How things work! (PHYS 121)

Daniel M. Sussman

November 20, 2020

This is a set of lecture notes prepared for PHYS 121: How things work (Emory University, Fall 2020). There are undoubtedly typos and errors in this document: please feel free to email any corrections to:

daniel.m.sussman@emory.edu

Contents

Preface 6
Sources 6

1 Quantitative thinking, estimation, and units! 7
1.1 A brief introduction... 7
1.1.1 Symmetry 7
1.1.2 Scale 9
1.2 Units – what's the fuss? 10
1.2.1 Physical quantities have dimensions 10
1.3 Estimation and "Fermi problems" 12
1.3.1 Scientific notation and orders of magnitude 12
1.3.2 Order of magnitude estimates 12
1.3.3 Fermi estimation 14
1.3.4 Case study: trees and paper 15
1.3.5 Practice examples! 17
1.4 Dimensional analysis 17
1.4.1 Basics of dimensional analysis 18
1.4.2 Case study 1: The Trinity detonation 18
1.4.3 Case study 2: Can you estimate how fast your (car / train / plane) is going be watching raindrops rolling on the side windows? 21

2 The laws of motion, Part 1! 25
2.1 Inertia, friction, and why our intuition is often wrong 26
2.2 Skating 27
2.2.1 Acceleration and forces 28
2.2.2 Frames of reference 29
2.3 Throwing stuff 30
2.3.1 Gravity 30
2.3.2 Free fall 31
2.3.3 Motion of a constantly accelerating object 31
2.3.4 Throwing stuff 32
2.4 Ramps 33
2.4.1 "Support" or "Normal" forces 34
2.4.2 Energy and work 36
2.4.3 Doing work 37

2.4.4 Using a ramp! . . . 38

3 The laws of motion, Part 2! **39**
3.1 Seesaws, levers, and mobiles . . . 39
3.1.1 Balancing a seesaw . . . 39
3.1.2 Newton's Laws of rotational motion . . . 41
3.1.3 Torques . . . 43
3.2 Wheels . . . 45
3.2.1 Friction . . . 45
3.2.2 Wheels . . . 46
3.2.3 Rotational motion, work, and energy . . . 47
3.3 Bumper cars . . . 48
3.3.1 Linear momentum . . . 48
3.3.2 Angular momentum . . . 49
3.3.3 Impulses . . . 50

4 Mechanical objects, Part 1! **52**
4.1 Springs and scales . . . 52
4.1.1 Measuring weight . . . 52
4.1.2 Stretching and compressing a spring . . . 54
4.2 Bouncing balls . . . 57
4.3 Roller coasters . . . 59
4.3.1 The feeling of acceleration . . . 59
4.3.2 Riding a carousel in a playground . . . 60
4.3.3 Riding a rollercoaster . . . 62

5 Mechanical objects, Part 2! **64**
5.1 Bicycles . . . 64
5.1.1 Static stability – when do object fall over? . . . 64
5.1.2 Dynamic stability – what keeps a bicycle upright while it's moving? . . . 66
5.1.3 Choosing the right gear . . . 68
5.2 Satellites and planets . . . 69
5.2.1 Orbiting around . . . 69
5.2.2 Kepler's Laws . . . 71
5.3 Rocket ships . . . 73
5.3.1 Blasting off... . . . 73
5.3.2 ... into space! . . . 73

6 Fluids! **75**
6.1 Gases . . . 75
6.1.1 Air pressure and the ideal gas law . . . 76
6.2 Floating and buoyancy . . . 78
6.3 Liquids and water pressure . . . 80
6.3.1 Pressure inside a liquid . . . 80
6.3.2 Water seeks its own level . . . 82

7 Fluids and motion! 84
7.1 Flowing water . . . 84
7.1.1 Viscosity . . . 84
7.2 Flow through a hose . . . 85
7.2.1 Flow through a straight hose . . . 85
7.2.2 Dynamic pressure variations: going around the bend . . . 87
7.3 Turbulent vs smooth flow . . . 87
7.4 Aerodynamics . . . 89
7.4.1 Slow-moving stuff and laminar airflow . . . 89
7.4.2 Faster-moving stuff and turbulent airflow . . . 90
7.4.3 Curveball! . . . 92

8 Heat and and insulation! 95
8.1 Stoves . . . 95
8.1.1 Thermal energy in solids (aka, "stuff on fire") . . . 96
8.1.2 Warming an object up . . . 97
8.2 How heat flows . . . 98
8.2.1 Conduction of heat . . . 98
8.2.2 Convection of heat . . . 98
8.2.3 Radiation . . . 99
8.3 Insulation . . . 100
8.3.1 Limiting thermal conduction . . . 101
8.3.2 Limiting convective heat loss . . . 101
8.3.3 Limiting radiative heat loss . . . 102

9 Thermodynamics! 104
9.1 The laws of thermodynamics! . . . 104
9.1.1 The "zeroth" law of thermodynamics . . . 105
9.1.2 The first law of thermodynamics . . . 106
9.1.3 The second law of thermodynamics . . . 106
9.1.4 The third law of thermodynamics . . . 108
9.2 Engines and automobiles! . . . 108
9.2.1 Work, heat, and temperature in a piston . . . 108
9.2.2 Abstract engines . . . 110
9.2.3 An ideal gas engine . . . 110
9.2.4 More realistic engines . . . 112
9.3 Air conditioners! . . . 113
9.3.1 Phase changes from solids, liquids, and gases . . . 113
9.3.2 Vapor-compression refrigerator . . . 115

10 Mechanical Waves! 117
10.1 Clocks . . . 117
10.1.1 Harmonic oscillators again . . . 117
10.1.2 Resonance and natural frequencies . . . 119
10.1.3 Clocks . . . 119

10.2 Musical instruments . . . 121
10.2.1 Vibrations, oscillations, harmonics . . . 121
10.2.2 Other musical instruments . . . 122
10.2.3 Sound waves . . . 124
10.3 Water waves . . . 126
10.3.1 Capillary waves . . . 126
10.3.2 Gravity waves . . . 127
10.3.3 The tides . . . 128

11 Light! 130
11.1 What is light? . . . 130
11.1.1 Historical interlude . . . 130
11.1.2 Electromagnetic radiation . . . 131
11.2 Mirrors! . . . 133
11.3 Rainbows! . . . 135
11.3.1 Refraction of light . . . 135
11.3.2 Chromatic aberration . . . 137
11.3.3 Sun and rain . . . 138
11.4 The sky is blue! . . . 140
11.4.1 Tyndall and Rayleigh . . . 140
11.4.2 Einstein . . . 141
11.4.3 What the sky looks like . . . 141

12 Random walks! 143
12.1 Random walks . . . 144
12.2 Fermi estimation revisited . . . 146
12.3 Masks and COVID-19 . . . 147
12.3.1 Particles, big and small . . . 147
12.3.2 A mask is not a sieve . . . 149
12.3.3 Random walks and impact capture . . . 150

Preface

So, what is this course about? I like to think that *physics*, at its heart, is the idea that the world around us is understandable, and so in PHYS 121 we'll be spending time trying to understand a little bit about how things work. We'll discuss the science behind everyday things – motion, music, materials, and much more – and see how amazing complex behavior can follow from just a few simple, general ideas. The material is structure around investigating everyday objects and phenomena in the world – bumper cars, basketballs, balloons, bicycles... – and we'll spend time thinking about the physical principles the help us understand how things work.

But that's not the *goal* of this course. I want you to come away from this class knowing what it means to think like a physicist. No matter what your background or your major is, a central value of a liberal arts education is being able to think from many different perspectives. An artist (for example) and a physicist might approach the same question in very different ways, and even if ultimately a physicist's approach doesn't resonate with you, I hope you'll walk away with an enriched perspective for understanding another way of seeing the world.

So, in a nutshell, what *does* it mean to think like a physicist? As we'll see over the coming months, it means thinking that Nature is organized by *symmetry* and *scale*; it means thinking about what aspects of a system are *conserved*; and, importantly, it means combining *qualitative* with *quantitative* arguments to try to estimate how something will behave. I'm excited to convey how beautiful I think this way of looking at the world is, and I hope you're excited, too!

Sources used

These notes are certainly not original, drawing on a variety of existing resources. Most of the structure of the course closely follows selections of Louis Bloomfield's book (item 1, below), with some additional context, examples, and material from other sources, including:

1. How things work: The physics of everyday life (Louis Bloomfield)
2. Thinking Physics (Lewis Carroll Epstein) – Keep the physics, ignore the politics
3. Tom Bing's lecture notes.
4. Biological Physics: Energy, Information, Life (Philip Nelson)
5. Tom Murphy's website

Chapter 1

Quantitative thinking, estimation, and units!

1.1 A brief introduction...

In the preface I made the claim that *physics* is really the idea that the world around us is understandable, and I meant that in a *quantitative* sense: it is the optimistic view that not only can we slowly piece together the underlying reasons that we observe the phenomena in the world around us that we do, but also that mathematical reasoning and quantitative thinking is the right "language" in which to express that understanding. This course will, of course, eschew the heavy mathematical formalism that accompanies modern research in physics, but we will not shy away from talking about the mathematical concepts that help us organize a physical worldview.

1.1.1 Symmetry

One such mathematical concept is that of *symmetry*. We all have an intuitive sense of what this word means in everyday use – a harmonious balance in the composition of things ranging from architecture to music to social interactions – but when we use it in physics we have a precise definition in mind: something is *symmetric* if we can apply some *transformation* to it any have it remain the same. One such type of transformation – rotations – is illustrated in Fig. 1.1, by investigating a few different shapes. Some symmetries are discrete, some are continuous, and shapes can be more or less symmetrical compared to each other.

This kind of symmetry is close to what we mean when we use the word colloquially, but there are other important types of symmetry, too! Suppose I want to test gravity by throwing a ball up in the air and then watching it come back down. I could do the same experiment in different spot in the room – say, six feet to the left – and the results should be exactly the same, right? This is *also* a symmetry – I've transformed *where in space I am*, and I see *how everything moves is unchanged*. This example is called *translational symmetry*. Deep down, why are solids different from liquids, and why are they different "phases of matter"? It's because the atoms and molecules composing them *break* symmetries – like rotational and translational symmetries – in different ways.

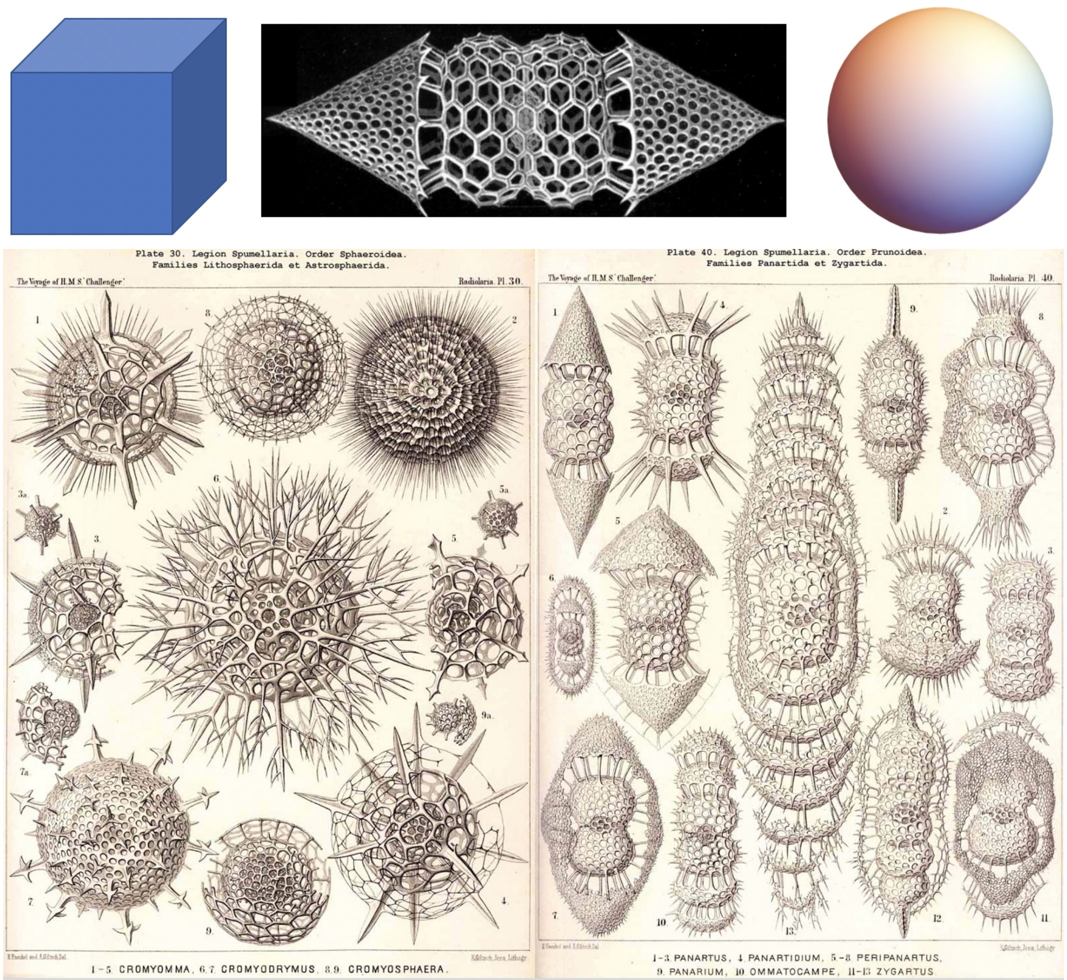

Figure 1.1: **Objects with different rotational symmetry** (Top Left) A cube! You can rotate it in certain ways and have it look the same, but most rotations give you a different view of it. (Top Center) An Illustration of the mineral skeleton of a tiny ocean protozoa, from *Kunstformen der Natur* (Ernst Haeckel, 1899)! You can rotate the skeleton about its long axis and have it look the same, or you could do a few 180-degree flips, but other rotations give you a different view. (Top Right) A sphere! You can rotate it about any axis by any amount, and it always looks like a sphere! (Bottom) More radiolarian skeletons, just because nature is amazing. From the Report of the Scientific Results of the Voyage Of The H.M.S. Challenger, During the Years 1873-76; Zoology, Volume XVIII, part 40

Figure 1.2: **Emmy Noether,** faced sexism and anti-semitism, and reformed our understanding of mathematics and physics.

One of the heroes of physics, Emmy Noether (Fig. 1.2), taught us that there is an incredibly deep connection between *symmetries* and *conservation laws*. Perhaps you've heard of things like "conservation of energy" or "conservation of momentum" – and if not, we will certainly be seeing these ideas in the coming chapters! – and wondered "but *why* are these things conserved?" Symmetry!

1.1.2 Scale

Scale – how big something is – also helps organize our understanding of the world around us. We have a sense that small things affect big things. To state the obvious: electrons, protons, and neutrons conspire to build up atoms and molecules; atoms and molecules make up different materials; different materials make up objects (from architecture to organisms), and so on. Moving from one level to the next on this hierarchy of scale can be difficult – how does a chemist go from an understanding electronic structure to synthesizing a new material? How does a sculptor go from an understanding how different metals can be shaped into a piece of art? – but the hierarchy is there.

Crucially, the hierarchy tends to go only in one direction – big things don't affect small things[1]. At least equally important: the hierarchy tends to only go one step at a time. Small things affect big things, but they don't usually affect *really* big things! This is why architects don't need to know about quantum mechanics, and, in essence, why the scientific project is possible.

[1]You will have noticed, I'm sure, that Emory lacks a Department of Astrology

1.2 Units – what's the fuss?

Throughout this class we'll make use of ideas involving *dimensions* and *units*. This sounds boring. Often, if you've seen these ideas before, it's in the context of making sure you've written your answer in the units your teacher wanted. Which probably had a feel similar to brushing your teeth and eating your broccoli. That type of presentation is a real shame because really, deep down, *dimensional analysis* is a way to organize or classify different numbers, and the abstractions they represent. This not only helps you check your work: it can help you remember what's going on in a formula, it can help you quickly estimate how big or small something should be, and it can even help you guess the physical laws underpinning the world! This section will give you a little taste of all of these uses for dimensional analysis.

1.2.1 Physical quantities have dimensions

The *dimensions* of a quantity tells us what sort of thing that quantity is. When someone asks how tall you are, you don't just say "5," because *height* is not just a pure number, it's a measure of distance, or length (in this case, the distance from the sole of your foot to the crown of your head). Each dimension can be measured by any number of different *units*, and the choice of which units we use is pretty arbitrary[2]. In this set of notes we'll use a few different notations to represent units and dimensions. We'll use funny "blackboard" letters to stand for an abstract dimension; e.g., $\mathbb{L}$ stands for the dimension of length. We'll use a slightly different font for units, so we'll abbreviate the meter symbol as $\texttt{m}$. And if we have some variable, say, x, and we want to know not the value but the *dimensions* of that variable, we'll use brackets; e.g., if x is a distance we could write $[x] = \mathbb{L}$.

Do I have more examples of dimensions and units? Glad you asked! For instance:

1. Length, as we said above, has dimensions of $\mathbb{L}$; In Imperial[3] units it's measured in feet, $\texttt{ft}$, and in SI[4] units it's measured in meters, $\texttt{m}$. Easy peasy.
2. Mass has dimensions of $\mathbb{M}$, and in SI units is measured in kilograms, $\texttt{kg}$.
3. Time has dimensions of $\mathbb{T}$, and in SI units is measured in seconds, $\texttt{s}$.
4. Speed – our first composite dimensional abstraction! – has dimensions of $\mathbb{L}\mathbb{T}^{-1}$ (equivalently, $\mathbb{L}/\mathbb{T}$), and in SI units is measured in "meters per second", $\texttt{m s}^{-1}$.
5. Acceleration has dimensions of $\mathbb{L}\mathbb{T}^{-2}$, and in SI units is measured in "meters per second squared", $\texttt{m s}^{-2}$.
6. Force has dimensions of $\mathbb{M}\mathbb{L}\mathbb{T}^{-2}$, and in SI units is measured in "kilogram meters per second squared", $\texttt{kg m s}^{-2}$. This combination comes up so much we abbreviate the whole things as a "Newton", $\texttt{N}$, where $1\ \texttt{N} = 1\ \texttt{kg m s}^{-2}$.

[2] And, really, just governed by the fact that if we all *agree on the same set of units*, it makes it a lot easier to talk to each other.

[3] Just like imperialism – not particularly favored these days

[4] Systèm International d'Unités

7. Energy has dimensions of $\mathbb{M}\mathbb{L}^2\mathbb{T}^{-2}$, and in SI units is measured in "kilogram meters squared per second squared", $\texttt{kg m}^2\ \texttt{s}^{-2}$. You could call this whole thing "a Newton-meter, $\texttt{N m}$," which is (again) such an important combination that we abbreviate the whole thing as a "Joule," $\texttt{J}$, where $1\texttt{J} = 1\texttt{kg m}^2\texttt{s}^{-2}$.

