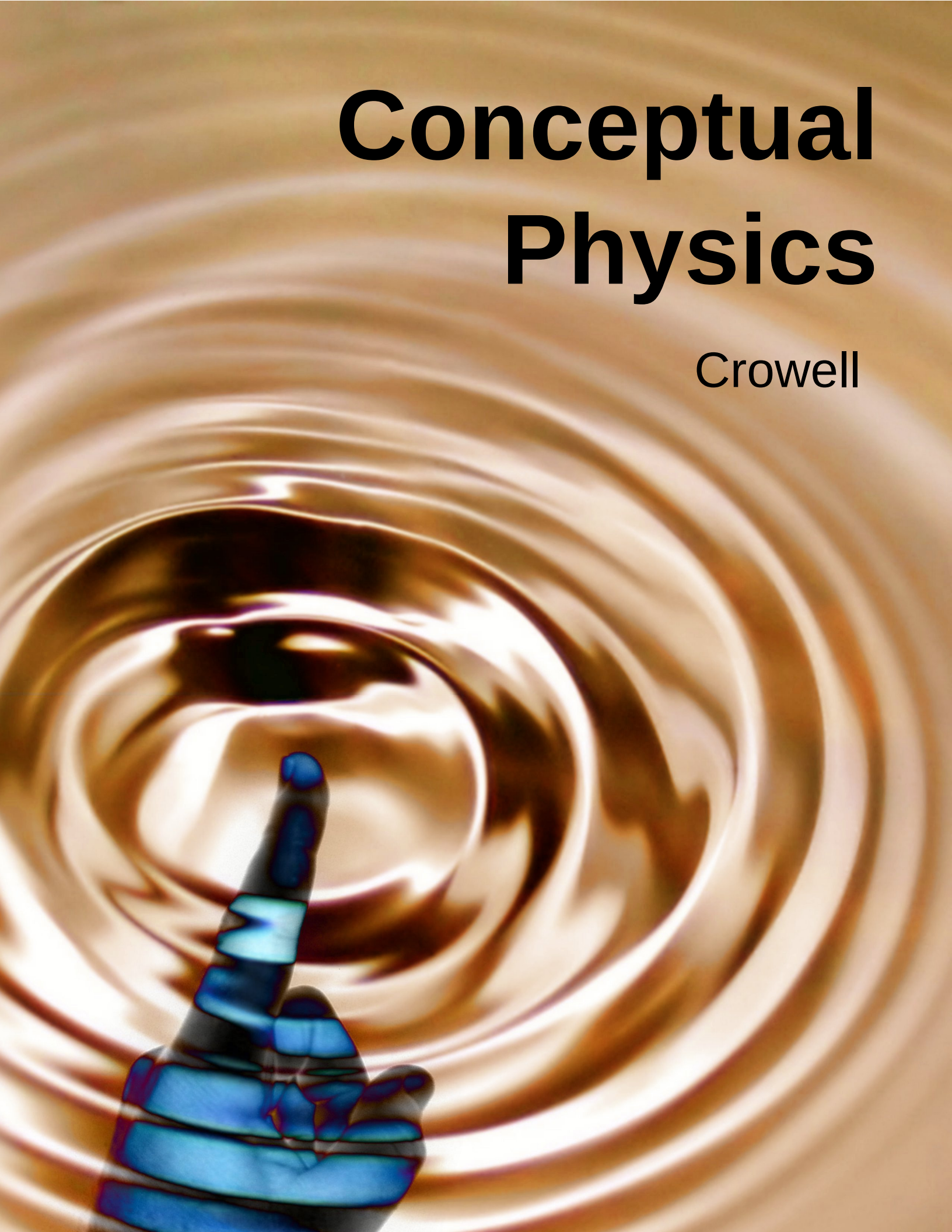


# Conceptual Physics

Crowell



copyright 2006 Benjamin Crowell

rev. February 22, 2017



This book is licensed under the Creative Commons Attribution-ShareAlike license, version 3.0, <http://creativecommons.org/licenses/by-sa/3.0/>, except for those photographs and drawings of which I am not the author, as listed in the photo credits. If you agree to the license, it grants you certain privileges that you would not otherwise have, such as the right to copy the book, or download the digital version free of charge from [www.lightandmatter.com](http://www.lightandmatter.com). At your option, you may also copy this book under the GNU Free Documentation License version 1.2, <http://www.gnu.org/licenses/fdl.txt>, with no invariant sections, no front-cover texts, and no back-cover texts.

## Brief Contents

1	Conservation of Mass and Energy	7
2	Conservation of Momentum	37
3	Conservation of Angular Momentum	61
4	Relativity	71
5	Electricity	93
6	Fields	111
7	The Ray Model of Light	131
8	Waves	161

For a semester-length course, all seven chapters can be covered. For a shorter course, the book is designed so that chapters 1, 2, and 5 are the only ones that are required for continuity; any of the others can be included or omitted at the instructor's discretion, with the only constraint being that chapter 6 requires chapter 4.

# Contents



## 1 Conservation of Mass and Energy

1.1 Symmetry and Conservation Laws . . . . .	7
1.2 Conservation of Mass . . . . .	9
1.3 Review of the Metric System and Conversions . . . . .	11
The metric system, 11.—Scientific notation, 12.—Conversions, 13.	
1.4 Conservation of Energy . . . . .	15
Energy, 15.—The principle of inertia, 16.—Kinetic and gravitational energy, 20.—Energy in general, 21.	
1.5 Newton's Law of Gravity . . . . .	25
1.6 Noether's Theorem for Energy . . . . .	27
1.7 Equivalence of Mass and Energy . . . . .	29
Mass-energy, 29.—The correspondence principle, 31.	
Problems . . . . .	33



## 2 Conservation of Momentum

2.1 Translation Symmetry . . . . .	38
2.2 The Principle of Inertia . . . . .	39
Symmetry and inertia, 39.	
2.3 Momentum . . . . .	40
Conservation of momentum, 40.—	

Momentum compared to kinetic energy, 45.—Force, 46.—Motion in two dimensions, 49.

2.4 Newton's Triumph . . . . .	53
2.5 Work . . . . .	57
Problems . . . . .	59

## 3 Conservation of Angular Momentum

3.1 Angular Momentum . . . . .	61
3.2 Torque . . . . .	64
Torque distinguished from force, 66.	
3.3 Noether's Theorem for Angular Momentum . . . . .	67
Problems . . . . .	68



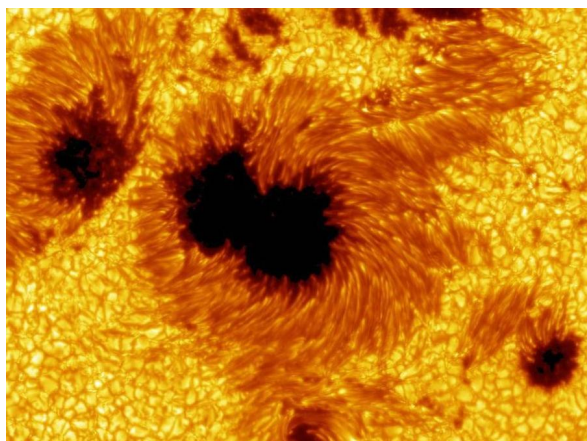
## 4 Relativity

4.1 Relativity According To Einstein . . . . .	71
Time is not absolute, 71.—Distortion of time and space, 73.—The $\gamma$ factor, 77.	
4.2 Speeds In Relativity . . . . .	80
The universal speed $c$ , 80.—Velocities don't simply add and subtract., 80.—A universal speed limit, 80.—Light travels at $c$ , 80.—The Michelson-Morley experiment, 81.	
4.3 Dynamics . . . . .	85
Momentum, 85.—Equivalence of mass and energy, 87.	
Problems . . . . .	89



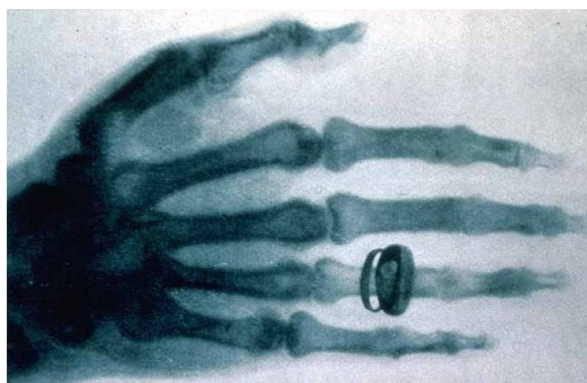
## 5 Electricity

5.1 The Quest for the Atomic Force . . .	94
5.2 Charge, Electricity and Magnetism . . .	95
Charge, 95.—Conservation of charge, 97.—	
Electrical forces involving neutral objects,	
97.—The atom, and subatomic particles,	
98.—Electric current, 98.	
5.3 Circuits . . . . .	100
5.4 Voltage . . . . .	101
The volt unit, 101.	
5.5 Resistance . . . . .	104
Applications, 105.	
Problems . . . . .	109



## 6 Fields

6.1 Farewell To the Mechanical Universe	111
Time delays in forces exerted at a distance, 112.—More evidence that fields of force are real: they carry energy., 113.—The gravitational field, 113.—Sources and sinks, 114.—The electric field, 115.	
6.2 Electromagnetism . . . . .	115
Magnetic interactions, 115.—Relativity requires magnetism, 116.—Magnetic fields, 119.	
6.3 Induction. . . . .	122
Electromagnetic waves, 125.	
Problems . . . . .	127



## 7 The Ray Model of Light

7.1 Light Rays . . . . .	131
The nature of light, 132.—Interaction of light with matter, 135.—The ray model of light, 136.—Geometry of specular reflection, 138.	
7.2 Applications . . . . .	140
The inverse-square law, 140.—Parallax, 142.	
7.3 The Principle of Least Time for Reflection . . . . .	146
7.4 Images by Reflection . . . . .	147
A virtual image, 147.—Curved mirrors, 148.—A real image, 149.—Images of images, 151.	
Problems . . . . .	156



## 8 Waves

8.1 Vibrations . . . . .	161
8.2 Wave Motion . . . . .	164
1. Superposition, 164.—2. The medium is not transported with the wave., 166.—3. A wave's velocity depends on the medium., 167.—Wave patterns, 168.	
8.3 Sound and Light Waves . . . . .	168
Sound waves, 169.—Light waves, 170.	
8.4 Periodic Waves . . . . .	170
Period and frequency of a periodic wave, 170.—Graphs of waves as a function of position, 171.—Wavelength, 172.—Wave velocity related to frequency and wavelength, 172.	
Problems . . . . .	175

Appendix 1: Photo Credits 177

Appendix 2: Hints and Solutions 179



# Chapter 1

## Conservation of Mass and Energy

### 1.1 Symmetry and Conservation Laws

Even before history began, people must already have noticed certain facts about the sky. The sun and moon both rise in the east and set in the west. Another fact that can be settled to a fair degree of accuracy using the naked eye is that the apparent sizes of the sun and moon don't change noticeably. (There is an optical illusion that makes the moon appear bigger when it's near the horizon, but you can easily verify that it's nothing more than an illusion by checking its angular size against some standard, such as your pinkie held at arm's length.) If the sun and moon were varying their distances from us, they would appear to get bigger and smaller, and since they don't appear to change in size, it appears, at least approximately, that they always stay at the same distance from us.

From observations like these, the ancients constructed a scientific *model*, in which the sun and moon traveled around the earth in perfect circles. Of course, we now know that the earth isn't the center of the universe, but that doesn't mean the model wasn't useful. That's the way science always works. Science never aims to reveal the ultimate reality. Science only tries to make models of reality that have predictive power.

Our modern approach to understanding physics revolves around the concepts of *symmetry* and *conservation laws*, both of which are demonstrated by this example.

The sun and moon were believed to move in circles, and a circle is a very symmetric shape. If you rotate a circle about its center, like a spinning wheel, it doesn't change. Therefore, we say that the circle is *symmetric* with respect to rotation about its center. The ancients thought it was beautiful that the universe seemed to have this type of symmetry built in, and they became very attached to the idea.

A *conservation law* is a statement that some number stays the same with the passage of time. In our example, the distance between the sun and the earth is conserved, and so is the distance between the moon and the earth. (The ancient Greeks were even able to



a / Due to the rotation of the earth, everything in the sky appears to spin in circles. In this time-exposure photograph, each star appears as a streak.

determine that earth-moon distance.)



b / Emmy Noether (1882-1935). The daughter of a prominent German mathematician, she did not show any early precocity at mathematics — as a teenager she was more interested in music and dancing. She received her doctorate in 1907 and rapidly built a world-wide reputation, but the University of Göttingen refused to let her teach, and her colleague Hilbert had to advertise her courses in the university’s catalog under his own name. A long controversy ensued, with her opponents asking what the country’s soldiers would think when they returned home and were expected to learn at the feet of a woman. Allowing her on the faculty would also mean letting her vote in the academic senate. Said Hilbert, “I do not see that the sex of the candidate is against her admission as a privatdozent [instructor]. After all, the university senate is not a bathhouse.” She was finally admitted to the faculty in 1919. A Jew, Noether fled Germany in 1933 and joined the faculty at Bryn Mawr in the U.S.

In our example, the symmetry and the conservation law both give the same information. Either statement can be satisfied only by a circular orbit. That isn’t a coincidence. Physicist Emmy Noether showed on very general mathematical grounds that for physical theories of a certain type, every symmetry leads to a corresponding conservation law. Although the precise formulation of Noether’s theorem, and its proof, are too mathematical for this book, we’ll see many examples like this one, in which the physical content of the theorem is fairly straightforward.



c / In this scene from *Swan Lake*, the choreography has a symmetry with respect to left and right.

The idea of perfect circular orbits seems very beautiful and intuitively appealing. It came as a great disappointment, therefore, when the astronomer Johannes Kepler discovered, by the painstaking analysis of precise observations, that orbits such as the moon’s were actually ellipses, not circles. This is the sort of thing that led the biologist Huxley to say, “The great tragedy of science is the slaying of a beautiful theory by an ugly fact.” The lesson of the story, then, is that symmetries are important and beautiful, but we can’t decide which symmetries are right based only on common sense or aesthetics; their validity has to be determined based on observations and experiments.



d / C.S. Wu at Columbia University in 1963.

As a more modern example, consider the symmetry between right and left. For example, we observe that a top spinning clockwise has exactly the same behavior as a top spinning counterclockwise. This kind of observation led physicists to believe, for hundreds of years, that the laws of physics were perfectly symmetric with respect to right and left. This mirror symmetry appealed to physicists’ common sense. However, experiments by Chien-Shiung Wu et al. in 1957 showed that right-left symmetry was violated in certain types of nuclear reactions. Physicists were thus forced to change their opinions about what constituted common sense.



## 1.2 Conservation of Mass

We intuitively feel that matter shouldn't appear or disappear out of nowhere: that the amount of matter should be a conserved quantity. If that was to happen, then it seems as though atoms would have to be created or destroyed, which doesn't happen in any physical processes that are familiar from everyday life, such as chemical reactions. On the other hand, I've already cautioned you against believing that a law of physics must be true just because it seems appealing. The laws of physics have to be found by experiment, and there seem to be experiments that are exceptions to the conservation of matter. A log weighs more than its ashes. Did some matter simply disappear when the log was burned?

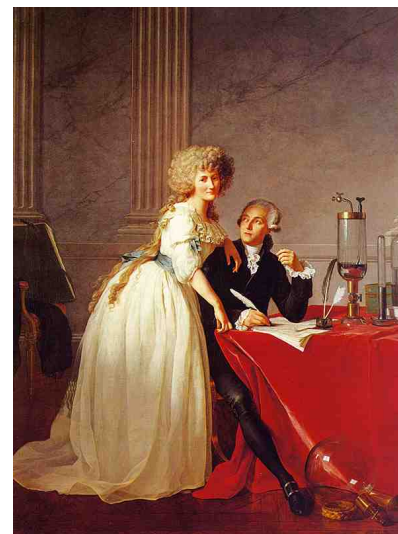
The French chemist Antoine-Laurent Lavoisier was the first scientist to realize that there were no such exceptions. Lavoisier hypothesized that when wood burns, for example, the supposed loss of weight is actually accounted for by the escaping hot gases that the flames are made of. Before Lavoisier, chemists had almost never weighed their chemicals to quantify the amount of each substance that was undergoing reactions. They also didn't completely understand that gases were just another state of matter, and hadn't tried performing reactions in sealed chambers to determine whether gases were being consumed from or released into the air. For this they had at least one practical excuse, which is that if you perform a gas-releasing reaction in a sealed chamber with no room for expansion, you get an explosion! Lavoisier invented a balance that was capable of measuring milligram masses, and figured out how to do reactions in an upside-down bowl in a basin of water, so that the gases could expand by pushing out some of the water. In one crucial experiment, Lavoisier heated a red mercury compound, which we would now describe as mercury oxide ( $\text{HgO}$ ), in such a sealed chamber. A gas was produced (Lavoisier later named it "oxygen"), driving out some of the water, and the red compound was transformed into silvery liquid mercury metal. The crucial point was that the total mass of the entire apparatus was exactly the same before and after the reaction. Based on many observations of this type, Lavoisier proposed a general law of nature, that matter is always conserved.

### *self-check A*

In ordinary speech, we say that you should "conserve" something, because if you don't, pretty soon it will all be gone. How is this different from the meaning of the term "conservation" in physics?   ▷ Answer,

p. 179

Although Lavoisier was an honest and energetic public official, he was caught up in the Terror and sentenced to death in 1794. He requested a fifteen-day delay of his execution so that he could complete some experiments that he thought might be of value to the Republic. The judge, Coffinhal, infamously replied that "the state



Portrait of Monsieur Lavoisier and His Wife, by Jacques-Louis David, 1788. Lavoisier invented the concept of conservation of mass. The husband is depicted with his scientific apparatus, while in the background on the left is the portfolio belonging to Madame Lavoisier, who is thought to have been a student of David's.



f / Example 1.

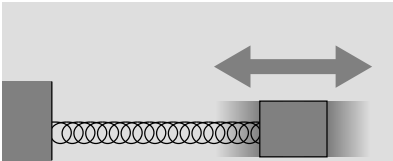
has no need of scientists.” As a scientific experiment, Lavoisier decided to try to determine how long his consciousness would continue after he was guillotined, by blinking his eyes for as long as possible. He blinked twelve times after his head was chopped off. Ironically, Judge Coffinhal was himself executed only three months later, falling victim to the same chaos.

*A stream of water*

*example 1*

The stream of water is fatter near the mouth of the faucet, and skinnier lower down. This can be understood using conservation of mass. Since water is being neither created nor destroyed, the mass of the water that leaves the faucet in one second must be the same as the amount that flows past a lower point in the same time interval. The water speeds up as it falls, so the two quantities of water can only be equal if the stream is narrower at the bottom.

Physicists are no different than plumbers or ballerinas in that they have a technical vocabulary that allows them to make precise distinctions. A pipe isn’t just a pipe, it’s a PVC pipe. A jump isn’t just a jump, it’s a grand jeté. We need to be more precise now about what we really mean by “the amount of matter,” which is what we’re saying is conserved. Since physics is a mathematical science, definitions in physics are usually definitions of numbers, and we define these numbers *operationally*. An operational definition is one that spells out the steps required in order to *measure* that quantity. For example, one way that an electrician knows that current and voltage are two different things is that she knows she has to do completely different things in order to measure them with a meter.



g / The time for one cycle of vibration is related to the object’s mass.

If you ask a room full of ordinary people to define what is meant by mass, they’ll probably propose a bunch of different, fuzzy ideas, and speak as if they all pretty much meant the same thing: “how much space it takes up,” “how much it weighs,” “how much matter is in it.” Of these, the first two can be disposed of easily. If we were to define mass as a measure of how much space an object occupied, then mass wouldn’t be conserved when we squished a piece of foam rubber. Although Lavoisier did use weight in his experiments, weight also won’t quite work as the ultimate, rigorous definition, because weight is a measure of how hard gravity pulls on an object, and gravity varies in strength from place to place. Gravity is measurably weaker on the top of a mountain than at sea level, and much weaker on the moon. The reason this didn’t matter to Lavoisier was that he was doing all his experiments in one location. The third proposal is better, but how exactly should we define “how much matter?” To make it into an operational definition, we could do something like figure g. A larger mass is harder to whip back and forth — it’s harder to set into motion, and harder to stop once it’s started. For this reason, the vibration of the mass on the spring will take a longer time if the mass is greater. If we put two different



h / Astronaut Tamara Jernigan measures her mass aboard the Space Shuttle. She is strapped into a chair attached to a spring, like the mass in figure g. (NASA)

masses on the spring, and they both take the same time to complete one oscillation, we can define them as having the same mass.

Since I started this chapter by highlighting the relationship between conservation laws and symmetries, you're probably wondering what symmetry is related to conservation of mass. I'll come back to that at the end of the chapter.

When you learn about a new physical quantity, such as mass, you need to know what units are used to measure it. This will lead us to a brief digression on the metric system, after which we'll come back to physics.

## 1.3 Review of the Metric System and Conversions

### The metric system

Every country in the world besides the U.S. has adopted a system of units known colloquially as the “metric system.” Even in the U.S., the system is used universally by scientists, and also by many engineers. This system is entirely decimal, thanks to the same eminently logical people who brought about the French Revolution. In deference to France, the system's official name is the *Système International*, or SI, meaning International System. (The phrase “SI system” is therefore redundant.)

The metric system works with a single, consistent set of prefixes (derived from Greek) that modify the basic units. Each prefix stands for a power of ten, and has an abbreviation that can be combined with the symbol for the unit. For instance, the meter is a unit of distance. The prefix kilo- stands for 1000, so a kilometer, 1 km, is a thousand meters.

In this book, we'll be using a flavor of the metric system, the SI, in which there are three basic units, measuring distance, time, and mass. The basic unit of distance is the meter (m), the one for time is the second (s), and for mass the kilogram (kg). Based on these units, we can define others, e.g., m/s (meters per second) for the speed of a car, or kg/s for the rate at which water flows through a pipe. It might seem odd that we consider the basic unit of mass to be the kilogram, rather than the gram. The reason for doing this is that when we start defining other units starting from the basic three, some of them come out to be a more convenient size for use in everyday life. For example, there is a metric unit of force, the newton (N), which is defined as the push or pull that would be able to change a 1-kg object's velocity by 1 m/s, if it acted on it for 1 s. A newton turns out to be about the amount of force you'd use to pick up your keys. If the system had been based on the gram instead of the kilogram, then the newton would have been a thousand times

smaller, something like the amount of force required in order to pick up a breadcrumb.

The following are the most common metric prefixes. You should memorize them.

prefix	meaning	example
kilo- k	1000	60 kg = a person's mass
centi- c	1/100	28 cm = height of a piece of paper
milli- m	1/1000	1 ms = time for one vibration of a guitar string playing the note D

The prefix centi-, meaning 1/100, is only used in the centimeter; a hundredth of a gram would not be written as 1 cg but as 10 mg. The centi- prefix can be easily remembered because a cent is 1/100 of a dollar. The official SI abbreviation for seconds is "s" (not "sec") and grams are "g" (not "gm").

You may also encounter the prefixes mega- (a million) and micro- (one millionth).

### Scientific notation

Most of the interesting phenomena in our universe are not on the human scale. It would take about 1,000,000,000,000,000,000 bacteria to equal the mass of a human body. When the physicist Thomas Young discovered that light was a wave, scientific notation hadn't been invented, and he was obliged to write that the time required for one vibration of the wave was 1/500 of a millionth of a millionth of a second. Scientific notation is a less awkward way to write very large and very small numbers such as these. Here's a quick review.

Scientific notation means writing a number in terms of a product of something from 1 to 10 and something else that is a power of ten. For instance,

$$\begin{aligned} 32 &= 3.2 \times 10^1 \\ 320 &= 3.2 \times 10^2 \\ 3200 &= 3.2 \times 10^3 \quad \dots \end{aligned}$$

Each number is ten times bigger than the last.

Since  $10^1$  is ten times smaller than  $10^2$ , it makes sense to use the notation  $10^0$  to stand for one, the number that is in turn ten times smaller than  $10^1$ . Continuing on, we can write  $10^{-1}$  to stand for 0.1, the number ten times smaller than  $10^0$ . Negative exponents are used for small numbers:

$$\begin{aligned} 3.2 &= 3.2 \times 10^0 \\ 0.32 &= 3.2 \times 10^{-1} \\ 0.032 &= 3.2 \times 10^{-2} \quad \dots \end{aligned}$$



A common source of confusion is the notation used on the displays of many calculators. Examples:

$3.2 \times 10^6$  (written notation)  
 $3.2\text{E}+6$  (notation on some calculators)  
 $3.2^6$  (notation on some other calculators)

The last example is particularly unfortunate, because  $3.2^6$  really stands for the number  $3.2 \times 3.2 \times 3.2 \times 3.2 \times 3.2 \times 3.2 = 1074$ , a totally different number from  $3.2 \times 10^6 = 3200000$ . The calculator notation should never be used in writing. It's just a way for the manufacturer to save money by making a simpler display.

*self-check B*

A student learns that  $10^4$  bacteria, standing in line to register for classes at Paramecium Community College, would form a queue of this size:

\_\_\_\_\_

The student concludes that  $10^2$  bacteria would form a line of this length:

\_\_\_\_\_

Why is the student incorrect?

▷ Answer, p. 179

## Conversions

I suggest you avoid memorizing lots of conversion factors between SI units and U.S. units. Suppose the United Nations sends its black helicopters to invade California (after all who wouldn't rather live here than in New York City?), and institutes water fluoridation and the SI, making the use of inches and pounds into a crime punishable by death. I think you could get by with only two mental conversion factors:

$$1 \text{ inch} = 2.54 \text{ cm}$$

An object with a weight on Earth of 2.2 pounds-force has a mass of 1 kg.

The first one is the present definition of the inch, so it's exact. The second one is not exact, but is good enough for most purposes. (U.S. units of force and mass are confusing, so it's a good thing they're not used in science. In U.S. units, the unit of force is the pound-force, and the best unit to use for mass is the slug, which is about 14.6 kg.)

More important than memorizing conversion factors is understanding the right method for doing conversions. Even within the SI, you may need to convert, say, from grams to kilograms. Different people have different ways of thinking about conversions, but the method I'll describe here is systematic and easy to understand. The idea is that if 1 kg and 1000 g represent the same mass, then

we can consider a fraction like

$$\frac{10^3 \text{ g}}{1 \text{ kg}}$$

to be a way of expressing the number one. This may bother you. For instance, if you type 1000/1 into your calculator, you will get 1000, not one. Again, different people have different ways of thinking about it, but the justification is that it helps us to do conversions, and it works! Now if we want to convert 0.7 kg to units of grams, we can multiply kg by the number one:

$$0.7 \text{ kg} \times \frac{10^3 \text{ g}}{1 \text{ kg}}$$

If you're willing to treat symbols such as "kg" as if they were variables as used in algebra (which they're really not), you can then cancel the kg on top with the kg on the bottom, resulting in

$$0.7 \cancel{\text{kg}} \times \frac{10^3 \text{ g}}{1 \cancel{\text{kg}}} = 700 \text{ g.}$$

To convert grams to kilograms, you would simply flip the fraction upside down.

One advantage of this method is that it can easily be applied to a series of conversions. For instance, to convert one year to units of seconds,

$$\begin{aligned} 1 \cancel{\text{year}} \times \frac{365 \cancel{\text{days}}}{1 \cancel{\text{year}}} \times \frac{24 \cancel{\text{hours}}}{1 \cancel{\text{day}}} \times \frac{60 \cancel{\text{min}}}{1 \cancel{\text{hour}}} \times \frac{60 \text{ s}}{1 \cancel{\text{min}}} &= \\ &= 3.15 \times 10^7 \text{ s.} \end{aligned}$$

*Should that exponent be positive or negative?*

A common mistake is to write the conversion fraction incorrectly. For instance the fraction

$$\frac{10^3 \text{ kg}}{1 \text{ g}} \quad (\text{incorrect})$$

does not equal one, because  $10^3$  kg is the mass of a car, and 1 g is the mass of a raisin. One correct way of setting up the conversion factor would be

$$\frac{10^{-3} \text{ kg}}{1 \text{ g}} \quad (\text{correct}).$$

You can usually detect such a mistake if you take the time to check your answer and see if it is reasonable.

If common sense doesn't rule out either a positive or a negative exponent, here's another way to make sure you get it right. There are big prefixes, like kilo-, and small ones, like milli-. In the example above, we want the top of the fraction to be the same as the bottom. Since  $k$  is a big prefix, we need to *compensate* by putting a small number like  $10^{-3}$  in front of it, not a big number like  $10^3$ .

### Discussion question

**A** Each of the following conversions contains an error. In each case, explain what the error is.

(a)  $1000 \text{ kg} \times \frac{1 \text{ kg}}{1000 \text{ g}} = 1 \text{ g}$

(b)  $50 \text{ m} \times \frac{1 \text{ cm}}{100 \text{ m}} = 0.5 \text{ cm}$

## 1.4 Conservation of Energy

### Energy

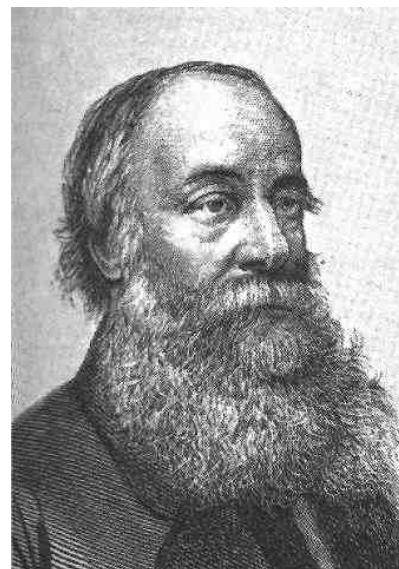
Consider the hockey puck in figure i. If we release it at rest, we expect it to remain at rest. If it did start moving all by itself, that would be strange: it would have to pick some direction in which to move, and why would it pick that direction rather than some other one? If we observed such a phenomenon, we would have to conclude that that direction in space was somehow special. It would be the favored direction in which hockey pucks (and presumably other objects as well) preferred to move. That would violate our intuition about the symmetry of space, and this is a case where our intuition is right: a vast number of experiments have all shown that that symmetry is a correct one. In other words, if you secretly pick up the physics laboratory with a crane, and spin it around gently with all the physicists inside, all their experiments will still come out the same, regardless of the lab's new orientation. If they don't have windows they can look out of, or any other external cues (like the Earth's magnetic field), then they won't notice anything until they hang up their lab coats for the evening and walk out into the parking lot.

Another way of thinking about it is that a moving hockey puck would have some *energy*, whereas a stationary one has none. I haven't given you an operational definition of energy yet, but we'll gradually start to build one up, and it will end up fitting in pretty well with your general idea of what energy means from everyday life. Regardless of the mathematical details of how you would actually calculate the energy of a moving hockey puck, it makes sense that a puck at rest has zero energy. It starts to look like energy is conserved. A puck that initially has zero energy must continue to have zero energy, so it can't start moving all by itself.

You might conclude from this discussion that we have a new example of Noether's theorem: that the symmetry of space with respect to different directions must be equivalent, in some mysterious way, to conservation of energy. Actually that's not quite right, and the possible confusion is related to the fact that we're not going to deal with the full, precise mathematical statement of Noether's theorem. In fact, we'll see soon that conservation of energy is really more closely related to a different symmetry, which is symmetry



i / A hockey puck is released at rest. If it spontaneously scooted off in some direction, that would violate the symmetry of all directions in space.



j / James Joule (1818-1889) discovered the law of conservation of energy.

with respect to the passage of time.

### The principle of inertia

Now there's one very subtle thing about the example of the hockey puck, which wouldn't occur to most people. If we stand on the ice and watch the puck, and we don't see it moving, does that mean that it really is at rest in some absolute sense? Remember, the planet earth spins once on its axis every 24 hours. At the latitude where I live, this results in a speed of about 800 miles per hour, or something like 400 meters per second. We could say, then that the puck wasn't really staying at rest. We could say that it was really in motion at a speed of 400 m/s, and remained in motion at that same speed. This may be inconsistent with our earlier description, but it is still consistent with the same description of the laws of physics. Again, we don't need to know the relevant formula for energy in order to believe that if the puck keeps the same speed (and its mass also stays the same), it's maintaining the same energy.

In other words, we have two different *frames of reference*, both equally valid. The person standing on the ice measures all velocities relative to the ice, finds that the puck maintained a velocity of zero, and says that energy was conserved. The astronaut watching the scene from deep space might measure the velocities relative to her own space station; in her frame of reference, the puck is moving at 400 m/s, but energy is still conserved.

This probably seems like common sense, but it wasn't common sense to one of the smartest people ever to live, the ancient Greek philosopher Aristotle. He came up with an entire system of physics based on the premise that there is one frame of reference that is special: the frame of reference defined by the dirt under our feet. He believed that all motion had a tendency to slow down unless a force was present to maintain it. Today, we know that Aristotle was wrong. One thing he was missing was that he didn't understand the concept of friction as a force. If you kick a soccer ball, the reason it eventually comes to rest on the grass isn't that it "naturally" wants to stop moving. The reason is that there's a frictional force from the grass that is slowing it down. (The energy of the ball's motion is transformed into other forms, such as heat and sound.) Modern people may also have an easier time seeing his mistake, because we have experience with smooth motion at high speeds. For instance, consider a passenger on a jet plane who stands up in the aisle and inadvertently drops his bag of peanuts. According to Aristotle, the bag would naturally slow to a stop, so it would become a life-threatening projectile in the cabin! From the modern point of view, the cabin can just as well be considered to be at rest.



k / Why does Aristotle look so sad? Is it because he's realized that his entire system of physics is wrong?



l / The jets are at rest. The Empire State Building is moving.



m / Galileo Galilei was the first physicist to state the principle of inertia (in a somewhat different formulation than the one given here). His contradiction of Aristotle had serious consequences. He was interrogated by the Church authorities and convicted of teaching that the earth went around the sun as a matter of fact and not, as he had promised previously, as a mere mathematical hypothesis. He was placed under permanent house arrest, and forbidden to write about or teach his theories. Immediately after being forced to recant his claim that the earth revolved around the sun, the old man is said to have muttered defiantly “and yet it does move.”



The *principle of inertia* says, roughly, that all frames of reference are equally valid:

### The principle of inertia

The results of experiments don't depend on the straight-line, constant-speed motion of the apparatus.

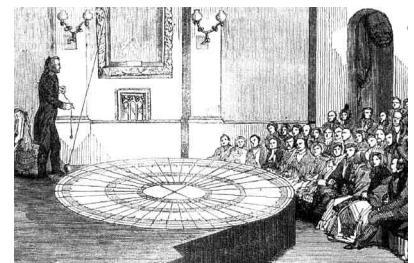
Speaking slightly more precisely, the principle of inertia says that if frame B moves at constant speed, in a straight line, relative to frame A, then frame B is just as valid as frame A, and in fact an observer in frame B will consider B to be at rest, and A to be moving. The laws of physics will be valid in both frames. The necessity for the more precise formulation becomes evident if you think about examples in which the motion changes its speed or direction. For instance, if you're in a car that's accelerating from rest, you feel yourself being pressed back into your seat. That's very different from the experience of being in a car cruising at constant speed, which produces no physical sensation at all. A more extreme example of this is shown in figure n on page 18.

A frame of reference moving at constant speed in a straight line is known as an inertial frame of reference. A frame that changes its speed or direction of motion is called noninertial. The principle of inertia applies only to inertial frames. The frame of reference defined by an accelerating car is noninertial, but the one defined by a car cruising at constant speed in a straight line is inertial.

### Foucault's pendulum

### example 2

Earlier, I spoke as if a frame of reference attached to the surface of the rotating earth was just as good as any other frame of reference. Now, with the more exact formulation of the principle of inertia, we can see that that isn't quite true. A point on the earth's surface moves in a circle, whereas the principle of inertia refers only to motion in a straight line. However, the curve of the motion is so gentle that under ordinary conditions we don't notice that the local dirt's frame of reference isn't quite inertial. The first demonstration of the noninertial nature of the earth-fixed frame of reference was by Léon Foucault using a very massive pendulum



o / Foucault demonstrates his pendulum to an audience at a lecture in 1851.



n / This Air Force doctor volunteered to ride a rocket sled as a medical experiment. The obvious effects on his head and face are not because of the sled's speed but because of its rapid changes in speed: increasing in 2 and 3, and decreasing in 5 and 6. In 4 his speed is greatest, but because his speed is not increasing or decreasing very much at this moment, there is little effect on him.

(figure o) whose oscillations would persist for many hours without becoming imperceptible. Although Foucault did his demonstration in Paris, it's easier to imagine what would happen at the north pole: the pendulum would keep swinging in the same plane, but the earth would spin underneath it once every 24 hours. To someone standing in the snow, it would appear that the pendulum's plane of motion was twisting. The effect at latitudes less than 90 degrees turns out to be slower, but otherwise similar. The Foucault pendulum was the first definitive experimental proof that the earth really did spin on its axis, although scientists had been convinced of its rotation for a century based on more indirect evidence about the structure of the solar system.

People have a strong intuitive belief that there is a state of absolute rest, and that the earth's surface defines it. But Copernicus proposed as a mathematical assumption, and Galileo argued as a matter of physical reality, that the earth spins on its axis, and also circles the sun. Galileo's opponents objected that this was impossible, because we would observe the effects of the motion. They said, for example, that if the earth was moving, then you would never be able to jump up in the air and land in the same place again — the earth would have moved out from under you. Galileo realized

that this wasn't really an argument about the earth's motion but about physics. In one of his books, which were written in the form of dialogues, he has the three characters debate what would happen if a ship was cruising smoothly across a calm harbor and a sailor climbed up to the top of its mast and dropped a rock. Would it hit the deck at the base of the mast, or behind it because the ship had moved out from under it? This is the kind of experiment referred to in the principle of inertia, and Galileo knew that it would come out the same regardless of the ship's motion. His opponents' reasoning, as represented by the dialog's stupid character Simplicio, was based on the assumption that once the rock lost contact with the sailor's hand, it would naturally start to lose its forward motion. In other words, they didn't even believe in the idea that motion naturally continues unless a force acts to stop it.

But the principle of inertia says more than that. It says that motion isn't even real: to a sailor standing on the deck of the ship, the deck and the masts and the rigging are not even moving. People on the shore can tell him that the ship and his own body are moving in a straight line at constant speed. He can reply, "No, that's an illusion. I'm at rest. The only reason you think I'm moving is because you and the sand and the water are moving in the opposite direction." The principle of inertia says that straight-line, constant-speed motion is a matter of opinion. Thus things can't "naturally" slow down and stop moving, because we can't even agree on which things are moving and which are at rest.

If observers in different frames of reference disagree on velocities, it's natural to want to be able to convert back and forth. For motion in one dimension, this can be done by simple addition.

---

*A sailor running on the deck* *example 3*

▷ A sailor is running toward the front of a ship, and the other sailors say that in their frame of reference, fixed to the deck, his velocity is 7.0 m/s. The ship is moving at 1.3 m/s relative to the shore. How fast does an observer on the beach say the sailor is moving?

▷ They see the ship moving at 7.0 m/s, and the sailor moving even faster than that because he's running from the stern to the bow. In one second, the ship moves 1.3 meters, but he moves 1.3+7.0 m, so his velocity relative to the beach is 8.3 m/s.

The only way to make this rule give consistent results is if we define velocities in one direction as positive, and velocities in the opposite direction as negative.

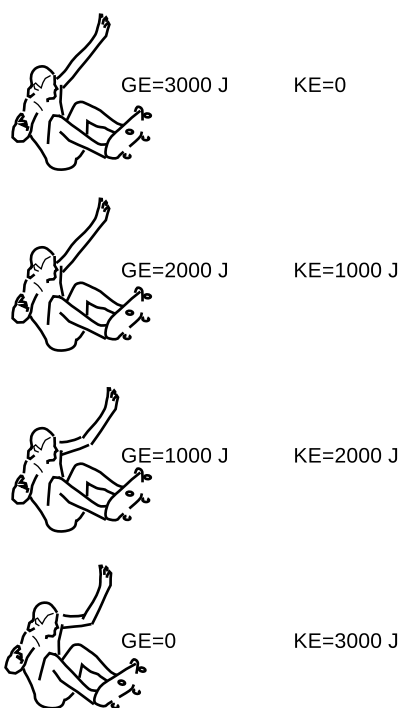
---

*Running back toward the stern* *example 4*

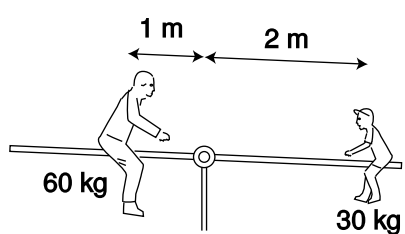
▷ The sailor of example 3 turns around and runs back toward the stern at the same speed relative to the deck. How do the other sailors describe this velocity mathematically, and what do



p / The skater has converted all his kinetic energy into gravitational energy on the way up the side of the pool. Photo by J.D. Rogge, www.sonic.net/~shawn.



q / As the skater free-falls, his gravitational energy is converted into kinetic energy.



r / Example 5.

observers on the beach say?

▷ Since the other sailors described his original velocity as positive, they have to call this negative. They say his velocity is now  $-7.0$  m/s. A person on the shore says his velocity is  $1.3 + (-7.0) = -5.7$  m/s.

### Kinetic and gravitational energy

Now suppose we drop a rock. The rock is initially at rest, but then begins moving. This seems to be a violation of conservation of energy, because a moving rock would have more energy. But actually this is a little like the example of the burning log that seems to violate conservation of mass. Lavoisier realized that there was a second form of mass, the mass of the smoke, that wasn't being accounted for, and proved by experiments that mass *was*, after all, conserved once the second form had been taken into account. In the case of the falling rock, we have two forms of energy. The first is the energy it has because it's moving, known as *kinetic energy*. The second form is a kind of energy that it has because it's interacting with the planet earth via gravity. This is known as *gravitational energy*.<sup>1</sup> The earth and the rock attract each other gravitationally, and the greater the distance between them, the greater the gravitational energy — it's a little like stretching a spring.

The SI unit of energy is the joule (J), and in those units, we find that lifting a 1-kg mass through a height of 1 m requires 9.8 J of energy. This number, 9.8 joules per meter per kilogram, is a measure of the strength of the earth's gravity near its surface. We notate this number, known as the gravitational field, as  $g$ , and often round it off to 10 for convenience in rough calculations. If you lift a 1-kg rock to a height of 1 m above the ground, you're giving up 9.8 J of the energy you got from eating food, and changing it into gravitational energy stored in the rock. If you then release the rock, it starts transforming the energy into kinetic energy, until finally when the rock is just about to hit the ground, all of that energy is in the form of kinetic energy. That kinetic energy is then transformed into heat and sound when the rock hits the ground.

Stated in the language of algebra, the formula for gravitational energy is

$$GE = mgh,$$

where  $m$  is the mass of an object,  $g$  is the gravitational field, and  $h$  is the object's height.

A lever	example 5
Figure r shows two sisters on a seesaw. The one on the left has twice as much mass, but she's at half the distance from the center. No energy input is needed in order to tip the seesaw. If	

<sup>1</sup>You may also see this referred to in some books as gravitational potential energy.



the girl on the left goes up a certain distance, her gravitational energy will increase. At the same time, her sister on the right will drop twice the distance, which results in an equal decrease in energy, since her mass is half as much. In symbols, we have

$$(2m)gh$$

for the gravitational energy gained by the girl on the left, and

$$mg(2h)$$

for the energy lost by the one on the right. Both of these equal  $2mgh$ , so the amounts gained and lost are the same, and energy is conserved.

Looking at it another way, this can be thought of as an example of the kind of experiment that you'd have to do in order to arrive at the equation  $GE = mgh$  in the first place. If we didn't already know the equation, this experiment would make us suspect that it involved the product  $mh$ , since that's what's the same for both girls.

Once we have an equation for one form of energy, we can establish equations for other forms of energy. For example, if we drop a rock and measure its final velocity,  $v$ , when it hits the ground, we know how much GE it lost, so we know that's how much KE it must have had when it was at that final speed. Here are some imaginary results from such an experiment.

<b>m</b> (kg)	<b>v</b> (m/s)	<b>energy</b> (J)
1.00	1.00	0.50
1.00	2.00	2.00
2.00	1.00	1.00

Comparing the first line with the second, we see that doubling the object's velocity doesn't just double its energy, it quadruples it. If we compare the first and third lines, however, we find that doubling the mass only doubles the energy. This suggests that kinetic energy is proportional to mass times the square of velocity,  $mv^2$ , and further experiments of this type would indeed establish such a general rule. The proportionality factor equals 0.5 because of the design of the metric system, so the kinetic energy of a moving object is given by

$$KE = \frac{1}{2}mv^2.$$

### Energy in general

By this point, I've casually mentioned several forms of energy: kinetic, gravitational, heat, and sound. This might be disconcerting, since we can get thoroughly messed up if we don't realize that a



t / The spinning coin slows down. It looks like conservation of energy is violated, but it isn't.



s / A vivid demonstration that heat is a form of motion. A small amount of boiling water is poured into the empty can, which rapidly fills up with hot steam. The can is then sealed tightly, and soon crumples. This can be explained as follows. The high temperature of the steam is interpreted as a high average speed of random motions of its molecules. Before the lid was put on the can, the rapidly moving steam molecules pushed their way out of the can, forcing the slower air molecules out of the way. As the steam inside the can thinned out, a stable situation was soon achieved, in which the force from the less dense steam molecules moving at high speed balanced against the force from the more dense but slower air molecules outside. The cap was put on, and after a while the steam inside the can reached the same temperature as the air outside. The force from the cool, thin steam no longer matched the force from the cool, dense air outside, and the imbalance of forces crushed the can.

certain form of energy is important in a particular situation. For instance, the spinning coin in figure t gradually loses its kinetic energy, and we might think that conservation of energy was therefore being violated. However, whenever two surfaces rub together, friction acts to create heat. The correct analysis is that the coin's kinetic energy is gradually converted into heat.

One way of making the proliferation of forms of energy seem less scary is to realize that many forms of energy that seem different on the surface are in fact the same. One important example is that heat is actually the kinetic energy of molecules in random motion, so where we thought we had two forms of energy, in fact there is only one. Sound is also a form of kinetic energy: it's the vibration of air molecules.

This kind of unification of different types of energy has been a process that has been going on in physics for a long time, and at this point we've gotten it down to the point where there really only appear to be four forms of energy:

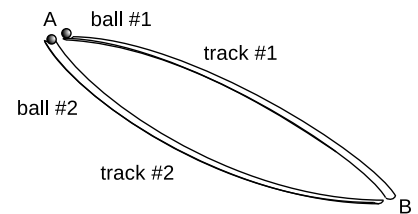
1. kinetic energy
2. gravitational energy
3. electrical energy
4. nuclear energy

We don't even encounter nuclear energy in everyday life (except in the sense that sunlight originates as nuclear energy), so really for most purposes the list only has three items on it. Of these three, electrical energy is the only form that we haven't talked about yet. The interactions between atoms are all electrical, so this form of energy is what's responsible for all of chemistry. The energy in the food you eat, or in a tank of gasoline, are forms of electrical energy.

*You take the high road and I'll take the low road.*      *example 6*

▷ Figure u shows two ramps which two balls will roll down. Compare their final speeds, when they reach point B. Assume friction is negligible.

▷ Each ball loses some gravitational energy because of its decreasing height above the earth, and conservation of energy says that it must gain an equal amount of kinetic energy (minus a little heat created by friction). The balls lose the same amount of height, so their final speeds must be equal.



u / Example 6.

*The birth of stars*      *example 7*

Orion is the easiest constellation to find. You can see it in the winter, even if you live under the light-polluted skies of a big city. Figure v shows an interesting feature of this part of the sky that you can easily pick out with an ordinary camera (that's how I took the picture) or a pair of binoculars. The three stars at the top are Orion's belt, and the stuff near the lower left corner of the picture is known as his sword — to the naked eye, it just looks like three more stars that aren't as bright as the stars in the belt. The middle "star" of the sword, however, isn't a star at all. It's a cloud of gas, known as the Orion Nebula, that's in the process of collapsing due to gravity. Like the pool skater on his way down, the gas is losing gravitational energy. The results are very different, however. The skateboard is designed to be a low-friction device, so nearly all of the lost gravitational energy is converted to kinetic energy, and very little to heat. The gases in the nebula flow and rub against each other, however, so most of the gravitational energy is converted to heat. This is the process by which stars are born: eventually the core of the gas cloud gets hot enough to ignite nuclear reactions.



v / Example 7.

### Lifting a weight

example 8

▷ At the gym, you lift a mass of 40 kg through a height of 0.5 m. How much gravitational energy is required? Where does this energy come from?

▷ The strength of the gravitational field is 10 joules per kilogram per meter, so after you lift the weight, its gravitational energy will be greater by  $10 \times 40 \times 0.5 = 200$  joules.

Energy is conserved, so if the weight gains gravitational energy, something else somewhere in the universe must have lost some. The energy that was used up was the energy in your body, which came from the food you'd eaten. This is what we refer to as "burning calories," since calories are the units normally used to describe the energy in food, rather than metric units of joules.

In fact, your body uses up even more than 200 J of food energy, because it's not very efficient. The rest of the energy goes into heat, which is why you'll need a shower after you work out. We can summarize this as

food energy  $\rightarrow$  gravitational energy + heat.



w / Example 10.



x / Example 10.

### Lowering a weight

example 9

▷ After lifting the weight, you need to lower it again. What's happening in terms of energy?

▷ Your body isn't capable of accepting the energy and putting it back into storage. The gravitational energy all goes into heat. (There's nothing fundamental in the laws of physics that forbids this. Electric cars can do it — when you stop at a stop sign, the car's kinetic energy is absorbed back into the battery, through a generator.)

