CHAPTER ONE

Our Idea of the Physical World

The earth was formed about four billion years ago. During the most recent few hundred thousand years, that is, during the most recent ten-thousandth of the existence of the earth, humans have evolved nervous systems that allow us to sense a little bit, actually a very tiny fraction, of what is going on around us. Our physiological mechanisms of communication—speaking, drawing, writing—have also evolved. As a result of that ability to communicate over time and location, we have been able to accumulate knowledge and understanding of more of the world we live in and to develop means—microscopes, telescopes, radar, X-rays, MRIs—that allow us to sense much more than our physiologies provide.

By far the most immediately useful information about the physical world comes to us directly by means of our senses, especially hearing, vision, and touch. We believe we can sense almost everything that's going on around us, but our senses provide us with an astonishingly small fraction of the information that we are actually imbedded in, and we have generated our conception of the physical world on the basis of the extremely limited range of things in the physical world that can be detected by our physiology.

For instance, we talk as if there are *things,* objects, around us that are fixed and solid—that table, this book—things that are *there.* We say we see this book, but we are actually interacting not with the book but rather with the light reflected from the book. Further, the properties of the book are not at all what our senses tell us. It is made of gigantic quantities of tiny bits, "subatomic particles," that are constantly in motion, with big spaces between them and forces that pull the bits together and push them apart. We talk as though in between objects there is just space, maybe filled with air and sometimes light, but the spaces are actually packed with streams of waves of all kinds of energies, which we know about only because of the accumulation of scientific information gathered from devices that can detect things we can't. Therefore, from our experiences, each of us has put together a concept of the world that is based on a severely restricted portion of the information that is actually present in the physical world, and most of the physiological mechanisms that have evolved in us support that often misleading concept of the world.

Our visual systems have evolved a way of sensing light, which will be discussed at considerable length in the following chapters, and it is a good example of how severely limited our view of the physical world really is. The world is permeated with electromagnetic waves of all kinds. The waves emitted by a typical AM radio transmitter have wavelengths from about 1 meter to about 10,000 meters; X-rays have wavelengths around one ten-thousandth of a millionth (not a typo) of a meter; and various sources, for instance the sun, emit wavelengths at ranges in between. Our eyes have evolved in such a way that we can detect only wavelengths from about 0.4 millionths of a meter to about 0.8 millionths of a meter.

The gap in the middle of figure 1.1 is the way that the range of visible wavelengths is frequently presented. (The row of numbers at the top, labeled "wavelength in meters," is in what is called scientific notation: 10 raised to various exponents. For example, 10^{-12} means 0.000000000001, ten with 11 zeros in front of it.)

The whole width of the drawing represents a range of wavelengths of familiar sources, from gamma rays to radio, and that narrow strip near the middle that is stretched out below represents the range of

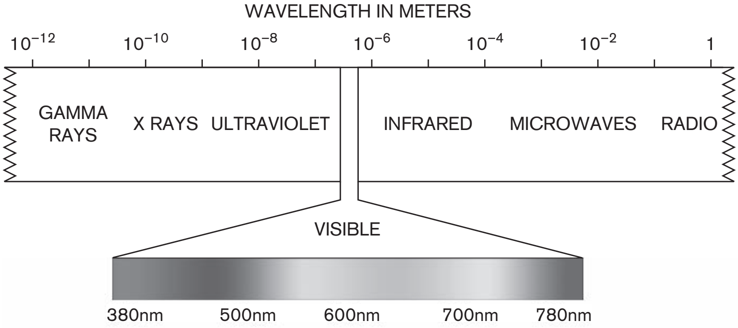

Figure 1.1. Range of wavelengths from gamma rays to radio, with the range of visible wavelengths expanded.

visible wavelengths. That diagram is correct but extremely misleading. Note that the numbers given for wavelength represent what is called a logarithmic scale. That is, each equal space, such as between 10^{-2} (0.01) meters and 1 meter, represents not an equal increase but a 100-fold increase. The distances or lengths in the world we experience are not on a logarithmic scale, and very few people can look at such a scale and understand what it really means.