What do you if someone gives you a physical quantity in one set of units, but you want to express it in a different set of units? Why, you just multiply by one! What does this look like? For instance, you might know that

$$1\ \texttt{m} = 3.28084\ \texttt{ft},$$

which you can re-write as

$$\frac{3.28084\ \texttt{ft}}{1\ \texttt{m}} = 1.$$

So, when I tell you that I'm 2.5 meters tall, but you're more comfortable with the Imperial system commonly used in the US, you can quickly work out that I'm saying I'm

$$2.5\texttt{m} = 2.5\ \texttt{m} \times 1 = 2.5\ \texttt{m} \times \frac{3.28084\ \texttt{ft}}{1\ \texttt{m}} = 8.2021\ \texttt{ft},$$

which, at least the last time I checked, is not true.

Similarly, if I tell you that I have[5] "Fifty stone of stones in a cubic cubit," you can work out[6] that I have rocks at a density of

$$\rho = \frac{50\ \texttt{stone}}{(1\ \texttt{cubit})^3} = \frac{50\ \texttt{stone}}{(1\ \texttt{cubit})^3} \times \frac{6.35\ \texttt{kg}}{1\ \texttt{stone}} \times \left(\frac{1\ \texttt{cubit}}{0.457\ \texttt{m}}\right)^3 \approx 3322.2\ \frac{\texttt{kg}}{\texttt{m}^3}.$$

(From which you could guess that the rocks are probably gabbro (Fig. 1.3) or some other dense, igneous rock, but you couldn't be quite sure.)

[5] A mass density, $[\rho] = \mathbb{M}\mathbb{L}^{-3}$

[6] after wondering why on earth I would choose to quote densities in such outdated terms

Figure 1.3: **Gabbro.** (Left) A small piece of the dense, igneous rock, with typical densities between 2700 - 3500 $\texttt{kg}/\texttt{m}^3$. (Right) Zuma Rock, a substantially larger piece of gabbro.

1.3 Estimation and "Fermi problems"

1.3.1 Scientific notation and orders of magnitude

As a quick review (or introduction, if you haven't seen it before), scientific notation is a way of compactly writing wextremely large or small numbers without wasting lots of ink. In everyday language we have special words for large amounts, so for instance: I could write that in 2018 the US federal government spent[7] \$4,109,042,000,000, or I could just say that is was about "4.1 trillion dollars." Scientific notation is just a way of systematizing this idea in the general case where we don't have pre-existing words for various amounts: we simply write down a number between 1 and 10 and multiply it by an appropriate power of 10. Thus,

$$4{,}100{,}000{,}000{,}000 \to 4.1 \times 10^{12}.$$

Similarly, if I want to tell you the mass of a proton I could say it's

$$m_{proton} = 0.0000000000000000000000000016726219 \ \mathtt{kg},$$

but it is much easier to understand, and compare to other numbers, if I write it like

$$m_{proton} = 1.6726219 \times 10^{-27} \ \mathtt{kg}.$$

This comes in handy, because we often want to compare, or simultaneously reason about, quantities that have very different *orders of magnitude*, a term which essentially refers to what power of ten is associated with it.

Indeed, often as we go through our lives and want to quantitatively think about the world around us, know the order of magnitude of a number is a lot more important than knowing the precise number. If I'm about to get hit by a car, I'm not so concerned about whether that car's velocity is precisely 24 mph or 22 mph or whatever – but I desperately want to know if the number is 20 or 200 or 2. Scientific notation quickly tells us the order of magnitude: roughly, 2.4×10^4 is about 10^4 – ten thousand – and I really need to know more precisely I can look at the number multiplying the power of ten. Scientific notation also makes it easy to multiply numbers together, just remember that exponents add: $10^x \times 10^y = 10^{x+y}$.

1.3.2 Order of magnitude estimates

As in the example of a car about to hit me, we often want to know not precise answers to quantitative questions, but rough, "order-of-magnitude" approximate answers. For instance, suppose we want to quickly estimate how many people attended the presidential inauguration in 2009 (shown in Fig. 1.4). Surely we're not going to comb through the photo and try to count individual people – if we even could, do we really care about the precise number? How can we just get an order of magnitude estimate?

One strategy would be something like the following. In the dense part of those crowds, people are pretty close together. How close? More than a few inches, but definitely less than a meter, so let's say that each person is occupying about $1/4 \ \mathtt{m}^2$. The National Mall, all told,

[7] Govinfo.gov source

Figure 1.4: **Crowd at the 2009 presidential inauguration ceremony** Photograph from the Washington Monunment. Jewel Samad/AFP/Getty Images.

is about 146 acres[8] (and 1 acre is about 4000 $\mathtt{m}^2$), and let's say about half of the surface area of the national mall is covered in dense crowd. Putting that all together:

$$\frac{1\ \texttt{person}}{0.25\ \mathtt{m}^2} \times \frac{146}{2}\ \texttt{acres} \times \frac{4000\ \mathtt{m}^2}{1\ \texttt{acre}} \approx 1.1 \times 10^6\ \texttt{persons}.$$

So, about a million people. We could easily be wrong about the precise number, but we can be pretty confident in the *order of magnitude* of our estimate. Maybe we're a factor of two off, but not a factor of ten! The great thing about this kind of estimation is that it is a fast way to guess rough figures, and you can always go back and refine your assumptions if, say, you want more and more accurate estimates! Are dense crowds really covering "half" of the National Mall? Look at more satellite photos! Is a crowd density of a person per quarter of a square meter reasonable? We can look at studies of typical crowds to get a better sense and a more refined estimate!

This kind of approach is very typical in bringing a quantitative perspective to bear on the world. If someone tells you a number, you can quickly ask yourself "Does the order of magnitude of that number make sense? And is that number big or small compared to other relevant numbers?

[8] A fact I remember from visiting it a few years ago – I could also estimate this by remembering that it took over 20 minutes to walk from one end to the other (so it's more than a mile long), and that the width felt like at least two soccer-fields-worth. We could also more coarsely guess by looking at the image, as we talked about in class!

1.3.3 Fermi estimation

The above example – how many people were in some particular crowd – was a simple example of a class of what are sometimes called "Fermi problems[9]." In such problems we often want to get a sense – an order of magnitude sense! – of the answer to some question that we, a priori, have very little idea of or have almost no data to work with. Some examples are at the end of this subsection, and sometimes they have the feel of interview questions meant to test your general reasoning skills: "How many soccer balls could fit in a school bus?" or "How many grains of sand are there on earth?" or "How many piano tuners are there in the city of Chicago?" These are great types of questions to practice on, because they force us to break a question down, reason about it quantitatively, and make an estimate not about a precise answer but about a relative frame of reference of what the answer has to, more-or-less, be. Later on, if you go to the trouble of making a precise and difficult calculation, this frame of reference helps anchor you; if you're way off, you either made a mistake in your calculation, or perhaps one of your assumptions in making your original estimate was wrong. Either way, you learn something important!

A systematic method for estimation problems

1. What can we assume about the problem? What kinds of pieces of information do we need? Can we break down the problem into sub-problems? In the crowd example, we broke down the problem of counting people into multiplying a density times an area.

2. How big are each of those factors? And what do you do if you don't have a good sense of the answer to that. Here's a **common trick**: come up with a underestimate and an overestimate of your answer, and take the *geometric average*[10] of those numbers! Taking the geometric average is like averaging the *order of magnitude*: $\sqrt{10^{-3} \times 10^{7}} = 10^{(7-3)/2} = 10^{2}$. That's what I did for the crowd density problem: I figured people were more than a few inches apart (so my underestimate was "0.1 $\mathtt{m}^2$") and less than a meter apart (so my overestimate was "1 $\mathtt{m}^2$"), so my estimate of the crowd density was

$$\frac{1\ \mathtt{person}}{\sqrt{(0.1\ \mathtt{m}^2 \times 1.0\ \mathtt{m}^2)}} = \frac{1\ \mathtt{person}}{0.316\ \mathtt{m}^2} \approx \frac{1\ \mathtt{person}}{0.25\ \mathtt{m}^2},$$

   where at the end I rounded to about a quarter of a square meter. The point is that, to the accuracy of these calculations, that kind of rounding *just doesn't matter* and, since it makes the math easier, we go for it.

3. Combine the factors you got above into an order-of-magnitude answer (feel free to round a bit, use scientific notation to help).

---

[9] After Enrico Fermi, a prominent Italian-American physicist of the mid-20th century

[10] I.e., if you're estimates are x and y, guess $\sqrt{xy}$! Why do this, instead of just taking the normal average? The normal average, $(x+y)/2$, would bias you towards your overestimate! For instance, suppose you feel like some answer is not smaller than 1 and not bigger than 1 million. The normal average – 500,000.5 – feels like you're really just plugging in your guess for the biggest it could be, whereas the geometric average – 1,000 – feels like it's accounting for both ranges of your estimate.

4. Interpret your answer, and compare it with what you know about the world! If you get a very different answer than someone else, is it because you made different assumptions? Or is it because you made different estimates on the same set of assumptions?

5. Refine each step of this process. This, after all, is science! Go back, question your assumptions, make better and better estimates, until you are satisfied with your level of quantitative understanding of the problem.

By the way – you might be wondering: why can't I just use the "geometric averaging" trick *at the very beginning of the process* and just jump to the answer? The issue is that your guess will be very sensitive to the *quality* of your underestimates and overestimates, so your answer could be very far off. The goal in Step 1 above is to not just break the main problem down into arbitrary sub-problems, but rather to break it down into subproblems that you have a better intuition for, or will be better able to make good approximations for.

1.3.4 Case study: trees and paper

To illustrate this method, let's work through it in the context of the following question: **How many trees are needed for all toilet paper production in the US in a year?** Where to even begin, right? At first glance, it would be easy to give up, or just guess wildly – is it a 100,000 trees? A billion trees? Probably not either, but I'm just flailing – after all, I know basically nothing about trees, or industrial production methods, or the toilet paper habits of anybody other than myself. Fortunately, with our quantitative approach there is no reason to be deterred from coming up with a more reasonable.

Assumptions and problem break-down: Okay, trees used to produce toilet paper in the whole US every year... Let's assume that the amount produced for use in the US is basically the same as the amount consumed. Maybe the US is a big net exporter of toilet paper – or maybe it's an importer – but I don't have a great reason to think either way, so I'll assume I'm just trying to estimate consumption.

With this, I guess it would be helpful to think about (a) how many people are there in the US, (b) how much toilet paper each person uses per year, and (c) how much "tree" it takes to produce some typical unit of toilet paper.

How big are those different factors: Well:

(a) I know that the US population is about 330 million people – I actually don't know if that's only US citizens or if it's everybody who lives in the US, but it doesn't matter – the answer will be the same order of magnitude, and citizenship status doesn't meaningfully affect toilet paper use (unlike, say, what country you are actually *living in*, which does!).

(b) Hmmm... I'm not really sure exactly how much each person uses per year! But, as a rough estimate, it's probably not that different from "1-2 standard rolls every week." We split the different and call it 70 rolls per year.

(c) Now I *really* don't know! But let's combine our estimates from daily life with use our geometric averaging technique: A roll of toilet paper is about half a pound, which seems like a reasonable starting point. How much does a tree weigh? First of all, I don't even know what kinds of trees are used to make toilet paper. Let's pretend it's made out of pine trees, so I'm picturing a pine tree which is about 40 feet tall and 2 feet in diameter[11]. Real trees have complicated shapes, but if I just pretend a tree is a cylinder, that's a tree whose volume of wood is $\approx 125\ \mathtt{ft}^3$. How much does a cubic foot of pine wood weigh? From days in my youth spent chopping firewood, I'm just going to guess about 30 pounds. How efficiently is tree mass converted into toilet paper mass? Again, no idea! But for most industrial processes I think a 10% yield would be bad, and a 99% yield would be unreasonably effective, so I'll guess that $\approx 30\%$ of tree mass is converted to toilet paper mass.

Combine the factors you got above into an order-of-magnitude answer: Putting all of that together:

$$(3.3\times 10^8\ \mathtt{people})\times\frac{70\ \mathtt{rolls}}{1\ \mathtt{person\ year}}\times\frac{0.5\ \mathtt{lbs}}{1\ \mathtt{roll}}\times\frac{1\ \mathtt{tree}}{125\ \mathtt{ft}^3}\times\frac{1\ \mathtt{ft}^3}{30\ \mathtt{lbs}}\times\frac{100\ \mathtt{lbs\ tree}}{30\ \mathtt{lbs\ paper}}$$
$$\approx 1.0\times 10^7\ \frac{\mathtt{trees}}{\mathtt{year}}.$$

Interpret your answer: 10 million trees per year used in the US just to produce toilet paper! Honestly, that answer is a bit larger than I (secretly) expected. On the other hand, I bet the US has a few hundred *billion* total trees, so I like that my answer came out to less than a percent of all trees in the US used every year – that seems like a good sanity check! Now, how do some of my estimates compare with other people? Honestly, I picked this particular case study because the numbers I found on the internet are all over the place! In the same source I found, for instance, estimates that said you could get anywhere from 200 - 2000 rolls out of a typical tree. Our estimate (2250 rolls per tree) is at the upper end of that. On the other hand, I found a different article claiming we use the equivalent of 27,000 trees per day *worldwide* in toilet paper – which would be about 10 million trees per year: if correct, that would mean we have a strong *overestimate*, since we only considered the US. On the other other hand, yet another article quotes someone as estimating that the toilet paper industry in the US pulps 15 million trees per year – right in the same ballpark as our guess. The delightful Toilet Paper fun facts webpage[12] makes some self-contradictory claims, implicitly guessing that the answer to our question should be either 1.8 billion trees *or* 40 million (without noticing that it is making multiple very different predictions).

Refine: Not that I want to, but if needed we could definitely refine a lot of the estimates I used above! A particular source of uncertain is in that estimate of "how many rolls can you get out of a tree?" I picked specific dimensions and guessed the density of a species of tree, but without much knowledge. In industry, what age of tree is typically harvested? Are some species used more than others? How efficient *is* the wood pulping process at converting wood to paper pulp? And so on. In many ways, the goal is not to get precisely the right answer – as we see from the above, it is apparently hard to even look up what the official answer

[11]Even as a scientist, sometimes it's easier for me to think in the units I was brought up in!

[12]I cannot believe I actually went to this website in the course of preparing to teach this class. You're welcome?

is in this case – but rather to start expanding our understanding by feeling our way around the order of magnitude we think the answer ought to be!

1.3.5 Practice examples!

One beautiful thing about Fermi estimation problems is that there is basically no shortage of them you can come up with – none of us are experts in everything, after all! Here are a handful of random examples you might ask yourself:

1. How many total hairs are there on the heads of everyone in the city of Atlanta?
2. What is the total length of grass that grows on a well-watered suburban lawn every summer? How much water does that family spend to get the grass to grow?
3. How many people in the world are looking at Snapchat *right now*?
4. Oh no – an asteroid the size of Texas just left the asteroid belt and will hit Earth in 18 days! How big a bomb should we hit it with to knock it sufficiently far off course?
5. How many times *in the history of the world* has a toe been stubbed?
6. How many miles can you drive before a one-molecule-thick layer of rubber gets worn off the car tires?
7. How many m&m's are in that jar?

Try a few out, or, better yet, think of your own question, and come up with an estimate? Some of the questions above might be hard to start answer without at least some extra knowledge – you may or may not know how big is an atom, for instance – but think critically and see if you can make a good guess anyway! One of the goals, here, is to start getting used to thinking about the world with numbers attached, and see how much more you can understand in the process!

1.4 Dimensional analysis

Knowing that physical quantities have dimensions attached to them, we can immediately start deducing relationships about things in the physical world. If I throw a baseball and watch it land some distance away, what happens if I throw another one twice as hard? What determines how long it takes for a pendulum to swing back and forth (and, implicitly, why were they so useful in clocks?)? Even before we understand the detailed workings of a system, we can already extract lots (and sometimes basically all) of the relevant information about what we expect to happen, just by thinking about dimensions and units. Let's see how this type of reasoning works!

1.4.1 Basics of dimensional analysis

So far, this is a very straightforward way of thinking about dimensions and units. What we've seen above is that there is a kind of "arithmetic" to units and dimensions: If you multiply together dimensional quantities the dimensions also multiply together. It is also the case that in an equation, every separate term has to have the *same* dimensions in order for that equation to make sense. That is, a combination like

$$1\ \mathtt{m} + 1\ \mathtt{kg}$$

doesn't make any sense! This is the first step in dimensional analysis, and it can help you check your work, and also deduce physically relevant information from different expressions.

For instance, Newton's theory of gravity tells us that the force experienced by two objects due to gravitational interactions is

$$F = G\frac{m_1 m_2}{r^2}.$$

In this expression m_1 and m_2 stand for the *masses* of the two objects, and r stands for the distance between them. G is some constant, and we can work out the dimensions of G by insisting that the *dimensions* of both sides of the equation are the same (otherwise, as we said, the equation would be meaningless!). We know from our definitions about that $[F] = \mathbb{M}\mathbb{L}\mathbb{T}^{-2}$, so we must have

$$\mathbb{M}\mathbb{L}\mathbb{T}^{-2} = \left[G\frac{m_1 m_2}{r^2}\right] = [G]\,\mathbb{M}^2\mathbb{L}^{-2}.$$

Evidently, then,

$$[G] = \mathbb{M}^{-1}\mathbb{L}^3\mathbb{T}^{-2}.$$

The version of dimensional analysis presented above feels, perhaps, a bit dry: a book-keeping technique to make sure we don't accidentally write down an equation incorrectly. As promised, though, it can also be a gateway[13] to physical understanding. Let's illustrate this with a pair of examples. **The following two case studies are not meant to be easy or obvious – they are probably harder and more subtle than anything else in this class!** I hope they give you a sense, though, of just how powerful thinking about units and dimensions can be! As we go through this, though, focus on the *logic* of our approach, and make sure you ask yourself what important assumptions we are making (and how those assumptions could be wrong)!