### Absorption and emission of light

example 10

Light has energy. Light can be absorbed by matter and transformed into heat, but the reverse is also possible: an object can glow, transforming some of its heat energy into light. Very hot objects, like a candle flame or a welding torch, will glow in the visible part of the spectrum, as in figure w.

Objects at lower temperatures will also emit light, but in the infrared part of the spectrum, i.e., the part of the rainbow lying beyond the red end, which humans can't see. The photos in figure x were taken using a camera that is sensitive to infrared light. The cyclist locked his rear brakes suddenly, and skidded to a stop. The kinetic energy of the bike and his body are rapidly transformed into heat by the friction between the tire and the floor. In the first panel, you can see the glow of the heated strip on the floor, and in the second panel, the heated part of the tire.

*Heavy objects don't fall faster* *example 11*

Stand up now, take off your shoe, and drop it alongside a much less massive object such as a coin or the cap from your pen.

Did that surprise you? You found that they both hit the ground at the same time. Aristotle wrote that heavier objects fall faster than lighter ones. He was wrong, but Europeans believed him for thousands of years, partly because experiments weren't an accepted way of learning the truth, and partly because the Catholic Church gave him its posthumous seal of approval as its official philosopher.

Heavy objects and light objects have to fall the same way, because conservation laws are additive — we find the total energy of an object by adding up the energies of all its atoms. If a single atom falls through a height of one meter, it loses a certain amount of gravitational energy and gains a corresponding amount of kinetic energy. Kinetic energy relates to speed, so that determines how fast it's moving at the end of its one-meter drop. (The same reasoning could be applied to any point along the way between zero meters and one.)

Now what if we stick two atoms together? The pair has double the mass, so the amount of gravitational energy transformed into kinetic energy is twice as much. But twice as much kinetic energy is exactly what we need if the pair of atoms is to have the same speed as the single atom did. Continuing this train of thought, it doesn't matter how many atoms an object contains; it will have the same speed as any other object after dropping through the same height.

## 1.5 Newton's Law of Gravity

Why does the gravitational field on our planet have the particular value it does? For insight, let's compare with the strength of gravity elsewhere in the universe:

location	$g$ (joules per kg per m)
asteroid Vesta (surface)	0.3
earth's moon (surface)	1.6
Mars (surface)	3.7
earth (surface)	9.8
Jupiter (cloud-tops)	26
sun (visible surface)	270
typical neutron star (surface)	$10^{12}$
black hole (center)	infinite according to some theories, on the order of $10^{52}$ according to others

A good comparison is Vesta versus a neutron star. They're roughly the same size, but they have vastly different masses — a



teaspoonful of neutron star matter would weigh a million tons! The different mass must be the reason for the vastly different gravitational fields. (The notation  $10^{12}$  means 1 followed by 12 zeroes.) This makes sense, because gravity is an attraction between things that have mass.

The mass of an object, however, isn't the only thing that determines the strength of its gravitational field, as demonstrated by the difference between the fields of the sun and a neutron star, despite their similar masses. The other variable that matters is distance. Because a neutron star's mass is compressed into such a small space (comparable to the size of a city), a point on its surface is within a fairly short distance from every part of the star. If you visited the surface of the sun, however, you'd be millions of miles away from most of its atoms.

As a less exotic example, if you travel from the seaport of Guayaquil, Ecuador, to the top of nearby Mt. Cotopaxi, you'll experience a slight reduction in gravity, from 9.7806 to 9.7624 J/kg/m. This is because you've gotten a little farther from the planet's mass. Such differences in the strength of gravity between one location and another on the earth's surface were first discovered because pendulum clocks that were correctly calibrated in one country were found to run too fast or too slow when they were shipped to another location.

The general equation for an object's gravitational field was discovered by Isaac Newton, by working backwards from the observed motion of the planets:<sup>2</sup>

$$g = \frac{GM}{d^2},$$

where  $M$  is the mass of the object,  $d$  is the distance from the object, and  $G$  is a constant that is the same everywhere in the universe. This is known as Newton's law of gravity.<sup>3</sup> This type of relationship, in which an effect is inversely proportional to the square of the distance from the object creating the effect, is known as an inverse square law. For example, the intensity of the light from a candle obeys an inverse square law, as discussed in subsection 7.2.1 on page 140.

*self-check C*

Mars is about twice as far from the sun as Venus. Compare the strength of the sun's gravitational field as experienced by Mars with the strength of the field felt by Venus. ▷ Answer, p. 179

Newton's law of gravity really gives the field of an individual atom, and the field of a many-atom object is the sum of the fields of the atoms. Newton was able to prove mathematically that this scary sum has an unexpectedly simple result in the case of a spherical object such as a planet: the result is the same as if all the object's



y / Isaac Newton (1642-1727)

<sup>2</sup>Example 12 on page 50 shows the type of reasoning that Newton had to go through.

<sup>3</sup>This is not the form in which Newton originally wrote the equation.

mass had been concentrated at its center.

Newton showed that his theory of gravity could explain the orbits of the planets, and also finished the project begun by Galileo of driving a stake through the heart of Aristotelian physics. His book on the motion of material objects, the *Mathematical Principles of Natural Philosophy*, was uncontradicted by experiment for 200 years, but his other main work, *Optics*, was on the wrong track due to his conviction that light was composed of particles rather than waves. He was an avid alchemist, an embarrassing fact that modern scientists would like to forget. Newton was on the winning side of the revolution that replaced King James II with William and Mary of Orange, which led to a lucrative post running the English royal mint; he worked hard at what could have been a sinecure, and took great satisfaction from catching and executing counterfeiters. Newton's personal life was less happy, as we'll see in chapter 5.

*Newton's apple*

*example 12*

A charming legend attested to by Newton's niece is that he first conceived of gravity as a universal attraction after seeing an apple fall from a tree. He wondered whether the force that made the apple fall was the same one that made the moon circle the earth rather than flying off straight. Newton had astronomical data that allowed him to calculate that the gravitational field the moon experienced from the earth was  $1/3600$  as strong as the field on the surface of the earth.<sup>4</sup> (The moon has its own gravitational field, but that's not what we're talking about.) The moon's distance from the earth is 60 times greater than the earth's radius, so this fit perfectly with an inverse-square law:  $60 \times 60 = 3600$ .

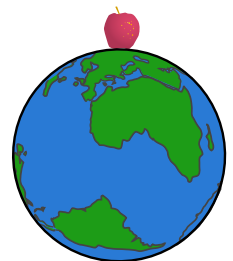
60

## 1.6 Noether's Theorem for Energy

Now we're ready for our first full-fledged example of Noether's theorem. Conservation of energy is a law of physics, and Noether's theorem says that the laws of physics come from symmetry. Specifically, Noether's theorem says that every symmetry implies a conservation law. Conservation of energy comes from a symmetry that we haven't even discussed yet, but one that is simple and intuitively appealing: as time goes by, the universe doesn't change the way it works. We'll call this time symmetry.

We have strong evidence for time symmetry, because when we see a distant galaxy through a telescope, we're seeing light that has taken billions of years to get here. A telescope, then, is like a time machine. For all we know, alien astronomers with advanced technology may be observing our planet right now,<sup>5</sup> but if so, they're

1



<sup>4</sup>See example 12 on page 50.

<sup>5</sup>Our present technology isn't good enough to let us pick the planets of other solar systems out from the glare of their suns, except in a few exceptional cases.

z / Example 12.

seeing it not as it is now but as it was in the distant past, perhaps in the age of the dinosaurs, or before life even evolved here. As we observe a particularly distant, and therefore ancient, supernova, we see that its explosion plays out in exactly the same way as those that are closer, and therefore more recent.

Now suppose physics really does change from year to year, like politics, pop music, and hemlines. Imagine, for example, that the “constant”  $G$  in Newton’s law of gravity isn’t quite so constant. One day you might wake up and find that you’ve lost a lot of weight without dieting or exercise, simply because gravity has gotten weaker since the day before.

If you know about such changes in  $G$  over time, it’s the ultimate insider information. You can use it to get as rich as Croesus, or even Bill Gates. On a day when  $G$  is low, you pay for the energy needed to lift a large mass up high. Then, on a day when gravity is stronger, you lower the mass back down, extracting its gravitational energy. The key is that the energy you get back out is greater than what you originally had to put in. You can run the cycle over and over again, always raising the weight when gravity is weak, and lowering it when gravity is strong. Each time, you make a profit in energy. Everyone else thinks energy is conserved, but your secret technique allows you to keep on increasing and increasing the amount of energy in the universe (and the amount of money in your bank account).

The scheme can be made to work if anything about physics changes over time, not just gravity. For instance, suppose that the mass of an electron had one value today, and a slightly different value tomorrow. Electrons are one of the basic particles from which atoms are built, so on a day when the mass of electrons is low, every physical object has a slightly lower mass. In problem 14 on page 35, you’ll work out a way that this could be used to manufacture energy out of nowhere.

Sorry, but it won’t work. Experiments show that  $G$  doesn’t change measurably over time, nor does there seem to be any time variation in any of the other rules by which the universe works.<sup>6</sup> If archaeologists find a copy of this book thousands of years from now, they’ll be able to reproduce all the experiments you’re doing in this course.

I’ve probably convinced you that if time symmetry was violated, then conservation of energy wouldn’t hold. But does it work the

---

<sup>6</sup>In 2002, there have been some reports that the properties of atoms as observed in distant galaxies are slightly different than those of atoms here and now. If so, then time symmetry is weakly violated, and so is conservation of energy. However, this is a revolutionary claim, and it needs to be examined carefully. The change being claimed is large enough that, if it’s real, it should be detectable from one year to the next in ultra-high-precision laboratory experiments here on earth.

other way around? If time symmetry is valid, must there be a law of conservation of energy? Logically, that's a different question. We may be able to prove that if A is false, then B must be false, but that doesn't mean that if A is true, B must be true as well. For instance, if you're not a criminal, then you're presumably not in jail, but just because someone is a criminal, that doesn't mean he is in jail — some criminals never get caught.

Noether's theorem does work the other way around as well: if physics has a certain symmetry, then there must be a certain corresponding conservation law. This is a stronger statement. The full-strength version of Noether's theorem can't be proved without a model of light and matter more detailed than the one currently at our disposal.

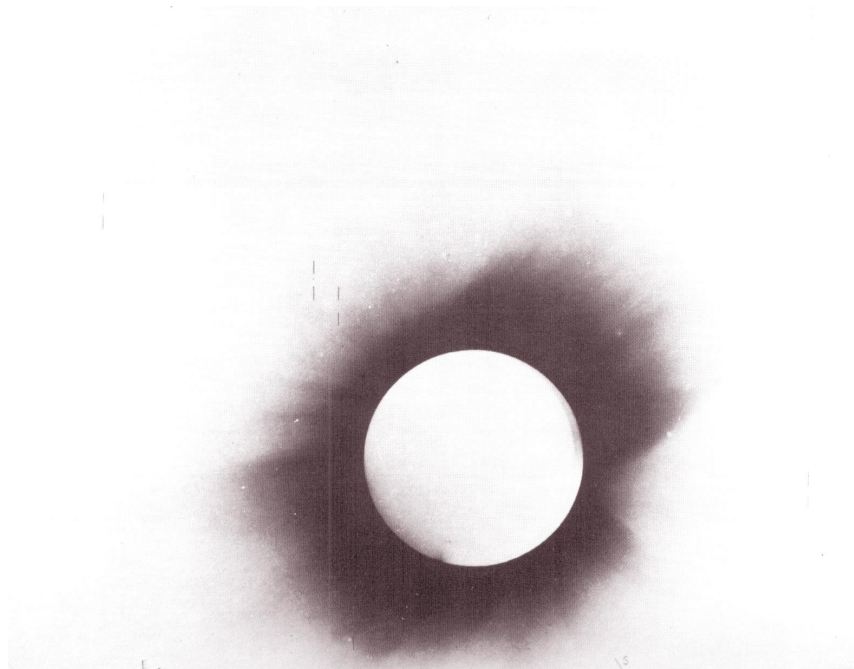
## 1.7 Equivalence of Mass and Energy

### Mass-energy

You've encountered two conservation laws so far: conservation of mass and conservation of energy. If conservation of energy is a consequence of symmetry, is there a deeper reason for conservation of mass?

Actually they're not even separate conservation laws. Albert Einstein found, as a consequence of his theory of relativity, that mass and energy are equivalent, and are not separately conserved — one can be converted into the other. Imagine that a magician waves his wand, and changes a bowl of dirt into a bowl of lettuce. You'd be impressed, because you were expecting that both dirt and lettuce would be conserved quantities. Neither one can be made to vanish, or to appear out of thin air. However, there are processes that can change one into the other. A farmer changes dirt into lettuce, and a compost heap changes lettuce into dirt. At the most fundamental level, lettuce and dirt aren't really different things at all; they're just collections of the same kinds of atoms — carbon, hydrogen, and so on.

We won't examine relativity in detail in this book, but mass-energy equivalence is an inevitable implication of the theory, and it's the only part of the theory that most people have heard of, via the famous equation  $E = mc^2$ . This equation tells us how much energy is equivalent to how much mass: the conversion factor is the square of the speed of light,  $c$ . Since  $c$  is a big number, you get a really really big number when you multiply it by itself to get  $c^2$ . This means that even a small amount of mass is equivalent to a very large amount of energy.



aa / Example 13.

## LIGHTS ALL ASKEW IN THE HEAVENS

**Men of Science More or Less  
Agog Over Results of Eclipse  
Observations.**

**EINSTEIN THEORY TRIUMPHS**

**Stars Not Where They Seemed  
or Were Calculated to be,  
but Nobody Need Worry.**

**A BOOK FOR 12 WISE MEN**

**No More in All the World Could  
Comprehend It, Said Einstein When  
His Daring Publishers Accepted It.**

ab / A New York Times headline from November 10, 1919, describing the observations discussed in example 13.

### *Gravity bending light*

*example 13*

Gravity is a universal attraction between things that have mass, and since the energy in a beam of light is equivalent to some very small amount of mass, we expect that light will be affected by gravity, although the effect should be very small. The first experimental confirmation of relativity came in 1919 when stars next to the sun during a solar eclipse were observed to have shifted a little from their ordinary position. (If there was no eclipse, the glare of the sun would prevent the stars from being observed.) Starlight had been deflected by the sun's gravity. Figure aa is a photographic negative, so the circle that appears bright is actually the dark face of the moon, and the dark area is really the bright corona of the sun. The stars, marked by lines above and below them, appeared at positions slightly different than their normal ones.

### *Black holes*

*example 14*

A star with sufficiently strong gravity can prevent light from leaving. Quite a few black holes have been detected via their gravitational forces on neighboring stars or clouds of gas and dust.

Because mass and energy are like two different sides of the same coin, we may speak of mass-energy, a single conserved quantity, found by adding up all the mass and energy, with the appropriate conversion factor:  $E + mc^2$ .



▷ An iron nail is left in a cup of water until it turns entirely to rust. The energy released is about 500,000 joules. In theory, would a sufficiently precise scale register a change in mass? If so, how much?

▷ The energy will appear as heat, which will be lost to the environment. The total mass-energy of the cup, water, and iron will indeed be lessened by 500,000 joules. (If it had been perfectly insulated, there would have been no change, since the heat energy would have been trapped in the cup.) The speed of light in metric units is  $c = 3 \times 10^8$  meters per second (scientific notation for 3 followed by 8 zeroes), so converting to mass units, we have

$$\begin{aligned} m &= \frac{E}{c^2} \\ &= \frac{500,000}{(3 \times 10^8)^2} \\ &= 0.000000000006 \text{ kilograms.} \end{aligned}$$

(The design of the metric system is based on the meter, the kilogram, and the second. The joule is designed to fit into this system, so the result comes out in units of kilograms.) The change in mass is too small to measure with any practical technique. This is because the square of the speed of light is such a large number in metric units.

### The correspondence principle

The realization that mass and energy are not separately conserved is our first example of a general idea called the correspondence principle. When Einstein came up with relativity, conservation of energy had been accepted by physicists for decades, and conservation of mass for over a hundred years.

Does an example like this mean that physicists don't know what they're talking about? There is a recent tendency among social scientists to deny that the scientific method even exists, claiming that science is no more than a social system that determines what ideas to accept based on an in-group's criteria. If science is an arbitrary social ritual, it would seem difficult to explain its effectiveness in building such useful items as airplanes, CD players and sewers. If voodoo and astrology were no less scientific in their methods than chemistry and physics, what was it that kept them from producing anything useful? This silly attitude was effectively skewered in a famous hoax carried out in 1996 by New York University physicist Alan Sokal. Sokal wrote an article titled "Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity," and got it accepted by a cultural studies journal called *Social Text*.<sup>7</sup>

<sup>7</sup>The paper appeared in *Social Text* #46/47 (1996) pp. 217-

The scientific content of the paper is a carefully constructed soup of mumbo jumbo, using technical terms to create maximum confusion; I can't make heads or tails of it, and I assume the editors and peer reviewers at *Social Text* understood even less. The physics, however, is mixed in with cultural relativist statements designed to appeal to them — “. . . the truth claims of science are inherently theory-laden and self-referential” — and footnoted references to academic articles such as “Irigaray's and Hayles' exegeses of gender encoding in fluid mechanics . . . and . . . Harding's comprehensive critique of the gender ideology underlying the natural sciences in general and physics in particular. . . .” On the day the article came out, Sokal published a letter explaining that the whole thing had been a parody — one that apparently went over the heads of the editors of *Social Text*.

What keeps physics from being merely a matter of fashion is that it has to agree with experiments and observations. If a theory such as conservation of mass or conservation of energy became accepted in physics, it was because it was supported by a vast number of experiments. It's just that experiments never have perfect accuracy, so a discrepancy such as the tiny change in the mass of the rusting nail in example 15 was undetectable. The old experiments weren't all wrong. They were right, within their limitations. If someone comes along with a new theory he claims is better, it must still be consistent with all the same experiments. In computer jargon, it must be backward-compatible. This is called the correspondence principle: new theories must be compatible with old ones in situations where they are both applicable. The correspondence principle tells us that we can still use an old theory within the realm where it works, so for instance I'll typically refer to conservation of mass and conservation of energy in this book rather than conservation of mass-energy, except in cases where the new theory is actually necessary.

Ironically, the extreme cultural relativists want to attack what they see as physical scientists' arrogant claims to absolute truth, but what they fail to understand is that science only claims to be able to find partial, provisional truth. The correspondence principle tells us that each of today's scientific truths can be superseded tomorrow by another truth that is more accurate and more broadly applicable. It also tells us that today's truth will not lose any value when that happens.

---

252. The full text is available on Professor Sokal's web page at [www.physics.nyu.edu/faculty/sokal/](http://www.physics.nyu.edu/faculty/sokal/).

## Problems

### Key

✓ A computerized answer check is available online.

∫ A problem that requires calculus.

★ A difficult problem.

**1** Convert 134 mg to units of kg, writing your answer in scientific notation. ▷ Solution, p. 180

**2** Compute the following things. If they don't make sense because of units, say so.

(a) 3 cm + 5 cm

(b) 1.11 m + 22 cm

(c) 120 miles + 2.0 hours

(d) 120 miles / 2.0 hours

**3** Your backyard has brick walls on both ends. You measure a distance of 23.4 m from the inside of one wall to the inside of the other. Each wall is 29.4 cm thick. How far is it from the outside of one wall to the outside of the other? Pay attention to significant figures.

**4** The speed of light is  $3.0 \times 10^8$  m/s. Convert this to furlongs per fortnight. A furlong is 220 yards, and a fortnight is 14 days. An inch is 2.54 cm. ✓

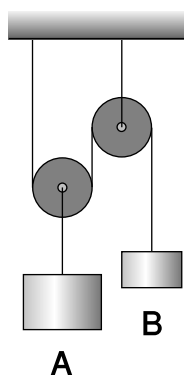
**5** Express each of the following quantities in micrograms:

(a) 10 mg, (b)  $10^4$  g, (c) 10 kg, (d)  $100 \times 10^3$  g, (e) 1000 ng. ✓

**6** In the last century, the average age of the onset of puberty for girls has decreased by several years. Urban folklore has it that this is because of hormones fed to beef cattle, but it is more likely to be because modern girls have more body fat on the average and possibly because of estrogen-mimicking chemicals in the environment from the breakdown of pesticides. A hamburger from a hormone-implanted steer has about 0.2 ng of estrogen (about double the amount of natural beef). A serving of peas contains about 300 ng of estrogen. An adult woman produces about 0.5 mg of estrogen per day (note the different unit!). (a) How many hamburgers would a girl have to eat in one day to consume as much estrogen as an adult woman's daily production? (b) How many servings of peas? ✓

**7** You jump up straight up in the air. When do you have the greatest gravitational energy? The greatest kinetic energy? (Based on a problem by Serway and Faughn.)

**8** Anya and Ivan lean over a balcony side by side. Anya throws a penny downward with an initial speed of 5 m/s. Ivan throws a penny upward with the same speed. Both pennies end up on the ground below. Compare their kinetic energies and velocities on impact.

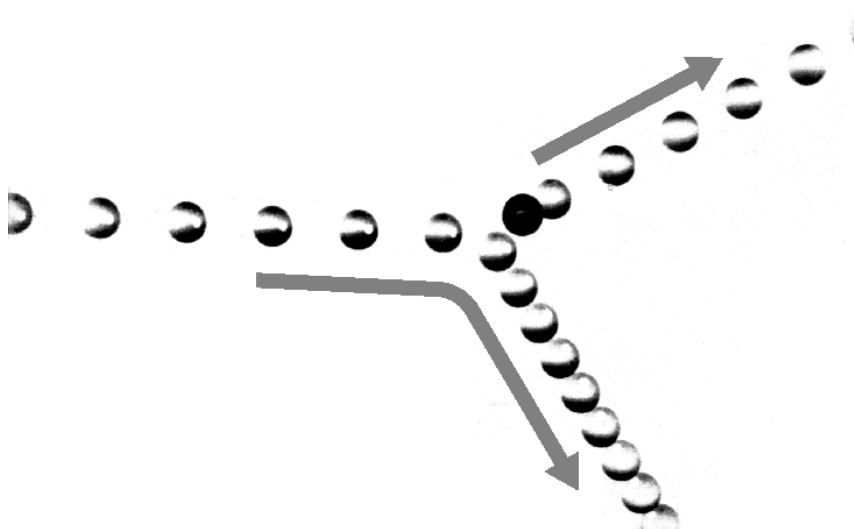


Problem 9.

- 9** (a) If weight B moves down by a certain amount, how much does weight A move up or down?  
 (b) What should the ratio of the two weights be if they are to balance? Explain in terms of conservation of energy.

- 10** (a) You release a magnet on a tabletop near a big piece of iron, and the magnet leaps across the table to the iron. Does the magnetic energy increase, or decrease? Explain.  
 (b) Suppose instead that you have two repelling magnets. You give them an initial push towards each other, so they decelerate while approaching each other. Does the magnetic energy increase, or decrease? Explain.

- 11** For an astronaut sealed inside a space suit, getting rid of body heat can be difficult. Suppose an astronaut is performing vigorous physical activity, expending 200 watts of power. An energy of 200 kJ is enough to raise her body temperature by  $1^{\circ}\text{C}$ . If none of the heat can escape from her space suit, how long will it take before her body temperature rises by  $6^{\circ}\text{C}$  ( $11^{\circ}\text{F}$ ), an amount sufficient to kill her? Express your answer in units of minutes.  $\checkmark$



Problem 12.

- 12** The multiframe photograph below shows a collision between two pool balls. The ball that was initially at rest shows up as a dark image in its initial position, because its image was exposed several times before it was struck and began moving. By making measurements on the figure, determine whether or not energy appears to have been conserved in the collision. What systematic effects would limit the accuracy of your test? (From an example in PSSC Physics.)

**13** How high above the surface of the earth should a rocket be in order to have 1/100 of its normal weight? Express your answer in units of earth radii. ✓

**14** As suggested on page 28, imagine that the mass of the electron rises and falls over time. (Since all electrons are identical, physicists generally talk about “the electron” collectively, as in “the modern man wants more than just beer and sports.”) The idea is that all electrons are increasing and decreasing their masses in unison, and at any given time, they’re all identical. They’re like a litter of puppies whose weights are all identical on any given day, but who all change their weights in unison from one month to the next. Suppose you were the only person who knew about these small day-to-day changes in the mass of the electron. Find a plan for violating conservation of energy and getting rich.

**15** A typical balance like the ones used in school classes can be read to an accuracy of about plus or minus 0.1 grams, or  $10^{-4}$  kg. What if the laws of physics had been designed around a different value of the speed of light? To make mass-energy equivalence detectable in example 15 on page 31 using an ordinary balance, would  $c$  have to be smaller than it is in our universe, or bigger? Find the value of  $c$  for which the effect would be just barely detectable. ✓

**16** (a) A free neutron (as opposed to a neutron bound into an atomic nucleus) is unstable, and decays radioactively into a proton, an electron, and a particle called an antineutrino, which fly off in three different directions. The masses are as follows:

neutron	$1.67495 \times 10^{-27}$ kg
proton	$1.67265 \times 10^{-27}$ kg
electron	$0.00091 \times 10^{-27}$ kg
antineutrino	negligible

Find the energy released in the decay of a free neutron. ✓

(b) Neutrons and protons make up essentially all of the mass of the ordinary matter around us. We observe that the universe around us has no free neutrons, but lots of free protons (the nuclei of hydrogen, which is the element that 90% of the universe is made of). We find neutrons only inside nuclei along with other neutrons and protons, not on their own.

If there are processes that can convert neutrons into protons, we might imagine that there could also be proton-to-neutron conversions, and indeed such a process does occur sometimes in nuclei that contain both neutrons and protons: a proton can decay into a neutron, a positron, and a neutrino. A positron is a particle with the same properties as an electron, except that its electrical charge is positive (see chapter 5). A neutrino, like an antineutrino, has negligible mass.

Although such a process can occur within a nucleus, explain why



it cannot happen to a free proton. (If it could, hydrogen would be radioactive, and you wouldn't exist!)

- 17** (a) A 1.0 kg rock is released from rest, and drops 1.0 m. Find the amount of gravitational energy released. ✓
- (b) Find the rock's kinetic energy at the end of its fall. ✓
- (c) Find the rock's velocity at the end of its fall. ✓



## Chapter 2

# Conservation of Momentum

Fantasy novelist T.H. White invented a wonderful phrase that has since entered into popular culture: “Everything not forbidden is compulsory.” Originally intended as a satire of totalitarianism, it was taken up by physicist Murray Gell-Mann as a metaphor for physics. What he meant was that the laws of physics forbid all the impossible things, and what’s left over is what really happens. Conservation of mass and energy prevent many things from happening. Objects can’t disappear into thin air, and you can’t run your car forever without putting gas in it.

Some other processes are impossible, but not forbidden by these two conservation laws. In the martial arts movie *Crouching Tiger, Hidden Dragon*, those who have received mystical enlightenment are able to violate the laws of physics. Some of the violation, such as their ability to fly, are obvious, but others are a little more subtle. The rebellious young heroine/antiheroine Jen Yu gets into an argument while sitting at a table in a restaurant. A young tough, Iron Arm Lu, comes running toward her at full speed, and she puts up one arm and effortlessly makes him bounce back, without even getting out of her seat or bracing herself against anything. She does all this between bites. It’s impossible, but how do we know it’s impossible? It doesn’t violate conservation of mass, because neither character’s mass changes. It conserves energy as well, since the rebounding Lu has the same energy he started with.

Suppose you live in a country where the only laws are prohibi-

tions against murder and robbery. One day someone covers your house with graffiti, and the authorities refuse to prosecute, because no crime was committed. You're convinced of the need for a new law against vandalism. Similarly, the story of Jen Yu and Iron Arm Lu shows that we need a new conservation law.

## 2.1 Translation Symmetry

The most fundamental laws of physics are conservation laws, and Noether's theorem tells us that conservation laws are the way they are because of symmetry. Time symmetry is responsible for conservation of energy, but time is like a river with only two directions, past and future. What's impossible about Lu's motion is the abrupt reversal in the direction of his motion in *space*, but neither time symmetry nor energy conservation tell us anything about directions in space. When you put gas in your car, you don't have to decide whether you want to buy north gas or south gas, east, west, up or down gas. Energy has no direction. What we need is a new conserved quantity that has a direction in space, and such a conservation law can only come from a symmetry that relates to space. Since we've already had some luck with time symmetry, which says that the laws of physics are the same at all times, it seems reasonable to turn now to the possibility of a new type of symmetry, which would state that the laws of physics are the same in all places in space. This is known as translation symmetry, where the word "translation" is being used in a mathematical sense that means sliding something around without rotating it.

Translation symmetry would seem reasonable to most people, but you'll see that it ends up producing some very surprising results. To see how, it will be helpful to imagine the consequences of a violation of translation symmetry. What if, like the laws of nations, the laws of physics were different in different places? What would happen, and how would we detect it? We could try doing the same experiment in two different places and comparing the results, but it's even easier than that. Tap your finger on this spot on the page

×

and then wait a second and do it again. Did both taps occur at the same point in space? You're probably thinking that's a silly question; am I just checking whether you followed my directions? Not at all. Consider the whole scene from the point of view of a Martian who is observing it through a powerful telescope from her home planet. (You didn't draw the curtains, did you?) From her point of view, the earth is spinning on its axis and orbiting the sun, at speeds measured in thousands of kilometers per hour. According to her, your second finger tap happened at a point in space about 30 kilometers from the first. If you want to impress the Martians and

win the Martian version of the Nobel Prize for detecting a violation of translation symmetry, all you have to do is perform a physics experiment twice in the same laboratory, and show that the result is different.

But who's to say that the Martian point of view is the right one? It gets a little thorny now. How do you know that what you detected was a violation of translation symmetry at all? Maybe it was just a violation of time symmetry. The Martian Nobel committee isn't going to give you the prize based on an experiment this ambiguous. A possible scheme for resolving the ambiguity would be to wait a year and do the same experiment a third time. After a year, the earth will have completed one full orbit around the sun, and your lab will be back in the same spot in space. If the third experiment comes out the same as the first one, then you can make a strong argument that what you've detected is an asymmetry of space, not time. There's a problem, however. You and the Martians agree that the earth is back in the same place after a year, but what about an observer from another solar system, whose planet orbits a different star? This observer says that our whole solar system is in motion. To him, the earth's motion around our sun looks like a spiral or a corkscrew, since the sun is itself moving.

## 2.2 The Principle of Inertia

### Symmetry and inertia

This story shows that translation symmetry is closely related to the relative nature of motion, as expressed by the principle of inertia. Riding in a train on a long, straight track at constant speed, how can you even tell you're in motion? You can look at the scenery outside, but that's irrelevant, because we could argue that the trees and cows are moving while you stand still. (The Martians say both train and scenery are moving.) The real point is whether you can detect your motion without reference to any external object. You can hear the repetitive thunk-thunk-thunk as the train passes from one piece of track to the next, but again this is just a reference to an external object — all that proves is that you're moving relative to the tracks, but is there any way to tell that you're moving in some *absolute* sense? Assuming no interaction with the outside world, is there any experiment you can do that will give a different result when the train is in motion than when it's at rest? You could if translation symmetry was violated. If the laws of physics were different in different places, then as the train moved it would pass through them. "Riding over" these regions would be like riding over the pieces of track, but you would be able to detect the transition from one region to the next simply because experiments inside the train came out different, without referring to any external objects. Rather than the thunk-thunk-thunk of the rails, you would detect

increases and decreases in some quantity such as the gravitational constant  $G$ , or the speed of light, or the mass of the electron.

We can therefore conclude that the following two hypotheses are closely related.

### **The principle of inertia**

The results of experiments don't depend on the straight-line, constant-speed motion of the apparatus.

### **Translation symmetry**

The laws of physics are the same at every point in space. Specifically, experiments don't give different results just because you set up your apparatus in a different place.

#### *A state of absolute rest*

#### *example 1*

Suppose that translation symmetry is violated. The laws of physics are different in one region of space than in another. Cruising in our spaceship, we monitor the fluctuations in the laws of physics by watching the needle on a meter that measures some fundamental quantity such as the gravitational constant. We make a short blast with the ship's engines and turn them off again. Now we see that the needle is wavering more slowly, so evidently it's taking us more time to move from one region to the next. We keep on blasting with the ship's engines until the fluctuations stop entirely. Now we know that we're in a state of absolute rest. The violation of translation symmetry logically resulted in a violation of the principle of inertia.

#### *self-check A*

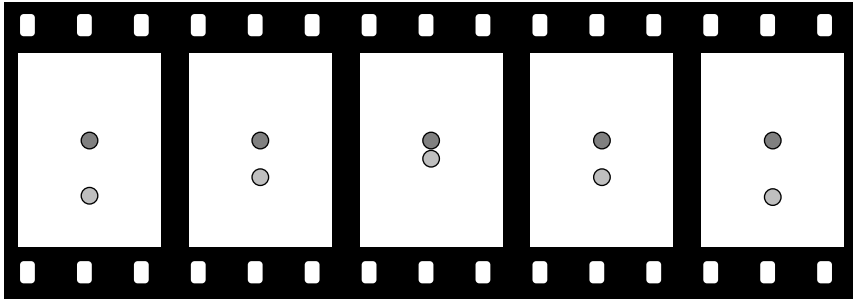
Suppose you do an experiment to see how long it takes for a rock to drop one meter. This experiment comes out different if you do it on the moon. Does this violate translation symmetry? ▷ Answer, p. 179

## **2.3 Momentum**

### **Conservation of momentum**

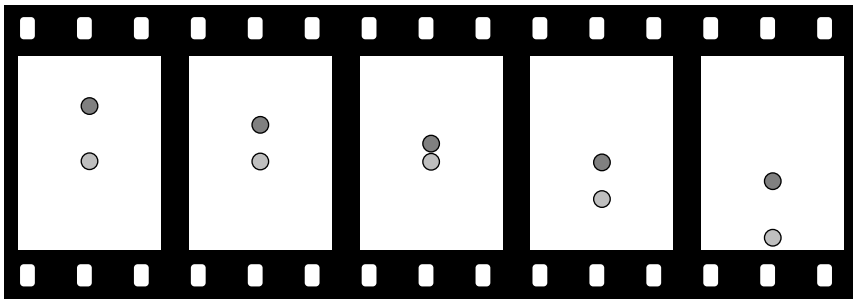
Let's return to the impossible story of Jen Yu and Iron Arm Lu on page 37. For simplicity, we'll model them as two identical, featureless pool balls,  $a$ . This may seem like a drastic simplification, but even a collision between two human bodies is really just a series of many collisions between atoms. The film shows a series of instants in time, viewed from overhead. The light-colored ball comes in, hits the darker ball, and rebounds. It seems strange that the dark ball has such a big effect on the light ball without experiencing any consequences itself, but how can we show that this is really impossible?





a / How can we prove that this collision is impossible?

We can show it's impossible by looking at it in a different frame of reference, b. This camera follows the light ball on its way in, so in this frame the incoming light ball appears motionless. (If you ever get hauled into court on an assault charge for hitting someone, try this defense: "Your honor, in my fist's frame of reference, it was his face that assaulted my knuckles!") After the collision, we let the camera keep moving in the same direction, because if we didn't, it wouldn't be showing us an inertial frame of reference. To help convince yourself that figures a and b represent the same motion seen in two different frames, note that both films agree on the distances between the balls at each instant. After the collision, frame b shows the light ball moving twice as fast as the dark ball; an observer who prefers frame a explains this by saying that the camera that produced film b was moving one way, while the ball was moving the opposite way.



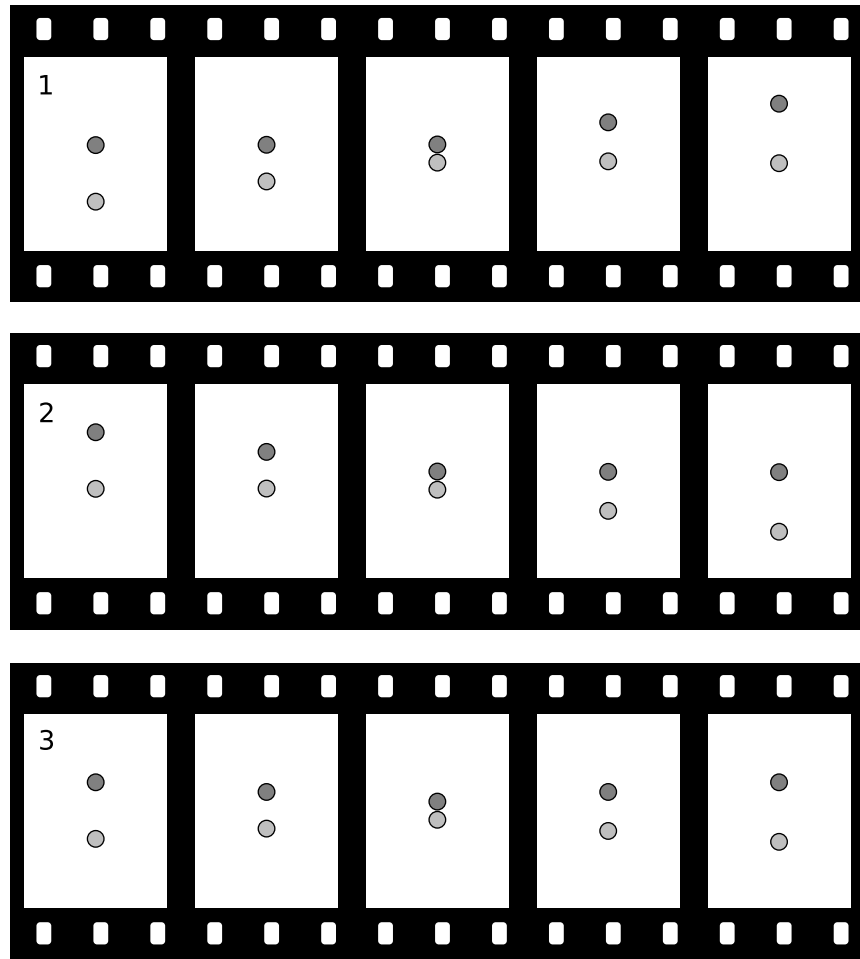
b / The collision of figure a is viewed in a different frame of reference.

Figures a and b record the same events, so if one is impossible, the other is too. But figure b is definitely impossible, because it violates conservation of energy. Before the collision, the only kinetic energy is the dark ball's. After the collision, light ball suddenly has some energy, but where did that energy come from? It can only have come from the dark ball. The dark ball should then have lost some energy, which it hasn't, since it's moving at the same speed as before.

Figure c shows what really does happen. This kind of behavior is familiar to anyone who plays pool. In a head-on collision, the

incoming ball stops dead, and the target ball takes all its energy and flies away. In  $c/1$ , the light ball hits the dark ball. In  $c/2$ , the camera is initially following the light ball; in this frame of reference, the dark ball hits the light one (“Judge, his face hit my knuckles!”). The frame of reference shown in  $c/3$  is particularly interesting. Here the camera always stays at the midpoint between the two balls. This is called the center-of-mass frame of reference.

$c/$  This is what really happens. Three films represent the same collision viewed in three different frames of reference. Energy is conserved in all three frames. Figure d shows a less cumbersome way of representing the same thing.



*self-check B*

In each picture in figure  $c/1$ , mark an  $x$  at the point half-way in between the two balls. This series of five  $x$ 's represents the motion of the camera that was used to make the bottom film. How fast is the camera moving? Does it represent an inertial frame of reference? ▷ Answer, p. 179

What's special about the center-of-mass frame is its symmetry. In this frame, both balls have the same initial speed. Since they start out with the same speed, and they have the same mass, there's no reason for them to behave differently from each other after the collision. By symmetry, if the light ball feels a certain effect from the dark ball, the dark ball must feel the same effect from the light

ball.

This is exactly like the rules of accounting. Let's say two big corporations are doing business with each other. If Glutcorp pays a million dollars to Slushco, two things happen: Glutcorp's bank account goes down by a million dollars, and Slushco's rises by the same amount. The two companies' books have to show transactions on the same date that are equal in size, but one is positive (a payment) and one is negative. What if Glutcorp records  $-1,000,000$  dollars, but Slushco's books say  $+920,000$ ? This indicates that a law has been broken; the accountants are going to call the police and start looking for the employee who's driving a new 80,000-dollar Jaguar. Money is supposed to be conserved.

In figure c, let's define velocities as positive if the motion is toward the top of the page. In figure c/1 let's say the incoming light ball's velocity is 1 m/s.

	velocity (meters per second)		
	before the collision	after the collision	change
●	0	1	+1
○	1	0	-1

The books balance. The light ball's payment,  $-1$ , matches the dark ball's receipt,  $+1$ . Everything also works out fine in the center of mass frame,  $c/3$ :

	velocity (meters per second)		
	before the collision	after the collision	change
●	$-0.5$	$+0.5$	+1
○	$+0.5$	$-0.5$	-1

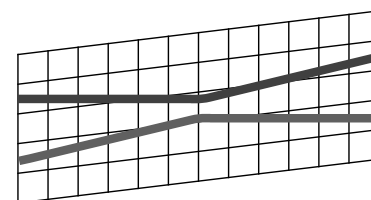
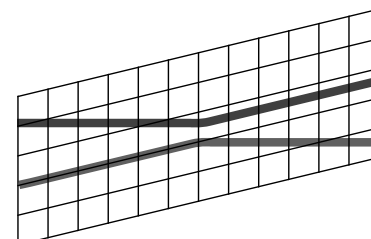
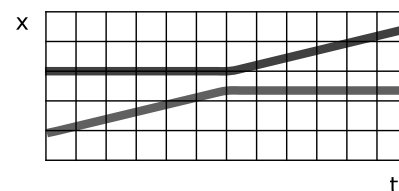
*self-check C*

Make a similar table for figure c/2. What do you notice about the change in velocity when you compare the three tables? ▷ Answer, p. 179

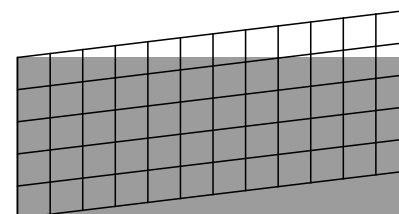
Accounting works because money is conserved. Apparently, something is also conserved when the balls collide. We call it momentum. Momentum is not the same as velocity, because conserved quantities have to be additive. Our pool balls are like identical atoms, but atoms can be stuck together to form molecules, people, and planets. Because conservation laws work by addition, two atoms stuck together and moving at a certain velocity must have double the momentum that a single atom would have had. We therefore define momentum as velocity multiplied by mass.

**Conservation of momentum**

The quantity defined by momentum  $= mv$  is conserved.



d / A less cumbersome representation of figure c. The collision is shown as a graph of position,  $x$ , versus time,  $t$ . By distorting the graph-paper grid, we can describe the same collision in the two other frames of reference. Cf. figure e.



e / A visual shorthand for describing the relationship between two frames of reference as in figure d: the gray rectangle represents the original, undistorted, graph paper, while the superimposed grid shows a different frame.

This is our second example of Noether's theorem:

symmetry		conserved quantity
time symmetry	⇒	mass-energy
translation symmetry	⇒	momentum

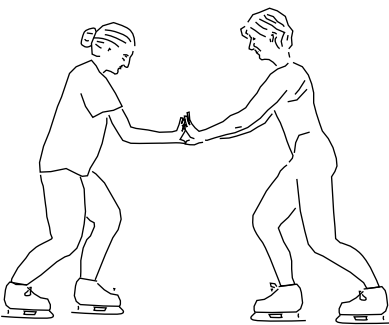
*Conservation of momentum for pool balls* *example 2*

- ▷ Is momentum conserved in figure c/1?
- ▷ We have to check whether the total initial momentum is the same as the total final momentum.

$$\begin{aligned} &\text{dark ball's initial momentum} + \text{light ball's initial momentum} \\ &\qquad\qquad\qquad =? \\ &\text{dark ball's final momentum} + \text{light ball's final momentum} \end{aligned}$$

Yes, momentum was conserved:

$$0 + mv = mv + 0$$



f / Example 3.

*Ice skaters push off from each other* *example 3*

If the ice skaters in figure f have equal masses, then left-right (mirror) symmetry implies that they moved off with equal speeds in opposite directions. Let's check that this is consistent with conservation of momentum:

$$\begin{aligned} &\text{left skater's initial momentum} + \text{right skater's initial momentum} \\ &\qquad\qquad\qquad =? \\ &\text{left skater's final momentum} + \text{right skater's final momentum} \end{aligned}$$

Momentum was conserved:

$$0 + 0 = m \times (-v) + mv$$

This is an interesting example, because if these had been pool balls instead of people, we would have accused them of violating conservation of energy. Initially there was zero kinetic energy, and at the end there wasn't zero. (Note that the energies at the end don't cancel, because kinetic energy is always positive, regardless of direction.) The mystery is resolved because they're people, not pool balls. They both ate food, and they therefore have chemical energy inside their bodies:

$$\text{food energy} \rightarrow \text{kinetic energy} + \text{kinetic energy} + \text{heat}$$

*Unequal masses* *example 4*

▷ Suppose the skaters have unequal masses: 50 kg for the one on the left, and 55 kg for the other. The more massive skater, on the right, moves off at 1.0 m/s. How fast does the less massive skater go?

▷ Their momenta (plural of momentum) have to be the same amount, but with opposite signs. The less massive skater must have a greater velocity if her momentum is going to be as much as the more massive one's.

$$\begin{aligned}
 0 + 0 &= (50 \text{ kg})(-v) + (55 \text{ kg})(1.0 \text{ m/s}) \\
 (50 \text{ kg})(v) &= (55 \text{ kg})(1.0 \text{ m/s}) \\
 v &= \frac{(55 \text{ kg})}{50 \text{ kg}}(1.0 \text{ m/s}) \\
 &= 1.1 \text{ m/s}
 \end{aligned}$$

**Momentum compared to kinetic energy**

Momentum and kinetic energy are both measures of the amount of motion, and a sideshow in the Newton-Leibniz controversy over who invented calculus was an argument over which quantity was the “true” measure of motion. The modern student can certainly be excused for wondering why we need both quantities, when their complementary nature was not evident to the greatest minds of the 1700’s. The following table highlights their differences.

<b>Kinetic energy . . .</b>	<b>Momentum . . .</b>
has no direction in space.	has a direction in space.
is always positive, and cannot cancel out.	cancels with momentum in the opposite direction.
can be traded for forms of energy that do not involve motion. KE is not a conserved quantity by itself.	is always conserved.
is quadrupled if the velocity is doubled.	is doubled if the velocity is doubled.

Here are some examples that show the different behaviors of the two quantities.

*A spinning coin* *example 5*

A spinning coin has zero total momentum, because for every moving point, there is another point on the opposite side that cancels its momentum. It does, however, have kinetic energy.



g / Example 5.

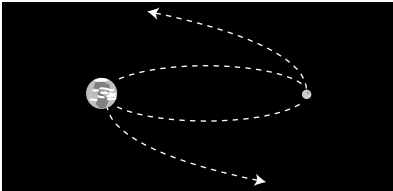
*Momentum and kinetic energy in firing a rifle* *example 6*

The rifle and bullet have zero momentum and zero kinetic energy to start with. When the trigger is pulled, the bullet gains some momentum in the forward direction, but this is canceled by the rifle's

backward momentum, so the total momentum is still zero. The kinetic energies of the gun and bullet are both positive numbers, however, and do not cancel. The total kinetic energy is allowed to increase, because both objects' kinetic energies are destined to be dissipated as heat — the gun's "backward" kinetic energy does not refrigerate the shooter's shoulder!

*The wobbly earth* *example 7*

As the moon completes half a circle around the earth, its motion reverses direction. This does not involve any change in kinetic energy, because the moon doesn't speed up or slow down, nor is there any change in gravitational energy, because the moon stays at the same distance from the earth.<sup>1</sup> The reversed velocity does, however, imply a reversed momentum, so conservation of momentum tells us that the earth must also change its momentum. In fact, the earth wobbles in a little "orbit" about a point below its surface on the line connecting it and the moon. The two bodies' momenta always point in opposite directions and cancel each other out.



h / Example 8.

*The earth and moon get a divorce* *example 8*

Why can't the moon suddenly decide to fly off one way and the earth the other way? It is not forbidden by conservation of momentum, because the moon's newly acquired momentum in one direction could be canceled out by the change in the momentum of the earth, supposing the earth headed the opposite direction at the appropriate, slower speed. The catastrophe is forbidden by conservation of energy, because both their kinetic energies would have increased greatly.

*Momentum and kinetic energy of a glacier* *example 9*

A cubic-kilometer glacier would have a mass of about  $10^{12}$  kg — 1 followed by 12 zeroes. If it moves at a speed of 0.00001 m/s, then its momentum would be 10,000,000 kg·m/s. This is the kind of heroic-scale result we expect, perhaps the equivalent of the space shuttle taking off, or all the cars in LA driving in the same direction at freeway speed. Its kinetic energy, however, is only 50 joules, the equivalent of the calories contained in a poppy seed or the energy in a drop of gasoline too small to be seen without a microscope. The surprisingly small kinetic energy is because kinetic energy is proportional to the square of the velocity, and the square of a small number is an even smaller number.

## Force

<sup>1</sup>Actually these statements are both only approximately true. The moon's orbit isn't exactly a circle.



### Definition of force

When momentum is being transferred, we refer to the rate of transfer as the force.<sup>2</sup> The metric unit of force is the newton (N). The relationship between force and momentum is like the relationship between power and energy, or the one between your cash flow and your bank balance:

conserved quantity		rate of transfer	
name	units	name	units
energy	joules (J)	power	watts (W)
momentum	kg·m/s	force	newtons (N)

*A bullet*

*example 10*

▷ A bullet emerges from a gun with a momentum of 1.0 kg·m/s, after having been acted on for 0.01 seconds by the force of the gases from the explosion of the gunpowder. What was the force on the bullet?

