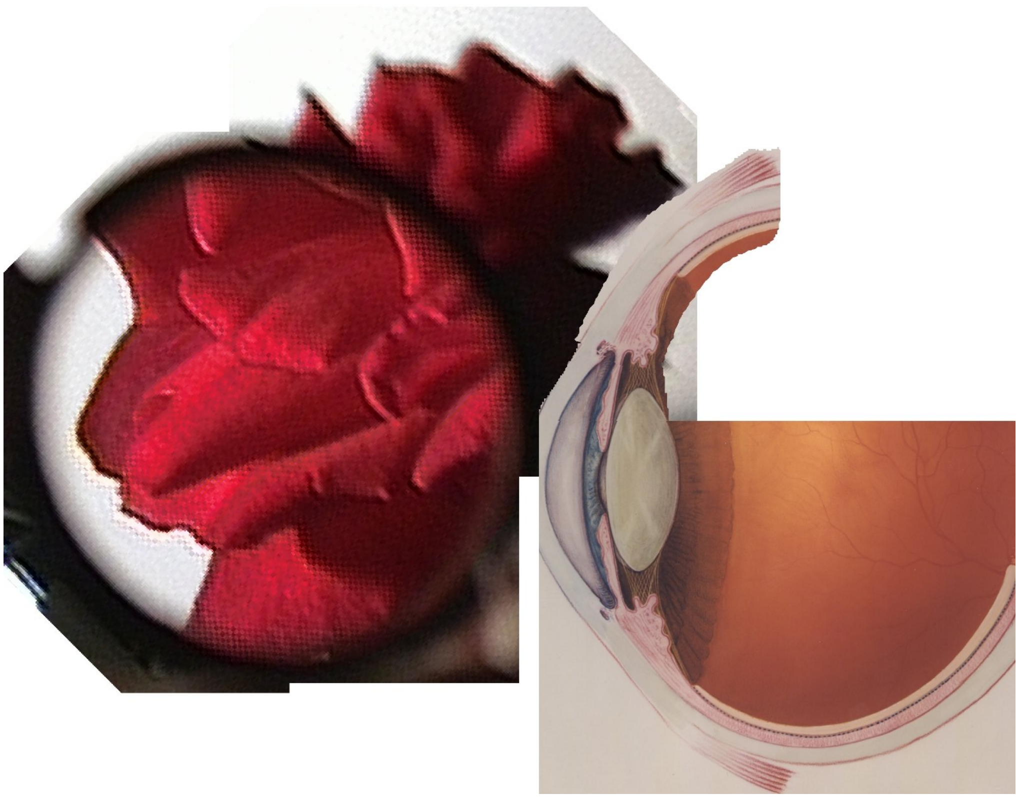
Optics

Benjamin Crowell

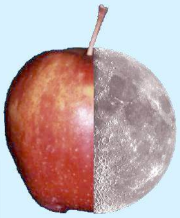


Book 5 in the Light and Matter series of free introductory physics textbooks
www.lightandmatter.com

Optics

The **Light and Matter** series of introductory physics textbooks:

1 Newtonian Physics
2 Conservation Laws
3 Vibrations and Waves
4 Electricity and Magnetism
5 Optics
6 The Modern Revolution in Physics



Optics

Benjamin Crowell

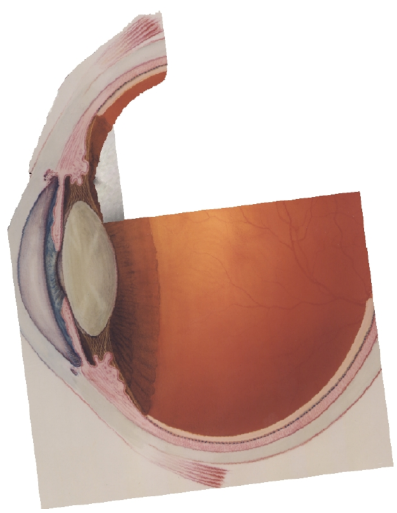
www.lightandmatter.com

Light and Matter
Fullerton, California
www.lightandmatter.com

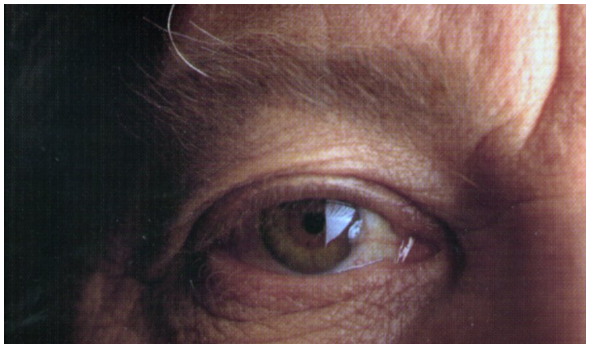
edition 2.2
rev. 6th January 2007

ISBN 0-9704670-5-2

Brief Contents



Contents



5 Wave Optics

Chapter 1
The Ray Model of Light

Ads for one Macintosh computer bragged that it could do an arithmetic calculation in less time than it took for the light to get from the screen to your eye. We find this impressive because of the contrast between the speed of light and the speeds at which we interact with physical objects in our environment. Perhaps it shouldn't surprise us, then, that Newton succeeded so well in explaining the motion of objects, but was far less successful with the study of light.

These books are billed as the Light and Matter series, but only now, in the fifth of the six volumes, are we ready to focus on light. If you are reading the series in order, then you know that the climax of our study of electricity and magnetism was discovery that light is an electromagnetic wave. Knowing this, however, is not the same as knowing everything about eyes and telescopes. In fact, the full description of light as a wave can be rather cumbersome. We will instead spend most of this book making use of a simpler model of light, the ray model, which does a fine job in most practical situations. Not only that, but we will even backtrack a little and start with a discussion of basic ideas about light and vision that predated the discovery of electromagnetic waves.

1.1 The Nature of Light

The cause and effect relationship in vision

Despite its title, this chapter is far from your first look at light. That familiarity might seem like an advantage, but most people have never thought carefully about light and vision. Even smart people who have thought hard about vision have come up with incorrect ideas. The ancient Greeks, Arabs and Chinese had theories of light and vision, all of which were mostly wrong, and all of which were accepted for thousands of years.

One thing the ancients did get right is that there is a distinction between objects that emit light and objects that don't. When you see a leaf in the forest, it's because three different objects are doing their jobs: the leaf, the eye, and the sun. But luminous objects like the sun, a flame, or the filament of a light bulb can be seen by the eye without the presence of a third object. Emission of light is often, but not always, associated with heat. In modern times, we are familiar with a variety of objects that glow without being heated, including fluorescent lights and glow-in-the-dark toys.

How do we see luminous objects? The Greek philosophers Pythagoras (b. ca. 560 BC) and Empedocles of Acragas (b. ca. 492 BC), who unfortunately were very influential, claimed that when you looked at a candle flame, the flame and your eye were both sending out some kind of mysterious stuff, and when your eye's stuff collided with the candle's stuff, the candle would become evident to your sense of sight.

Bizarre as the Greek "collision of stuff theory" might seem, it had a couple of good features. It explained why both the candle and your eye had to be present for your sense of sight to function. The theory could also easily be expanded to explain how we see nonluminous objects. If a leaf, for instance, happened to be present at the site of the collision between your eye's stuff and the candle's stuff, then the leaf would be stimulated to express its green nature, allowing you to perceive it as green.

Modern people might feel uneasy about this theory, since it suggests that greenness exists only for our seeing convenience, implying a human precedence over natural phenomena. Nowadays, people would expect the cause and effect relationship in vision to be the other way around, with the leaf doing something to our eye rather than our eye doing something to the leaf. But how can you tell? The most common way of distinguishing cause from effect is to determine which happened first, but the process of seeing seems to occur too quickly to determine the order in which things happened.

Certainly there is no obvious time lag between the moment when you move your head and the moment when your reflection in the mirror moves.

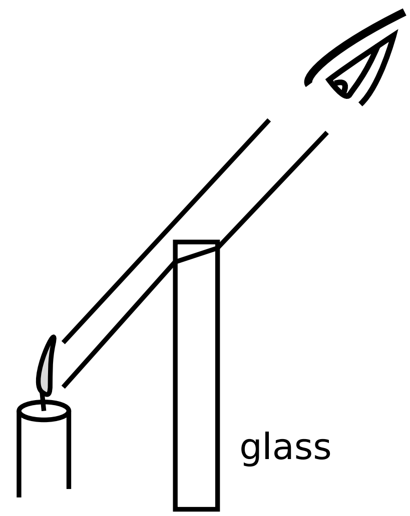
Today, photography provides the simplest experimental evidence that nothing has to be emitted from your eye and hit the leaf in order to make it "greenify." A camera can take a picture of a leaf even if there are no eyes anywhere nearby. Since the leaf appears green regardless of whether it is being sensed by a camera, your eye, or an insect's eye, it seems to make more sense to say that the leaf's greenness is the cause, and something happening in the camera or eye is the effect.

Light is a thing, and it travels from one point to another.

Another issue that few people have considered is whether a candle's flame simply affects your eye directly, or whether it sends out light which then gets into your eye. Again, the rapidity of the effect makes it difficult to tell what's happening. If someone throws a rock at you, you can see the rock on its way to your body, and you can tell that the person affected you by sending a material substance your way, rather than just harming you directly with an arm motion, which would be known as "action at a distance." It is not easy to do a similar observation to see whether there is some "stuff" that travels from the candle to your eye, or whether it is a case of action at a distance.

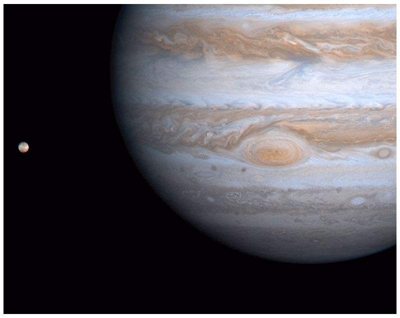
Newtonian physics includes both action at a distance (e.g. the earth's gravitational force on a falling object) and contact forces such as the normal force, which only allow distant objects to exert forces on each other by shooting some substance across the space between them (e.g., a garden hose spraying out water that exerts a force on a bush).

One piece of evidence that the candle sends out stuff that travels to your eye is that as in figure a, intervening transparent substances can make the candle appear to be in the wrong location, suggesting that light is a thing that can be bumped off course. Many people would dismiss this kind of observation as an optical illusion, however. (Some optical illusions are purely neurological or psychological effects, although some others, including this one, turn out to be caused by the behavior of light itself.)

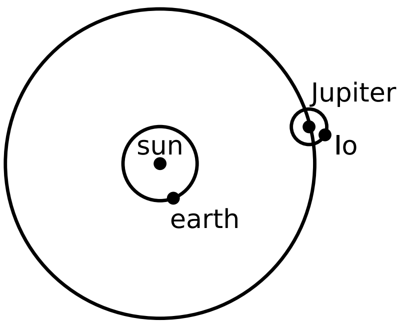


a / Light from a candle is bumped off course by a piece of glass. Inserting the glass causes the apparent location of the candle to shift. The same effect can be produced by taking off your eyeglasses and looking at which you see near the edge of the lens, but a flat piece of glass works just as well as a lens for this purpose.