1.4.2 Case study 1: The Trinity detonation

On July 16, 1945, the first detonation of an atomic bomb[14] was conducted in the deserts of New Mexico. The details of this test detonation, including how powerful the explosion was, were classified for a long time. Before they were fully declassified, though, *photographs* of the test were published, a few of which are reproduced in Fig. 1.5. Actually, those photographs,

[13] or shortcut

[14] "Polla ta deina, kouden anthropou deinoteron pelei..." Sophocles' *Antigone*, (332-333),

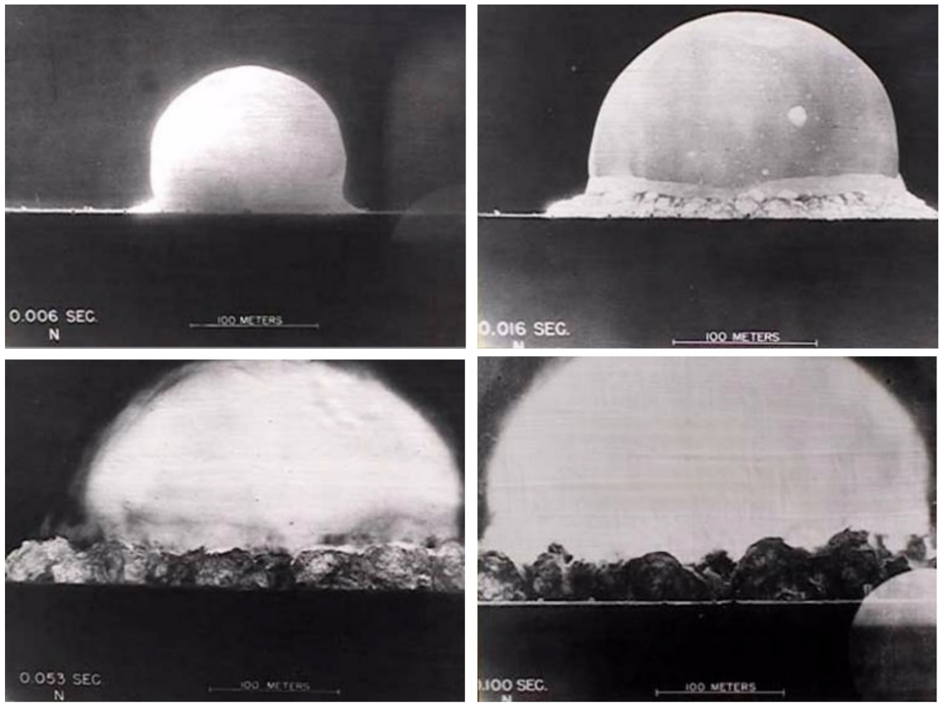

Figure 1.5: **Photos of the Trinity test** Four photographs, with the time from the initial detonation, and an extremely helpful scale bar allowing us to measure the size of the blast.

combined with dimensional analysis and some physical intuition[15], contain all the information we need to estimate how powerful the bomb was. How is that possible?

Let's think systematically about the information we have. From the photographs, the thing we can easily measure is *how big the blast is* as a function of time. Let's *assume* the blast is basically a sphere of radius r, and write this measurement as

$$r = f(t).$$

So far, we have no idea what that function f is, but we can add some more physical intuition to the situation. We've written $f(t)$, but our intuition tells us that the function should depend on a lot more than just time, right? It should also depend on the *energy* released in the explosion (bigger and smaller bombs presumably have bigger and smaller blast sizes, all else being equal), which we'll call E. Additionally, the size of the explosion should probably depend on the density of the air in the environment (does this make sense? What do you think the relative sizes of an explosion would be in air vs in water?), which we'll call ρ.

Now we're getting somewhere! We now have

$$r = f(t, E, \rho),$$

that is, the radius should be a function of time, energy, and density. *Even though we still have no idea what the function f is*, dimensional analysis gives us a powerful tool wrapped

[15]Here we're essentially following an argument due to G.I. Taylor, who made this estimate without having access to the classified data.

up in the seeming innocuous statement that the dimensions of both sides of that equation *have to be the same* in order for the equation to make sense! And remember, we can't just add different dimensional quantities together – adding quantities with different units doesn't make sense – but we can *multiply* dimensional quantities to try to build up a composite expression.

So, we know that $[r] = \mathbb{L}$, which means $[f(t, E, \rho)] = \mathbb{L}$, too! That means we should figure out combinations of time, energy, and density can give us dimensions of length. Here's a systematic way to do the algebra:

1. We want $[f(t, E, \rho)] = \mathbb{L}$. We know that $[t] = \mathbb{T}$, $[\rho] = \mathbb{M}\mathbb{L}^{-3}$, and $[E] = \mathbb{M}\mathbb{L}^2\mathbb{T}^{-2}$.

2. We know we'll have to *multiply* different powers of t, ρ, and E together in order to get something which has dimensions of $\mathbb{L}$. Let's write that equation as

$$\mathbb{L} = [t]^x [E]^y [\rho]^z .$$

3. We *substitute* the dimensions we know for each of those quantities, and rearrange slightly, to get

$$\mathbb{L} = \mathbb{T}^{x-2y}\mathbb{M}^{z+y}\mathbb{L}^{2y-3z}.$$

4. Compare exponents on the two sides of the above equation (if it helps, you can think of the left hand side, $\mathbb{L}$ as $\mathbb{T}^0\mathbb{M}^0\mathbb{L}^1$), and realize we have three equations in three unknowns:

$$\begin{aligned} x - 2y &= 0 \\ y + z &= 0. \\ 2y - 3z &= 1 \end{aligned}$$

5. This system has the solution

$$x = \frac{2}{5}, \quad y = \frac{1}{5}, \quad z = \frac{-1}{5}.$$

6. Putting that all together, we have deduced that the function f, *to be dimensionally correct*, needs to basically be this funny product of its variables:

$$r = f(t, E, \rho) = C \times E^{1/5}\rho^{-1/5}t^{2/5},$$

where C is some dimensionless constant, which we do not know the value of.

Rearranging the above equation to solve for the energy of the explosion, we get

$$E = \frac{1}{C}\frac{r^5\rho}{t^2}.$$

Finally, we can make our estimate! We first *guess*[16] that $C \approx 1$. We then *measure* from the photograph that at $t = 0.006$ s the radius of the blast is $r \approx 77$ m, and we know that the density of air is about $\rho \approx 1.2\ \mathtt{kg\ m^{-3}}$. Putting this all together:

$$E \approx (77\ \mathtt{m})^5 \times (1.2\ \frac{\mathtt{kg}}{\mathtt{m}^3}) \times \frac{1}{(0.006\ \mathtt{s})^2} \approx 9.0 \times 10^{13}\ \mathtt{J}.$$

In the units bombs are typically expressed in ($1\ \mathtt{kiloton\ of\ TNT} = 4.184 \times 10^{12}\ \mathtt{J}$), this is an energy release of about 21.6 `kilotons of TNT`.

The value from the most recent thorough analysis? $E = 22.1 \pm 2.7$ `kilotons of TNT`. Not too bad for knowing only dimensional analysis and nothing about explosions, fireballs, and shockwaves! This kind of "back-of-the-envelope" physical estimation is something we'll see throughout this class; this precise calculation of blast radius has been used to try to estimate everything from bombs to supernovae to the August 4, 2020 explosion in Beirut.

1.4.3 Case study 2: Can you estimate how fast your (car / train / plane) is going be watching raindrops rolling on the side windows?

But this class is about seeing physics in the everyday world around us, so let's turn to a more prosaic example. Have you ever stared out the side window of a car while it was raining and watched the rain drops roll diagonally across the window, as shown in Fig. 1.6? Or on an airplane? It makes sense that the *angle* of the tracks depends on how fast the vehicle is moving – if your car is stopped the drops roll straight down, and on a fast-moving plane they move almost horizontally – but can we use dimensional analysis to estimate what the relationship between angle and speed really is[17]? Of course we can! This section may seem more difficult than the previous one, but don't worry! We'll go into a deeper exposition of things like force diagrams and fluid flows later in this course – for now let's just sit back and enjoy the ride (and focus on following the logic of our approach)!

The schematic illustration in Fig. 1.6 shows what's happening from the point of view of one of the droplets. The train is moving with some velocity, v_{train} to the left, and as the air whooshes past it the droplet feels a drag force, F_{drag}, from being "dragged" through the air. Gravity is trying to pull the droplet straight down with a force $F_{gravity}$, and as the drop moves on a diagonal trajectory (with angle θ) there is a friction force, $F_{adhesion}$, stemming from the adhesion between the water and the window and trying to slow the droplet down.

So, what do we do? We want to be able to estimate the speed of the train, (I'll drop the subscripts where possible) v, by observing the angle of the droplet's tracks, θ. We don't really care about the velocity with which the droplet, in its streams or in its fits and starts, makes it's way across the window, so we ignore v_{drop}. Similarly, the force of adhesion *slows down* the droplet, but it simply opposes the motion of the droplet, no matter what direction it tries to go. So, $F_{adhesion}$ doesn't affect the relationship between v and θ. That leaves only

[16] There are some physical reasons that this should be the case, but in dimensional analysis we often do guess that the prefactors aren't *too* different from unity

[17] And does it matter how hard it's raining? Or how big the droplets are?

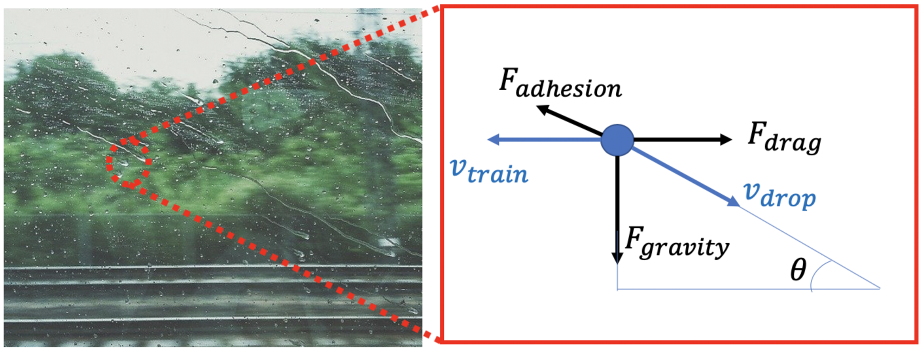

Figure 1.6: **Rain on a train window** (Left) A photo of diagonal rain tracks on a fast-moving train. (Right) A schematic representation of a droplet, the velocity of it and the train, and the forces the droplet experiences. Image from pixabay.

gravity and drag forces in play, so we conclude (with the help of the figure) that

$$\tan\theta = \frac{F_{gravity}}{F_{drag}}.$$

The force due to gravity is quite simple – as you may have seen and as we'll see in more detail later in this course, there is a typical acceleration due to gravity for an object near the surface of the Earth, and it is $g = 9.81\mathtt{m}\ \mathtt{s}^{-2}$, and the force due to gravity on the raindrop of mass m is simply

$$F_{gravity} = mg.$$

But what about F_{drag}? As the poor droplet is sticking to the window, being buffeted by the air flowing past it, what physical quantities matter most in determining the net drag force the droplet feels? Some nature things come to mind (see if you can anticipate all the items on this list before looking!):

1. The droplet radius, certainly, where $[r] = \mathbb{L}$.
2. The speed of the train, where $[v] = \mathbb{L}\mathbb{T}^{-1}$.
3. It makes sense that the density of the air might matter, since that helps tell us how many air molecules the droplet is bumping into as it is whisked along. As before, $[\rho] = \mathbb{M}\mathbb{L}^{-3}$
4. Finally, we must not forget about the *viscosity* of the air! The friction applied by a fluid flowing around an object is measured by the fluid's viscosity (think about the difference in how honey flows compared to water, for instance), which we denote by the symbol η, and where $[\eta] = \mathbb{M}\mathbb{L}^{-1}\mathbb{T}^{-1}$.

Uh oh – this problem feels a bit different from the one where we estimated the energy of a blast! Back then, we only had one way of smushing together the physically important quantities (time, energy, and density) into a combination with dimensions of $\mathbb{L}$. Here, though, *there is more than one way to combine the above four quantities into something with dimensions of force*! What are we to do?

One way to express the fact that there are multiple ways of combining those four ingredients is to *form a dimensionless quantity* out of some of them. Let's take a particular combination, often called the "Reynolds number[18]:"

$$Re = \frac{\rho r v}{\eta}.$$

This combination of quantities, as you can check for yourself, does not have any dimensions associated with it – it is a pure, simple number! In contrast, we can put together a different combination of quantities that has the dimensions we want – dimensions of a force – like $\rho r^2 v^2$.

So, dimensional analysis doesn't get us all the way to our answer, but it *does* tell us that we can write the drag force on our droplet as

$$F_{drag} = (\rho r^2 v^2) f(Re),$$

where $f(Re)$ is some function of the Reynolds number that, a priori, we know nothing about[19]. That's a bit unfortunate, because Re itself depends on v – we wanted a relationship between v and θ, but our ignorance of this function f is getting in the way!

As it turns out, folks have *measured* this function $f(Re)$, and they have *observed* a proportionality

$$f(Re) \propto \begin{cases} 1/Re & \text{for } Re \lesssim 10 \\ 1 & \text{for } Re > 100 \end{cases}.$$

Knowing this makes our life a lot easier! Taking number for air ($\rho \approx 1.2\ \mathtt{kg/m^3}$, $\eta \approx 1.5 \times 10^{-5}\ \mathtt{kg/(m\ s)}$), and assuming a raindrop that is half a centimeter ($r \approx 5 \times 10^{-3}\ \mathtt{m}$), we find that we are in the "simple" range of the function f's behavior, with $Re > 100$, as long as $v > 0.25\ \mathtt{m/s}$; that's barely more than half a mile per hour – not a big ask!

So, for any reasonable plane, train, or automobile velocity we replace the whole function $f(Re)$ with some constant[20], c_1, and write our estimate for the drag force:

$$F_{drag} = c_1 \rho r^2 v^2.$$

It took us a while, but now we're basically done. Going all the way back to when we related the angle of the droplet trajectories to the forces due to drag and gravity, we have

$$\tan\theta = \frac{F_{gravity}}{F_{drag}} = \frac{mg}{c_1 \rho r^2 v^2}.$$

Rearranging to solve for the velocity of the train, we get

$$v_{train} = \sqrt{\frac{mg}{c_1 \rho r^2 \tan\theta}} \propto \frac{\sqrt{r}}{\sqrt{\tan\theta}}.$$

[18] A number we'll meet in greater detail in the second half of this course – it comes up all the time in problems involving flowing liquids and gasses. Roughly speaking, small Reynolds numbers tend to correspond to situations with smooth flows, and high Reynolds numbers to situations with a lot of turbulence. That bumpy airplane ride? High Reynolds numbers.

[19] This makes our dimensional analysis arguments very slippery! the function f could be anything, after all.

[20] Which, again, we don't know off-hand, but which we can measure

So there you go: you estimate the angle, you estimate the size of the droplet, and you make sure you do at least one real measurement so you know the constant of proportionality. Do all that, and you'll be able to estimate speed from rain drop tracks for the rest of your left.

This example also illustrates one of the shortcomings of dimensional analysis. Because we didn't do a full calculation, we are often left with some quantities – those constants C and c_1 in the above examples, and also the function $f(Re)$ – that we don't know how to estimate. We can be sure they're unitless, but not much more than that. Thus, when relying on dimensional analysis one often needs to bring in extra information – say, a *measurement* of the system at least once – in order to determine unknown quantities.

Stop and reflect! We've come a long way already! Just by considering the dimensions of different physical quantities, insisting that different parts of an equation have to have dimensions that match to make sense, and adding some "common sense" physical reasoning, we've found that we're able to predict all sort of relationships between things in the physical world. And we didn't have to be experts in explosions, or in aerodynamics, or anything like that – dimensional analysis let us put in a bare minimum of physical understanding and get out quite a lot!

Make sure you pause and think about, for instance, the implications of our calculation in the last section. What happens to a trajectory if the train keeps going at the same speed, but a raindrop bumps into another raindrop on the window and, hence, get's bigger (for instance)?

Of course, dimensional analysis it's not the be-all-end-all of techniques: we made many assumptions along the way, and these assumptions will break down. Ultimately, to know the correct answers to things like "what was the energy of the explosion?" or "what is the relationship between train speed and droplet trajectory?" – and to have confidence in those answers – we will need to write down real theories that account for all of the details. But as a *first effort* at understanding a physical system, thinking about dimensional analysis and about units is an incredibly powerful tool to get rough but quantitative relationships between quantities.

Looking for more? Here's a question for you: How big is a mammal? Suppose all you know are really basic facts about the world – how big an atom is, how big the earth is, the energy in a typical covalent bond in a carbon-centric molecule. Is it possible to combine these sorts of quantities with physical reasoning to estimate how big animals *ought to be*? A fun thing to think about! If you're interested, here's a link to just such a calculation (due to William H. Press) that combines dimensional analysis with physical intuition to make an estimate of *what the answer probably had to be*!

Chapter 2

The laws of motion, Part 1!

In this class we're trying to gain a quantitative understanding of the world around us, and explore the physical laws that govern the behavior of the world around us. In this chapter we will cover Newton's Laws of Motion, a compact set of principles that are an attempt to understand the field of physics called "Classical Mechanics." Classical mechanics is the study of how objects (from, say, projectiles to planets) move, and if you've had a physics class before it probably exposed you to the ideas we're about to explore. Even though it may seem quite simple, this chapter will introduce many of the most important ideas about basic physical motion – forces, accelerations, the conservation of energy – that we'll repeatedly use in seemingly very different contexts throughout our exploration of the physics of the everyday world.

Before we get started, we need to talk about the difference between ***scalar*** and ***vector*** quantities: physical quantities can be either type, both come up all the time in the physics, and so it's crucial to internalize what these terms mean.

"***Scalar*** quantities"? That's just a technical term for quantities that can be defined by a single number (together with the appropriate units), and many physical quantities have this character. For example, if you want to tell me the *temperature* outside, you can just say "30 degrees Celsius[1]." If you ask me how long our classes are, I can say "75 minutes." And if I ask you how far away the earth is from the sun, you can say "About 93 million miles away[2]."

So, scalars are simple enough: just a number (or, you might say, a "magnitude") and whatever units you need. "***Vector*** quantities," in contrast, require both a *magnitude* and a *direction* from some reference point. A classic example is *position*: if you want to specify the position of an object, you need to give me a reference point to start, a direction to go away from that reference point, and an amount to move in that direction. For instance, in Fig. 2.1 I can tell you where the top of the hourglass is by saying "27% of the painting's total width to the left of the top of the candle's flame." I have a reference point (the top of the flame), a direction (left), and a measure of how far to move in that direction (a specified fraction of

[1]A unit named after Anders Celsius, a Swedish physicist of the 18th century. Did you know that Celsius' original temperature scale was *backwards*, so that "0 C" was the boiling point of water and "100 C" was the freezing point? It was a strange convention, and later scientists kept the name and the meaning of 1 degree Celsius, but flipped the scale so that larger numbers corresponded to hotter temperatures

[2]"and that's why it looks so small"

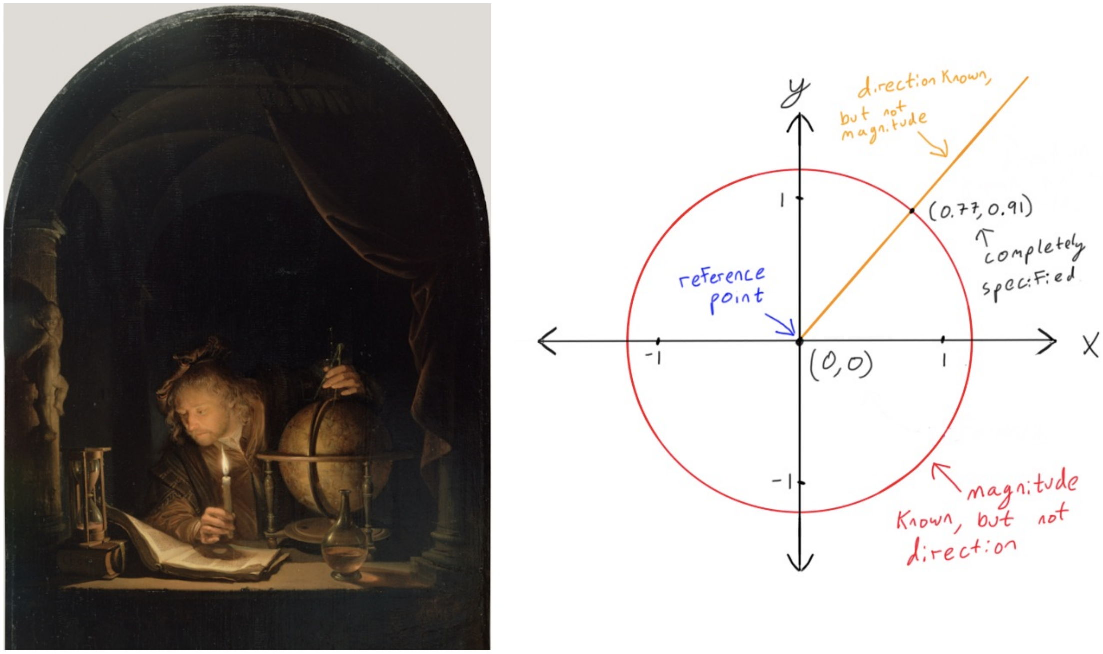

Figure 2.1: **Positions are vector quantities** (Left) "Astronomer by Candlelight," Gerrit Dou, circa late 1650s. Image courtesy of The J. Paul Getty Museum, Los Angeles (Right) An abstract representation specifying a position in a two-dimensional space.

the painting's size). If I had left out any of those three pieces of information, the position would *not* be completely specified!