▷ The force is<sup>3</sup>

$$\frac{1.0}{0.01} = 100 \text{ newtons.}$$

There's no new physics happening here, just a definition of the word "force." Definitions are neither right nor wrong, and just because the Chinese call it 力 instead, that doesn't mean they're incorrect. But when Isaac Newton first started using the term "force" according to this technical definition, people already had some definite ideas about what the word meant.

In some cases Newton's definition matches our intuition. In example 10, we divided by a small time, and the result was a big force; this is intuitively reasonable, since we expect the force on the bullet to be strong.

### Forces occur in equal-strength pairs

In other situations, however, our intuition rebels against reality.

*Extra protein*

*example 11*

▷ While riding my bike fast down a steep hill, I pass through a cloud of gnats, and one of them goes into my mouth. Compare my force on the gnat to the gnat's force on me.

▷ Momentum is conserved, so the momentum gained by the gnat equals the momentum lost by me. Momentum conservation holds true at every instant over the fraction of a second that it takes for

<sup>2</sup>This definition is known as Newton's second law of motion. Don't memorize that!

<sup>3</sup>This is really only an estimate of the average force over the time it takes for the bullet to move down the barrel. The force probably starts out stronger than this, and then gets weaker because the gases expand and cool.

the collision to happen. The rate of transfer of momentum out of me must equal the rate of transfer into the gnat. Our forces on each other have the same strength, but they're in opposite directions.

Most people would be willing to believe that the momentum gained by the gnat is the same as the momentum lost by me, but they would not believe that the forces are the same strength. Nevertheless, the second statement follows from the first merely as a matter of definition. Whenever two objects, A and B, interact, A's force on B is the same strength as B's force on A, and the forces are in opposite directions.<sup>4</sup>

$$(A \text{ on } B) = -(B \text{ on } A)$$

Using the metaphor of money, suppose Alice and Bob are adrift in a life raft, and pass the time by playing poker. Money is conserved, so if they count all the money in the boat every night, they should always come up with the same total. A completely equivalent statement is that their cash flows are equal and opposite. If Alice is winning five dollars per hour, then Bob must be losing at the same rate.

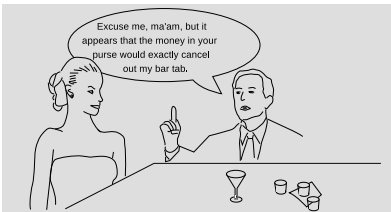
This statement about equal forces in opposite directions implies to many students a kind of mystical principle of equilibrium that explains why things don't move. That would be a useless principle, since it would be violated every time something moved.<sup>5</sup> The ice skaters of figure f on page 44 make forces on each other, and their forces are equal in strength and opposite in direction. That doesn't mean they won't move. They'll both move — in opposite directions.

The fallacy comes from trying to add things that it doesn't make sense to add, as suggested by the cartoon in figure i. We only add forces that are acting on the same object. It doesn't make sense to say that the skaters' forces on each other add up to zero, because it doesn't make sense to add them. One is a force on the left-hand skater, and the other is a force on the right-hand skater.

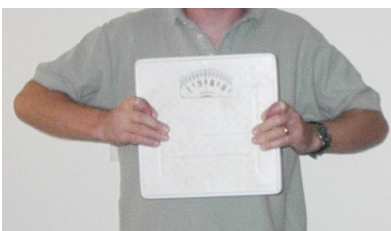
In figure j, my fingers' force and my thumbs' force are both acting on the bathroom scale. It does make sense to add these forces, and they may possibly add up to zero, but that's not guaranteed by the laws of physics. If I throw the scale at you, my thumbs' force is stronger than my fingers', and the forces no longer cancel:

$$(\text{fingers on scale}) \neq -(\text{thumbs on scale}).$$

What's guaranteed by conservation of momentum is a whole differ-



i / It doesn't make sense to add his debts to her assets.



j / I squeeze the bathroom scale. It does make sense to add my fingers' force to my thumbs', because they both act on the same object — the scale.

<sup>4</sup>This is called Newton's third law. Don't memorize that name!

<sup>5</sup>During the Scopes monkey trial, William Jennings Bryan claimed that every time he picked his foot up off the ground, he was violating the law of gravity.

ent relationship:

$$\begin{aligned}(\text{fingers on scale}) &= -(\text{scale on fingers}) \\ (\text{thumbs on scale}) &= -(\text{scale on thumbs})\end{aligned}$$

### *The force of gravity*

How much force does gravity make on an object? From everyday experience, we know that this force is proportional to the object's mass.<sup>6</sup> Let's find the force on a one-kilogram object. If we release this object from rest, then after it has fallen one meter, its kinetic energy equals the strength of the gravitational field,

$$10 \text{ joules per kilogram per meter} \times 1 \text{ kilogram} \times 1 \text{ meter} = 10 \text{ joules.}$$

Using the equation for kinetic energy and doing a little simple algebra, we find that its final velocity is 4.4 m/s. It starts from 0 m/s, and ends at 4.4 m/s, so its average velocity is 2.2 m/s, and the time takes to fall one meter is therefore  $(1 \text{ m})/(2.2 \text{ m/s})=0.44$  seconds. Its final momentum is 4.4 units, so the force on it was evidently

$$\frac{4.4}{0.44} = 10 \text{ newtons.}$$

This is like one of those card tricks where the magician makes you go through a bunch of steps so that you end up revealing the card you had chosen — the result is just equal to the gravitational field, 10, but in units of newtons! If algebra makes you feel warm and fuzzy, you may want to replay the derivation using symbols and convince yourself that it had to come out that way. If not, then I hope the numerical result is enough to convince you of the general fact that the force of gravity on a one-kilogram mass equals  $g$ . For masses other than one kilogram, we have the handy-dandy result that

$$(\text{force of gravity on a mass } m) = mg.$$

In other words,  $g$  can be interpreted not just as the gravitational energy per kilogram per meter of height, but also as the gravitational force per kilogram.

## **Motion in two dimensions**

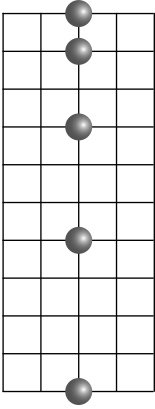
### *Projectile motion*

Galileo was an innovator in more than one way. He was arguably the inventor of open-source software: he invented a mechanical calculating device for certain engineering applications, and rather than keeping the device's design secret as his competitors did, he made it public, but charged students for lessons in how to use it. Not

---

<sup>6</sup>This follows from the additivity of forces.

only that, but he was the first physicist to make money as a military consultant. Galileo understood projectiles better than anyone else, because he understood the principle of inertia. Even if you're not planning on a career involving artillery, projectile motion is a good thing to learn about because it's an example of how to handle motion in two or three dimensions.

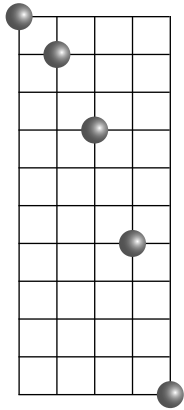


k / A ball is falling (or rising).

Figure k shows a ball in the process of falling — or rising, it really doesn't matter which. Let's say the ball has a mass of one kilogram, each square in the grid is 10 meters on a side, and the positions of the ball are shown at time intervals of one second. The earth's gravitational force on the ball is 10 newtons, so with each second, the ball's momentum increases by 10 units, and its speed also increases by 10 m/s. The ball falls 10 m in the first second, 20 m in the next second, and so on.

*self-check D*

What would happen if the ball's mass was 2 kilograms?   ▷ Answer, p. 179



l / The same ball is viewed in a frame of reference that is moving horizontally.

Now let's look at the ball's motion in a new frame of reference, l, which is moving at 10 meters per second to the left compared to the frame of reference used in figure k. An observer in this frame of reference sees the ball as moving to the right by 10 meters every second. The ball traces an arc of a specific mathematical type called a parabola:

- 1 step over and 1 step down
- 1 step over and 2 steps down
- 1 step over and 3 steps down
- 1 step over and 4 steps down
- ...

It doesn't matter which frame of reference is the "real" one. Both diagrams show the possible motion of a projectile. The interesting point here is that the vertical force of gravity has no effect on the horizontal motion, and the horizontal motion also has no effect on what happens in the vertical motion. The two are completely independent. If the sun is directly overhead, the motion of the ball's shadow on the ground seems perfectly natural: there are no horizontal forces, so it either sits still or moves at constant velocity. (Zero force means zero rate of transfer of momentum.) The same is true if we shine a light from one side and cast the ball's shadow on the wall. Both shadows obey the laws of physics.



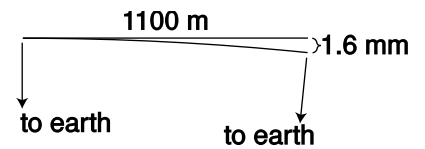
m / The drops of water travel in parabolic arcs.

*The moon*

*example 12*

In example 12 on page 27, I promised an explanation of how Newton knew that the gravitational field experienced by the moon due to the earth was 1/3600 of the one we feel here on the earth's surface. The radius of the moon's orbit had been known since ancient times, so Newton knew its speed to be 1,100 m/s (ex-

pressed in modern units). If the earth's gravity wasn't acting on the moon, the moon would fly off straight, along the straight line shown in figure o, and it would cover 1,100 meters in one second. We observe instead that it travels the arc of a circle centered on the earth. Straightforward geometry shows that the amount by which the arc drops below the straight line is 1.6 millimeters. Near the surface of the earth, an object falls 5 meters in one second,<sup>7</sup> which is indeed about 3600 times greater than 1.6 millimeters.



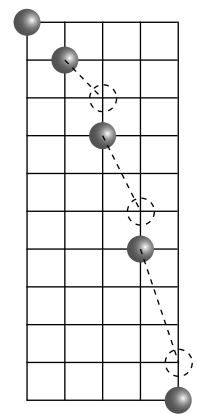
o / Example 12.

The tricky part about this argument is that although I said the path of a projectile was a parabola, in this example it's a circle. What's going on here? What's different here is that as the moon moves 1,100 meters, it changes its position relative to the earth, so down is now in a new direction. We'll discuss circular motion more carefully soon, but in this example, it really doesn't matter. The curvature of the arc is so gentle that a parabola and a circle would appear almost identical. (Actually the curvature is so gentle — 1.6 millimeters over a distance of 1,100 meters! — that if I had drawn the figure to scale, you wouldn't have even been able to tell that it wasn't straight.)

As an interesting historical note, Newton claimed that he first did this calculation while confined to his family's farm during the plague of 1666, and found the results to "answer pretty nearly." His notebooks, however, show that although he did the calculation on that date, the result didn't quite come out quite right, and he became uncertain about whether his theory of gravity was correct as it stood or needed to be modified. Not until 1675 did he learn of more accurate astronomical data, which convinced him that his theory didn't need to be tinkered with. It appears that he rewrote his own life story a little bit in order to make it appear that his work was more advanced at an earlier date, which would have helped him in his dispute with Leibniz over priority in the invention of calculus.

*The memory of motion*

There's another useful way of thinking about motion along a curve. In the absence of a force, an object will continue moving in the same speed and in the same direction. One of my students invented a wonderful phrase for this: the memory of motion. Over the first second of its motion, the ball in figure n moved 1 square over and 1 square down, which is 10 meters and 10 meters. The default for the next one-second interval would be to repeat this, ending up at the location marked with the first dashed circle. The earth's 10-newton gravitational force on the ball, however, changes the vertical part of the ball's momentum by 10 units. The ball actually ends up



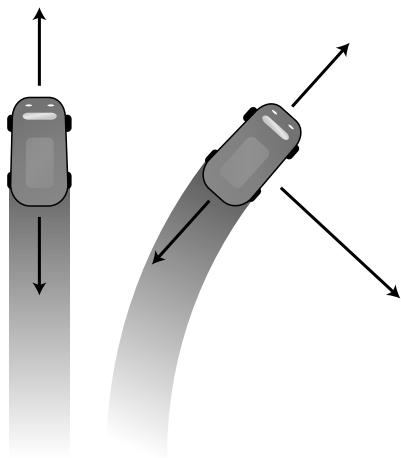
n / The memory of motion: the default would be for the ball to continue doing what it was already doing. The force of gravity makes it deviate downward, ending up one square below the default.

<sup>7</sup>Its initial speed is 0, and its final speed is 10 m/s, so its average speed is 5 m/s over the first second of falling.

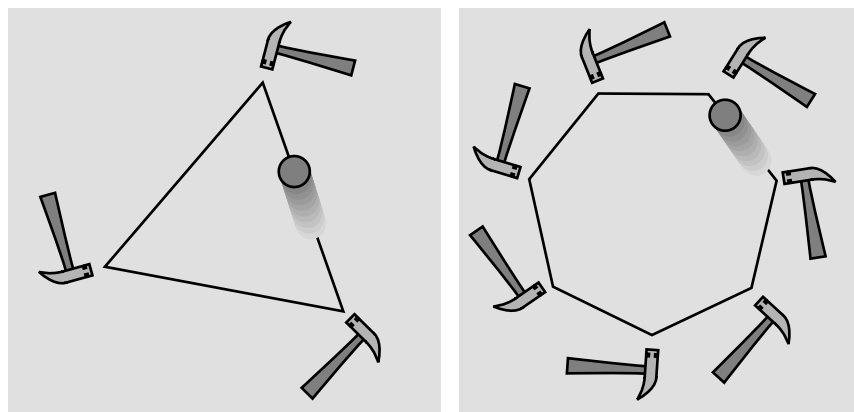
10 meters (1 square) below the default.

### Circular motion

Figure q shows how to apply the memory-of-motion idea to circular motion. It should convince you that only an inward force is needed to produce circular motion. One of the reasons Newton was the first to make any progress in analyzing the motion of the planets around the sun was that his contemporaries were confused on this point. Most of them thought that in addition to an attraction from the sun, a second, forward force must exist on the planets, to keep them from slowing down. This is incorrect Aristotelian thinking; objects don't naturally slow down. Car 1 in figure p only needs a forward force in order to cancel out the backward force of friction; the total force on it is zero. Similarly, the forward and backward forces on car 2 are canceling out, and the only force left over is the inward one. There's no friction in the vacuum of outer space, so if car 2 was a planet, the backward force wouldn't exist; the forward force wouldn't exist either, because the only force would be the force of the sun's gravity.



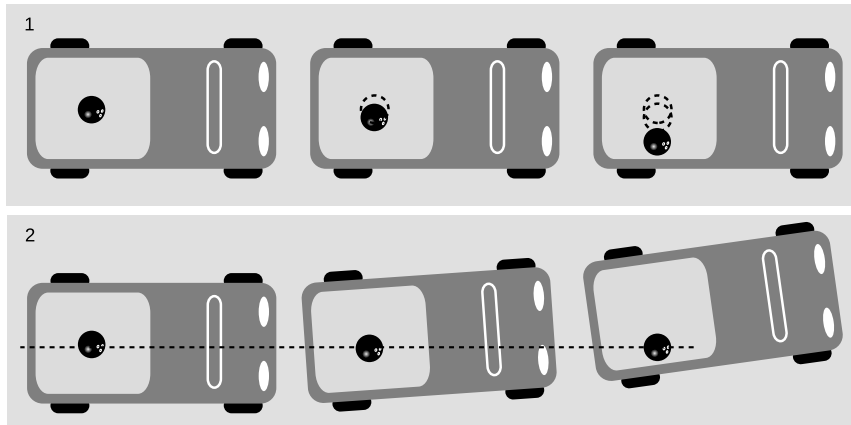
p/The forces on car 1 cancel, and the total force on it is zero. The forward and backward forces on car 2 also cancel. Only the inward force remains.



q/A large number of gentle taps gives a good approximation to circular motion. A steady inward force would give exactly circular motion.

One confusing thing about circular motion is that it often tempts us psychologically to adopt a noninertial frame of reference. Figure r shows a bowling ball in the back of a turning pickup truck. Each panel gives a view of the same events from a different frame of reference. The frame of reference r/1, attached to the turning truck, is noninertial, because it changes the direction of its motion. The ball violates conservation of energy by accelerating from rest for no apparent reason. Is there some mysterious outward force that is slamming the ball into the side of the truck's bed? No. By analyzing everything in a proper inertial frame of reference, r/2, we see that it's the truck that swerves and hits the ball. That makes sense,

because the truck is interacting with the asphalt.



1 / A bowling ball is in the back of a pickup truck turning left. The motion is viewed first in a frame that turns along with the truck, 1, and then in an inertial frame, 2.

## 2.4 Newton's Triumph

Isaac Newton's greatest triumph was his explanation of the motion of the planets in terms of universal physical laws. It was a tremendous psychological revolution: for the first time, both heaven and earth were seen as operating automatically according to the same rules.

Newton wouldn't have been able to figure out *why* the planets move the way they do if it hadn't been for the astronomer Tycho Brahe (1546-1601) and his protege Johannes Kepler (1571-1630), who together came up with the first simple and accurate description of *how* the planets actually do move. The difficulty of their task is suggested by figure 2.4.1, which shows how the relatively simple orbital motions of the earth and Mars combine so that as seen from earth Mars appears to be staggering in loops like a drunken sailor.

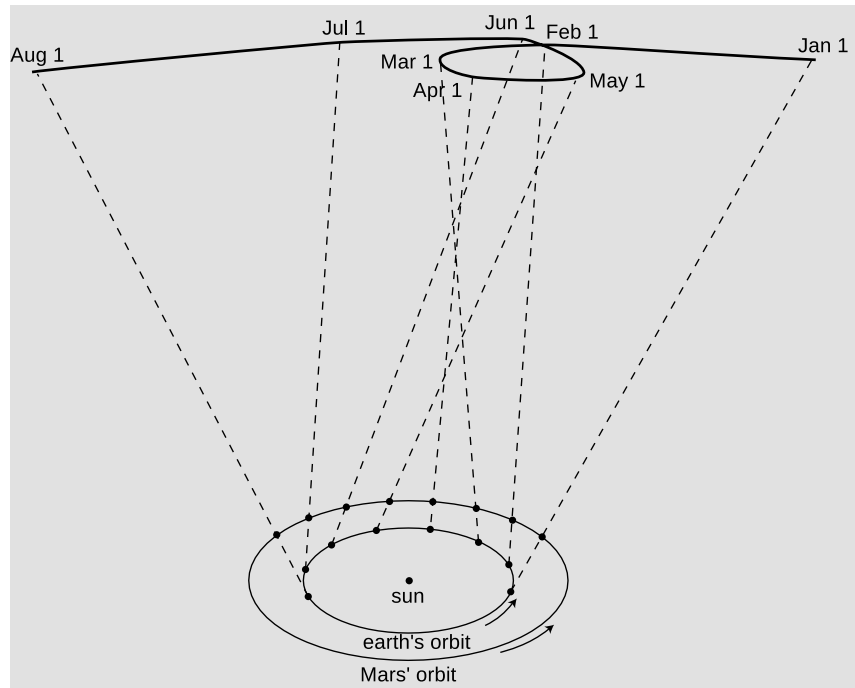
Brahe, the last of the great naked-eye astronomers, collected extensive data on the motions of the planets over a period of many years, taking the giant step from the previous observations' accuracy of about 10 minutes of arc (10/60 of a degree) to an unprecedented 1 minute. The quality of his work is all the more remarkable considering that his observatory consisted of four giant brass protractors mounted upright in his castle in Denmark. Four different observers would simultaneously measure the position of a planet in order to check for mistakes and reduce random errors.

With Brahe's death, it fell to his former assistant Kepler to try



tycho / Tycho Brahe made his name as an astronomer by showing that the bright new star, today called a supernova, that appeared in the skies in 1572 was far beyond the Earth's atmosphere. This, along with Galileo's discovery of sunspots, showed that contrary to Aristotle, the heavens were not perfect and unchanging. Brahe's fame as an astronomer brought him patronage from King Frederick II, allowing him to carry out his historic high-precision measurements of the planets' motions. A contradictory character, Brahe enjoyed lecturing other nobles about the evils of dueling, but had lost his own nose in a youthful duel and had it replaced with a prosthesis made of an alloy of gold and silver. Willing to endure scandal in order to marry a peasant, he nevertheless used the feudal powers given to him by the king to impose harsh forced labor on the inhabitants of his parishes. The result of their work, an Italian-style palace with an observatory on top, surely ranks as one of the most luxurious science labs ever built. He died of a ruptured bladder after falling from a wagon on the way home from a party — in those days, it was considered rude to leave the dinner table to relieve oneself.





s / As the Earth and Mars revolve around the sun at different rates, the combined effect of their motions makes Mars appear to trace a strange, looped path across the background of the distant stars.

to make some sense out of the volumes of data. Kepler, in contradiction to his late boss, had formed a prejudice, a correct one as it turned out, in favor of the theory that the earth and planets revolved around the sun, rather than the earth staying fixed and everything rotating about it. Although motion is relative, it is not just a matter of opinion what circles what. The earth's rotation and revolution about the sun make it a noninertial reference frame, which causes detectable violations of Newton's laws when one attempts to describe sufficiently precise experiments in the earth-fixed frame. Although such direct experiments were not carried out until the 19th century, what convinced everyone of the sun-centered system in the 17th century was that Kepler was able to come up with a surprisingly simple set of mathematical and geometrical rules for describing the planets' motion using the sun-centered assumption. After 900 pages of calculations and many false starts and dead-end ideas, Kepler finally synthesized the data into the following three laws:

#### **Kepler's elliptical orbit law**

The planets orbit the sun in elliptical orbits with the sun at one focus.

#### **Kepler's equal-area law**

The line connecting a planet to the sun sweeps out equal areas in equal amounts of time.

### Kepler's law of periods

Let  $T$ , called the planet's period, be the time required for a planet to orbit the sun, and let  $a$  be the long axis of the ellipse. Then  $T^2$  is proportional to  $a^3$ .

Although the planets' orbits are ellipses rather than circles, most are very close to being circular. The earth's orbit, for instance, is only flattened by 1.7% relative to a circle. In the special case of a planet in a circular orbit, the two foci (plural of "focus") coincide at the center of the circle, and Kepler's elliptical orbit law thus says that the circle is centered on the sun. The equal-area law implies that a planet in a circular orbit moves around the sun with constant speed. For a circular orbit, the law of periods then amounts to a statement that  $T^2$  is proportional to  $r^3$ , where  $r$  is the radius. If all the planets were moving in their orbits at the same speed, then the time for one orbit would only increase with the circumference of the circle, so we would have a simple proportionality between  $T$  and  $r$ . Since this is not the case, we can interpret the law of periods to mean that different planets orbit the sun at different speeds. In fact, the outer planets move more slowly than the inner ones.

#### Jupiter and Uranus

#### example 13

▷ The planets Jupiter and Uranus have very nearly circular orbits, and the radius of Uranus's orbit is about four times greater than that of Jupiter's orbit. Compare their orbital periods.

▷ If all the planets moved at the same speed, then it would take Uranus four times longer to complete the four-times-greater circumference of its orbit. However, the law of periods tells us that this isn't the case. We expect Uranus to take *more* than four times as long to orbit the sun.

The law of periods is stated as a proportionality, and proportionalities are statements about quantities in proportion to one another, i.e., about division. We're given information about Uranus's orbital radius divided by Jupiter's, and what we should expect to get out is information about Uranus's period divided by Jupiter's. Let's call the latter ratio  $y$ . Then we're looking for a number  $y$  such that

$$y^2 = 4^3,$$

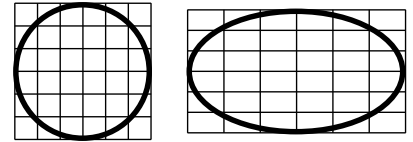
i.e.,

$$y \times y = 4 \times 4 \times 4$$

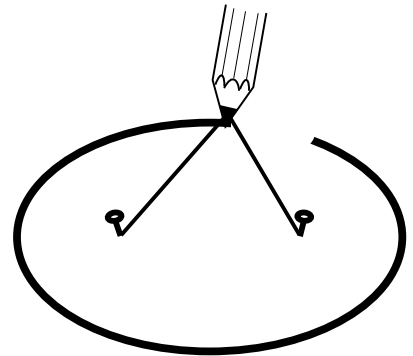
$$y \times y = 64$$

$$y = 8$$

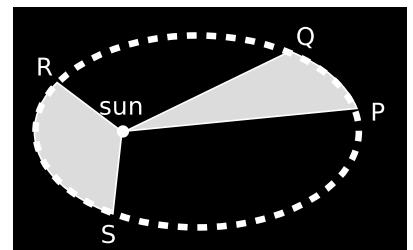
The law of periods predicts that Uranus's period will be eight times greater than Jupiter's, which is indeed what is observed (to within the precision to be expected since the given figure of 4 was just



u / An ellipse is a circle that has been distorted by shrinking and stretching along perpendicular axes.



v / An ellipse can be constructed by tying a string to two pins and drawing like this with the pencil stretching the string taut. Each pin constitutes one focus of the ellipse.



w / If the time interval taken by the planet to move from P to Q is equal to the time interval from R to S, then according to Kepler's equal-area law, the two shaded areas are equal. The planet is moving faster during interval RS than it did during PQ, which Newton later determined was due to the sun's gravitational force accelerating it. The equal-area law predicts exactly how much it will speed up.

stated roughly as a whole number, for convenience in calculation).

What Newton discovered was the reasons why Kepler's laws were true: he showed that they followed from his laws of motion. From a modern point of view, conservation laws are more fundamental than Newton's laws, so rather than following Newton's approach, it makes more sense to look for the reasons why Kepler's laws follow from conservation laws. The equal-area law is most easily understood as a consequence of conservation of angular momentum, which is a new conserved quantity to be discussed in chapter 3. The proof of the elliptical orbit law is a little too mathematical to be appropriate for this book, but the interested reader can find the proof in chapter 15 of my online book *Light and Matter*.

x / Connecting Kepler's law of periods to the laws of physics.



The law of periods follows directly from the physics we've already covered. Consider the example of Jupiter and Uranus. We want to show that the result of example 13 is the only one that's consistent with conservation of energy and momentum, and Newton's law of gravity. Since Uranus takes eight times longer to cover four times the distance, it's evidently moving at half Jupiter's speed. In figure x, the distance Jupiter covers from A to B is therefore twice the distance Uranus covers, over the same time, from D to E. If there hadn't been any gravitational force from the sun, Jupiter would have ended up at C, and Uranus at F. The distance from B to C is a measure of how much force acted on Jupiter, and likewise for the very small distance from E to F. We find that BC is 16 millimeters on this scale drawing, and EF is 1 mm, but this is exactly what we expect from Newton's law of gravity: quadrupling the distance should give 1/16 the force.

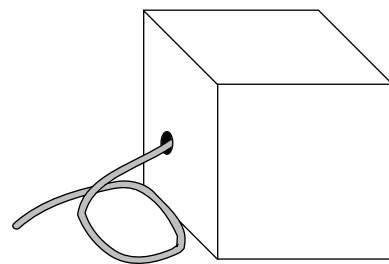
## 2.5 Work

Imagine a black box<sup>8</sup>, containing a gasoline-powered engine, which is designed to reel in a steel cable of length  $d$ , exerting a certain force  $F$ .

If we use this black box to lift a weight, then by the time it has pulled in its whole cable, it will have lifted the weight through a height  $d$ . The force  $F$  is barely capable of lifting a weight  $m$  if  $F = mg$ , and if it does this, then the upward force from the cable exactly cancels the downward force of gravity, so the weight will rise at constant speed, without changing its kinetic energy. Only gravitational energy is transferred into the weight, and the amount of gravitational energy is  $mgd$ , which equals  $Fd$ . By conservation of energy, this must also be the amount of energy lost from the chemical energy of the gasoline inside the box.<sup>9</sup>

Now what if we use the black box to pull a plow? The energy increase in the outside world is of a different type than before: mainly heat created by friction between the dirt and the ploughshare. The box, however, only communicates with the outside world via the hole through which its cable passes. The amount of chemical energy lost by the gasoline can therefore only depend on  $F$  and  $d$ , so again the amount of energy transferred must equal  $Fd$ .

The same reasoning can in fact be applied no matter what the cable is being used to do. There must always be a transfer of energy from the box to the outside world that is equal to  $Fd$ . In general, when energy is transferred, we refer to the amount of energy transferred as *work*,  $W$ . If, as in the example of the black box, the motion of the object to which the force is applied is in the same direction as the force, then  $W = Fd$ .



y / The black box does work by reeling in its cable.



z / The baseball pitcher put kinetic energy into the ball, so he did work on it. To do the greatest possible amount of work, he applied the greatest possible force over the greatest possible distance.

If the motion is in the opposite direction compared to the force, then  $W = -Fd$ ; the negative work is to be interpreted as energy removed from the object to which the force was applied. For ex-

<sup>8</sup>“Black box” is a traditional engineering term for a device whose inner workings we don’t care about.

<sup>9</sup>For conceptual simplicity, we ignore the transfer of heat energy to the outside world via the exhaust and radiator. In reality, the sum of these energies plus the useful kinetic energy transferred would equal  $W$ .

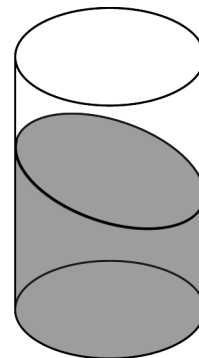
ample, if Superman gets in front of an oncoming freight train, and brings it to a stop, he's decreased its energy rather than increasing it. In a normal gasoline-powered car, stepping on the brakes takes away the car's kinetic energy (doing negative work on it), and turns it into heat in the brake shoes. In an electric or hybrid-electric car, the car's kinetic energy is transformed back into electrical energy to be used again.

## Problems

### Key

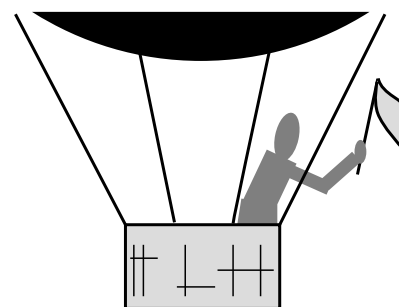
- ✓ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ★ A difficult problem.

**1** The beer bottle shown in the figure is resting on a table in the dining car of a train. The tracks are straight and level. What can you tell about the motion of the train? Can you tell whether the train is currently moving forward, moving backward, or standing still? Can you tell what the train's speed is?



Problem 1.

**2** You're a passenger in the open basket hanging under a hot-air balloon. The balloon is being carried along by the wind at a constant velocity. If you're holding a flag in your hand, will the flag wave? If so, which way? (Based on a question from PSSC Physics.)



Problem 2.

**3** Driving along in your car, you take your foot off the gas, and your speedometer shows a reduction in speed. Describe an inertial frame in which your car was *speeding up* during that same period of time.

**4** If all the air molecules in the room settled down in a thin film on the floor, would that violate conservation of momentum as well as conservation of energy?

**5** A bullet flies through the air, passes through a paperback book, and then continues to fly through the air beyond the book. When is there a force? When is there energy?

**6** (a) Continue figure l farther to the left, and do the same for the numerical table in the text.

(b) Sketch a smooth curve (a parabola) through all the points on the figure, including all the ones from the original figure and all the ones you added. Identify the very top of its arc.

(c) Now consider figure k. Is the highest point shown in the figure the top of the ball's up-down path? Explain by comparing with your results from parts a and b.

**7** Criticize the following statement about the top panel of figure c on page 42: In the first few pictures, the light ball is moving up and to the right, while the dark ball moves directly to the right.

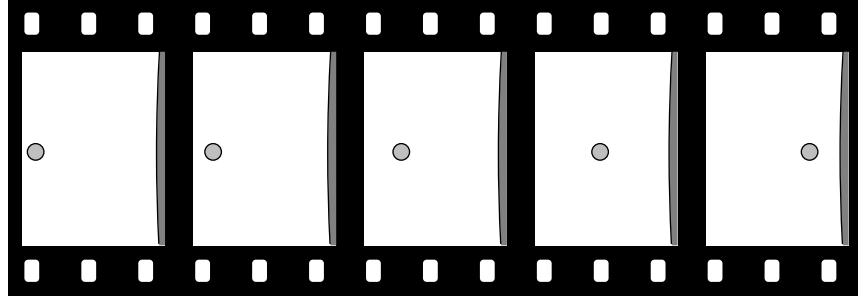
**8** Figure ac on page 60 shows a ball dropping to the surface of the earth. Energy is conserved: over the whole course of the film, the gravitational energy between the ball and the earth decreases by 1 joule, while the ball's kinetic energy increases by 1 joule.

(a) How can you tell *directly from the figure* that the ball's speed isn't staying the same?

(b) Draw what the film would look like if the camera was following the ball.

- (c) Explain how you can tell that in this new frame of reference, energy is not conserved.
- (d) Does this violate the strong principle of inertia? Isn't every frame of reference supposed to be equally valid?

Problem 8.



- 9** Two cars with different masses each have the same kinetic energy. (a) If both cars have the same brakes, capable of supplying the same force, how will the stopping distances compare? Explain. (b) Compare the times required for the cars to stop.

- 10** In each of the following situations, is the work being done positive, negative, or zero? (a) a bull paws the ground; (b) a fishing boat pulls a net through the water behind it; (c) the water resists the motion of the net through it; (d) you stand behind a pickup truck and lower a bale of hay from the truck's bed to the ground. Explain. [Based on a problem by Serway and Faughn.]



A bull paws the ground, as in problem 10.

- 11** Weiping lifts a rock with a weight of 1.0 N through a height of 1.0 m, and then lowers it back down to the starting point. Bubba pushes a table 1.0 m across the floor at constant speed, requiring a force of 1.0 N, and then pushes it back to where it started. (a) Compare the total work done by Weiping and Bubba. (b) Check that your answers to part a make sense, using the definition of work: work is the transfer of energy. In your answer, you'll need to discuss what specific type of energy is involved in each case.





A tornado touches down in Spring Hill, Kansas, May 20, 1957.

## Chapter 3

# Conservation of Angular Momentum

### 3.1 Angular Momentum

“Sure, and maybe the sun won’t come up tomorrow.” Of course, the sun only appears to go up and down because the earth spins, so the cliché should really refer to the unlikelihood of the earth’s stopping its rotation abruptly during the night. Why can’t it stop? It wouldn’t violate conservation of momentum, because the earth’s rotation doesn’t add anything to its momentum. While California spins in one direction, some equally massive part of India goes the opposite way, canceling its momentum. A halt to Earth’s rotation would entail a drop in kinetic energy, but that energy could simply be converted into some other form, such as heat.

Other examples along these lines are not hard to find. A hydro-



a / A figure skater pulls in her arms so that she can execute a spin more rapidly.

gen atom spins at the same rate for billions of years. A high-diver who is rotating when he comes off the board does not need to make any physical effort to continue rotating, and indeed would be unable to stop rotating before he hit the water.

These observations have the hallmarks of a conservation law, but what numerical measure of rotational motion is conserved? Car engines and old-fashioned LP records have speeds of rotation measured in rotations per minute (r.p.m.), but the number of rotations per minute (or per second) is not a conserved quantity. For example, the twirling figure skater in figure a can pull her arms in to increase her r.p.m.'s.

The example of the figure skater suggests that this conserved quantity depends on distance from the axis of rotation. We'll notate this distance as  $r$ , since, for an object moving in a circle around an axis of rotation, its distance from the axis equals the radius of the circle.

Once we realize that  $r$  is a variable that matters, it becomes clear that the examples we've been considering were all examples that would be fairly complicated mathematically, because different parts of these objects' masses have different values of  $r$ . For example, the figure skater's front teeth are farther from the axis than her back teeth. That suggests that instead of objects with complicated shapes, we should consider the simplest possible example, which is a single particle, of mass  $m$ , traveling in a circle of radius  $r$  at speed  $v$ . Experiments show that the conserved quantity in this situation is

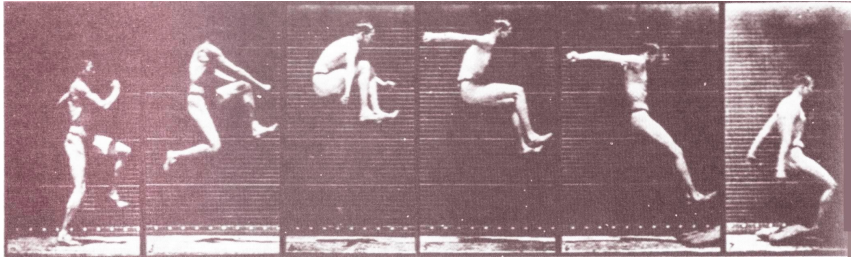
$$\pm mvr.$$

We call this quantity *angular momentum*. The symbol  $\pm$  indicates that angular momentum has a positive or negative sign to represent the direction of rotation; for example, in a given problem, we could choose to represent clockwise angular momenta as positive numbers, and counterclockwise ones as negative. In this equation, the only velocity that matters is velocity that is perpendicular to the radius line; motion parallel to the radius line, i.e., directly in or out, is neither clockwise nor counterclockwise.

*A figure skater pulls her arms in* *example 1*

When the skater in figure a pulls her arms in, she is decreasing  $r$  for all the atoms in her arms. It would violate conservation of angular momentum if she then continued rotating at the same speed, i.e., taking the same amount of time for each revolution, because her arms would be closer to the axis of rotation and therefore have a smaller  $r$  (as well as a smaller  $v$  because they would be completing a smaller circle in the same time). This is impossible because it would violate conservation of angular momentum. If her total angular momentum is to remain constant, the decrease in angular momentum for her arms must be com-

compensated for by an overall increase in her rate of rotation. That is, by pulling her arms in, she substantially reduces the time for each rotation.

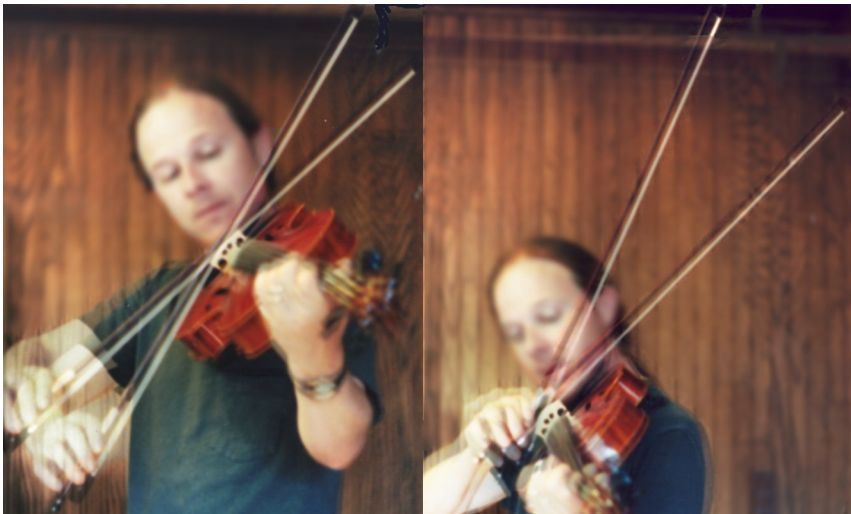


b / Example 2: An early photograph of an old-fashioned long-jump.

*A longjump*

*example 2*

In figure b, the jumper wants to get his feet out in front of him so he can keep from doing a “face plant” when he lands. Bringing his feet forward would involve a certain quantity of counterclockwise rotation, but he didn’t start out with any rotation when he left the ground. Suppose we consider counterclockwise as positive and clockwise as negative. The only way his legs can acquire some positive rotation is if some other part of his body picks up an equal amount of negative rotation. This is why he swings his arms up behind him, clockwise.

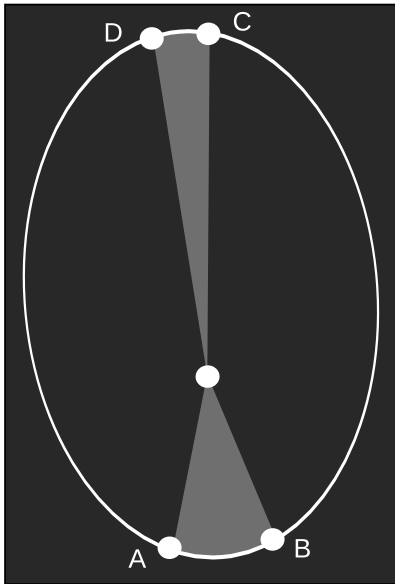


c / Example 3.

*Changing the axis*

*example 3*

An object’s angular momentum can be different depending on the axis about which it rotates, because  $r$  is defined relative to the axis. Figure c shows two double-exposure photographs a viola player tipping the bow in order to cross from one string to another. Much more angular momentum is required when playing near the bow’s handle, called the frog, as in the panel on the right; not



d / Example 4.

only are most of the atoms in the bow are at greater distances,  $r$ , from the axis of rotation, but the ones in the tip also have more velocity,  $v$ . It is difficult for the player to quickly transfer a large angular momentum into the bow, and then transfer it back out just as quickly. This is one of the reasons that string players tend to stay near the middle of the bow as much as possible.

*Kepler's equal-area law*

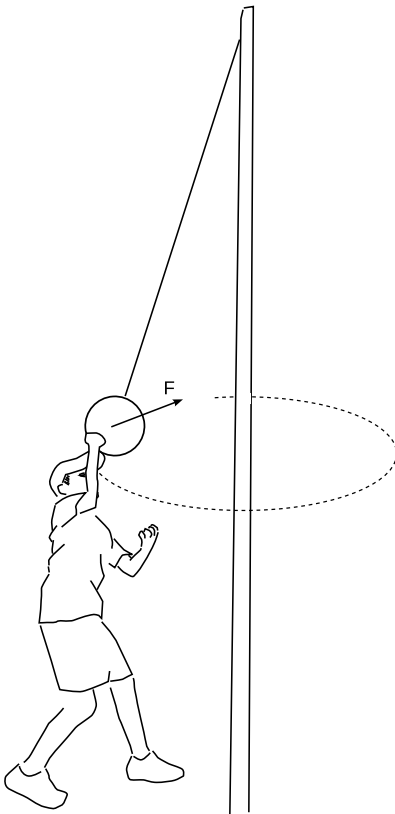
*example 4*

The hypothetical planet in figure d has an orbit in which its closest approach to the sun is at half the distance compared to the point at which it recedes the farthest. Since angular momentum,  $mvr$ , is conserved, and the planet's mass is constant, the quantity  $vr$  must be the same at both ends of the orbit. Doubling  $r$  therefore requires cutting  $v$  in half. If the time interval from A to B is the same as that from C to D, then the distance from C to D must be half as much. But this is exactly what Kepler's equal area law requires, since the triangular pie wedge on top needs to have half the width to compensate for its doubled height. In other words, the equal area law is a direct consequence of conservation of angular momentum.

**Discussion question**

**A** Conservation of plain old momentum,  $p$ , can be thought of as the greatly expanded and modified descendant of Galileo's original principle of inertia, that no force is required to keep an object in motion. The principle of inertia is counterintuitive, and there are many situations in which it appears superficially that a force *is* needed to maintain motion, as maintained by Aristotle. Think of a situation in which conservation of angular momentum,  $L$ , also seems to be violated, making it seem incorrectly that something external must act on a closed system to keep its angular momentum from "running down."

**B** The figure is a strobe photo of a pendulum bob, taken from underneath the pendulum looking straight up. The black string can't be seen in the photograph. The bob was given a slight sideways push when it was released, so it did not swing in a plane. The bright spot marks the center, i.e., the position the bob would have if it hung straight down at us. Does the bob's angular momentum appear to remain constant if we consider the center to be the axis of rotation? What if we choose a different axis?



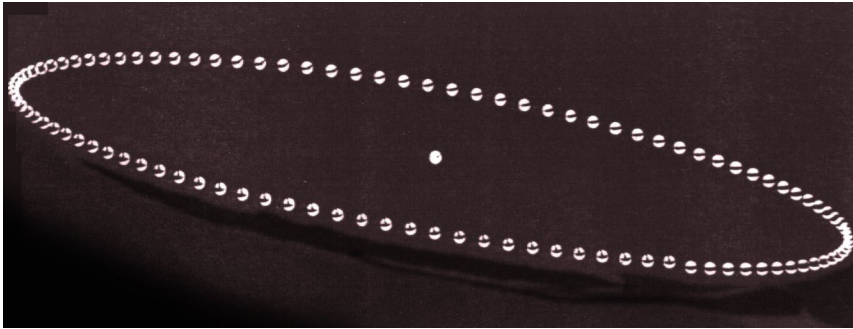
e / The boy makes a torque on the tetherball.

**3.2 Torque**

Force is the rate of transfer of momentum. The equivalent in the case of angular momentum is called *torque* (rhymes with "fork"):

$$\text{torque} = \frac{\text{amount of angular momentum transferred}}{\text{time taken to transfer it}}$$

Where force tells us how hard we are pushing or pulling on something, torque indicates how hard we are twisting on it.



Discussion question B.

Have you ever had the experience of trying to open a door by pushing on the wrong side, the side near the hinge? It's difficult to do, which apparently indicates that a given amount of force produces less torque when it's applied close to the axis of rotation. To try to pin down this relationship more precisely, let's imagine hitting a tetherball, e. The boy applies a force  $F$  to the ball for a short time  $t$ , accelerating the ball from rest to a velocity  $v$ . Since force is the rate of transfer of momentum, we have

$$F = \frac{mv}{t},$$

and multiplying both sides by  $r$  gives

$$Fr = \frac{mvr}{t}.$$

But  $\pm mvr$  is simply the amount of angular momentum he's given the ball, so  $\pm mvr/t$  also equals the amount of torque he applied. The result of this example is

$$\text{torque} = \pm Fr,$$

where the plus or minus sign indicates whether torque would tend to create clockwise or counterclockwise motion. This equation applies more generally, with the caveat that  $F$  should only include the part of the force perpendicular to the radius line.

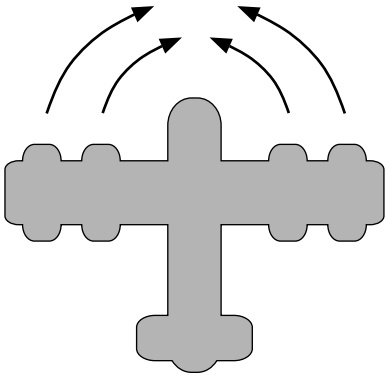
*self-check A*

There are four equations on this page. Which ones are important, and likely to be useful later? ▷ Answer, p. 179

To summarize, we've learned three conserved quantity, each of which has a rate of transfer:

conserved quantity		rate of transfer	
name	units	name	units
energy	joules (J)	power	watts (W)
momentum	kg·m/s	force	newtons (N)
angular momentum	kg·m <sup>2</sup> /s	torque	newton-meters (N·m)





f / The plane's four engines produce zero total torque but not zero total force.

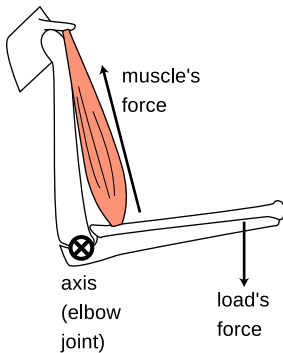
### Torque distinguished from force

Of course a force is necessary in order to create a torque — you can't twist a screw without pushing on the wrench — but force and torque are two different things. One distinction between them is direction. We use positive and negative signs to represent forces in the two possible directions along a line. The direction of a torque, however, is clockwise or counterclockwise, not a linear direction.

The other difference between torque and force is a matter of leverage. A given force applied at a door's knob will change the door's angular momentum twice as rapidly as the same force applied halfway between the knob and the hinge. The same amount of force produces different amounts of torque in these two cases.

It is possible to have a zero total torque with a nonzero total force. An airplane with four jet engines, f, would be designed so that their forces are balanced on the left and right. Their forces are all in the same direction, but the clockwise torques of two of the engines are canceled by the counterclockwise torques of the other two, giving zero total torque.

Conversely, we can have zero total force and nonzero total torque. A merry-go-round's engine needs to supply a nonzero torque on it to bring it up to speed, but there is zero total force on it. If there was not zero total force on it, its center of mass would accelerate!



g / Example 5: the biceps muscle flexes the arm.

#### A lever

#### example 5

Figure g shows an example of a lever within your arm. Different muscles are used to flex and extend the arm, because muscles work only by contraction. The biceps flexes it.

There are three forces acting on the forearm: the force from the biceps, the force at the elbow joint, and the force from the load being lifted. Because the elbow joint is motionless, it is natural to define our torques using the joint as the axis. The situation now becomes quite simple, because the upper arm bone's force exerted at the elbow has  $r = 0$ , and therefore creates no torque. We can ignore it completely. In general, we would call this the fulcrum of the lever.

If we restrict ourselves to the case in which the forearm rotates with constant angular momentum, then we know that the total torque on the forearm is zero, so the torques from the muscle and the load must be opposite in sign and equal in absolute value:

$$r_{\text{muscle}}F_{\text{muscle}} = r_{\text{load}}F_{\text{load}},$$

where  $r_{\text{muscle}}$ , the distance from the elbow joint to the biceps' point of insertion on the forearm, is only a few cm, while  $r_{\text{load}}$  might be 30 cm or so. The force exerted by the muscle must therefore be about ten times the force exerted by the load. We thus see that this lever is a force reducer. In general, a lever may

be used either to increase or to reduce a force.

Why did our arms evolve so as to reduce force? In general, your body is built for compactness and maximum speed of motion rather than maximum force. This is the main anatomical difference between us and the Neanderthals (their brains covered the same range of sizes as those of modern humans), and it seems to have worked for us.

As with all machines, the lever is incapable of changing the amount of mechanical work we can do. A lever that increases force will always reduce motion, and vice versa, leaving the amount of work unchanged.

