The Hubble Space Telescope was placed into orbit with faulty optics in 1990. Its main mirror was supposed to have been nearly parabolic, since it is an astronomical telescope, meant for producing images of objects at infinity. However, contractor Perkin Elmer had delivered a faulty mirror, which produced aberrations. The large photo shows astronauts putting correcting mirrors in place in 1993. The two small photos show images produced by the telescope before and after the fix.

12.4 Refraction

Economists normally consider free markets to be the natural way of judging the monetary value of something, but social scientists also use questionnaires to gauge the relative value of privileges, disadvantages, or possessions that cannot be bought or sold. They ask people to imagine that they could trade one thing for another and ask which they would choose. One interesting result is that the average light-skinned person in the U.S. would rather lose an arm than suffer the racist treatment routinely endured by African-Americans. Even more impressive is the value of sight. Many prospective parents can imagine without too much fear having a deaf child, but would have a far more difficult time coping with raising a blind one.

So great is the value attached to sight that some have imbued it with mystical aspects. Joan of Arc saw visions, and my college has a “vision statement.” Christian fundamentalists who perceive a conflict between evolution and their religion have claimed that the eye is such a perfect device that it could never have arisen through a process as helter-skelter as evolution, or that it could not have evolved because half of an eye would be useless. In fact, the structure of an eye is fundamentally dictated by physics, and it has arisen separately by evolution somewhere between eight and 40 times, depending on which biologist you ask. We humans have a version of the eye that can be traced back to the evolution of a light-sensitive “eye spot” on the head of an ancient invertebrate. A sunken pit then developed so that the eye would only receive light from one direction, allowing the organism to tell where the light was coming from. (Modern flatworms have this type of eye.) The top of the pit then became partially covered, leaving a hole, for even greater directionality (as in the nautilus). At some point the cavity became filled with jelly, and this jelly finally became a lens, resulting in the
general type of eye that we share with the bony fishes and other vertebrates. Far from being a perfect device, the vertebrate eye is marred by a serious design flaw due to the lack of planning or intelligent design in evolution: the nerve cells of the retina and the blood vessels that serve them are all in front of the light-sensitive cells, blocking part of the light. Squids and other molluscs, whose eyes evolved on a separate branch of the evolutionary tree, have a more sensible arrangement, with the light-sensitive cells out in front.

12.4.1 Refraction

The fundamental physical phenomenon at work in the eye is that when light crosses a boundary between two media (such as air and the eye’s jelly), part of its energy is reflected, but part passes into the new medium. In the ray model of light, we describe the original ray as splitting into a reflected ray and a transmitted one (the one that gets through the boundary). Of course the reflected ray goes in a direction that is different from that of the original one, according to the rules of reflection we have already studied. More surprisingly — and this is the crucial point for making your eye focus light — the transmitted ray is bent somewhat as well. This bending phenomenon is called refraction. The origin of the word is the same as that of the word “fracture,” i.e., the ray is bent or “broken.” (Keep in mind, however, that light rays are not physical objects that can really be “broken.”) Refraction occurs with all waves, not just light waves.

The actual anatomy of the eye, b, is quite complex, but in essence it is very much like every other optical device based on refraction. The rays are bent when they pass through the front surface of the eye, c. Rays that enter farther from the central axis are bent more, with the result that an image is formed on the retina. There is only one slightly novel aspect of the situation. In most human-built optical devices, such as a movie projector, the light is bent as it passes into a lens, bent again as it reemerges, and then reaches a focus beyond the lens. In the eye, however, the “screen” is inside the eye, so the rays are only refracted once, on entering the jelly, and never emerge again.

A common misconception is that the “lens” of the eye is what does the focusing. All the transparent parts of the eye are made of fairly similar stuff, so the dramatic change in medium is when a ray crosses from the air into the eye (at the outside surface of the cornea). This is where nearly all the refraction takes place. The lens medium differs only slightly in its optical properties from the rest of the eye, so very little refraction occurs as light enters and exits the lens. The lens, whose shape is adjusted by muscles attached to it, is only meant for fine-tuning the focus to form images of near or far objects.
Refraction

What are the rules governing refraction? The first thing to observe is that just as with reflection, the new, bent part of the ray lies in the same plane as the normal (perpendicular) and the incident ray, d.

If you try shooting a beam of light at the boundary between two substances, say water and air, you’ll find that regardless of the angle at which you send in the beam, the part of the beam in the water is always closer to the normal line, e. It doesn’t matter if the ray is entering the water or leaving, so refraction is symmetric with respect to time-reversal, f.

If, instead of water and air, you try another combination of substances, say plastic and gasoline, again you’ll find that the ray’s angle with respect to the normal is consistently smaller in one and larger in the other. Also, we find that if substance A has rays closer to normal than in B, and B has rays closer to normal than in C, then A has rays closer to normal than C. This means that we can rank-order all materials according to their refractive properties. Isaac Newton did so, including in his list many amusing substances, such as “Danzig vitriol” and “a pseudo-topazius, being a natural, pellucid, brittle, hairy stone, of a yellow color.” Several general rules can be inferred from such a list:

- Vacuum lies at one end of the list. In refraction across the interface between vacuum and any other medium, the other medium has rays closer to the normal.

- Among gases, the ray gets closer to the normal if you increase the density of the gas by pressurizing it more.

- The refractive properties of liquid mixtures and solutions vary in a smooth and systematic manner as the proportions of the mixture are changed.

- Denser substances usually, but not always, have rays closer to the normal.

The second and third rules provide us with a method for measuring the density of an unknown sample of gas, or the concentration of a solution. The latter technique is very commonly used, and the CRC Handbook of Physics and Chemistry, for instance, contains extensive tables of the refractive properties of sugar solutions, cat urine, and so on.

Snell’s law

The numerical rule governing refraction was discovered by Snell, who must have collected experimental data something like what is
The relationship between the angles in refraction.

Snell further found that if media A and B gave a constant $K_{AB}$ and media B and C gave a constant $K_{BC}$, then refraction at an interface between A and C would be described by a constant equal to the product, $K_{AC} = K_{AB}K_{BC}$. This is exactly what one would expect if the constant depended on the ratio of some number characterizing one medium to the number characteristic of the second medium. This number is called the index of refraction of the medium, written as $n$ in equations. Since measuring the angles would only allow him to determine the ratio of the indices of refraction of two media, Snell had to pick some medium and define it as having $n = 1$. He chose to define vacuum as having $n = 1$. (The index of refraction of air at normal atmospheric pressure is 1.0003, so for most purposes it is a good approximation to assume that air has $n = 1$.) He also had to decide which way to define the ratio, and he chose to define it so that media with their rays closer to the normal would have larger indices of refraction. This had the advantage that denser media would typically have higher indices of refraction, and for this reason the index of refraction is also referred to as the optical density. Written in terms of indices of refraction, Snell’s equation becomes

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{n_2}{n_1},$$

but rewriting it in the form

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

[relationship between angles of rays at the interface between media with indices of refraction $n_1$ and $n_2$; angles are defined with respect to the normal]

makes us less likely to get the 1’s and 2’s mixed up, so this is the way most people remember Snell’s law. A few indices of refraction are given in the back of the book.

self-check E
(1) What would the graph look like for two substances with the same index of refraction?

(2) Based on the graph, when does refraction at an air-water interface change the direction of a ray most strongly?  

Answer, p. 1066
Finding an angle using Snell’s law example 10

A submarine shines its searchlight up toward the surface of the water. What is the angle $\alpha$ shown in the figure?

The tricky part is that Snell’s law refers to the angles with respect to the normal. Forgetting this is a very common mistake. The beam is at an angle of $30^\circ$ with respect to the normal in the water. Let’s refer to the air as medium 1 and the water as 2. Solving Snell’s law for $\theta_1$, we find

$$\theta_1 = \sin^{-1} \left( \frac{n_2}{n_1} \sin \theta_2 \right).$$

As mentioned above, air has an index of refraction very close to 1, and water’s is about 1.3, so we find $\theta_1 = 40^\circ$. The angle $\alpha$ is therefore $50^\circ$.

The index of refraction is related to the speed of light.

What neither Snell nor Newton knew was that there is a very simple interpretation of the index of refraction. This may come as a relief to the reader who is taken aback by the complex reasoning involving proportionality that led to its definition. Later experiments showed that the index of refraction of a medium was inversely proportional to the speed of light in that medium. Since $c$ is defined as the speed of light in vacuum, and $n = 1$ is defined as the index of refraction of vacuum, we have

$$n = \frac{c}{v}.$$

[$n =$ medium’s index of refraction, $v =$ speed of light in that medium, $c =$ speed of light in a vacuum]

Many textbooks start with this as the definition of the index of refraction, although that approach makes the quantity’s name somewhat of a mystery, and leaves students wondering why $c/v$ was used rather than $v/c$. It should also be noted that measuring angles of refraction is a far more practical method for determining $n$ than direct measurement of the speed of light in the substance of interest.

A mechanical model of Snell’s law

Why should refraction be related to the speed of light? The mechanical model shown in the figure may help to make this more plausible. Suppose medium 2 is thick, sticky mud, which slows down the car. The car’s right wheel hits the mud first, causing the right side of the car to slow down. This will cause the car to turn to the right until it moves far enough forward for the left wheel to cross into the mud. After that, the two sides of the car will once again be moving at the same speed, and the car will go straight.

Of course, light isn’t a car. Why should a beam of light have anything resembling a “left wheel” and “right wheel?” After all,
the mechanical model would predict that a motorcycle would go straight, and a motorcycle seems like a better approximation to a ray of light than a car. The whole thing is just a model, not a description of physical reality.

A derivation of Snell’s law

However intuitively appealing the mechanical model may be, light is a wave, and we should be using wave models to describe refraction. In fact Snell’s law can be derived quite simply from wave concepts. Figure j shows the refraction of a water wave. The water in the upper left part of the tank is shallower, so the speed of the waves is slower there, and their wavelengths is shorter. The reflected part of the wave is also very faintly visible.

In the close-up view on the right, the dashed lines are normals to the interface. The two marked angles on the right side are both equal to $\theta_1$, and the two on the left to $\theta_2$.