The right hand side of Fig. 2.1 shows this more abstractly by imagining the position of a point in the two-dimensional plane. The *origin* is our reference point. The red circle shows all points a given *distance* away from the origin, and the origin line shows points a specified *direction* away from the origin. Only the black point at the intersection of the line and the circle *completely specifies the position of a point away from the origin.*

Many physical quantities have this vector character to them: positions (as we've just seen), velocities, accelerations, and forces, for instance, are all vector quantities. For example, if you run 1000 m North in 100 seconds in straight line at constant speed, your velocity during that time is 10 m/s North. In these lecture notes we'll use the convention that a symbol with a little arrow over it represents a vector (e.g., a vector position $\vec{x}$, or a vector force $\vec{F}$): the arrow reminds us that the quantity in question has both a magnitude and a direction.

2.1 Inertia, friction, and why our intuition is often wrong

A basic question for us to think about: suppose we have an object at rest – say, a book on a table – and we do not subject it to any outside influences (i.e, we don't push or pull on it). What happens? Does it stay still? Does it start to move? Similarly, what if we push on the object with a constant force: does it start moving at a constant speed? At an accelerating speed? What if we have, not a book on a table, but an apple falling from a tree?

These elementary questions are deceptively difficult, in part because they refer to an extremely *idealized* version of how objects exist in the world: on Earth objects are never really from *all* outside influences – they're always pushing on other objects, or sliding against them, or rolling over them, or bumping into air molecules, etc. Friction, in particular, gets in the way of our physical intuition, here. It's why even someone as smart as Aristotle[3] got the answers to the above questions profoundly wrong: he believed that the velocity of an object would be proportional to the force exerted on it. This theory makes apparently correct predictions for sliding objects (like the book on a table), but it also predicts that heavier objects should fall faster than lighter objects.

It was not until Galileo[4] that more better solution was proposed: an object that begins in a stationary state will *remain* stationary if no forces act upon it, and an object that is already moving will *continue* moving in the same direction at the same rate until some force acts upon it.

This property of objects – a fact about how our universe works! – is called *inertia*. This property is sometimes summarized as "an object in motion tends to remain in motion; an object at rest tends to remain at rest," and sometimes it is framed as an object's resistance to changes in its own velocity[5]. Inertia was not discovered for so many thousands of years because of *friction*. When you try to slide in shoes on a basketball court, friction quickly slows you down even if you were running at full speed; when you push on a massive object lying flat on the ground, the presence of friction means you can be exerting quite a lot of force without generating any motion whatsoever.

2.2 Skating

Friction, in other words, is a nuisance to us at this stage in our understanding: to uncover how inertia works, we need to somehow get rid of it. One way we could do this would be to imagine ourselves floating in outer space; a more terrestrial method would be to strap on a pair of ice skates. Skates don't completely eliminate friction, of course – they just make it much smaller than it is when we're wearing shoes – but they help with our intuition. Wearing skates, one push will send you gliding for a good distance (before friction finally wins and grinds you to a halt). In this section we'll go even farther, and imagine we're wearing *completely perfect* ice skates that somehow experience no friction whatsoever; in the same vein, we'll neglect any air resistance you might experience. Now a single push will send you gliding indefinitely – you'll move at a constant, steady speed in precisely the same direction until you either exert more force or you bump into another object.

With this scenario in mind, we can restate the property of *inertia* in terms of how an object's velocity behaves:

---

[3]Whose writings influenced scholarship for over a thousand years! On the other hand, he also thought that women had fewer teeth than men, so... how smart could he have been?

[4]or, arguably, Avicenna about 500 years earlier than Galileo!

[5]Newton himself, writing in his particular style, defined the idea as "The vis insita, or innate force of matter, is a power of resisting, by which every body, as much as in it lies, endeavours to perserve in its present state, whether it be of rest, or of moving uniformly forward in a right line." He liked commas.

> **Newton's 1st Law of Motion**: An object will continue to move at a constant velocity, unless acted upon by a force.

Note that this statement also covers the case of objects at rest: objects at rest simply have a constant velocity $\vec{v} = 0$.

To summarize: our intuition might say that we need to continually push an object in order to keep it moving, but the laws of Nature say that an object will move at a constant velocity when no force acts upon it. This apparent conflict is resolved by the presence of other forces acting on the object – like friction, or air resistance – that have a tendency to slow objects down.

2.2.1 Acceleration and forces

Newton's 1st Law says that if no forces are acting upon our hypothetical ice skater, then that skater will move at a constant velocity, with neither direction nor speed changing. What if forces *do* act, though? We all know: if something pushes on us, our *velocity changes*, and in physics we call the rate at which an object's velocity changes with time that object's *acceleration*. Just like position, velocity, and force, acceleration is a *vector* quantity: it has both a magnitude *and a direction*. In the English language we have specialized words that refer to accelerations that point in the same direction as an object's velocity or directly against that direction: in the first case an object moves faster and faster and we say it is *accelerating*, and in the second case the object slows down and we say it is *decelerating*. In physics we typically have no need for this multiplication of basic terminology, and we say that *any* change in velocity is an acceleration. Since velocity is both a speed and a direction, that means that speeding up, slowing down, and turning *are all* types of acceleration.

So, how much do objects (or skaters) accelerate when acted on by forces? At any moment in time, an object only has a single vector quantity for its acceleration, so what we need to know is the total force, or *net force* an object is experiencing. If you experience a total force – say, of 1 `N` to the right – it doesn't matter if that force is the result of you pushing against the ground, or you being pulled by a whole team of horses working together, or if your in a tug-of-war with a whole group of people pulling in different directions: your acceleration will point in the direction of the net force, and the amount you accelerate will depend on the magnitude of the net force. This idea is illustrated in Fig. fig:iceBlocks.

What is the precise mathematical relationship? It's

$$\vec{a} = \frac{\vec{F}_{net}}{m},$$

that is, the acceleration of an object (a vector) is equal to the net force (a vector) divided by the object's mass[6] (a scalar). This is

> **Newton's 2nd Law of Motion**: The vector sum of the forces, $\vec{F}$, on a object is equal to the mass, m, of that object multiplied by the acceleration, $\vec{a}$, of that object.

[6]By the way: does it surprise you that the *same* quantity – the mass of an object, m – appears both as a measure of inertia (how much an object resists changes in its velocity) *and* in the expression we saw in Chapter 1 for the force of gravity between two objects? Did it have to be that way?

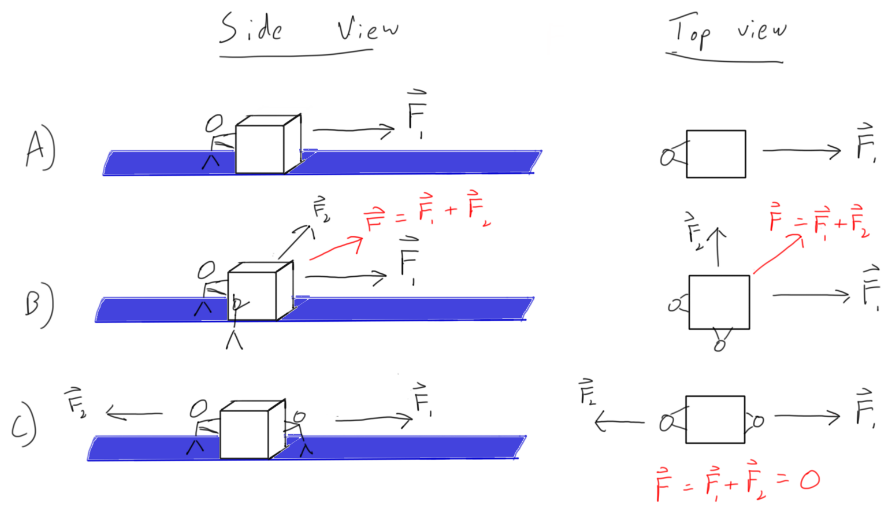

Figure 2.2: **Applying forces to an object on ice** (A) A single person pushes a block to the right; the total force is to the right, and the object accelerates to the right. (B) One person is pushing to the right, and a different person is pushing the same object up. The total force is the sum of the two forces, and so it points diagonally up and to the right. The acceleration would be in the direction of the total force. (C) Two people push equally hard in opposite directions. The total force is the sum of the individual forces, *and is zero*! As a result, the object experiences *no net force* and so it doesn't accelerate at all.

This law is often summarized as $\vec{F} = m\vec{a}$, but I actually prefer the (equivalent) formulation I gave above, since it helps us keep cause and effect straight: forces are the cause, and accelerations are what result from forces.

This law tells us that the more massive an object is, the more force we need to apply to get that object to accelerate by the same amount. This probably feels intuitive: your arms can only output a certain maximum force, and with that maximum force you can throw a baseball which starts at rest to a much higher velocity (that is, you can accelerate it more!) than you could a bowling ball. Similarly, if you strap a jet engine to a car it will be able to accelerate *much* quicker than if those same engines were trying to move a massive plane.

2.2.2 Frames of reference

Before we move away from ice skating, we should pause for a moment and discuss different *frames of reference*. As an example: suppose I am skating on a frozen river[7], and for a while I simply coast. From my perspective, I could imagine that I am perfectly still, and the scenery is moving past me. Someone standing still on the riverbank, though, would say *they* are perfectly still – as is the scenery – and I'm the one moving with a constant (and non-zero!) velocity.

Who's right? Fortunately, we both are! In physics we would say that each of us are observing the world from a different "*inertial frame of reference*," which is a way of saying

[7]I did grow up in the northeastern US, after all

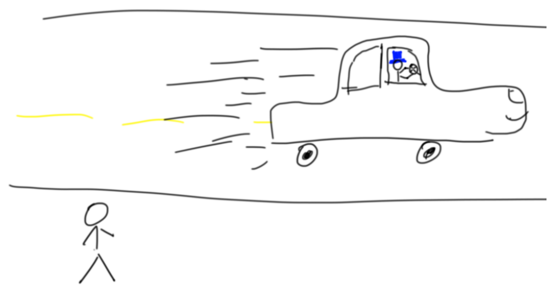

Figure 2.3: **Inertial frames of reference** You watch your friend and her cool hat speed past you in a car, moving to the right. She sees herself (and her hat!) as motionless, with you moving opposite to the direction she is driving.

"from the point of view of an object that is not accelerating and, hence, is moving according to Newton's 1st law." Remarkably, all of the laws of physics work perfectly well in any inertial reference frame. Watching some event from your own inertial reference frame, you will see objects moving, and then accelerating in response to forces, and generally obeying the laws of physics. Somebody else, watching from *their* inertial reference frame, will *also* see objects moving, and accelerating in response to forces, and generally obeying the laws of physics. But you might disagree on physical measurements!

For example, suppose you watch your friend, wearing her extremely cool hat, zoom past in a car going 48 `km/hr` to the right (Fig. 2.3). From your point of view, that hat has a velocity of 48 `km/hr` to the right. To your friend, though, that hat is sitting quietly on top of her head – with a velocity of zero! – as you speed backwards at 48 `km/hr`. Importantly: you both agree that everything is obeying Newton's Laws – neither of you think that the hat is experiencing any net force, and neither of you see the hat accelerate, for instance.

The fact that the laws of physics work perfectly fine in any inertial reference frame means we're free to *choose* the frame of reference that makes our lives the easiest. Usually this will be so obvious that we do it without even thinking we're making a choice: we'll select a frame that makes objects and their motion as simple as possible. Sometimes we'll have to think more carefully about what choice is best for the task in front of us, though – we'll see a few examples in future chapters.

2.3 Throwing stuff

2.3.1 Gravity

We've already casually tossed around the idea of gravity, but what is it? It is one of the fundamental forces that govern the behavior of the universe, and it dictates that there is an *attractive force between every pair of objects*; that is, between every pair of objects there is a force that is trying to bring those objects together. This force is stronger between objects that are more massive, and it gets weaker the farther apart those objects are. Although gravity is omnipresent, it is also *extraordinarily* weak, and in our daily life there is only one object that is both massive enough and close enough that we obviously experience the effects of gravity: the Earth.

Through gravity, the Earth exerts a downward force on every object near its surface, and we call the force with which an object is attract to the Earth's center its *weight*. Near the surface of the earth, the *acceleration due to gravity*, $\vec{g}$, has a value of about 9.8 $\mathtt{m/s}^2$ and points straight down. An object's weight is

$$\vec{w} = m\vec{g},$$

that is, it depends on both the *mass* of the object and the acceleration due to gravity. Thus, when an astronaut is on the surface of the moon their *mass* stays the same, but their *weight* is quite different[8].

2.3.2 Free fall

We see, then, that weight is a measure of force (not just of mass), and using weight we can easily figure out what happens when an object is falling. When an object is falling (and neglecting any air resistance), the *only* force acting on it is its weight. From Newton's 2nd Law, we see that the object's acceleration is[9]

$$\vec{a} = \frac{\vec{F}_{net}}{m} = \frac{\vec{w}}{m} = \frac{m\vec{g}}{m} = \vec{g}.$$

In summary: weight, as a force given by an objects mass times $\vec{g}$, is important to think about when an object *isn't* just freely falling; when an object *is* free falling, $\vec{g}$ tells us what its acceleration is. Note that in the above equation for an object's acceleration, the mass of the object cancelled out at the end. If you drop two lead balls of different mass from the Nieuwe Kerk, they accelerate in the same way and hit the ground at the same time.

2.3.3 Motion of a constantly accelerating object

While acceleration is important, we often want to know *other* vector quantities, like where an object is (position) and how it's moving (velocity). These questions are easy to answer for an object that has constant acceleration – such as an object in free fall. If you wait a certain amount of time, t, then the "final" velocity at that time is given by

$$\vec{v}_f = \vec{v}_i + \vec{a}t,$$

that is, the final velocity is equal to the *initial velocity* plus the *acceleration* multiplied by the *time elapsed.*

[8]The moon, after all, is less massive than the Earth!

[9]By the way: we're going to be confronted with equations more and more in this and the following chapters. Don't panic! Just read them as English-language sentences and think about what they're trying to tell us. In this example, I would read this as "The acceleration is equal to the net force over the mass, but since the only force is weight, this is just equal to the weight over the mass. Weight is just "$m\vec{g}$," so substituting we have acceleration equals mass times the acceleration due to gravity divided by mass. That's just the acceleration due to gravity."

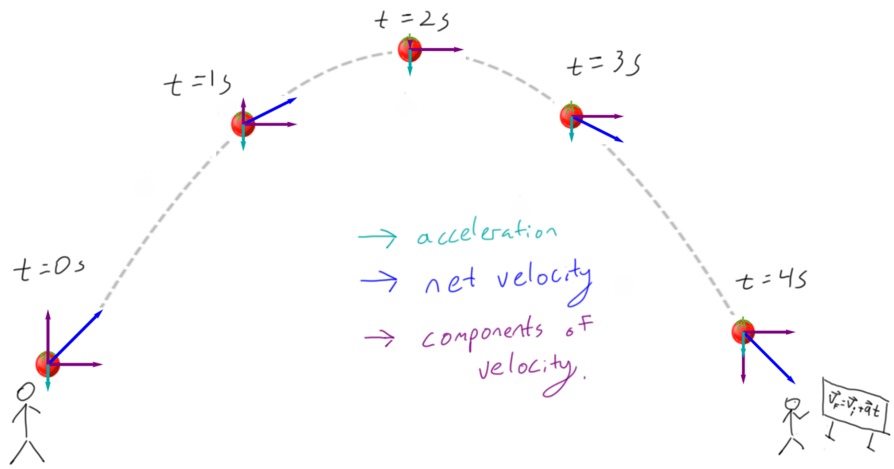

Figure 2.4: **Throwing a tomato** You find yourself sitting in a particularly uninspiring lecture, and you want to find a way to communicate your displeasure with your professor. Fortunately, you have brought a tomato for just such an occasion. After applying the necessary force to get it going, at time $t = 0$ **s** the tomato leaves your hand. At every time thereafter (until, that is, it collides with its target), the tomato feels only the force of gravity, and constantly accelerates downwards. The horizontal component of velocity of the tomato stays constant, while the vertical component initially gets smaller, reaches zero at the peak of the trajectory, then becomes increasingly negative. The illustration shows snapshots of the tomato at various time points; the cyan arrow shows the constant acceleration due to gravity, the blue arrow shows the current velocity of the tomato, and the purple arrow shows the horizontal and vertical components of velocity of the tomato.

Great, but where does the object end up? What is its position? Given a constant acceleration, the object's velocity is constantly changing, but it does so in a smooth and uniform way from its initial to its final velocity. Thus, its *average* velocity is just

$$\vec{v}_{ave} = \frac{1}{2}\left(\vec{v}_i + \vec{v}_f\right) = \vec{v}_i + \frac{1}{2}\vec{a}t.$$

Given this average velocity, the *position* of the object is just the average velocity multiplied by the total time (plus wherever the object started from!). As an equation:

$$\vec{x}_f = \vec{x}_i + \vec{v}_i t + \frac{1}{2}\vec{a}t^2.$$

Pause and reflect! Think about the above equation in terms of *dimensional analysis*. Do the units of everything agree? What could you have predicted each of the terms above by thinking about units and dimensions? What couldn't you have predicted?