### Discussion questions

**A** You whirl a rock over your head on the end of a string, and gradually pull in the string, eventually cutting the radius in half. What happens to the rock's angular momentum? What changes occur in its speed, the time required for one revolution, and its acceleration? Why might the string break?

**B** A helicopter has, in addition to the huge fan blades on top, a smaller propeller mounted on the tail that rotates in a vertical plane. Why?

**C** The photo shows an amusement park ride whose two cars rotate in opposite directions. Why is this a good design?

## 3.3 Noether's Theorem for Angular Momentum

Suppose a sunless planet is sitting all by itself in interstellar space, not rotating. Then, one day, it decides to start spinning. This doesn't necessarily violate conservation of energy, because it could have energy stored up, e.g., the heat in a molten core, which could be converted into kinetic energy. It does violate conservation of angular momentum, but even if we didn't already know about that law of physics, the story would seem odd. How would it decide which axis to spin around? If it was to spontaneously start spinning about some axis, then that axis would have to be a special, preferred direction in space. That is, space itself would have to have some asymmetry to it.

In reality, as I've already mentioned on page 15, experiments show to a very high degree of precision that the laws of physics are completely symmetric with respect to different directions. The story of the planet that abruptly starts spinning is an example of Noether's theorem, applied to angular momentum. We now have three such examples:

<b>symmetry</b>		<b>conserved quantity</b>
time symmetry	⇒	mass-energy
translation symmetry	⇒	momentum
rotational symmetry	⇒	angular momentum



Discussion question C.



## Problems

### Key

- ✓ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ★ A difficult problem.

**1** You are trying to loosen a stuck bolt on your RV using a big wrench that is 50 cm long. If you hang from the wrench, and your mass is 55 kg, what is the maximum torque you can exert on the bolt? ✓

**2** A physical therapist wants her patient to rehabilitate his injured elbow by laying his arm flat on a table, and then lifting a 2.1 kg mass by bending his elbow. In this situation, the weight is 33 cm from his elbow. He calls her back, complaining that it hurts him to grasp the weight. He asks if he can strap a bigger weight onto his arm, only 17 cm from his elbow. How much mass should she tell him to use so that he will be exerting the same torque? (He is raising his forearm itself, as well as the weight.) ✓

**3** An object is observed to have constant angular momentum. Can you conclude that no torques are acting on it? Explain. [Based on a problem by Serway and Faughn.]

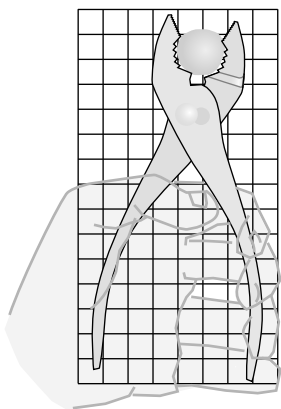
**4** The figure shows scale drawing of a pair of pliers being used to crack a nut, with an appropriately reduced centimeter grid. Warning: do not attempt this at home; it is bad manners. If the force required to crack the nut is 300 N, estimate the force required of the person's hand. ▷ Solution, p. 180

**5** Two horizontal tree branches on the same tree have equal diameters, but one branch is twice as long as the other. Give a quantitative comparison of the torques where the branches join the trunk. [Thanks to Bong Kang.]

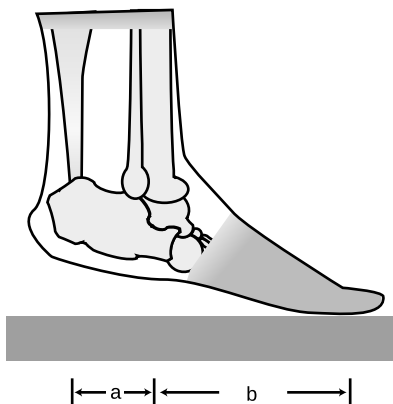
**6** (a) Alice says Cathy's body has zero momentum, but Bob says Cathy's momentum is nonzero. Nobody is lying or making a mistake. How is this possible? Give a concrete example.

(b) Alice and Bob agree that Dong's body has nonzero momentum, but disagree about Dong's angular momentum, which Alice says is zero, and Bob says is nonzero. Explain.

**7** A person of weight  $W$  stands on the ball of one foot. Find the tension in the calf muscle and the force exerted by the shinbones on the bones of the foot, in terms of  $W$ ,  $a$ , and  $b$ . (The tension is a measure of how tight the calf muscle has been pulled; it has units of newtons, and equals the amount of force applied by the muscle where it attaches to the heel.) For simplicity, assume that all the forces are at 90-degree angles to the foot. Suggestion: Write down an equation that says the total force on the foot is zero, and another equation saying that the total torque on the foot is zero; solve the



Problem 4.



Problem 7.

two equations for the two unknowns.

✓





The clock took up two seats, and two tickets were bought for it under the name of "Mr. Clock."

## Chapter 4

# Relativity

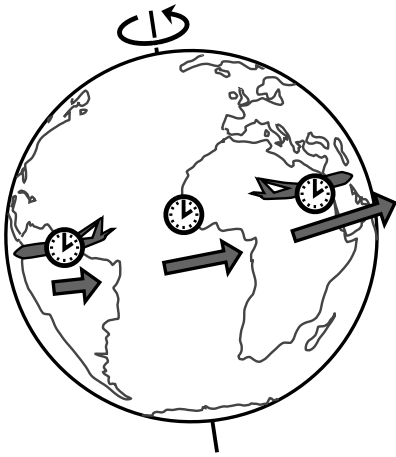
### 4.1 Relativity According To Einstein

#### Time is not absolute

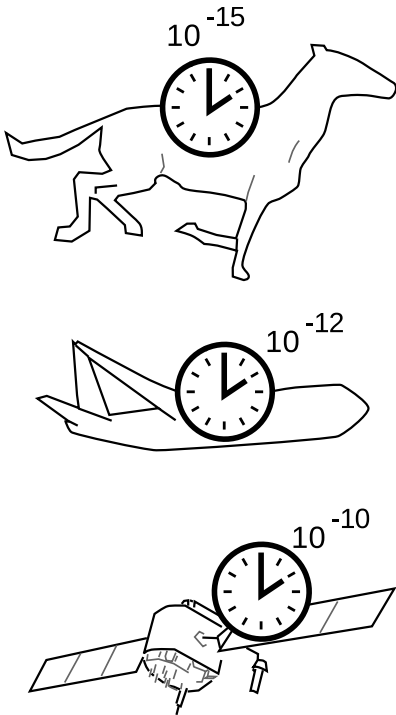
So far we've been discussing relativity according to Galileo and Newton, but there is also relativity according to Einstein. When Einstein first began to develop the theory of relativity, around 1905, the only real-world observations he could draw on were ambiguous and indirect. Today, the evidence is part of everyday life. For example, every time you use a GPS receiver, a, you're using Einstein's theory of relativity. Somewhere between 1905 and today, technology became good enough to allow conceptually *simple* experiments that students in the early 20th century could only discuss in terms like "Imagine that we could..." A good jumping-on point is 1971. In that year, J.C. Hafele and R.E. Keating, shown in the photo above, brought atomic clocks aboard commercial airliners, and went around the world, once from east to west and once from west to east. The clocks were capable of keeping time to within a few nanoseconds. (A nanosecond, abbreviated ns, is one billionth of a second.) Hafele



a / This Global Positioning System (GPS) system, running on a smartphone attached to a bike's handlebar, depends on Einstein's theory of relativity. Time flows at a different rates aboard a GPS satellite than it does on the bike, and the GPS software has to take this into account.



b / All three clocks are moving to the east. Even though the west-going plane is moving to the west relative to the air, the air is moving to the east due to the earth's rotation.



c / The correspondence principle requires that the relativistic distortion of time become small for small velocities. The effects are so small that we have to describe them in scientific notation (p. 12). For example,  $10^{-15}$  means 0.000000000000001, which is a *hundred thousand times* smaller than  $10^{-10}$ .

and Keating observed that there was a discrepancy between the times measured by the traveling clocks and the times measured by similar clocks that stayed home at the U.S. Naval Observatory in Washington. The east-going clock lost time, ending up off by  $-59$  nanoseconds, while the west-going one gained 273 ns.

### Causality

It is reassuring that the effects on time were small compared to the three-day lengths of the plane trips. There was therefore no opportunity for paradoxical scenarios such as one in which the east-going experimenter arrived back in Washington before he left and then convinced himself not to take the trip. A theory that maintains this kind of orderly relationship between cause and effect is said to satisfy causality.

### Time affected by motion and gravity

Hafele and Keating were testing specific quantitative predictions of relativity, and they verified them to within their experiment's error bars. Let's work backward instead, and inspect the empirical results for clues as to how time works.

The two traveling clocks experienced effects in opposite directions, and this suggests that the rate at which time flows depends on the motion of the observer. The east-going clock was moving in the same direction as the earth's rotation, so its velocity relative to the earth's center was greater than that of the clock that remained in Washington, while the west-going clock's velocity was correspondingly reduced. The fact that the east-going clock fell behind, and the west-going one got ahead, shows that the effect of motion is to make time go more slowly. This effect of motion on time was predicted by Einstein in his original 1905 paper on relativity, written when he was 26.

If this had been the only effect in the Hafele-Keating experiment, then we would have expected to see effects on the two flying clocks that were equal in size. In fact, the two effects are unequal in size:  $-59$  ns and 273 ns. This turns out to be because there was a second effect involved, a gravitational one, simply due to the planes' being up in the air. The gravitational effects are beyond the scope of this book.

### The correspondence principle

The effects that Hafele and Keating observed were small. This makes sense: the version of relativity worked out by Galileo (sections 2.2 and 2.3, pp. 39-53) had already been thoroughly tested by experiments under a wide variety of conditions, so a new theory like Einstein's relativity must agree with Galileo's to a good approximation, within the Galilean theory's realm of applicability. This is an example of the correspondence principle (p. 31). The behavior of

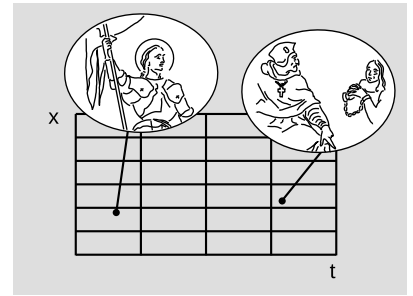
the three clocks in the Hafele-Keating experiment shows that the amount of time distortion increases as the speed of the clock's motion increases. Newton lived in an era when the fastest mode of transportation was a galloping horse, and the best pendulum clocks would accumulate errors of perhaps a minute over the course of several days. A horse is much slower than a jet plane, so the distortion of time would have had a relative size of only  $\sim 10^{-15}$  — much smaller than the clocks were capable of detecting. At the speed of a passenger jet, the effect is about  $10^{-12}$ , and state-of-the-art atomic clocks in 1971 were capable of measuring that. A GPS satellite travels much faster than a jet airplane, and the effect on the satellite turns out to be  $\sim 10^{-10}$ . The general idea here is that all physical laws are approximations, and approximations aren't simply right or wrong in different situations. Approximations are better or worse in different situations, and the question is whether a particular approximation is good enough in a given situation to serve a particular purpose. The faster the motion, the worse the Newtonian approximation of absolute time. Whether the approximation is good enough depends on what you're trying to accomplish. The correspondence principle says that the approximation must have been good enough to explain all the experiments done in the centuries before Einstein came up with relativity.

By the way, don't get an inflated idea of the importance of the Hafele-Keating experiment. Relativity had already been confirmed by a vast and varied body of experiments decades before 1971. The only reason I'm giving such a prominent role to this experiment is that it is conceptually very direct.

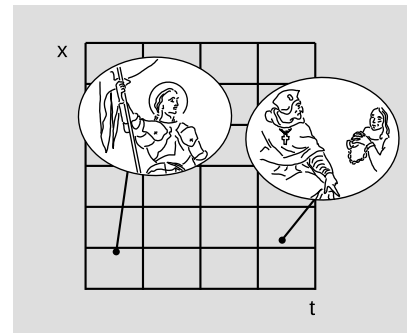
### Distortion of time and space

Relativity says that when two observers are in different frames of reference, each observer considers the other one's perception of time to be distorted. We'll also see that something similar happens to their observations of distances, so both space and time are distorted. What exactly is this distortion? How do we even conceptualize it?

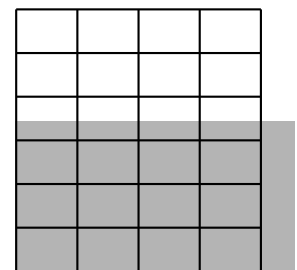
The idea isn't really as radical as it might seem at first. We can visualize the structure of space and time using a graph with position and time on its axes. These graphs were introduced on p. 43 in figures d and e, but we're going to look at them in a slightly different way. Before, we used them to describe the motion of objects. The grid underlying the graph was merely the stage on which the actors played their parts. Now the background comes to the foreground: it's time and space themselves that we're studying. We don't necessarily need to have a line or a curve drawn on top of the grid to represent a particular object. We may, for example, just want to talk about events, depicted as points on the graph as in figure d. A distortion of the Cartesian grid underlying the graph can arise for perfectly ordinary reasons that Isaac Newton would have read-



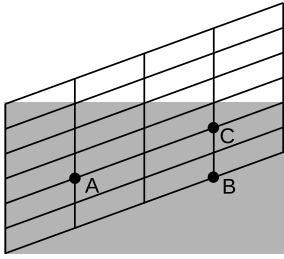
d / Two events are given as points on a graph of position versus time. Joan of Arc helps to restore Charles VII to the throne. At a later time and a different position, Joan of Arc is sentenced to death.



e / A change of units distorts an  $x$ - $t$  graph. This graph depicts exactly the same events as figure d. The only change is that the  $x$  and  $t$  coordinates are measured using different units, so the grid is compressed in  $t$  and expanded in  $x$ .



f / A convention we'll use to represent a distortion of time and space.



g / A Galilean version of the relationship between two frames of reference, as introduced in figure e, p. 43. As in all such graphs in this chapter, the original coordinates, represented by the gray rectangle, have a time axis that goes to the right, and a distance axis that goes straight up.

ily accepted. For example, we can simply change the units used to measure time and position, as in figure e.

We’re going to have quite a few examples of this type, so I’ll adopt the convention shown in figure f for depicting them; this convention was originally introduced in figure e on p. 43. Figure f summarizes the relationship between figures d and e in a more compact form. The gray rectangle represents the original coordinate grid of figure d, while the grid of black lines represents the new version from figure e. Omitting the grid from the gray rectangle makes the diagram easier to decode visually.

Our goal of unraveling the mysteries of special relativity amounts to nothing more than finding out how to draw a diagram like f in the case where the two different sets of coordinates represent measurements of time and space made by two different observers, each in motion relative to the other. Galileo and Newton thought they knew the answer to this question, but their answer turned out to be only approximately right. To avoid repeating the same mistakes, we need to clearly spell out what we think are the basic properties of time and space that will be a reliable foundation for our reasoning.

Experiments show that:

1. The laws of physics have translation symmetry (section 2.1), time symmetry (section 1.6), and rotational symmetry (p. 15 and section 3.3).
2. The principle of inertia holds (p. 16).
3. Causality holds, in the sense described on page 72.
4. Time depends on the state of motion of the observer.

If it were not for property 4, we could imagine that figure g would give the correct transformation between frames of reference in motion relative to one another. Let’s say that observer 1, whose grid coincides with the gray rectangle, is a hitch-hiker standing by the side of a road. Event A is a raindrop hitting his head, and



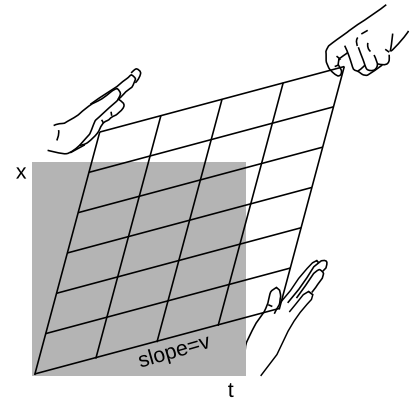
event B is another raindrop hitting his head. He says that A and B occur at the same location in space. Observer 2 is a motorist who drives by without stopping; to him, the passenger compartment of his car is at rest, while the asphalt slides by underneath. He says that A and B occur at different points in space, because during the time between the first raindrop and the second, the hitch-hiker has moved backward. On the other hand, observer 2 says that events A and C occur in the same place, while the hitch-hiker disagrees. The slope of the grid-lines is simply the velocity of the relative motion of each observer relative to the other. (Recall that slope is defined as the rise over the run. On these graphs of distance versus time, the slope is the distance traveled divided by the elapsed time.)

Figure g has familiar, comforting, and eminently sensible behavior, but it also happens to be wrong, because it violates property 4. The distortion of the coordinate grid has only moved the vertical lines up and down, so both observers agree that events like B and C are simultaneous. If this was really the way things worked, then all observers could synchronize all their clocks with one another for once and for all, and the clocks would never get out of sync. This contradicts the results of the Hafele-Keating experiment, in which all three clocks were initially synchronized in Washington, but later went out of sync because of their different states of motion.

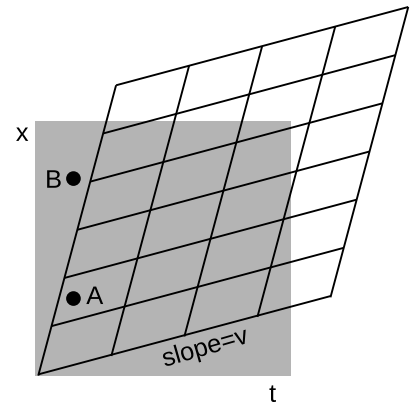
Based on properties 1-4, there is only one possible way to modify g, which is the one shown in h.<sup>1</sup> This distortion is the one that Einstein predicted in 1905, and is known as the Lorentz transformation, after Hendrik Lorentz (1853-1928). The distortion is a kind of smooching and stretching, as suggested by the hands. Also, we've already seen in figures d-f on page 73 that we're free to stretch or compress everything as much as we like in the horizontal and vertical directions, because this simply corresponds to choosing different units of measurement for time and distance. In figure h I've chosen units that give the whole drawing a convenient symmetry about a 45-degree diagonal line. Ordinarily it wouldn't make sense to talk about a 45-degree angle on a graph whose axes had different units. But in relativity, the symmetric appearance of the transformation tells us that space and time ought to be treated on the same footing, and measured in the same units.

The exact size and shape of the parallelogram are controlled by the requirements that (i) the slope labeled in the figure corresponds properly to the velocity; (ii) the units are the special ones described

<sup>1</sup>For a proof that no other version is possible, see ch. 23 of my free online book *Light and Matter*.



h / In the units that are most convenient for relativity, the transformation has symmetry about a 45-degree diagonal line.



i / Interpretation of the Lorentz transformation. The slope indicated in the figure gives the relative velocity of the two frames of reference. Events A and B that were simultaneous in frame 1 are not simultaneous in frame 2, where event A occurs to the right of the  $t = 0$  line represented by the left edge of the grid, but event B occurs to its left.

above; and (iii) the area of the parallelogram is the same as the area of the original square.<sup>2</sup>

---

<sup>2</sup>The equal-area property is proved in *Light and Matter*.

## The $\gamma$ factor

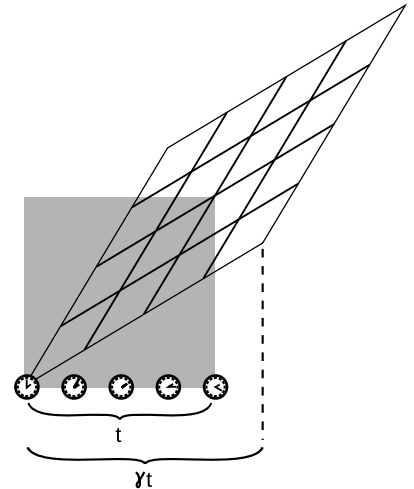
We've seen the experimental evidence that motion changes the rate of flow of time, and this effect is correctly reproduced by the Lorentz transformation.

### Time dilation

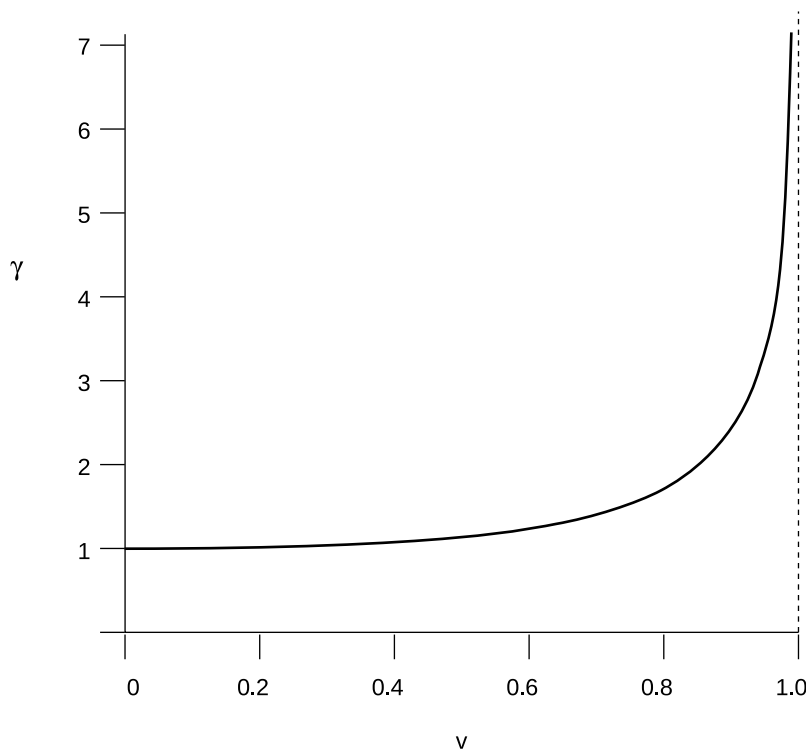
A clock runs fastest in the frame of reference of an observer who is at rest relative to the clock.

We define the factor  $\gamma$  (Greek letter gamma) as in figure j. An observer in motion relative to the clock at speed  $v$  perceives the clock as running more slowly by a factor of  $\gamma$ . For example, if  $\gamma$  equals 2, then the observer says the clock runs at half its normal speed.

Figure k shows an example of how we can use properties (i)-(iii) on p. 75 to find the value of  $\gamma$  for a given velocity  $v$  of the clock and the observer relative to one another. By plotting many such points,<sup>3</sup> we get the graph shown in figure l.



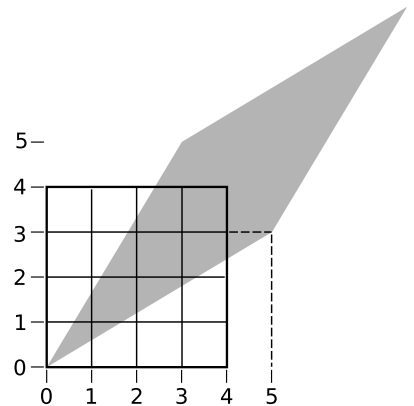
j / The  $\gamma$  factor.



l / The behavior of the  $\gamma$  factor. (The velocity is in the special units described on p. 75. More on these units in section 4.2.)

For small velocities, the graph is nearly flat at  $\gamma \approx 1$ , meaning that

<sup>3</sup>To avoid the tedious work of drawing many figures like k, one can use algebra and geometry to derive the equation  $\gamma = 1/\sqrt{1 - v^2}$ .



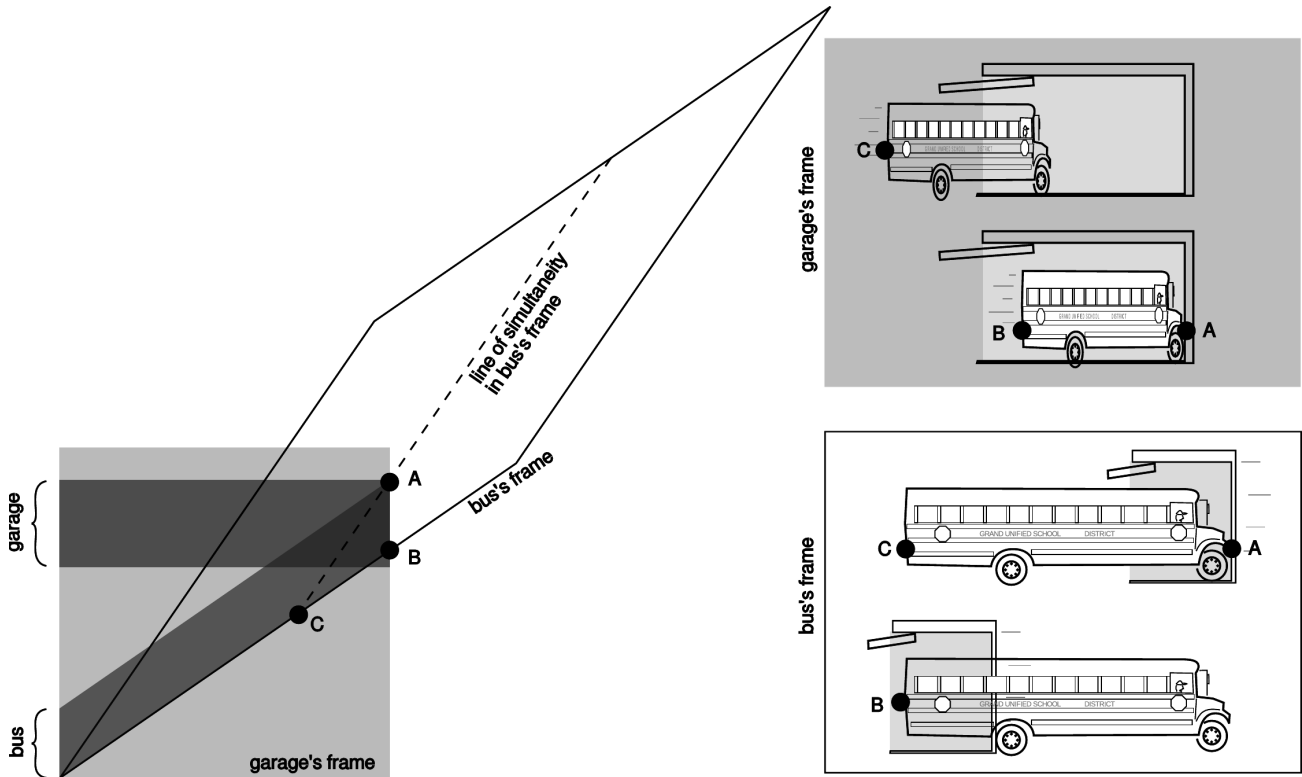
k / An example in which the  $\gamma$  factor is numerically simple to work out. The bottom edge of the parallelogram rises 3 units and goes 5 units to the right, so its slope is  $3/5$ . This slope represents the speed of one frame of reference relative to the other. It's easy to verify that the square and the parallelogram have the same area, because one diagonal of the square has been stretched to twice its original length, the other smushed down by a half. The bottom-right corner of the square is at a time of 4 units, while the corresponding corner of the parallelogram is at 5. As defined in figure j, the ratio of these times is the value of  $\gamma = 5/4$ .

there is very little time dilation. This is required by the correspondence principle.

Distances are also distorted:

**Length contraction**

A meter-stick appears longest to an observer who is at rest relative to it. An observer moving relative to the meter-stick at  $v$  observes the stick to be shortened by a factor of  $\gamma$ .



m / Example 1: In the garage's frame of reference, the bus is moving, and can fit in the garage due to its length contraction. In the bus's frame of reference, the garage is moving, and can't hold the bus due to its length contraction.

*The garage paradox* example 1

One of the most famous of all the so-called relativity paradoxes has to do with our incorrect feeling that simultaneity is well defined. The idea is that one could take a schoolbus and drive it at relativistic speeds into a garage of ordinary size, in which it normally would not fit. Because of the length contraction, the bus would supposedly fit in the garage. The driver, however, will perceive the *garage* as being contracted and thus even less able to contain the bus.

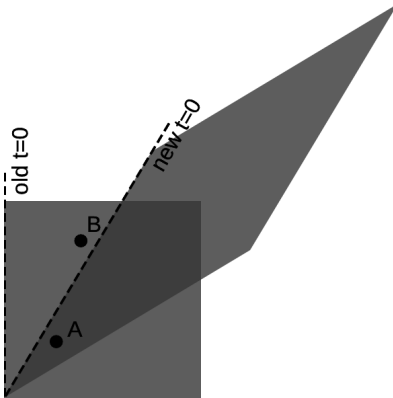
The paradox is resolved when we recognize that the concept of fitting the bus in the garage "all at once" contains a hidden as-

sumption, the assumption that it makes sense to ask whether the front and back of the bus can *simultaneously* be in the garage. Observers in different frames of reference moving at high relative speeds do not necessarily agree on whether things happen simultaneously. As shown in figure m, the person in the garage's frame can shut the door at an instant B he perceives to be simultaneous with the front bumper's arrival A at the back wall of the garage, but the driver would not agree about the simultaneity of these two events, and would perceive the door as having shut long after she plowed through the back wall.

## 4.2 Speeds In Relativity

### The universal speed $c$

Let's think a little more about the role of the 45-degree diagonal in the Lorentz transformation. Slopes on these graphs are interpreted as velocities. This line has a slope of 1 in our special relativistic units, but that slope corresponds to some number, call it  $c$ , in ordinary units of meters per second. Now note what happens when we perform a Lorentz transformation: this particular line gets stretched, but the new version of the line lies right on top of the old one, and its slope stays the same. In other words, if one observer says that something has a velocity equal to  $c$ , every other observer will agree on that velocity as well.



A proof that causality imposes a universal speed limit. In the original frame of reference, represented by the square, event A happens a little before event B. In the new frame, shown by the parallelogram, A happens after  $t = 0$ , but B happens before  $t = 0$ ; that is, B happens before A. The time ordering of the two events has been reversed. This can only happen because events A and B are very close together in time and fairly far apart in space. The line segment connecting A and B has a slope greater than 1, meaning that if we wanted to be present at both events, we would have to travel at a speed greater than  $c$  (which equals 1 in the units used on this graph). You will find that if you pick any two points for which the slope of the line segment connecting them is less than 1, you can never get them to straddle the new  $t = 0$  line in this funny, time-reversed way. Since different observers disagree on the time order of events like A and B, causality requires that information never travel from A to B or from B to A; if it did, then we would have time-travel paradoxes. The conclusion is that  $c$  is the maximum speed of cause and effect in relativity.

### Velocities don't simply add and subtract.

This is counterintuitive, since we expect velocities to add and subtract in relative motion. If a dog is running away from me at 5 m/s relative to the sidewalk, and I run after it at 3 m/s, the dog's velocity in my frame of reference is 2 m/s. According to everything we have learned about motion, the dog must have different speeds in the two frames: 5 m/s in the sidewalk's frame and 2 m/s in mine. But velocities are measured by dividing a distance by a time, and both distance and time are distorted by relativistic effects, so we actually shouldn't expect the ordinary arithmetic addition of velocities to hold in relativity; it's an approximation that's valid at velocities that are small compared to  $c$ .

### A universal speed limit

For example, suppose Janet takes a trip in a spaceship, and accelerates until she is moving at  $0.6c$  relative to the earth. She then launches a space probe in the forward direction at a speed relative to her ship of  $0.6c$ . We might think that the probe was then moving at a velocity of  $1.2c$ , but in fact the answer is still less than  $c$  (problem 1, page 89). This is an example of a more general fact about relativity, which is that  $c$  represents a universal speed limit. This is required by causality, as shown in figure n.

### Light travels at $c$ .

Now consider a beam of light. We're used to talking casually about the "speed of light," but what does that really mean? Motion is relative, so normally if we want to talk about a velocity, we have to specify what it's measured relative to. A sound wave has a certain speed relative to the air, and a water wave has its own speed relative to the water. If we want to measure the speed of an ocean wave, for example, we should make sure to measure it in a frame of reference at rest relative to the water. But light isn't a vibration of a physical medium; it can propagate through the near-perfect vacuum of outer space, as when rays of sunlight travel to earth. This seems like a

paradox: light is supposed to have a specific speed, but there is no way to decide what frame of reference to measure it in. The way out of the paradox is that light must travel at a velocity equal to  $c$ . Since all observers agree on a velocity of  $c$ , regardless of their frame of reference, everything is consistent.

### The Michelson-Morley experiment

The constancy of the speed of light had in fact already been observed when Einstein was an 8-year-old boy, but because nobody could figure out how to interpret it, the result was largely ignored. In 1887 Michelson and Morley set up a clever apparatus to measure any difference in the speed of light beams traveling east-west and north-south. The motion of the earth around the sun at 110,000 km/hour (about 0.01% of the speed of light) is to our west during the day. Michelson and Morley believed that light was a vibration of a mysterious medium called the ether, so they expected that the speed of light would be a fixed value relative to the ether. As the earth moved through the ether, they thought they would observe an effect on the velocity of light along an east-west line. For instance, if they released a beam of light in a westward direction during the day, they expected that it would move away from them at less than the normal speed because the earth was chasing it through the ether. They were surprised when they found that the expected 0.01% change in the speed of light did not occur.

#### *The ring laser gyroscope*

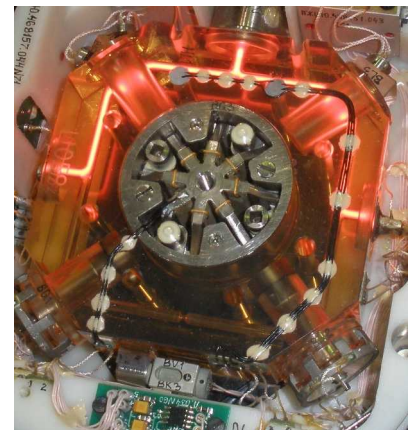
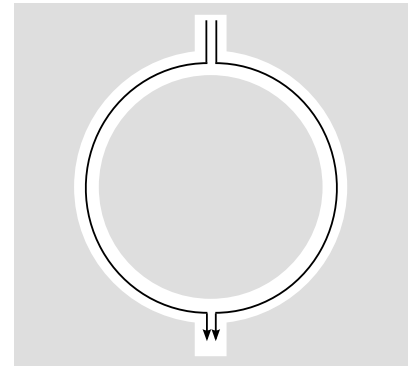
*example 2*

If you've flown in a jet plane, you can thank relativity for helping you to avoid crashing into a mountain or an ocean. Figure 0 shows a standard piece of navigational equipment called a ring laser gyroscope. A beam of light is split into two parts, sent around the perimeter of the device, and reunited. Since the speed of light is constant, we expect the two parts to come back together at the same time. If they don't, it's evidence that the device has been rotating. The plane's computer senses this and notes how much rotation has accumulated.

#### *No frequency-dependence*

*example 3*

Relativity has only one universal speed, so it requires that all light waves travel at the same speed, regardless of their frequency and wavelength. Presently the best experimental tests of the invariance of the speed of light with respect to wavelength come from astronomical observations of gamma-ray bursts, which are sudden outpourings of high-frequency light, believed to originate from a supernova explosion in another galaxy. One such observation, in 2009,<sup>4</sup> found that the times of arrival of all the different frequencies in the burst differed by no more than 2 seconds out of a total time in flight on the order of ten billion years!



o / A ring laser gyroscope.

<sup>4</sup><http://arxiv.org/abs/0908.1832>

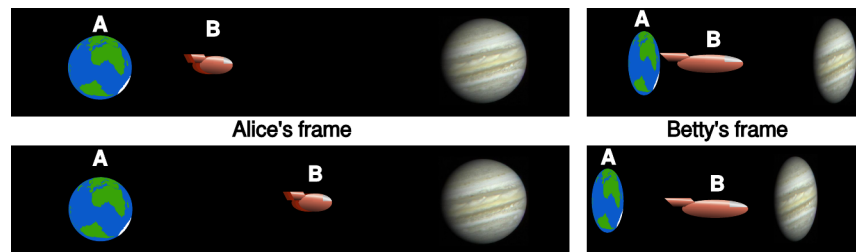
*An interstellar road trip*

*example 4*

Because the distances between the stars are so vast, it's convenient to measure them in light-years rather than kilometers. A light-year is defined as the distance traveled by light in one year. If we adopt the year as our unit of time, and the light-year as our unit of distance, then the speed of light is 1, i.e., these units qualify as the kind of "special units" that we've been assuming in all the graphs.

Suppose that Alice stays on earth while her twin Betty heads off in a spaceship for Tau Ceti, a nearby star. Tau Ceti is 12 light-years away, so even though Betty travels at 87% of the speed of light, it will take her a long time to get there: 14 years, according to Alice.

p / Example 4.



Betty experiences time dilation. At this speed, her  $\gamma$  is 2.0, so that the voyage will only seem to her to last 7 years. But there is perfect symmetry between Alice's and Betty's frames of reference, so Betty agrees with Alice on their relative speed; Betty sees herself as being at rest, while the sun and Tau Ceti both move backward at 87% of the speed of light. How, then, can she observe Tau Ceti to get to her in only 7 years, when it should take 14 years to travel 12 light-years at this speed?

We need to take into account length contraction. Betty sees the distance between the sun and Tau Ceti to be shrunk by a factor of 2. The same thing occurs for Alice, who observes Betty and her spaceship to be foreshortened.



## Discussion questions

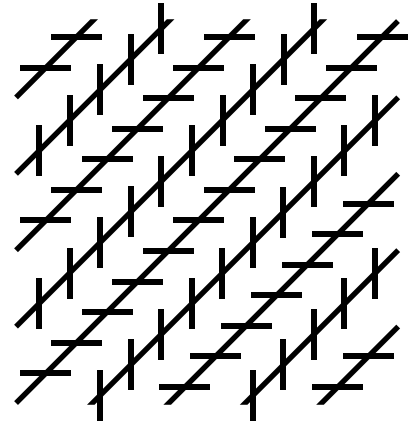
**A** A person in a spaceship moving at 99.99999999% of the speed of light relative to Earth shines a flashlight forward through dusty air, so the beam is visible. What does she see? What would it look like to an observer on Earth?

**B** A question that students often struggle with is whether time and space can really be distorted, or whether it just seems that way. Compare with optical illusions or magic tricks. How could you verify, for instance, that the lines in the figure are actually parallel? Are relativistic effects the same, or not?

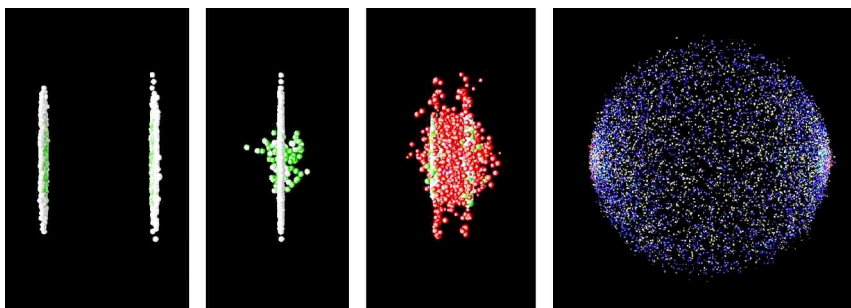
**C** On a spaceship moving at relativistic speeds, would a lecture seem even longer and more boring than normal?

**D** Mechanical clocks can be affected by motion. For example, it was a significant technological achievement to build a clock that could sail aboard a ship and still keep accurate time, allowing longitude to be determined. How is this similar to or different from relativistic time dilation?

**E** The figure shows an artist's rendering of the length contraction for the collision of two gold nuclei at relativistic speeds in the RHIC accelerator in Long Island, New York, which went on line in 2000. The gold nuclei would appear nearly spherical (or just slightly lengthened like an American football) in frames moving along with them, but in the laboratory's frame, they both appear drastically foreshortened as they approach the point of collision. The later pictures show the nuclei merging to form a hot soup, in which experimenters hope to observe a new form of matter. What would the shapes of the two nuclei look like to a microscopic observer riding on the left-hand nucleus? To an observer riding on the right-hand one? Can they agree on what is happening? If not, why not — after all, shouldn't they see the same thing if they both compare the two nuclei side-by-side at the same instant in time?



Discussion question B

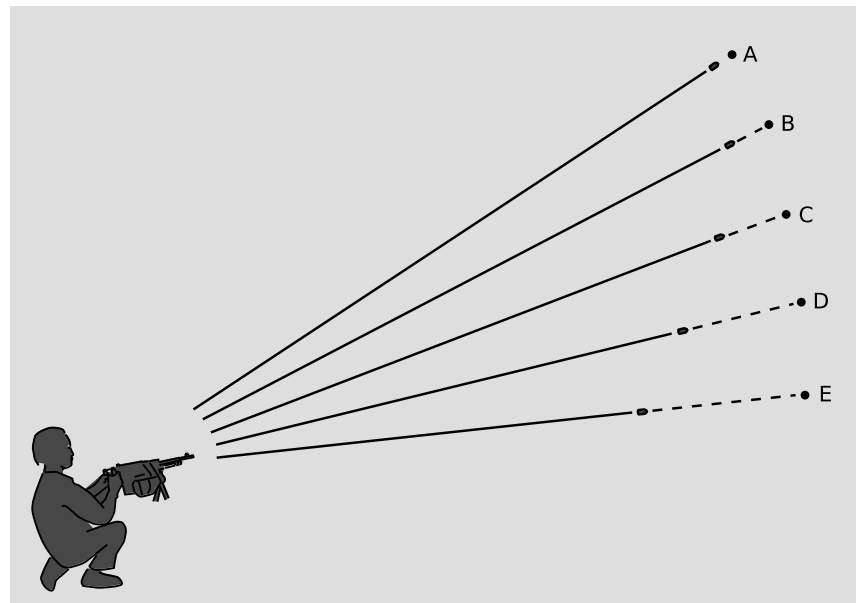


Discussion question E: colliding nuclei show relativistic length contraction.

**F** If you stick a piece of foam rubber out the window of your car while driving down the freeway, the wind may compress it a little. Does it make sense to interpret the relativistic length contraction as a type of strain that pushes an object's atoms together like this? How does this relate to discussion question E?

**G** The machine-gunner in the figure sends out a spray of bullets. Suppose that the bullets are being shot into outer space, and that the distances traveled are trillions of miles (so that the human figure in the diagram is not to scale). After a long time, the bullets reach the points shown with dots which are all equally far from the gun. Their arrivals at those points are events A through E, which happen at different times. Sketch these events on a position-time graph. The chain of impacts extends across space at a speed greater than  $c$ . Does this violate special relativity?

Discussion question G.

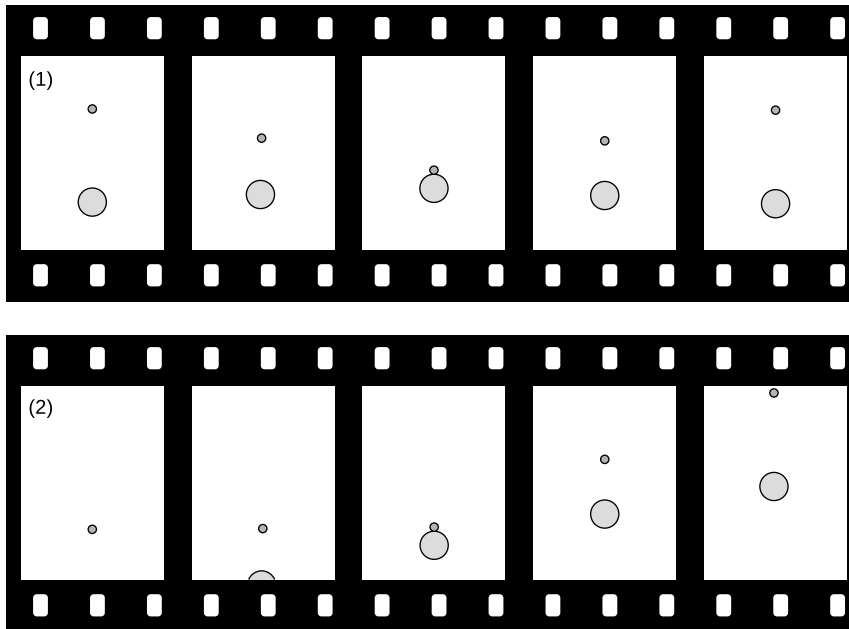


### 4.3 Dynamics

So far we have said nothing about how to predict motion in relativity. Do Newton's laws still work? Do conservation laws still apply? The answer is yes, but many of the definitions need to be modified, and certain entirely new phenomena occur, such as the conversion of mass to energy and energy to mass, as described by the famous equation  $E = mc^2$ .

#### Momentum

Here's a flawed scheme for traveling faster than the speed of light. The basic idea can be demonstrated by dropping a ping-pong ball and a baseball stacked on top of each other like a snowman. They separate slightly in mid-air, and the baseball therefore has time to hit the floor and rebound before it collides with the ping-pong ball, which is still on the way down. The result is a surprise if you haven't seen it before: the ping-pong ball flies off at high speed and hits the ceiling! A similar fact is known to people who investigate the scenes of accidents involving pedestrians. If a car moving at 90 kilometers per hour hits a pedestrian, the pedestrian flies off at nearly double that speed, 180 kilometers per hour. Now suppose the car was moving at 90 percent of the speed of light. Would the pedestrian fly off at 180% of  $c$ ?



1 / An unequal collision, viewed in the center-of-mass frame, 1, and in the frame where the small ball is initially at rest, 2. The motion is shown as it would appear on the film of an old-fashioned movie camera, with an equal amount of time separating each frame from the next. Film 1 was made by a camera that tracked the center of mass, film 2 by one that was initially tracking the small ball, and kept on moving at the same speed after the collision.

To see why not, we have to back up a little and think about where this speed-doubling result comes from. For any collision, there is a special frame of reference, the center-of-mass frame, in which the two colliding objects approach each other, collide, and rebound with their velocities reversed. In the center-of-mass frame, the total

momentum of the objects is zero both before and after the collision.

Figure r/1 shows such a frame of reference for objects of very unequal mass. Before the collision, the large ball is moving relatively slowly toward the top of the page, but because of its greater mass, its momentum cancels the momentum of the smaller ball, which is moving rapidly in the opposite direction. The total momentum is zero. After the collision, the two balls just reverse their directions of motion. We know that this is the right result for the outcome of the collision because it conserves both momentum and kinetic energy, and everything not forbidden is compulsory, i.e., in any experiment, there is only one possible outcome, which is the one that obeys all the conservation laws.

*self-check A*

How do we know that momentum and kinetic energy are conserved in figure r/1? ▷ Answer, p. 179

Let's make up some numbers as an example. Say the small ball has a mass of 1 kg, the big one 8 kg. In frame 1, let's make the velocities as follows:

	before the collision	after the collision
◦	-0.8	0.8
●	0.1	-0.1

Figure r/2 shows the same collision in a frame of reference where the small ball was initially at rest. To find all the velocities in this frame, we just add 0.8 to all the ones in the previous table.

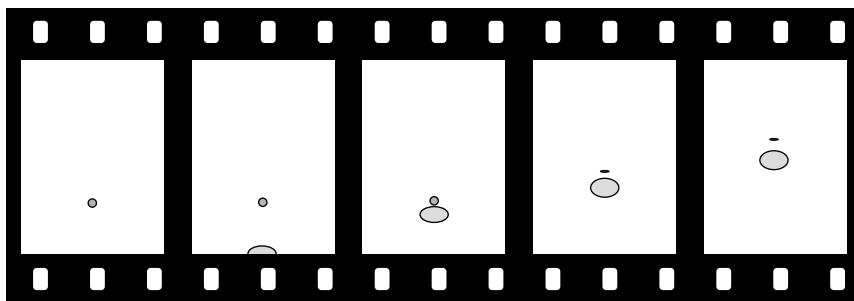
	before the collision	after the collision
◦	0	1.6
●	0.9	0.7

In this frame, as expected, the small ball flies off with a velocity, 1.6, that is almost twice the initial velocity of the big ball, 0.9.

If all those velocities were in meters per second, then that's exactly what happened. But what if all these velocities were in units of the speed of light? Now it's no longer a good approximation just to add velocities. We need to combine them according to the relativistic rules. For instance, reasoning similar to that in problem 1, page 89 tells us that combining a velocity of 0.8 times the speed of light with another velocity of 0.8 results in 0.98, not 1.6. The results are very different:

	before the collision	after the collision
◦	0	0.98
●	0.83	0.76

We can interpret this as follows. Figure r/1 is one in which the big ball is moving fairly slowly. This is very nearly the way the scene would be seen by an ant standing on the big ball. According



s / An 8-kg ball moving at 83% of the speed of light hits a 1-kg ball. The balls appear foreshortened due to the relativistic distortion of space.

to an observer in frame  $s$ , however, both balls are moving at nearly the speed of light after the collision. Because of this, the balls appear foreshortened, but the distance between the two balls is also shortened. To this observer, it seems that the small ball isn't pulling away from the big ball very fast.

Now here's what's interesting about all this. The outcome shown in figure  $r/2$  was supposed to be the only one possible, the only one that satisfied both conservation of energy and conservation of momentum. So how can the *different* result shown in figure  $s$  be possible? The answer is that relativistically, momentum must not equal  $mv$ . The old, familiar definition is only an approximation that's valid at low speeds. If we observe the behavior of the small ball in figure  $s$ , it looks as though it somehow had some extra inertia. It's as though a football player tried to knock another player down without realizing that the other guy had a three-hundred-pound bag full of lead shot hidden under his uniform — he just doesn't seem to react to the collision as much as he should. This extra inertia is described by redefining momentum as

$$\text{momentum} = m\gamma v.$$

At very low velocities,  $\gamma$  is close to 1, and the result is very nearly  $mv$ , as demanded by the correspondence principle. But at very high velocities,  $\gamma$  gets very big — the small ball in figure  $s$  has a  $\gamma$  of 5.0, and therefore has five times more inertia than we would expect nonrelativistically.