If, instead, we consider actual lengths or distances, not their logarithms, and we represent the range between AM radio and X-ray wavelengths as the distance from New York to Los Angeles, then the wavelengths we are able to see would be represented on that scale as a distance of less than *an eighth of an inch.* Science had to invent instruments to detect wavelengths represented by the rest of that distance.

Within that extremely restricted visible range of wavelengths, we have evolved physiological devices, called the rod and cone systems, that actually sense different sub-ranges. The rod system is sensitive to one sub-range and the cones to three different sub-ranges, providing us with vision at very low light levels (rods) and in color (cones). Those mechanisms will be discussed in detail in later chapters. We are also limited in the range of brightnesses over which we can see. To understand that limitation, a different aspect of electromagnetic waves will be considered.

A LITTLE BACKGROUND ABOUT LIGHT

To combine, and modify a little, things that Einstein, Bohr, and Feynman have said, "If you think you understand light, you haven't thought deeply enough." Light will be discussed a lot in this book without trying to explain it. But it will be helpful, and not entirely wrong or misleading, to think of light and all other forms of electromagnetic radiation, such as radio waves and X-rays, in the following way.

Water Waves

Try partly filling the bathtub and dropping a pea into the middle of the water. Waves will of course radiate out from where the pea was dropped because the molecules of water under the pea will be pushed down, which will push the molecules next to them away and up, making a rising hump, and since water has the same properties in every direction, it will form in all directions, making a ring. Then those risen molecules will be higher than the rest of the water, so they will push down on their neighbors, making a new ring of humps, and the ring will expand. Meanwhile, the molecules that were pushed down by the pea will be pushed back up by their neighbors, and, having momentum, will keep going up (but not as far as they went down, because of friction among them). Then they will fall back down, starting the cycle over again, each time moving up and down a little less, until the ripples die out.

Why do waves seem to get smaller as they move away from their source? To detect a wave, like the one in the bathtub, the wave has to be detected over some finite part of it. For example, a cork intersects the wave over the width of the cork and detects the wave by its up-and-down motion. Similarly, if you look at a wave to try to determine its height, you only make the height judgment by watching a short stretch of the wave. Each wave forms a circle that expands as it travels away from its source, so the farther a wave has traveled, the greater is its radius and the smaller is the proportion of the wave that will be detected.

Because the circumference of a circle increases in direct proportion to its radius (circumference = pi × radius × 2), the proportion of the energy in a water wave that is detected is inversely proportional to the distance it has traveled. (That's not quite true with a water wave in a bathtub because the molecules of water exert a little friction on each other, which uses up some of the energy that the pea transferred to the water. A wave in the ocean can gather energy from wind and differences in water temperature, and so usually doesn't get smaller as it travels.)

That is a very crude and not quite accurate description of why water waves radiate out from a dropped pea, making a little group of rings that decrease in height as they travel away from the center, but it's a start on an explanation of electromagnetic waves.

Now suppose you want to detect or sense the water wave. You might put a cork in the water and measure how it moves up and down as the wave passes. The distance the cork travels up and then down is a measure of the strength, really the *amplitude,* of the wave at that particular place, and the number of times the cork moves up and down each second is the *frequency* of the wave at that point. If you have two corks and move one farther away from the center of the wave than the other until the two corks are moving up and down together, as in figure 1.2, the distance between the corks is what's called the *wavelength* of the wave.

Here are some more words and concepts that are useful in understanding waves. A *force* is a push or pull against an object. *Energy* is basically defined as the amount of whatever it is that moves something against a force. For the water wave, since the water molecules are pushed and pulled by the force of gravity, the wave must carry energy, but figuring out how much is a little tricky. You can weigh the cork and measure its amplitude of movement, which gives you a measure of energy, but that is just sampling the energy in the short segment of the wave that intersects the cork. That is just the energy that hits the cork. To have a measure of the energy of the entire wave, all the vertical movements should be added around the entire circumference of the wave.