Trigonometry gives

$$\sin \theta_1 = \frac{\lambda_1}{h}$$

and

$$\sin \theta_2 = \frac{\lambda_2}{h}.$$  

Eliminating $h$ by dividing the equations, we find

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{\lambda_1}{\lambda_2}.$$  

The frequencies of the two waves must be equal or else they would get out of step, so by $v = f\lambda$ we know that their wavelengths are
proportional to their velocities. Combining \( \lambda \propto v \) with \( v \propto 1/n \) gives \( \lambda \propto 1/n \), so we find

\[
\frac{\sin \theta_1}{\sin \theta_2} = \frac{n_2}{n_1},
\]

which is one form of Snell’s law.

**Ocean waves near and far from shore**

Ocean waves are formed by winds, typically on the open sea, and the wavefronts are perpendicular to the direction of the wind that formed them. At the beach, however, you have undoubtedly observed that waves tend to come in with their wavefronts very nearly (but not exactly) parallel to the shoreline. This is because the speed of water waves in shallow water depends on depth: the shallower the water, the slower the wave. Although the change from the fast-wave region to the slow-wave region is gradual rather than abrupt, there is still refraction, and the wave motion is nearly perpendicular to the normal in the slow region.

**Color and refraction**

In general, the speed of light in a medium depends both on the medium and on the wavelength of the light. Another way of saying it is that a medium’s index of refraction varies with wavelength. This is why a prism can be used to split up a beam of white light into a rainbow. Each wavelength of light is refracted through a different angle.

**How much light is reflected, and how much is transmitted?**

In section 6.2 we developed an equation for the percentage of the wave energy that is transmitted and the percentage reflected at a boundary between media. This was only done in the case of waves in one dimension, however, and rather than discuss the full three-dimensional generalization it will be more useful to go into some qualitative observations about what happens. First, reflection happens only at the interface between two media, and two media with the same index of refraction act as if they were a single medium. Thus, at the interface between media with the same index of refraction, there is no reflection, and the ray keeps going straight. Continuing this line of thought, it is not surprising that we observe very little reflection at an interface between media with similar indices of refraction.

The next thing to note is that it is possible to have situations where no possible angle for the refracted ray can satisfy Snell’s law. Solving Snell’s law for \( \theta_2 \), we find

\[
\theta_2 = \sin^{-1} \left( \frac{n_1}{n_2} \sin \theta_1 \right),
\]

and if \( n_1 \) is greater than \( n_2 \), then there will be large values of \( \theta_1 \) for which the quantity \( (n_1/n_2) \sin \theta \) is greater than one, meaning
that your calculator will flash an error message at you when you try to take the inverse sine. What can happen physically in such a situation? The answer is that all the light is reflected, so there is no refracted ray. This phenomenon is known as total internal reflection, and is used in the fiber-optic cables that nowadays carry almost all long-distance telephone calls. The electrical signals from your phone travel to a switching center, where they are converted from electricity into light. From there, the light is sent across the country in a thin transparent fiber. The light is aimed straight into the end of the fiber, and as long as the fiber never goes through any turns that are too sharp, the light will always encounter the edge of the fiber at an angle sufficiently oblique to give total internal reflection. If the fiber-optic cable is thick enough, one can see an image at one end of whatever the other end is pointed at.

Alternatively, a bundle of cables can be used, since a single thick cable is too hard to bend. This technique for seeing around corners is useful for making surgery less traumatic. Instead of cutting a person wide open, a surgeon can make a small “keyhole” incision and insert a bundle of fiber-optic cable (known as an endoscope) into the body.

Since rays at sufficiently large angles with respect to the normal may be completely reflected, it is not surprising that the relative amount of reflection changes depending on the angle of incidence, and is greatest for large angles of incidence.

Discussion Questions

A What index of refraction should a fish have in order to be invisible to other fish?

B Does a surgeon using an endoscope need a source of light inside the body cavity? If so, how could this be done without inserting a light bulb through the incision?

C A denser sample of a gas has a higher index of refraction than a less dense sample (i.e., a sample under lower pressure), but why would it not make sense for the index of refraction of a gas to be proportional to density?

D The earth’s atmosphere gets thinner and thinner as you go higher in altitude. If a ray of light comes from a star that is below the zenith, what will happen to it as it comes into the earth’s atmosphere?

E Does total internal reflection occur when light in a denser medium encounters a less dense medium, or the other way around? Or can it occur in either case?

12.4.2 Lenses

Figures n/1 and n/2 show examples of lenses forming images. There is essentially nothing for you to learn about imaging with lenses that is truly new. You already know how to construct and use ray diagrams, and you know about real and virtual images. The
The concept of the focal length of a lens is the same as for a curved mirror. The equations for locating images and determining magnifications are of the same form. It’s really just a question of flexing your mental muscles on a few examples. The following self-checks and discussion questions will get you started. I’ve also made a video that demonstrates some applications and how to explain them with ray diagrams: https://youtu.be/gL8awy6PWLQ.

self-check F

(1) In figures n/1 and n/2, classify the images as real or virtual.

(2) Glass has an index of refraction that is greater than that of air. Consider the topmost ray in figure n/1. Explain why the ray makes a slight left turn upon entering the lens, and another left turn when it exits.

(3) If the flame in figure n/2 was moved closer to the lens, what would happen to the location of the image? Answer, p. 1066

Discussion Questions

A In figures n/1 and n/2, the front and back surfaces are parallel to each other at the center of the lens. What will happen to a ray that enters near the center, but not necessarily along the axis of the lens? Draw a BIG ray diagram, and show a ray that comes from off axis.

In discussion questions B-F, don’t draw ultra-detailed ray diagrams as in A.

B Suppose you wanted to change the setup in figure n/1 so that the location of the actual flame in the figure would instead be occupied by an image of a flame. Where would you have to move the candle to achieve this? What about in n/2?

C There are three qualitatively different types of image formation that can occur with lenses, of which figures n/1 and n/2 exhaust only two. Figure out what the third possibility is. Which of the three possibilities can result in a magnification greater than one? Cf. problem 10, p. 831.
D Classify the examples shown in figure o according to the types of images delineated in discussion question C.

E In figures n/1 and n/2, the only rays drawn were those that happened to enter the lenses. Discuss this in relation to figure o.

F In the right-hand side of figure o, the image viewed through the lens is in focus, but the side of the rose that sticks out from behind the lens is not. Why?

12.4.3 The lensmaker’s equation

The focal length of a spherical mirror is simply \( r/2 \), but we cannot expect the focal length of a lens to be given by pure geometry, since it also depends on the index of refraction of the lens. Suppose we have a lens whose front and back surfaces are both spherical. (This is no great loss of generality, since any surface with a sufficiently shallow curvature can be approximated with a sphere.) Then if the lens is immersed in a medium with an index of refraction of 1, its focal length is given approximately by

\[
f = \left[ (n - 1) \left| \frac{1}{r_1} \pm \frac{1}{r_2} \right| \right]^{-1},
\]

where \( n \) is the index of refraction and \( r_1 \) and \( r_2 \) are the radii of curvature of the two surfaces of the lens. This is known as the lensmaker’s equation. In my opinion it is not particularly worthy
of memorization. The positive sign is used when both surfaces are curved outward or both are curved inward; otherwise a negative sign applies. The proof of this equation is left as an exercise to those readers who are sufficiently brave and motivated.

12.4.4 Dispersion

For most materials, we observe that the index of refraction depends slightly on wavelength, being highest at the blue end of the visible spectrum and lowest at the red. For example, white light disperses into a rainbow when it passes through a prism, q. Even when the waves involved aren’t light waves, and even when refraction isn’t of interest, the dependence of wave speed on wavelength is referred to as dispersion. Dispersion inside spherical raindrops is responsible for the creation of rainbows in the sky, and in an optical instrument such as the eye or a camera it is responsible for a type of aberration called chromatic aberration (subsection 12.3.3 and problem 28). As we’ll see in subsection 13.3.2, dispersion causes a wave that is not a pure sine wave to have its shape distorted as it travels, and also causes the speed at which energy and information are transported by the wave to be different from what one might expect from a naive calculation. The microscopic reasons for dispersion of light in matter are discussed in optional subsection 12.4.6.

12.4.5 * The principle of least time for refraction

We have seen previously how the rules governing straight-line motion of light and reflection of light can be derived from the principle of least time. What about refraction? In the figure, it is indeed plausible that the bending of the ray serves to minimize the time required to get from a point A to point B. If the ray followed the unbent path shown with a dashed line, it would have to travel a longer distance in the medium in which its speed is slower. By bending the correct amount, it can reduce the distance it has to cover in the slower medium without going too far out of its way. It is true that Snell’s law gives exactly the set of angles that minimizes the time required for light to get from one point to another. The proof of this fact is left as an exercise (problem 38, p. 836).
12.4.6 *Microscopic description of refraction*

Given that the speed of light is different in different media, we’ve seen two different explanations (on p. 806 and in subsection 12.4.5 above) of why refraction must occur. What we haven’t yet explained is why the speed of light does depend on the medium.

A good clue as to what’s going on comes from the figure s. The relatively minor variation of the index of refraction within the visible spectrum was misleading. At certain specific frequencies, $n$ exhibits wild swings in the positive and negative directions. After each such swing, we reach a new, lower plateau on the graph. These frequencies are resonances. For example, the visible part of the spectrum lies on the left-hand tail of a resonance at about $2 \times 10^{15}$ Hz, corresponding to the ultraviolet part of the spectrum. This resonance arises from the vibration of the electrons, which are bound to the nuclei as if by little springs. Because this resonance is narrow, the effect on visible-light frequencies is relatively small, but it is stronger at the blue end of the spectrum than at the red end. Near each resonance, not only does the index of refraction fluctuate wildly, but the glass becomes nearly opaque; this is because the vibration becomes very strong, causing energy to be dissipated as heat. The “staircase” effect is the same one visible in any resonance, e.g., figure k on p. 184: oscillators have a finite response for $f \ll f_0$, but the response approaches zero for $f \gg f_0$.

So far, we have a qualitative explanation of the frequency-variation of the loosely defined “strength” of the glass’s effect on a light wave, but we haven’t explained why the effect is observed as a change in speed, or why each resonance is an up-down swing rather than a single positive peak. To understand these effects in more detail, we need to consider the phase response of the oscillator. As shown in the bottom panel of figure j on p. 185, the phase response reverses itself as we pass through a resonance.

