativity, won the Nobel Prize for explaining the photoelectric effect, which was an important step in the transition from classical to quantum theory. Einstein’s prize demonstrates the importance of the explanation of the photoelectric effect in modern physics.

Einstein said that light is not composed of waves, but rather of photons or quanta of light. In the photoelectric effect, a photon acts like a particle rather than a wave. So Einstein said that a beam of light is composed of many photons, each of which is a discrete particle. (As discussed in detail later, these are not particles in the classical sense of a particle.) As shown in Figure 4.3, one photon “hits” one electron and ejects it from the metal. The process is in

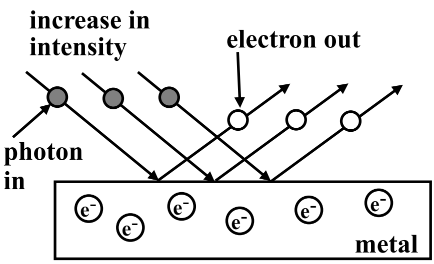
FIGURE 4.3. *Einstein described light as composed of discreet quanta of light “particles” called photons. In the photoelectric effect, one photon hits one electron and knocks it out of the metal.*



some sense like the cue ball in a game of pool hitting a stationary ball and sending it across the table. The cue ball hitting the stationary ball transfers energy to it in the form of kinetic energy, that is, energy of motion. The collision causes the cue ball to give up energy and the target ball to gain energy. A light beam is composed of many photons, but one photon ejects one electron from the metal.

When the intensity of light is increased, the light beam is composed of more photons. As illustrated in Figure 4.4, more photons impinging on the metal can hit and eject more electrons from the metal. Because one photon hits one electron, increasing the intensity of the light beam does not change the speed of the electron that is ejected. In pool, the speed of a target ball is determined by how fast the cue ball was moving. Imagine two cue balls were simultaneously shot at the same speed at two different target balls. After being hit, the two target balls would move with the same speed. When more photons of a particular color hit the metal, more electrons come out, but all with the same speed. In contrast to the wave picture, increased intensity does not produce a harder hit on an electron; increased intensity only produces more photons hitting more

FIGURE 4.4. *An increase in the intensity of a light beam corresponds to the beam being composed of more photons. More photons can hit and eject more electrons, so an increase in intensity results in more electrons flying out of the metal.*



electrons. Each photon hits an electron with the same impact whether there are many or few. Therefore, electrons come out with the same speed independent of the intensity.

RED LIGHT EJECTS SLOWER ELECTRONS THAN BLUE LIGHT

To explain why changing the color of the light to red (longer wavelength, lower energy) caused electrons to be ejected with a lower speed, Einstein used a formula first presented by Planck (Max Karl Ernst Ludwig Planck, 1858–1947). Planck first introduced the idea that energy comes in discreet units, called quanta, while he was explaining another phenomenon involving light, called black body radiation. When a piece of metal or other material is heated to a high temperature it will glow; it is emitting light. If it is quite hot, it will glow red. An example is the heating element of an electric stove or space heater when turned to high. As its temperature is increased, the color shifts toward blue. This is not only true of a piece of metal but also of stars. Red stars are relatively cool. A yellow star, such as our own sun, is hotter. A blue star is very hot. In 1900, classical physics could not explain the amount of light that came out at each color from a hot object. Planck found the explanation that still stands today by introducing a new concept, that the electrons in a piece of metal could only "oscillate" at certain discreet frequencies. The energy steps between these frequencies are called quanta. Planck won the Nobel Prize in Physics in 1918 "in recognition of the services he rendered to the advancement of Physics by his discovery of energy quanta." Planck's discovery of energy quanta led to the name Quantum Mechanics.

In his work, Planck introduced the formula that related the frequency of the electrons to their energy, $E = h\nu$, where ν is the frequency as discussed in Chapter 3, and h is called Planck's constant. In the equation, $h = 6.6 \times 10^{-34}$ J-s, J is the unit of energy

Joule, and s is seconds. In the formula, the units of ν are Hz or 1/s; so h times ν gives the units of energy, J. In his description of black body radiation, Planck postulated that the energy E could only change in discreet steps. It could be hν, of 2hν, or 3hν, and so on, but the energy could not have values between these integer step changes. The recognition that energy changes in discreet quanta at the atomic level marked the beginning of quantum mechanics.