Figure 1.2. Water waves with corks marking crests. The distance between corks is the wavelength.

Light and Other Electromagnetic Waves

Suppose somebody hands you two objects, one to hold in each hand, and you find that you have to make an effort to hold them apart; they keep trying to get together. Or you have to exert effort to keep them from flying apart. The experience of that effort is what physicists call *force*. You have to exert a force to keep them from coming together or moving apart.

It is critically important to be clear about the subtle distinction between how we conceive of the world and what we actually observe. For example, the first few sentences in the preceding paragraph could have said "A force is strength or energy as an attribute of physical action or movements," or "A force is a push or pull upon an object resulting from the object's interaction with another object," both quotes from definitions of "force" on the internet. But that kind of sentence implies that there are forces, things, out there. No one has seen piles of forces lying around. Instead, scientists have observed the behaviors of objects, have had experiences, and have given the experiences names. Making this important distinction often requires somewhat awkward and wordy sentences, but it is worth it. In fiction brevity is elegant. In explanations, it can sometimes be confusing.

If you have exerted force over some distance, you have expended *energy*. The definition of energy is the exertion of a force over a distance. An object is said to have *potential* energy if there is a force acting on it but it doesn't move, and *kinetic* energy if there is a force acting on it and it does move. When electromagnetic energy, for example light, travels through a vacuum—think of it as fast-traveling packets of energy—nothing pushes against it and it doesn't push against anything, but if it hits an object, it will transfer some or all of its energy to the object, making the object move (or move in a different way than it was already moving), and that requires energy. So we say that a beam of light traveling in a vacuum has potential energy, and if it hits something, some or all of that energy becomes kinetic energy.

We describe matter as made up of things we call atoms. Some of the things all atoms contain are called electrons, and others are called protons; atoms of different materials normally contain different numbers of electrons and protons. Protons exert forces against each other but attract electrons, and electrons exert forces against each other but attract protons.

Usually, the forces among electrons and protons in the atoms that make up any object are balanced. However, it's easy to spoil that balance. For instance, walk across a carpet on a dry day, shuffling your feet a little. That will rub off some of the electrons from the atoms in the carpet and attach them to your feet, giving you more electrons than are balanced by your protons, and those extra electrons, pushing against each other, will spread over your whole body. Then if you touch a doorknob or your dog, unless they happen to have the same imbalance between electrons and protons, your electrons will push the extra electrons to the knob or dog in the form of a spark of "static electricity."

When an object has more electrons than it would at balance, it is said to have a negative charge. Too few electrons create a positive charge. All these words have been leading up to describing what creates an electromagnetic wave. *Whenever an object has a charge and it moves, a particular kind of wave, an electromagnetic wave, is emitted* and travels off. If you walk across a carpet, picking up electrons, and then wave your hand back and forth, you generate an electromagnetic wave. The *frequency* of the wave equals the number of times you wave back and forth per second, and the *amplitude* of the wave is proportional to the distance your hand travels in each cycle.

The electromagnetic waves your charged hand makes travel away at the speed of light (very slightly slower in air than in a vacuum), so each movement back and then forth creates a wave that goes back and then forth over some distance, the *wavelength.* For instance, if your hand made one complete cycle back and forth in a hundredth of a second (so the frequency of the electromagnetic wave is a hundred cycles per second) and it travels 300,000,000 meters per second, its wavelength is 300,000,000/100 = 3,000,000 meters.

(We don't actually observe the wave traveling from the source to its detector. We observe that when a charge moves, some time later a detector responds, and there is a delay between the movement and the detection that depends on the distance between the two. Whatever it is that traveled from the source to the detector exhibits some properties of waves that we can observe, for instance water waves—and some properties of particles, as will be discussed below.)