Suppose that a plane wave is normally incident on the left side of a thin sheet of glass, t/1, at $f \ll f_0$. The light wave observed on the
right side consists of a superposition of the incident wave consisting of \(E_0\) and \(B_0\) with a secondary wave \(E^*\) and \(B^*\) generated by the oscillating charges in the glass. Since the frequency is far below resonance, the response \(q x\) of a vibrating charge \(q\) is in phase with the driving force \(E_0\). The current is the derivative of this quantity, and therefore 90 degrees ahead of it in phase. The magnetic field generated by a sheet of current has been analyzed in subsection 11.2.1, and the result, shown in figure e on p. 692, is just what we would expect from the right-hand rule. We find, t/1, that the secondary wave is 90 degrees ahead of the incident one in phase. The incident wave still exists on the right side of the sheet, but it is superposed with the secondary one. Their addition is shown in t/2 using the complex number representation introduced in subsection 10.5.7. The superposition of the two fields lags behind the incident wave, which is the effect we would expect if the wave had traveled more slowly through the glass.

In the case \(f \gg f_0\), the same analysis applies except that the phase of the secondary wave is reversed. The transmitted wave is advanced rather than retarded in phase. This explains the dip observed in figure s after each spike.

All of this is in accord with our understanding of relativity, ch. 7, in which we saw that the universal speed \(c\) was to be understood fundamentally as a conversion factor between the units used to measure time and space — not as the speed of light. Since \(c\) isn’t defined as the speed of light, it’s of no fundamental importance whether light has a different speed in matter than it does in vacuum. In fact, the picture we’ve built up here is one in which all of our electromagnetic waves travel at \(c\); propagation at some other speed is only what appears to happen because of the superposition of the \((E_0, B_0)\) and \((E^*, B^*)\) waves, both of which move at \(c\).

But it is worrisome that at the frequencies where \(n < 1\), the speed of the wave is greater than \(c\). According to special relativity, information is never supposed to be transmitted at speeds greater than \(c\), since this would produce situations in which a signal could be received before it was transmitted! This difficulty is resolved in subsection 13.3.2, where we show that there are two different velocities that can be defined for a wave in a dispersive medium, the phase velocity and the group velocity. The group velocity is the velocity at which information is transmitted, and it is always less than \(c\).
12.5 Wave optics

Electron microscopes can make images of individual atoms, but why will a visible-light microscope never be able to? Stereo speakers create the illusion of music that comes from a band arranged in your living room, but why doesn’t the stereo illusion work with bass notes? Why are computer chip manufacturers investing billions of dollars in equipment to etch chips with x-rays instead of visible light?

The answers to all of these questions have to do with the subject of wave optics. So far this book has discussed the interaction of light waves with matter, and its practical applications to optical devices like mirrors, but we have used the ray model of light almost exclusively. Hardly ever have we explicitly made use of the fact that light is an electromagnetic wave. We were able to get away with the simple ray model because the chunks of matter we were discussing, such as lenses and mirrors, were thousands of times larger than a wavelength of light. We now turn to phenomena and devices that can only be understood using the wave model of light.

12.5.1 Diffraction

Figure a shows a typical problem in wave optics, enacted with water waves. It may seem surprising that we don’t get a simple pattern like figure b, but the pattern would only be that simple if the wavelength was hundreds of times shorter than the distance between the gaps in the barrier and the widths of the gaps.

Wave optics is a broad subject, but this example will help us to pick out a reasonable set of restrictions to make things more manageable:

(1) We restrict ourselves to cases in which a wave travels through a uniform medium, encounters a certain area in which the medium has different properties, and then emerges on the other side into a second uniform region.

(2) We assume that the incoming wave is a nice tidy sine-wave pattern with wavefronts that are lines (or, in three dimensions, planes).

(3) In figure a we can see that the wave pattern immediately beyond the barrier is rather complex, but farther on it sorts itself out into a set of wedges separated by gaps in which the water is still. We will restrict ourselves to studying the simpler wave patterns that occur farther away, so that the main question of interest is how intense the outgoing wave is at a given angle.

The kind of phenomenon described by restriction (1) is called diffraction. Diffraction can be defined as the behavior of a wave when it encounters an obstacle or a nonuniformity in its medium. In general, diffraction causes a wave to bend around obstacles and
make patterns of strong and weak waves radiating out beyond the obstacle. Understanding diffraction is the central problem of wave optics. If you understand diffraction, even the subset of diffraction problems that fall within restrictions (2) and (3), the rest of wave optics is icing on the cake.

Diffraction can be used to find the structure of an unknown diffracting object: even if the object is too small to study with ordinary imaging, it may be possible to work backward from the diffraction pattern to learn about the object. The structure of a crystal, for example, can be determined from its x-ray diffraction pattern.

Diffraction can also be a bad thing. In a telescope, for example, light waves are diffracted by all the parts of the instrument. This will cause the image of a star to appear fuzzy even when the focus has been adjusted correctly. By understanding diffraction, one can learn how a telescope must be designed in order to reduce this problem — essentially; it should have the biggest possible diameter.

There are two ways in which restriction (2) might commonly be violated. First, the light might be a mixture of wavelengths. If we simply want to observe a diffraction pattern or to use diffraction as a technique for studying the object doing the diffracting (e.g., if the object is too small to see with a microscope), then we can pass the light through a colored filter before diffracting it.

A second issue is that light from sources such as the sun or a lightbulb does not consist of a nice neat plane wave, except over very small regions of space. Different parts of the wave are out of step with each other, and the wave is referred to as incoherent. One way of dealing with this is shown in figure c. After filtering to select a certain wavelength of red light, we pass the light through a small pinhole. The region of the light that is intercepted by the pinhole is so small that one part of it is not out of step with another. Beyond the pinhole, light spreads out in a spherical wave; this is analogous to what happens when you speak into one end of a paper towel roll and the sound waves spread out in all directions from the other end. By the time the spherical wave gets to the double slit it has spread out and reduced its curvature, so that we can now think of it as a simple plane wave.

If this seems laborious, you may be relieved to know that modern technology gives us an easier way to produce a single-wavelength, coherent beam of light: the laser.

The parts of the final image on the screen in c are called diffraction fringes. The center of each fringe is a point of maximum brightness, and halfway between two fringes is a minimum.

**Discussion Question**

- **A** Why would x-rays rather than visible light be used to find the structure
of a crystal? Sound waves are used to make images of fetuses in the womb. What would influence the choice of wavelength?

### 12.5.2 Scaling of diffraction

This chapter has “optics” in its title, so it is nominally about light, but we started out with an example involving water waves. Water waves are certainly easier to visualize, but is this a legitimate comparison? In fact the analogy works quite well, despite the fact that a light wave has a wavelength about a million times shorter. This is because diffraction effects scale uniformly. That is, if we enlarge or reduce the whole diffraction situation by the same factor, including both the wavelengths and the sizes of the obstacles the wave encounters, the result is still a valid solution.

This is unusually simple behavior! In subsection 0.2.2 we saw many examples of more complex scaling, such as the impossibility of bacteria the size of dogs, or the need for an elephant to eliminate heat through its ears because of its small surface-to-volume ratio, whereas a tiny shrew’s life-style centers around conserving its body heat.

Of course water waves and light waves differ in many ways, not just in scale, but the general facts you will learn about diffraction are applicable to all waves. In some ways it might have been more appropriate to insert this chapter after section 6.2 on bounded waves, but many of the important applications are to light waves, and you would probably have found these much more difficult without any background in optics.

Another way of stating the simple scaling behavior of diffraction is that the diffraction angles we get depend only on the unitless ratio $\frac{\lambda}{d}$, where $\lambda$ is the wavelength of the wave and $d$ is some dimension of the diffracting objects, e.g., the center-to-center spacing between the slits in figure a. If, for instance, we scale up both $\lambda$ and $d$ by a factor of 37, the ratio $\frac{\lambda}{d}$ will be unchanged.

### 12.5.3 The correspondence principle

The only reason we don’t usually notice diffraction of light in everyday life is that we don’t normally deal with objects that are comparable in size to a wavelength of visible light, which is about a millionth of a meter. Does this mean that wave optics contradicts ray optics, or that wave optics sometimes gives wrong results? No. If you hold three fingers out in the sunlight and cast a shadow with them, either wave optics or ray optics can be used to predict the straightforward result: a shadow pattern with two bright lines where the light has gone through the gaps between your fingers. Wave optics is a more general theory than ray optics, so in any case where ray optics is valid, the two theories will agree. This is an example of a general idea enunciated by the physicist Niels Bohr, called the correspondence principle: when flaws in a physical theory
lead to the creation of a new and more general theory, the new theory must still agree with the old theory within its more restricted area of applicability. After all, a theory is only created as a way of describing experimental observations. If the original theory had not worked in any cases at all, it would never have become accepted.

In the case of optics, the correspondence principle tells us that when \( \lambda/d \) is small, both the ray and the wave model of light must give approximately the same result. Suppose you spread your fingers and cast a shadow with them using a coherent light source. The quantity \( \lambda/d \) is about \( 10^{-4} \), so the two models will agree very closely. (To be specific, the shadows of your fingers will be outlined by a series of light and dark fringes, but the angle subtended by a fringe will be on the order of \( 10^{-4} \) radians, so they will be too tiny to be visible.

**self-check G**

What kind of wavelength would an electromagnetic wave have to have in order to diffract dramatically around your body? Does this contradict the correspondence principle?

Answer, p. 1066

### 12.5.4 Huygens’ principle

Returning to the example of double-slit diffraction, it is as though all the sets of ripples have been blocked except for two. It is a rather surprising mathematical fact, however, that Huygens’ principle gives the right result in the case of an unobstructed linear wave. A theoretically infinite number of circular wave patterns somehow conspire to add together and produce the simple linear wave motion with which we are familiar.

Since Huygens’ principle is equivalent to the principle of superposition, and superposition is a property of waves, what Huygens had created was essentially the first wave theory of light. However, he imagined light as a series of pulses, like hand claps, rather than as a sinusoidal wave.