A more convincing way to decide in which category light belongs is to find out if it takes time to get from the candle to your eye; in Newtonian physics, action at a distance is supposed to be instantaneous. The fact that we speak casually today of "the speed of light" implies that at some point in history, somebody succeeded in showing that light did not travel infinitely fast. Galileo tried, and failed, to detect a finite speed for light, by arranging with a person in a distant tower to signal back and forth with lanterns. Galileo uncovered his lantern, and when the other person saw the light, he uncovered his lantern. Galileo was unable to measure any time lag that was significant compared to the limitations of human reflexes.



b / An image of Jupiter and its moon Io (left) from the Cassini probe.



c / The earth is moving toward Jupiter and Io. Since the distance is shrinking, it is taking less and less time for the light to get to us from Io, and Io appears to circle Jupiter more quickly than normal. Six months later, the earth will be on the opposite side of the sun, and receding from Jupiter and Io, so Io will appear to revolve around Jupiter more slowly.

The first person to prove that light's speed was finite, and to determine it numerically, was Ole Roemer, in a series of measurements around the year 1675. Roemer observed Io, one of Jupiter's moons, over a period of several years. Since Io presumably took the same amount of time to complete each orbit of Jupiter, it could be thought of as a very distant, very accurate clock. A practical and accurate pendulum clock had recently been invented, so Roemer could check whether the ratio of the two clocks' cycles, about 42.5 hours to 1 orbit, stayed exactly constant or changed a little. If the process of seeing the distant moon was instantaneous, there would be no reason for the two to get out of step. Even if the speed of light was finite, you might expect that the result would be only to offset one cycle relative to the other. The earth does not, however, stay at a constant distance from Jupiter and its moons. Since the distance is changing gradually due to the two planets' orbital motions, a finite speed of light would make the "Io clock" appear to run faster as the planets drew near each other, and more slowly as their separation increased. Roemer did find a variation in the apparent speed of Io's orbits, which caused Io's eclipses by Jupiter (the moments when Io passed in front of or behind Jupiter) to occur about 7 minutes early when the earth was closest to Jupiter, and 7 minutes late when it was farthest. Based on these measurements, Roemer estimated the speed of light to be approximately 2×10^8 m/s, which is in the right ballpark compared to modern measurements of 3×10^8 m/s. (I'm not sure whether the fairly large experimental error was mainly due to imprecise knowledge of the radius of the earth's orbit or limitations in the reliability of pendulum clocks.)

Light can travel through a vacuum.

Many people are confused by the relationship between sound and light. Although we use different organs to sense them, there are some similarities. For instance, both light and sound are typically emitted in all directions by their sources. Musicians even use visual metaphors like "tone color," or "a bright timbre" to describe sound. One way to see that they are clearly different phenomena is to note their very different velocities. Sure, both are pretty fast compared to a flying arrow or a galloping horse, but as we have seen, the speed of light is so great as to appear instantaneous in most situations. The speed of sound, however, can easily be observed just by watching a group of schoolchildren a hundred feet away as they clap their hands to a song. There is an obvious delay between when you see their palms come together and when you hear the clap.

The fundamental distinction between sound and light is that sound is an oscillation in air pressure, so it requires air (or some other medium such as water) in which to travel. Today, we know that outer space is a vacuum, so the fact that we get light from the sun, moon and stars clearly shows that air is not necessary for the propagation of light.

Discussion Questions

A If you observe thunder and lightning, you can tell how far away the storm is. Do you need to know the speed of sound, of light, or of both?

B When phenomena like X-rays and cosmic rays were first discovered, suggest a way one could have tested whether they were forms of light.

C Why did Roemer only need to know the radius of the earth's orbit, not Jupiter's, in order to find the speed of light?

1.2 Interaction of Light with Matter

Absorption of light

The reason why the sun feels warm on your skin is that the sunlight is being absorbed, and the light energy is being transformed into heat energy. The same happens with artificial light, so the net result of leaving a light turned on is to heat the room. It doesn't matter whether the source of the light is hot, like the sun, a flame, or an incandescent light bulb, or cool, like a fluorescent bulb. (If your house has electric heat, then there is absolutely no point in fastidiously turning off lights in the winter; the lights will help to heat the house at the same dollar rate as the electric heater.)

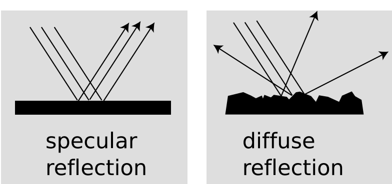
This process of heating by absorption is entirely different from heating by thermal conduction, as when an electric stove heats spaghetti sauce through a pan. Heat can only be conducted through matter, but there is vacuum between us and the sun, or between us and the filament of an incandescent bulb. Also, heat conduction can only transfer heat energy from a hotter object to a colder one, but a cool fluorescent bulb is perfectly capable of heating something that had already started out being warmer than the bulb itself.

How we see nonluminous objects

Not all the light energy that hits an object is transformed into heat. Some is reflected, and this leads us to the question of how we see nonluminous objects. If you ask the average person how we see a light bulb, the most likely answer is "The light bulb makes light, which hits our eyes." But if you ask how we see a book, they are likely to say "The bulb lights up the room, and that lets me see the book." All mention of light actually entering our eyes has mysteriously disappeared.

Most people would disagree if you told them that light was reflected from the book to the eye, because they think of reflection as something that mirrors do, not something that a book does. They associate reflection with the formation of a reflected image, which does not seem to appear in a piece of paper.

d / Two self-portraits of the author, one taken in a mirror and one with a piece of aluminum foil.



e / Specular and diffuse reflection.

Imagine that you are looking at your reflection in a nice smooth piece of aluminum foil, fresh off the roll. You perceive a face, not a piece of metal. Perhaps you also see the bright reflection of a lamp over your shoulder behind you. Now imagine that the foil is just a little bit less smooth. The different parts of the image are now a little bit out of alignment with each other. Your brain can still recognize a face and a lamp, but it's a little scrambled, like a Picasso painting. Now suppose you use a piece of aluminum foil that has been crumpled up and then flattened out again. The parts of the image are so scrambled that you cannot recognize an image. Instead, your brain tells you you're looking at a rough, silvery surface.

Mirror-like reflection at a specific angle is known as specular reflection, and random reflection in many directions is called diffuse reflection. Diffuse reflection is how we see nonluminous objects. Specular reflection only allows us to see images of objects other than the one doing the reflecting. In top part of figure d, imagine that the rays of light are coming from the sun. If you are looking down at the reflecting surface, there is no way for your eye-brain system to tell that the rays are not really coming from a sun down below you.

Figure f shows another example of how we can't avoid the conclusion that light bounces off of things other than mirrors. The lamp is one I have in my house. It has a bright bulb, housed in a completely opaque bowl-shaped metal shade. The only way light can get out of the lamp is by going up out of the top of the bowl. The fact that I can read a book in the position shown in the figure means that light must be bouncing off of the ceiling, then bouncing off of the book, then finally getting to my eye.

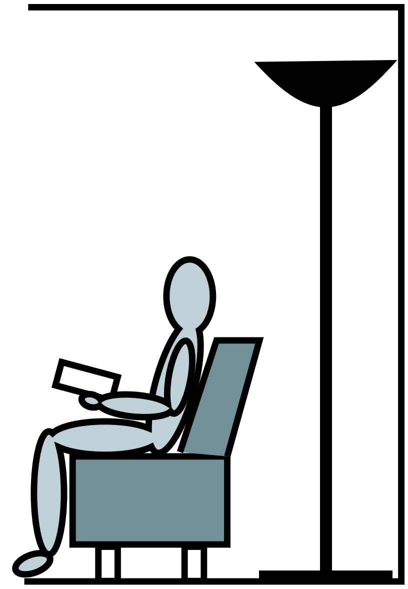
This is where the shortcomings of the Greek theory of vision become glaringly obvious. In the Greek theory, the light from the bulb and my mysterious "eye rays" are both supposed to go to the book, where they collide, allowing me to see the book. But we now have a total of four objects: lamp, eye, book, and ceiling. Where does the ceiling come in? Does it also send out its own mysterious "ceiling rays," contributing to a three-way collision at the book? That would just be too bizarre to believe!

The differences among white, black, and the various shades of gray in between is a matter of what percentage of the light they absorb and what percentage they reflect. That's why light-colored clothing is more comfortable in the summer, and light-colored upholstery in a car stays cooler that dark upholstery.

Numerical measurement of the brightness of light

We have already seen that the physiological sensation of loudness relates to the sound's intensity (power per unit area), but is not directly proportional to it. If sound A has an intensity of 1 $\mathrm{nW/m^2}$, sound B is 10 $\mathrm{nW/m^2}$, and sound C is 100 $\mathrm{nW/m^2}$, then the increase in loudness from C to B is perceived to be the same as the increase from A to B, not ten times greater. That is, the sensation of loudness is logarithmic.

The same is true for the brightness of light. Brightness is related to power per unit area, but the psychological relationship is a logarithmic one rather than a proportionality. For doing physics, it's the power per unit area that we're interested in. The relevant unit is $\mathrm{W/m^2}$. One way to determine the brightness of light is to measure the increase in temperature of a black object exposed to the light. The light energy is being converted to heat energy, and the amount of heat energy absorbed in a given amount of time can be related to the power absorbed, using the known heat capacity of the object. More practical devices for measuring light intensity, such as the light meters built into some cameras, are based on the conversion of light into electrical energy, but these meters have to be calibrated somehow against heat measurements.



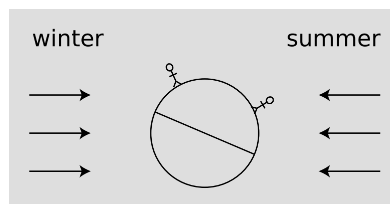
f / Light bounces off of the ceiling, then off of the book.

Discussion Questions

A The curtains in a room are drawn, but a small gap lets light through, illuminating a spot on the floor. It may or may not also be possible to see the beam of sunshine crossing the room, depending on the conditions. What's going on?

B Laser beams are made of light. In science fiction movies, laser beams are often shown as bright lines shooting out of a laser gun on a spaceship. Why is this scientifically incorrect?

C A documentary film-maker went to Harvard's 1987 graduation ceremony and asked the graduates, on camera, to explain the cause of the seasons. Only two out of 23 were able to give a correct explanation, but you now have all the information needed to figure it out for yourself, assuming you didn't already know. The figure shows the earth in its winter and summer positions relative to the sun. Hint: Consider the units used to measure the brightness of light, and recall that the sun is lower in the sky in winter, so its rays are coming in at a shallower angle.

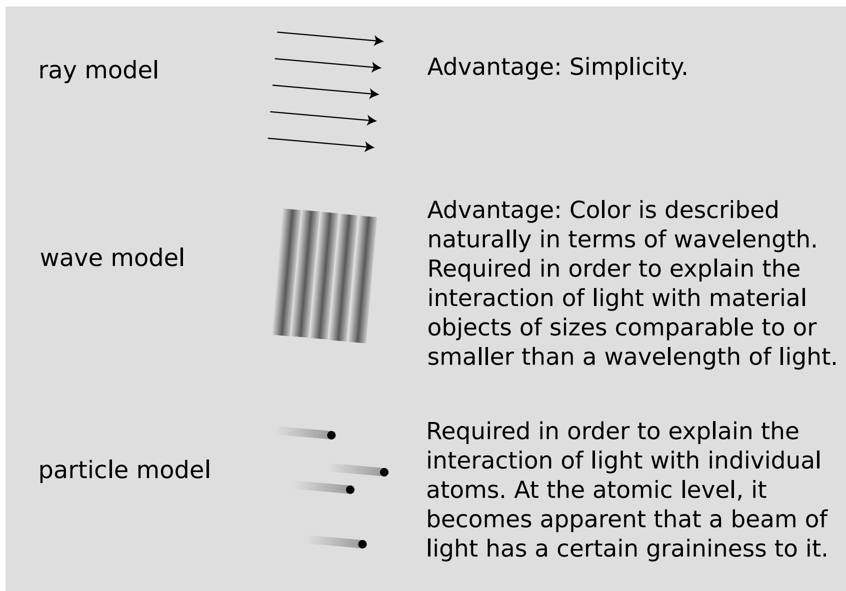


g / Discussion question C.