Radio broadcasting stations have electronic devices that cause electrons to move back and forth and emit wavelengths in the neighborhood of one meter. Cell phones transmit and receive signals at about three-tenths of a meter. Wi-Fi is about one-tenth of a meter. Visible electromagnetic radiation, which we call light, has wavelengths between about 0.4 millionths of a meter and 0.8 millionths of a meter. X-rays have wavelengths in the neighborhood of one ten-thousandth of a millionth of a meter. All electromagnetic waves have the same basic properties; they are waves of energy of different frequencies and amplitudes, traveling extremely fast. Our space is crammed full of electromagnetic waves traveling in all directions, but, as shown in figure 1.1, our physiology is capable of sensing only a very narrow range of them.

In many ways, electromagnetic energy acts as water waves do. However, when electromagnetic waves are being detected, they act as though they consist of a stream of separate packets of energy, like bullets. Suppose the amplitude of the light wave, where it hits the detector, is extremely small. The kinds of devices used to detect light contain a material that converts the energy in light into what we can think of as shifts in the position of one or more electrons in the molecules of the material, and the physics of that material permits electrons to shift only among a limited number of positions. (These "positions" are usually called energy levels.) As a result, an extremely weak wave may not deliver enough energy to shift even one electron, and the presence of the wave may not be detected.

If, as it intersects the detector, the wave has a little more energy, it may trigger the shift of one or maybe two electrons, and as the energy contained in the portion of the wave intersecting the detector increases, the *number* of electron shifts or jumps will increase. If the energy in a wave is too small to shift an electron in a detector, there is no way to know whether or not the wave is actually present, so we say there was no wave, or if it has enough energy to shift some number of electrons but not enough to shift that number plus one, then we say that the amount of light, when it's being detected, increases in steps. It acts as though it consists of particles of energy.

The physics of this process does not allow the shifting of a fraction of an electron. Therefore, the intensity of a light wave must be described or measured as a whole number of shifts per unit of time. (The reason some subatomic particles have charges—and as a result the shifts in electrons are constrained to a limited number of positions or levels—is one of those deep spots into which we will try to avoid stepping.) So when light is detected, it is as if the light is a stream of particles of energy, unlike the way a wave behaves. These particles of energy are called *quanta,* and the kinds of quanta that we can see, quanta of light, are usually called *photons.*

A light wave traveling though empty space expands as a sphere. For it to be detected, its energy must be absorbed over an area, that is, over a two-dimensional detector such as a solar cell, a pixel in a camera, or a photodetector in the eye. The area of the surface of a sphere increases in proportion to the square of its radius. Therefore, so long as the area of the detector doesn't change, the energy it detects will be inversely proportional to the square of the distance between it and the source. Imagine how much energy is being emitted by the nearest visible star (other than our sun), which is the third-brightest star in the night sky and about 25,000,000,000,000 miles away. For us to see it, it must emit enough energy *in all directions* so that when it passes through a hole six or so millimeters in diameter (your pupil), it stimulates your retina strongly enough that you can see it.

Not only is human vision limited to a small range of wavelengths, it is also limited to a range of *intensities.* If you've been in darkness for about a half hour or more, a small source of light has to deliver about five or ten photons to your eye in about a tenth of a second for you to reliably detect it. At the lower end of the range of intensities of human vision, if you've been in darkness for a while, you could detect the approach of an albino lion on a moonless night with a cloudy sky.

At the other end of the range, after you've been looking for a few seconds, you can see well on a beach or snow field at noon, when the illumination is at least a billion times as intense. So human vision operates over a range of intensities of at least 1,000,000,000-fold.

But in order for us to do what we call seeing, we need information not just about whether or not light is falling on our retinas, but also about the distribution of the light in the scene. (Image formation and the ways that our eyes perform it will be discussed in chapter 8.)