The history is interesting. Isaac Newton loved the atomic theory of matter so much that he searched enthusiastically for evidence that
light was also made of tiny particles. The paths of his light particles would correspond to rays in our description; the only significant difference between a ray model and a particle model of light would occur if one could isolate individual particles and show that light had a “graininess” to it. Newton never did this, so although he thought of his model as a particle model, it is more accurate to say he was one of the builders of the ray model.

Almost all that was known about reflection and refraction of light could be interpreted equally well in terms of a particle model or a wave model, but Newton had one reason for strongly opposing Huygens’ wave theory. Newton knew that waves exhibited diffraction, but diffraction of light is difficult to observe, so Newton believed that light did not exhibit diffraction, and therefore must not be a wave. Although Newton’s criticisms were fair enough, the debate also took on the overtones of a nationalistic dispute between England and continental Europe, fueled by English resentment over Leibniz’s supposed plagiarism of Newton’s calculus. Newton wrote a book on optics, and his prestige and political prominence tended to discourage questioning of his model.

Thomas Young (1773-1829) was the person who finally, a hundred years later, did a careful search for wave interference effects with light and analyzed the results correctly. He observed double-slit diffraction of light as well as a variety of other diffraction effects, all of which showed that light exhibited wave interference effects, and that the wavelengths of visible light waves were extremely short. The crowning achievement was the demonstration by the experimentalist Heinrich Hertz and the theorist James Clerk Maxwell that light was an electromagnetic wave. Maxwell is said to have related his discovery to his wife one starry evening and told her that she was the only other person in the world who knew what starlight was.

12.5.5 Double-slit diffraction

Let’s now analyze double-slit diffraction, k, using Huygens’ principle. The most interesting question is how to compute the angles such as X and Z where the wave intensity is at a maximum, and the in-between angles like Y where it is minimized. Let’s measure all our angles with respect to the vertical center line of the figure, which was the original direction of propagation of the wave.

If we assume that the width of the slits is small (on the order of the wavelength of the wave or less), then we can imagine only a single set of Huygens ripples spreading out from each one, l. White lines represent peaks, black ones troughs. The only dimension of the diffracting slits that has any effect on the geometric pattern of the overlapping ripples is then the center-to-center distance, d, between the slits.
We know from our discussion of the scaling of diffraction that there must be some equation that relates an angle like $\theta_Z$ to the ratio $\lambda/d$,

$$\frac{\lambda}{d} \leftrightarrow \theta_Z.$$  

If the equation for $\theta_Z$ depended on some other expression such as $\lambda + d$ or $\lambda^2/d$, then it would change when we scaled $\lambda$ and $d$ by the same factor, which would violate what we know about the scaling of diffraction.

Along the central maximum line, $X$, we always have positive waves coinciding with positive ones and negative waves coinciding with negative ones. (I have arbitrarily chosen to take a snapshot of the pattern at a moment when the waves emerging from the slit are experiencing a positive peak.) The superposition of the two sets of ripples therefore results in a doubling of the wave amplitude along this line. There is constructive interference. This is easy to explain, because by symmetry, each wave has had to travel an equal number of wavelengths to get from its slit to the center line, $m$: Because both sets of ripples have ten wavelengths to cover in order to reach the point along direction $X$, they will be in step when they get there.

At the point along direction $Y$ shown in the same figure, one wave has traveled ten wavelengths, and is therefore at a positive extreme, but the other has traveled only nine and a half wavelengths, so it is at a negative extreme. There is perfect cancellation, so points along this line experience no wave motion.

But the distance traveled does not have to be equal in order to get constructive interference. At the point along direction $Z$, one wave has gone nine wavelengths and the other ten. They are both at a positive extreme.

**self-check H**

At a point half a wavelength below the point marked along direction $X$, carry out a similar analysis.  \(\triangleright\) Answer, p. 1066

To summarize, we will have perfect constructive interference at any point where the distance to one slit differs from the distance to the other slit by an integer number of wavelengths. Perfect destructive interference will occur when the number of wavelengths of path length difference equals an integer plus a half.

Now we are ready to find the equation that predicts the angles of the maxima and minima. The waves travel different distances to get to the same point in space, $n$. We need to find whether the waves are in phase (in step) or out of phase at this point in order to predict whether there will be constructive interference, destructive interference, or something in between.

One of our basic assumptions in this chapter is that we will only be dealing with the diffracted wave in regions very far away from the
object that diffracts it, so the triangle is long and skinny. Most real-world examples with diffraction of light, in fact, would have triangles with even skinner proportions than this one. The two long sides are therefore very nearly parallel, and we are justified in drawing the right triangle shown in figure o, labeling one leg of the right triangle as the difference in path length, \( L - L' \), and labeling the acute angle as \( \theta \). (In reality this angle is a tiny bit greater than the one labeled \( \theta \) in figure n.)

The difference in path length is related to \( d \) and \( \theta \) by the equation

\[
\frac{L - L'}{d} = \sin \theta.
\]

Constructive interference will result in a maximum at angles for which \( L - L' \) is an integer number of wavelengths,

\[
L - L' = m\lambda. \quad \text{[condition for a maximum; } m \text{ is an integer]}
\]

Here \( m \) equals 0 for the central maximum, -1 for the first maximum to its left, +2 for the second maximum on the right, etc. Putting all the ingredients together, we find \( m\lambda/d = \sin \theta \), or

\[
\frac{\lambda}{d} = \frac{\sin \theta}{m}. \quad \text{[condition for a maximum; } m \text{ is an integer]}
\]

Similarly, the condition for a minimum is

\[
\frac{\lambda}{d} = \frac{\sin \theta}{m}. \quad \text{[condition for a minimum; } m \text{ is an integer plus 1/2]}
\]

That is, the minima are about halfway between the maxima.

As expected based on scaling, this equation relates angles to the unitless ratio \( \lambda/d \). Alternatively, we could say that we have proven the scaling property in the special case of double-slit diffraction. It was inevitable that the result would have these scaling properties, since the whole proof was geometric, and would have been equally valid when enlarged or reduced on a photocopying machine!

Counterintuitively, this means that a diffracting object with smaller dimensions produces a bigger diffraction pattern, p.
**Double-slit diffraction of blue and red light** example 12
Blue light has a shorter wavelength than red. For a given double-slit spacing \(d\), the smaller value of \(\lambda/d\) for leads to smaller values of \(\sin \theta\), and therefore to a more closely spaced set of diffraction fringes, as shown in figure q.

**The correspondence principle** example 13
Let’s also consider how the equations for double-slit diffraction relate to the correspondence principle. When the ratio \(\lambda/d\) is very small, we should recover the case of simple ray optics. Now if \(\lambda/d\) is small, \(\sin \theta\) must be small as well, and the spacing between the diffraction fringes will be small as well. Although we have not proven it, the central fringe is always the brightest, and the fringes get dimmer and dimmer as we go farther from it. For small values of \(\lambda/d\), the part of the diffraction pattern that is bright enough to be detectable covers only a small range of angles. This is exactly what we would expect from ray optics: the rays passing through the two slits would remain parallel, and would continue moving in the \(\theta = 0\) direction. (In fact there would be images of the two separate slits on the screen, but our analysis was all in terms of angles, so we should not expect it to address the issue of whether there is structure within a set of rays that are all traveling in the \(\theta = 0\) direction.)

**Spacing of the fringes at small angles** example 14
At small angles, we can use the approximation \(\sin \theta \approx \theta\), which is valid if \(\theta\) is measured in radians. The equation for double-slit diffraction becomes simply

\[
\frac{\lambda}{d} = \frac{\theta}{m},
\]

which can be solved for \(\theta\) to give

\[
\theta = \frac{m\lambda}{d}.
\]

The difference in angle between successive fringes is the change in \(\theta\) that results from changing \(m\) by plus or minus one,

\[
\Delta \theta = \frac{\lambda}{d}.
\]

For example, if we write \(\theta_7\) for the angle of the seventh bright fringe on one side of the central maximum and \(\theta_8\) for the neighboring one, we have

\[
\theta_8 - \theta_7 = \frac{8\lambda}{d} - \frac{7\lambda}{d} = \frac{\lambda}{d},
\]

and similarly for any other neighboring pair of fringes.

---

Section 12.5 Wave optics 821
Although the equation \( \frac{\lambda}{d} = \sin \frac{\theta}{m} \) is only valid for a double slit, it is still a guide to our thinking even if we are observing diffraction of light by a virus or a flea’s leg: it is always true that

1. large values of \( \frac{\lambda}{d} \) lead to a broad diffraction pattern, and
2. diffraction patterns are repetitive.

In many cases the equation looks just like \( \frac{\lambda}{d} = \sin \frac{\theta}{m} \) but with an extra numerical factor thrown in, and with \( d \) interpreted as some other dimension of the object, e.g., the diameter of a piece of wire.

### 12.5.6 Repetition

Suppose we replace a double slit with a triple slit, \( s \). We can think of this as a third repetition of the structures that were present in the double slit. Will this device be an improvement over the double slit for any practical reasons?

The answer is yes, as can be shown using figure u. For ease of visualization, I have violated our usual rule of only considering points very far from the diffracting object. The scale of the drawing is such that a wavelength is one cm. In u/1, all three waves travel an integer number of wavelengths to reach the same point, so there is a bright central spot, as we would expect from our experience with the double slit. In figure u/2, we show the path lengths to a new point. This point is farther from slit A by a quarter of a wavelength, and correspondingly closer to slit C. The distance from slit B has hardly changed at all. Because the paths lengths traveled from slits A and C differ by half a wavelength, there will be perfect destructive interference between these two waves. There is still some uncanceled wave intensity because of slit B, but the amplitude will be three times less than in figure u/1, resulting in a factor of 9 decrease in brightness. Thus, by moving off to the right a little, we have gone from the bright central maximum to a point that is quite dark.

Now let’s compare with what would have happened if slit C had been covered, creating a plain old double slit. The waves coming from slits A and B would have been out of phase by 0.23 wavelengths, but this would not have caused very severe interference. The point in figure u/2 would have been quite brightly lit up.

To summarize, we have found that adding a third slit narrows down the central fringe dramatically. The same is true for all the other fringes as well, and since the same amount of energy is con-
Single-slit diffraction of water waves.