2.3.4 Throwing stuff

Even though we were (perhaps) thinking about freely falling objects, we wrote the above equations as *vector* equations, because the same expressions apply regardless of whether an object is being dropped, or thrown straight up and then falling back down, or thrown at some angle. An amazing simplification implied by these vector equations is that we can treat the different "components" of an object's motion – its position/velocity/acceleration in the horizontal/vertical directions, say – independently.

Suppose you drop a rock off a bridge into the river below: once you let go the only force acting on the rock is its weight, and the acceleration is just $\vec{g}$ (pointing down, at 9.8 $\mathtt{ms}^2$). Suppose instead you throw a baseball away from you: you apply a lot of force to get its initial velocity pointing in a particular direction, but once you let go the only force acting on the baseball is its weight, and the acceleration is just $\vec{g}$ (pointing down, at 9.8 $\mathtt{ms}^2$). The *vertical* components of the baseball's motion is affected by gravity, but the *horizontal* components are not. This idea is illustrated in Fig. 2.4.

2.4 Ramps

In this section we'll discuss, among other things, ramps! Why is it that, with the help of a ramp, you can change the elevation of an object so heavy that you otherwise couldn't lift it? Why do we move faster when biking down a steep hill than a less-steep hill? And why is it so much harder to bike uphill, for that matter?

Let's start out by thinking more about the *weight* of an object: an object's weight is the force with which the Earth's gravitational field is pulling that object down towards the Earth's center. When an object – say, an apple – is in free-fall, that force causes an acceleration, but what about when an object is *not* in free fall? You know from experience that a book on a table still has a weight – if you try to pick it up you have to exert a force to do so, and more massive objects require more force – but certainly you don't see the book accelerating: it's not moving at all! So, what's going on?

What's going on is

> **Newton's 3rd Law of Motion**: For every force that one object exerts on a second object, the second object simultaneously exerts a force on the first object which is equal in magnitude and opposite in direction.

This is sometimes summarized as "for every action, there is an equal but opposite reaction," but I think the statement in terms of forces is more explicit and more easily understood.

To see this, we have to think about all of the forces involved, which are illustrated in Fig. 2.5. When sitting on the table, the book is still being pulled down by gravity, and in turn it is exerting a downward force *on the table*! Newton's 3rd law says that, since the book is exerting a force on the table, the table *exerts an equal but opposite-in-direction force on the book*!

People sometimes misunderstand Newton's 3rd Law by using the following sort of reasoning: "I push on an object, and the object pushes exactly as hard back, so these forces cancel and nothing happens." No! The key is that the equal-but-opposite forces in Newton's 3rd Law *act on different objects* – you exert a force on the object, and the object exerts a force on you! What *happens* to you depends on the total set of forces acting on you, and what happens to the object depends on the total set of forces acting on the object.

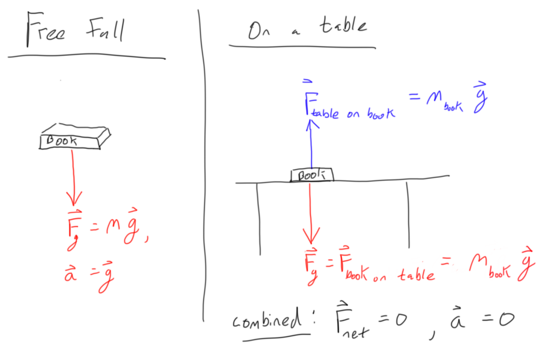

Figure 2.5: **Newton's 3rd Law** (Left) When in free fall, the book feels only the force of gravity, and accelerates with $\vec{a} = \vec{g}$. Not shown: the Earth feels an equal and opposite force, but since it is much more massive, it barely accelerates at all. (Right) When it is sitting on a table, the Earth still exerts a downward force on the book. In turn, the book is pushing down on the table with a force given by its weight, and the table exerts an *equal and opposite* force back upwards on the book! The net force on the book is zero, so it does not accelerate.

To help get this intuition, let's think again about being on a pair of skates and pushing agains a wall. I exert a force on the wall, and the wall exerts a force on me. What are the forces acting on me? Well, in the horizontal direction the wall is pushing on me, and there is no friction to resist that motion, so I move away from the wall! What forces are acting on the wall? Well, in the opposite horizontal direction I'm pushing on it, but the wall (by design!) has lots of anchors and supports that resist that force, so that the total force on the wall is zero, so it doesn't move!

2.4.1 "Support" or "Normal" forces

Okay, so, when an object is resting on the ground (or on a table), it is not accelerating downward because as it pushes down on the ground with its weight, the ground pushes back on the object. What is the nature of these "support" forces that stop objects from sinking into the ground, or through a table?

Because two objects cannot occupy the same point in physical space at the same time, their surfaces repel each other – they push each other apart – whenever they come into contact. The *direction* of this force is *normal*[10] to the surface: if you put an object on flat, horizontal ground, the normal force the ground exerts on the object is straight up in the vertical direction. On the other hand, if you put an object on a sloped surface, the normal force only exerts a force back on the object *normal to its own surface* – this is illustrated in Fig. 2.6.

What about the *magnitude* of the normal force? We can figure this out by observing the behavior of the world around us! When we place a book on a table, the book comes to rest: it is in a state where it is not accelerating (neither flying back up into the air nor sinking into the table). Since there is no acceleration, that *must* mean that the normal force is perfectly balancing the effects of gravity. The magnitude of the normal force, then, is exactly the same

[10] which is just a math term that means "perpendicular"

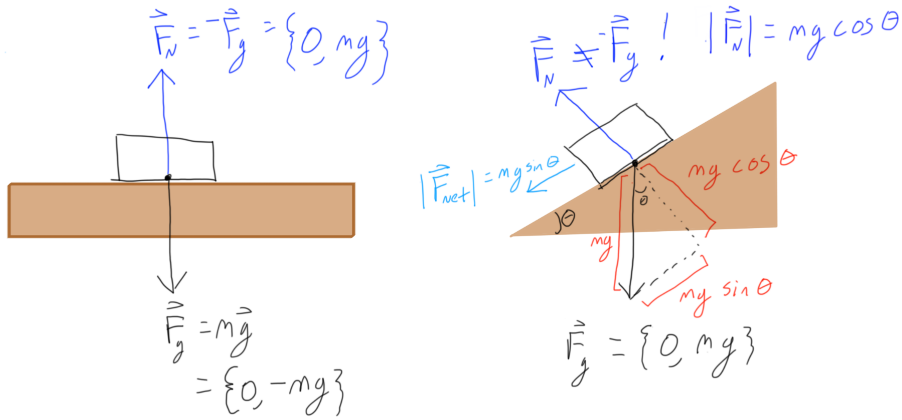

Figure 2.6: **Normal force of an object resting on the ground** *Normal* or *support* forces act in a direction perpendicular to the surface an object is resting on. (Left) On flat ground, an object pushes on a surface with its weight straight down, and the object pushes back with a normal force whose magnitude is the same as the weight: $|\vec{F}_N| = mg$. (Right) On a slope, gravity pulls straight down, and hence the weight of an object points *straight down* on the surface, as well. The normal force, though, only compensates for the *component of the weight in a direction perpendicular to the surface*. As a result, $|\vec{F}_N| = mg\cos\theta$, and there is a net force pointing down the slope with magnitude $|\vec{F}_{Net}| = mg\sin\theta$. If friction isn't strong enough, or if nobody is there to hold the object in place, the object will accelerate down the slope since it has a net force still acting on it!

as the downward weight of the object. If you now jump up and stand on top of the book, the table quickly supplies a normal force which balances the combined weight of both you and the book[11].

What if your object is on a sloped surface (say, you lift up one end of the table)? Again, surfaces only exert forces perpendicular to themselves: the *book* only exerts a portion of its weight normal to its surface on the slope, and so the normal force supplied by the slope is *not* equal in magnitude to the object's full weight. This, too, is illustrated in Fig. 2.6.

From the diagram we already have a sense of how ramps help us moving things up and down. Neglecting friction, we can see that on the horizontal surface we can easily push the block to the left or right, since there are no opposing forces, causing it to accelerate in the direction of our push. If we want to lift a block straight up, we have to supply a force that is *at least as large* as the weight of the block, so that the net force acting on the block is upwards and it starts accelerating in that direction. On a ramp, though, we see that we do not need to supply a force equal to the weight of the block to push it "uphill" – we only need a portion of that force, and the block will accelerate in a direction which is not entirely upwards but nevertheless causes the block to gain some elevation.

Of course, we have to supply that force over a longer distance – the length of the ramp,

[11]Each surface has a *maximum normal force* it can supply: eventually the table will break, or you will sink into the ground a bit; this maximum normal force is a property of the material supplying the normal force.

rather than the elevation we gain – but for most people the constraint is the maximum force they can supply at any given time.

2.4.2 Energy and work

We can explore an alternate way of understanding why ramps make our lives easier by finally talking about *energy* and *work*. The definitions of these concepts might seem a bit circular at first, but they should become clear as we talk more about them!

First, energy is a physical quantity which, as we saw in Chapter 1 has dimensions of $\mathbb{M}\mathbb{L}^2\mathbb{T}^{-2}$, which is something with the dimensions of a forces times a length. Energy is defined as the capacity to do work, or to heat an object up, and work refers to the process of transferring energy. Energy can take many different forms – moving objects have *kinetic energy*, rubber bands have *elastic energy*, the molecules in gasoline stores *chemical energy*, lifting an object up gives it *gravitational potential energy*, and so on – but it plays a very special role in physics because it is a **conserved quantity**. That is, energy cannot be created or destroyed. You can, however, move energy around: you can transfer it from one object to another, and you can convert it from one form into another. This process of transferring energy is called *work*.[12]

In the first part of this class, *kinetic energy* and *gravitational potential energy* are going to be the two forms of energy we'll focus our attention on. Kinetic energy is an energy that moving objects have, allowing a moving object to do work on whatever it hits. The amount of kinetic energy an object has is

$$\text{kinetic energy} = \frac{1}{2}mv^2.$$

Gravitational potential energy is a measure of the energy an object has by virtue of being a certain distance away from the Earth. The amount of gravitational potential energy an object has due to its interaction with the earth is

$$\text{gravitational potential energy } = mgh,$$

the mass of the object times the (magnitude of the) acceleration due to gravity times its height away from the Earth.

If you lift a book a certain distance above the table, there is an energy stored in the gravitational forces between the book and the Earth; if you let go, that potential energy gets converted to kinetic energy! The farther the book falls, the faster it moves downward, until it crashes and does work on the table. What if you throw an object straight up into the air? Immediately after you release it, it has a certain amount of kinetic energy:

$$\text{kinetic energy} = \frac{1}{2}mv^2.$$

[12] Bloomfield's textbook has a nice analogy: in a fictitious fixed-total-currancy world, energy is like "money" and work is like "spending." You can have money in the form of dollars or euros or yen or pesos, you can convert from one form of money to another, and by spending money you can also transfer your money to someone else. It probably feels similarly circular to define money as "the capacity to spend," but you could certainly do so. Also, probably best not to ask an economist about this analogy.

Gravity is constantly making the object accelerate downward, and at the top of the trajectory the object has *zero velocity*. Since we know that energy is conserved, all of the kinetic energy must have been converted to gravitational potential energy, allowing us to figure out the relationship between the velocity with which we threw the object and the height it eventually reaches! Gravity is still acting, though, so the object continues accelerating downward, until it eventually crashes into the ground.

Question: From the conservation of energy, what is the relationship between the initial velocity and the maximum height? What can you say about the velocity of the object in the instant before it hits the ground?

2.4.3 Doing work

So, what is involved doing work on an object? We can do work by *applying a force to an object as it moves in the direction of that force.* For instance, we throw a baseball by using our muscles to apply a force to the baseball over a certain distance (say, somewhat behind your body to the reach of our outstretched arm), ultimately converting chemical energy stored in how our body breaks down the food we consume into a combination of body heat and the baseball's kinetic energy.

The definition of the *work you do on an object* is "work equals force times distance traversed while applying the force[13],"

$$W = \vec{F} \cdot \vec{d}.$$

Work, like energy is a scalar physical quantity, and it depends on the dot-product of two vector quantities, a force and a distance over which that force was applied[14]. That dot product means that if you're applying a force but the object doesn't move in exactly the same direction, the work done on the object only depends on *the component of the force in the direction of motion.* If you're pushing and pushing and that massive boulder isn't moving? You're not actually doing work *on the boulder*! It sure feels like you're "working" in the everyday sense of the word, and you're certainly converting energy from one form to another (food energy into heat), but you're not doing work *on* the rock.

Finally, remember that forces come in equal and opposite pairs – this is a key to realizing why energy is ultimately conserved! When you lift an object up, you apply a force in the upward direction, and the object moves in the upward direction, so you're doing positive work on the object. At the same time, the object is pushing *downward* on your hand as your hand moves *upward*, so the object actually does *the same amount of negative work* on you! You do positive work, it does negative work, and the total amount of energy in the universe is the same.

[13]This assumes you're applying a constant force over the distance involved; if you apply a varying force we need to bring in calculus to get the right expression, so we'll mostly ignore that situation

[14]Remember that $\vec{a} \cdot \vec{b} = |a||b| \cos\theta$: the dot product is the product of the magnitude of the two vectors times the cosine of the angle between them

2.4.4 Using a ramp!

We can finally explain, using the concepts of energy and work, why using a ramp makes it "easier" to move an object up! Let's suppose we're trying to bring a new couch up to the second story of a building: the weight of, say, a 50 $\mathtt{kg}$ couch is 490 $\mathtt{N}$ (weight being mg, and we want to change it's elevation by 3 $\mathtt{m}$.

One option would be to attach a set of ropes and haul it directly up in the air: how much work do we have to do on the couch in this scenario? We would need to supply a force pointing straight up that is *just* enough to be more than the weight of the object; from our formulas we could say

$$W = \vec{F} \cdot \vec{d} = 490 \ \mathtt{N} \times 3 \ \mathtt{m} = 1470 \ \mathtt{J}.$$

Note that this is the same as the change in the couch's gravitational potential energy!

What if, instead, we use a ramp? Let's suppose we're using a ramp whose angle, is $\theta = 5.7$ degrees (or about 0.1 radians) – which is like a ramp which goes 10 units of distance over for every one unit of distance it goes up. Now, to push the couch (and, again, neglecting friction), we only need to supply a force of $mg \sin\theta \approx 49$ $\mathtt{N}$, but we need to do so over 10 times the distance. Consulting our formula, we see that in this case we've done

$$W = \vec{F} \cdot \vec{d} = 49 \ \mathtt{N} \times 30 \ \mathtt{m} = 1470 \ \mathtt{J},$$

exactly the same amount of work!

In the absence of friction, the amount of work you need to do doesn't depend on the *way* you lift the couch up to the second floor; all that matters is the total change in gravitational potential energy of the couch you're trying to make. You could use a ramp, you could use a pulley, you could chop the couch up and throw the pieces to the second floor – the amount of work in lifting the couch is the same, even if the *forces* you need to apply in the different cases are very different.

Pause and reflect! Think about that statement in terms of the conservation of energy! We are trying to change the elevation of a massive object – the couch, and when we do so, the couch gains gravitational potential energy. We can turn some of that potential energy into kinetic energy by dropping the couch out the window onto the ground below, and the speed with which the couch falls – which then tells us how much kinetic energy it acquires while falling – doesn't depend on how we brought the couch up to the second floor! We don't need to know all of the details of every object's "history" to know their physical motions, because energy is conserved!

Pause and reflect! At the beginning of this section I asked "Why do we move faster when biking down a steep hill than a less-steep hill?" But wait a second: If I have a short steep hill or a long gradual hill with the same total change in elevation, what's the important difference? Isn't the change in the gravitational potential energy of a cyclist at the top and bottom of the two hills the same, and so shouldn't they gain the same amount of kinetic energy? What aspect of the physical world have we left out of our idealized set up.

Chapter 3

The laws of motion, Part 2!

In this chapter, we'll continue our exploration of the basic laws of motion, but we'll start making things a bit more complicated. We'll see how to think about the motion of spinning and rotating objects, we'll think a bit more about what friction is and where it comes from, and we'll introduce two more conserved quantities related to motion.

3.1 Seesaws, levers, and mobiles

In the last chapter we learned a bit about ramps, and how they made things easier to lift up into the air: they didn't let us get away with doing any less *work*, but they let us apply a small force over a large distance instead of applying a large force over a short distance. In this section we'll learn about a different device that pulls the same trick: the seesaw (or, more technically, the lever).

3.1.1 Balancing a seesaw

If you've ever played on a seesaw, you probably have a good intuition for what it's like to balance a seesaw with a friend. As depicted in Fig. 3.1, when you both weight the same, everything is pretty easy: you sit in the seats at the end of the seesaw, and the balance is good. When you weigh different amounts it's a little bit trickier, but you've probably figured out that if the heavier person scoots closer to the pivot – or if the lighter person scoots all the way to the very edge and leans backwards – you can still get the seesaw to balance. But if you and your friend weigh different amounts and you sit at the same distance from the pivot, things are not so fun. The seesaw *rotates* about the pivot point as the seesaw with the heavier end crashes to the ground, leaving the lighter person stranded high above the ground until the heavier person deigns to let them down.

In the last chapter we learned about "translational motion," the motion of the center of mass of an object through space, and the seesaw is providing an example of "rotational motion," the motion of an object around a fixed point. In the case of the seesaw, the pivot point is exerting forces on the board so that the seesaw as a whole doesn't wander around the playground, but the board can rotate. Of course, it is possible for an object to to have both translational and rotational motion at the same time. One example would be throwing

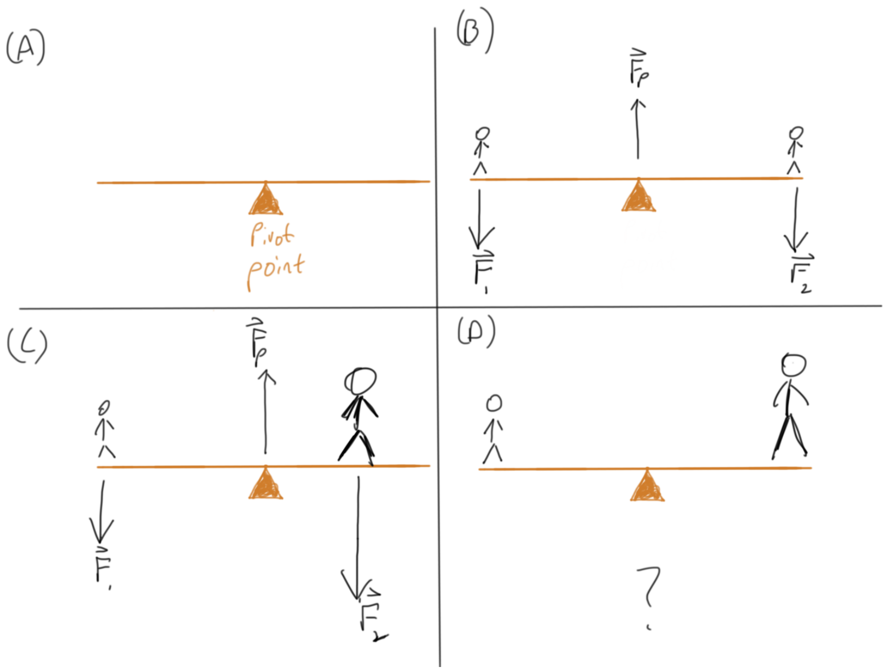

Figure 3.1: **Balancing a seesaw** (A) A seesaw with its pivot exactly in the center. (B) Two people of equal mass stand equally far from the pivot. The pivot is *not accelerating*, so we know the upward force from the pivot exactly balances the downward forces from the people's weights. (C) Two people of unequal weight can still balance a seesaw by standing at different distances from the pivot. (D) When two people of unequal weight stand at the same distance from the pivot what happens? The seesaw starts to rotate!