Einstein proposed that Planck's formula also applied to photons, so that the energy of a photon was determined by its frequency ν as $E = h\nu$. Using this formula, Einstein explained the reason that red light generates slower electrons than blue light. Red light is lower frequency than blue light. Therefore a red photon is lower energy than a blue photon. In the pool ball analogy, a blue photon hits the electron harder than a red photon, and therefore, the blue photon produces an electron that has a higher speed than a red photon. With this picture, it is clear why using redder and redder light produces slower and slower electrons emerging from the metal.

VERY RED LIGHT DOES NOT EJECT ELECTRONS

The one observation left to explain is why do the electrons stop coming out of the metal when the light is tuned far enough to the red? Einstein resolved this as well. When an electron is ejected from a metal by a photon, it has a certain kinetic energy. Kinetic energy means the energy associated with its motion. The higher the energy, the faster the electron moves. The kinetic energy is E_k where the subscript k stands for kinetic. The formula for kinetic energy is given by $E_k = \frac{1}{2}mV^2$, where m is the mass and V is the velocity. Then the velocity of an electron that emerges from a metal is related to its energy, which in turn is related to the energy of the photon that knocked it out of the metal. A higher energy photon will give the electron more kinetic energy, and the electron will move faster (have a larger V). As mentioned, electrons are held in a metal by a binding energy, call it E_b, where the subscript b stands for binding. Therefore, some of the energy that is carried by the photon has to go into overcoming the binding energy. The kinetic energy of the electron that comes out of the metal is just the photon energy, $E = h\nu$, minus the binding energy, E_b. Thus, the electron's kinetic energy is $E_k = h\nu - E_b$. For an electron to be ejected from the metal, the photon energy $h\nu$ must be larger than the binding energy E_b. As the light is tuned further and further to the red (longer wavelength, λ), ν becomes smaller and smaller because $\nu = c/\lambda$, where c is the speed of light. At some red enough color, $h\nu$ is less than E_b, and electrons are no longer ejected from the metal. Turning up the intensity causes more photons to impinge on the metal, but none of these photons has enough energy to eject an electron.

The fact that electrons stop coming out of the metal when the photon is tuned far enough to red (has low enough energy) can be understood by thinking of the child's game, Red Rover. In Red Rover, a line of kids stands across a field holding hands. A kid on the other team runs at the line. If he runs very fast (high energy), he breaks through the line and keeps going, although he is slowed down. If he runs somewhat slower, he will still break through the line. However, if he runs slow enough, he will not break through the line because his energy is insufficient to overcome the binding energy of the hands holding the line together.

HOW FAST IS AN EJECTED ELECTRON

It is interesting to get a feel for how fast an electron moves when it is ejected from a piece of metal. Different metals have different binding energies called work functions. A binding energy for a metal can be determined by tuning the color redder and redder and seeing the wavelength of light at which photons cannot eject electrons. For a metal with a small binding energy, a typical cutoff wavelength for electron ejection is 800 nm. For $\lambda = 800$ nm, $\nu = 3.75 \times 10^{14}$ Hz, and $E_b = h\nu = 2.48 \times 10^{-19}$ J. If we shine green light on the metal with a wavelength of 525 nm, the energy of the photon is 3.77×10^{-19} J. The kinetic energy of the electron that will be ejected from the metal is $E_k = h\nu - E_b = 1.30 \times 10^{-19}$ J. We can find out how fast the electron is moving using $E_k = \frac{1}{2}m_e V^2 = 1.30 \times 10^{-19}$ J, where m_e is the electron mass, $m_e = 9.11 \times 10^{-31}$ kg (kg is kilograms, that is, 1000 grams). Multiplying the equation for E_k by 2 and dividing by m_e gives $V^2 = 2(1.30 \times 10^{-19}\ \text{J})/m_e = (2.60 \times 10^{-19}\ \text{J})/(9.11 \times 10^{-31}\ \text{kg}) = 2.85 \times 10^{11}\ \text{m}^2/\text{s}^2$. This value is the square of the velocity. Taking the square root, $V = 5.34 \times 10^5$ m/s, which is about one million miles per hour. In this example of the photoelectric effect, the ejected electrons are really moving.

Classical electromagnetic theory describing light as waves seems to work perfectly in describing a vast array of phenomena including interference, but it can't come close to explaining the photoelectric effect. Einstein explains the photoelectric effect, but now light can't be waves, so what happens to the classical description of interference? Reconciling the photoelectric effect and interference brings us to the cusp of quantum theory and back to Schrödinger's Cats.