Single-slit diffraction of red light. Note the double width of the central maximum.

A pretty good simulation of the single-slit pattern of figure v, made by using three motors to produce overlapping ripples from three neighboring points in the water.
things we’ve learned about diffraction. We know based on scaling arguments that the angular sizes of features in the diffraction pattern must be related to the wavelength and the width, \( a \), of the slit by some relationship of the form

\[
\frac{\lambda}{a} \leftrightarrow \theta.
\]

This is indeed true, and for instance the angle between the maximum of the central fringe and the maximum of the next fringe on one side equals \( 1.5\frac{\lambda}{a} \). Scaling arguments will never produce factors such as the 1.5, but they tell us that the answer must involve \( \lambda/a \), so all the familiar qualitative facts are true. For instance, shorter-wavelength light will produce a more closely spaced diffraction pattern.

An important scientific example of single-slit diffraction is in telescopes. Images of individual stars, as in figure y, are a good way to examine diffraction effects, because all stars except the sun are so far away that no telescope, even at the highest magnification, can image their disks or surface features. Thus any features of a star’s image must be due purely to optical effects such as diffraction. A prominent cross appears around the brightest star, and dimmer ones surround the dimmer stars. Something like this is seen in most telescope photos, and indicates that inside the tube of the telescope there were two perpendicular struts or supports. Light diffracted around these struts. You might think that diffraction could be eliminated entirely by getting rid of all obstructions in the tube, but the circles around the stars are diffraction effects arising from single-slit diffraction at the mouth of the telescope’s tube! (Actually we have not even talked about diffraction through a circular opening, but the idea is the same.) Since the angular sizes of the diffracted images depend on \( \lambda/a \), the only way to improve the resolution of the images is to increase the diameter, \( a \), of the tube. This is one of the main reasons (in addition to light-gathering power) why the best telescopes must be very large in diameter.

**self-check J**

What would this imply about radio telescopes as compared with visible-light telescopes?

▷ Answer, p. 1066

Double-slit diffraction is easier to understand conceptually than single-slit diffraction, but if you do a double-slit diffraction experiment in real life, you are likely to encounter a complicated pattern like figure aa/1, rather than the simpler one, 2, you were expecting. This is because the slits are fairly big compared to the wavelength of the light being used. We really have two different distances in our pair of slits: \( d \), the distance between the slits, and \( w \), the width of each slit. Remember that smaller distances on the object the light diffracts around correspond to larger features of the diffraction pattern. The pattern 1 thus has two spacings in it: a short spac-
ing corresponding to the large distance \( d \), and a long spacing that relates to the small dimension \( w \).

aa / 1. A diffraction pattern formed by a real double slit. The width of each slit is fairly big compared to the wavelength of the light. This is a real photo. 2. This idealized pattern is not likely to occur in real life. To get it, you would need each slit to be so narrow that its width was comparable to the wavelength of the light, but that’s not usually possible. This is not a real photo. 3. A real photo of a single-slit diffraction pattern caused by a slit whose width is the same as the widths of the slits used to make the top pattern.

Discussion Question
A Why is it optically impossible for bacteria to evolve eyes that use visible light to form images?

12.5.8 Coherence

Up until now, we have avoided too much detailed discussion of two facts that sometimes make interference and diffraction effects unobservable, and that historically made them more difficult to discover. First there is the fact that white light is a mixture of all the visible wavelengths. This is why, for example, the thin-film interference pattern of a soap bubble looks like a rainbow. To simplify things, we need a source of light that is monochromatic, i.e., contains only a single wavelength or a small range of wavelengths. We could do this either by filtering a white light source or by using a source of light that is intrinsically monochromatic, such as a laser or some gas discharge tubes.

But even with a monochromatic light source, we encounter a separate issue, which is that most light sources do not emit light waves that are perfect, infinitely long sine waves. Sunlight and candlelight, for example, can be thought of as being composed of separate little spurts of light, referred to as wave packets or wave trains. Each wave packet is emitted by a separate atom of the gas. It contains some number of wavelengths, and it has no fixed phase relationship to any other wave packet. The wave trains emitted by a laser are much longer, but still not infinitely long.
As an example of an experiment that can show these effects, figure ab/1 shows a thin-film interference pattern created by the air wedge between two pieces of very flat glass, where the top piece is placed at a very small angle relative to the bottom one, ab/2. The phase relationship between the two reflected waves is determined by the extra distance traveled by the ray that is reflected by the bottom plate (as well as the fact that one of the two reflections will be inverting).

If the angle is opened up too much, ab/3, we will no longer see fringes where the air layer is too thick. This is because the incident wave train has only a certain length, and the extra distance traveled is now so great that the two reflected wave trains no longer overlap in space. In general, if the incident wave trains are \( n \) wavelengths long, then we can see at most \( n \) bright and \( n \) dark fringes. The fact that about 18 fringes are visible in ab/1 shows that the light source used (let’s say a sodium gas discharge tube) made wave trains at least 18 wavelengths in length.

In real-world light sources, the wave packets may not be as neat and tidy as the ones in figure ab. They may not look like sine waves with clean cut-offs at the ends, and they may overlap one another. The result will look more like the examples in figure ac. Such a wave pattern has a property called its coherence length \( L \). On scales small compared to \( L \), the wave appears like a perfect sine wave. On scales large compared to \( L \), we lose all phase correlations. For example, the middle wave in figure ac has \( L \approx 5\lambda \). If we pick two points within this wave separated by a distance of \( \lambda \) in the left-right direction, they are likely to be very nearly in phase. But if the

As an example of an experiment that can show these effects, figure ab/1 shows a thin-film interference pattern created by the air wedge between two pieces of very flat glass, where the top piece is placed at a very small angle relative to the bottom one, ab/2. The phase relationship between the two reflected waves is determined by the extra distance traveled by the ray that is reflected by the bottom plate (as well as the fact that one of the two reflections will be inverting).

If the angle is opened up too much, ab/3, we will no longer see fringes where the air layer is too thick. This is because the incident wave train has only a certain length, and the extra distance traveled is now so great that the two reflected wave trains no longer overlap in space. In general, if the incident wave trains are \( n \) wavelengths long, then we can see at most \( n \) bright and \( n \) dark fringes. The fact that about 18 fringes are visible in ab/1 shows that the light source used (let’s say a sodium gas discharge tube) made wave trains at least 18 wavelengths in length.

In real-world light sources, the wave packets may not be as neat and tidy as the ones in figure ab. They may not look like sine waves with clean cut-offs at the ends, and they may overlap one another. The result will look more like the examples in figure ac. Such a wave pattern has a property called its coherence length \( L \). On scales small compared to \( L \), the wave appears like a perfect sine wave. On scales large compared to \( L \), we lose all phase correlations. For example, the middle wave in figure ac has \( L \approx 5\lambda \). If we pick two points within this wave separated by a distance of \( \lambda \) in the left-right direction, they are likely to be very nearly in phase. But if the
Light could take many different paths from A to B. If the light comes from a flame or a gas discharge tube, then this lack of a phase relationship would be because the parts of the wave at these large separations from one another probably originated from different atoms in the source.

**12.5.9 The principle of least time**

In subsection 12.1.5 and 12.4.5, we saw how in the ray model of light, both refraction and reflection can be described in an elegant and beautiful way by a single principle, the principle of least time. We can now justify the principle of least time based on the wave model of light. Consider an example involving reflection, ad. Starting at point A, Huygens’ principle for waves tells us that we can think of the wave as spreading out in all directions. Suppose we imagine all the possible ways that a ray could travel from A to B. We show this by drawing 25 possible paths, of which the central one is the shortest. Since the principle of least time connects the wave model to the ray model, we should expect to get the most accurate results when the wavelength is much shorter than the distances involved — for the sake of this numerical example, let’s say that a wavelength is 1/10 of the shortest reflected path from A to B. The table, 2, shows the distances traveled by the 25 rays.

Note how similar are the distances traveled by the group of 7 rays, indicated with a bracket, that come closest to obeying the principle of least time. If we think of each one as a wave, then all 7 are again nearly in phase at point B. However, the rays that are farther from satisfying the principle of least time show more rapidly changing distances; on reuniting at point B, their phases are a random jumble, and they will very nearly cancel each other out. Thus, almost none of the wave energy delivered to point B goes by these longer paths. Physically we find, for instance, that a wave pulse emitted at A is observed at B after a time interval corresponding very nearly to the shortest possible path, and the pulse is not very “smeared out” when it gets there. The shorter the wavelength compared to the dimensions of the figure, the more accurate these approximate statements become.

Instead of drawing a finite number of rays, such as 25, what happens if we think of the angle, θ, of emission of the ray as a continuously varying variable? Minimizing the distance $L$ requires

$$\frac{dL}{d\theta} = 0.$$  

Because $L$ is changing slowly in the vicinity of the angle that satisfies the principle of least time, all the rays that come out close to this angle have very nearly the same $L$, and remain very nearly in phase when they reach B. This is the basic reason why the discrete
table, ad/2, turned out to have a group of rays that all traveled nearly the same distance.

As discussed in subsection 12.1.5, the principle of least time is really a principle of least or greatest time. This makes perfect sense, since $dL/d\theta = 0$ can in general describe either a minimum or a maximum.

The principle of least time is very general. It does not apply just to refraction and reflection — it can even be used to prove that light rays travel in a straight line through empty space, without taking detours! This general approach to wave motion was used by Richard Feynman, one of the pioneers who in the 1950's reconciled quantum mechanics with relativity. A very readable explanation is given in a book Feynman wrote for laypeople, QED: The Strange Theory of Light and Matter.
Problems

The symbols $\sqrt{\cdot}$, $\equiv$, etc. are explained on page 846.

1. Draw a ray diagram showing why a small light source (a candle, say) produces sharper shadows than a large one (e.g., a long fluorescent bulb).

2. A Global Positioning System (GPS) receiver is a device that lets you figure out where you are by receiving timed radio signals from satellites. It works by measuring the travel time for the signals, which is related to the distance between you and the satellite. By finding the ranges to several different satellites in this way, it can pin down your location in three dimensions to within a few meters. How accurate does the measurement of the time delay have to be to determine your position to this accuracy?