Figure 3.2: **Defining angular position** (Left) We pick a particular seat at a particular time as a point of reference for our angles. (Right) after some time, that seat has rotated by θ, relative to our point of reference.

a ball with spin, as when a baseball pitcher throws a curveball: the ball is moving through space from the pitcher's mound towards the batter, and at the same time the ball is spinning about its center of mass. Here's a link to a video of another example: as the jugglers through chainsaws back and forth, the *centers of mass* of the chainsaws follow a trajectory much like that of the thrown tomato in the previous chapter, but at the same time the chainsaws are rotating about their centers of mass, too.

3.1.2 Newton's Laws of rotational motion

Once we describe it in the right language, it turns out that rotational motion follows *the same laws* as translation motion does – how convenient that Nature is organized so simply for us! To set up this language, we first talk about an object's *angular position*. Just as a regular position required us to specify a point of reference, a direction, and a magnitude, so does angular position. Let's consider Fig. 3.2, where we will try to describe the angular position of a big Ferris wheel. We can specify a point of reference for the angle arbitrarily: we pick one of the seats on the edge of the wheel at some point in time, and use that to define a reference orientation. After the wheel has rotated a bit, we look at the new position of that same seat, and we can write down the *magnitude* of angular motion: this could be measured in *degrees* (out of 360°) or in *radians* (out of 2π) that the wheel had to rotate to go from one position to the other.

What about the *direction* of rotation, though? Did the Ferris wheel in Fig. 3.2 rotate clockwise or counterclockwise? It depends on which direction we're looking at it from! We've identified the point about which the ferris wheel is rotating, but there is still this ambiguity; we get around this ambiguity by picking a particular convention and always sticking to it. We define the *axis of rotation* as the line in space about which the Ferris wheel is rotating (in Fig. 3.2, the axis is a line that looks like it's going straight into the computer screen, since we see a face-on view of the Ferris wheel), and then we use the *right hand rule* to determine the direction of rotation. This convention says: define a positive direction of rotation if the axis of rotation goes from our eye towards the point of rotation and if the object appears to be moving clockwise[1]. We'll use this convention, but the important thing isn't *what* the convention is, it's why we need a convention in the first place! We need to be able to specify the direction of rotational motion, and we need to have a way to remove any ambiguity about rotational motion that comes from the way we happen to be looking at a rotating object.

Now that we have angular positions, we can define angular velocities by the change in the angular position with time and angular accelerations by the change in the angular velocity with time. That's all we need to make a complete analogy between the laws of motion in the last chapter for translation motion and the laws of motion for rotational motion.

First, just as *inertia* was a fact of how objects in the universe behave, so is *rotational inertia*: an object that's rotating tends to keep rotating, and an objects that's not rotating tends to remain not rotating.

What about Newton's 1st Law, which said that objects with no forces wouldn't experience a change in velocity? Well, here we have

> A rigid object that is not wobbling and is not subject to any outside torques rotates at a constant angular velocity.

Constant angular velocity means the object will turn through the same angle about the same axis of rotation in every set amount of time. *Torques* here are the equivalent of *forces* for rotational motion: when you try to spin an object you are applying a force on an object away from its center of mass to try to get it to spin, like when you're unscrewing a jar of peanut butter (or when you're an octopus trying to escape from inside a jar!. There are some caveats in the above law you may have noticed – objects that aren't rigid (i.e., that can change their shape) and objects that are wobbling have more complicated motion – we'll talk about these complications later in the class!

So, how do objects respond to torques? The analog of Newton's 2nd Law says

> The angular acceleration of a rigid object that is not wobbling multiplied by its rotational mass is equal to the *net torque* exerted on that object.

We've got the same caveats as above, and we've also introduced the idea of the *rotational mass* of an object. Just as mass is a measure of how hard it is to change an object's velocity, rotational mass is a measure of how hard it is to change an object's angular velocity. The rotational mass of an object, often called the moment of inertia, depends both on the mass of the object and the way that mass is distributed throughout the object. It also depends on the axis about which you're trying to rotate the object: you may have noticed that it is easier to spin a tennis racket around its handle than it is to try to flip it end-over-end. Written as an equation, this rotational version of the second law looks like

$$\vec{\tau}_{net} = I\vec{\alpha}.$$

It's conventional to use the greek letter τ for torques and α for accelerations. Notice that from this equation, the acceleration of an object points in the same direction as the torque. Since we have a sense for what directions of angular accelerations are (clockwise or counterclockwise

[1] It's called the right-hand rule because if you point your thumb in the direction of the axis of rotation, your fingers will curl in the direction of motion.

rotation about a particular axis of rotation), and we know what it's like to exert a force on an object, can you anticipate the relationship between forces and torques that we'll see in the next subsection?

While you're thinking about that, it's worthwhile to briefly comment on the units and dimensions of quantities related to rotational motion. Confusion sometimes arises because the *radian*[2] is said to be a "natural unit" that doesn't carry dimensions with it: $[\theta] = 1$. Angular velocity (often denoted by $\vec{\omega}$ has units of radians per second, so $[\vec{\omega}] = \mathbb{T}^{-1}$. Angular acceleration, then, has units of radians per second squared, and $[\vec{\alpha}] = \mathbb{T}^{-2}$. Since rotational mass depends on both mass and how mass is positioned within an object, it turns out to have units of $\texttt{kg m}^2$ and hence dimensions of $[I] = \mathbb{M}\mathbb{L}^2$. By looking at the equation above, you should be able to deduce from all of this that torques have units of "Newton-meters," and that $[\vec{\tau}] = \mathbb{M}\mathbb{L}^2\mathbb{T}^{-2}$. You'll notice, I hope, that this is *not* the same dimensions as forces have!

3.1.3 Torques

Alright, we've danced around what a torque actually *is* in the last little bit, so how do we actually change an object's angular acceleration? What do we do to get an object to rotate? Let's go back to the seesaw, now shown in Fig. 3.3. If you've ever stood on a seesaw immediately above the pivot, you know that doing so doesn't cause the seesaw to tilt (or, more likely, you discovered that shifting your weight even slightly from one leg to the other as you straddle the pivot point causes the seesaw to tilt back and forth. You also know that if you just grab the board and pull directly away from the pivot point, the seesaw won't rotate. But if you stand at one end of a seesaw and there's nothing to balance it out on the other side of the pivot point, the seesaw will certainly start rotating: you've used the force due to gravity to exert a torque on the seesaw and caused the seesaw to experience an angular acceleration.

How can we combine all of these observations of how we intuitively know torques behave into a simple mathematical language? First, we define the *lever arm*, which is a vector quantity that goes from the pivot to the location that a force acts on an object. You may have noticed – say, if you have ever assembled Ikea furniture in your dorm room – that the longer this lever arm is the greater amount of torque you can supply. It's also the case that the more *force* you apply the more torque you apply – is that jar hard to open? Apply a greater force!

Combining everything above the relationship between torque and force can be obtained by saying that the torque on an object is equal to the lever arm times the amount of force perpendicular to the lever arm. As an equation we would write

$$\vec{\tau} = \vec{r} \times \vec{F},$$

where here the $\times$ symbol is standing for the "cross product" (as opposed to the "dot product" we saw in the previous chapter. If you remember the definition of the cross product from your math classes, that's great! If not, all we'll need in this class are the following two things. First, you can get the direction of the torque by using the right-hand rule: point your index

[2] 1 complete rotation about an axis is composed of 2π radians

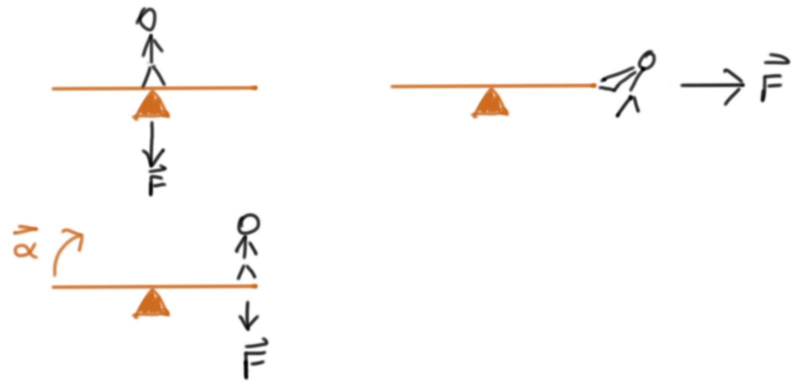

Figure 3.3: **Trying to rotate a seesaw** (Top row) Two ways of failing to rotate a seesaw: standing directly over the pivot point, or pulling in a direction directly away from it. (Bottom) A successful approach: exerting a force *at an angle to* a line from the pivot point to where the force is exerted gets the seesaw to start rotating.

finger in the direction of the lever arm (from the pivot to the force), bend your middle finger in the direction of the force, and your thumb will point in the direction of the torque. Second, the magnitude of the torque is given by

$$|\vec{\tau}| = |\vec{r}||\vec{F}|\sin\theta,$$

that is, the magnitude of torque is equal to the magnitude of the lever arm time the magnitude of the force times the sine of the angle between $\vec{r}$ and $\vec{F}$.

Finally, of course, there is the equivalent of Newton's 3rd Law:

> For every torque that one objects exerts on a second object, the second objects exerts a torque which is equal in magnitude but opposite in direction on the first object.

Of course, the consequences of these torques can be very different. A person stands on a seesaw and exerts a torque, causing the seesaw to rotate. The seesaw exerts a torque back on the person, but does the person rotate about their center of mass because of it? What is the direction of the lever arm for the person.

Pause and reflect! In all of the above pictures, I drew the seesaws as initially horizontal, and initially motionless... When would that matter?

Pause and reflect! From the discussion of forces and torques above, do you see why levers give a "mechanical advantage" – letting us trade off force and distance – in the same way that ramps do? How does a claw hammer work to pull out nails? How do Calder's kinetic sculptures work, and what can you tell about the masses of the different parts of the mobile just from the picture?

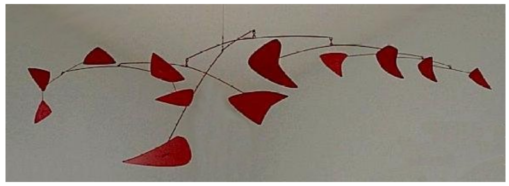

Figure 3.4: ***Red Mobile*,** Alexander Calder, 1956. Photograph from the Montreal Museum of Fine Arts.

3.2 Wheels

We've seen two of the canonical six[3] "simple machines" already – the ramp and the lever – that exploit the laws of motion to manipulate forces and make our lives easier. In this section, we'll cover the wheel! One of the main things wheels help us do is *reduce friction* that an object encounters as it moves; here's an experiment suggested by the textbook: First, grab a book and push it around on a table. How quickly does it stop after you stop pushing it? How much force do you need to exert to get it to move? Does the answer to that depend on how fast you push? Next, lay down a few parallel pencils (or pens, if they're that simple cylindrical shape), put the book on top, and roll the book forwards and backwards on top of the pencils. How do your answers to the above questions change?

3.2.1 Friction

What wheels really help us do is reduce friction, but what is friction? Friction is a force that *opposes the relative motion of two objects in contact with each other*. We know that an object accelerates according to the net force on an object, so why is it we can push on a heavy object on flat ground and not have it move? It's because the interaction between the object and the surface it's resting on exert a force *in the direction that opposes* the motion you are trying to impart. If two surfaces (of an object and the ground, or two objects pressed against each other) are moving relative to each other, friction acts on both objects in directions that will tend to bring the two surfaces to the same mutual velocity.

You may be wondering, by the way: if an object is flat and on flat ground, with gravity and the normal force pointing vertically, how can friction point *horizontally*, to oppose our efforts to push the object around? The answer, remarkably, comes from the *microscopic* shape of the interfaces, which can never be totally smooth and flat. We know some of the details of how this works, but not all of them, and understanding the precise details of how this works is actually *still* a topic of active research! So there's another example of how asking a simple question – "Why do I need to push harder to get some surfaces to move across each other?" – brings you to the edge of current scientific research!

You may also be wondering, at this point, *where is the energy going?* Energy is conserved, and when I do *work* to slide a bookcase across the ground, I see that the bookcase is not gaining kinetic energy (after all, it stops moving basically as soon as I stop pushing it). Where did the energy go? It mostly went into *heat*! When you slide two objects past each

[3]As identified by scientists in the Renaissance. Note, by the way, that the word "scientist" didn't appear until the 1830s.

other you heat them both up, and thermal energy (which is essentially the random wiggling and jiggling of the microscopic particles that make up matter) is another form that energy can take! Some of the energy also goes into deforming the material, maybe breaking some of the chemical bonds that kept atoms and molecules in specific places relative to each other.

Frictional forces always oppose relative motion of two surfaces, and their strength depends primarily on three factors: how tightly the surfaces are pushed against each other, how "slippery" the surfaces are, and whether the surfaces are actually moving relative to each other. You've probably encountered all of these features before! If you try to push an empty bookcase across a floor its easier than pushing a full bookcase – the empty bookcase has less total mass, so it pushes down against the floor with less force. Pushing a person standing on ice is a lot easier than pushing the same person standing on asphalt – even when they're wearing the same shoes, the ice is more slippery. And maybe you've noticed when trying to push something that it feels harder to get something moving in the first place than it does to keep it moving.

That's because there are two types of friction, *static friction* and *dynamic friction* (also. called "sliding friction"). Dynamic friction acts when two surfaces are already moving relative to each other, and static friction is what we need to consider when two surfaces at currently moving at the same velocity, and a force is trying to accelerate one relative to the other. These two types of friction are sometimes talked about simultaneously by talking about an objects *traction* – the largest amount of friction force an object can obtain from a surface it's in contact with. Traction, of course, is neither good nor bad: the traction of your bookcase makes it hard to move around, but it also keeps it in its place once you've moved it into position (even if, say, you live in a house where the floors aren't perfectly level!). Similarly, when you are pushing the bookcase around in the first place, your feet against the floor provide *you* the traction you need in order to exert a horizontal force on the bookcase – hence why we don't usually push heavy objects around while only wearing socks.

3.2.2 Wheels

So, how does rolling help? Make a fist with one hand, and press it against your other open palm, and try to drag your fist across your palm. As your fist slides, it is experiencing sliding friction, and you will feel your skin heating up (also the principle behind rubbing your hands together to keep them warm in the winter). If you don't force your fist to drag, though, you can press just as hard but allow your fist to *roll*. Your fist still experiences static friction – the static friction provides a *torque*! – but now it rotates, and you do not feel the same frictional heating. Rollers act in the same way. An object resting on rollers can be pushed easily along the ground, with static friction exerting a torque on the rollers to make them rotate. Rollers, though, are kind of annoying to use: as the object traveling across them moves forward it leaves the roller behind, so you're always racing to put new rollers in front of the object. Better is to introduce a *wheel and axel*, as in Fig. 3.5. As, say, a cart and buggy on wheels rolls along, the cart moves forward with the wheels turning so that the point where the wheel and ground meet doesn't slide – so it loses almost no energy to frictional heating! In a simple wheel and axel set-up, the wheel does have to slide against the axel (causing both heat and wear-and-tear), which is why axels are almost always lubricated somehow to reduce the sliding friction there. Another trick we use all the time is to add *rolling bearings*,

Figure 3.5: **Wheels!** (Left) A schematic of a wheel and axel. (Right) To reduce friction, we put rolling bearings in between the wheel's hub and the axel

for instance by making more complicated hubs where the wheel and axel meet. Now now the wheel is rolling against the ground *and* against the ball/roller bearings separating it from the axel; with good bearings you can reduce the overwhelming majority of sliding friction.

3.2.3 Rotational motion, work, and energy

With all of this rotating and spinning, we need a way to think about the energy associated with these more complicated motions. We saw in the last chapter that *translational kinetic energy* could be expressed as

$$K_T = \frac{1}{2}mv^2,$$

half of the mass times the square of the speed of a moving object[4]. Well, rotational kinetic energy can be written so that it looks the same:

$$K_R = \frac{1}{2}I\omega^2,$$

half of the *angular mass* times the square of the *angular speed*. The total kinetic energy of an object is the sum of both the translational and rotational pieces. For instance, when you're cruising on your bicycle, the total kinetic energy is the *translational kinetic energy* of you and your bicycle moving at some velocity plus the *rotational kinetic energy* of the wheels spinning! When you do work to get your bicycle moving, you have to do work to get both of these quantities going.

Speaking of work, we saw that for translational motion, work was a force times a distance. For rotational work, its a *torque* times an *angle* over which that torque is exerted:

$$W = \vec{\tau} \cdot \vec{\theta}.$$

Just like how exerting a force on an object that doesn't move means you aren't doing work on that object, if you exert a torque but the object doesn't rotate you aren't doing work on that object.

[4]Note: the square of a vector is $v^2 = \vec{v} \cdot \vec{v}$, which is the same as the square of the magnitude of the vector. Hence, "velocity squared" = "speed squared."

Finally, we've already talked about work but not *power*. Power is a physical quantity describing work done in an amount of time:

$$P = W/t,$$

and the SI units of power are *Joules per second*, `J/s`, also known as *watts*, `W`. You've seen this unit on lightbulbs and appliances, but if you like cars[5] you might be more familiar with a "horsepower[6]," defined to be 745.7 `W`. Let's get some order-of-magnitude feel for these units. A liter of water has a mass of about one kilogram and hence a weight of ~ 10 `N`. So, a Joule is the energy used in lifting a liter of water about 10 centimeters up (equivalently, if you prefer imperial units, not that different from lifting a quart of water about 4 inches up), so a watt is the power produced to lift a liter of water 10 centimeters up in one second. Automotively, your Ferrari F8 Spider has an engine that can output up to 710 `horsepower` = 530000 `W` and the car has a mass of about 1400 `kg` – add in your own mass and some luggage and let's say it's moving a weight of 15000 `N`. Apparently, then, the car's engine is powerful enough to lift it upwards by 35 meters every second. No problem tackling even very steep hills, I guess!