1.3 The Ray Model of Light

Models of light

Note how I've been casually diagramming the motion of light with pictures showing light rays as lines on the page. More formally, this is known as the ray model of light. The ray model of light seems natural once we convince ourselves that light travels through space, and observe phenomena like sunbeams coming through holes in clouds. Having already been introduced to the concept of light as an electromagnetic wave, you know that the ray model is not the ultimate truth about light, but the ray model is simpler, and in any case science always deals with models of reality, not the ultimate nature of reality. The following table summarizes three models of light.

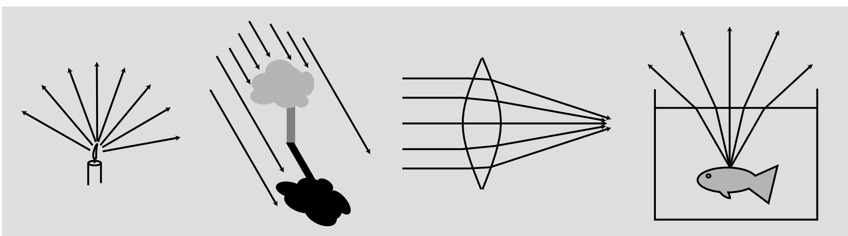


h / Three models of light.

The ray model is a generic one. By using it we can discuss the path taken by the light, without committing ourselves to any specific description of what it is that is moving along that path. We will use the nice simple ray model for most of this book, and with it we can analyze a great many devices and phenomena. Not until the last chapter will we concern ourselves specifically with wave optics, although in the intervening chapters I will sometimes analyze the same phenomenon using both the ray model and the wave model.

Note that the statements about the applicability of the various models are only rough guides. For instance, wave interference effects are often detectable, if small, when light passes around an obstacle that is quite a bit bigger than a wavelength. Also, the criterion for when we need the particle model really has more to do with energy scales than distance scales, although the two turn out to be related.

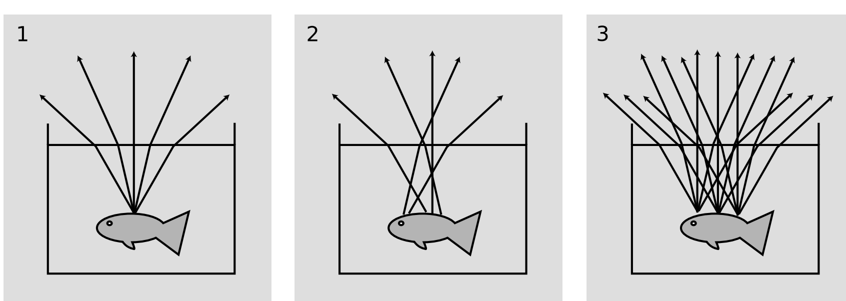
The alert reader may have noticed that the wave model is required at scales smaller than a wavelength of light (on the order of a micrometer for visible light), and the particle model is demanded on the atomic scale or lower (a typical atom being a nanometer or so in size). This implies that at the smallest scales we need *both* the wave model and the particle model. They appear incompatible, so how can we simultaneously use both? The answer is that they are not as incompatible as they seem. Light is both a wave and a particle, but a full understanding of this apparently nonsensical statement is a topic for the following book in this series.



i / Examples of ray diagrams.

Ray diagrams

Without even knowing how to use the ray model to calculate anything numerically, we can learn a great deal by drawing ray diagrams. For instance, if you want to understand how eyeglasses help you to see in focus, a ray diagram is the right place to start. Many students under-utilize ray diagrams in optics and instead rely on rote memorization or plugging into formulas. The trouble with memorization and plug-ins is that they can obscure what's really going on, and it is easy to get them wrong. Often the best plan is to do a ray diagram first, then do a numerical calculation, then check that your numerical results are in reasonable agreement with what you expected from the ray diagram.



j / 1. Correct. 2. Incorrect: implies that diffuse reflection only gives one ray from each reflecting point. 3. Correct, but unnecessarily complicated

Figure j shows some guidelines for using ray diagrams effectively. The light rays bend when then pass out through the surface of the water (a phenomenon that we'll discuss in more detail later). The rays appear to have come from a point above the goldfish's actual location, an effect that is familiar to people who have tried spear-fishing.

- A stream of light is not really confined to a finite number of narrow lines. We just draw it that way. In j/1, it has been necessary to choose a finite number of rays to draw (five), rather than the theoretically infinite number of rays that will diverge from that point.

- There is a tendency to conceptualize rays incorrectly as objects. In his Optics, Newton goes out of his way to caution the reader against this, saying that some people "consider ... the refraction of ... rays to be the bending or breaking of them in their passing out of one medium into another." But a ray is a record of the path traveled by light, not a physical thing that can be bent or broken.

- In theory, rays may continue infinitely far into the past and future, but we need to draw lines of finite length. In j/1, a judicious choice has been made as to where to begin and end the rays. There is no point in continuing the rays any farther than shown, because nothing new and exciting is going to happen to them. There is also no good reason to start them earlier, before being reflected by the fish, because the direction of the diffusely reflected rays is random anyway, and unrelated to the direction of the original, incoming ray.

- When representing diffuse reflection in a ray diagram, many students have a mental block against drawing many rays fanning out from the same point. Often, as in example j/2, the problem is the misconception that light can only be reflected in one direction from one point.

- Another difficulty associated with diffuse reflection, example j/3, is the tendency to think that in addition to drawing many rays coming out of one point, we should also be drawing many rays coming from many points. In j/1, drawing many rays coming out of one point gives useful information, telling us, for instance, that the fish can be seen from any angle. Drawing many sets of rays, as in j/3, does not give us any more useful information, and just clutters up the picture in this example. The only reason to draw sets of rays fanning out from more than one point would be if different things were happening to the different sets.

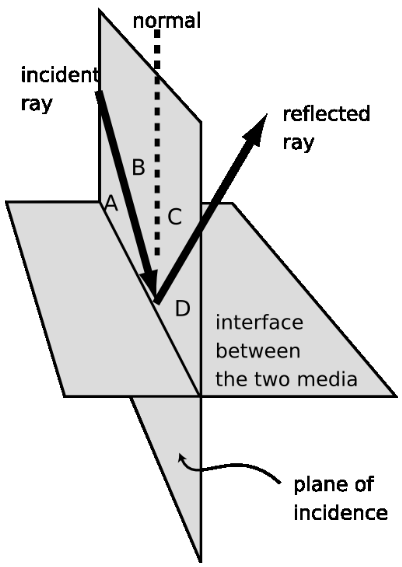
Discussion Question

A Suppose an intelligent tool-using fish is spear-hunting for humans. Draw a ray diagram to show how the fish has to correct its aim. Note that although the rays are now passing from the air to the water, the same rules apply: the rays are closer to being perpendicular to the surface when they are in the water, and rays that hit the air-water interface at a shallow angle are bent the most.

1.4 Geometry of Specular Reflection

To change the motion of a material object, we use a force. Is there any way to exert a force on a beam of light? Experiments show that electric and magnetic fields do not deflect light beams, so apparently light has no electric charge. Light also has no mass, so until the twentieth century it was believed to be immune to gravity as well. Einstein predicted that light beams would be very slightly deflected by strong gravitational fields, and he was proved correct by observations of rays of starlight that came close to the sun, but obviously that's not what makes mirrors and lenses work!

If we investigate how light is reflected by a mirror, we will find that the process is horrifically complex, but the final result is surprisingly simple. What actually happens is that the light is made of electric and magnetic fields, and these fields accelerate the electrons in the mirror. Energy from the light beam is momentarily transformed into extra kinetic energy of the electrons, but because the electrons are accelerating they re-radiate more light, converting their kinetic energy back into light energy. We might expect this to result in a very chaotic situation, but amazingly enough, the electrons move together to produce a new, reflected beam of light, which obeys two simple rules:



k / The geometry of specular reflection.

- The angle of the reflected ray is the same as that of the incident ray.
- The reflected ray lies in the plane containing the incident ray and the normal (perpendicular) line. This plane is known as the plane of incidence.

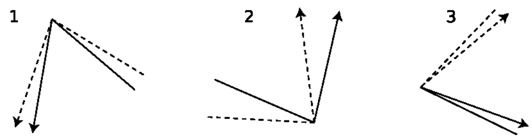
The two angles can be defined either with respect to the normal, like angles B and C in the figure, or with respect to the reflecting surface, like angles A and D. There is a convention of several hundred years' standing that one measures the angles with respect to the normal, but the rule about equal angles can logically be stated either as B=C or as A=D.

The phenomenon of reflection occurs only at the boundary between two media, just like the change in the speed of light that passes from one medium to another. As we have seen in book 3 of this series, this is the way all waves behave.

Most people are surprised by the fact that light can be reflected back from a less dense medium. For instance, if you are diving and you look up at the surface of the water, you will see a reflection of yourself.

self-check A

Each of these diagrams is supposed to show two different rays being reflected from the same point on the same mirror. Which are correct, and which are incorrect?



▷ Answer, p. 106

Reversibility of light rays

The fact that specular reflection displays equal angles of incidence and reflection means that there is a symmetry: if the ray had come in from the right instead of the left in the figure above, the angles would have looked exactly the same. This is not just a pointless detail about specular reflection. It's a manifestation of a very deep and important fact about nature, which is that the laws of physics do not distinguish between past and future. Cannonballs and planets have trajectories that are equally natural in reverse, and so do light rays. This type of symmetry is called time-reversal symmetry.

Typically, time-reversal symmetry is a characteristic of any process that does not involve heat. For instance, the planets do not experience any friction as they travel through empty space, so there is no frictional heating. We should thus expect the time-reversed versions of their orbits to obey the laws of physics, which they do. In contrast, a book sliding across a table does generate heat from friction as it slows down, and it is therefore not surprising that this type of motion does not appear to obey time-reversal symmetry. A book lying still on a flat table is never observed to spontaneously start sliding, sucking up heat energy and transforming it into kinetic energy.

Similarly, the only situation we've observed so far where light does not obey time-reversal symmetry is absorption, which involves heat. Your skin absorbs visible light from the sun and heats up, but we never observe people's skin to glow, converting heat energy into visible light. People's skin does glow in infrared light, but that doesn't mean the situation is symmetric. Even if you absorb infrared, you don't emit visible light, because your skin isn't hot enough to glow in the visible spectrum.

These apparent heat-related asymmetries are not actual asymmetries in the laws of physics. The interested reader may wish to learn more about this from the optional thermodynamics chapter of book 2 in this series.

Ray tracing on a computer *example 1*

A number of techniques can be used for creating artificial visual scenes in computer graphics. Figure l shows such a scene, which was created by the brute-force technique of simply constructing a very detailed ray diagram on a computer. This technique requires a great deal of computation, and is therefore too slow to be used for video games and computer-animated movies. One trick for speeding up the computation is to exploit the reversibility of light rays. If one was to trace every ray emitted by every illuminated surface, only a tiny fraction of those would actually end up passing into the virtual "camera," and therefore almost all of the computational effort would be wasted. One can instead start a ray at the camera, trace it backward in time, and see where it would have come from. With this technique, there is no wasted effort.

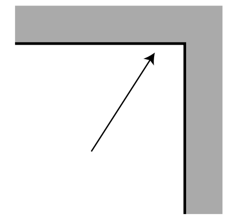
l / This photorealistic image of a nonexistent countertop was produced completely on a computer, by computing a complicated ray diagram.