3. Estimate the frequency of an electromagnetic wave whose wavelength is similar in size to an atom (about a nm). Referring back to figure o on p. 732, in what part of the electromagnetic spectrum would such a wave lie (infrared, gamma-rays, . . .)?

4. The Stealth Bomber is designed with flat, smooth surfaces. Why would this make it difficult to detect using radar?

5. The natives of planet Wumpus play pool using light rays on an eleven-sided table with mirrors for bumpers, shown in the figure on the next page. Trace this shot accurately with a ruler to reveal the hidden message. To get good enough accuracy, you’ll need to photocopy the page (or download the book and print the page) and construct each reflection using a protractor.

Problem 5.
6. The figure on the next page shows a curved (parabolic) mirror, with three parallel light rays coming toward it. One ray is approaching along the mirror’s center line. (a) Continue the light rays until they are about to undergo their second reflection. To get good enough accuracy, you’ll need to photocopy the page (or download the book and print the page) and draw in the normal at each place where a ray is reflected. What do you notice? (b) Make up an example of a practical use for this device. (c) How could you use this mirror with a small lightbulb to produce a parallel beam of light rays going off to the right? Solution, p. 1050

7. A man is walking at 1.0 m/s directly towards a flat mirror. At what speed is his separation from his image decreasing?

8. If a mirror on a wall is only big enough for you to see yourself from your head down to your waist, can you see your entire body by backing up? Test this experimentally and come up with an explanation for your observations, including a ray diagram.

Note that when you do the experiment, it’s easy to confuse yourself if the mirror is even a tiny bit off of vertical. One way to check yourself is to artificially lower the top of the mirror by putting a piece of tape or a post-it note where it blocks your view of the top of your head. You can then check whether you are able to see more of yourself both above and below by backing up.
9 In section 12.2 we’ve only done examples of mirrors with hollowed-out shapes (called concave mirrors). Now draw a ray diagram for a curved mirror that has a bulging outward shape (called a convex mirror). (a) How does the image’s distance from the mirror compare with the actual object’s distance from the mirror? From this comparison, determine whether the magnification is greater than or less than one. (b) Is the image real, or virtual? Could this mirror ever make the other type of image?

10 As discussed in question 9, there are two types of curved mirrors, concave and convex. Make a list of all the possible combinations of types of images (virtual or real) with types of mirrors (concave and convex). (Not all of the four combinations are physically possible.) Now for each one, use ray diagrams to determine whether increasing the distance of the object from the mirror leads to an increase or a decrease in the distance of the image from the mirror.

Draw BIG ray diagrams! Each diagram should use up about half a page of paper.

Some tips: To draw a ray diagram, you need two rays. For one of these, pick the ray that comes straight along the mirror’s axis, since its reflection is easy to draw. After you draw the two rays and locate the image for the original object position, pick a new object position that results in the same type of image, and start a new ray diagram, in a different color of pen, right on top of the first one. For the two new rays, pick the ones that just happen to hit the mirror at the same two places; this makes it much easier to get the result right without depending on extreme accuracy in your ability to draw the reflected rays.

11 If the user of an astronomical telescope moves her head closer to or farther away from the image she is looking at, does the magnification change? Does the angular magnification change? Explain. (For simplicity, assume that no eyepiece is being used.)

> Solution, p. 1050

12 In figure g/2 in on page 784, only the image of my forehead was located by drawing rays. Either photocopy the figure or download the book and print out the relevant page. On this copy of the figure, make a new set of rays coming from my chin, and locate its image. To make it easier to judge the angles accurately, draw rays from the chin that happen to hit the mirror at the same points where the two rays from the forehead were shown hitting it. By comparing the locations of the chin’s image and the forehead’s image, verify that the image is actually upside-down, as shown in the original figure.

13 The figure shows four points where rays cross. Of these, which are image points? Explain.
Here’s a game my kids like to play. I sit next to a sunny window, and the sun reflects from the glass on my watch, making a disk of light on the wall or floor, which they pretend to chase as I move it around. Is the spot a disk because that’s the shape of the sun, or because it’s the shape of my watch? In other words, would a square watch make a square spot, or do we just have a circular image of the circular sun, which will be circular no matter what?

Apply the equation $M = d_i/d_o$ to the case of a flat mirror.

Use the method described in the text to derive the equation relating object distance to image distance for the case of a virtual image produced by a converging mirror.

Find the focal length of the mirror in problem 6.

Rank the focal lengths of the mirrors in the figure, from shortest to longest. Explain.

(a) A converging mirror with a focal length of 20 cm is used to create an image, using an object at a distance of 10 cm. Is the image real, or is it virtual? (b) How about $f = 20$ cm and $d_o = 30$ cm? (c) What if it was a diverging mirror with $f = 20$ cm and $d_o = 10$ cm? (d) A diverging mirror with $f = 20$ cm and $d_o = 30$ cm?

Make up a numerical example of a virtual image formed by a converging mirror with a certain focal length, and determine the magnification. (You will need the result of problem 16.) Make sure to choose values of $d_o$ and $f$ that would actually produce a virtual image, not a real one. Now change the location of the object a little bit and redetermine the magnification, showing that it changes.

At my local department store, the cosmetics department sells hand mirrors advertised as giving a magnification of 5 times. How would you interpret this?

(a) Suppose a Newtonian telescope is being used for astronomical observing. Assume for simplicity that no eyepiece is used, and assume a value for the focal length of the mirror that would be reasonable for an amateur instrument that is to fit in a closet. Is the angular magnification different for objects at different distances? For example, you could consider two planets, one of which is twice as far as the other.

(a) Find a case where the magnification of a curved mirror is infinite. Is the angular magnification infinite from any realistic viewing position? (b) Explain why an arbitrarily large magnification can’t be achieved by having a sufficiently small value of $d_o$. 

Problem 18.
22. A concave surface that reflects sound waves can act just like a converging mirror. Suppose that, standing near such a surface, you are able to find a point where you can place your head so that your own whispers are focused back on your head, so that they sound loud to you. Given your distance to the surface, what is the surface’s focal length? 

23. The figure shows a device for constructing a realistic optical illusion. Two mirrors of equal focal length are put against each other with their silvered surfaces facing inward. A small object placed in the bottom of the cavity will have its image projected in the air above. The way it works is that the top mirror produces a virtual image, and the bottom mirror then creates a real image of the virtual image. (a) Show that if the image is to be positioned as shown, at the mouth of the cavity, then the focal length of the mirrors is related to the dimension \( h \) via the equation

\[
\frac{1}{f} = \frac{1}{h} + \frac{1}{h + \left(\frac{1}{h} - \frac{1}{f}\right)^{-1}}.
\]

(b) Restate the equation in terms of a single variable \( x = h/f \), and show that there are two solutions for \( x \). Which solution is physically consistent with the assumptions of the calculation? 

24. (a) A converging mirror is being used to create a virtual image. What is the range of possible magnifications? (b) Do the same for the other types of images that can be formed by curved mirrors (both converging and diverging). 

25. A diverging mirror of focal length \( f \) is fixed, and faces down. An object is dropped from the surface of the mirror, and falls away from it with acceleration \( g \). The goal of the problem is to find the maximum velocity of the image.

(a) Describe the motion of the image verbally, and explain why we should expect there to be a maximum velocity.

(b) Use arguments based on units to determine the form of the solution, up to an unknown unitless multiplicative constant.

(c) Complete the solution by determining the unitless constant.

26. Diamond has an index of refraction of 2.42, and part of the reason diamonds sparkle is that this encourages a light ray to undergo many total internal reflections before it emerges. (a) Calculate the critical angle at which total internal reflection occurs in diamond. (b) Explain the interpretation of your result: Is it measured from the normal, or from the surface? Is it a minimum angle for total internal reflection, or is it a maximum? How would the critical angle have been different for a substance such as glass or plastic, with a lower index of refraction?
27 Suppose a converging lens is constructed of a type of plastic whose index of refraction is less than that of water. How will the lens’s behavior be different if it is placed underwater?

▷ Solution, p. 1052

28 There are two main types of telescopes, refracting (using a lens) and reflecting (using a mirror as in figure i on p. 786). (Some telescopes use a mixture of the two types of elements: the light first encounters a large curved mirror, and then goes through an eyepiece that is a lens. To keep things simple, assume no eyepiece is used.) What implications would the color-dependence of focal length have for the relative merits of the two types of telescopes? Describe the case where an image is formed of a white star. You may find it helpful to draw a ray diagram.

▷ Solution, p. 1052

29 Based on Snell’s law, explain why rays of light passing through the edges of a converging lens are bent more than rays passing through parts closer to the center. It might seem like it should be the other way around, since the rays at the edge pass through less glass — shouldn’t they be affected less? In your answer:

- Include a ray diagram showing a huge, full-page, close-up view of the relevant part of the lens.

- Make use of the fact that the front and back surfaces aren’t always parallel; a lens in which the front and back surfaces are always parallel doesn’t focus light at all, so if your explanation doesn’t make use of this fact, your argument must be incorrect.

- Make sure your argument still works even if the rays don’t come in parallel to the axis or from a point on the axis.

▷ Solution, p. 1052

30 When you take pictures with a camera, the distance between the lens and the film or chip has to be adjusted, depending on the distance at which you want to focus. This is done by moving the lens. If you want to change your focus so that you can take a picture of something farther away, which way do you have to move the lens? Explain using ray diagrams. [Based on a problem by Eric Mazur.]
31 When swimming underwater, why is your vision made much clearer by wearing goggles with flat pieces of glass that trap air behind them? [Hint: You can simplify your reasoning by considering the special case where you are looking at an object far away, and along the optic axis of the eye.]

Solution, p. 1053

32 An object is more than one focal length from a converging lens. (a) Draw a ray diagram. (b) Using reasoning like that developed in section 12.3, determine the positive and negative signs in the equation $1/f = \pm 1/d_i \pm 1/d_o$. (c) The images of the rose in section 4.2 were made using a lens with a focal length of 23 cm. If the lens is placed 80 cm from the rose, locate the image.

Solution, p. 1053

33 The figure shows four lenses. Lens 1 has two spherical surfaces. Lens 2 is the same as lens 1 but turned around. Lens 3 is made by cutting through lens 1 and turning the bottom around. Lens 4 is made by cutting a central circle out of lens 1 and recessing it.