### Equivalence of mass and energy

Now we're ready to see why mass and energy must be equivalent as claimed in the famous  $E = mc^2$ . So far we've only considered collisions in which none of the kinetic energy is converted into any other form of energy, such as heat or sound. Let's consider what happens if a blob of putty moving at velocity  $v$  hits another blob that is initially at rest, sticking to it. The nonrelativistic result is that to obey conservation of momentum the two blobs must fly off together at  $v/2$ . Half of the initial kinetic energy has been converted to heat.<sup>5</sup>

<sup>5</sup>A double-mass object moving at half the speed does not have the same kinetic energy. Kinetic energy depends on the square of the velocity, so cutting

Relativistically, however, an interesting thing happens. A hot object has more momentum than a cold object! This is because the relativistically correct expression for momentum is  $m\gamma v$ , and the more rapidly moving atoms in the hot object have higher values of  $\gamma$ . In our collision, the final combined blob must therefore be moving a little more slowly than the expected  $v/2$ , since otherwise the final momentum would have been a little greater than the initial momentum. To an observer who believes in conservation of momentum and knows only about the overall motion of the objects and not about their heat content, the low velocity after the collision would seem to be the result of a magical change in the mass, as if the mass of two combined, hot blobs of putty was more than the sum of their individual masses.

Now we know that the masses of all the atoms in the blobs must be the same as they always were. The change is due to the change in  $\gamma$  with heating, not to a change in mass. The heat energy, however, seems to be acting as if it was equivalent to some extra mass.

But this whole argument was based on the fact that heat is a form of kinetic energy at the atomic level. Would  $E = mc^2$  apply to other forms of energy as well? Suppose a rocket ship contains some electrical energy stored in a battery. If we believed that  $E = mc^2$  applied to forms of kinetic energy but not to electrical energy, then we would have to believe that the pilot of the rocket could slow the ship down by using the battery to run a heater! This would not only be strange, but it would violate the principle of relativity, because the result of the experiment would be different depending on whether the ship was at rest or not. The only logical conclusion is that all forms of energy are equivalent to mass. Running the heater then has no effect on the motion of the ship, because the total energy in the ship was unchanged; one form of energy (electrical) was simply converted to another (heat).

The equation  $E = mc^2$  tells us how much energy is equivalent to how much mass: the conversion factor is the square of the speed of light,  $c$ . Since  $c$  a big number, you get a really really big number when you multiply it by itself to get  $c^2$ . This means that even a small amount of mass is equivalent to a very large amount of energy.

We've already seen several examples of applications of  $E = mc^2$ , on page 30.

---

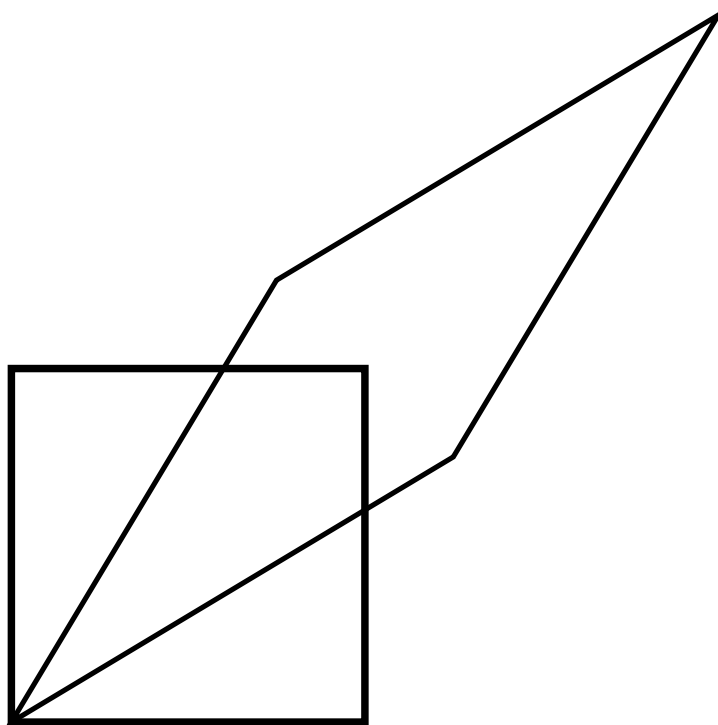
the velocity in half reduces the energy by a factor of 1/4, which, multiplied by the doubled mass, makes 1/2 the original energy.

## Problems

### Key

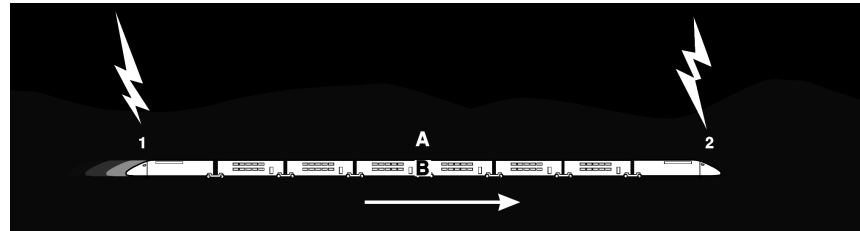
- ✓ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ★ A difficult problem.

1 The figure illustrates a Lorentz transformation using the conventions described on p. 43. For simplicity, the transformation chosen is one that lengthens one diagonal by a factor of 2. Since Lorentz transformations preserve area, the other diagonal is shortened by a factor of 2. Let the original frame of reference, depicted with the square, be A, and the new one B. (a) By measuring with a ruler on the figure, show that the velocity of frame B relative to frame A is  $0.6c$ . (b) Print out a copy of the page. With a ruler, draw a third parallelogram that represents a second successive Lorentz transformation, one that lengthens the long diagonal by another factor of 2. Call this third frame C. Use measurements with a ruler to determine frame C's velocity relative to frame A. Does it equal double the velocity found in part a? Explain why it should be expected to turn out the way it does. ✓

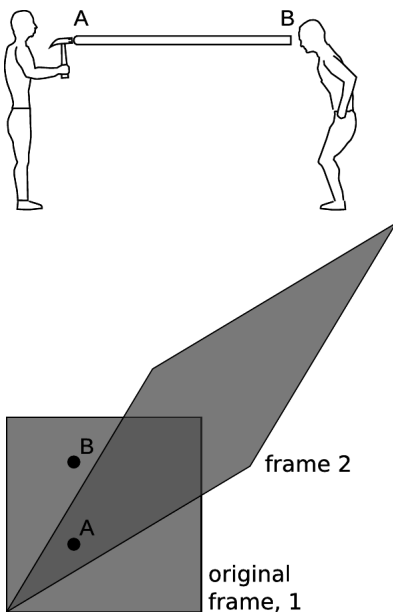


- 2 Astronauts in three different spaceships are communicating with each other. Those aboard ships A and B agree on the rate at which time is passing, but they disagree with the ones on ship C.
- Alice is aboard ship A. How does she describe the motion of her own ship, in its frame of reference?
  - Describe the motion of the other two ships according to Alice.
  - Give the description according to Betty, whose frame of reference is ship B.
  - Do the same for Cathy, aboard ship C.

3 Figure e on p. 43 shows a convention for representing a Lorentz transformation using a parallelogram. Recall that on these graphs, the slope of the parallelogram's bottom edge represents the velocity, and that special units are assumed in which the speed of light equals 1. What would happen to the diagram if the velocity equaled the speed of light?



Problem 4.



Problem 5.

4 The figure shows a famous thought experiment devised by Einstein. A train is moving at constant velocity to the right when bolts of lightning strike the ground near its front and back. Alice, standing on the dirt at the midpoint of the flashes, observes that the light from the two flashes arrives simultaneously, so she says the two strikes must have occurred simultaneously. Bob, meanwhile, is sitting aboard the train, at its middle. He passes by Alice at the moment when Alice later figures out that the flashes happened. Later, he receives flash 2, and then flash 1. He infers that since both flashes traveled half the length of the train, flash 2 must have occurred first. How can this be reconciled with Alice's belief that the flashes were simultaneous? Explain using a graph. Note that the light from the flashes will move at velocity  $c$  or  $-c$ , represented by lines at 45-degree angles.

5 The rod in the figure is perfectly rigid. At event A, the hammer strikes one end of the rod. At event B, the other end moves. Since the rod is perfectly rigid, it can't compress, so A and B are simultaneous. In frame 2, B happens before A. Did the motion at the right end *cause* the person on the left to decide to pick up the hammer and use it?



**6** Suppose that the starship Enterprise from Star Trek has a mass of  $8.0 \times 10^7$  kg, about the same as the Queen Elizabeth 2. Suppose that it was moving at half the speed of light. Read its  $\gamma$  off of the graph in figure 1 on p. 77, and use this to compute its energy. Compare with the total energy content of the world's nuclear arsenals, which is about  $10^{21}$  J.  $\checkmark$

**7** In the graph in figure 1 on p. 77, the  $\gamma$  factor blows up to infinity as the velocity approaches the speed of light. Recall that force is the rate of change of momentum, and that relativistic momentum is given by  $m\gamma v$ . Based on these ideas, what would happen if we applied a constant force to an object for a very long time? Would it eventually go faster than the speed of light?





## Chapter 5

# Electricity

Where the telescope ends, the microscope begins. Which of the two has the grander view?  
*Victor Hugo*

His father died during his mother's pregnancy. Rejected by her as a boy, he was packed off to boarding school when she remarried. He himself never married, but in middle age he formed an intense relationship with a much younger man, a relationship that he terminated when he underwent a psychotic break. Following his early scientific successes, he spent the rest of his professional life mostly in frustration over his inability to unlock the secrets of alchemy.

The man being described is Isaac Newton, but not the triumphant Newton of the standard textbook hagiography. Why dwell on the sad side of his life? To the modern science educator, Newton's lifelong obsession with alchemy may seem an embarrassment, a distraction from his main achievement, the creation the modern science of mechanics. To Newton, however, his alchemical researches were naturally related to his investigations of force and motion. What was radical about Newton's analysis of motion was its universality: it succeeded in describing both the heavens and the earth with the same equations, whereas previously it had been assumed that the sun, moon, stars, and planets were fundamentally different from earthly objects. But Newton realized that if science was to describe all of nature in a unified way, it was not enough to unite the human scale with the scale of the universe: he would not be satisfied until

he fit the microscopic universe into the picture as well.

It should not surprise us that Newton failed. Although he was a firm believer in the existence of atoms, there was no more experimental evidence for their existence than there had been when the ancient Greeks first posited them on purely philosophical grounds. Alchemy labored under a tradition of secrecy and mysticism. Newton had already almost single-handedly transformed the fuzzyheaded field of “natural philosophy” into something we would recognize as the modern science of physics, and it would be unjust to criticize him for failing to change alchemy into modern chemistry as well. The time was not ripe. The microscope was a new invention, and it was cutting-edge science when Newton’s contemporary Hooke discovered that living things were made out of cells.

## 5.1 The Quest for the Atomic Force

Newton was not the first of the age of reason. He was the last of the magicians.

*John Maynard Keynes*

Nevertheless it will be instructive to pick up Newton’s train of thought and see where it leads us with the benefit of modern hindsight. In uniting the human and cosmic scales of existence, he had reimagined both as stages on which the actors were objects (trees and houses, planets and stars) that interacted through attractions and repulsions. He was already convinced that the objects inhabiting the microworld were atoms, so it remained only to determine what kinds of forces they exerted on each other.

His next insight was no less brilliant for his inability to bring it to fruition. He realized that the many human-scale forces — friction, sticky forces, the normal forces that keep objects from occupying the same space, and so on — must all simply be expressions of a more fundamental force acting between atoms. Tape sticks to paper because the atoms in the tape attract the atoms in the paper. My house doesn’t fall to the center of the earth because its atoms repel the atoms of the dirt under it.

Here he got stuck. It was tempting to think that the atomic force was a form of gravity, which he knew to be universal, fundamental, and mathematically simple. Gravity, however, is always attractive, so how could he use it to explain the existence of both attractive and repulsive atomic forces? The gravitational force between objects of ordinary size is also extremely small, which is why we never notice cars and houses attracting us gravitationally. It would be hard to understand how gravity could be responsible for anything as vigorous as the beating of a heart or the explosion of gunpowder. Newton went on to write a million words of alchemical notes filled with speculation about some other force, perhaps a “divine force” or “vegetative force” that would for example be carried by the sperm

to the egg.

Luckily, we now know enough to investigate a different suspect as a candidate for the atomic force: electricity. Electric forces are often observed between objects that have been prepared by rubbing (or other surface interactions), for instance when clothes rub against each other in the dryer. A useful example is shown in figure 5.1/1: stick two pieces of tape on a tabletop, and then put two more pieces on top of them. Lift each pair from the table, and then separate them. The two top pieces will then repel each other, 5.1/2, as will the two bottom pieces. A bottom piece will attract a top piece, however, 5.1/3. Electrical forces like these are similar in certain ways to gravity, the other force that we already know to be fundamental:

- Electrical forces are *universal*. Although some substances, such as fur, rubber, and plastic, respond more strongly to electrical preparation than others, all matter participates in electrical forces to some degree. There is no such thing as a “nonelectric” substance. Matter is both inherently gravitational and inherently electrical.
- Experiments show that the electrical force, like the gravitational force, is an *inverse square* force. That is, the electrical force between two spheres is proportional to  $1/r^2$ , where  $r$  is the center-to-center distance between them.

Furthermore, electrical forces make more sense than gravity as candidates for the fundamental force between atoms, because we have observed that they can be either attractive or repulsive.

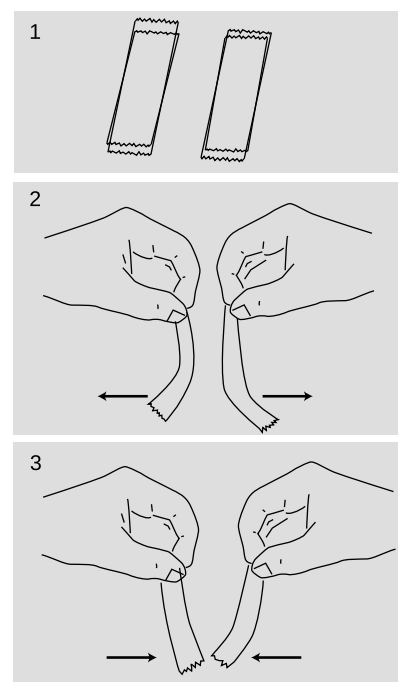
## 5.2 Charge, Electricity and Magnetism

### Charge

“Charge” is the technical term used to indicate that an object participates in electrical forces. This is to be distinguished from the common usage, in which the term is used indiscriminately for anything electrical. For example, although we speak colloquially of “charging” a battery, you may easily verify that a battery has no charge in the technical sense, e.g., it does not exert any electrical force on a piece of tape that has been prepared as described in section 5.1.

#### *Two types of charge*

We can easily collect reams of data on electrical forces between different substances that have been charged in different ways. We find for example that cat fur prepared by rubbing against rabbit fur will attract glass that has been rubbed on silk. How can we make any sense of all this information? A vast simplification is



a / Four pieces of tape are prepared, 1, as described in the text. Depending on which combination is tested, the interaction can be either repulsive, 2, or attractive, 3.

achieved by noting that there are really only two types of charge. Suppose we pick cat fur rubbed on rabbit fur as a representative of type A, and glass rubbed on silk for type B. We will now find that there is no “type C.” Any object electrified by any method is either A-like, attracting things A attracts and repelling those it repels, or B-like, displaying the same attractions and repulsions as B. The two types, A and B, always display opposite interactions. If A displays an attraction with some charged object, then B is guaranteed to undergo repulsion with it, and vice-versa.

### *The coulomb*

Although there are only two types of charge, each type can come in different amounts. The metric unit of charge is the coulomb (rhymes with “drool on”), defined as follows:

One Coulomb (C) is the amount of charge such that a force of  $9.0 \times 10^9$  N occurs between two pointlike objects with charges of 1 C separated by a distance of 1 m.

The notation for an amount of charge is  $q$ . The numerical factor in the definition is historical in origin, and is not worth memorizing. The definition is stated for pointlike, i.e., very small, objects, because otherwise different parts of them would be at different distances from each other.

### *A model of two types of charged particles*

Experiments show that all the methods of rubbing or otherwise charging objects involve two objects, and both of them end up getting charged. If one object acquires a certain amount of one type of charge, then the other ends up with an equal amount of the other type. Various interpretations of this are possible, but the simplest is that the basic building blocks of matter come in two flavors, one with each type of charge. Rubbing objects together results in the transfer of some of these particles from one object to the other. In this model, an object that has not been electrically prepared may actually possess a great deal of *both* types of charge, but the amounts are equal and they are distributed in the same way throughout it. Since type A repels anything that type B attracts, and vice versa, the object will make a total force of zero on any other object. The rest of this chapter fleshes out this model and discusses how these mysterious particles can be understood as being internal parts of atoms.

### *Use of positive and negative signs for charge*

Because the two types of charge tend to cancel out each other’s forces, it makes sense to label them using positive and negative signs, and to discuss the *total* charge of an object. It is entirely arbitrary which type of charge to call negative and which to call positive. Benjamin Franklin decided to describe the one we’ve been calling

“A” as negative, but it really doesn’t matter as long as everyone is consistent with everyone else. An object with a total charge of zero (equal amounts of both types) is referred to as electrically *neutral*.

*self-check A*

Criticize the following statement: “There are two types of charge, attractive and repulsive.”

▷ Answer, p.

180

*Coulomb’s law*

A large body of experimental observations can be summarized as follows:

Coulomb’s law: The magnitude of the force acting between pointlike charged objects at a center-to-center distance  $r$  is given by the equation

$$|\mathbf{F}| = k \frac{|q_1||q_2|}{r^2},$$

where the constant  $k$  equals  $9.0 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2$ . The force is attractive if the charges are of different signs, and repulsive if they have the same sign.

**Conservation of charge**

An even more fundamental reason for using positive and negative signs for electrical charge is that experiments show that with the signs defined this way, the total amount of charge is a conserved quantity. This is why we observe that rubbing initially uncharged substances together always has the result that one gains a certain amount of one type of charge, while the other acquires an equal amount of the other type. Conservation of charge seems natural in our model in which matter is made of positive and negative particles. If the charge on each particle is a fixed property of that type of particle, and if the particles themselves can be neither created nor destroyed, then conservation of charge is inevitable.

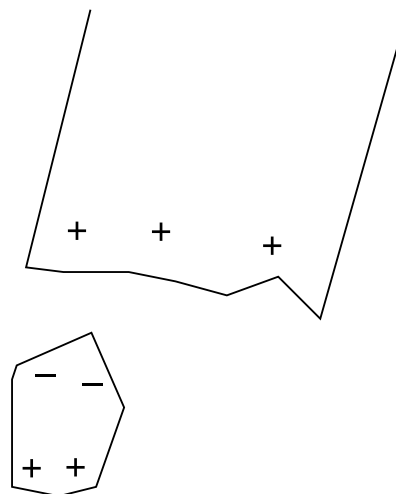
**Electrical forces involving neutral objects**

As shown in figure 5.2.3, an electrically charged object can attract objects that are uncharged. How is this possible? The key is that even though each piece of paper has a total charge of zero, it has at least some charged particles in it that have some freedom to move. Suppose that the tape is positively charged, 5.2.3. Mobile particles in the paper will respond to the tape’s forces, causing one end of the paper to become negatively charged and the other to become positive. The attraction between the paper and the tape is now stronger than the repulsion, because the negatively charged end is closer to the tape.

*self-check B*

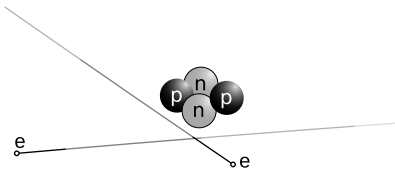


b / A charged piece of tape attracts uncharged pieces of paper from a distance, and they leap up to it.



c / The paper has zero total charge, but it does have charged particles in it that can move.





d / Examples of the construction of atoms: hydrogen (top) and helium (bottom). On this scale, the electrons' orbits would be the size of a college campus.

What would have happened if the tape was negatively charged? ▷  
 Answer, p. 180

### The atom, and subatomic particles

I once had a student whose father had been an electrician. He told me that his father had never really believed that an electrical current in a wire could be carried by moving electrons, because the wire was solid, and it seemed to him that physical particles moving through it would eventually have drilled so many holes through it that it would have crumbled. It may sound as though I'm trying to make fun of the father, but actually he was behaving very much like the model of the skeptical scientist: he didn't want to make hypotheses that seemed more complicated than would be necessary in order to explain his observations. Physicists before about 1905 were in exactly the same situation. They knew all about electrical circuits, and had even invented radio, but knew absolutely nothing about subatomic particles. In other words, it hardly ever matters that electricity really is made of charged particles, and it hardly ever matters what those particles are. Nevertheless, it may avoid some confusion to give a brief review of how an atom is put together:

	<i>charge</i>	<i>mass in units of the proton's mass</i>	<i>location in atom</i>
proton	$+e$	1	in nucleus
neutron	0	1.001	in nucleus
electron	$-e$	1/1836	orbiting nucleus

The symbol  $e$  in this table is an abbreviation for  $1.60 \times 10^{-19}$  C. The physicist Robert Millikan discovered in 1911 that any material object (he used oil droplets) would have a charge that was a multiple of this number, and today we interpret that as being a consequence of the fact that matter is made of atoms, and atoms are made of particles whose charges are plus and minus this amount.

### Electric current

If the fundamental phenomenon is the motion of charged particles, then how can we define a useful numerical measurement of it? We might describe the flow of a river simply by the velocity of the water, but velocity will not be appropriate for electrical purposes because we need to take into account how much charge the moving particles have, and in any case there are no practical devices sold at Radio Shack that can tell us the velocity of charged particles. Experiments show that the intensity of various electrical effects is related to a different quantity: the number of coulombs of charge that pass by a certain point per second. By analogy with the flow of water, this quantity is called the electric *current*,  $I$ . Its units of coulombs/second are more conveniently abbreviated as amperes,  $1 \text{ A} = 1 \text{ C/s}$ . (In informal speech, one usually says "amps.")



e / André Marie Ampère (1775-1836).

The main subtlety involved in this definition is how to account for the two types of charge. The stream of water coming from a hose is made of atoms containing charged particles, but it produces none of the effects we associate with electric currents. For example, you do not get an electrical shock when you are sprayed by a hose. This type of experiment shows that the effect created by the motion of one type of charged particle can be canceled out by the motion of the opposite type of charge in the same direction. In water, every oxygen atom with a charge of  $+8e$  is surrounded by eight electrons with charges of  $-e$ , and likewise for the hydrogen atoms.

We therefore refine our definition of current as follows:

**definition of electric current**

When charged particles are exchanged between regions of space A and B, the electric current flowing from A to B is

$$I = \frac{\text{change in B's charge}}{t},$$

where the transfer occurs over a period of time  $t$ .

In the garden hose example, your body picks up equal amounts of positive and negative charge, resulting in no change in your total charge, so the electrical current flowing into you is zero.

*Ions moving across a cell membrane* *example 1*

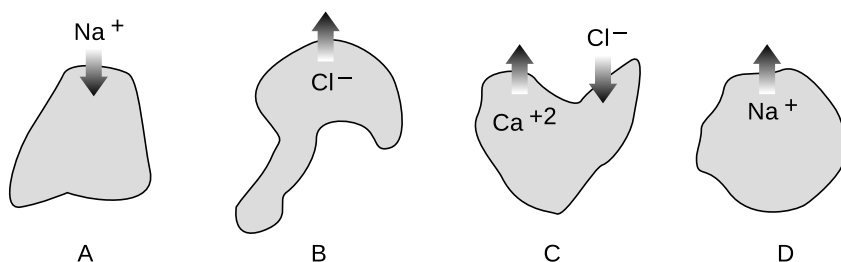
▷ Figure f shows ions, labeled with their charges, moving in or out through the membranes of four cells. If the ions all cross the membranes during the same interval of time, how would the currents into the cells compare with each other?

▷ Cell A has positive current going into it because its charge is increased, i.e., has a positive change in its charge.

Cell B has the same current as cell A, because by losing one unit of negative charge it also ends up increasing its own total charge by one unit.

Cell C's total charge is reduced by three units, so it has a large negative current going into it.

Cell D loses one unit of charge, so it has a small negative current into it.



f / Example 1

It may seem strange to say that a negatively charged particle going one way creates a current going the other way, but this is quite ordinary. As we will see, currents flow through metal wires via the motion of electrons, which are negatively charged, so the direction of motion of the electrons in a circuit is always opposite to the direction of the current. Of course it would have been convenient if Benjamin Franklin had defined the positive and negative signs of charge the opposite way, since so many electrical devices are based on metal wires.

*Number of electrons flowing through a lightbulb*      *example 2*

▷ If a lightbulb has 1.0 A flowing through it, how many electrons will pass through the filament in 1.0 s?

▷ We are only calculating the number of electrons that flow, so we can ignore the positive and negative signs. Solving for (charge) =  $It$  gives a charge of 1.0 C flowing in this time interval. The number of electrons is

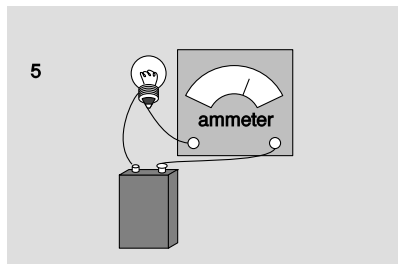
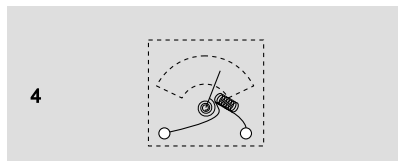
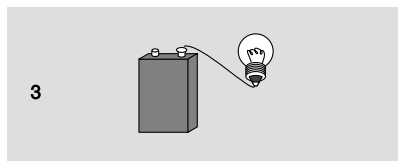
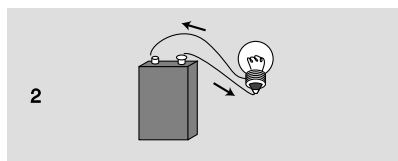
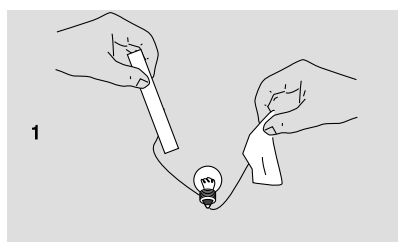
$$\begin{aligned} \text{number of electrons} &= \text{coulombs} \times \frac{\text{electrons}}{\text{coulomb}} \\ &= \text{coulombs} / \frac{\text{coulombs}}{\text{electron}} \\ &= 1.0 \text{ C} / e \\ &= 6.2 \times 10^{18} \end{aligned}$$

That's a lot of electrons!

### 5.3 Circuits

How can we put electric currents to work? The only method of controlling electric charge we have studied so far is to charge different substances, e.g., rubber and fur, by rubbing them against each other. Figure g/1 shows an attempt to use this technique to light a lightbulb. This method is unsatisfactory. True, current will flow through the bulb, since electrons can move through metal wires, and the excess electrons on the rubber rod will therefore come through the wires and bulb due to the attraction of the positively charged fur and the repulsion of the other electrons. The problem is that after a zillionth of a second of current, the rod and fur will both have run out of charge. No more current will flow, and the lightbulb will go out.

Figure g/2 shows a setup that works. The battery pushes charge through the circuit, and recycles it over and over again. (We will have more to say later in this chapter about how batteries work.) This is called a *complete circuit*. Today, the electrical use of the word “circuit” is the only one that springs to mind for most people, but the original meaning was to travel around and make a round trip, as when a circuit court judge would ride around the boondocks, dispensing justice in each town on a certain date.



g / 1. Static electricity runs out quickly. 2. A practical circuit. 3. An open circuit. 4. How an ammeter works. 5. Measuring the current with an ammeter.

Note that an example like g/3 does not work. The wire will quickly begin acquiring a net charge, because it has no way to get rid of the charge flowing into it. The repulsion of this charge will make it more and more difficult to send any more charge in, and soon the electrical forces exerted by the battery will be canceled out completely. The whole process would be over so quickly that the filament would not even have enough time to get hot and glow. This is known as an *open circuit*. Exactly the same thing would happen if the complete circuit of figure g/2 was cut somewhere with a pair of scissors, and in fact that is essentially how an ordinary light switch works: by opening up a gap in the circuit.

The definition of electric current we have developed has the great virtue that it is easy to measure. In practical electrical work, one almost always measures current, not charge. The instrument used to measure current is called an *ammeter*. A simplified ammeter, g/4, simply consists of a coiled-wire magnet whose force twists an iron needle against the resistance of a spring. The greater the current, the greater the force. Although the construction of ammeters may differ, their use is always the same. We break into the path of the electric current and interpose the meter like a tollbooth on a road, g/5. There is still a complete circuit, and as far as the battery and bulb are concerned, the ammeter is just another segment of wire.

Does it matter where in the circuit we place the ammeter? Could we, for instance, have put it in the left side of the circuit instead of the right? Conservation of charge tells us that this can make no difference. Charge is not destroyed or “used up” by the lightbulb, so we will get the same current reading on either side of it. What is “used up” is energy stored in the battery, which is being converted into heat and light energy.

## 5.4 Voltage

### The volt unit

Electrical circuits can be used for sending signals, storing information, or doing calculations, but their most common purpose by far is to manipulate energy, as in the battery-and-bulb example of the previous section. We know that lightbulbs are rated in units of watts, i.e., how many joules per second of energy they can convert into heat and light, but how would this relate to the flow of charge as measured in amperes? By way of analogy, suppose your friend, who didn't take physics, can't find any job better than pitching bales of hay. The number of calories he burns per hour will certainly depend on how many bales he pitches per minute, but it will also be proportional to how much mechanical work he has to do on each bale. If his job is to toss them up into a hayloft, he will get tired a lot more quickly than someone who merely tips bales off a loading dock



h / Alessandro Volta (1745-1827).

into trucks. In metric units,

$$\frac{\text{joules}}{\text{second}} = \frac{\text{haybales}}{\text{second}} \times \frac{\text{joules}}{\text{haybale}}.$$

Similarly, the rate of energy transformation by a battery will not just depend on how many coulombs per second it pushes through a circuit but also on how much mechanical work it has to do on each coulomb of charge:

$$\frac{\text{joules}}{\text{second}} = \frac{\text{coulombs}}{\text{second}} \times \frac{\text{joules}}{\text{coulomb}}$$

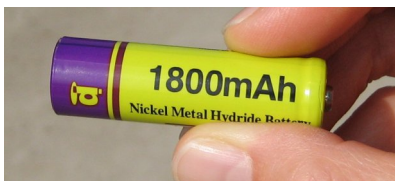
or

$$\text{power} = \text{current} \times \text{work per unit charge}.$$

Units of joules per coulomb are abbreviated as *volts*, 1 V=1 J/C, named after the Italian physicist Alessandro Volta. Everyone knows that batteries are rated in units of volts, but the voltage concept is more general than that; it turns out that voltage is a property of every point in space.

To gain more insight, let's think again about the analogy with the haybales. It took a certain number of joules of gravitational energy to lift a haybale from one level to another. Since we're talking about gravitational energy, it really makes more sense to talk about units of mass, rather than using the haybale as our measure of the quantity of matter. The gravitational version of voltage would then be joules per kilogram. Gravitational energy equals  $mgh$ , but if we calculate how much of that we have *per kilogram*, we're canceling out the  $m$ , giving simply  $gh$ . For any point in the Earth's gravitational field, we can assign a number,  $gh$ , which tells us how hard it is to get a given amount of mass to that point. For instance, the top of Mount Everest would have a big value of  $gh$ , because of the big height. That tells us that it's expensive in terms of energy to lift a given amount of mass from some reference level (sea level, say) to the top of Mount Everest.

Voltage does the same thing, but using electrical energy. We can visualize an electrical circuit as being like a roller-coaster. The battery is like the part of the roller-coaster where they lift you up to the top. The height of this initial hill is analogous to the voltage of the battery. When you roll downhill later, that's like a lightbulb. In the roller-coaster, the initial gravitational energy is turned into heat and sound as the cars go down the hill. In our circuit, the initial electrical energy is turned into heat by the lightbulb (and the hot filament of the lightbulb then glows, turning the heat into light).



i / Example 3.

**Energy stored in a battery**

**example 3**

▷ The 1.2 V rechargeable battery in figure i is labeled 1800 milliamp-hours. What is the maximum amount of energy the battery can store?

▷ An ampere-hour is a unit of current multiplied by a unit of time. Current is charge per unit time, so an ampere-hour is in fact a funny unit of *charge*:

$$\begin{aligned}(1 \text{ A})(1 \text{ hour}) &= (1 \text{ C/s})(3600 \text{ s}) \\ &= 3600 \text{ C}\end{aligned}$$

1800 milliamp-hours is therefore  $1800 \times 10^{-3} \times 3600 \text{ C} = 6.5 \times 10^3 \text{ C}$ . That's a huge number of charged particles, but the total loss of electrical energy will just be their total charge multiplied by the voltage difference across which they move:

$$\begin{aligned}\text{energy} &= (6.5 \times 10^3 \text{ C})(1.2 \text{ V}) \\ &= 7.8 \text{ kJ}\end{aligned}$$

Using the definition of voltage,  $V$ , we can rewrite the equation power = current  $\times$  work per unit charge more concisely as  $P = IV$ .

---

*Units of volt-amps* *example 4*

- ▷ Doorbells are often rated in volt-amps. What does this combination of units mean?
- ▷ Current times voltage gives units of power,  $P = IV$ , so volt-amps are really just a nonstandard way of writing watts. They are telling you how much power the doorbell requires.

---

*Power dissipated by a battery and bulb* *example 5*

- ▷ If a 9.0-volt battery causes 1.0 A to flow through a lightbulb, how much power is dissipated?
- ▷ The voltage rating of a battery tells us what voltage difference  $\Delta V$  it is designed to maintain between its terminals.

$$\begin{aligned}P &= I \Delta V \\ &= 9.0 \text{ A} \cdot \text{V} \\ &= 9.0 \frac{\text{C}}{\text{s}} \cdot \frac{\text{J}}{\text{C}} \\ &= 9.0 \text{ J/s} \\ &= 9.0 \text{ W}\end{aligned}$$

The only nontrivial thing in this problem was dealing with the units. One quickly gets used to translating common combinations like  $\text{A} \cdot \text{V}$  into simpler terms.

### Discussion questions

- A** In the roller-coaster metaphor, what would a high-voltage roller coaster be like? What would a high-current roller coaster be like?
- B** Criticize the following statements:

“He touched the wire, and 10000 volts went through him.”

“That battery has a charge of 9 volts.”

“You used up the charge of the battery.”

**C** When you touch a 9-volt battery to your tongue, both positive and negative ions move through your saliva. Which ions go which way?

**D** I once touched a piece of physics apparatus that had been wired incorrectly, and got a several-thousand-volt voltage difference across my hand. I was not injured. For what possible reason would the shock have had insufficient power to hurt me?

## 5.5 Resistance

What’s the physical difference between a 100-watt lightbulb and a 200-watt one? They both plug into a 110-volt outlet, so according to the equation  $P = IV$ , the only way to explain the double power of the 200-watt bulb is that it must pull in, or “draw,” twice as much current. By analogy, a fire hose and a garden hose might be served by pumps that give the same pressure (voltage), but more water will flow through the fire hose, because there’s simply more water in the hose that can flow. Likewise, a wide, deep river could flow down the same slope as a tiny creek, but the number of liters of water flowing through the big river is greater. If you look at the filaments of a 100-watt bulb and a 200-watt bulb, you’ll see that the 200-watt bulb’s filament is thicker. In the charged-particle model of electricity, we expect that the thicker filament will contain more charged particles that are available to flow. We say that the thicker filament has a lower electrical resistance than the thinner one.



k / Georg Simon Ohm (1787-1854).

j / A fat pipe has less resistance than a skinny pipe.



Although it’s harder to pump water rapidly through a garden hose than through a fire hose, we could always compensate by using a higher-pressure pump. Similarly, the amount of current that will flow through a lightbulb depends not just on its resistance but also on how much of a voltage difference is applied across it. For many substances, including the tungsten metal that lightbulb filaments are made of, we find that the amount of current that flows is pro-

portional to the voltage difference applied to it, so that the ratio of voltage to current stays the same. We then use this ratio as a numerical definition of resistance,

$$R = \frac{V}{I},$$

which is known as Ohm's law. The units of resistance are ohms, symbolized with an uppercase Greek letter Omega,  $\Omega$ . Physically, when a current flows through a resistance, the result is to transform electrical energy into heat. In a lightbulb filament, for example, the heat is what causes the bulb to glow.

Ohm's law states that many substances, including many solids and some liquids, display this kind of behavior, at least for voltages that are not too large. The fact that Ohm's law is called a "law" should not be taken to mean that all materials obey it, or that it has the same fundamental importance as the conservation laws, for example. Materials are called *ohmic* or *nonohmic*, depending on whether they obey Ohm's law.

On an intuitive level, we can understand the idea of resistance by making the sounds "hhhhh" and "ffffff." To make air flow out of your mouth, you use your diaphragm to compress the air in your chest. The pressure difference between your chest and the air outside your mouth is analogous to a voltage difference. When you make the "h" sound, you form your mouth and throat in a way that allows air to flow easily. The large flow of air is like a large current. Dividing by a large current in the definition of resistance means that we get a small resistance. We say that the small resistance of your mouth and throat allows a large current to flow. When you make the "f" sound, you increase the resistance and cause a smaller current to flow. In this mechanical analogy, resistance is like friction: the air rubs against your lips. Mechanical friction converts mechanical forms of energy to heat, as when you rub your hands together. Electrical friction — resistance — converts electrical energy to heat.

If objects of the same size and shape made from two different ohmic materials have different resistances, we can say that one material is more resistive than the other, or equivalently that it is less conductive. Materials, such as metals, that are very conductive are said to be good *conductors*. Those that are extremely poor conductors, for example wood or rubber, are classified as *insulators*. There is no sharp distinction between the two classes of materials. Some, such as silicon, lie midway between the two extremes, and are called semiconductors.

## Applications

### *Superconductors*

All materials display some variation in resistance according to temperature (a fact that is used in thermostats to make a ther-



mometer that can be easily interfaced to an electric circuit). More spectacularly, most metals have been found to exhibit a sudden change to *zero* resistance when cooled to a certain critical temperature. They are then said to be superconductors. A current flowing through a superconductor doesn't create any heat at all.

Theoretically, superconductors should make a great many exciting devices possible, for example coiled-wire magnets that could be used to levitate trains. In practice, the critical temperatures of all metals are very low, and the resulting need for extreme refrigeration has made their use uneconomical except for such specialized applications as particle accelerators for physics research.

But scientists have recently made the surprising discovery that certain ceramics are superconductors at less extreme temperatures. The technological barrier is now in finding practical methods for making wire out of these brittle materials. Wall Street is currently investing billions of dollars in developing superconducting devices for cellular phone relay stations based on these materials. In 2001, the city of Copenhagen replaced a short section of its electrical power trunks with superconducting cables, and they are now in operation and supplying power to customers.

There is currently no satisfactory theory of superconductivity in general, although superconductivity in metals is understood fairly well. Unfortunately I have yet to find a fundamental explanation of superconductivity in metals that works at the introductory level.

#### *Constant voltage throughout a conductor*

The idea of a superconductor leads us to the question of how we should expect an object to behave if it is made of a very good conductor. Superconductors are an extreme case, but often a metal wire can be thought of as a perfect conductor, for example if the parts of the circuit other than the wire are made of much less conductive materials. What happens if the resistance equals zero in the equation

$$R = \frac{V}{I} \quad ?$$

The result of dividing two numbers can only be zero if the number on top equals zero. This tells us that if we pick any two points in a perfect conductor, the voltage difference between them must be zero. In other words, the entire conductor must be at the same voltage. Using the water metaphor, a perfect conductor is like a perfectly calm lake or canal, whose surface is flat. If you take an eyedropper and deposit a drop of water anywhere on the surface, it doesn't flow away, because the water is still. In electrical terms, a charge located anywhere in the interior of a perfect conductor will always feel a total electrical force of zero.

Suppose, for example, that you build up a static charge by scuff-

ing your feet on a carpet, and then you deposit some of that charge onto a doorknob, which is a good conductor. How can all that charge be in the doorknob without creating any electrical force at any point inside it? The only possible answer is that the charge moves around until it has spread itself into just the right configuration. In this configuration, the forces exerted by all the charge on any charged particle within the doorknob exactly cancel out.

We can explain this behavior if we assume that the charge placed on the doorknob eventually settles down into a stable equilibrium. Since the doorknob is a conductor, the charge is free to move through it. If it was free to move and any part of it did experience a nonzero total force from the rest of the charge, then it would move, and we would not have an equilibrium.

It also turns out that charge placed on a conductor, once it reaches its equilibrium configuration, is entirely on the surface, not on the interior. We will not prove this fact formally, but it is intuitively reasonable (see discussion question B).

### Short circuits

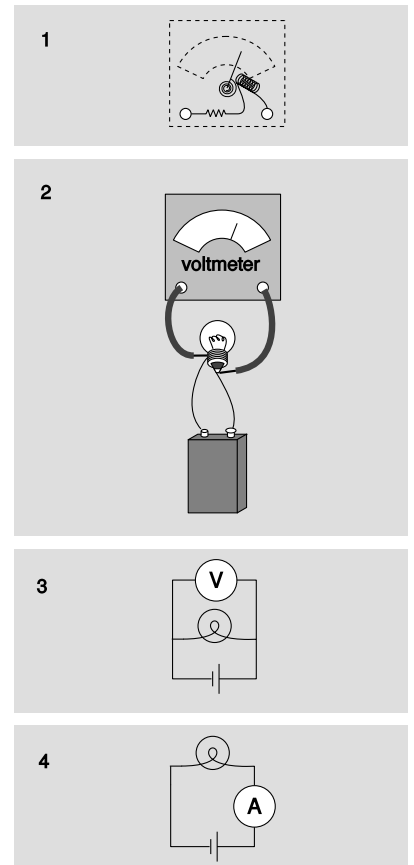
So far we have been assuming a perfect conductor. What if it's a good conductor, but not a perfect one? Then we can solve for

$$V = IR.$$

An ordinary-sized current will make a very small result when we multiply it by the resistance of a good conductor such as a metal wire. The voltage throughout the wire will then be nearly constant. If, on the other hand, the current is extremely large, we can have a significant voltage difference. This is what happens in a short-circuit: a circuit in which a low-resistance pathway connects the two sides of a voltage source. Note that this is much more specific than the popular use of the term to indicate any electrical malfunction at all. If, for example, you short-circuit a 9-volt battery as shown in the figure, you will produce perhaps a thousand amperes of current, leading to a very large value of  $P = IV$ . The wire gets hot!

### The voltmeter

A voltmeter is nothing more than an ammeter with an additional high-value resistor through which the current is also forced to flow,  $1/1$ . Ohm's law relates the current through the resistor directly to the voltage difference across it, so the meter can be calibrated in units of volts based on the known value of the resistor. The voltmeter's two probes are touched to the two locations in a circuit between which we wish to measure the voltage difference,  $1/2$ . Note how cumbersome this type of drawing is, and how difficult it can be to tell what is connected to what. This is why electrical drawing are usually shown in schematic form. Figure  $1/3$  is a schematic representation of figure  $1/2$ .



1/1. A simplified diagram of how a voltmeter works. 2. Measuring the voltage difference across a lightbulb. 3. The same setup drawn in schematic form. 4. The setup for measuring current is different.

The setups for measuring current and voltage are different. When we're measuring current, we're finding "how much stuff goes through," so we place the ammeter where all the current is forced to go through it. Voltage, however, is not "stuff that goes through," it is a measure of electrical energy. If an ammeter is like the meter that measures your water use, a voltmeter is like a measuring stick that tells you how high a waterfall is, so that you can determine how much energy will be released by each kilogram of falling water. We don't want to force the water to go through the measuring stick! The arrangement in figure 1/3 is a parallel circuit: one which in there are "forks in the road" where some of the current will flow one way and some will flow the other. Figure 1/4 is said to be wired in series: all the current will visit all the circuit elements one after the other.

If you inserted a voltmeter incorrectly, in series with the bulb and battery, its large internal resistance would cut the current down so low that the bulb would go out. You would have severely disturbed the behavior of the circuit by trying to measure something about it.

Incorrectly placing an ammeter in parallel is likely to be even more disconcerting. The ammeter has nothing but wire inside it to provide resistance, so given the choice, most of the current will flow through it rather than through the bulb. So much current will flow through the ammeter, in fact, that there is a danger of burning out the battery or the meter or both! For this reason, most ammeters have fuses or circuit breakers inside. Some models will trip their circuit breakers and make an audible alarm in this situation, while others will simply blow a fuse and stop working until you replace it.

### **Discussion questions**

**A** In figure g/4 on page 100, what would happen if you had the ammeter on the left rather than on the right?

**B** Imagine a charged doorknob, as described on page 107. Why is it intuitively reasonable to believe that all the charge will end up on the surface of the doorknob, rather than on the interior?

## Problems

### Key

✓ A computerized answer check is available online.

∫ A problem that requires calculus.

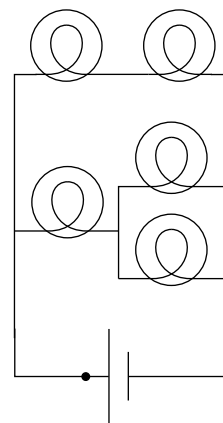
★ A difficult problem.

**1** A hydrogen atom consists of an electron and a proton. For our present purposes, we'll think of the electron as orbiting in a circle around the proton.

The subatomic particles called muons behave exactly like electrons, except that a muon's mass is greater by a factor of 206.77. Muons are continually bombarding the Earth as part of the stream of particles from space known as cosmic rays. When a muon strikes an atom, it can displace one of its electrons. If the atom happens to be a hydrogen atom, then the muon takes up an orbit that is on the average 206.77 times closer to the proton than the orbit of the ejected electron. How many times greater is the electric force experienced by the muon than that previously felt by the electron?

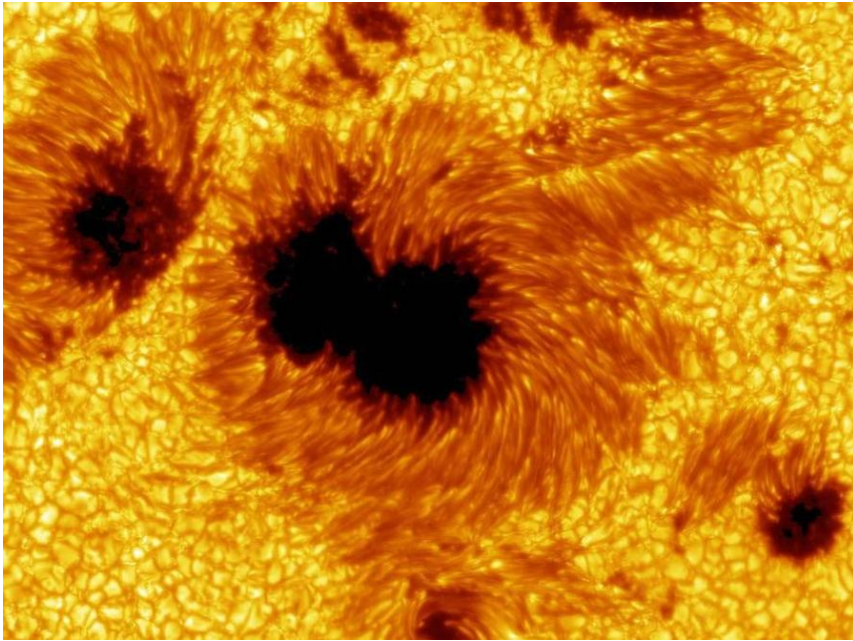
**2** The figure shows a circuit containing five lightbulbs connected to a battery. Suppose you're going to connect one probe of a voltmeter to the circuit at the point marked with a dot. How many unique, nonzero voltage differences could you measure by connecting the other probe to other wires in the circuit? Visualize the circuit using the same waterfall metaphor.

**3** The lightbulbs in the figure are all identical. If you were inserting an ammeter at various places in the circuit, how many unique currents could you measure? If you know that the current measurement will give the same number in more than one place, only count that as one unique current.



Problems 2 and 3.





This sunspot is a product of the sun's magnetic fields. The darkest region in the center is about the size of our planet.

## Chapter 6

# Fields

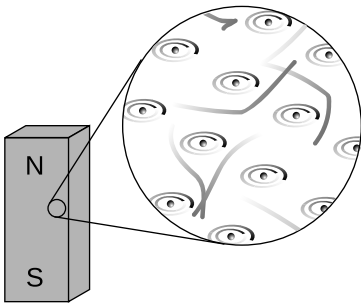
### 6.1 Farewell To the Mechanical Universe

As late as 1900, physicists generally conceived of the universe in mechanical terms. Newton had revealed the solar system as a collection of material objects interacting through forces that acted at a distance. By 1900, evidence began to accumulate for the existence of atoms as real things, and not just as imaginary models of reality. In this microscopic realm, the same (successful) Newtonian picture tended to be transferred over to the microscopic world. Now the actors on the stage were atoms rather than planets, and the forces were electrical rather than gravitational, but it seemed to be a variation on the same theme. Some physicists, however, began to realize that the old mechanical picture wouldn't quite work. At a deeper level, the operation of the universe came to be understood in terms of *fields*, the general idea being embodied fairly well in "The Force" from the Star Wars movies: "... an energy field created by all living things. It surrounds us, penetrates us, and binds the galaxy together." Substitute "massive" for "living," and you have a fairly good description of the gravitational field, which I first casually mentioned on page 20. Substitute "charged" instead, and

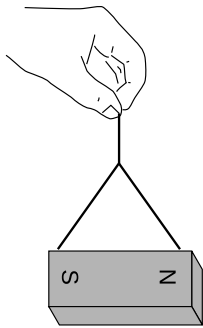
it's a depiction of the electric field.