Discussion Questions

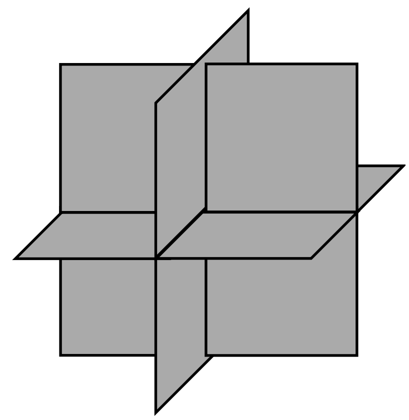
A If a light ray has a velocity vector with components c_x and c_y, what will happen when it is reflected from a surface that lies along the y axis? Make sure your answer does not imply a change in the ray's speed.

B Generalizing your reasoning from discussion question A, what will happen to the velocity components of a light ray that hits a corner, as shown in the figure, and undergoes two reflections?

C Three pieces of sheet metal arranged perpendicularly as shown in the figure form what is known as a radar corner. Let's assume that the radar corner is large compared to the wavelength of the radar waves, so that the ray model makes sense. If the radar corner is bathed in radar rays, at least some of them will undergo three reflections. Making a further generalization of your reasoning from the two preceding discussion questions, what will happen to the three velocity components of such a ray? What would the radar corner be useful for?



m / Discussion question B.



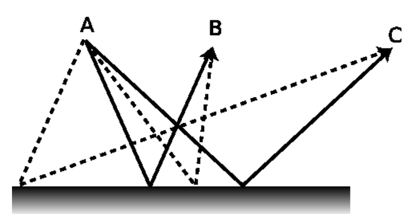
n / Discussion question C.

1.5 ⋆ The Principle of Least Time for Reflection

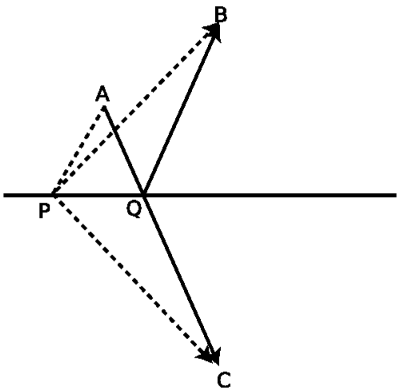
We had to choose between an unwieldy explanation of reflection at the atomic level and a simpler geometric description that was not as fundamental. There is a third approach to describing the interaction of light and matter which is very deep and beautiful. Emphasized by the twentieth-century physicist Richard Feynman, it is called the principle of least time, or Fermat's principle.

Let's start with the motion of light that is not interacting with matter at all. In a vacuum, a light ray moves in a straight line. This can be rephrased as follows: of all the conceivable paths light could follow from P to Q, the only one that is physically possible is the path that takes the least time.

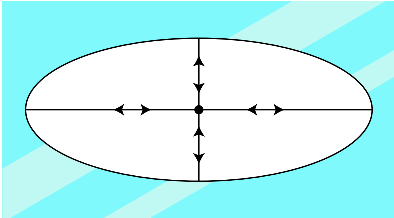
What about reflection? If light is going to go from one point to another, being reflected on the way, the quickest path is indeed the one with equal angles of incidence and reflection. If the starting and ending points are equally far from the reflecting surface, o, it's not hard to convince yourself that this is true, just based on symmetry. There is also a tricky and simple proof, shown in figure p, for the more general case where the points are at different distances from the surface.



o / The solid lines are physically possible paths for light rays traveling from A to B and from A to C. They obey the principle of least time. The dashed lines do not obey the principle of least time, and are not physically possible.



p / Paths AQB and APB are two conceivable paths that a ray could follow to get from A to B with one reflection, but only AQB is physically possible. We wish to prove that the path AQB, with equal angles of incidence and reflection, is shorter than any other path, such as APB. The trick is to construct a third point, C, lying as far below the surface as B lies above it. Then path AQC is a straight line whose length is the same as AQB's, and path APC has the same length as path APB. Since AQC is straight, it must be shorter than any other path such as APC that connects A and C, and therefore AQB must be shorter than any path such as APB.



q / Light is emitted at the center of an elliptical mirror. There are four physically possible paths by which a ray can be reflected and return to the center.

Not only does the principle of least time work for light in a vacuum and light undergoing reflection, we will also see in a later chapter that it works for the bending of light when it passes from one medium into another.

Although it is beautiful that the entire ray model of light can be reduced to one simple rule, the principle of least time, it may seem a little spooky to speak as if the ray of light is intelligent, and has carefully planned ahead to find the shortest route to its destination. How does it know in advance where it's going? What if we moved the mirror while the light was en route, so conditions along its planned path were not what it "expected?" The answer is that the principle of least time is really a shortcut for finding certain results of the wave model of light, which is the topic of the last chapter of this book.

There are a couple of subtle points about the principle of least time. First, the path does not have to be the quickest of all possible paths; it only needs to be quicker than any path that differs infinitesimally from it. In figure p, for instance, light could get from A to B either by the reflected path AQB or simply by going straight from A to B. Although AQB is not the shortest possible path, it cannot be shortened by changing it infinitesimally, e.g., by moving Q a little to the right or left. On the other hand, path APB is physically impossible, because it is possible to improve on it by moving point P infinitesimally to the right.

It's not quite right to call this the principle of *least* time. In figure q, for example, the four physically possible paths by which a ray can return to the center consist of two shortest-time paths and two longest-time paths. Strictly speaking, we should refer to the *principle of least or greatest time*, but most physicists omit the niceties, and assume that other physicists understand that both maxima and minima are possible.

Summary

Selected Vocabulary

absorption	what happens when light hits matter and gives up some of its energy
reflection	what happens when light hits matter and bounces off, retaining at least some of its energy
specular reflection	reflection from a smooth surface, in which the light ray leaves at the same angle at which it came in
diffuse reflection	reflection from a rough surface, in which a single ray of light is divided up into many weaker reflected rays going in many directions
normal	the line perpendicular to a surface at a given point

Notation

c	the speed of light

Summary

We can understand many phenomena involving light without having to use sophisticated models such as the wave model or the particle model. Instead, we simply describe light according to the path it takes, which we call a ray. The ray model of light is useful when light is interacting with material objects that are much larger than a wavelength of light. Since a wavelength of visible light is so short compared to the human scale of existence, the ray model is useful in many practical cases.

We see things because light comes from them to our eyes. Objects that glow may send light directly to our eyes, but we see an object that doesn't glow via light from another source that has been reflected by the object.

Many of the interactions of light and matter can be understood by considering what happens when light reaches the boundary between two different substances. In this situation, part of the light is reflected (bounces back) and part passes on into the new medium. This is not surprising — it is typical behavior for a wave, and light is a wave. Light energy can also be absorbed by matter, i.e., converted into heat.

A smooth surface produces specular reflection, in which the reflected ray exits at the same angle with respect to the normal as that of the incoming ray. A rough surface gives diffuse reflection, where a single ray of light is divided up into many weaker reflected rays going in many directions.

Problems

Key

- √ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ⋆ A difficult problem.

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).

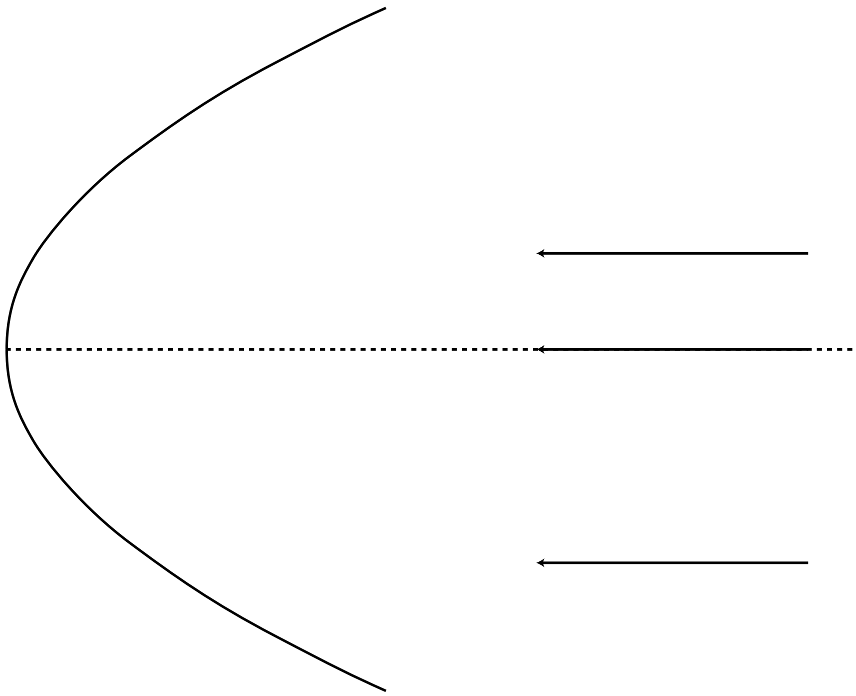
2 A Global Positioning System (GPS) receiver is a device that lets you figure out where you are by receiving timed radio signals from satellites. It works by measuring the travel time for the signals, which is related to the distance between you and the satellite. By finding the ranges to several different 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 your electricity and magnetism text, in what part of the electromagnetic spectrum would such a wave lie (infrared, gamma-rays,...)?

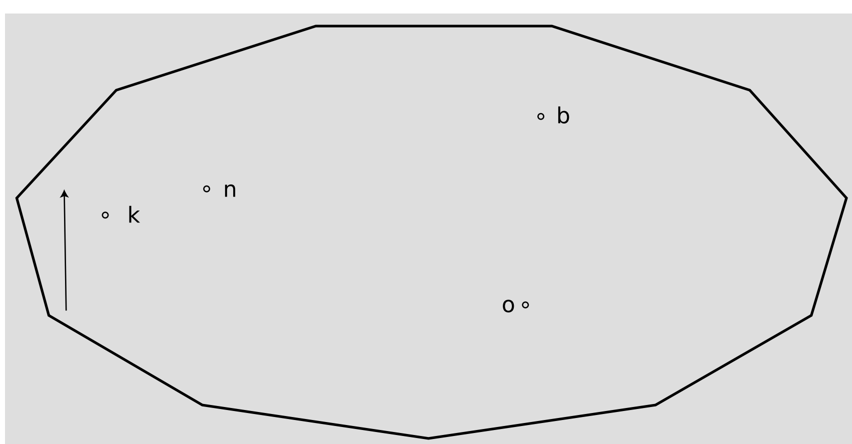
4 The Stealth bomber is designed with flat, smooth surfaces. Why would this make it difficult to detect via radar?

5 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) Trace the drawing accurately, and 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?

6 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 draw in the normal at each place where the ray strikes a bumper.



Problem 5.



Problem 6.

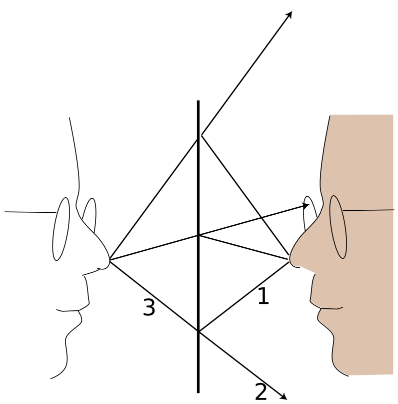
Narcissus, by Michelangelo Caravaggio, ca. 1598.