(a) A parallel beam of light enters lens 1 from the left, parallel to its axis. Reasoning based on Snell’s law, will the beam emerging from the lens be bent inward, or outward, or will it remain parallel to the axis? Explain your reasoning. As part of your answer, make a huge drawing of one small part of the lens, and apply Snell’s law at both interfaces. Recall that rays are bent more if they come to the interface at a larger angle with respect to the normal.

(b) What will happen with lenses 2, 3, and 4? Explain. Drawings are not necessary.

Solution, p. 1053

34 The drawing shows the anatomy of the human eye, at twice life size. Find the radius of curvature of the outer surface of the cornea by measurements on the figure, and then derive the focal length of the air-cornea interface, where almost all the focusing of light occurs. You will need to use physical reasoning to modify the lensmaker’s equation for the case where there is only a single refracting surface. Assume that the index of refraction of the cornea is essentially that of water.

Solution, p. 1053

35 An object is less than one focal length from a converging lens. (a) Draw a ray diagram. (b) Using reasoning like that developed in section 12.3, determine the positive and negative signs in the equation $1/f = \pm 1/d_i \pm 1/d_o$. (c) The images of the rose in section 4.2 were made using a lens with a focal length of 23 cm. If the lens is placed 10 cm from the rose, locate the image.

36 Nearsighted people wear glasses whose lenses are diverging. (a) Draw a ray diagram. For simplicity pretend that there is no eye behind the glasses. (b) Using reasoning like that developed in section 12.3, determine the positive and negative signs in the equation $1/f = \pm 1/d_i \pm 1/d_o$. (c) If the focal length of the lens is
37  (a) Light is being reflected diffusely from an object 1.000 m underwater. The light that comes up to the surface is refracted at the water-air interface. If the refracted rays all appear to come from the same point, then there will be a virtual image of the object in the water, above the object’s actual position, which will be visible to an observer above the water. Consider three rays, A, B and C, whose angles in the water with respect to the normal are $\theta_i = 0.000^\circ$, $1.000^\circ$ and $20.000^\circ$ respectively. Find the depth of the point at which the refracted parts of A and B appear to have intersected, and do the same for A and C. Show that the intersections are at nearly the same depth, but not quite. [Check: The difference in depth should be about 4 cm.]

(b) Since all the refracted rays do not quite appear to have come from the same point, this is technically not a virtual image. In practical terms, what effect would this have on what you see?

(c) In the case where the angles are all small, use algebra and trig to show that the refracted rays do appear to come from the same point, and find an equation for the depth of the virtual image. Do not put in any numerical values for the angles or for the indices of refraction — just keep them as symbols. You will need the approximation $\sin \theta \approx \tan \theta \approx \theta$, which is valid for small angles measured in radians.

38  Prove that the principle of least time leads to Snell’s law.

39  Two standard focal lengths for camera lenses are 50 mm (standard) and 28 mm (wide-angle). To see how the focal lengths relate to the angular size of the field of view, it is helpful to visualize things as represented in the figure. Instead of showing many rays coming from the same point on the same object, as we normally do, the figure shows two rays from two different objects. Although the lens will intercept infinitely many rays from each of these points, we have shown only the ones that pass through the center of the lens, so that they suffer no angular deflection. (Any angular deflection at the front surface of the lens is canceled by an opposite deflection at the back, since the front and back surfaces are parallel at the lens’s center.) What is special about these two rays is that they are aimed at the edges of one 35-mm-wide frame of film; that is, they show the limits of the field of view. Throughout this problem, we assume that $d_o$ is much greater than $d_i$.  

(a) Compute the angular width of the camera’s field of view when these two lenses are used.  

(b) Use small-angle approximations to find a simplified equation for the angular width of the field of view, $\theta$, in terms of the focal length, $f$, and the width of the film, $w$. Your equation should not have any trig functions in it. Compare the results of this approximation with your answers from part a.  

(c) Suppose that we are holding
constant the aperture (amount of surface area of the lens being used to collect light). When switching from a 50-mm lens to a 28-mm lens, how many times longer or shorter must the exposure be in order to make a properly developed picture, i.e., one that is not under- or overexposed? [Based on a problem by Arnold Arons.]

Solution, p. 1054

40 A nearsighted person is one whose eyes focus light too strongly, and who is therefore unable to relax the lens inside her eye sufficiently to form an image on her retina of an object that is too far away.

(a) Draw a ray diagram showing what happens when the person tries, with uncorrected vision, to focus at infinity.

(b) What type of lenses do her glasses have? Explain.

(c) Draw a ray diagram showing what happens when she wears glasses. Locate both the image formed by the glasses and the final image.

(d) Suppose she sometimes uses contact lenses instead of her glasses. Does the focal length of her contacts have to be less than, equal to, or greater than that of her glasses? Explain.

41 Fred’s eyes are able to focus on things as close as 5.0 cm. Fred holds a magnifying glass with a focal length of 3.0 cm at a height of 2.0 cm above a flatworm. (a) Locate the image, and find the magnification. (b) Without the magnifying glass, from what distance would Fred want to view the flatworm to see its details as well as possible? With the magnifying glass? (c) Compute the angular magnification.
It would be annoying if your eyeglasses produced a magnified or reduced image. Prove that when the eye is very close to a lens, and the lens produces a virtual image, the angular magnification is always approximately equal to 1 (regardless of whether the lens is diverging or converging).
44 The figure shows a diffraction pattern made by a double slit, along with an image of a meter stick to show the scale. Sketch the diffraction pattern from the figure on your paper. Now consider the four variables in the equation $\frac{\lambda}{d} = \sin \theta / m$. Which of these are the same for all five fringes, and which are different for each fringe? Which variable would you naturally use in order to label which fringe was which? Label the fringes on your sketch using the values of that variable.

45 Match gratings A-C with the diffraction patterns 1-3 that they produce. Explain.
46  The figure below shows two diffraction patterns. The top one was made with yellow light, and the bottom one with red. Could the slits used to make the two patterns have been the same?

47  The figure on p. 839 shows a diffraction pattern made by a double slit, along with an image of a meter stick to show the scale. The slits were 146 cm away from the screen on which the diffraction pattern was projected. The spacing of the slits was 0.050 mm. What was the wavelength of the light?

48  Why would blue or violet light be the best for microscopy?

49  The figure below shows two diffraction patterns, both made with the same wavelength of red light. (a) What type of slits made the patterns? Is it a single slit, double slits, or something else? Explain. (b) Compare the dimensions of the slits used to make the top and bottom pattern. Give a numerical ratio, and state which way the ratio is, i.e., which slit pattern was the larger one. Explain.

50  When white light passes through a diffraction grating, what is the smallest value of $m$ for which the visible spectrum of order $m$ overlaps the next one, of order $m + 1$? (The visible spectrum runs from about 400 nm to about 700 nm.)
Problem 51. This image of the Pleiades star cluster shows haloes around the stars due to the wave nature of light.

51 For star images such as the ones in figure y, estimate the angular width of the diffraction spot due to diffraction at the mouth of the telescope. Assume a telescope with a diameter of 10 meters (the largest currently in existence), and light with a wavelength in the middle of the visible range. Compare with the actual angular size of a star of diameter $10^9$ m seen from a distance of $10^{17}$ m. What does this tell you? 

Solution, p. 1054

52 The figure below shows three diffraction patterns. All were made under identical conditions, except that a different set of double slits was used for each one. The slits used to make the top pattern had a center-to-center separation $d = 0.50$ mm, and each slit was $w = 0.04$ mm wide. (a) Determine $d$ and $w$ for the slits used to make the pattern in the middle. (b) Do the same for the slits used to make the bottom pattern.

Solution, p. 1055
The beam of a laser passes through a diffraction grating, fans out, and illuminates a wall that is perpendicular to the original beam, lying at a distance of 2.0 m from the grating. The beam is produced by a helium-neon laser, and has a wavelength of 694.3 nm. The grating has 2000 lines per centimeter. (a) What is the distance on the wall between the central maximum and the maxima immediately to its right and left? (b) How much does your answer change when you use the small-angle approximations \( \theta \approx \sin \theta \approx \tan \theta \)?

Ultrasound, i.e., sound waves with frequencies too high to be audible, can be used for imaging fetuses in the womb or for breaking up kidney stones so that they can be eliminated by the body. Consider the latter application. Lenses can be built to focus sound waves, but because the wavelength of the sound is not all that small compared to the diameter of the lens, the sound will not be concentrated exactly at the geometrical focal point. Instead, a diffraction pattern will be created with an intense central spot surrounded by fainter rings. About 85% of the power is concentrated within the central spot. The angle of the first minimum (surrounding the central spot) is given by \( \sin \theta = \lambda/b \), where \( b \) is the diameter of the lens. This is similar to the corresponding equation for a single slit, but with a factor of 1.22 in front which arises from the circular shape of the aperture. Let the distance from the lens to the patient’s kidney stone be \( L = 20 \) cm. You will want \( f > 20 \) kHz, so that the sound is inaudible. Find values of \( b \) and \( f \) that would result in a usable design, where the central spot is small enough to lie within a kidney stone 1 cm in diameter.

Under what circumstances could one get a mathematically undefined result by solving the double-slit diffraction equation for \( \theta \)? Give a physical interpretation of what would actually be observed.

When ultrasound is used for medical imaging, the frequency may be as high as 5-20 MHz. Another medical application of ultrasound is for therapeutic heating of tissues inside the body; here, the frequency is typically 1-3 MHz. What fundamental physical reasons could you suggest for the use of higher frequencies for imaging?
57 Suppose we have a polygonal room whose walls are mirrors, and there a pointlike light source in the room. In most such examples, every point in the room ends up being illuminated by the light source after some finite number of reflections. A difficult mathematical question, first posed in the middle of the last century, is whether it is ever possible to have an example in which the whole room is not illuminated. (Rays are assumed to be absorbed if they strike exactly at a vertex of the polygon, or if they pass exactly through the plane of a mirror.)

The problem was finally solved in 1995 by G.W. Tokarsky, who found an example of a room that was not illuminable from a certain point. Figure 57 shows a slightly simpler example found two years later by D. Castro. If a light source is placed at either of the locations shown with dots, the other dot remains unilluminated, although every other point is lit up. It is not straightforward to prove rigorously that Castro’s solution has this property. However, the plausibility of the solution can be demonstrated as follows.