### Time delays in forces exerted at a distance

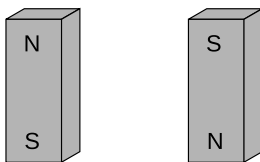
What convinced physicists that they needed this new concept of a field of force? Although we have been dealing mostly with electrical forces, let's start with a magnetic example. (In fact the main reason I've delayed a detailed discussion of magnetism for so long is that mathematical calculations of magnetic effects are handled much more easily with the concept of a field of force.) First a little background leading up to our example. A bar magnet, *a*, has an axis about which many of the electrons' orbits are oriented. The earth itself is also a magnet, although not a bar-shaped one. The interaction between the earth-magnet and the bar magnet, *b*, makes them want to line up their axes in opposing directions (in other words such that their electrons rotate in parallel planes, but with one set rotating clockwise and the other counterclockwise as seen looking along the axes). On a smaller scale, any two bar magnets placed near each other will try to align themselves head-to-tail, *c*.



a / A bar magnet's atoms are (partially) aligned.



b / A bar magnet interacts with our magnetic planet.



c / Magnets aligned north-south.

Now we get to the relevant example. It is clear that two people separated by a paper-thin wall could use a pair of bar magnets to signal to each other. Each person would feel her own magnet trying to twist around in response to any rotation performed by the other person's magnet. The practical range of communication would be very short for this setup, but a sensitive electrical apparatus could pick up magnetic signals from much farther away. In fact, this is not so different from what a radio does: the electrons racing up and down the transmitting antenna create forces on the electrons in the distant receiving antenna. (Both magnetic and electric forces are involved in real radio signals, but we don't need to worry about that yet.)

A question now naturally arises as to whether there is any time delay in this kind of communication via magnetic (and electric) forces. Newton would have thought not, since he conceived of physics in terms of instantaneous action at a distance. We now know, however, that there is such a time delay. If you make a long-distance phone call that is routed through a communications satellite, you should easily be able to detect a delay of about half a second over the signal's round trip of 50,000 miles. Modern measurements have shown that electric, magnetic, and gravitational forces all travel at the speed of light,  $3 \times 10^8$  m/s.<sup>1</sup> (In fact, we will soon discuss how light itself is made of electricity and magnetism.)

If it takes some time for forces to be transmitted through space, then apparently there is some *thing* that travels *through* space. The fact that the phenomenon travels outward at the same speed in all

<sup>1</sup>As discussed in chapter 4, one consequence of Einstein's theory of relativity is that material objects can never move faster than the speed of light. It can also be shown that signals or information are subject to the same limit.

directions strongly evokes wave metaphors such as ripples on a pond.

**More evidence that fields of force are real: they carry energy.**

The smoking-gun argument for this strange notion of traveling force ripples comes from the fact that they carry energy.

First suppose that the person holding the bar magnet on the right decides to reverse hers, resulting in configuration d. She had to do mechanical work to twist it, and if she releases the magnet, energy will be released as it flips back to c. She has apparently stored energy by going from c to d. So far everything is easily explained without the concept of a field of force.

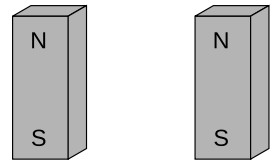
But now imagine that the two people start in position c and then simultaneously flip their magnets extremely quickly to position e, keeping them lined up with each other the whole time. Imagine, for the sake of argument, that they can do this so quickly that each magnet is reversed while the force signal from the other is still in transit. (For a more realistic example, we'd have to have two radio antennas, not two magnets, but the magnets are easier to visualize.) During the flipping, each magnet is still feeling the forces arising from the way the other magnet *used* to be oriented. Even though the two magnets stay aligned during the flip, the time delay causes each person to feel resistance as she twists her magnet around. How can this be? Both of them are apparently doing mechanical work, so they must be storing magnetic energy somehow. But in the traditional Newtonian conception of matter interacting via instantaneous forces at a distance, interaction energy arises from the relative positions of objects that are interacting via forces. If the magnets never changed their orientations relative to each other, how can any magnetic energy have been stored?

The only possible answer is that the energy must have gone into the magnetic force ripples crisscrossing the space between the magnets. Fields of force apparently carry energy across space, which is strong evidence that they are real things.

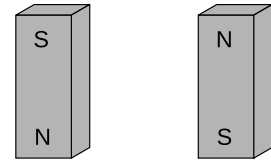
This is perhaps not as radical an idea to us as it was to our ancestors. We are used to the idea that a radio transmitting antenna consumes a great deal of power, and somehow spews it out into the universe. A person working around such an antenna needs to be careful not to get too close to it, since all that energy can easily cook flesh (a painful phenomenon known as an “RF burn”).

**The gravitational field**

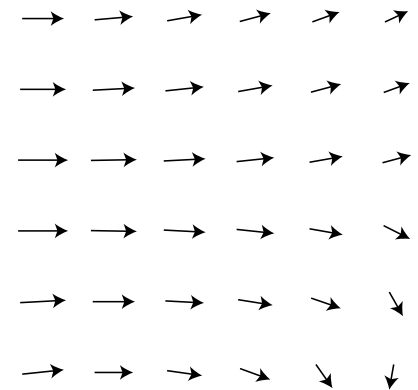
Given that fields of force are real, how do we define, measure, and calculate them? A fruitful metaphor will be the wind patterns experienced by a sailing ship. Wherever the ship goes, it will feel a certain amount of force from the wind, and that force will be in a certain direction. The weather is ever-changing, of course, but for



d / The second magnet is reversed.



e / Both magnets are reversed.



f / The wind patterns in a certain area of the ocean could be charted in a “sea of arrows” representation like this. Each arrow represents both the wind’s strength and its direction at a certain location.



now let's just imagine steady wind patterns. Definitions in physics are operational, i.e., they describe how to measure the thing being defined. The ship's captain can measure the wind's "field of force" by going to the location of interest and determining both the direction of the wind and the strength with which it is blowing. Charting all these measurements on a map leads to a depiction of the field of wind force like the one shown in the figure. This is known as the "sea of arrows" method of visualizing a field.

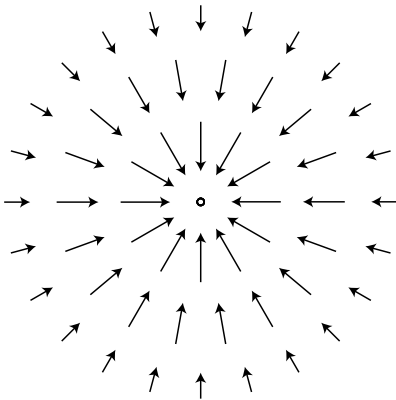
Now let's see how these concepts are applied to the fundamental force fields of the universe. We'll start with the gravitational field, which is the easiest to understand. We've already encountered the gravitational field,  $g$ , which we defined in terms of energy. Essentially,  $g$  was defined as the number that would make the equation  $GE = mgh$  give the right answer. However, we intuitively feel that the gravitational field has a direction associated with it: down! This can be more easily expressed via the following definition:

#### definition of the gravitational field

The gravitational field,  $g$ , at any location in space is found by placing a test mass  $m$  at that point. The field is then given by  $g = F/m$ , where  $F$  is the gravitational force on the test mass.

With this new definition, we get units of N/kg, rather than J/kg/m. These are in fact equivalent units.

The most subtle point about all this is that the gravitational field tells us about what forces *would* be exerted on a test mass by the earth, sun, moon, and the rest of the universe, *if* we inserted a test mass at the point in question. The field still exists at all the places where we didn't measure it.



$g$  / The gravitational field surrounding a clump of mass such as the earth.

#### Sources and sinks

If we make a sea-of-arrows picture of the gravitational fields surrounding the earth,  $g$ , the result is evocative of water going down a drain. For this reason, anything that creates an inward-pointing field around itself is called a sink. The earth is a gravitational sink. The term "source" can refer specifically to things that make outward fields, or it can be used as a more general term for both "outies" and "innies." However confusing the terminology, we know that gravitational fields are only attractive, so we will never find a region of space with an outward-pointing field pattern.

Knowledge of the field is interchangeable with knowledge of its sources (at least in the case of a static, unchanging field). If aliens saw the earth's gravitational field pattern they could immediately infer the existence of the planet, and conversely if they knew the mass of the earth they could predict its influence on the surrounding gravitational field.

## The electric field

The definition of the electric field is directly analogous to, and has the same motivation as, the definition of the gravitational field:

### definition of the electric field

The electric field,  $E$ , at any location in space is found by placing a test charge  $q$  at that point. The electric field vector is then given by  $E = F/q$ , where  $F$  is the electric force on the test charge.

Charges are what create electric fields. Unlike gravity, which is always attractive, electricity displays both attraction and repulsion. A positive charge is a source of electric fields, and a negative one is a sink.

## 6.2 Electromagnetism

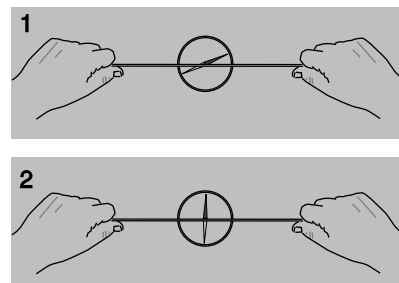
Think not that I am come to destroy the law, or the prophets: I am not come to destroy, but to fulfill. *Matthew 5:17*

### Magnetic interactions

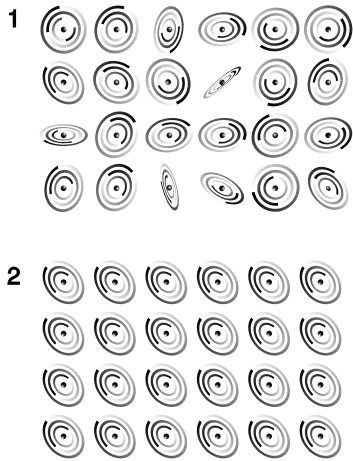
At this stage, you understand roughly as much about the classification of interactions as physicists understood around the year 1800. There appear to be three fundamentally different types of interactions: gravitational, electrical, and magnetic. Many types of interactions that appear superficially to be distinct — stickiness, chemical interactions, the energy an archer stores in a bow — are really the same: they're manifestations of electrical interactions between atoms. Is there any way to shorten the list any further? The prospects seem dim at first. For instance, we find that if we rub a piece of fur on a rubber rod, the fur does not attract or repel a magnet. The fur has an electric field, and the magnet has a magnetic field. The two are completely separate, and don't seem to affect one another. Likewise we can test whether magnetizing a piece of iron changes its weight. The weight doesn't seem to change by any measurable amount, so magnetism and gravity seem to be unrelated.

That was where things stood until 1820, when the Danish physicist Hans Christian Oersted was delivering a lecture at the University of Copenhagen, and he wanted to give his students a demonstration that would illustrate the cutting edge of research. He generated a current in a wire by making a short circuit across a battery, and held the wire near a magnetic compass. The idea was to give an example of how one could search for a previously undiscovered link between electricity (the electric current in the wire) and magnetism. One never knows how much to believe from these dramatic legends, but the story is<sup>2</sup> that the experiment he'd expected to turn out neg-

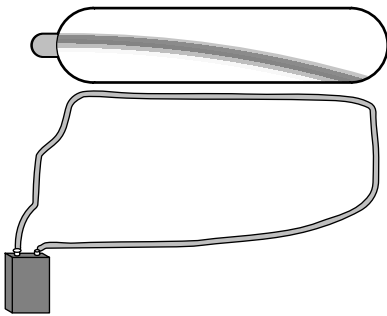
<sup>2</sup>Oersted's paper describing the phenomenon says that "The first experiments on the subject . . . were set on foot in the classes for electricity, galvanism, and



h / 1. When the circuit is incomplete, no current flows through the wire, and the magnet is unaffected. It points in the direction of the Earth's magnetic field. 2. The circuit is completed, and current flows through the wire. The wire has a strong effect on the magnet, which turns almost perpendicular to it. If the earth's field could be removed entirely, the compass would point exactly perpendicular to the wire; this is the direction of the wire's field.



i / A schematic representation of an unmagnetized material, 1, and a magnetized one, 2.



j / Magnetism is an interaction between moving charges and moving charges. The moving charges in the wire attract the moving charges in the beam of charged particles in the vacuum tube.

ative instead turned out positive: when he held the wire near the compass, the current in the wire caused the compass to twist!

People had tried similar experiments before, but only with static electricity, not with a moving electric current. For instance, they had hung batteries so that they were free to rotate in the earth’s magnetic field, and found no effect; since the battery was not connected to a complete circuit, there was no current flowing. With Oersted’s own setup, h, the effect was only produced when the “circuit was closed, but not when open, as certain very celebrated physicists in vain attempted several years ago.”<sup>3</sup>

Oersted was eventually led to the conclusion that magnetism was an interaction between moving charges and other moving charges, i.e., between one current and another. A permanent magnet, he inferred, contained currents on a microscopic scale that simply weren’t practical to measure with an ammeter. Today this seems natural to us, since we’re accustomed to picturing an atom as a tiny solar system, with the electrons whizzing around the nucleus in circles. As shown in figure i, a magnetized piece of iron is different from an unmagnetized piece because the atoms in the unmagnetized piece are jumbled in random orientations, whereas the atoms in the magnetized piece are at least partially organized to face in a certain direction.

Figure j shows an example that is conceptually simple, but not very practical. If you try this with a typical vacuum tube, like a TV or computer monitor, the current in the wire probably won’t be enough to produce a visible effect. A more practical method is to hold a magnet near the screen. We still have an interaction between moving charges and moving charges, but the swirling electrons in the atoms in the magnet are now playing the role played by the moving charges in the wire in figure j. Warning: if you do this, make sure your monitor has a demagnetizing button! If not, then your monitor may be permanently ruined.

### Relativity requires magnetism

So magnetism is an interaction between moving charges and moving charges. But how can that be? Relativity tells us that motion is a matter of opinion. Consider figure k. In this figure and in figure l, the dark and light coloring of the particles represents the fact that one particle has one type of charge and the other particle has the other type. Observer k/2 sees the two particles as flying through space side by side, so they would interact both electrically (simply because they’re charged) and magnetically (because they’re charges in motion). But an observer moving along with them, k/1, would say they were both at rest, and would expect only an elec-

magnetism, which were held by me in the winter just past,” but that doesn’t tell us whether the result was really a surprise that occurred in front of his students.

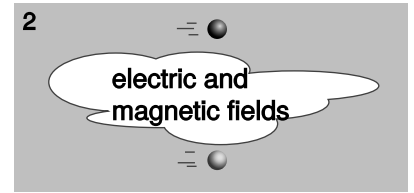
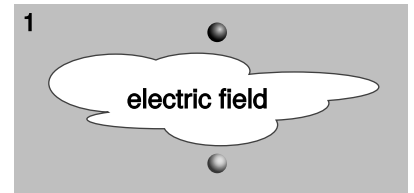
<sup>3</sup>All quotes are from the 1876 translation by J.E. Kempe.

trical interaction. This seems like a paradox. Magnetism, however, comes not to destroy relativity but to fulfill it. Magnetic interactions *must* exist according to the theory of relativity. To understand how this can be, consider how time and space behave in relativity. Observers in different frames of reference disagree about the lengths of measuring sticks and the speeds of clocks, but the laws of physics are valid and self-consistent in either frame of reference. Similarly, observers in different frames of reference disagree about what electric and magnetic fields there are, but they agree about concrete physical events. An observer in frame of reference k/1 says there are electric fields around the particles, and predicts that as time goes on, the particles will begin to accelerate towards one another, eventually colliding. She explains the collision as being due to the electrical attraction between the particles. A different observer, k/2, says the particles are moving. This observer also predicts that the particles will collide, but explains their motion in terms of both an electric field and a magnetic field. As we'll see shortly, the magnetic field is *required* in order to maintain consistency between the predictions made in the two frames of reference.

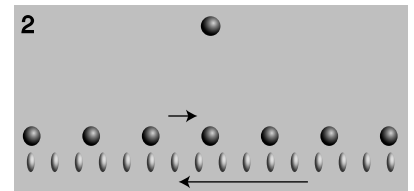
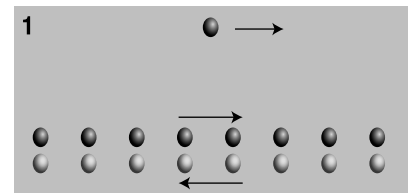
To see how this really works out, we need to find a nice simple example. An example like figure k is *not* easy to handle, because in the second frame of reference, the moving charges create fields that change over time at any given location, like when the V-shaped wake of a speedboat washes over a buoy. Examples like figure j are easier, because there is a steady flow of charges, and all the fields stay the same over time. Figure l/1 shows a simplified and idealized model of figure j. The charge by itself is like one of the charged particles in the vacuum tube beam of figure j, and instead of the wire, we have two long lines of charges moving in opposite directions. Note that, as discussed in discussion question C on page 104, the currents of the two lines of charges do not cancel out. The dark balls represent particles with one type of charge, and the light balls have the other type. Because of this, the total current in the “wire” is double what it would be if we took away one line.

As a model of figure j, figure l/1 is partly realistic and partly unrealistic. In a real piece of copper wire, there are indeed charged particles of both types, but it turns out that the particles of one type (the protons) are locked in place, while only some of the other type (the electrons) are free to move. The model also shows the particles moving in a simple and orderly way, like cars on a two-lane road, whereas in reality most of the particles are organized into copper atoms, and there is also a great deal of random thermal motion. The model's unrealistic features aren't a problem, because the point of this exercise is only to find one particular situation that shows magnetic effects must exist based on relativity.

What electrical force does the lone particle in figure l/1 feel? Since the density of “traffic” on the two sides of the “road” is equal,



k / One observer sees an electric field, while the other sees both an electric field and a magnetic one.



l / A model of a charged particle and a current-carrying wire, seen in two different frames of reference. The relativistic length contraction is highly exaggerated. The force on the lone particle is purely magnetic in 1, and purely electric in 2.

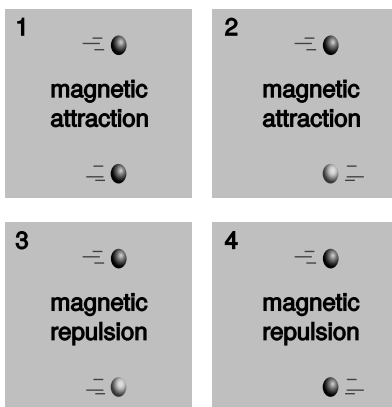
there is zero overall electrical force on the lone particle. Each “car” that attracts the lone particle is paired with a partner on the other side of the road that repels it. If we didn’t know about magnetism, we’d think this was the whole story: the lone particle feels no force at all from the wire.

Figure 1/2 shows what we’d see if we were observing all this from a frame of reference moving along with the lone charge. Here’s where the relativity comes in. Relativity tells us that moving objects appear contracted to an observer who is not moving along with them. Both lines of charge are in motion in both frames of reference, but in frame 1 they were moving at equal speeds, so their contractions were equal. In frame 2, however, their speeds are unequal. The dark charges are moving more slowly than in frame 1, so in frame 2 they are less contracted. The light-colored charges are moving more quickly, so their contraction is greater now. The “cars” on the two sides of the “road” are no longer paired off, so the electrical forces on the lone particle no longer cancel out as they did in 1/1. The lone particle is attracted to the wire, because the particles attracting it are more dense than the ones repelling it. Furthermore, the attraction felt by the lone charge must be purely electrical, since the lone charge is at rest in this frame of reference, and magnetic effects occur only between moving charges and other moving charges.

Now observers in frames 1 and 2 disagree about many things, but they do agree on concrete events. Observer 2 is going to see the lone particle drift toward the wire due to the wire’s electrical attraction, gradually speeding up, and eventually hit the wire. If 2 sees this collision, then 1 must as well. But 1 knows that the total electrical force on the lone particle is exactly zero. There must be some new type of force. She invents a name for this new type of force: magnetism. This was a particularly simple example, because the force was purely magnetic in one frame of reference, and purely electrical in another. In general, an observer in a certain frame of reference will measure a mixture of electric and magnetic fields, while an observer in another frame, in motion with respect to the first, says that the same volume of space contains a different mixture.

We therefore arrive at the conclusion that electric and magnetic phenomena aren’t separate. They’re different sides of the same coin. We refer to electric and magnetic interactions collectively as electromagnetic interactions. Our list of the fundamental interactions of nature now has two items on it instead of three: gravity and electromagnetism.

The basic rules for magnetic attractions and repulsions, shown in figure m, aren’t quite as simple as the ones for gravity and electricity. Rules m/1 and m/2 follow directly from our previous analysis of figure 1. Rules 3 and 4 are obtained by flipping the type of charge that the bottom particle has. For instance, rule 3 is like rule 1,

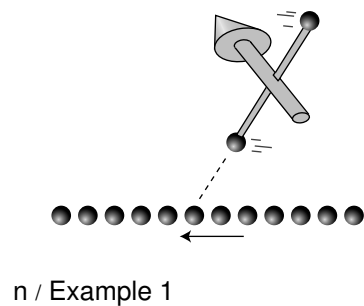


m / Magnetic interactions involving only two particles at a time. In these figures, unlike figure 1/1, there are electrical forces as well as magnetic ones. The electrical forces are not shown here. Don’t memorize these rules!

except that the bottom charge is now the opposite type. This turns the attraction into a repulsion. (We know that flipping the charge reverses the interaction, because that's the way it works for electric forces, and magnetic forces are just electric forces viewed in a different frame of reference.)

*A magnetic weathervane placed near a current. example 1*

Figure n shows a magnetic weathervane, consisting of two charges that spin in circles around the axis of the arrow. (The magnetic field doesn't cause them to spin; a motor is needed to get them to spin in the first place.) Just like the magnetic compass in figure h, the weathervane's arrow tends to align itself in the direction perpendicular to the wire. This is its preferred orientation because the charge close to the wire is attracted to the wire, while the charge far from the wire is repelled by it.

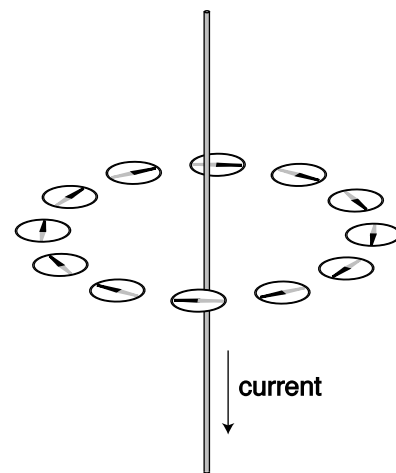


**Magnetic fields**

How should we define the magnetic field? When two objects attract each other gravitationally, their gravitational energy depends only on the distance between them, and it seems intuitively reasonable that we define the gravitational field arrows like a street sign that says "this way to lower gravitational energy." The same idea works fine for the electric field. But what if two charged particles are interacting magnetically? Their interaction doesn't just depend on the distance, but also on their motions.

We need some way to pick out some direction in space, so we can say, "this is the direction of the magnetic field around here." A natural and simple method is to define the magnetic field's direction according to the direction a compass points. Starting from this definition we can, for example, do experiments to show that the magnetic field of a current-carrying wire forms a circular pattern, o.

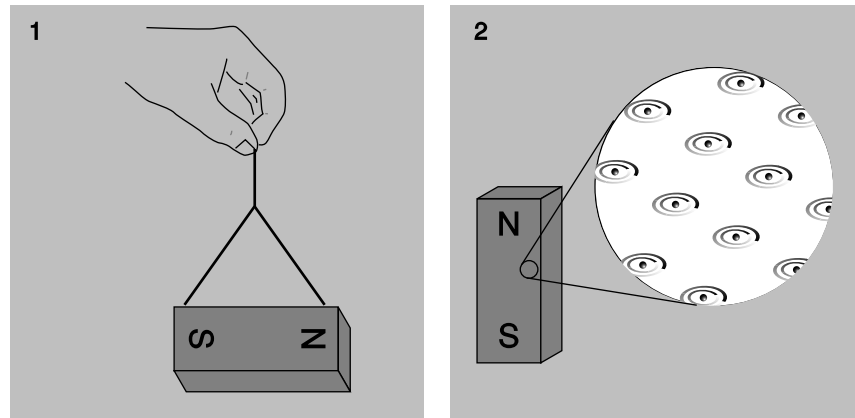
But is this the right definition? Unlike the definitions of the gravitational and electric fields' directions, it involves a particular human-constructed tool. However, compare figure h on page 115 with figure n on page 119. Note that both of these tools line themselves up along a line that's perpendicular to the wire. In fact, no matter how hard you try, you will never be able to invent any other electromagnetic device that will align itself with any other line. All you can do is make one that points in exactly the opposite direction, but along the same line. For instance, you could use paint to reverse the colors that label the ends of the magnetic compass needle, or you could build a weathervane just like figure n, but spinning like a left-handed screw instead of a right-handed one. The weathervane and the compass aren't even as different as they appear. Figure p shows their hidden similarities.



o / The magnetic field curls around the wire in circles. At each point in space, the magnetic compass shows the direction of the field.

Nature is trying to tell us something: there really is something

p / 1. The needle of a magnetic compass is nothing more than a bar magnet that is free to rotate in response to the earth's magnetic field. 2. A cartoon of the bar magnet's structure at the atomic level. Each atom is very much like the weathervane of figure n.



special about the direction the compass points. Defining the direction of the magnetic field in terms of this particular device isn't as arbitrary as it seems. The only arbitrariness is that we could have built up a whole self-consistent set of definitions that started by defining the magnetic field as being in the opposite direction.

*Head-to-tail alignment of bar magnets* example 2

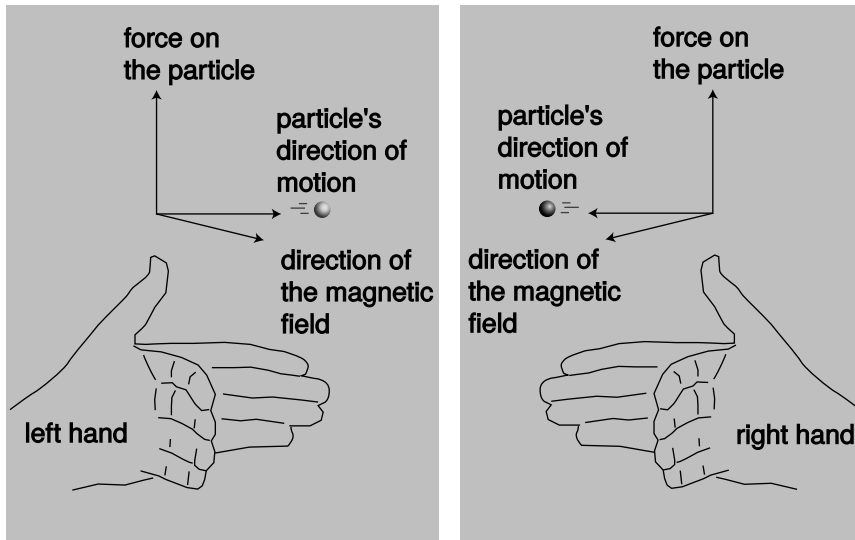
▷ If you let two bar magnets like the one in figure p interact, which way do they want to line up, head-to-head or head-to-tail?

▷ Each bar magnet contains a huge number of atoms, but that won't matter for our result; we can imagine this as an interaction between two individual atoms. For that matter, let's model the atoms as weathervanes like the one in figure n. Suppose we put two such weather vanes side by side, with their arrows both pointing away from us. From our point of view, they're both spinning clockwise. As one of the charges in the left-hand weather vane comes down on the right side, one of the charges in the right-hand vane comes up on the left side. These two charges are close together, so their magnetic interaction is very strong at this moment. Their interaction is repulsive, so this is an unstable arrangement of the two weathervanes.

On the other hand, suppose the left-hand weathervane is pointing away from us, while its partner on the right is pointing toward us. From our point of view, we see the one on the right spinning counterclockwise. At the moment when their charges come as close as possible, they're both on the way up. Their interaction is attractive, so this is a stable arrangement.

Translating back from our model to the original question about bar magnets, we find that bar magnets will tend to align themselves head-to-tail. This is easily verified by experiment.

If you go back and apply this definition to all the examples we've encountered so far, you'll find that there's a general rule: the force on



$q$  / The force on a charged particle moving through a magnetic field is perpendicular to both the field and its direction of motion. The relationship is right-handed for one type of charge, and left-handed for the other type.

a charged particle moving through a magnetic field is perpendicular to both the field and its direction of motion. A force perpendicular to the direction of motion is exactly what is required for circular motion, so we find that a charged particle in a vacuum will go in a circle around the magnetic field arrows (or perhaps a corkscrew pattern, if it also has some motion along the direction of the field). That means that magnetic fields tend to trap charged particles.

Figure r shows this principle in action. A beam of electrons is created in a vacuum tube, in which a small amount of hydrogen gas has been left. A few of the electrons strike hydrogen molecules, creating light and letting us see the path of the beam. A magnetic field is produced by passing a current (meter) through the circular coils of wire in front of and behind the tube. In the bottom figure, with the magnetic field turned on, the force perpendicular to the electrons' direction of motion causes them to move in a circle.

**Sunspots**

*example 3*

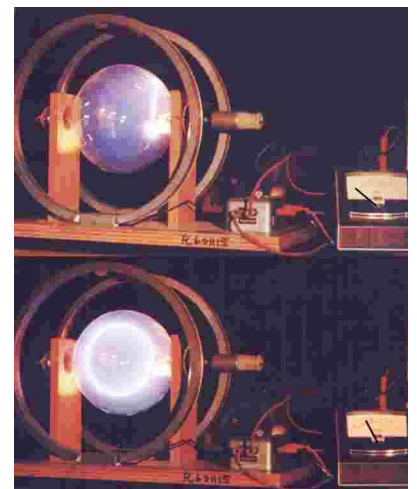
Sunspots, like the one shown in the photo on page 111, are places where the sun's magnetic field is unusually strong. Charged particles are trapped there for months at a time. This is enough time for the sunspot to cool down significantly, and it doesn't get heated back up because the hotter surrounding material is kept out by the same magnetic forces.

**The aurora and life on earth's surface**

*example 4*

A strong magnetic field seems to be one of the prerequisites for the existence of life on the surface of a planet. Energetic charged particles from the sun are trapped by our planet's magnetic field, and harmlessly spiral down to the earth's surface at the poles. In addition to protecting us, this creates the aurora, or "northern lights."

The astronauts who went to the moon were outside of the earth's



r / A beam of electrons circles around the magnetic field arrows.





s / Faraday on a British banknote.

protective field for about a week, and suffered significant doses of radiation during that time. The problem would be much more serious for astronauts on a voyage to Mars, which would take at least a couple of years. They would be subjected to intense radiation while in interplanetary space, and also while on Mars's surface, since Mars lacks a strong magnetic field.

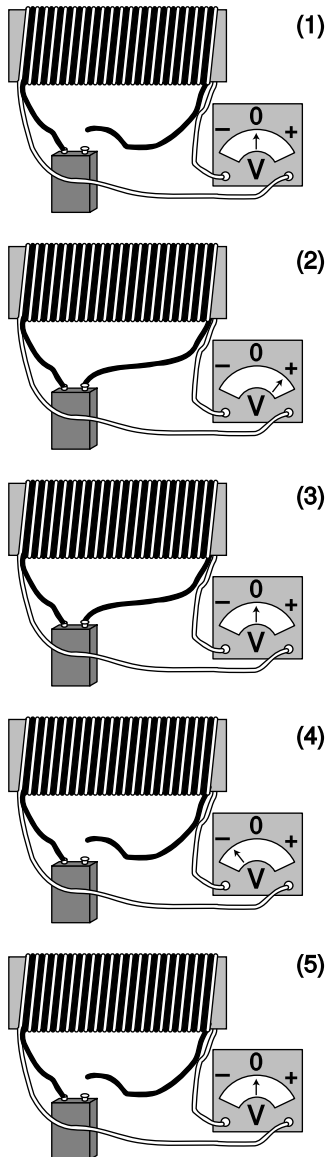
Features in one Martian rock have been interpreted by some scientists as fossilized bacteria. If single-celled life evolved on Mars, it has presumably been forced to stay below the surface. (Life on Earth probably evolved deep in the oceans, and most of the Earth's biomass consists of single-celled organisms in the oceans and deep underground.)

### 6.3 Induction

We've already seen that the electric and magnetic fields are closely related, since what one observer sees as one type of field, another observer in a different frame of reference sees as a mixture of both. The relationship goes even deeper than that, however. Figure t shows an example that doesn't even involve two different frames of reference. This phenomenon of induced electric fields — fields that are not due to charges — was a purely experimental accomplishment by Michael Faraday (1791-1867), the son of a blacksmith who had to struggle against the rigid class structure of 19th century England. Faraday, working in 1831, had only a vague and general idea that electricity and magnetism were related to each other, based on Oersted's demonstration, a decade before, that magnetic fields were caused by electric currents.

Figure t is a simplified drawing of the experiment, as described in Faraday's original paper: "Two hundred and three feet of copper wire ... were passed round a large block of wood; [another] two hundred and three feet of similar wire were interposed as a spiral between the turns of the first, and metallic contact everywhere prevented by twine [insulation]. One of these [coils] was connected with a galvanometer [voltmeter], and the other with a battery... When the contact was made, there was a sudden and very slight effect at the galvanometer, and there was also a similar slight effect when the contact with the battery was broken. But whilst the ... current was continuing to pass through the one [coil], no ... effect ... upon the other [coil] could be perceived, although the active power of the battery was proved to be great, by its heating the whole of its own coil [through ordinary resistive heating] ..."

From Faraday's notes and publications, it appears that the situation in figure t/3 was a surprise to him, and he probably thought it would be a surprise to his readers, as well. That's why he offered evidence that the current was still flowing: to show that the battery hadn't just died. The induction effect occurred during the short



t / Faraday's experiment, simplified and shown with modern equipment.

time it took for the black coil's magnetic field to be established,  $t/2$ . Even more counterintuitively, we get an effect, equally strong but in the opposite direction, when the circuit is *broken*,  $t/4$ . The effect occurs only when the magnetic field is changing: either ramping up or ramping down.

What are we really measuring here with the voltmeter? A voltmeter is nothing more than a resistor with an attachment for measuring the current through it. A current will not flow through a resistor unless there is some electric field pushing the electrons, so we conclude that the changing *magnetic field* has produced an *electric field* in the surrounding space. Since the white wire is not a perfect conductor, there must be electric fields in it as well. The remarkable thing about the circuit formed by the white wire is that as the electrons travel around and around, they are always being pushed forward by electric fields. That is, the electric field seems to form a curly pattern, like a whirlpool.

What Faraday observed was an example of the following principle:

#### the principle of induction

Any magnetic field that changes over time will create an electric field. The induced electric field is perpendicular to the magnetic field, and forms a curly pattern around it.

Any electric field that changes over time will create a magnetic field. The induced magnetic field is perpendicular to the electric field, and forms a curly pattern around it.

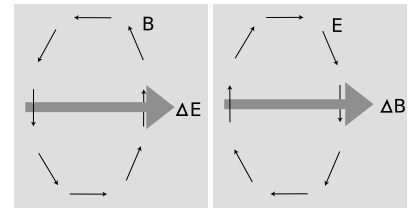
The first part was the one Faraday had seen in his experiment. The geometrical relationships are illustrated in figure u. In Faraday's setup, the magnetic field was pointing along the axis of the coil of wire, so the induced electric field made a curly pattern that circled around the circumference of the block.

#### The generator

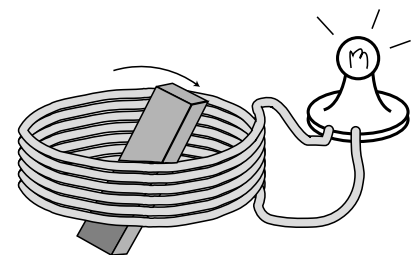
#### example 5

A basic generator, *v*, consists of a permanent magnet that rotates within a coil of wire. The magnet is turned by a motor or crank, (not shown). As it spins, the nearby magnetic field changes. This changing magnetic field results in an electric field, which has a curly pattern. This electric field pattern creates a current that whips around the coils of wire, and we can tap this current to light the lightbulb.

If the magnet was on a frictionless bearing, could we light the bulb for free indefinitely, thus violating conservation of energy? No. It's hard work to crank the magnet, and that's where the energy comes from. If we break the light-bulb circuit, it suddenly gets easier to crank the magnet! This is because the current in the



u / The geometry of induced fields. The induced field tends to form a whirlpool pattern around the change in the field producing it. The notation  $\Delta$  (Greek letter delta) stands for "change in." Note how the induced fields circulate in opposite directions.

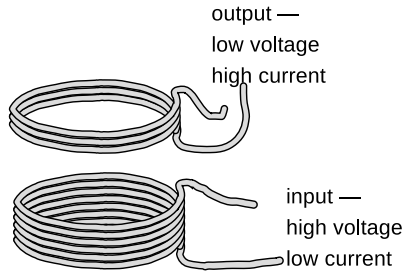


v / A generator.

coil sets up its own magnetic field, and that field exerts a torque on the magnet. If we stopped cranking, this torque would quickly make the magnet stop turning.

*self-check A*

When you're driving your car, the engine recharges the battery continuously using a device called an alternator, which is really just a generator. Why can't you use the alternator to start the engine if your car's battery is dead? ▷ Answer, p. 180



w / A transformer.

*The transformer*

*example 6*

It's more efficient for the electric company to transmit power over electrical lines using high voltages and low currents. However, we don't want our wall sockets to operate at 10000 volts! For this reason, the electric company uses a device called a transformer, w, to convert everything to lower voltages and higher currents inside your house. The coil on the input side creates a magnetic field. Transformers work with alternating current (currents that reverses its direction many times a second), so the magnetic field surrounding the input coil is always changing. This induces an electric field, which drives a current around the output coil.

Since the electric field is curly, an electron can keep gaining more and more energy by circling through it again and again. Thus the output voltage can be controlled by changing the number of turns of wire on the output side. In any case, conservation of energy guarantees that the amount of power on the output side must equal the amount put in originally,

$$(\text{input current}) \times (\text{input voltage}) = (\text{output current}) \times (\text{output voltage})$$

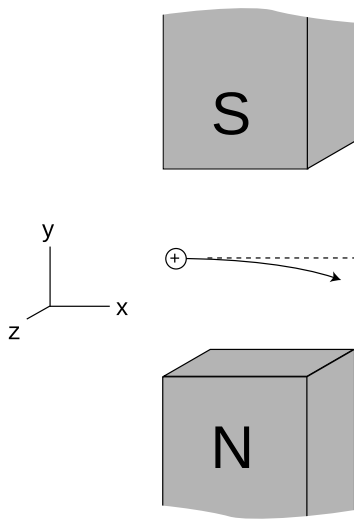
so no matter what factor the voltage is reduced by, the current is increased by the same factor. This is analogous to a lever. A crowbar allows you to lift a heavy boulder, but to move the boulder a centimeter, you may have to move your end of the lever a meter. The advantage in force comes with a disadvantage in distance. It's as though you were allowed to lift a small weight through a large height rather than a large weight through a small height. Either way, the energy you expend is the same.

*Fun with sparks*

*example 7*

Unplug a lamp while it's turned on, and watch the area around the wall outlet. You should see a blue spark in the air at the moment when the prongs of the plug lose contact with the electrical contacts inside the socket.

This is evidence that, as discussed on page 113, fields contain energy. Somewhere on your street is a transformer, one side of which is connected to the lamp's circuit. When the lamp is plugged in and turned on, there's a complete circuit, and current flows. As current flows through the coils in the transformer, a magnetic field is formed — remember, any time there's moving



x / Observer A sees a positively charged particle moves through a region of upward magnetic field, which we assume to be uniform, between the poles of two magnets. The resulting force along the z axis causes the particle's path to curve toward us.

charge, there will be magnetic fields. Because there is a large number turns in the coils, these fields are fairly strong, and store quite a bit of energy.

When you pull the plug, the circuit is no longer complete, and the current stops. Once the current has disappeared, there's no more magnetic field, which means that some energy has disappeared. Conservation of energy tells us that if a certain amount of energy disappears, an equal amount must reappear somewhere else. That energy goes into making the spark. (Once the spark is gone, its energy remains in the form of heat in the air.)

We now have two connections between electric and magnetic fields. One is the principle of induction, and the other is the idea that according to relativity, observers in different frames of reference must perceive different mixtures of magnetic and electric fields. At the time Faraday was working, relativity was still 70 years in the future, so the relativistic concepts weren't available — to him, his observations were just surprising empirical facts. But in fact, the relativistic idea about frames of reference has a logical connection to the idea of induction.

Figure x is a nice example that can be interpreted either way. Observer A is at rest with respect to the bar magnets, and sees the particle swerving off in the  $z$  direction, as it should according to the right-hand rule. Suppose observer B, on the other hand, is moving to the right along the  $x$  axis, initially at the same speed as the particle. B sees the bar magnets moving to the left and the particle initially at rest but then accelerating along the  $z$  axis in a straight line. It is not possible for a magnetic field to start a particle moving if it is initially at rest, since magnetism is an interaction of moving charges with moving charges. B is thus led to the inescapable conclusion that there is an electric field in this region of space, which points along the  $z$  axis. In other words, what A perceives as a pure magnetic field, B sees as a mixture of electric and magnetic fields. This is what we expect based on the relativistic arguments, but it's also what's required by the principle of induction. In B's frame of reference, there's initially no magnetic field, but then a couple of bar magnets come barging in and create one. This is a change in the magnetic field, so the principle of induction predicts that there must be an electric field as well.

### Electromagnetic waves

Theorist James Clerk Maxwell was the first to work out the principle of induction (including the detailed numerical and geometric relationships, which we won't go into here). Legend has it that it was on a starry night that he first realized the most important implication of his equations: light itself is an electromagnetic wave, a ripple spreading outward from a disturbance in the electric and magnetic fields. He went for a walk with his wife, and told her

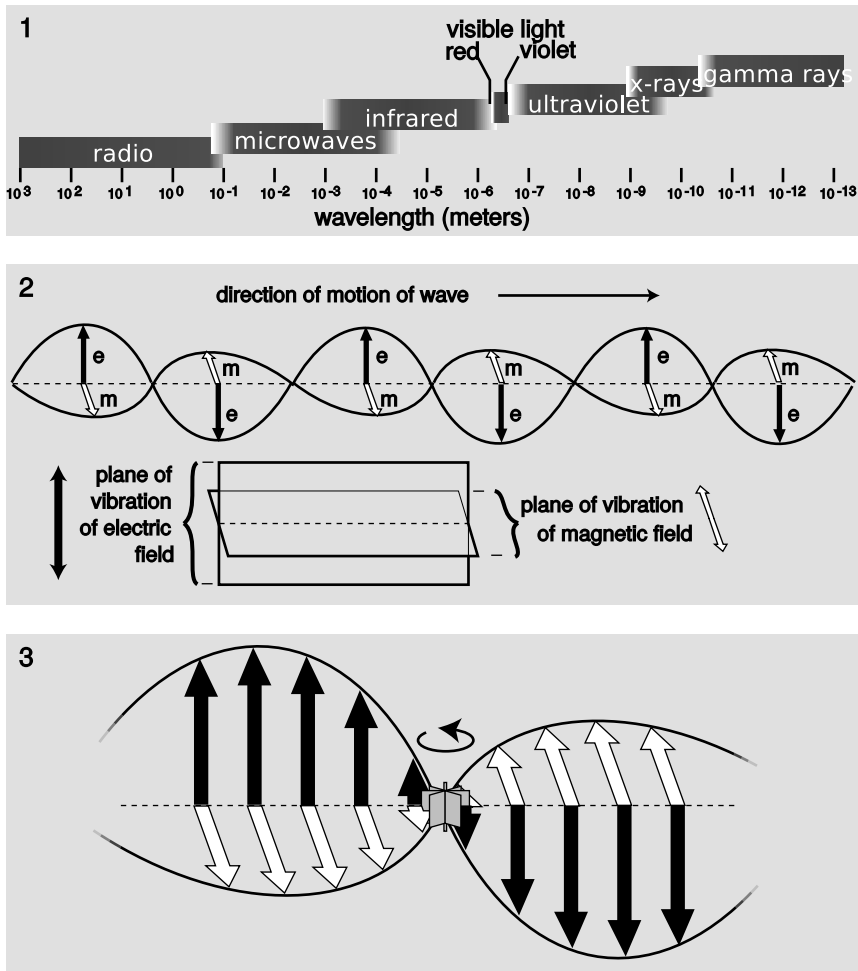


Figure x / James Clerk Maxwell (1831-1879)

she was the only other person in the world who really knew what starlight was.

The principle of induction tells us that there can be no such thing as a purely electric or purely magnetic wave. As an electric wave washes over you, you feel an electric field that changes over time. By the principle of induction, there must also be a magnetic field accompanying it. It works the other way, too. It may seem a little spooky that the electric field causes the magnetic field while the magnetic field causes the electric field, but the waves themselves don't seem to worry about it.

The distance from one ripple to the next is called the wavelength of the light. Light with a certain wavelength (about quarter a millionth of a meter) is at the violet end of the rainbow spectrum, while light with a somewhat longer wavelength (about twice as long) is red. Figure z/1 shows the complete spectrum of light waves. Maxwell's equations predict that all light waves have the same structure, regardless of wavelength and frequency, so even though radio and x-rays, for example, hadn't been discovered, Maxwell predicted that such waves would have to exist. Maxwell's 1865 prediction passed an important test in 1888, when Heinrich Hertz published the results of experiments in which he showed that radio waves could be manipulated in the same ways as visible light waves. Hertz showed, for example, that radio waves could be reflected from a flat surface, and that the directions of the reflected and incoming waves were related in the same way as with light waves, forming equal angles with the normal. Likewise, light waves can be focused with a curved, dish-shaped mirror, and Hertz demonstrated the same thing with radio waves using a metal dish.



z / Panel 1 shows the electromagnetic spectrum. Panel 2 shows how an electromagnetic wave is put together. Imagine that this is a radio wave, with a wavelength of a few meters. If you were standing inside the wave as it passed through you, you could theoretically hold a compass in your hand, and it would wiggle back and forth as the magnetic field pattern (white arrows) washed over you. (The vibration would actually be much too rapid to detect this way.) Similarly, you'd experience an electric field alternating between up and down. Panel 3 shows how this relates to the principle of induction. The changing electric field (black arrows) should create a curly magnetic field (white). Is it really curly? Yes, because if we inserted a paddlewheel that responded to electric fields, the field would make the paddlewheel spin counterclockwise as seen from above. Similarly, the changing magnetic field (white) makes an electric field (black) that curls in the clockwise direction as seen from the front.

## Problems

### Key

- ✓ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ★ A difficult problem.

1 Albert Einstein wrote, “What really interests me is whether God had any choice in the creation of the world.” What he meant by this is that if you randomly try to imagine a set of rules — the laws of physics — by which the universe works, you’ll almost certainly come up with rules that don’t make sense. For instance, we’ve seen that if you tried to omit magnetism from the laws of physics, electrical interactions wouldn’t make sense as seen by observers in different frames of reference; magnetism is required by relativity.

The magnetic interaction rules in figure m are consistent with the time-reversal symmetry of the laws of physics. In other words, the rules still work correctly if you reverse the particles’ directions of motion. Now you get to play God (and fail). Suppose you’re going to make an alternative version of the laws of physics by reversing

the direction of motion of only *one* of the eight particles. You have eight choices, and each of these eight choices would result in a new set of physical laws. We can imagine eight alternate universes, each governed by one of these eight sets. Prove that *all* of these modified sets of physical laws are impossible, either because they are self-contradictory, or because they violate time-reversal symmetry.

**2** The purpose of this problem is to show that the magnetic interaction rules shown in figure m can be simplified by stating them in terms of current. Recall that, as discussed in discussion question C on page 104, one type of charge moving in a particular direction produces the same current as the other type of charge moving in the opposite direction. Let's say arbitrarily that the current made by the dark type of charged particle is in the direction it's moving, while a light-colored particle produces a current in the direction opposite to its motion. Redraw all four panels of figure m, replacing each picture of a moving light or dark particle with an arrow showing the direction of the current it makes. Show that the rules for attraction and repulsion can now be made much simpler, and state the simplified rules explicitly.

**3** Physicist Richard Feynman originated a new way of thinking about charge: a charge of a certain type is equivalent to a charge of the opposite type that happens to be moving backward in time! An electron moving backward in time is an antielectron — a particle that has the same mass as an electron, but whose charge is opposite. Likewise we have antiprotons, and antimatter made from antiprotons and antielectrons. Antielectrons occur naturally everywhere around you due to natural radioactive decay and radiation from outer space. A small number of antihydrogen atoms has even been created in particle accelerators!

Show that, for each rule for magnetic interactions shown in m, the rule is still valid if you replace one of the charges with an opposite charge moving in the opposite direction (i.e., backward in time).

**4** Refer to figure r on page 121. Electrons have the type of charge I've been representing with light-colored spheres.

(a) As the electrons in the beam pass over the top of the circle, what is the direction of the force on them? Use what you know about circular motion.

(b) From this information, use figure q on page 121 to determine the direction of the magnetic field (left, right, up, down, into the page, or out of the page).