Chapter 2
Images by Reflection

Infants are always fascinated by the antics of the Baby in the Mirror. Now if you want to know something about mirror images that most people don't understand, try this. First bring this page closer wand closer to your eyes, until you can no longer focus on it without straining. Then go in the bathroom and see how close you can get your face to the surface of the mirror before you can no longer easily focus on the image of your own eyes. You will find that the shortest comfortable eye-mirror distance is much less than the shortest comfortable eye-paper distance. This demonstrates that the image of your face in the mirror acts as if it had depth and existed in the space *behind* the mirror. If the image was like a flat picture in a book, then you wouldn't be able to focus on it from such a short distance.

In this chapter we will study the images formed by flat and curved mirrors on a qualitative, conceptual basis. Although this type of image is not as commonly encountered in everyday life as images formed by lenses, images formed by reflection are simpler to understand, so we discuss them first. In chapter 3 we will turn to a more mathematical treatment of images made by reflection. Surprisingly, the same equations can also be applied to lenses, which are the topic of chapter 4.

2.1 A Virtual Image



a / An image formed by a mirror.

We can understand a mirror image using a ray diagram. Figure a shows several light rays, 1, that originated by diffuse reflection at the person's nose. They bounce off the mirror, producing new rays, 2. To anyone whose eye is in the right position to get one of these rays, they appear to have come from a behind the mirror, 3, where they would have originated from a single point. This point is where the tip of the image-person's nose appears to be. A similar analysis applies to every other point on the person's face, so it looks as though there was an entire face behind the mirror. The customary way of describing the situation requires some explanation:

Customary description in physics: There is an image of the face behind the mirror.

Translation: The pattern of rays coming from the mirror is exactly the same as it would be if there was a face behind the mirror. Nothing is really behind the mirror.

This is referred to as a *virtual* image, because the rays do not actually cross at the point behind the mirror. They only appear to have originated there.

self-check A

Imagine that the person in figure a moves his face down quite a bit — a couple of feet in real life, or a few inches on this scale drawing. Draw a new ray diagram. Will there still be an image? If so, where is it visible from?

▷ Answer, p. 106

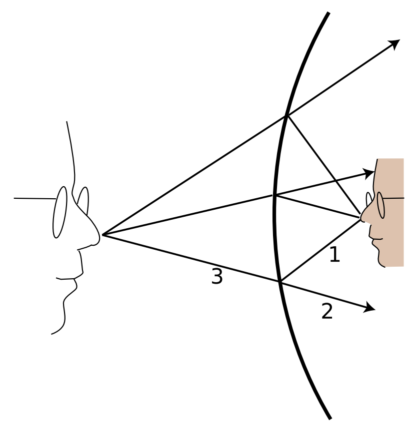
The geometry of specular reflection tells us that rays 1 and 2 are at equal angles to the normal (the imaginary perpendicular line piercing the mirror at the point of reflection). This means that ray 2's imaginary continuation, 3, forms the same angle with the mirror as ray 3. Since each ray of type 3 forms the same angles with the mirror as its partner of type 1, we see that the distance of the image from the mirror is the same as the actual face from the mirror, and lies directly across from it. The image therefore appears to be the same size as the actual face.

Discussion Question

A The figure shows an object that is off to one side of a mirror. Draw a ray diagram. Is an image formed? If so, where is it, and from which directions would it be visible?

2.2 Curved Mirrors

An image in a flat mirror is a pretechnological example: even animals can look at their reflections in a calm pond. We now pass to our first nontrivial example of the manipulation of an image by technology: an image in a curved mirror. Before we dive in, let's consider why this is an important example. If it was just a question of memorizing a bunch of facts about curved mirrors, then you would rightly rebel against an effort to spoil the beauty of your liberally educated brain by force-feeding you technological trivia. The reason this is an important example is not that curved mirrors are so important in and of themselves, but that the results we derive for curved bowl-shaped mirrors turn out to be true for a large class of other optical devices, including mirrors that bulge outward rather than inward, and lenses as well. A microscope or a telescope is simply a combination of lenses or mirrors or both. What you're really learning about here is the basic building block of all optical devices from movie projectors to octopus eyes.

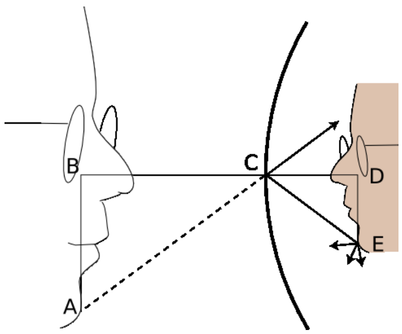


b / An image formed by a curved mirror.

Because the mirror in figure b is curved, it bends the rays back closer together than a flat mirror would: we describe it as *converging*. Note that the term refers to what it does to the light rays, not to the physical shape of the mirror's surface . (The surface itself would be described as *concave*. The term is not all that hard to remember, because the hollowed-out interior of the mirror is like a cave.) It is surprising but true that all the rays like 3 really do converge on a point, forming a good image. We will not prove this fact, but it is true for any mirror whose curvature is gentle enough and that is symmetric with respect to rotation about the perpendicular line passing through its center (not asymmetric like a potato chip). The old-fashioned method of making mirrors and lenses is by grinding them in grit by hand, and this automatically tends to produce an almost perfect spherical surface.

Bending a ray like 2 inward implies bending its imaginary continuation 3 outward, in the same way that raising one end of a seesaw causes the other end to go down. The image therefore forms deeper behind the mirror. This doesn't just show that there is extra distance between the image-nose and the mirror; it also implies that the image itself is bigger from front to back. It has been *magnified* in the front-to-back direction.

It is easy to prove that the same magnification also applies to the image's other dimensions. Consider a point like E in figure c. The trick is that out of all the rays diffusely reflected by E, we pick the one that happens to head for the mirror's center, C. The equal-angle property of specular reflection plus a little straightforward geometry easily leads us to the conclusion that triangles ABC and CDE are the same shape, with ABC being simply a scaled-up version of CDE. The magnification of depth equals the ratio BC/CD, and the up-down magnification is AB/DE. A repetition of the same proof shows that the magnification in the third dimension (out of the page) is also the same. This means that the image-head is simply a larger version of the real one, without any distortion. The scaling factor is called the magnification, M. The image in the figure is magnified by a factor $M = 1.9$.



c / The image is magnified by the same factor in depth and in its other dimensions.

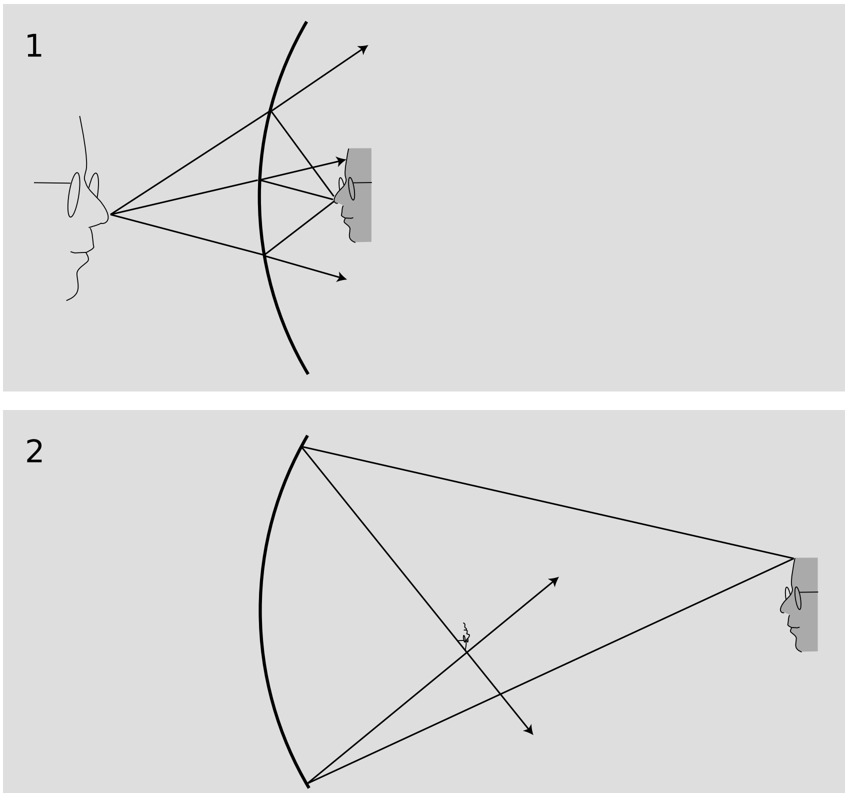
Note that we did not explicitly specify whether the mirror was a sphere, a paraboloid, or some other shape. However, we assumed that a focused image would be formed, which would not necessarily be true, for instance, for a mirror that was asymmetric or very deeply curved.

2.3 A Real Image

If we start by placing an object very close to the mirror, d/1, and then move it farther and farther away, the image at first behaves as we would expect from our everyday experience with flat mirrors, receding deeper and deeper behind the mirror. At a certain point, however, a dramatic change occurs. When the object is more than a certain distance from the mirror, d/2, the image appears upside-down and in *front* of the mirror.

Here's what's happened. The mirror bends light rays inward, but when the object is very close to it, as in d/1, the rays coming from a given point on the object are too strongly diverging (spreading) for the mirror to bring them back together. On reflection, the rays are still diverging, just not as strongly diverging. But when the object is sufficiently far away, d/2, the mirror is only intercepting the rays that came out in a narrow cone, and it is able to bend these enough so that they will reconverge.

Note that the rays shown in the figure, which both originated at the same point on the object, reunite when they cross. The point where they cross is the image of the point on the original object. This type of image is called a *real image*, in contradistinction to the virtual images we've studied before. The use of the word "real" is perhaps unfortunate. It sounds as though we are saying the image was an actual material object, which of course it is not.



d / 1. A virtual image. 2. A real image. As you'll verify in homework problem 6, the image is upside-down

The distinction between a real image and a virtual image is an important one, because a real image can projected onto a screen or photographic film. If a piece of paper is inserted in figure d/2 at the location of the image, the image will be visible on the paper (provided the object is bright and the room is dark). Your eye uses a lens to make a real image on the retina.

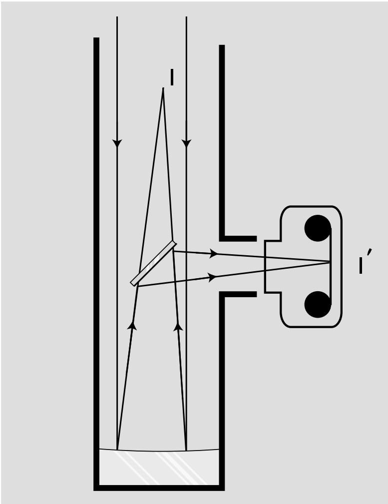
self-check B

Sketch another copy of the face in figure d/1, even farther from the mirror, and draw a ray diagram. What has happened to the location of the image? ▷ Answer, p. 106

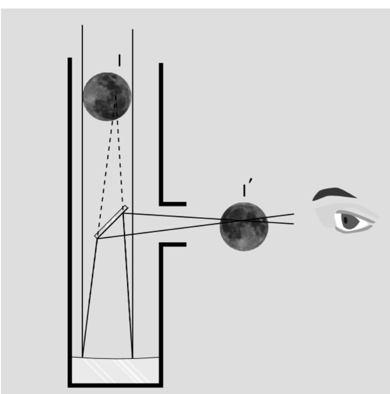
2.4 Images of Images

If you are wearing glasses right now, then the light rays from the page are being manipulated first by your glasses and then by the lens of your eye. You might think that it would be extremely difficult to analyze this, but in fact it is quite easy. In any series of optical elements (mirrors or lenses or both), each element works on the rays furnished by the previous element in exactly the same manner as if the image formed by the previous element was an actual object.

Figure e shows an example involving only mirrors. The Newto-



e / A Newtonian telescope being used with a camera.