3.3 Bumper cars

In this section we'll talk about "bumper cars," or really any sets of colliding objects. Why is it that a stationary object moves after a moving object collides with it? When two bumper cars crash together, what aspects of "motion" are transferred between them, and why? What governs the behavior of billiard balls as they bounce off of each other and off the rails? We'll start learning about *linear momentum* and *angular momentum* in this section, two more conserved physical quantities that govern the motion of objects!

3.3.1 Linear momentum

We already know that *energy is conserved*, and when two objects collide they typically exchange some of this *scalar* quantity: they might trade some of their kinetic energy, or deform each other a bit, or heat each other up, all in such a way that the total amount of energy is the same before and after the collision. As a scalar quantity, energy doesn't have a direction. But there is also something *directional* about how collisions work, but you probably already have a sense that when two things collide they exchange some property of their motion that depends on the directions involved: if you're stopped at a red light and hits your car from behind, your car does lurch backwards, or to the left, or to the right – it lurches forwards.

The quantity that's being exchanged in these kinds of collisions is *linear momentum*, which we'll often just call *momentum* – a conserved *vector* quantity associated with the translational motion of moving objects. As we discussed in class, you probably have a sense that the momentum an object has depends on both its mass and its velocity: it's harder to

[5]I do not. Have you see the traffic in Atlanta?

[6]A term more or less made up by James Watt in the 1700s so that he could compare the output of his steam engines to a horse, which were the... workhorses of the day. Note that this figurative sense of "workhorse" is only attested as far back as 1908 by the OED

stop a car than a bicycle moving at the same speed, and it's harder to stop a fast-moving bicycle than a slow-moving one! Written as an equation, the momentum of an object, $\vec{p}$ is

$$\vec{p} = m\vec{v}.$$

So, the momentum of an object points in the same direction as the velocity, and its magnitude is the object's mass times the magnitude of velocity.

Quick question: We'll meet angular momentum in just a little bit, but based on what we've said about the relationship between translational and rotational motion, what do you think the formula for angular momentum will look like?

Here's a classic "intro physics" puzzle: Suppose you're stranded in the center of a frozen lake, with perfectly slippery (i.e., frictionless) ice beneath your feet. Since the ice is frictionless, you have zero traction – if you try to take a step you'll just fall flat on your face. So, how do you get safely to the shore? The answer[7] to this is illustrated in Fig. 3.6.

3.3.2 Angular momentum

In addition to collisions involving linear momentum, you've probably seen collisions that send two objects spinning away from each other. Often, when objects collide at a glancing angle – rather than directly head on – they end up spinning as well as translating as they move away from each other. Well, just like linear momentum is a conserved quantity associated with translating through space, *angular momentum* is another vector conserved quantity associated with objects rotating about some pivot point (often the center of mass of that object). From everything we've said above, you probably aren't surprised to learn that angular momentum, $\vec{L}$, is a product of the angular versions of mass and velocity:

$$\vec{L} = I\vec{\omega}.$$

So, $\vec{L}$ is another conserved quantity: just like energy and linear momentum, it can't be created or destroyed, you can only transfer it between objects. And just like linear momentum, there's a *direction* associated with angular momentum: it points in the same direction as the angular velocity $\vec{\omega}$.

A classical illustration of angular momentum is the way figure skaters perform very fast spins on the ice. At first they start rotating with their limbs extended, setting up a spin with a certain amount of angular momentum, $\vec{L}$. They they bring their arms and legs very close to the axis around which they're rotating: what this does is *reduce their moment of inertia*[8]. But, by the conservation of angular momentum, $\vec{L}$ has to stay the same! Since I went down, that means $\vec{\omega}$, the skater's angular velocity, has to increase!

[7] Answer: you take off a shoe (or a hat, or whatever) and throw it horizontally! Momentum is conserved, and at first the total momentum is zero – nothing is moving! You then do some work to throw your shoe, so it has momentum $m_{shoe}\vec{v}_{shoe}$. Momentum must be conserved, though, which means: $m_{shoe}\vec{v}_{shoe}+m_{you}\vec{v}_{you} = 0$. Rearranging this equation, you find yourself moving at $\vec{v}_{you} = -\frac{m_{shoe}}{m_{you}}\vec{v}_{shoe}$. That is, you move in the opposite direction, with a speed which is less than that of the shoe by the ratio of the masses

[8] Remember: I depends both the the skater's mass and the way that mass is positioned relative to the axis of rotation!

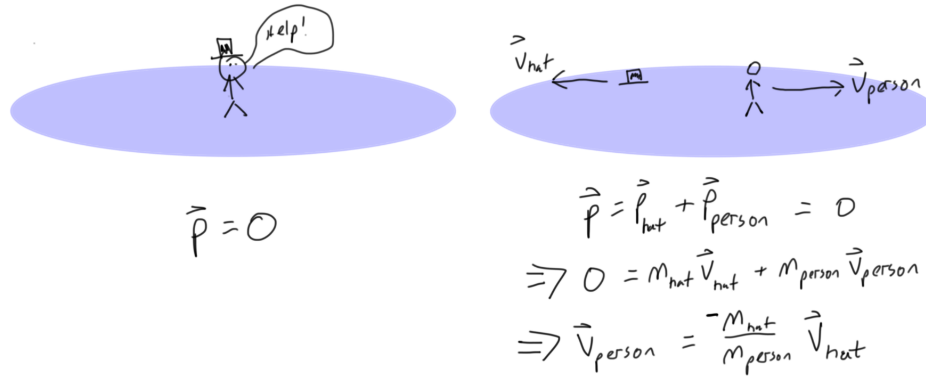

Figure 3.6: **Escaping from an icy lake** (Left) A person stands, despondent, at the center of a frozen lake with no friction. With sufficient traction to exert a horizontal force to start walking, the person feels doomed. (Right) Remembering that *momentum is conserved*, the person throws a top hat to the left, and by conservation of momentum, starts drifting very slowly to the right. If only our hero had brought along a heavier object.

3.3.3 Impulses

Okay, so both the momentum and angular momentum of a system is conserved, but we see the velocities of objects changing all of the time – how is momentum transferred and exchanged between objects? Momentum is transferred by *impulses*; an impulse is a *force exerted on an object for a duration of time.* The notation for impulses imagines them as changes in momentum (and the symbol we most often use for "change of a quantity" is Δ), hence:

$$\Delta\vec{p} = t \cdot \vec{F}.$$

So, an impulse is a force times a time: the longer and harder you push on something the more momentum it will have. Just like we can do equivalent amounts of work by either doing large or small forces over small or large distances (respectively), you can see from the equation about that you can get the *same impulse* by applying large forces for a short amount of time, or small forces for a long amount of time.

During the collision of two objects, the objects exchange impulses, and they do so *for the duration of the collision.* The fact that the duration of the collision matters explains why cars have bumpers and crumple zones, or why sports players wear padded equipment. These soft materials allow a collision to take place over *longer intervals of time* and so the impulse – the exchange of momentum between the objects, can involve smaller forces (since the same total change of momentum happens over a longer time)!

By thinking about impulses, together with Newton's 3rd law, you can understand why momentum is conserved: Objects in a collision exert equal but opposite forces over the same duration! Thus, they receive impulses that are equal in magnitude but opposite in direction, too: the momentum gained by one car in a collision must be equal to the momentum lost by another.

Pause and reflect: You're very mildly annoyed by your friend, and you want to throw a lightweight object at them. Within reach you can choose between a beanbag (which will hit your friend and then come to a stop, dropping straight down (due to gravity) after the collision) and a bouncy ball (which will hit your friend and then rebound (bouncing back towards you a bit)). Assuming the beanbag and the bouncy ball have the same mass and you throw them at the same velocity, which will transfer more momentum to your friend during the collision?

Okay, what about transferring angular momentum? We'll we've been setting up this language where we already know the answer: just as impulses are forces times time, angular impulses are torques times time:

$$\Delta\vec{L} = t \cdot \vec{\tau}.$$

The more torque you exert on an object for a longer time, the more you change its angular momentum. This explains why glancing collisions impart rotational motion: the force of the collision is off center, so there is a torque, $\vec{\tau} = \vec{r} \times \vec{F}$ involved for the duration of the collision!

We've already alluded to one of the subtle differences between angular momentum and linear momentum, at least as it relates to angular and translational velocity. A car moving at some velocity has a momentum, and since momentum is conserved and the car can't change it's mass on its own, the car keeps moving at the same velocity (until it collides with something, or you apply the brakes, or friction takes over...). On the other hand, objects that aren't rigid – i.e., objects that can change their shape – can change their angular mass! We saw that with the example of the figure skater. So, it's important to keep in mind this distinction between angular *momentum*, which is conserved, and angular *velocity*, which is not!

That covers it for now – we'll learn more about collisions in the next chapter as we learn about springs, bouncing balls, and roller-coasters!

Chapter 4

Mechanical objects, Part 1!

In the last chapters we've talked about the laws of motion and introduced three very important conserved quantities: energy, linear momentum, and angular momentum. These concepts will be crucial tools as we start understanding the behavior of the world around us, and in this chapter we'll introduce a few more crucial concepts. We'll spend some time talking about springs, we'll revisit the ideas of colliding objects, and we'll talk more about acceleration, and why roller coasters are fun.

4.1 Springs and scales

We've talked a bunch about the mass of various objects, but how do we *measure* mass? We've been measuring amounts of stuff for a very long time, but the specific concept of mass has only been around in its modern form since the mid-to-late 1600s. Much easier is to measure the *weight* of an object, which we know is a measure of the force due to gravity acting on an object.

4.1.1 Measuring weight

So, how do we measure weight? Fig. 4.1 shows two methods. On the left is a classic set of balance scales, and thanks to the last chapter we understand exactly how they work! Balance scales are basically seesaws, with a beam of material pivoting about its center that has two baskets of the same mass that hang equal distances away from the pivot. The *torques* exerted by these hanging baskets are equal, and so the beam does not have any *angular acceleration*. What happens when we put something in one of the baskets? The torques are no longer balanced, and the beam starts to turn; to measure the mass of unknown objects, you keep a collection of *carefully calibrated* measures: objects whose weight you trust have been correctly measured by someone else, and you add these to the other basket until the scales are balanced again. At this point you know the measures you added have the same total weight as the unknown object, since the torques are again balanced.

Much more common, though, are scales based on *stretching or compressing springs*, shown on the right of Fig. 4.1. The basic principle is that stretching or compressing a spring causes the spring to exert a *restoring force*. Every spring has a length that it will be when it is

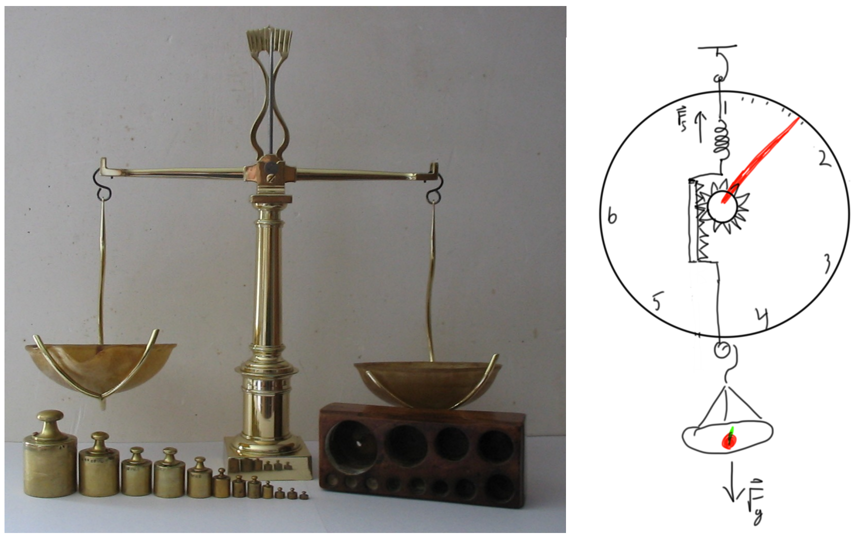

Figure 4.1: **Measuring weight** (Left) A balance scale, whose working principle involves keeping a set of fixed measures handy and exerting equal torques on a beam. (Right) A cartoonishly drawn hanging spring scale. When an object is placed in the hanging basket, it exerts a force downward on the basket. This force *stretches a spring* until the spring eventually comes to a new length where it exerts an equal force upward. The scale is calibrated so that the amount the spring stretches gets converted (via gears) to the force the spring is exerting. The astute reader will notice that in this diagram I've drawn the gears in such a way that the dial would spin the wrong way!

relaxed and no forces are acting on it – this is called the spring's *equilibrium length* (or "rest length"). The force that a spring exerts is called a restoring force because it points in a direction to try to "restore" the spring to its equilibrium length: pull on a spring to try to stretch it out and it pulls back; push on a spring to try to compress it and it pushes back!

An important point about using a spring scale, though, is that it doesn't *directly measure weight!* The weight of an object is the force gravity exerts on it, but a scale is measuring how much force *the scale exerts on the object.* We know from Newton's second law that if the object is not accelerating and the scale is level this distinction doesn't matter: gravity exerts a force straight down on the object, the object pushes against the scale with this weight, and since the object isn't accelerating the scale must be exerting the same magnitude of force (in the upwards direction).

If you are accelerating, though, you'd better not trust your scale! For instance, if you stand on a bathroom scale and jump up and down, you'll see the scale report numbers that bounce all over the place: as you jump and land you're accelerating upwards, which changes the net force acting on the scale; since the scale isn't accelerating, that means the force the scale is exerting on you is similarly changing.

And, of course, since scales are really measuring *forces* and not *masses* of objects, their reports depend on where you do them. In a previous homework we saw that the acceleration due to gravity is smaller on the moon than on the Earth; a spring scale on the moon would report your weight as about a sixth of what it is on the Earth. Even different places on the Earth will give slightly different readings: the Earth is not really a sphere – it bulges out in the middle – and the acceleration due to gravity at the equator is about 0.995 times the acceleration due to gravity at the poles[1].

4.1.2 Stretching and compressing a spring

So, that's the principle behind spring scales... but you know, part of the ethos of this class is to think about the world with numbers attached! So, first, try to play around yourself and see what you think the relationship between stretching or compressing a spring is and the restoring force the spring exerts back on you. You could surreptitiously do this using a hanging scale at a grocery store, or by deconstructing a pen to get the small spring inside... springs are cheap and all around us!

Let's start getting some language and notation to talk about the springs easily. We've already talked about a spring's equilibrium length – so called because when a spring is at this rest length it exerts no forces on either of it's ends, and we say that something is in *mechanical equilibrium* if it is experiencing no net force. So: if you leave a spring alone, it comes to some natural length. We'll often want to talk about the distortion of a spring *away* from it's equilibrium rest length. Often one end of a spring is fixed in place, so it is natural to describe the distortion by talking about the *position of the non-fixed end*, which is a vector quantity. Even if both ends of a spring can move, though, the overall distortion

[1] A great reason to take your honeymoon in Hawaii, for instance: by being slightly closer to the equator than the mainland US, when you first step on a scale you'll think you've lost a little weight – a great excuse to order an extra serving of delicious poke! Here's a local gravity calculator if you're curious to get a sense of how much the acceleration due to gravity varies across different locations on the Earth.

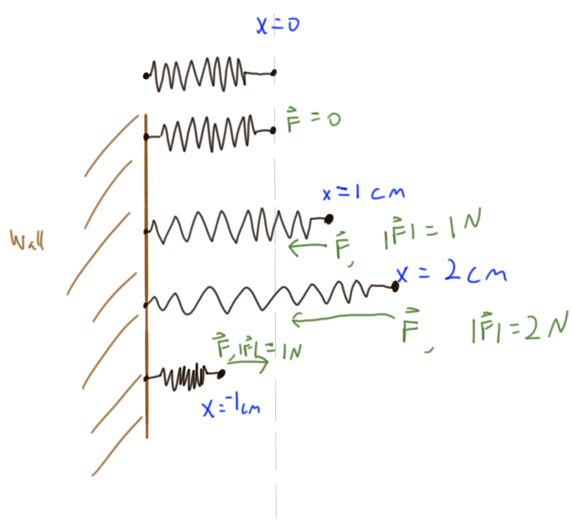

Figure 4.2: **Springs and restoring forces** Five identical (if poorly drawn) springs. At the top, a spring not attached to anything, doing nothing. Below, when attached to the wall but at its equilibrium (rest) length, the spring exerts no forces. We'll call the position of the right hand end of the spring $x = 0$ at the rest length. Next, we discover that if we pull the end of the spring 1 `cm` to the right, there is a force whose magnitude is 1 `N` *to the left*. Next, we discover that if we pull the end of the spring 2 `cm` to the right, there is a force whose magnitude is 2 `N` *to the left*. Finally, we discover that if we push the end of the spring 1 `cm` to the *left*, there is a force whose magnitude is 1 `N` *to the right*. Apparently these springs have $k = 100\ \mathtt{kg/s}^2$.

of the spring can still be written as a vector (how much is it stretched or compressed, and in which direction?).

What you might have found when you experimented on your own, and what is depicted in Fig. 4.2, is that springs exert restoring forces, and that *those restoring forces are proportional to the deformation of the spring!* That is, the farther you pull a spring, the harder the spring pulls back! In the late 1600s Robert Hooke[2] discovered the precise relationship between spring deformation and force, and it's as simple as can be. Each spring has a "spring constant," k, which characterizes how stiff the spring is[3], and *Hooke's Law* says that the restoring force is

[2]Scientists of this era had an odd relationship with telling the world know about their work. They often wanted to simultaneously *not tell anyone what they had discovered*, but also be able to prove that they were the first to discover something (should someone come along later and try to take credit). This led to all sorts of weird language games – a common one would involve sending a letter to a friend that would contain a random jumble of letters; these letters would be an anagram containing the essence of their discovery, should they need to reveal it. In this case, Hooke's first announced his law for springs in 1676 as "ceiiinosssttuv." Two years later he revealed this to be an anagram of "ut tensio, sic vis" – Latin for "as the extensions, so the force." Hooke, on a different occasion, told the world about the shape of a hanging rope by writing "abcccddeeeeeefggiiiiiiii-illmmmmnnnnnooprrsssttttttuuuuuuuux," which was very enlightening to his readers, I'm sure.

[3]The spring constant depends on things like the material the spring is made out of, how it's shaped, and how long the spring is

equal to negative k times the distortion; as an equation:

$$\vec{F} = -k\vec{x}.$$

This equation is illustrated in Fig. 4.2.

You might be wondering why we're talking so much about springs – I mean, they're cool and all, but are they that important? Secretly, yes! First, it turns out that Hooke's law is surprisingly general, and it doesn't just apply to coiled up springs. It turns out that many objects will respond to a distortion away from their *equilibrium shape* with restoring forces that are proportional to that distortion. If you sit in a comfy chair, the cushions will push up with a force proportional to how far you are squishing it down; if you stretch a rubber band, you need to pull proportionally to the amount you want to stretch it; if you drive across a bridge, the bridge will bend down (very slightly) under that weight while exerting an upward force proportional to the amount it bends.