Suppose the light source is placed at the right-hand dot. Locate all the images formed by single reflections. Note that they form a regular pattern. Convince yourself that none of these images illuminates the left-hand dot. Because of the regular pattern, it becomes plausible that even if we form images of images, images of images of images, etc., none of them will ever illuminate the other dot.

There are various other versions of the problem, some of which remain unsolved. The book by Klee and Wagon gives a good introduction to the topic, although it predates Tokarsky and Castro’s work.

References:

58 A mechanical linkage is a device that changes one type of motion into another. The most familiar example occurs in a gasoline car’s engine, where a connecting rod changes the linear motion of the piston into circular motion of the crankshaft. The top panel of the figure shows a mechanical linkage invented by Peaucellier in 1864, and independently by Lipkin around the same time. It consists of six rods joined by hinges, the four short ones forming a rhombus. Point O is fixed in space, but the apparatus is free to rotate about O. Motion at P is transformed into a different motion at P′ (or vice versa).

Geometrically, the linkage is a mechanical implementation of
Problem 59.

the ancient problem of inversion in a circle. Considering the case in which the rhombus is folded flat, let the \( k \) be the distance from \( O \) to the point where \( P \) and \( P' \) coincide. Form the circle of radius \( k \) with its center at \( O \). As \( P \) and \( P' \) move in and out, points on the inside of the circle are always mapped to points on its outside, such that \( rr' = k^2 \). That is, the linkage is a type of analog computer that exactly solves the problem of finding the inverse of a number \( r \). Inversion in a circle has many remarkable geometrical properties, discussed in H.S.M. Coxeter, *Introduction to Geometry*, Wiley, 1961.

If a pen is inserted through a hole at \( P \), and \( P' \) is traced over a geometrical figure, the Peaucellier linkage can be used to draw a kind of image of the figure.

A related problem is the construction of pictures, like the one in the bottom panel of the figure, called anamorphs. The drawing of the column on the paper is highly distorted, but when the reflecting cylinder is placed in the correct spot on top of the page, an undistorted image is produced inside the cylinder. (Wide-format movie technologies such as Cinemascope are based on similar principles.)

Show that the Peaucellier linkage does *not* convert correctly between an image and its anamorph, and design a modified version of the linkage that does. Some knowledge of analytic geometry will be helpful.

59 The figure shows a lens with surfaces that are curved, but whose thickness is constant along any horizontal line. Use the lens-maker’s equation to prove that this “lens” is not really a lens at all.

▶ Solution, p. 1055

60 Under ordinary conditions, gases have indices of refraction only a little greater than that of vacuum, i.e., \( n = 1 + \epsilon \), where \( \epsilon \) is some small number. Suppose that a ray crosses a boundary between a region of vacuum and a region in which the index of refraction is \( 1 + \epsilon \). Find the maximum angle by which such a ray can ever be deflected, in the limit of small \( \epsilon \).

▶ Hint, p. 1037

61 A converging mirror has focal length \( f \). An object is located at a distance \((1 + \epsilon)f\) from the mirror, where \( \epsilon \) is small. Find the distance of the image from the mirror, simplifying your result as much as possible by using the assumption that \( \epsilon \) is small.

▶ Answer, p. 1069
The intensity of a beam of light is defined as the power per unit area incident on a perpendicular surface. Suppose that a beam of light in a medium with index of refraction \( n \) reaches the surface of the medium, with air on the outside. Its incident angle with respect to the normal is \( \theta \). (All angles are in radians.) Only a fraction \( f \) of the energy is transmitted, the rest being reflected. Because of this, we might expect that the transmitted ray would always be less intense than the incident one. But because the transmitted ray is refracted, it becomes narrower, causing an additional change in intensity by a factor \( g > 1 \). The product of these factors \( I = fg \) can be greater than one. The purpose of this problem is to estimate the maximum amount of intensification.

We will use the small-angle approximation \( \theta \ll 1 \) freely, in order to make the math tractable. In our previous studies of waves, we have only studied the factor \( f \) in the one-dimensional case where \( \theta = 0 \). The generalization to \( \theta \neq 0 \) is rather complicated and depends on the polarization, but for unpolarized light, we can use Schlick’s approximation,

\[
f(\theta) = f(0)(1 - \cos \theta)^5,
\]

where the value of \( f \) at \( \theta = 0 \) is found as in problem 17 on p. 395.

(a) Using small-angle approximations, obtain an expression for \( g \) of the form \( g \approx 1 + P \theta^2 \), and find the constant \( P \). Answer, p. 1069

(b) Find an expression for \( I \) that includes the two leading-order terms in \( \theta \). We will call this expression \( I_2 \). Obtain a simple expression for the angle at which \( I_2 \) is maximized. As a check on your work, you should find that for \( n = 1.3, \theta = 63^\circ \). (Trial-and-error maximization of \( I \) gives 60°.)

(c) Find an expression for the maximum value of \( I_2 \). You should find that for \( n = 1.3 \), the maximum intensification is 31%.

In an experiment to measure the unknown index of refraction \( n \) of a liquid, you send a laser beam from air into a tank filled with the liquid. Let \( \phi \) be the angle of the beam relative to the normal while in the air, and let \( \theta \) be the angle in the liquid. You can set \( \phi \) to any value you like by aiming the laser from an appropriate direction, and you measure \( \theta \) as a result. We wish to plan such an experiment so as to minimize the error \( dn \) in the result of the experiment, for a fixed error \( d\theta \) in the measurement of the angle in the liquid. We assume that there is no significant contribution to the error from uncertainty in the index of refraction of air (which is very close to 1) or from the angle \( \phi \). Find \( dn \) in terms of \( d\theta \), and determine the optimal conditions.

Solution, p. 1055
Zahra likes to play practical jokes on the friends she goes hiking with. One night, by a blazing camp fire, she stealthily uses a lens of focal length $f$ to gather light from the fire and make a hot spot on Becky's neck. (a) Using the method of section 12.3.2, p. 794, draw a ray diagram and set up the equation for the image location, inferring the correct plus and minus signs from the diagram. (b) Let $A$ be the distance from the lens to the campfire, and $B$ the distance from the lens to Becky’s neck. Consider the following nine possibilities:

\[
\begin{array}{ccc}
B \\
< f & = f & > f \\
A = f & & \\
> f & & \\
\end{array}
\]

By reasoning about your equation from part a, determine which of these are possible and which are not.  

Key to symbols:

- easy
- typical
- challenging
- difficult
- very difficult

✓ An answer check is available at www.lightandmatter.com.
Exercises

Exercise 12A: Exploring Images With a Curved Mirror

Equipment:

- concave mirrors with deep curvature
- concave mirrors with gentle curvature
- convex mirrors

1. Obtain a curved mirror from your instructor. If it is silvered on both sides, make sure you’re working with the concave side, which bends light rays inward. Look at your own face in the mirror. Now change the distance between your face and the mirror, and see what happens. Explore the full range of possible distances between your face and the mirror.

In these observations you’ve been changing two variables at once: the distance between the object (your face) and the mirror, and the distance from the mirror to your eye. In general, scientific experiments become easier to interpret if we practice isolation of variables, i.e., only change one variable while keeping all the others constant. In parts 2 and 3 you’ll form an image of an object that’s not your face, so that you can have independent control of the object distance and the point of view.

2. With the mirror held far away from you, observe the image of something behind you, over your shoulder. Now bring your eye closer and closer to the mirror. Can you see the image with your eye very close to the mirror? See if you can explain your observation by drawing a ray diagram.
3. Now imagine the following new situation, but \textit{don’t actually do it yet}. Suppose you lay the mirror face-up on a piece of tissue paper, put your finger a few cm above the mirror, and look at the image of your finger. As in part 2, you can bring your eye closer and closer to the mirror. Will you be able to see the image with your eye very close to the mirror? Draw a ray diagram to help you predict what you will observe.

Prediction:____________________

Now test your prediction. If your prediction was incorrect, see if you can figure out what went wrong, or ask your instructor for help.

4. For parts 4 and 5, it’s more convenient to use concave mirrors that are more gently curved; obtain one from your instructor. Lay the mirror on the tissue paper, and use it to create an image of the overhead lights on a piece of paper above it and a little off to the side. What do you have to do in order to make the image clear? Can you explain this observation using a ray diagram?
5. Now imagine the following experiment, but don’t do it yet. What will happen to the image on the paper if you cover half of the mirror with your hand?

Prediction:

Test your prediction. If your prediction was incorrect, can you explain what happened?

6. Now imagine forming an image with a convex mirror (one that bulges outward), and that therefore bends light rays away from the central axis (i.e., is diverging). Draw a typical ray diagram.

Is the image real, or virtual? Will there be more than one type of image?

Prediction:

Test your prediction.
**Exercise 12B: Object and Image Distances**

Equipment:
- optical benches
- converging mirrors
- illuminated objects

1. Set up the optical bench with the mirror at zero on the centimeter scale. Set up the illuminated object on the bench as well.

2. Each group will locate the image for their own value of the object distance, by finding where a piece of paper has to be placed in order to see the image on it. (The instructor will do one point as well.) Note that you will have to tilt the mirror a little so that the paper on which you project the image doesn’t block the light from the illuminated object.

Is the image real or virtual? How do you know? Is it inverted, or uninverted?

Draw a ray diagram.

3. Measure the image distance and write your result in the table on the board. Do the same for the magnification.

4. What do you notice about the trend of the data on the board? Draw a second ray diagram with a different object distance, and show why this makes sense. Some tips for doing this correctly: (1) For simplicity, use the point on the object that is on the mirror’s axis. (2) You need to trace two rays to locate the image. To save work, don’t just do two rays at random angles. You can either use the on-axis ray as one ray, or do two rays that come off at the same angle, one above and one below the axis. (3) Where each ray hits the mirror, draw the normal line, and make sure the ray is at equal angles on both sides of the normal.

5. We will find the mirror’s focal length from the instructor’s data-point. Then, using this focal length, calculate a theoretical prediction of the image distance, and write it on the board next to the experimentally determined image distance.