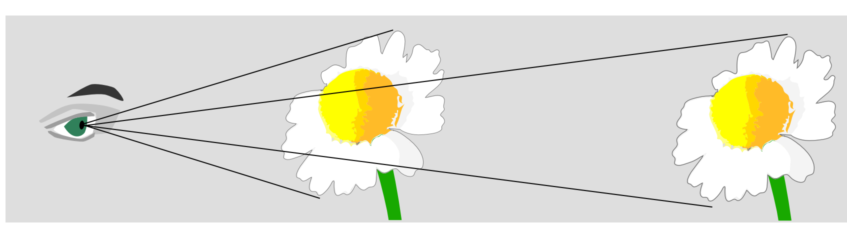
f / A Newtonian telescope being used for visual rather than photographic observing. In real life, an eyepiece lens is normally used for additional magnification, but this simpler setup will also work.

nian telescope, invented by Isaac Newton, consists of a large curved mirror, plus a second, flat mirror that brings the light out of the tube. (In very large telescopes, there may be enough room to put a camera or even a person inside the tube, in which case the second mirror is not needed.) The tube of the telescope is not vital; it is mainly a structural element, although it can also be helpful for blocking out stray light. The lens has been removed from the front of the camera body, and is not needed for this setup. Note that the two sample rays have been drawn parallel, because an astronomical telescope is used for viewing objects that are extremely far away. These two "parallel" lines actually meet at a certain point, say a crater on the moon, so they can't actually be perfectly parallel, but they are parallel for all practical purposes since we would have to follow them upward for a quarter of a million miles to get to the point where they intersect.

The large curved mirror by itself would form an image I, but the small flat mirror creates an image of the image, I′. The relationship between I and I′ is exactly the same as it would be if I was an actual object rather than an image: I and I′ are at equal distances from the plane of the mirror, and the line between them is perpendicular to the plane of the mirror.

One surprising wrinkle is that whereas a flat mirror used by itself forms a virtual image of an object that is real, here the mirror is forming a real image of virtual image I. This shows how pointless it would be to try to memorize lists of facts about what kinds of images are formed by various optical elements under various circumstances. You are better off simply drawing a ray diagram.

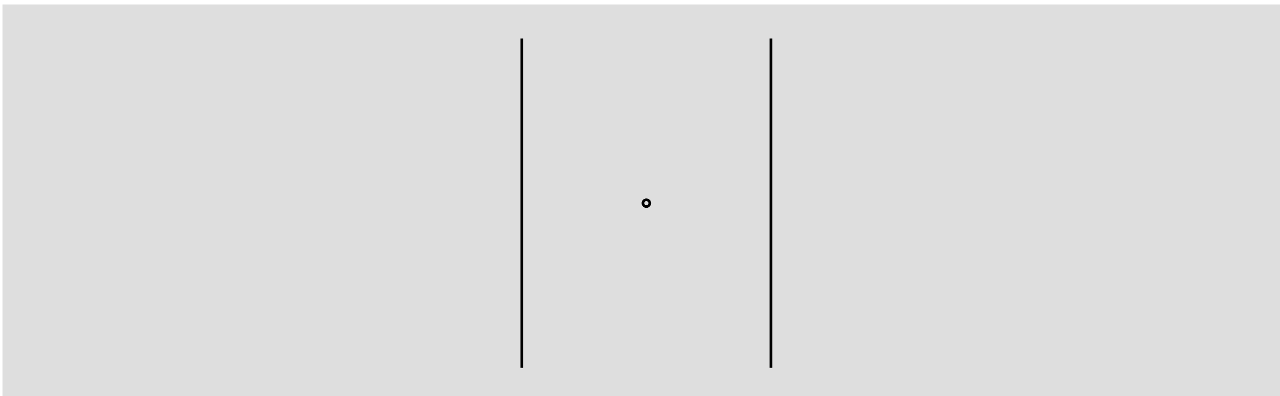
Although the main point here was to give an example of an image of an image, figure f shows an interesting case where we need to make the distinction between *magnification* and *angular magnification*. If you are looking at the moon through this telescope, then the images I and I′ are much *smaller* than the actual moon. Otherwise, for example, image I would not fit inside the telescope! However, these images are very close to your eye compared to the actual moon. The small size of the image has been more than compensated for by the shorter distance. The important thing here is the amount of *angle* within your field of view that the image covers, and it is this angle that has been increased. The factor by which it is increased is called the *angular magnification*, M_a.



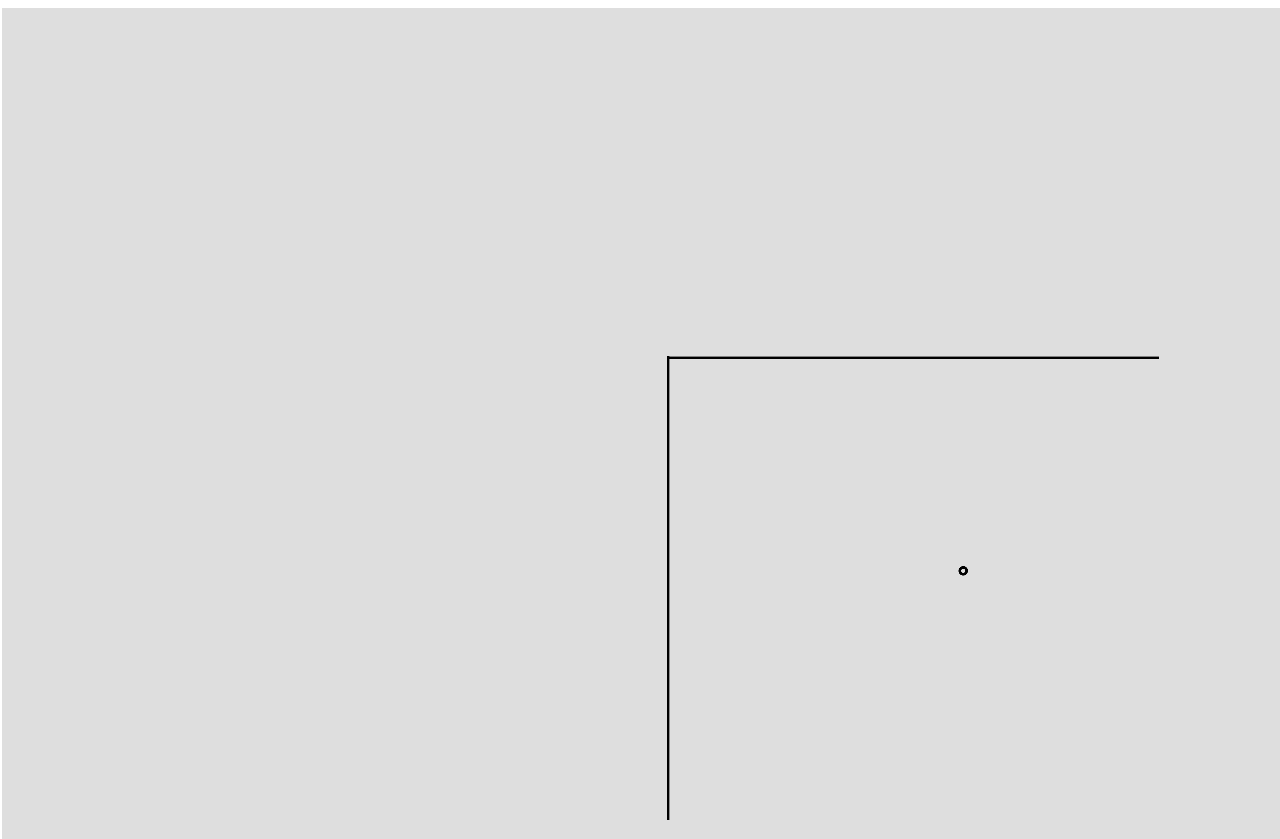
g / The angular size of the flower depends on its distance from the eye.

Discussion Questions

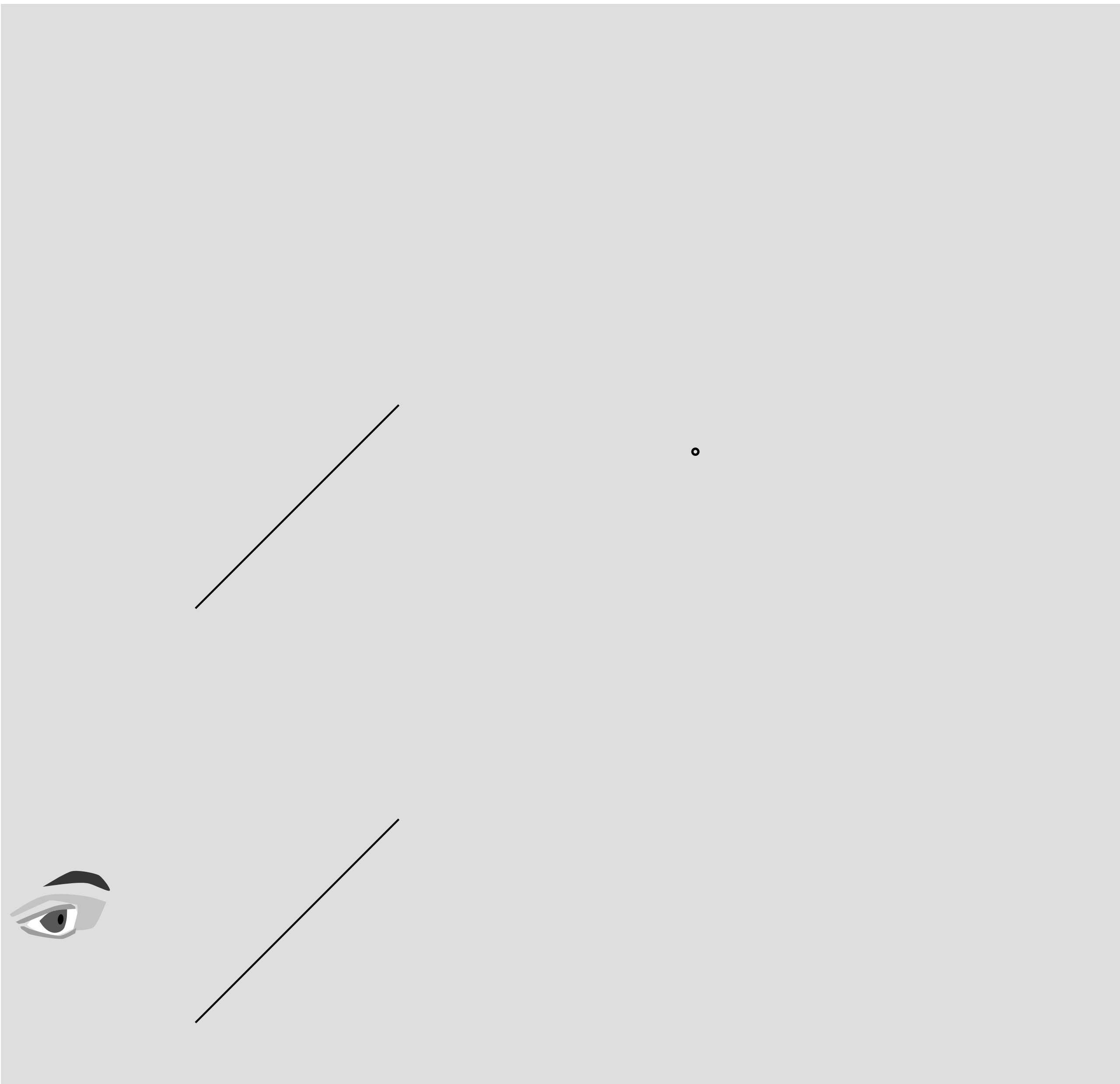
A Locate the images of you that will be formed if you stand between two parallel mirrors.