**5** You can't use a light wave to see things that are smaller than the wavelength of the light.

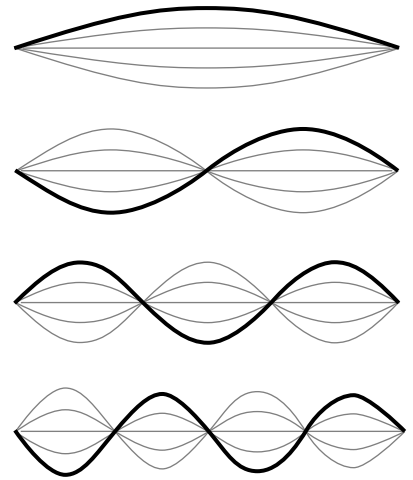
(a) Referring to figure z on page 127, what color of light do you think would be the best to use for microscopy?

(b) The size of an atom is about  $10^{-10}$  meters. Can visible light be

used to make images of individual atoms?

**6** You know how a microwave gets some parts of your food hot, but leaves other parts cold? Suppose someone is trying to convince you of the following explanation for this fact: *The microwaves inside the oven form a stationary wave pattern, like the vibrations of a clothesline or a guitar string. The food is heated unevenly because the wave crests are a certain distance apart, and the parts of the food that get heated the most are the ones where there's a crest in the wave pattern.* Use the wavelength scale in figure z on page 127 as a way of checking numerically whether this is a reasonable explanation.

**7** This book begins and ends with the topic of light. Give an example of how the correspondence principle applies here, referring to a concrete observation from a lab.



Stationary wave patterns on a clothesline (problem 6).







## Chapter 7

# The Ray Model of Light

### 7.1 Light Rays

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.

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. In this chapter, we'll instead make use of a simpler model of light, the ray model, which does a fine job in most practical situations. Not only that, but we'll even backtrack a little and start with a discussion of basic ideas about light and vision that predated the discovery of electromagnetic waves.

## 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.*

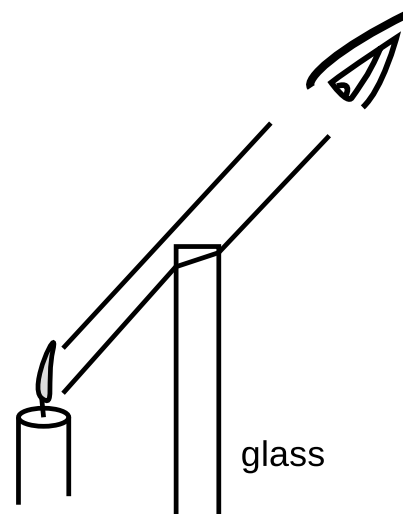
Another issue that few people have considered is whether a candle’s flame simply affects your eye directly, or whether it sends out light which then gets into your eye. Again, the rapidity of the effect makes it difficult to tell what’s happening. If someone throws a rock at you, you can see the rock 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.

Our description of the physics of material objects included both action at a distance (e.g., the earth’s gravitational force on a falling object) and contact forces such as friction.

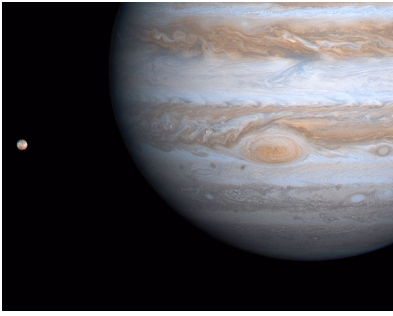
One piece of evidence that the candle sends out stuff that travels to your eye is that as in figure a, intervening transparent substances can make the candle appear to be in the wrong location, suggesting that light is a thing that can be bumped off course. Many people would dismiss this kind of observation 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 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 Newton’s picture of the universe, action at a distance was 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.

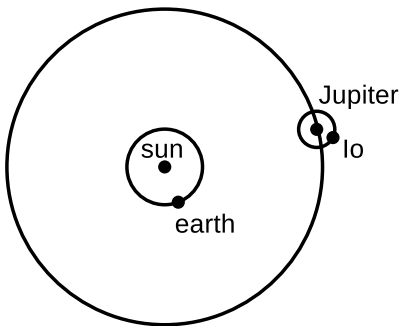
The first person to prove that light’s speed was finite, and to determine it numerically, was Ole Roemer, in a series of measure-



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.



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.

ments 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  $2 \times 10^8$  m/s, which is in the right ballpark compared to modern measurements of  $3 \times 10^8$  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’ve 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.

## 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*

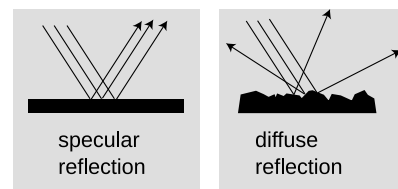
Not all the light energy that hits an object is transformed into heat. Some is reflected, and this leads us to the question of how we see nonluminous objects. If you ask the average person how we see a light bulb, the most likely answer is "The light bulb makes light, which hits our eyes." But if you 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.

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.

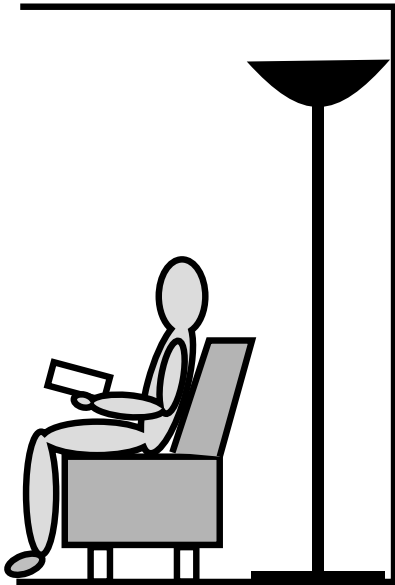


d / Two self-portraits of the author, one taken in a mirror and one with a piece of aluminum foil.



e / Specular and diffuse reflection.

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.



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

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.

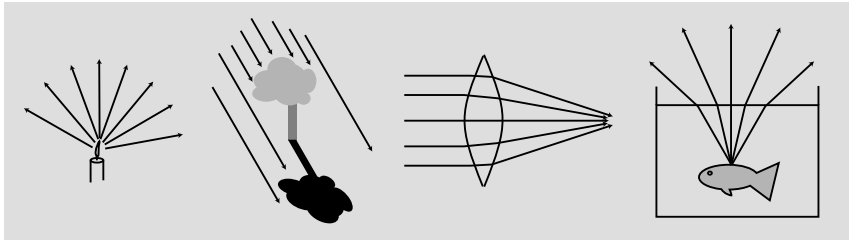
This is where the shortcomings of the Greek theory of vision become glaringly obvious. In the Greek theory, the light from the bulb and my mysterious "eye rays" are both supposed to go to the book, where they collide, allowing me to see the book. But we now have a total of four objects: lamp, eye, book, 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 than dark upholstery.

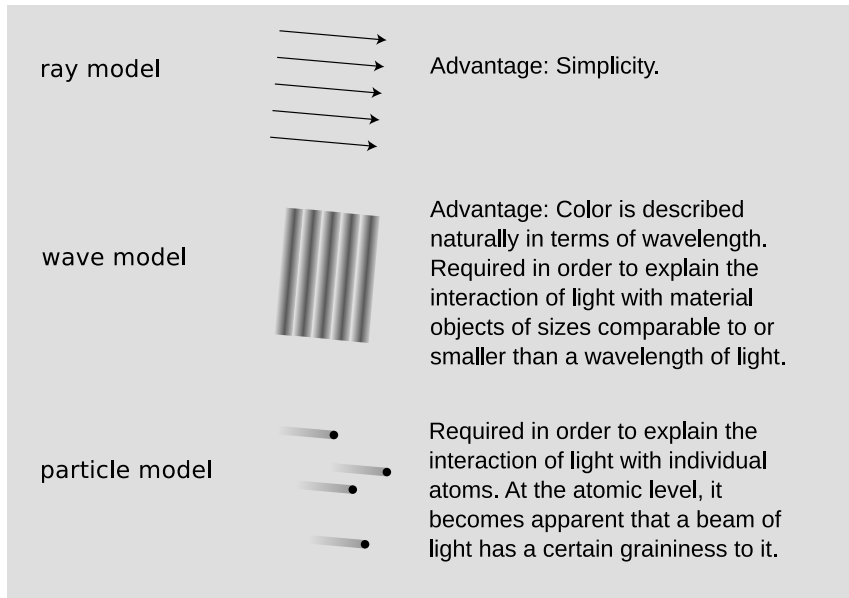
## The ray model of light

### *Models of light*

Note how I've been casually diagramming the motion of light with pictures showing light rays as lines on the page. Figure g shows some more examples. 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. If you've read chapter 6, you've already been introduced to the concept of light as an electromagnetic wave, and 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. Figure h summarizes three models of light.



g / Examples of ray diagrams.



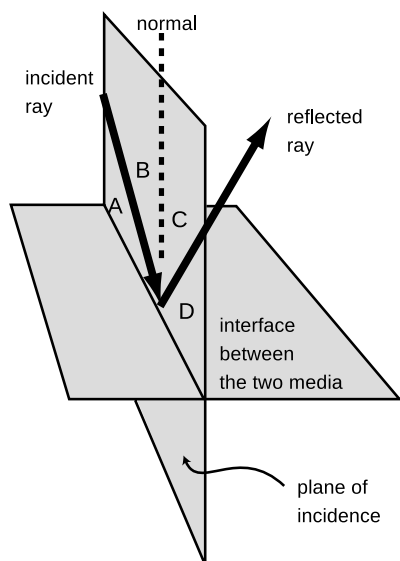
h / Three models of light.

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 rest of this chapter, and with it we can analyze a great many devices and phenomena.

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,





i / The geometry of specular reflection.

and a full understanding of this apparently nonsensical statement would lead us to a discussion of the quantum physics revolution of the twentieth century.

### Geometry of specular reflection

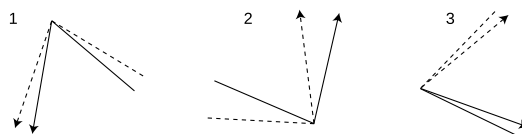
Specular reflection obeys two simple geometrical rules:

- 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  $A=D$ .

#### 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. 180

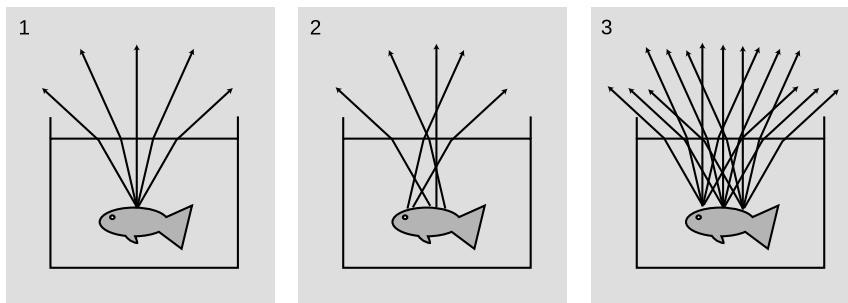
#### Ray diagrams

Figure j shows some guidelines for using ray diagrams effectively. The light rays bend when they 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.



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

### Discussion question

- 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?
- D** 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?
- E** 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?
- F** 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.

## 7.2 Applications

### The inverse-square law

Energy is conserved, so a ray of light should theoretically be able to cross an infinite distance without losing any of its intensity, provided that it's traveling through empty space, so that there's no matter that it can give its energy away to. In that case, why does a distant candle appear dim? Likewise, our sun is just a star like any other star, but it appears much brighter because it's so much closer to us. Why are the other stars so dim if not because their light gets "tired," or "wears out?" It's not that the light rays are stopping, it's that they're getting spread out more thinly. The light comes out of the source in all directions, and if you're very far away, only a tiny percentage of the light will go into your eye. (If all the light from a star went into your eye, you'd be in trouble.)

k / The light is four times dimmer at twice the distance.

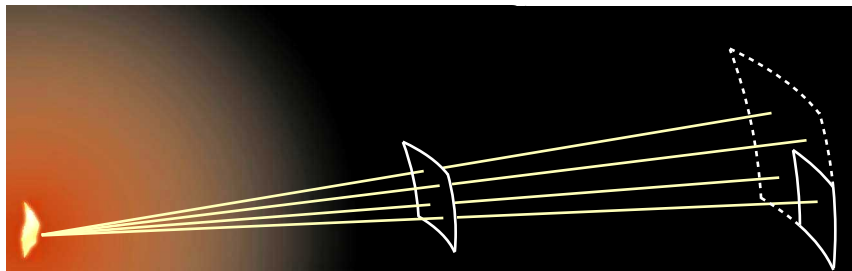


Figure k shows what happens if you double your distance from the source. The light from the flame spreads out in all directions. We pick four representative rays from among those that happen to pass through the nearer square. Of these four, only one passes through the square of equal area at twice the distance. If the two equal-area squares were people's eyes, then only one fourth of the light would go into the more distant person's eye.

Another way of thinking about it is that the light that passed through the first square spreads out and makes a bigger square; at double the distance, the square is twice as wide and twice as tall, so its area is  $2 \times 2 = 4$  times greater. The same light has been spread out over four times the area.

In general, the rule works like this:

$$\begin{aligned} \text{distance} \times 2 &\Rightarrow \text{brightness} \times \frac{1}{4} \\ \text{distance} \times 3 &\Rightarrow \text{brightness} \times \frac{1}{9} \\ \text{distance} \times 4 &\Rightarrow \text{brightness} \times \frac{1}{16} \end{aligned}$$

To get the 4, we multiplied 2 by itself, 9 came from multiplying 3 by itself, and so on. Multiplying a number by itself is called squaring it, and dividing one by a number is called inverting it, so a relationship like this is known as an inverse square law. Inverse square laws are very common in physics: they occur whenever something is spreading out in all directions from a point. Physicists already knew about this kind of inverse square law, for light, before Newton found out that the force of gravity varied as an inverse square, so his law of gravity made sense to them intuitively, and they were ready to accept it. However, Newton's law of gravity doesn't describe gravity as a substance that physically travels outward through space, so it's only a rough analogy. (One modern hypothesis about gravity is that the messages of gravitational attraction between two objects are actually carried by little particles, called gravitons, but nobody has ever detected a graviton directly.)

*self-check B*

Alice is one meter from the candle, while Bob is at a distance of five meters. How many times dimmer is the light at Bob's location?  $\triangleright$

Answer, p. 180

---

*An example with sound*

*example 1*

$\triangleright$  Four castaways are adrift in an open boat, and are yelling to try to attract the attention of passing ships. If all four of them yell at once, how much is their range increased compared to the range they would have if they took turns yelling one at a time?

$\triangleright$  This is an example involving sound. Although sound isn't the same as light, it does spread out in all directions from a source, so it obeys the inverse-square law. In the previous examples, we



1/ The same lens is shown with its diaphragm set to three different apertures.

knew the distance and wanted to find the intensity (brightness). Here, we know about the intensity (loudness), and we want to find out about the distance. Rather than taking a number and multiplying it by itself to find the answer, we need to reverse the process, and find the number that, when multiplied by itself, gives four. In other words, we're computing the square root of four, which is two. They will double their range, not quadruple it.

*Astronomical distance scales* *example 2*

The nearest star, Alpha Centauri,<sup>1</sup> is about 10,000,000,000,000,000 times dimmer than our sun when viewed from our planet. If we assume that Alpha Centauri's true brightness is roughly the same as that of our own sun, then we can find the distance to Alpha Centauri by taking the square root of this number. Alpha Centauri's distance from us is equal to about 100,000,000 times our distance from the sun.

*Pupils and camera diaphragms* *example 3*

In bright sunlight, your pupils contract to admit less light. At night they dilate, becoming bigger "light buckets." Your perception of brightness depends not only on the true brightness of the source and your distance from it, but also on how much area your pupils present to the light. Cameras have a similar mechanism, which is easy to see if you detach the lens and its housing from the body of the camera, as shown in the figure. Here, the diameter of the largest aperture is about ten times greater than that of the smallest aperture. Making a circle ten times greater in radius increases its area by a factor of 100, so the light-gathering power of the camera becomes 100 times greater. (Many people expect that the area would only be ten times greater, but if you start drawing copies of the small circle inside the large circle, you'll see that ten are not nearly enough to fill in the entire area of the larger circle. Both the width and the height of the bigger circle are ten times greater, so its area is 100 times greater.)

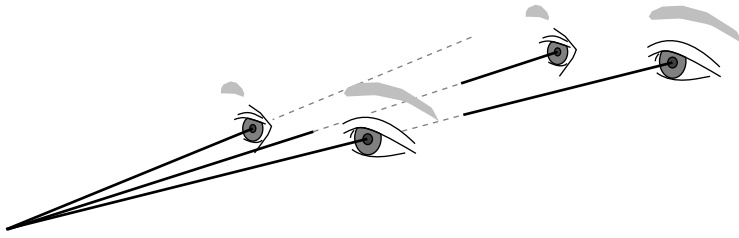
### Parallax

Example 2 on page 142 showed how we can use brightness to determine distance, but your eye-brain system has a different method. Right now, you can tell how far away this page is from your eyes. This sense of depth perception comes from the fact that your two eyes show you the same scene from two different perspectives. If you wink one eye and then the other, the page will appear to shift back and forth a little.

If you were looking at a fly on the bridge of your nose, there would be an angle of nearly  $180^\circ$  between the ray that went into

---

<sup>1</sup>Sticklers will note that the nearest star is really our own sun, and the second nearest is the burned-out cinder known as Proxima Centauri, which is Alpha Centauri's close companion.



m / At double the distance, the parallax angle is approximately halved.

your left eye and the one that went into your right. Your brain would know that this large angle implied a very small distance. This is called the parallax angle. Objects at greater distances have smaller parallax angles, and when the angles are small, it's a good approximation to say that the angle is inversely proportional to the distance. In figure m, the parallax angle is almost exactly cut in half when the person moves twice as far away.

Parallax can be observed in other ways than with a pair of eyeballs. As a child, you noticed that when you walked around on a moonlit evening, the moon seemed to follow you. The moon wasn't really following you, and this isn't even a special property of the moon. It's just that as you walk, you expect to observe a parallax angle between the same scene viewed from different positions of your whole head. Very distant objects, including those on the Earth's surface, have parallax angles too small to notice by walking back and forth. In general, rays coming from a very distant object are nearly parallel.

If your baseline is long enough, however, the small parallaxes of even very distant objects may be detectable. In the nineteenth century, nobody knew how tall the Himalayas were, or exactly where their peaks were on a map, and the Andes were generally believed to be the tallest mountains in the world. The Himalayas had never been climbed, and could only be viewed from a distance. From down on the plains of India, there was no way to tell whether they were very tall mountains very far away, or relatively low ones that were much closer. British surveyor George Everest finally established their true distance, and astounding height, by observing the same peaks through a telescope from different locations far apart.

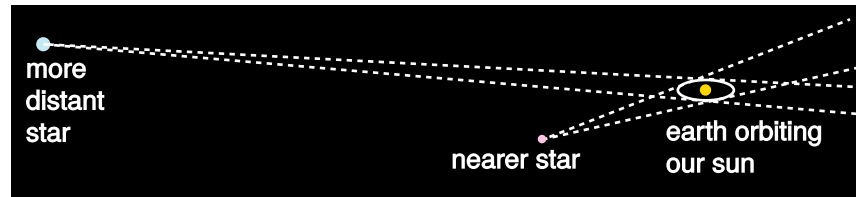
An even more spectacular feat of measurement was carried out by Hipparchus over twenty-one centuries ago. By measuring the parallax of the moon as observed from Alexandria and the Hellespont, he determined its distance to be about 90 times the radius of the earth.<sup>2</sup>

---

<sup>2</sup>The reason this was a hard measurement was that accurate clocks hadn't been invented, so there was no easy way to synchronize the two observations, and the desired effect would be masked by the apparent motion of the moon across the sky as it rose and set. Hipparchus's trick was to do the measurement

The earth circles the sun,  $n$ , and we can therefore determine the distances to a few hundred of the nearest stars by making observations six months apart, so that the baseline for the parallax measurement is the diameter of the earth's orbit. For these stars, the distances derived from parallax can be checked against the ones found by the method of example 2 on page 142. They do check out, which verifies the assumption that the stars are objects analogous to our sun.

$n$  / The nearer star has a larger parallax angle. By measuring the parallax angles, we can determine the distances to both stars. (The scale on this drawing is not realistic. If the earth's orbit was really this size, the nearest stars would be several kilometers away.)



### *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

---

during a solar eclipse, so that people at both locations would know they were in sync.



infrared, you don't emit visible light, because your skin isn't hot enough to glow in the visible spectrum.

*Ray tracing on a computer*

*example 4*

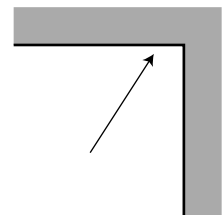
A number of techniques can be used for creating artificial visual scenes in computer graphics. Figure o shows such a scene, which was created by the brute-force technique of simply constructing a very detailed ray diagram on a computer. This technique requires a great deal of computation, and is therefore 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.



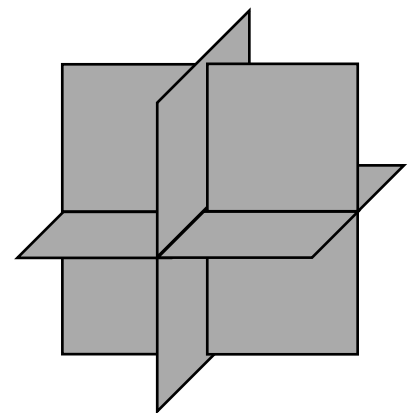
o / This photorealistic image of a nonexistent countertop was produced completely on a computer, by computing a complicated ray diagram.

**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

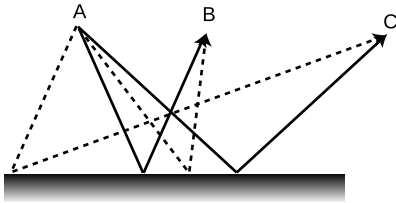


p / Discussion question B.

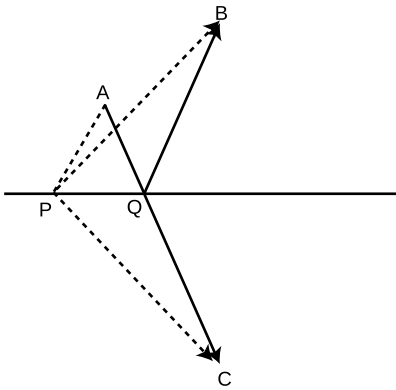


q / Discussion question C.





r / 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.



s / 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.

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?

### 7.3 ★ The Principle of Least Time for Reflection

There is another way of stating the rules of reflection that is very simple and beautiful, and turns out to have deep consequences and apply much more broadly, not just to reflection. 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, r, 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 s, for the more general case where the points are at different distances from the surface.

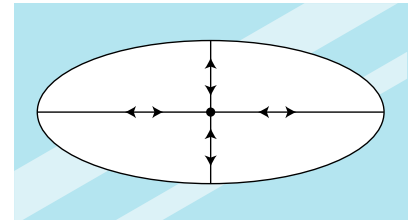
Not only does the principle of least time work for light in a vacuum and light undergoing reflection, we will also see in a later chapter that it works for the bending of light when it passes from one medium into another.

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 an approximate shortcut for finding certain results of the wave model of light.

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 s, 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 t, 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.



t / 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.

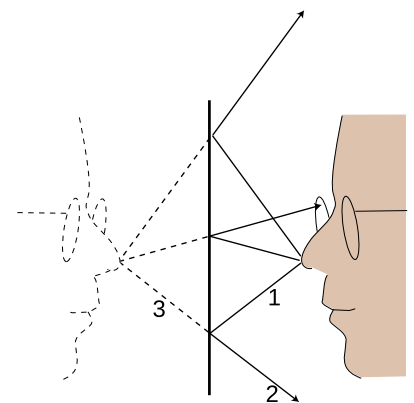
## 7.4 Images by Reflection

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 and 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.

### A virtual image

We can understand a mirror image using a ray diagram. Figure u 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:



u / An image formed by a mirror.

**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 C*

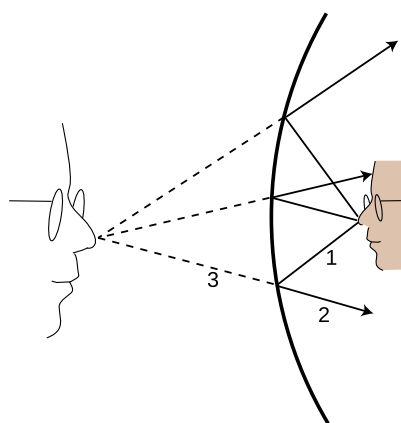
Imagine that the person in figure u moves his face down quite a bit — a couple of feet in real life, or a few inches on this scale drawing. Draw a new ray diagram. Will there still be an image? If so, where is it visible from? ▷ Answer, p. 180

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 1. 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.

### 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.

Because the mirror in figure v 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



v / An image formed by a curved mirror.

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.

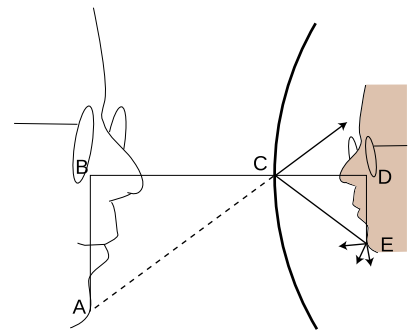
It is easy to prove that the same magnification also applies to the image's other dimensions. Consider a point like E in figure w. The trick is that out of all the rays diffusely reflected by E, we pick the one that happens to head for the mirror's center, C. The equal-angle property of specular reflection 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  $M = 1.9$ .

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.

### A real image

If we start by placing an object very close to the mirror,  $x/1$ , 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,  $x/2$ , the image appears upside-down and in *front* of the mirror.

Here's what's happened. The mirror bends light rays inward, but

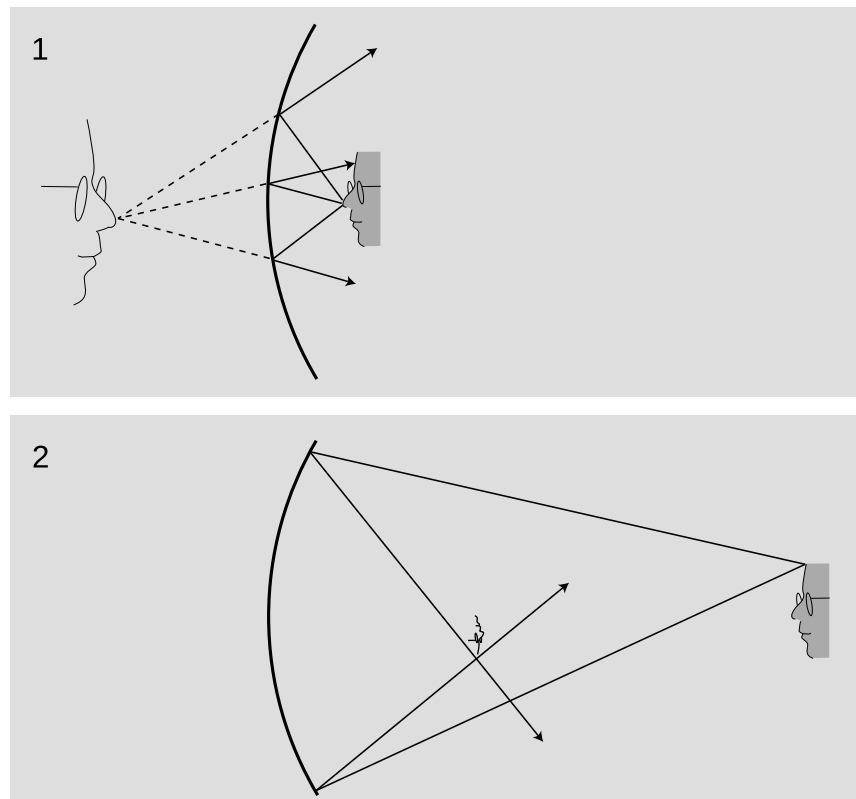


w / The image is magnified by the same factor in depth and in its other dimensions.

when the object is very close to it, as in  $x/1$ , 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,  $x/2$ , 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.

$x/1$ . A virtual image. 2. A real image. As you'll verify in homework problem 9, the image is upside-down



The distinction between a real image and a virtual image is an important one, because a real image can be projected onto a screen or photographic film. If a piece of paper is inserted in figure  $x/2$  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 D*

Sketch another copy of the face in figure  $x/1$ , even farther from the mir-

ror, and draw a ray diagram. What has happened to the location of the image?  
▷ Answer, p. 180

### Images of images

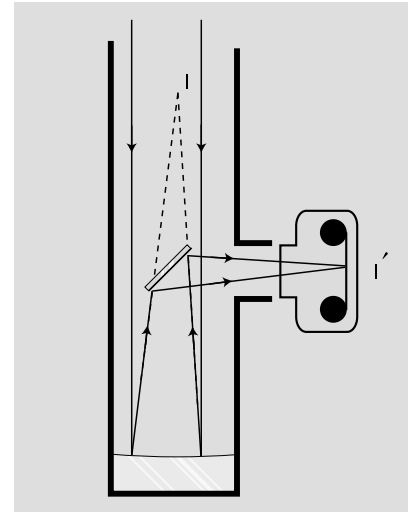
If you are wearing glasses right now, then the light rays from the page are being manipulated first by your glasses and then by the lens of your eye. You might think that it would be extremely difficult to analyze this, but in fact it is quite easy. In any series of optical elements (mirrors or lenses 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 y shows an example involving only mirrors. The Newtonian 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.

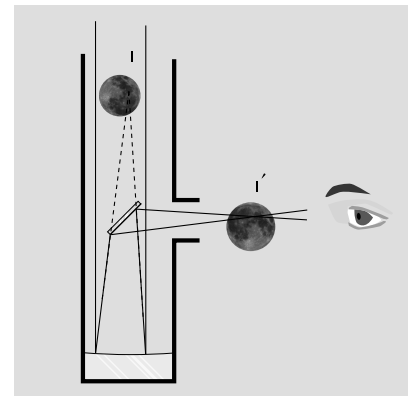
The large curved mirror by itself would form an image  $I$ , but the small flat mirror creates an image of the image,  $I'$ . The relationship between  $I$  and  $I'$  is exactly the same as it would be if  $I$  was an actual object rather than an image:  $I$  and  $I'$  are at equal distances from the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



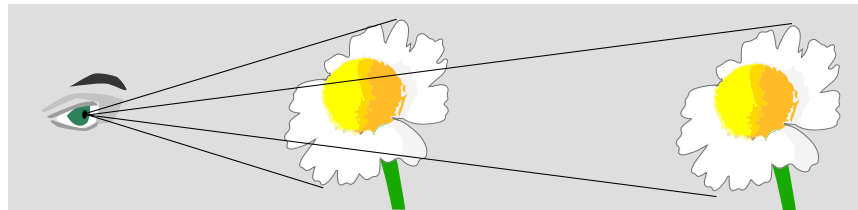
y / A Newtonian telescope being used with a camera.



z / 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.

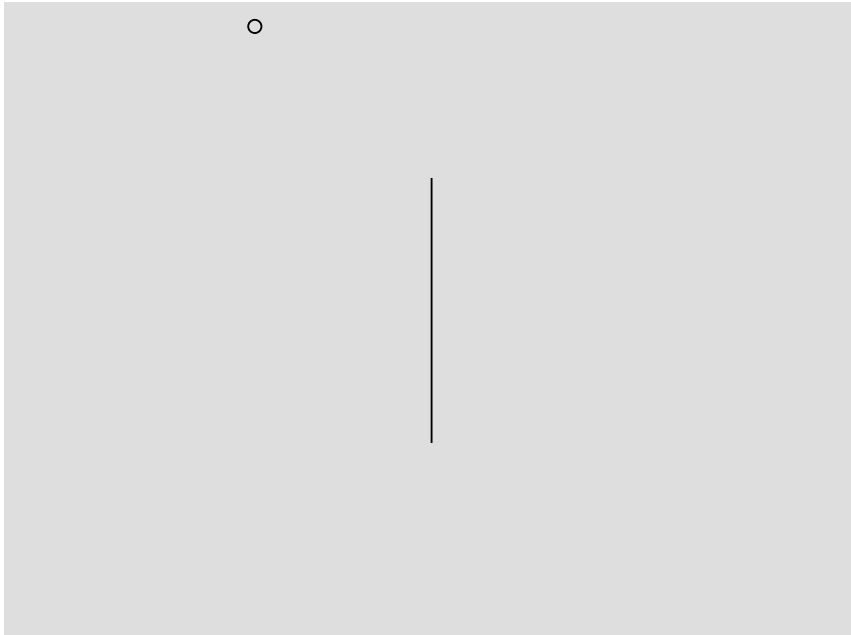
small size of the image has been more than compensated for by the shorter distance. The important thing here is the amount of *angle* within your field of view that the image covers, and it is this angle that has been increased. The factor by which it is increased is called the *angular magnification*,  $M_a$ .

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

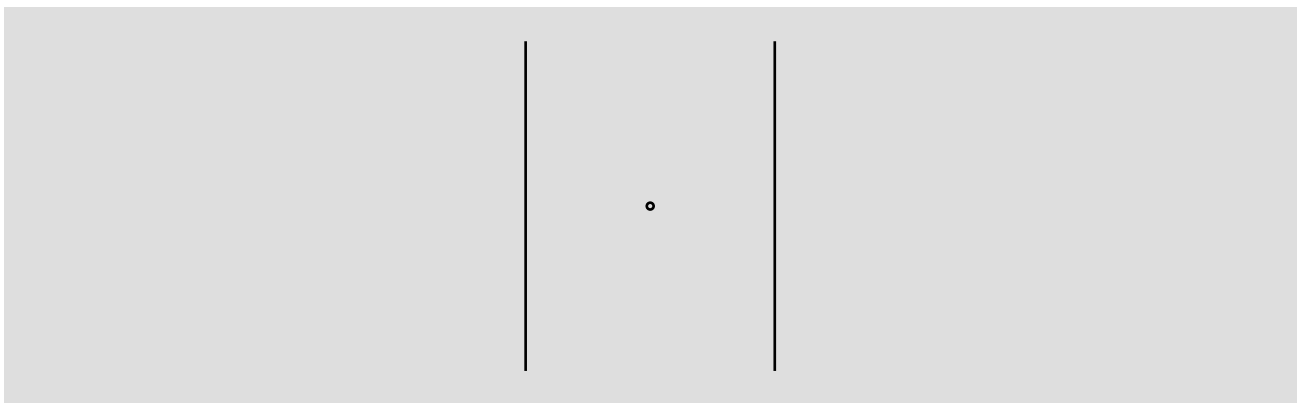


### Discussion questions

**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?

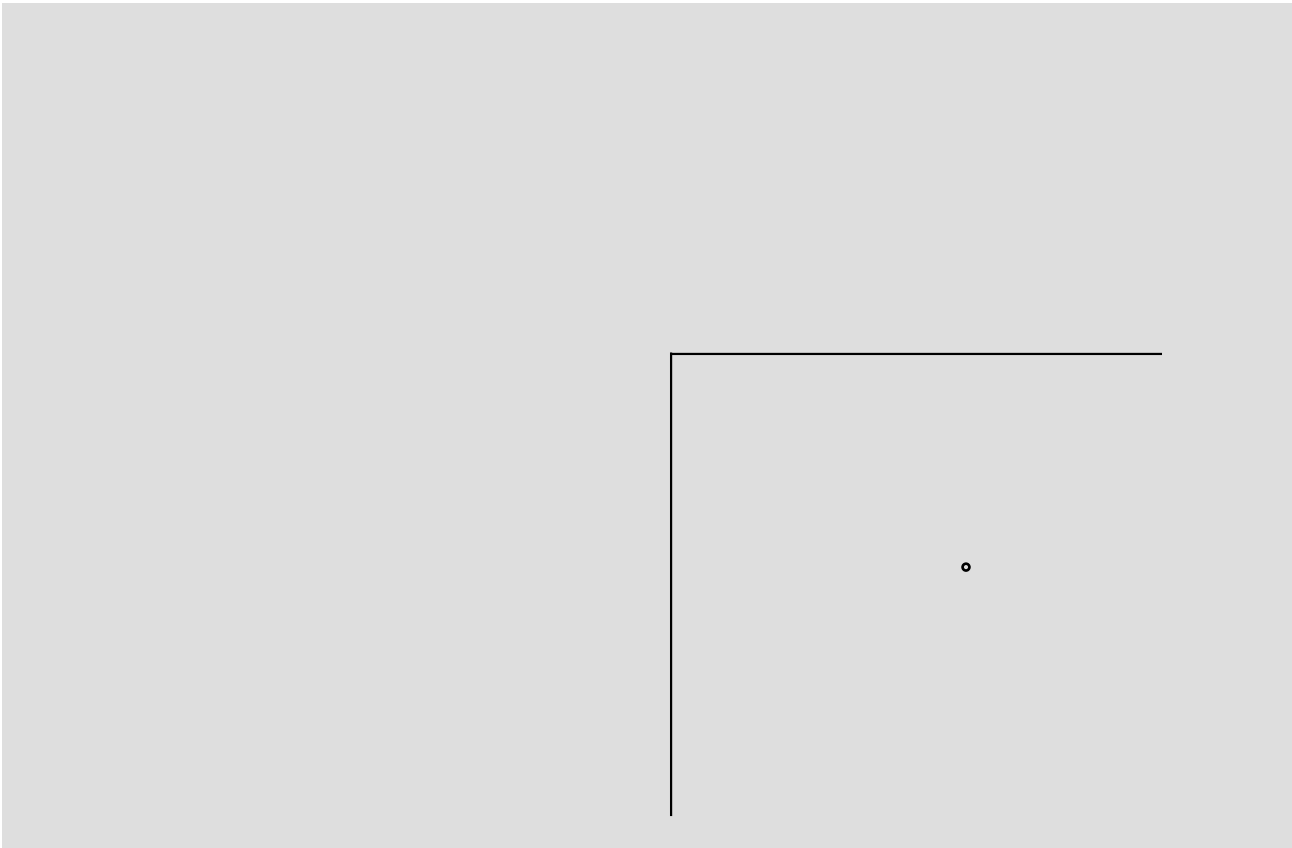


**B** Locate the images of you that will be formed if you stand between two parallel mirrors.

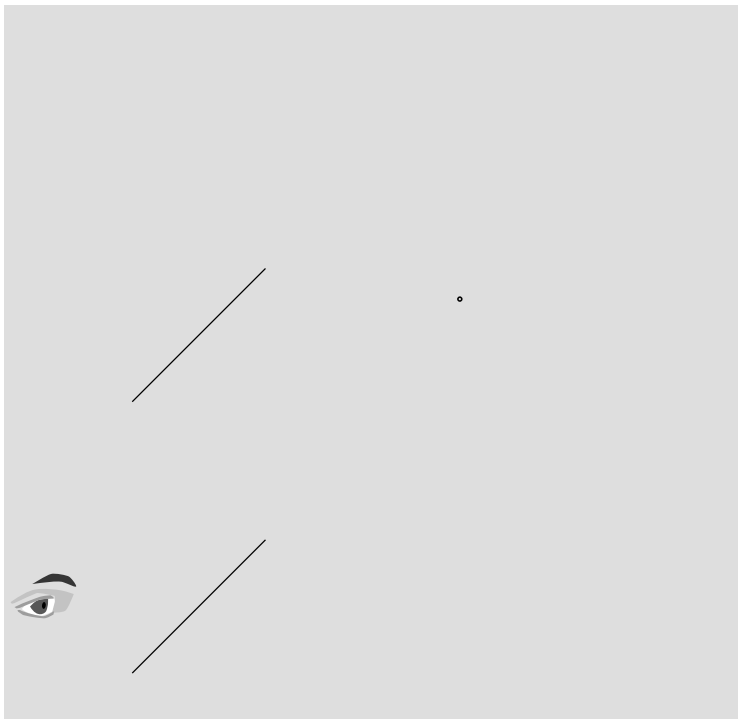




**C** Locate the images formed by two perpendicular mirrors, as in the figure. What happens if the mirrors are not perfectly perpendicular?



D Locate the images formed by the periscope.

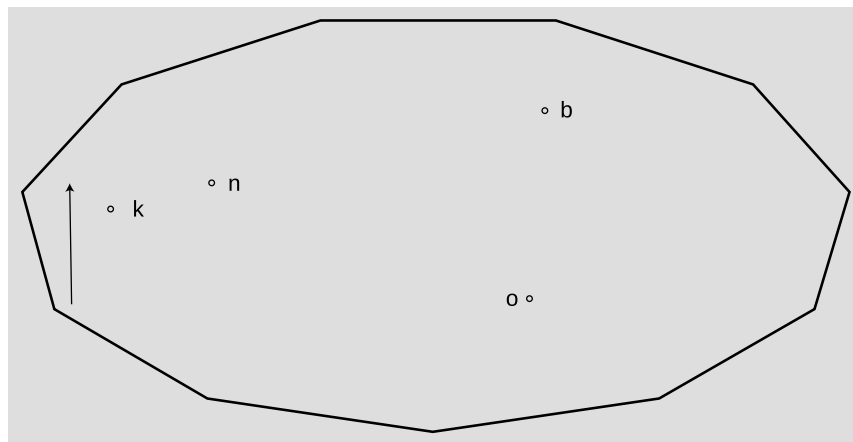


## Problems

### Key

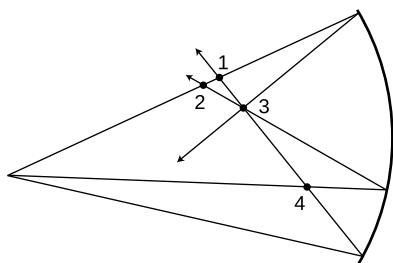
- ✓ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ★ A difficult problem.

**1** 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.



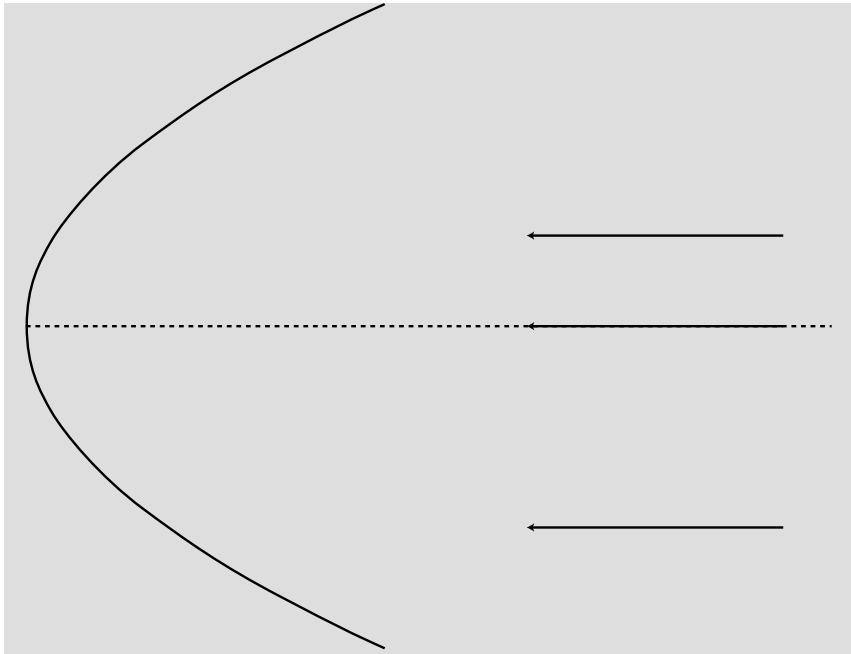
Problem 1.

**2** 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?



Problem 3.

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



**Problem 2.**

**4** 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?

**5** 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).

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

**7** 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.

**8** As discussed in question 4, 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.

**9** In figure x/2 in on page 150, 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.

**10** 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.)

**11** 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?

**12** A Global Positioning System (GPS) receiver is a device that lets you figure out where you are by measuring the time for radio signals to travel between you and the satellite, 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?

**13** Estimate the frequency of an electromagnetic wave whose wavelength is similar in size to an atom (about a nm). Referring back to figure z on p. 127, in what part of the electromagnetic spectrum would such a wave lie (infrared, gamma-rays, ...)?

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





“The Great Wave Off Kanagawa,” by Katsushika Hokusai (1760-1849).

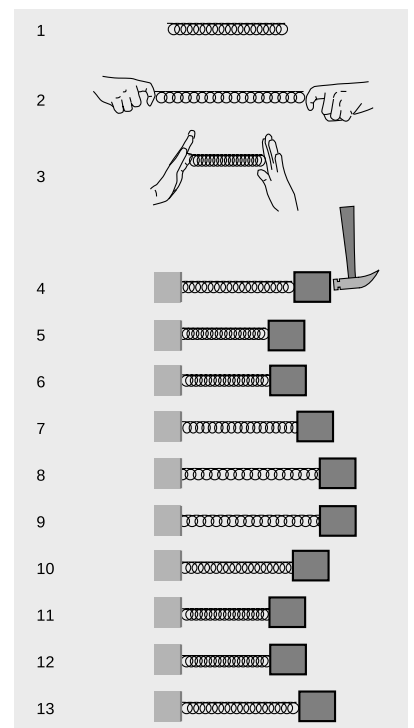
## Chapter 8

# Waves

If you’ve read chapter 6, you’ve been introduced to the idea that the universe isn’t really mechanical in nature. It’s made of fields of force. When a radio antenna makes a disturbance in the electric and magnetic fields, those disturbances travel outward like ripples on a pond. In other words, waves are fundamental to the way the universe works.

### 8.1 Vibrations

Your radio dial is calibrated in units of *frequency*, the simplest example of this concept is provided not by a wave but by a vibrating physical object such as a mass on the end of a spring, 8. With no forces on it, the spring assumes its equilibrium length, 8/1. It can be stretched, 2, or compressed, 3. We attach the spring to a wall on the left and to a mass on the right. If we now hit the mass with a hammer, 4, it oscillates as shown in the series of snapshots, 4-13. If we assume that the mass slides back and forth without friction and that the motion is one-dimensional, then conservation of energy proves that the motion must be repetitive. When the block comes back to its initial position again, 7, its potential energy is the same again, so it must have the same kinetic energy again. The motion is in the opposite direction, however. Finally, at 10, it returns to its



a / A spring has an equilibrium length, 1, and can be stretched, 2, or compressed, 3. A mass attached to the spring can be set into motion initially, 4, and will then vibrate, 4-13.



initial position with the same kinetic energy and the same direction of motion. The motion has gone through one complete cycle, and will now repeat forever in the absence of friction.

The usual physics terminology for motion that repeats itself over and over is periodic motion, and the time required for one repetition is called the period,  $T$ . One complete repetition of the motion is called a cycle.

We are used to referring to short-period sound vibrations as “high” in pitch, and it sounds odd to have to say that high pitches have low periods. It is therefore more common to discuss the rapidity of a vibration in terms of the number of vibrations per second, a quantity called the frequency,  $f$ . Since the period is the number of seconds per cycle and the frequency is the number of cycles per second, they are reciprocals of each other,

$$f = 1/T.$$

Units of inverse second,  $s^{-1}$ , are awkward in speech, so an abbreviation has been created. One Hertz, named in honor of a pioneer of radio technology, is one cycle per second. In abbreviated form,  $1 \text{ Hz} = 1 \text{ s}^{-1}$ . This is the familiar unit used for the frequencies on the radio dial.

*Frequency of a radio station*

*example 1*

▷ KMHD’s frequency is 89.1 MHz. What does this mean, and what period does this correspond to?

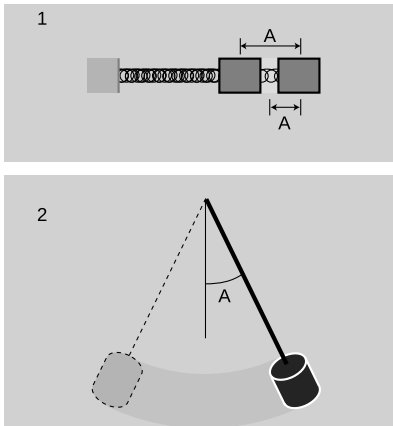
▷ The metric prefix M- is mega-, i.e., millions. The radio waves emitted by KMHD’s transmitting antenna vibrate 89.1 million times per second. This corresponds to a period of

$$T = 1/f = 1.12 \times 10^{-8} \text{ s}.$$

This example shows a second reason why we normally speak in terms of frequency rather than period: it would be painful to have to refer to such small time intervals routinely. I could abbreviate by telling people that KMHD’s period was 11.2 nanoseconds, but most people are more familiar with the big metric prefixes than with the small ones.

Units of frequency are also commonly used to specify the speeds of computers. The idea is that all the little circuits on a computer chip are synchronized by the very fast ticks of an electronic clock, so that the circuits can all cooperate on a task without getting ahead or behind. Adding two numbers might require, say, 30 clock cycles. Microcomputers these days operate at clock frequencies of about a gigahertz.

We have discussed how to measure how fast something vibrates, but not how big the vibrations are. The general term for this is



b / 1. The amplitude of the vibrations of the mass on a spring could be defined in two different ways. It would have units of distance. 2. The amplitude of a swinging pendulum would more naturally be defined as an angle.

amplitude,  $A$ . The definition of amplitude depends on the system being discussed, and two people discussing the same system may not even use the same definition. In the example of the block on the end of the spring, 8.1/1, the amplitude will be measured in distance units such as cm. One could work in terms of the distance traveled by the block from the extreme left to the extreme right, but it would be somewhat more common in physics to use the distance from the center to one extreme. The former is usually referred to as the peak-to-peak amplitude, since the extremes of the motion look like mountain peaks or upside-down mountain peaks on a graph of position versus time.

In other situations we would not even use the same units for amplitude. The amplitude of a child on a swing, or a pendulum, 8.1/2, would most conveniently be measured as an angle, not a distance, since her feet will move a greater distance than her head. The electrical vibrations in a radio receiver would be measured in electrical units such as volts or amperes.