For example, suppose you are outside on a clear night lit by a full moon, and the range of intensities falling on your retina in the image of the scene is from 1,000 photons per second in the dimmest place to

100,000 per second in the brightest place. Then suddenly the street lights are turned on. That will increase the illumination on the scene about 100-fold. During the first second or two after the street lights are turned on, everything will look equally bright. No detail will be visible. But then, photochemical processes (discussed in later chapters) will quickly shift the 100-fold range of vision upward, so that the range becomes from 100,000 to 10,000,000 photons per second, and you will be able to see detail again. That shift is called *light adaptation.* A change in illumination in the opposite direction—the street lights suddenly going off—produces the opposite effect, *dark adaptation,* but the downward shift in range is significantly slower.

Despite the processes of light and dark adaptation, the range of levels of illumination over which we humans can see is still somewhat limited—owls can see well at significantly lower light levels than we can.

The upper limit of the range of intensities over which our visual systems operate depends on the wavelength of the light. Two quite different mechanisms determine this limit. As explained in detail in chapter 3, when visual pigment molecules absorb light, they change their form to one that no longer absorbs light. They are said to be *bleached.* Therefore, for some wavelengths, the upper limit is reached when the light is so intense that essentially all the visual pigment molecules are in that bleached state. However, for most wavelengths—those outside or near the edges of the small range of vision indicated in figure 1.1—high intensities will cause the tissues of the retina to boil before they become great enough to bleach essentially all the pigment, setting an upper limit in a more permanent way.

Very few situations occur on Earth under which the natural levels of illumination exceed our normal range. Looking at the sun is one of them.

STARING AT THE SUN CAN CAUSE PERMANENT EYE DAMAGE.

Limitations in the range of operation exist in all of our sense modalities. For example, until undersea sound sensors were developed, we were unaware of the very long-wavelength (low-frequency) sounds emitted and sensed by whales and the very short-wavelength (high-frequency) clicks generated by crustaceans that abound in the depths of the ocean. Mechanisms to sense those frequencies had little survival value in the evolution of humans.

We call carbon monoxide odorless, but that is not a property of the gas, it is a limitation of our olfactory systems. If carbon monoxide occurred with much abundance on the earth, we would probably have evolved a biological detector. Glossy paper feels smooth, but under a microscope the roughness of its texture becomes obvious.

As discussed above, the range of all the wavelengths in the universe that we can actually use to see is extremely limited, and we will examine the reasons for this limitation at the molecular level in later chapters.

THINGS TO THINK ABOUT

There will be a few questions following each chapter. You might find answers to some of these questions in later chapters, but don't go looking for those answers now. Try to answer the questions yourself. You will probably be tempted to look for answers on the internet. Be warned that most of the "information" in cyberspace is simply wrong.

Often, we think we understand an issue until we try to explain it. How many times have teachers heard, "I know the answer but I just can't put it into words." Baloney! You don't know the answer until you *can* put it into words. Write down your answers to these questions or try to explain them to someone else.

1. Bees can see light of wavelengths shorter than we can see, that is, *ultraviolet*. Further, pictures of some flowers, when taken with a camera that is sensitive to ultraviolet light, show strong bull's-eye-like patterns that are not visible to the human eye. Do you think that bees' eyes and those flowers evolved together or that one came before the other, and if so, which came first? Can you be explicit about the reasoning behind your belief?

2. Here's another way to get a feeling for logarithmic scales. Suppose you could save a thousand dollars (10^3) every year in a box under your bed. How many years would it take you to save up a million dollars (10^6)? (In hundred-dollar bills, this would be a stack 43 inches high.) If you could save a million dollars a year, how many years would it take you to save up a billion dollars (10^9)? How high would that stack be? (The tallest building in the world, in 2017, is 32,664 inches high.)

RELEVANT READING

Even though it is more than fifty years old, for a clear explanation of the physical principles underlying optics (and for all of physics), there is nothing better than the Feynman Lectures:

Richard Phillips Feynman, Robert B. Leighton, and Matthew Sands. 1963. *The Feynman Lectures on Physics*. Addison-Wesley.