B Locate the images formed by two perpendicular mirrors, as in the figure. What happens if the mirrors are not perfectly perpendicular?



C Locate the images formed by the periscope.



Summary

Selected Vocabulary

real image	a place where an object appears to be, because the rays diffusely reflected from any given point on the object have been bent so that they come back together and then spread out again from the new point
virtual image	like a real image, but the rays don't actually cross again; they only appear to have come from the point on the image
converging	describes an optical device that brings light rays closer to the optical axis
diverging	bends light rays farther from the optical axis
magnification	the factor by which an image's linear size is increased (or decreased)
angular magnification	the factor by which an image's apparent angular size is increased (or decreased)
concave	describes a surface that is hollowed out like a cave
convex	describes a surface that bulges outward

Notation

M	the magnification of an image
M_a	the angular magnification of an image

Summary

A large class of optical devices, including lenses and flat and curved mirrors, operates by bending light rays to form an image. A real image is one for which the rays actually cross at each point of the image. A virtual image, such as the one formed behind a flat mirror, is one for which the rays only appear to have crossed at a point on the image. A real image can be projected onto a screen; a virtual one cannot.

Mirrors and lenses will generally make an image that is either smaller than or larger than the original object. The scaling factor is called the magnification. In many situations, the angular magnification is more important than the actual magnification.

Problems

Key

$\checkmark$ A computerized answer check is available online.
$\int$ A problem that requires calculus.
$\star$ A difficult problem.

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

2 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.

3 In this chapter 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?

4 As discussed in question 3, 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.

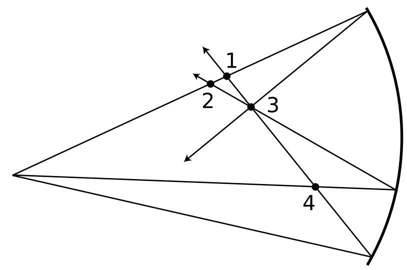
Draw BIG ray diagrams! Each diagram should use up about half a page of paper.

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.

5 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.)

6 In figure d in on page 35, 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.

7 The figure shows four points where rays cross. Of these, which are image points? Explain.



Problem 7.

8 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?

Breakfast Table, by Willem Clasz. de Heda, 17th century. The painting shows a variety of images, some of them distorted, resulting both from reflection and from refraction (ch. 4).

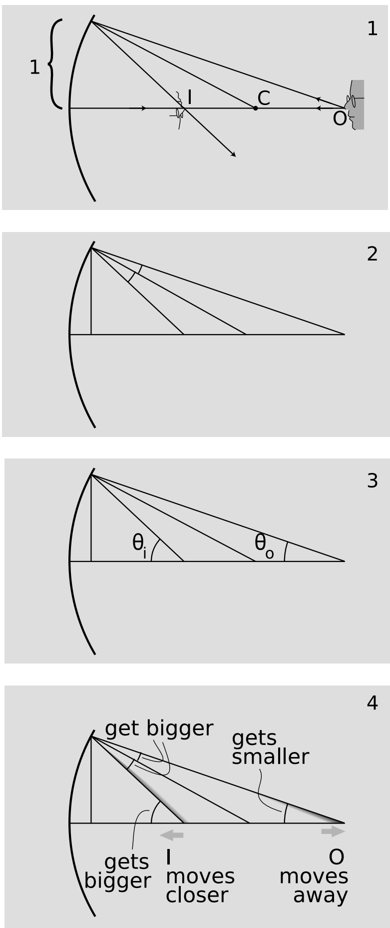
Chapter 3

Images, Quantitatively

It sounds a bit odd when a scientist refers to a theory as "beautiful," but to those in the know it makes perfect sense. One mark of a beautiful theory is that it surprises us by being simple. The mathematical theory of lenses and curved mirrors gives us just such a surprise. We expect the subject to be complex because there are so many cases: a converging mirror forming a real image, a diverging lens that makes a virtual image, and so on for a total of six possibilities. If we want to predict the location of the images in all these situations, we might expect to need six different equations, and six more for predicting magnifications. Instead, it turns out that we can use just one equation for the location of the image and one equation for its magnification, and these two equations work in all the different cases with no changes except for plus and minus signs. This is the kind of thing the physicist Eugene Wigner referred to as "the unreasonable effectiveness of mathematics." Sometimes we can find a deeper reason for this kind of unexpected simplicity, but sometimes it almost seems as if God went out of Her way to make the secrets of universe susceptible to attack by the human thought-tool called math.

3.1 A Real Image Formed by a Converging Mirror

Location of the image



a / The relationship between the object's position and the image's can be expressed in terms of the angles θ_o and θ_i.

We will now derive the equation for the location of a real image formed by a converging mirror. We assume for simplicity that the mirror is spherical, but actually this isn't a restrictive assumption, because any shallow, symmetric curve can be approximated by a sphere. The shape of the mirror can be specified by giving the location of its center, C. A deeply curved mirror is a sphere with a small radius, so C is close to it, while a weakly curved mirror has C farther away. Given the point O where the object is, we wish to find the point I where the image will be formed.

To locate an image, we need to track a minimum of two rays coming from the same point. Since we have proved in the previous chapter that this type of image is not distorted, we can use an on-axis point, O, on the object, as in figure a/1. The results we derive will also hold for off-axis points, since otherwise the image would have to be distorted, which we know is not true. We let one of the rays be the one that is emitted along the axis; this ray is especially easy to trace, because it bounces straight back along the axis again. As our second ray, we choose one that strikes the mirror at a distance of 1 from the axis. "One what?" asks the astute reader. The answer is that it doesn't really matter. When a mirror has shallow curvature, all the reflected rays hit the same point, so 1 could be expressed in any units you like. It could, for instance, be 1 cm, unless your mirror is smaller than 1 cm!

The only way to find out anything mathematical about the rays is to use the sole mathematical fact we possess concerning specular reflection: the incident and reflected rays form equal angles with respect to the normal, which is shown as a dashed line. Therefore the two angles shown in figure a/2 are the same, and skipping some straightforward geometry, this leads to the visually reasonable result that the two angles in figure a/3 are related as follows:

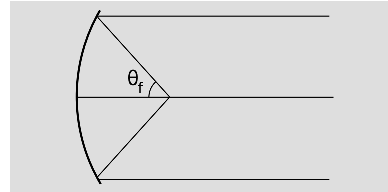
$$\theta_i + \theta_o = \text{constant}$$

(Note that θ_i and θ_o, which are measured from the image and the object, not from the eye like the angles we referred to in discussing angular magnification on page 36.) For example, move O farther from the mirror. The top angle in figure a/2 is increased, so the bottom angle must increase by the same amount, causing the image point, I, to move closer to the mirror. In terms of the angles shown in figure a/3, the more distant object has resulted in a smaller angle θ_o, while the closer image corresponds to a larger θ_i; One angle increases by the same amount that the other decreases, so their sum remains constant. These changes are summarized in figure a/4.

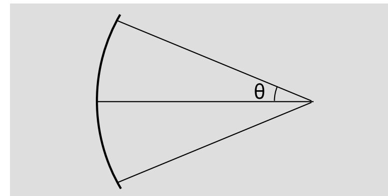
The sum $\theta_i + \theta_o$ is a constant. What does this constant represent? Geometrically, we interpret it as double the angle made by the dashed radius line. Optically, it is a measure of the strength of the mirror, i.e., how strongly the mirror focuses light, and so we call it the focal angle, θ_f,

$$\theta_i + \theta_o = \theta_f$$

Suppose, for example, that we wish to use a quick and dirty optical test to determine how strong a particular mirror is. We can lay it on the floor as shown in figure c, and use it to make an image of a lamp mounted on the ceiling overhead, which we assume is very far away compared to the radius of curvature of the mirror, so that the mirror intercepts only a very narrow cone of rays from the lamp. This cone is so narrow that its rays are nearly parallel, and θ_o is nearly zero. The real image can be observed on a piece of paper. By moving the paper nearer and farther, we can bring the image into focus, at which point we know the paper is located at the image point. Since $\theta_o \approx 0$, we have $\theta_i \approx \theta_f$, and we can then determine this mirror's focal angle either by measuring θ_i directly with a protractor, or indirectly via trigonometry. A strong mirror will bring the rays together to form an image close to the mirror, and these rays will form a blunt-angled cone with a large θ_i and θ_f.



b / The geometrical interpretation of the focal angle.



c / Example 1, an alternative test for finding the focal angle. The mirror is the same as in figure b.

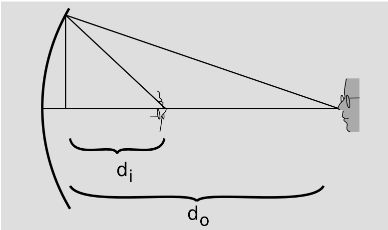
An alternative optical test — *example 1*

▷ Figure c shows an alternative optical test. Rather than placing the object at infinity as in figure b, we adjust it so that the image is right on top of the object. Points O and I coincide, and the rays are reflected right back on top of themselves. If we measure the angle θ shown in figure c, how can we find the focal angle?

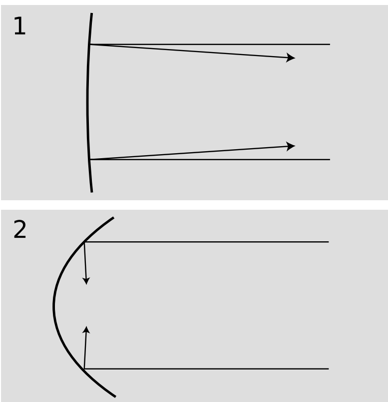
▷ The object and image angles are the same; the angle labeled θ in the figure equals both of them. We therefore have $\theta_i + \theta_o = \theta = \theta_f$. Comparing figures b and c, it is indeed plausible that the angles are related by a factor of two.

At this point, we could consider our work to be done. Typically, we know the strength of the mirror, and we want to find the image location for a given object location. Given the mirror's focal angle and the object location, we can determine θ_o by trigonometry, subtract to find $\theta_i = \theta_f - \theta_o$, and then do more trig to find the image location.

There is, however, a shortcut that can save us from doing so much work. Figure a/3 shows two right triangles whose legs of length 1 coincide and whose acute angles are θ_o and θ_i. These can be related by trigonometry to the object and image distances shown



d / The object and image distances



e / Mirror 1 is weaker than mirror 2. It has a shallower curvature, a longer focal length, and a smaller focal angle. It reflects rays at angles not much different than those that would be produced with a flat mirror.

in figure d:

$$\tan\theta_o = 1/d_o \qquad \tan\theta_i = 1/d_i$$

Ever since chapter 2, we've been assuming small angles. For small angles, we can use the small-angle approximation $\tan x \approx x$ (for x in radians), giving simply

$$\theta_o = 1/d_o \qquad \theta_i = 1/d_i \qquad .$$

We likewise define a distance called the focal length, f according to $\theta_f = 1/f$. In figure b, f is the distance from the mirror to the place where the rays cross. We can now reexpress the equation relating the object and image positions as

$$\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o} \qquad .$$

Figure e summarizes the interpretation of the focal length and focal angle.[1]

Which form is better, $\theta_f = \theta_i + \theta_o$ or $1/f = 1/d_i + 1/d_o$? The angular form has in its favor its simplicity and its straightforward visual interpretation, but there are two reasons why we might prefer the second version. First, the numerical values of the angles depend on what we mean by "one unit" for the distance shown as 1 in figure a/1. Second, it is usually easier to measure distances rather than angles, so the distance form is more convenient for number crunching. Neither form is superior overall, and we will often need to use both to solve any given problem.[2]

A searchlight *example 2*

Suppose we need to create a parallel beam of light, as in a searchlight. Where should we place the lightbulb? A parallel beam has zero angle between its rays, so $\theta_i = 0$. To place the lightbulb correctly, however, we need to know a distance, not an angle: the distance d_o between the bulb and the mirror. The problem involves a mixture of distances and angles, so we need to get everything in terms of one or the other in order to solve it. Since the goal is to find a distance, let's figure out the image distance corresponding to the given angle $\theta_i = 0$. These are related by $d_i = 1/\theta_i$, so we have $d_i = \infty$. (Yes, dividing by zero gives

[1] There is a standard piece of terminology which is that the "focal point" is the point lying on the optical axis at a distance from the mirror equal to the focal length. This term isn't particularly helpful, because it names a location where nothing normally happens. In particular, it is *not* normally the place where the rays come to a focus! — that would be the *image* point. In other words, we don't normally have $d_i = f$, unless perhaps $d_o = \infty$. A recent online discussion among some physics teachers (https://carnot.physics.buffalo.edu/archives, Feb. 2006) showed that many disliked the terminology, felt it was misleading, or didn't know it and would have misinterpreted it if they had come across it. That is, it appears to be what grammarians call a "skunked term" — a word that bothers half the population when it's used incorrectly, and the other half when it's used correctly.