In many physical examples of vibrations, the force that brings the vibrating object back to equilibrium gets stronger and stronger as the object gets fatter and farther from equilibrium, and the force is directly proportional to the distance from equilibrium. Most springs behave this way, for example, so for example we'd expect that the spring in figure 8 would make very nearly twice the force when stretched twice as much. We then define a spring constant,, which tells us how many newtons of force we get per meter of stretching. For example, the John Hancock Tower has a spring constant of about 200 MN/m (meganewtons per meter), meaning that the wind must exert a force of about 200 MN in order to make the tower sway by one meter. To make it sway by two meters, the force would have to be 400 MN.

When the force has this type of mathematical behavior, the resulting motion is known as *simple harmonic motion*. One surprising and useful fact about simple harmonic motion is that its frequency is independent of amplitude. Intuitively, we would expect that vibrations with a greater amplitude would take more time, i.e., have a lower frequency. However, when the amplitude is greater, the force accelerating the mass back toward the equilibrium position is also greater, and this turns out to compensate exactly for the need to travel a greater distance. Legend has it that Galileo first noticed this fact when he watched a chandelier swinging during a church service, and timed it against his pulse. Mathematically, the frequency of vibration is given by  $f = (1/2\pi)\sqrt{k/m}$ , where  $k$  is the spring constant, and  $m$  is the mass that is vibrating.



c / The John Hancock Tower in Boston vibrates naturally at a frequency of 0.14 Hz. Surprisingly, this frequency is the same regardless of the amplitude of the vibrations

d / The two circular patterns of ripples pass through each other. Unlike material objects, wave patterns can overlap in space, and when this happens they combine by addition.



## 8.2 Wave Motion

There are three main ways in which wave motion differs from the motion of objects made of matter.

### 1. Superposition

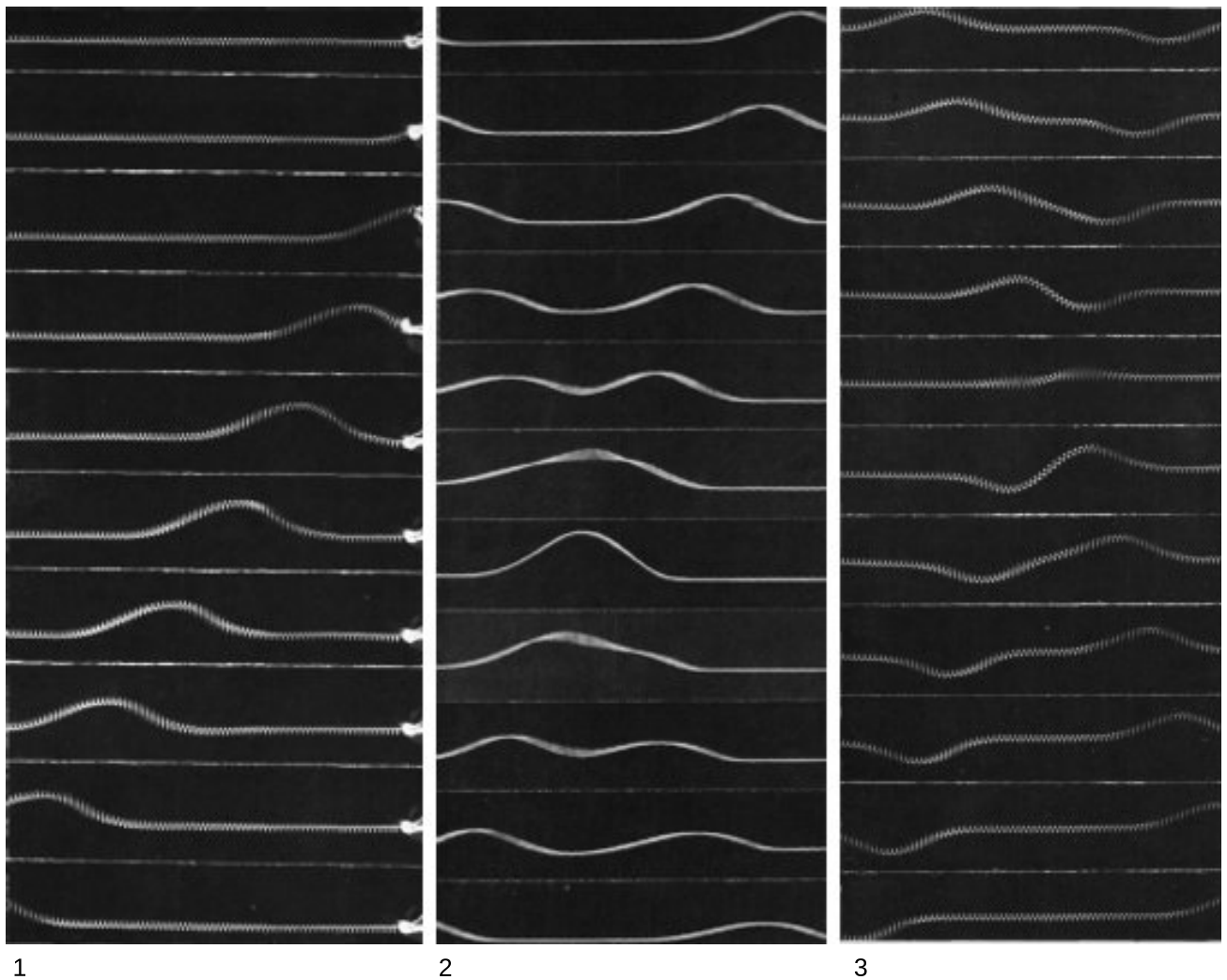
The first, and most profound, difference between wave motion and the motion of objects is that waves do not display any repulsion of each other analogous to the normal forces between objects that come in contact. Two wave patterns can therefore overlap in the same region of space, as shown in the figure at the top of the page. Where the two waves coincide, they add together. For instance, suppose that at a certain location in at a certain moment in time, each wave would have had a crest 3 cm above the normal water level. The waves combine at this point to make a 6-cm crest. We use negative numbers to represent depressions in the water. If both waves would have had a troughs measuring -3 cm, then they combine to make an extra-deep -6 cm trough. A +3 cm crest and a -3 cm trough result in a height of zero, i.e., the waves momentarily cancel each other out at that point. This additive rule is referred to as the principle of superposition, “superposition” being merely a fancy word for “adding.”

Superposition can occur not just with sinusoidal waves like the ones in the figure above but with waves of any shape. The figures on the following page show superposition of wave *pulses*. A pulse is simply a wave of very short duration. These pulses consist only of

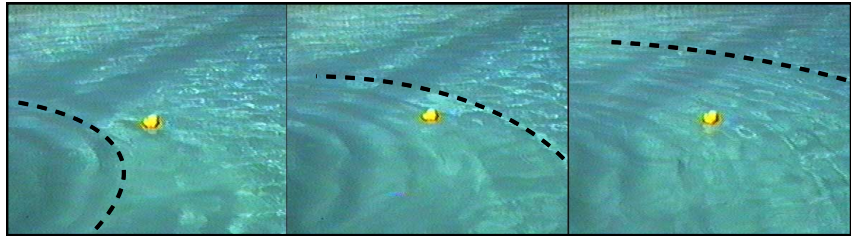
a single hump or trough. If you hit a clothesline sharply, you will observe pulses heading off in both directions. This is analogous to the way ripples spread out in all directions when you make a disturbance at one point on water. The same occurs when the hammer on a piano comes up and hits a string.

### Discussion question

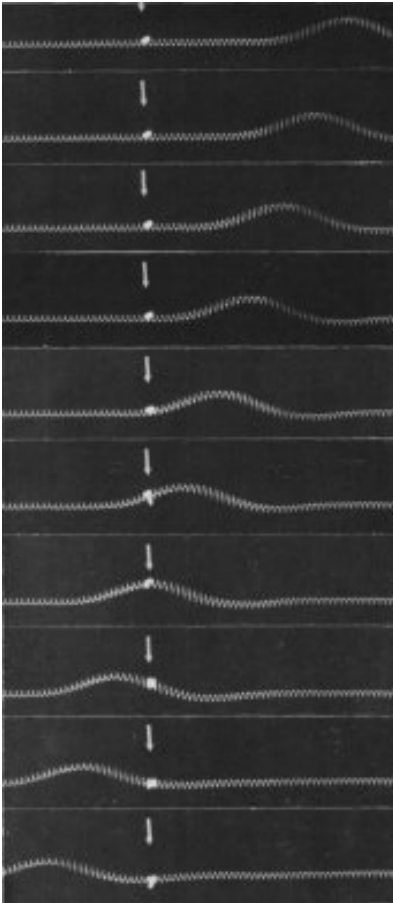
**A** In figure e, the fifth frame shows the spring just about perfectly flat. If the two pulses have essentially canceled each other out perfectly, then why does the motion pick up again? Why doesn't the spring just stay flat?



e / These pictures show the motion of wave pulses along a spring. To make a pulse, one end of the spring was shaken by hand. Movies were filmed, and a series of frame chosen to show the motion. 1. A pulse travels to the left. 2. Superposition of two colliding positive pulses. 3. Superposition of two colliding pulses, one positive and one negative.



f / As the wave pattern passes the rubber duck, the duck stays put. The water isn't moving forward with the wave.



g / As the wave pulse goes by, the ribbon tied to the spring is not carried along. The motion of the wave pattern is to the right, but the medium (spring) is moving up and down, not to the right.

## 2. The medium is not transported with the wave.

Figure f shows a series of water waves before it has reached a rubber duck (left), having just passed the duck (middle) and having progressed about a meter beyond the duck (right). The duck bobs around its initial position, but is not carried along with the wave. This shows that the water itself does not flow outward with the wave. If it did, we could empty one end of a swimming pool simply by kicking up waves! We must distinguish between the motion of the medium (water in this case) and the motion of the wave pattern through the medium. The medium vibrates; the wave progresses through space.

### self-check A

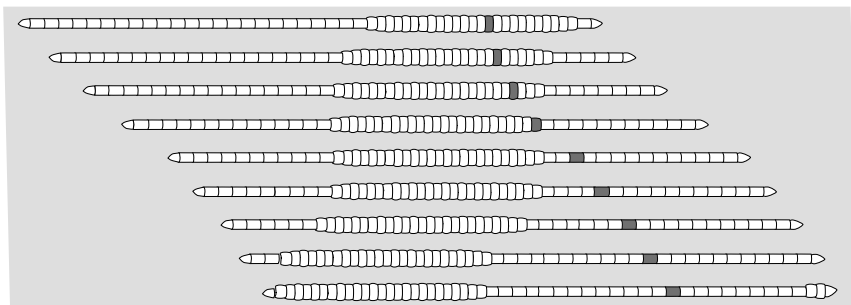
In figure g, you can detect the side-to-side motion of the spring because the spring appears blurry. At a certain instant, represented by a single photo, how would you describe the motion of the different parts of the spring? Other than the flat parts, do any parts of the spring have zero velocity?

▷ Answer, p. 180

### A worm

### example 2

The worm in the figure is moving to the right. The wave pattern, a pulse consisting of a compressed area of its body, moves to the left. In other words, the motion of the wave pattern is in the opposite direction compared to the motion of the medium.





### Surfing

### example 3

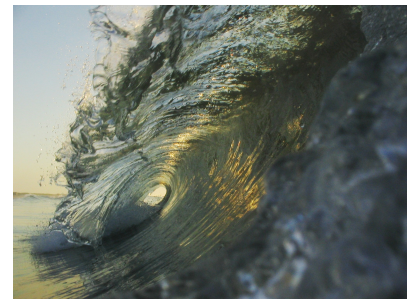
The incorrect belief that the medium moves with the wave is often reinforced by garbled secondhand knowledge of surfing. Anyone who has actually surfed knows that the front of the board pushes the water to the sides, creating a wake — the surfer can even drag his hand through the water, as in in figure h. If the water was moving along with the wave and the surfer, this wouldn't happen. The surfer is carried forward because forward is downhill, not because of any forward flow of the water. If the water was flowing forward, then a person floating in the water up to her neck would be carried along just as quickly as someone on a surfboard. In fact, it is even possible to surf down the back side of a wave, although the ride wouldn't last very long because the surfer and the wave would quickly part company.



h / Example 3. The surfer is dragging his hand in the water.

### 3. A wave's velocity depends on the medium.

A material object can move with any velocity, and can be sped up or slowed down by a force that increases or decreases its kinetic energy. Not so with waves. The magnitude of a wave's velocity depends on the properties of the medium (and perhaps also on the shape of the wave, for certain types of waves). Sound waves travel at about 340 m/s in air, 1000 m/s in helium. If you kick up water waves in a pool, you will find that kicking harder makes waves that are taller (and therefore carry more energy), not faster. The sound waves from an exploding stick of dynamite carry a lot of energy, but are no faster than any other waves. In the following section we will give an example of the physical relationship between the wave speed and the properties of the medium.



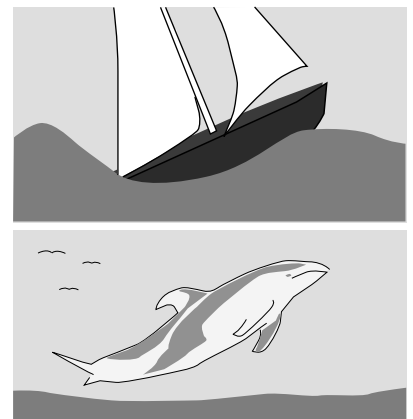
i / Example 4: a breaking wave.

### Breaking waves

### example 4

The velocity of water waves increases with depth. The crest of a wave travels faster than the trough, and this can cause the wave to break.

Once a wave is created, the only reason its speed will change is if it enters a different medium or if the properties of the medium change. It is not so surprising that a change in medium can slow down a wave, but the reverse can also happen. A sound wave traveling through a helium balloon will slow down when it emerges into the air, but if it enters another balloon it will speed back up again! Similarly, water waves travel more quickly over deeper water, so a wave will slow down as it passes over an underwater ridge, but speed up again as it emerges into deeper water.



j / Example 5. The boat has run up against a limit on its speed because it can't climb over its own wave. Dolphins get around the problem by leaping out of the water.

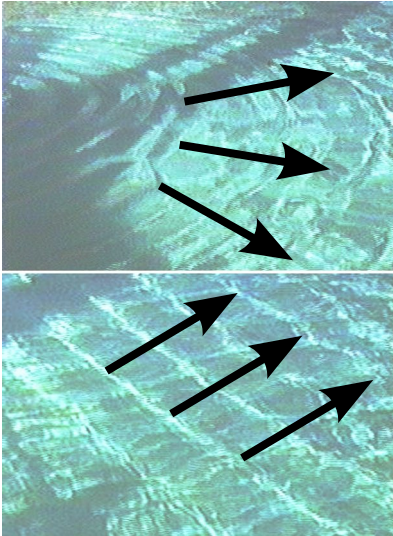
### Hull speed

### example 5

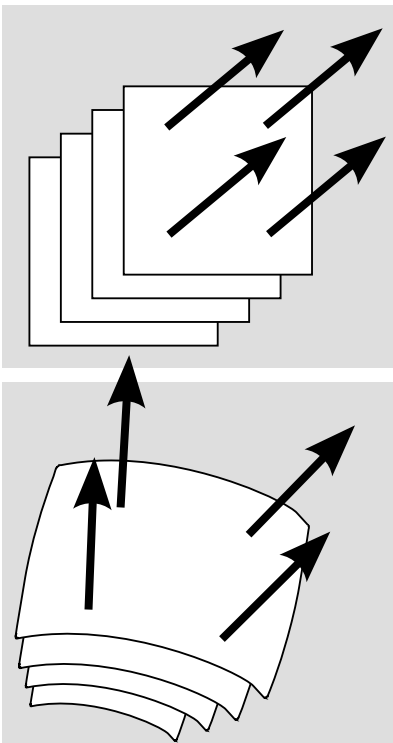
The speeds of most boats, and of some surface-swimming animals, are limited by the fact that they make a wave due to their motion through the water. The boat in figure j is going at the same speed as its own waves, and can't go any faster. No matter how

hard the boat pushes against the water, it can't make the wave move ahead faster and get out of the way. The wave's speed depends only on the medium. Adding energy to the wave doesn't speed it up, it just increases its amplitude.

A water wave, unlike many other types of wave, has a speed that depends on its shape: a broader wave moves faster. The shape of the wave made by a boat tends to mold itself to the shape of the boat's hull, so a boat with a longer hull makes a broader wave that moves faster. The maximum speed of a boat whose speed is limited by this effect is therefore closely related to the length of its hull, and the maximum speed is called the hull speed. Sailboats designed for racing are not just long and skinny to make them more streamlined — they are also long so that their hull speeds will be high.



k / Circular and linear wave patterns.



l / Plane and spherical wave patterns.

### Wave patterns

If the magnitude of a wave's velocity vector is preordained, what about its direction? Waves spread out in all directions from every point on the disturbance that created them. If the disturbance is small, we may consider it as a single point, and in the case of water waves the resulting wave pattern is the familiar circular ripple,  $k/1$ . If, on the other hand, we lay a pole on the surface of the water and wiggle it up and down, we create a linear wave pattern,  $k/2$ . For a three-dimensional wave such as a sound wave, the analogous patterns would be spherical waves and plane waves,  $l$ .

Infinitely many patterns are possible, but linear or plane waves are often the simplest to analyze, because the velocity vector is in the same direction no matter what part of the wave we look at. Since all the velocity vectors are parallel to one another, the problem is effectively one-dimensional. Throughout this chapter and the next, we will restrict ourselves mainly to wave motion in one dimension, while not hesitating to broaden our horizons when it can be done without too much complication.

### Discussion questions

- A [see above]
- B Sketch two positive wave pulses on a string that are overlapping but not right on top of each other, and draw their superposition. Do the same for a positive pulse running into a negative pulse.
- C A traveling wave pulse is moving to the right on a string. Sketch the velocity vectors of the various parts of the string. Now do the same for a pulse moving to the left.
- D In a spherical sound wave spreading out from a point, how would the energy of the wave fall off with distance?

## 8.3 Sound and Light Waves

## Sound waves

The phenomenon of sound is easily found to have all the characteristics we expect from a wave phenomenon:

- Sound waves obey superposition. Sounds do not knock other sounds out of the way when they collide, and we can hear more than one sound at once if they both reach our ear simultaneously.
- The medium does not move with the sound. Even standing in front of a titanic speaker playing earsplitting music, we do not feel the slightest breeze.
- The velocity of sound depends on the medium. Sound travels faster in helium than in air, and faster in water than in helium. Putting more energy into the wave makes it more intense, not faster. For example, you can easily detect an echo when you clap your hands a short distance from a large, flat wall, and the delay of the echo is no shorter for a louder clap.

Although not all waves have a speed that is independent of the shape of the wave, and this property therefore is irrelevant to our collection of evidence that sound is a wave phenomenon, sound does nevertheless have this property. For instance, the music in a large concert hall or stadium may take on the order of a second to reach someone seated in the nosebleed section, but we do not notice or care, because the delay is the same for every sound. Bass, drums, and vocals all head outward from the stage at 340 m/s, regardless of their differing wave shapes.

If sound has all the properties we expect from a wave, then what type of wave is it? It must be a vibration of a physical medium such as air, since the speed of sound is different in different media, such as helium or water. Further evidence is that we don't receive sound signals that have come to our planet through outer space. The roars and whooshes of Hollywood's space ships are fun, but scientifically wrong.<sup>1</sup>

We can also tell that sound waves consist of compressions and expansions, rather than sideways vibrations like the shimmying of a snake. Only compressional vibrations would be able to cause your

---

<sup>1</sup>Outer space is not a perfect vacuum, so it is possible for sound waves to travel through it. However, if we want to create a sound wave, we typically do it by creating vibrations of a physical object, such as the sounding board of a guitar, the reed of a saxophone, or a speaker cone. The lower the density of the surrounding medium, the less efficiently the energy can be converted into sound and carried away. An isolated tuning fork, left to vibrate in interstellar space, would dissipate the energy of its vibration into internal heat at a rate many orders of magnitude greater than the rate of sound emission into the nearly perfect vacuum around it.



eardrums to vibrate in and out. Even for a very loud sound, the compression is extremely weak; the increase or decrease compared to normal atmospheric pressure is no more than a part per million. Our ears are apparently very sensitive receivers!

### **Light waves**

Entirely similar observations lead us to believe that light is a wave, although the concept of light as a wave had a long and tortuous history. It is interesting to note that Isaac Newton very influentially advocated a contrary idea about light. The belief that matter was made of atoms was stylish at the time among radical thinkers (although there was no experimental evidence for their existence), and it seemed logical to Newton that light as well should be made of tiny particles, which he called corpuscles (Latin for “small objects”). Newton’s triumphs in the science of mechanics, i.e., the study of matter, brought him such great prestige that nobody bothered to question his incorrect theory of light for 150 years. One persuasive proof that light is a wave is that according to Newton’s theory, two intersecting beams of light should experience at least some disruption because of collisions between their corpuscles. Even if the corpuscles were extremely small, and collisions therefore very infrequent, at least some dimming should have been measurable. In fact, very delicate experiments have shown that there is no dimming.

The wave theory of light was entirely successful up until the 20th century, when it was discovered that not all the phenomena of light could be explained with a pure wave theory. It is now believed that both light and matter are made out of tiny chunks which have *both* wave and particle properties. For now, we will content ourselves with the wave theory of light, which is capable of explaining a great many things, from cameras to rainbows.

If light is a wave, what is waving? What is the medium that wiggles when a light wave goes by? It isn’t air. A vacuum is impenetrable to sound, but light from the stars travels happily through zillions of miles of empty space. Light bulbs have no air inside them, but that doesn’t prevent the light waves from leaving the filament. For a long time, physicists assumed that there must be a mysterious medium for light waves, and they called it the aether (not to be confused with the chemical). Supposedly the aether existed everywhere in space, and was immune to vacuum pumps. We now know that, as discussed in chapter 6, light can instead be explained as a wave pattern made up of electrical and magnetic fields.

## **8.4 Periodic Waves**

### **Period and frequency of a periodic wave**

You choose a radio station by selecting a certain frequency. We have already defined period and frequency for vibrations, but what

do they signify in the case of a wave? We can recycle our previous definition simply by stating it in terms of the vibrations that the wave causes as it passes a receiving instrument at a certain point in space. For a sound wave, this receiver could be an eardrum or a microphone. If the vibrations of the eardrum repeat themselves over and over, i.e., are periodic, then we describe the sound wave that caused them as periodic. Likewise we can define the period and frequency of a wave in terms of the period and frequency of the vibrations it causes. As another example, a periodic water wave would be one that caused a rubber duck to bob in a periodic manner as they passed by it.

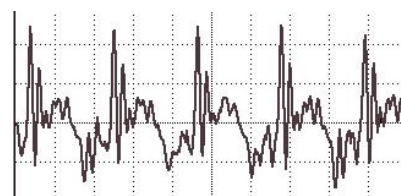
The period of a sound wave correlates with our sensory impression of musical pitch. A high frequency (short period) is a high note. The sounds that really define the musical notes of a song are only the ones that are periodic. It is not possible to sing a non-periodic sound like “sh” with a definite pitch.

The frequency of a light wave corresponds to color. Violet is the high-frequency end of the rainbow, red the low-frequency end. A color like brown that does not occur in a rainbow is not a periodic light wave. Many phenomena that we do not normally think of as light are actually just forms of light that are invisible because they fall outside the range of frequencies our eyes can detect. Beyond the red end of the visible rainbow, there are infrared and radio waves. Past the violet end, we have ultraviolet, x-rays, and gamma rays.

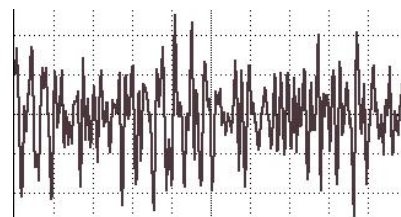
### Graphs of waves as a function of position

Some waves, like sound waves, are easy to study by placing a detector at a certain location in space and studying the motion as a function of time. The result is a graph whose horizontal axis is time. With a water wave, on the other hand, it is simpler just to look at the wave directly. This visual snapshot amounts to a graph of the height of the water wave as a function of *position*. Any wave can be represented in either way.

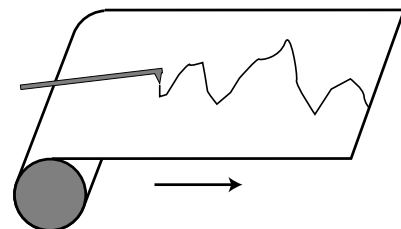
An easy way to visualize this is in terms of a strip chart recorder, an obsolescing device consisting of a pen that wiggles back and forth as a roll of paper is fed under it. It can be used to record a person’s electrocardiogram, or seismic waves too small to be felt as a noticeable earthquake but detectable by a seismometer. Taking the seismometer as an example, the chart is essentially a record of the ground’s wave motion as a function of time, but if the paper was set to feed at the same velocity as the motion of an earthquake wave, it would also be a full-scale representation of the profile of the actual wave pattern itself. Assuming, as is usually the case, that the wave velocity is a constant number regardless of the wave’s shape, knowing the wave motion as a function of time is equivalent to knowing it as a function of position.



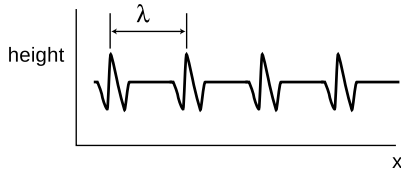
m / A graph of pressure versus time for a periodic sound wave, the vowel “ah.”



n / A similar graph for a non-periodic wave, “sh.”



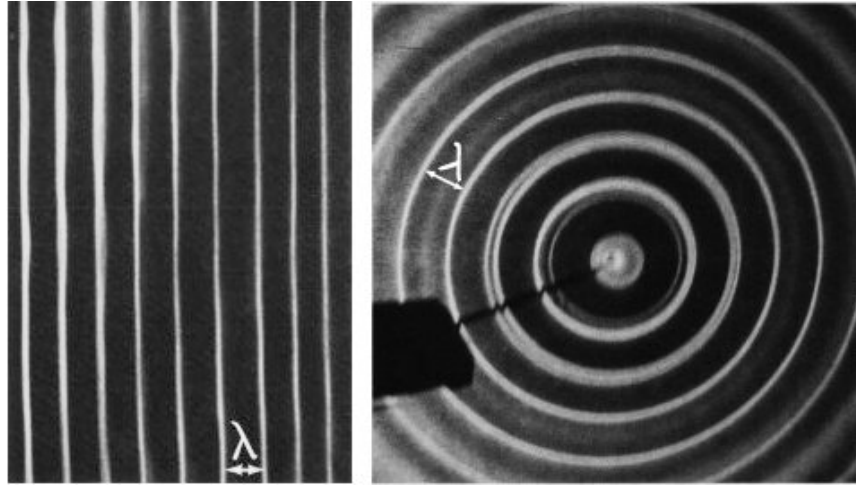
o / A strip chart recorder.



p / A water wave profile created by a series of repeating pulses.

## Wavelength

Any wave that is periodic will also display a repeating pattern when graphed as a function of position. The distance spanned by one repetition is referred to as one *wavelength*. The usual notation for wavelength is  $\lambda$ , the Greek letter lambda. Wavelength is to space as period is to time.



q / Wavelengths of linear and circular water waves.

## Wave velocity related to frequency and wavelength

Suppose that we create a repetitive disturbance by kicking the surface of a swimming pool. We are essentially making a series of wave pulses. The wavelength is simply the distance a pulse is able to travel before we make the next pulse. The distance between pulses is  $\lambda$ , and the time between pulses is the period,  $T$ , so the speed of the wave is the distance divided by the time,

$$v = \lambda/T.$$

This important and useful relationship is more commonly written in terms of the frequency,

$$v = f\lambda.$$

### Wavelength of radio waves

### example 6

▷ The speed of light is  $3.0 \times 10^8$  m/s. What is the wavelength of the radio waves emitted by KKJZ, a station whose frequency is 88.1 MHz?

▷ Solving for wavelength, we have

$$\begin{aligned}\lambda &= v/f \\ &= (3.0 \times 10^8 \text{ m/s}) / (88.1 \times 10^6 \text{ s}^{-1}) \\ &= 3.4 \text{ m}\end{aligned}$$

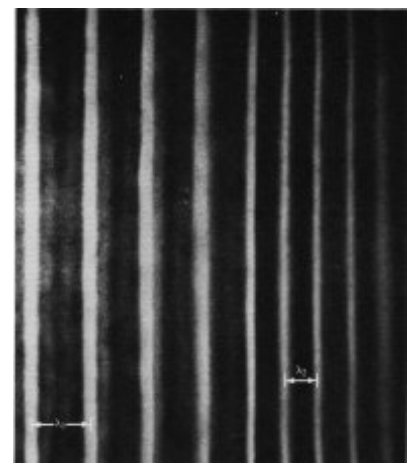
The size of a radio antenna is closely related to the wavelength of the waves it is intended to receive. The match need not be exact (since after all one antenna can receive more than one wavelength!), but the ordinary “whip” antenna such as a car’s is 1/4 of a wavelength. An antenna optimized to receive KKJZ’s signal would have a length of  $3.4 \text{ m}/4 = 0.85 \text{ m}$ .



r / Ultrasound, i.e., sound with frequencies higher than the range of human hearing, was used to make this image of a fetus. The resolution of the image is related to the wavelength, since details smaller than about one wavelength cannot be resolved. High resolution therefore requires a short wavelength, corresponding to a high frequency.

The equation  $v = f\lambda$  defines a fixed relationship between any two of the variables if the other is held fixed. The speed of radio waves in air is almost exactly the same for all wavelengths and frequencies (it is exactly the same if they are in a vacuum), so there is a fixed relationship between their frequency and wavelength. Thus we can say either “Are we on the same wavelength?” or “Are we on the same frequency?”

A different example is the behavior of a wave that travels from a region where the medium has one set of properties to an area where the medium behaves differently. The frequency is now fixed, because otherwise the two portions of the wave would otherwise get out of step, causing a kink or discontinuity at the boundary, which would be unphysical. (A more careful argument is that a kink or discontinuity would have infinite curvature, and waves tend



s / A water wave traveling into a region with a different depth changes its wavelength.

to flatten out their curvature. An infinite curvature would flatten out infinitely fast, i.e., it could never occur in the first place.) Since the frequency must stay the same, any change in the velocity that results from the new medium must cause a change in wavelength.

The velocity of water waves depends on the depth of the water, so based on  $\lambda = v/f$ , we see that water waves that move into a region of different depth must change their wavelength, as shown in figure s. This effect can be observed when ocean waves come up to the shore. If the deceleration of the wave pattern is sudden enough, the tip of the wave can curl over, resulting in a breaking wave.

## Problems

### Key

✓ A computerized answer check is available online.

∫ A problem that requires calculus.

★ A difficult problem.

**1** Many single-celled organisms propel themselves through water with long tails, which they wiggle back and forth. (The most obvious example is the sperm cell.) The frequency of the tail's vibration is typically about 10-15 Hz. To what range of periods does this range of frequencies correspond?

**2** (a) Pendulum 2 has a string twice as long as pendulum 1. If we define  $x$  as the distance traveled by the bob along a circle away from the bottom, how does the  $k$  of pendulum 2 compare with the  $k$  of pendulum 1? Give a numerical ratio. [Hint: the total force on the bob is the same if the angles away from the bottom are the same, but equal angles do not correspond to equal values of  $x$ .]

(b) Based on your answer from part (a), how does the period of pendulum 2 compare with the period of pendulum 1? Give a numerical ratio. ★

**3** The following is a graph of the height of a water wave as a function of *position*, at a certain moment in time.



Trace this graph onto another piece of paper, and then sketch below it the corresponding graphs that would be obtained if

(a) the amplitude and frequency were doubled while the velocity remained the same;

(b) the frequency and velocity were both doubled while the amplitude remained unchanged;

(c) the wavelength and amplitude were reduced by a factor of three while the velocity was doubled.

*Explain all your answers.* [Problem by Arnold Arons.]

**4** (a) The graph shows the height of a water wave pulse as a function of position. Draw a graph of height as a function of time for a specific point on the water. Assume the pulse is traveling to the right.

(b) Repeat part a, but assume the pulse is traveling to the left.

(c) Now assume the original graph was of height as a function of time, and draw a graph of height as a function of position, assuming the pulse is traveling to the right.



Problem 4.

(d) Repeat part c, but assume the pulse is traveling to the left. Explain all your answers. [Problem by Arnold Arons.]

**5** Suggest a quantitative experiment to look for any deviation from the principle of superposition for surface waves in water. Make it simple and practical.

**6** The musical note middle  $C$  has a frequency of 262 Hz. What are its period and wavelength? ✓

# Appendix 1: Photo Credits

Except as specifically noted below or in a parenthetical credit in the caption of a figure, all the illustrations in this book are by under my own copyright, and are copyleft licensed under the same license as the rest of the book.

In some cases it's clear from the date that the figure is public domain, but I don't know the name of the artist or photographer; I would be grateful to anyone who could help me to give proper credit. I have assumed that images that come from U.S. government web pages are copyright-free, since products of federal agencies fall into the public domain. When "PSSC Physics" is given as a credit, it indicates that the figure is from the second edition of the textbook entitled Physics, by the Physical Science Study Committee; these are used according to a blanket permission given in the later PSSC College Physics edition, which states on the copyright page that "The materials taken from the original and second editions and the Advanced Topics of PSSC PHYSICS included in this text will be available to all publishers for use in English after December 31, 1970, and in translations after December 31, 1975."

In a few cases, I have made use of images under the fair use doctrine. However, I am not a lawyer, and the laws on fair use are vague, so you should not assume that it's legal for you to use these images. In particular, fair use law may give you less leeway than it gives me, because I'm using the images for educational purposes, and giving the book away for free. Likewise, if the photo credit says "courtesy of ...," that means the copyright owner gave me permission to use it, but that doesn't mean you have permission to use it.

**Cover** *Wave*: Roger McLassus, GFDL 1.2.     *Hand and photomontage*: B. Crowell.

**Table of Contents** *Skateboarder*: Courtesy of J.D. Rogge, [www.sonic.net/~shawn](http://www.sonic.net/~shawn).  
*Figure skater*: Wikimedia Commons user Rosiemairianne, GFDL 2.     *Sunspot*: Royal Swedish Academy of Sciences.     The astronomers' web page at [http://www.solarphysics.kva.se/NatureNov2002/press\\_images\\_eng.html](http://www.solarphysics.kva.se/NatureNov2002/press_images_eng.html) states "All images are free for publication."  
*X-Ray*: 1896 image produced by Roentgen.     *Surfing*: Stan Shebs, GFDL licensed (Wikimedia Commons).

**7** *Star trails*: GFDL licensed, Wikipedia user Manfreed.     **8** *Emmy Noether*: Based on Noether's apparent age, the portrait must have been taken around 1900 or 1910, so it is in the public domain.     **8** *C.S. Wu*: Smithsonian Institution, believed to be public domain.     **8** *Swan Lake*: Peter Gerstbach, GFDL 1.2.     **9** *Portrait of Monsieur Lavoisier and His Wife*: Jacques-Louis David, 1788.     **10** *Astronaut*: NASA.     **15** *Hockey puck*: photo from Wikimedia Commons, user Robludwig, GFDL/CC-BY-SA.     **15** *Portrait of James Joule*: contemporary.     **16** *Aristotle*: Francesco Hayez, 1811.     **16** *Jets over New York*: U.S. Air Force, Tech. Sgt. Sean Mateo White, public domain work of the U.S. Government.     **17** *Galileo's trial*: Cristiano Banti (1857).     **18** *Rocket sled*: U.S. Air Force, public domain work of the U.S. Government.     **17** *Foucault and pendulum*: contemporary, ca. 1851.     **20** *Skateboarder*: Courtesy of J.D. Rogge, [www.sonic.net/~shawn](http://www.sonic.net/~shawn).     **24** *Welding*: William M. Plate, Jr., public-domain product of the U.S. Airforce, Wikimedia Commons.     **24** *Infrared photographs*: Courtesy of M. Vollmer and K.P. Möllmann, Univ. Appl. Sciences, Brandenburg, Germany, [www.fh-](http://www.fh-)



brandenburg.de/~piweb/projekte/thermo\_galerie\_eng.html. **26** *Newton*: Godfrey Kneller, 1702. **30** *Eclipse*: 1919, public domain. **30** *Newspaper headline*: 1919, public domain. **34** *Colliding balls*: PSSC Physics. **53** *Brahe*: public domain. **57** *Basebal pitch*: Wikipedia user Rick Dikeman, GFDL 1.2. **60** *Bull*: Photo by Wikimedia Commons user Bart Hiddink, CC-BY. **61** *Tornado*: NOAA Photo Library, NOAA Central Library; OAR/ERL/National Severe Storms Laboratory (NSSL); public-domain product of the U.S. government. **63** *Longjump*: Thomas Eakins, public domain. **65** *Pendulum*: PSSC Physics. **64** *Tetherball*: Line art by the author, based on a photo by The Chewonki Foundation (Flickr), CC-BY-SA 2.0 licensed. **83** *Colliding nuclei*: courtesy of RHIC. **84** *Machine gunner's body*: Redrawn from a public-domain photo by Cpl. Sheila Brooks. **84** *Machine gunner's head*: Redrawn from a sketch by Wenceslas Hollar, 17th century. **93** *Lightning*: C. Clark/NOAA photo library, public domain. **98** *Ampère*: Millikan and Gale, 1920. **102** *Volta*: Millikan and Gale, 1920. **104** *Ohm*: Millikan and Gale, 1920. **111** *Sunspot*: Royal Swedish Academy of Sciences. The astronomers' web page at [http://www.solarphysics.kva.se/NatureNov2002/press\\_images\\_eng.html](http://www.solarphysics.kva.se/NatureNov2002/press_images_eng.html) states "All images are free for publication.". **122** *Faraday banknote*: fair use. **125** *Maxwell*: 19th century photograph. **131** *Rays of sunlight*: Wikipedia user PiccoloNamek, GFDL 1.2. **134** *Jupiter and Io*: NASA/JPL/University of Arizona. **145** *Ray-traced image*: Gilles Tran, Wikimedia Commons, public domain. **152** *Flower*: Based on a photo by Wikimedia Commons user Fir0002, GFDL 1.2. **151** *Moon*: Wikimedia commons image. **161** *Painting of waves*: Katsushika Hokusai (1760-1849), public domain. **163** *John Hancock Tower*: Wikipedia user Sfoskett, GFDL 1.2. **165** *Superposition of pulses*: Photo from PSSC Physics. **166** *Marker on spring as pulse passes by*: PSSC Physics. **167** *Surfing (hand drag)*: Stan Shebs, GFDL licensed (Wikimedia Commons). **167** *Breaking wave*: Ole Kils, olekils at web.de, GFDL licensed (Wikipedia). **172** *Wavelengths of circular and linear waves*: PSSC Physics. **173** *Fetus*: Image of the author's daughter. **173** *Changing wavelength*: PSSC Physics.

# Appendix 2: Hints and Solutions

## Answers to Self-Checks

### Answers to Self-Checks for Chapter 1

**Page 9, self-check A:** A conservation law in physics says that the total amount always remains the same. You can't get rid of it even if you want to.

**Page 13, self-check B:** Exponents have to do with multiplication, not addition. The first line should be 100 times longer than the second, not just twice as long.

**Page 26, self-check C:** Doubling  $d$  makes  $d^2$  four times bigger, so the gravitational field experienced by Mars is four times weaker.

### Answers to Self-Checks for Chapter 2

**Page 40, self-check A:** No, it doesn't violate symmetry. Space-translation symmetry only says that space itself has the same properties everywhere. It doesn't say that all regions of space have the same stuff in them. The experiment on the earth comes out a certain way because that region of space has a planet in it. The experiment on the moon comes out different because that region of space has the moon in it. of the apparatus, which you forgot to take with you.

**Page 42, self-check B:** The camera is moving at half the speed at which the light ball is initially moving. After the collision, it keeps on moving at the same speed — your five x's all line on a straight line. Since the camera moves in a straight line with constant speed, it is showing an inertial frame of reference.

**Page 43, self-check C:** The table looks like this:

	velocity (meters per second)		
	before the collision	after the collision	change
●	-1	0	+1
○	0	-1	-1

Observers in all three frames agree on the changes in velocity, even though they disagree on the velocities themselves.

**Page 50, self-check D:** The motion would be the same. The force on the ball would be 20 newtons, so with each second it would gain 20 units of momentum. But 20 units of momentum for a 2-kilogram ball is still just 10 m/s of velocity.

### Answers to Self-Checks for Chapter 3

**Page 65, self-check A:** The definition of torque is important, and so is the equation  $F = \pm Fr$ . The two equations in between are just steps in a derivation of  $F = \pm Fr$ .

### Answers to Self-Checks for Chapter 4

**Page 86, self-check A:** The total momentum is zero before the collision. After the collision, the two momenta have reversed their directions, but they still cancel. Neither object has changed its kinetic energy, so the total energy before and after the collision is also the same.

## Answers to Self-Checks for Chapter 5

**Page 97, self-check A:** Either type can be involved in either an attraction or a repulsion. A positive charge could be involved in either an attraction (with a negative charge) or a repulsion (with another positive), and a negative could participate in either an attraction (with a positive) or a repulsion (with a negative).

**Page 97, self-check B:** It wouldn't make any difference. The roles of the positive and negative charges in the paper would be reversed, but there would still be a net attraction.

## Answers to Self-Checks for Chapter 6

**Page 124, self-check A:** An induced electric field can only be created by a *changing* magnetic field. Nothing is changing if your car is just sitting there. A point on the coil won't experience a changing magnetic field unless the coil is already spinning, i.e., the engine has already turned over.

## Answers to Self-Checks for Chapter 7

**Page 138, self-check A:** Only 1 is correct. If you draw the normal that bisects the solid ray, it also bisects the dashed ray.

**Page 141, self-check B:** He's five times farther away than she is, so the light he sees is 1/25 the brightness.

**Page 148, self-check C:** You should have found from your ray diagram that an image is still formed, and it has simply moved down the same distance as the real face. However, this new image would only be visible from high up, and the person can no longer see his own image.

**Page 150, self-check D:** Increasing the distance from the face to the mirror has decreased the distance from the image to the mirror. This is the opposite of what happened with the virtual image.

## Answers to Self-Checks for Chapter 8

**Page 166, self-check A:** The leading edge is moving up, the trailing edge is moving down, and the top of the hump is motionless for one instant.

## Solutions to Selected Homework Problems

### Solutions for chapter 1

**Page 33, problem 1:**

$$134 \text{ mg} \times \frac{10^{-3} \text{ g}}{1 \text{ mg}} \times \frac{10^{-3} \text{ kg}}{1 \text{ g}} = 1.34 \times 10^{-4} \text{ kg}$$

### Solutions for chapter 3

**Page 68, problem 4:** The pliers are not moving, so their angular momentum remains constant at zero, and the total torque on them must be zero. Not only that, but each half of the pliers must have zero total torque on it. This tells us that the magnitude of the torque at one end must be the same as that at the other end. The distance from the axis to the nut is about 2.5 cm, and the distance from the axis to the centers of the palm and fingers are about 8 cm. The angles are close enough to 90° that we can pretend they're 90 degrees, considering the rough nature of the other assumptions and measurements. The result is  $(300 \text{ N})(2.5 \text{ cm}) = (F)(8 \text{ cm})$ ,

or  $F = 90 \text{ N}$ .

# Index

- absorption, 135
- alchemy, 94
- ammeter, 101
- ampere (unit), 98
- amplitude
  - defined, 163
  - peak-to-peak, 163
- angular magnification, 151
- angular momentum, 62
  - introduction to, 61
- Aristotle, 16, 25
- aurora, 121
  
- black hole, 30
- Brahe, Tycho, 53
  
- Catholic Church, 25
- causality, 72
- centi- (metric prefix), 11
- charge, 95
  - conservation of, 97
- Church
  - Catholic, 25
- circuit, 100
  - complete, 100
  - open, 101
  - parallel, 108
  - series, 108
  - short, 107
- circular motion, 52
- complete circuit, 100
- conductor
  - defined, 105
- conservation
  - of charge, 97
  - of mass, 9
  - of momentum, 40
- conservation law, 7
- converging, 148
- conversions of units, 13
- Copernicus, 18
- correspondence principle
  - defined, 31
  - for mass-energy equivalence, 31
  - for relativistic momentum, 87
  - for time dilation, 72
- coulomb (unit), 96
- Coulomb's law, 97
- current
  - defined, 98
  
- diffuse reflection, 136
  
- Einstein, Albert, 29
- electric current
  - defined, 98
- electric field, 115
- electric forces, 95
- electrical energy, 23
- electromagnetism, 118
  - spectrum, 126
  - waves, 125
- Empedocles of Acragas, 132
- energy
  - electrical, 23
  - equivalence to mass, 29, 87
  - gravitational, 20
  - heat, 22
  - kinetic, 20
  - Noether's theorem, 27
  - nuclear, 23
  - sound, 22
- ether, 81
  
- Faraday, Michael, 122
- Fermat's principle, 146
- Feynman, Richard, 128
- field, 111
  - electric, 115
  - gravitational, 20, 113
    - Newton's law of gravity, 25
  - magnetic, 119
- force
  - definition, 47
  - of gravity, 49
  - pairs, 47
  - unit, 47
- Foucault, Léon, 17
- frame of reference, 17

- inertial, 17
- noninertial, 17, 53
- Franklin, Benjamin
  - definition of signs of charge, 96
- free fall, 25
- French Revolution, 11
- frequency, 161
  - defined, 162
- Galileo, 133
  - projectile motion, 50
- Galileo Galilei, 16
- garage paradox, 78
- Gates, Bill, 28
- Gell-Mann, Murray, 37
- generator, 123
- gravitational constant,  $G$ , 27
- gravitational energy, 20
- gravitational field, 113
  - Newton's law of gravity, 25
- gravity
  - force of, 49
  - Newton's law of, 25
- heat, 22
- Hertz, Heinrich, 126
- Hipparchus, 143
- Hooke, 94
- Hugo, Victor, 93
- images
  - formed by curved mirrors, 148
  - formed by plane mirrors, 147
  - of images, 151
  - real, 149
  - virtual, 147
- induction, 122
- inertia
  - principle of, 39
- inertia, principle of, 16
- insulator
  - defined, 105
- inverse-square law, 26, 140
- Io, 134
- joule (unit), 20
- Jupiter, 134
- Kepler's laws, 54
  - elliptical orbit law, 54
  - equal-area law, 54
  - law of periods, 54
- Kepler, Johannes, 8, 53
- Keynes, John Maynard, 94
- kilo- (metric prefix), 11
- kinetic energy, 20
  - compared to momentum, 45
- Lavoisier
  - Antoine-Laurent, 9
- least time, principle of, 146
- lever, 20
- light, 168
  - absorption of, 135
  - as an electromagnetic wave, 126
  - particle model of, 136
  - ray model of, 136
  - speed of, 133
  - wave model of, 136
- Lorentz transformation, 75
- Lorentz, Hendrik, 75
- magnetic field, 119
- magnetism
  - and relativity, 116
  - caused by moving charges, 116
  - magnetic field, 119
  - related to electricity, 118
- magnification
  - angular, 151
  - by a converging mirror, 149
- Mars
  - life on, 121
- mass
  - conservation of, 9
  - equivalence to energy, 29, 87
- mass-energy, 29
- Mathematical Principles of Natural Philosophy, 27
- Maxwell, James Clerk, 125, 126
- mega- (metric prefix), 11
- metric system
  - prefixes, 11
- Michelson-Morley experiment, 81
- micro- (metric prefix), 11
- milli- (metric prefix), 11
- model, 7
- momentum

- compared to kinetic energy, 45
  - conservation of, 40
  - relativistic, 85
- moon
  - distance to, 143
  - gravitational field experienced by, 27
  - orbit, 51
- motion
  - periodic, 162
- nano- (metric prefix), 11
- Neanderthals, 67
- neutral (electrically), 97
- newton (unit), 47
- Newton, Isaac, 93, 151
  - apple myth, 27
  - law of gravity, 25
  - third law, 47
- Noether's theorem, 8
  - for angular momentum, 67
  - for energy, 27
  - for momentum, 43
- Noether, Emmy, 8
- nuclear energy, 23
  
- Oersted, Hans Christian, 115
- ohm (unit), 105
- Ohm's law, 105
- ohmic
  - defined, 105
- open circuit, 101
- Optics, 27
- Orion Nebula, 23
  
- parabola, 50
- parallax, 142
- parallel circuit
  - defined, 108
- particle model of light, 136
- period
  - defined, 162
- Principia Mathematica, 27
- principle of superposition, 164
- projectile motion, 50
- pulse
  - defined, 164
- Pythagoras, 132
  
- ray diagrams, 138
- ray model of light, 136
- reflection
  - diffuse, 136
  - specular, 138
- relativity
  - and magnetism, 116
- resistance, 104
- retina, 150
- reversibility, 144
- RHIC accelerator, 83
- Roemer, 133
- rotational symmetry, 15
  
- schematics, 107
- scientific notation, 12
- sea-of-arrows representation, 114
- series circuit
  - defined, 108
- short circuit
  - defined, 107
- sinks in fields, 114
- Sokal, Alan, 31
- sound, 168
  - energy, 22
  - speed of, 167
- sources of fields, 114
- spring constant, 163
- sunspots, 121
- symmetry, 7
  - rotational, 15
  - time, 27
  - translation, 38
- Système International, 11
  
- time reversal, 144
- time symmetry, 27
- torque
  - defined, 64
- transformer, 124
  
- units, conversion of, 13
- velocity
  - addition of, 19
- vision, 132
- volt (unit)
  - defined, 102
- voltage, 101
  
- wave

electromagnetic, 126  
wave model of light, 136  
wavelength, 126  
work, 57  
Wu, Chien-Shiung, 8