[2] I would like to thank Fouad Ajami for pointing out the pedagogical advantages of using both equations side by side.

infinity. Don't be afraid of infinity. Infinity is a useful problem-solving device.) Solving the distance equation for d_o, we have

$$\begin{aligned} d_o &= (1/f - 1/d_i)^{-1} \\ &= (1/f - 0)^{-1} \\ &= (1/f)^{-1} \\ &= f \end{aligned}$$

The bulb has to be placed at a distance from the mirror equal to its focal point.

Diopters *example 3*

An equation like $d_i = 1/\theta_i$ really doesn't make sense in terms of units. Angles are unitless, since radians aren't really units, so the right-hand side is unitless. We can't have a left-hand side with units of distance if the right-hand side of the same equation is unitless. This is an artifact of my cavalier statement that the conical bundles of rays spread out to a distance of 1 from the axis where they strike the mirror, without specifying the units used to measure this 1. In real life, optometrists define the thing we're calling $\theta_i = 1/d_i$ as the "dioptric strength" of a lens or mirror, and measure it in units of inverse meters (m^{-1}), also known as diopters (1 D=1 m^{-1}).

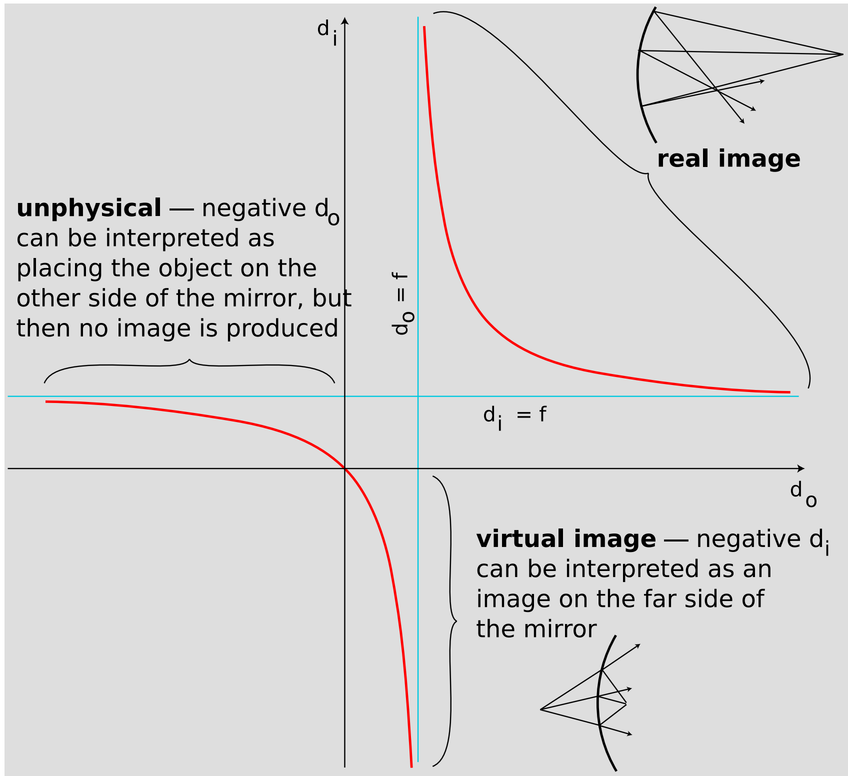
Magnification

We have already discussed in the previous chapter how to find the magnification of a virtual image made by a curved mirror. The result is the same for a real image, and we omit the proof, which is very similar. In our new notation, the result is $M = d_i/d_o$. A numerical example is given in section 3.2

3.2 Other Cases With Curved Mirrors

The equation $d_i =$ can easily produce a negative result, but we have been thinking of d_i as a distance, and distances can't be negative. A similar problem occurs with $\theta_i = \theta_f - \theta_o$ for $\theta_o > \theta_f$. What's going on here?

The interpretation of the angular equation is straightforward. As we bring the object closer and closer to the image, θ_o gets bigger and bigger, and eventually we reach a point where $\theta_o = \theta_f$ and $\theta_i = 0$. This large object angle represents a bundle of rays forming a cone that is very broad, so broad that the mirror can no longer bend them back so that they reconverge on the axis. The image angle $\theta_i = 0$ represents an outgoing bundle of rays that are parallel. The outgoing rays never cross, so this is not a real image, unless we want to be charitable and say that the rays cross at infinity. If we go on bringing the object even closer, we get a virtual image.



f / A graph of the image distance d_i as a function of the object distance d_o.

To analyze the distance equation, let's look at a graph of d_i as a function of d_o. The branch on the upper right corresponds to the case of a real image. Strictly speaking, this is the only part of the graph that we've proven corresponds to reality, since we never did any geometry for other cases, such as virtual images. As discussed in the previous section, making d_o bigger causes d_i to become smaller, and vice-versa.

Letting d_o be less than f is equivalent to $\theta_o > \theta_f$: a virtual image is produced on the far side of the mirror. This is the first example of Wigner's "unreasonable effectiveness of mathematics" that we have encountered in optics. Even though our proof depended on the assumption that the image was real, the equation we derived turns out to be applicable to virtual images, provided that we either interpret the positive and negative signs in a certain way, or else modify the equation to have different positive and negative signs.

self-check A

Interpret the three places where, in physically realistic parts of the graph, the graph approaches one of the dashed lines. [This will come more naturally if you have learned the concept of limits in a math class.] ▷ Answer, p. 106

A flat mirror *example 4*

We can even apply the equation to a flat mirror. As a sphere gets bigger and bigger, its surface is more and more gently curved. The planet Earth is so large, for example, that we cannot even perceive the curvature of its surface. To represent a flat mirror, we let the mirror's radius of curvature, and its focal length, become infinite. Dividing by infinity gives zero, so we have

$$1/d_o = -1/d_i \qquad ,$$

or

$$d_o = -d_i \qquad .$$

If we interpret the minus sign as indicating a virtual image on the far side of the mirror from the object, this makes sense.

It turns out that for any of the six possible combinations of real or virtual images formed by converging or diverging lenses or mirrors, we can apply equations of the form

$$\theta_f = \theta_i + \theta_o$$

and

$$\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o} \qquad ,$$

with only a modification of plus or minus signs. There are two possible approaches here. The approach we have been using so far is the more popular approach in American textbooks: leave the equation the same, but attach interpretations to the resulting negative or positive values of the variables. The trouble with this approach is that one is then forced to memorize tables of sign conventions, e.g. that the value of d_i should be negative when the image is a virtual image formed by a converging mirror. Positive and negative signs also have to be memorized for focal lengths. Ugh! It's highly unlikely that any student has ever retained these lengthy tables in his or her mind for more than five minutes after handing in the final exam in a physics course. Of course one can always look such things up when they are needed, but the effect is to turn the whole thing into an exercise in blindly plugging numbers into formulas.

As you have gathered by now, there is another method which I think is better, and which I'll use throughout the rest of this book. In this method, all distances and angles are *positive by definition*, and we put in positive and negative signs in the *equations* depending on the situation. (I thought I was the first to invent this method, but I've been told that this is known as the European sign convention, and that it's fairly common in Europe.) Rather than memorizing these signs, we start with the generic equations

$$\theta_f = \pm\theta_i \pm \theta_o$$
$$\frac{1}{f} = \pm\frac{1}{d_i} \pm \frac{1}{d_o} \qquad ,$$

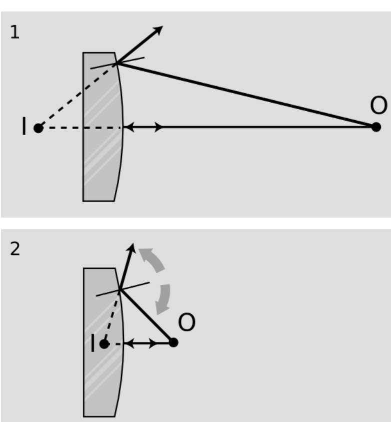
and then determine the signs by a two-step method that depends on ray diagrams. There are really only two signs to determine, not four; the signs in the two equations match up in the way you'd expect. The method is as follows:

1. Use ray diagrams to decide whether θ_o and θ_i vary in the same way or in opposite ways. (In other words, decide whether making θ_o greater results in a greater value of θ_i or a smaller one.) Based on this, decide whether the two signs in the angle equation are the same or opposite. If the signs are opposite, go on to step 2 to determine which is positive and which is negative.

2. If the signs are opposite, we need to decide which is the positive one and which is the negative. Since the focal angle is never negative, the smaller angle must be the one with a minus sign.

In step 1, many students have trouble drawing the ray diagram correctly. For simplicity, you should always do your diagram for a point on the object that is on the axis of the mirror, and let one of your rays be the one that is emitted along the axis and reflect straight back on itself, as in the figures in section 3.1. As shown in figure a/4 in section 3.1, there are four angles involved: two at the mirror, one at the object (θ_o), and one at the image (θ_i). Make sure to draw in the normal to the mirror so that you can see the two angles at the mirror. These two angles are equal, so as you change the object position, they fan out or fan in, like opening or closing a book. Once you've drawn this effect, you should easily be able to tell whether θ_o and θ_i change in the same way or in opposite ways.

Although focal lengths are always positive in the method used in this book, you should be aware that diverging mirrors and lenses are assigned negative focal lengths in the other method, so if you see a lens labeled $f = -30$ cm, you'll know what it means.



g / Example 5.

An anti-shoplifting mirror *example 5*

▷ Convenience stores often install a diverging mirror so that the clerk has a view of the whole store and can catch shoplifters. Use a ray diagram to show that the image is reduced, bringing more into the clerk's field of view. If the focal length of the mirror is 3.0 m, and the mirror is 7.0 m from the farthest wall, how deep is the image of the store?

▷ As shown in ray diagram g/1, d_i is less than d_o. The magnification, $M = d_i/d_o$, will be less than one, i.e., the image is actually reduced rather than magnified.

Apply the method outlined above for determining the plus and minus signs. Step 1: The object is the point on the opposite wall. As an experiment, g/2, move the object closer. I did these drawings using illustration software, but if you were doing them by hand, you'd want to make the scale much larger for greater accuracy. Also, although I split figure g into two separate drawings in order to make them easier to understand, you're less likely to make a mistake if you do them on top of each other.

The two angles at the mirror fan out from the normal. Increasing θ_o has