Of course, there's a limit to this behavior – stretch anything too far and it will break! Even before it breaks, though, most materials eventually show deviations from Hooke's law, so that the restoring force *is no longer linearly proportional to the distortion.* As long as you stay within each object's *elastic limit*, though, Hooke's law holds for all sorts of stuff.

So, Hooke's law covers forces that a spring exerts as it is deformed. How do we deform a spring? We do work! To stretch a spring, for instance, you have to exert a *force* through some *distance*, which is the definition of doing mechanical work. Since energy is conserved, that work must have gone into something, but what? It is being stored in the spring as *elastic energy* (or "elastic potential energy"); once you let go that elastic energy can be released, allowing the spring to do work on something else. Earlier we saw that the equation for gravitational potential energy was $U_g = mgh$ (using the letter U to stand for a type of potential energy); for a spring the *spring potential energy* is[4]

$$U_s = \frac{1}{2}kx^2.$$

This brings us to the second reason springs are so important, which has to do with their (idealized) behavior. Let's imagine a perfect spring scale, where the spring can compress and stretch with no friction. What happens when you step onto such a scale? Before you step onto it the scale is at its equilibrium length, but as you step onto it you are falling (a very short distance, I hope!), and you start to compress the spring. You're converting some gravitational potential energy into elastic potential energy, and by the time you are at the compression of the spring where the forces would balance, you've built up some velocity, which causes you to over-compress the spring. Eventually you come to rest, but now the spring is more compressed than it needs to be to support your weight, so you start accelerating upwards again! Again, by the time you get to the level where the force of the spring balances that of gravity, you've built up some velocity in the upward direction! You keep bouncing up and down forever in this scenario, with the same amount of total energy shuffling between your gravitational potential energy, your kinetic energy, and the spring's elastic potential energy!

[4]Those of you who know calculus might start to notice a connection between the potential energies we are writing down and the forces that correspond to them! The relationship where forces look like derivatives of potentials is... not an accident!

This is called *simple harmonic motion*, and in this example you are a *simple harmonic oscillator*. Harmonic oscillators, it turns out, are everywhere in the world around us: springs, sure, but also pendulums, the strings of your guitar, the electrons powering your computer as you read this, the shock absorbers of your car, the way your ear drum works... Harmonic oscillators are one of the foundational building blocks in how a physicist thinks about the world; we'll see more of them in later chapters!

Oh, and of course: why don't spring scales actually bounce around forever like in the above paragraph? Friction. The process of trading total potential energy and kinetic energy causes the spring itself to move around, and sliding friction is always siphoning energy away and turning it into thermal energy. Eventually you stop accelerating up and down and come to rest, at a place where there are no net forces acting on you: the spring is exerting exactly the same upward force on you as your weight exerts on the scale, and the scale finally tells you how much you weigh.

4.2 Bouncing balls

Let's return to thinking about collisions, in the context of how different balls bounce. In your own life you may have noticed that different objects bounce off the ground very differently – if you need a reminder here's a link to a (riveting) video of someone systematically recording how the balls used in various sports behave when you drop them from a fixed height. Watching the video, or thinking about your own experience, you know that when dropped from the same height you'll observe different subsequent bounces: a golf ball dropped on a hard surface will bounce up and down for a while, whereas a ball of wet clay (or, say, a hacky-sack) will land on the ground with a thud and not bounce at all.

But what's going on? Can a ball ever bounce higher than it is dropped from? When a ball is dropped and bounces several times before coming to rest, how does the maximum height the ball reaches change after every bounce?

You might have anticipated from the discussion of springs that each ball has an equilibrium shape: when no forces are acting on it a tennis ball is (roughly) a sphere, a rugby ball is... rubgy-ball-shaped, and so on. During a collision (with the ground, with a racket, with your foot that shape is *distorted*, and the ball acquires some elastic energy, and that elastic energy can be converted (partially) back into kinetic energy as the ball returns to its equilibrium shape. Here's a great video of a fast tennis serve taken in slow-motion – look at how much the tennis ball deforms as it collides with the racket, and then even oscillates for a little bit as it flies off!

It's sometimes convenient to define the *collision energy* as the kinetic energy absorbed during a collision and the *rebound energy* as the kinetic energy released as the ball rebounds and returns to its equilibrium shape. Since energy is conserved, these energies should be equal, right?

No! The total energy is conserved, but there are other forms of energy to consider. Some of the energy of the collision gets converted into heat. Also, that loud "thwack" of a racket hitting a tennis ball? Sound waves carry some of that collision energy away! The amount of collision energy that gets returned directly into rebound energy is a way of quantifying how well (or poorly) a ball bounces.

An ideal elastic ball would have a perfect $1:1$ ratio of rebound energy to collision energy, but this never occurs with real objects: real objects lose energy to heat, vibrations, sound, and so forth. Such ideal collisions are called *perfectly elastic collision*. The opposite extreme is a *perfectly inelastic collision*, which is one in which after a collision two objects stick together and begin moving at the same velocity together. In this scenario the maximum possible amount of kinetic energy gets converted into other forms of energy – what is the "maximum possible amount"? It's the maximum amount of kinetic energy that can be lost *while still obeying the conservation of momentum*!

Real balls typically experience *inelastic collisions*: some of the collision energy gets "wasted" and the ball bounces back to a different height. For example, a tennis ball will have about 55% of its kinetic energy after a collision (which you can see in the youtube video linked to above: it bounces back up to a little more than half of its original height, which is a reflection of the kinetic energy it has after it finishes colliding with the ground), whereas a baseball will only have about 30% of the collision energy converted into rebound energy.

This ratio of energies is one of the easiest ways to characterize how different balls bounce, but it's traditional to characterize ratios of *speed* instead (because we're better at thinking about speeds rather than energies). This characterization is called an objects *coefficient of restitution*, C_R:

$$C_R = \frac{\text{rebound speed of object}}{\text{collision speed of object}}.$$

Of course, you know that the speed is related to the collision energy in the first place, so it's easy to go back and forth between C_R and the ratio of energies.

What kinds of properties of an object go into determining C_R? What the object is made out of, of course, matters: some substances are more elastic than others. It's also the case that energy storage is more efficient (in terms of how much energy it takes to distort an object compared to the work you can get out of it) when energy is stored in compression rather than bending a surface (it turns out that bending most surfaces generates a fair amount of internal friction, so more collision energy goes into heat).

For instance, a ball of solid rubber will bounce really well, but a tennis ball (made out of a hollow rubber shell with felt on top and pressurized air inside) has a smaller coefficient of restitution: the rubber ball is just compressing an elastic material during a collision, but the rubber shell of the tennis ball bends a lot before it compresses. That's also why most sports balls – which are filled with pressurized air – really depend on being properly inflated! A lot of energy of a collision gets stored in compressing the air inside the ball, and when a ball is "flat" a good chunk of that energy gets wasted by friction as the surface of the ball bends. Go back and watch that slow-motion tennis racket video again, and now compare it to this video of a baseball being struck.

Finally, when two objects are both moving and collide with each other, how does that change how an object bounces? When a baseball hits a bat, what is the "collision speed"? Is it the speed of the ball, the speed of the bat, or something else? By thinking about *frames of reference*, we can see that the speed is the rate at which the two objects are approaching each other (i.e., the difference in their (vector) velocities)!

This might be intuitive: think about driving on a 2-lane, 2-way highway. If you and another car are both traveling at 100 `km/hr` and moving in opposite directions, you are approaching each other at 200 `km/hr`, and any collision would be devastating! If you're going 100 `km/hr` and the other car is going 95 `km/hr` and your traveling in the *same* direction, you're only approaching the other car at a relatively tame 5 `km/hr`, and a collision would be nothing to write home about[5] So, when a baseball bat hits a ball, the bat's velocity is in the opposite direction as the ball's velocity – just like our cars moving in opposite directions, the speeds add and the collision in impressive.

Let's put some numbers into this baseball-and-bat example. Suppose a major-league pitcher throws has thrown a fastball with a speed of 100 `miles/hr`. The batter swings, and just before the bat makes contact it is traveling at 65 `miles/hr`. In the frame of reference in which the bat is stationary, the ball approaches the bat at 165 `miles/hr`; it has a coefficient of restitution[6] of $C_R \approx 0.55$, so after the collision the ball flies away from the bat at $0.55 \times$ 165 `miles/hr` = 90.75 `miles/hr`. From the frame of reference of someone watching this unfold, the ball flies away from the bat at 90.75 miles per hour, but *the bat is still moving forward with some velocity*, so the "exit velocity" of the ball is even faster – it's the sum of the 90.75 we already calculated plus however fast the bat is still traveling! Apparently professionals sometimes reach exit velocities around 120 `miles/hr` – pretty fast!.

Even faster: a baseball breaking the speed of sound. Here's a youtube video of some folks building a (vacuum-pump-driven) air gun to watch what happens to baseballs traveling faster than a thousand miles per hour. The whole video is kind of a fun watch; I'd never seen a baseball just disintegrate before!

4.3 Roller coasters

In this last section, we're going to talk about what acceleration *feels* like, and why. On campus you get get a taste of this feeling by riding any of the elevators – perhaps you've noticed that when an elevator accelerates upward (when it starts at rest and then jolts to get to its smooth upward velocity) you feel slightly heavier for a brief moment, and when it accelerates downwards you feel a bit lighter.

4.3.1 The feeling of acceleration

You may have noticed that same feeling when you're driving very fast over the top of a hill – at the very top there's a brief moment where you suddenly feel as if you weigh a little bit less. Similarly, when you're driving fast and go around a turn perhaps you feel like you're being "pushed" outward – and remember that since velocity is a vector, changing your direction *also* corresponds to an acceleration. Similarly, when you are stopped at a red light and suddenly slam on the gas, you feel yourself pushed backwards into your seat.

Up until this point we've focused on how forces generate accelerations, from Newton's 2nd Law of Motion. But the reverse perspective is equally valid: if we observe an object accelerating, we can infer that a force is pushing or pulling on it. The details of how that force is exerted, say, on you, ultimately determines what you feel as you undergo acceleration.

[5]Please don't try this, though. The *collision itself* will be mild, but if either of you loses control of their car that would be... bad.

[6]Part of the rules of baseball involve *defining* a baseball by the way it bounces off of ash wood!

For instance, sitting in a car moving at constant velocity, you are not accelerating at all, and for the most part all you feel is the car seat pushing up against you (supplying a support force to balance your weight). When you push down on the gas, the car's engine converts some chemical energy from the gas into rotational motion of the tires; with the help of static friction to keep them rolling but not slipping, that additional rotational motion makes the car move faster. Your car seat is connected to the frame of the car, and *you* accelerate forward because the back of your seat is pushing you forward – applying a force to get you to accelerate as much as the rest of the car.

Remarkably, the feeling of acceleration is the same as the feeling of gravity: if you were to accelerate forward at 9.8 $\mathtt{m/s}^2$ you would feel the car seat pushing on your back in *precisely* the same way as if you were moving at zero velocity but lying on your back (in a car seat rotated by 90 degrees, I suppose).

Actually, how *could* they be different? When we're standing motionless on the ground what we feel is the ground pushing up with enough force to keep us motionless. It's not like the ground *knows* what is supplying our weight, it's just generating a support force. When we're accelerating, the sensations we feel are our nerves detecting the external forces on us and the internal forces the different parts of our body exert on each other.

Importantly, this is not a statement about the fallibility of biological sensations: it's a statement about the nature of the physical laws of the universe. A famous thought experiment involves locking a scientist in a compartment, with whatever scientific instruments she or he might desired. *No* measurement the scientist could conduct on anything inside the compartment would be able to answer the question "Am I in a gravitational field, or is the whole compartment accelerating?" This inability to distinguish between a gravitational force and a *non-inertial, i.e., an accelerating frame of reference* is an example of the *equivalence principle* that underlies Einstein's theory of general relativity.

When thinking about how accelerations "feel," you can summarize the above by characterizing your "feeling of acceleration" by an *apparent weight.* Standing motionless on the ground, you feel like you normally feel, and your apparent weight is the same as your actual weight, i.e., the force with which gravity is pulling down. If you're in a rocket ship blasting into orbit (as we'll consider in the next chapter), you're accelerating *upwards*, and your apparent weight has a magnitude given by the sum of your actual weight and your upwards acceleration (since accelerations are vectors, this is like taking the difference of the acceleration due to gravity and and your acceleration). If you're slamming on the gas in your car, you're accelerating forward, but it feels like you're being pulled down and back.

4.3.2 Riding a carousel in a playground

We have a good intuition of the feeling of acceleration when we travel in a straight line, but what about when we're turning, changing our velocity by changing our direction (rather than our speed)? Let's first think about this in the context of a playground carousel, shown in Fig. 4.3, where a rider travels around in a circle.

First thing's first: since the rider is traveling in a circle, the rider *must* be experiencing a net force! We know this from Newton's Law: without a net force the rider would be inertial and, hence, would travel in a straight line (which is what it means to have a constant non-zero velocity). Instead, the rider's has an acceleration – the direction of velocity is changing

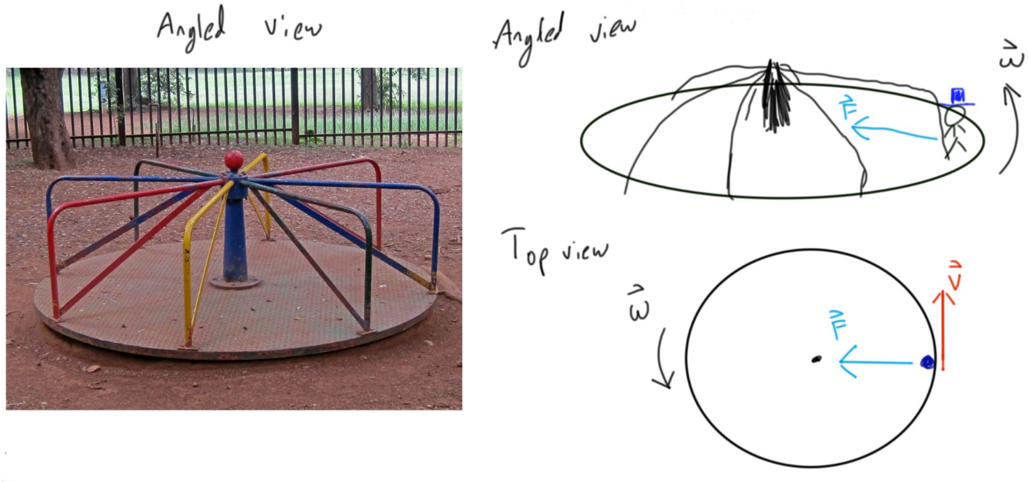

Figure 4.3: **Riding a carousel** (Left) An image of a random carousel in a playground. (Right) A schematic of someone riding a carousel. As the carousel rotates, the person holds on tight, supplying a force that points towards the center of the carousel. The top view shows that while the person's *instantaneous velocity* is pointing "tangent" to the circle (and that's the direction the person would fly off if they let go!), the force is pointing *inward* towards the center; this is *also* the direction of the person's acceleration as their velocity changes direction.

– but which way is the acceleration pointing? One way[7] to answer this is to think carefully about how the velocity is changing: the instantaneous velocity is always tangent to the circle. Viewed from the top and if the rotation is counterclockwise, when the rider is on the right side the velocity points up. This velocity would move the rider away from the pivot (if there were no other forces involved); since the rider stays at the same distance from the pivot, in fact, the rider must be accelerating *towards* the pivot!

This is an example of *uniform circular motion*: "uniform" because the angular speed of the carousel is constant (and, hence, the instantaneous speed of the rider is constant), and "circular motion" because... it's motion in a circle. For an object to execute uniform circular motion there must be a force pointing in the direction of the center of the circle; this force is called a *centripetal*[8] *force*. The centripetal force is not a new type of force, it's just the net force pointing towards a central pivot when an object is turning: it might be supplied by friction, or support forces, or a rider desperately holding onto one of those carousel bars.

We see that the direction of a centripetal force is always towards the center in uniform circular motion, what about the *magnitude*? You probably have a sense that it depends on the angular speed – a carousel turning faster is harder to hold onto – but angular speed on it's own doesn't have the right units! The magnitude also depends on how far away from the pivot point the rider is. Before looking, can you guess from dimensional analysis what the answer is?

[7] Another intuitive way is to imagine swinging an object by a rope around you in a circle: the rope is the only thing exerting forces on the object, and the only direction the force it exerts *can* pull is along it's own direction, i.e., towards you (the pivot in this example).

[8] which is derived from Latin roots meaning "center-seeking"

The answer is that the acceleration keeping an object in uniform circular motion is

$$\vec{a} = \frac{v^2}{r} = \omega^2 \cdot r.$$

That is, the acceleration is given by the speed squared divided by the radius of the circular motion, which is identcial to the angular speed squared *multiplied* by the radius. The force is, of course, the acceleration times the mass.

Having talked about centripetal forces, you might be wondering about *centrifgual* forces. Because, when we're turning in a car we feel like there's a force pushing us outward (remember: the feeling of acceleration is in a direction opposite to the actual acceleration!), people often think that there is such a force pushing the objects outward as they move in a circle. This is one of those caveats when we said that *from any inertial reference frame* physics looks the same... if you're in a rotating frame of reference (like when you're spinning around on a carousel) you are always accelerating, so you are *not* in an inertial reference frame! To you, it feels like there is a "centrifugal force" pushing you outward, and you have to hold on tight to exert a force inward so that you don't accelerate away from the pivot. An (inertial) observer standing to the side of the carousel can see, however, that no such additional force actually exists: you are simply exerting an inward force in order to constantly change the direction of your velocity so that you move in a circle instead of a straight line! Such "fictitious" forces that only show up when you analyze a situation from a non-inertial frame of reference are often called "pseudo-forces."

4.3.3 Riding a rollercoaster

There's not much else to say: riding rollercoasters is fun because they involve dramatic and often rapid changes in your velocity as you ride them! By twisting and turning and loop-the-looping along the track, a rollercoaster subjects you to accelerations in all sorts of different directions. Our experience of our own weight depends both on the acceleration due to gravity *and* the acceleration we undergo.

For instance, in a drop tower we are in free fall: our acceleration is *the same* as the acceleration due to gravity, and our apparent weight is zero! We feel weightless, and our seats and straps do not need to exert any forces on us. Similarly, your internal organs and connective tissue no longer need to support each other, contributing to your sense of bizarre weightlessness.

What about when you go through a loop-the-loop? Some representative snapshots of a trip around the loop are shown in Fig. 4.4; remember that the feeling of acceleration is opposite to your actual acceleration: over the course of your transit through the loop you feel yourself accelerating down, sideways, up, sideways, and down again.

The figure assumes that the motion of the rollercoaster is passive, starting at some velocity as it enters the loop, loosing some kinetic energy to gravitational potential energy as the coaster rises to the top of the loop, and then speeding up again as it begins its downward descent. It also draws the loop as a circle... in reality, most rollercoasters are designed to both be occasionally powered by the track the coaster is on, and the loops aren't usually circles: theme parks tend to design rollercoasters to maximize "fun," rather than to illustrate

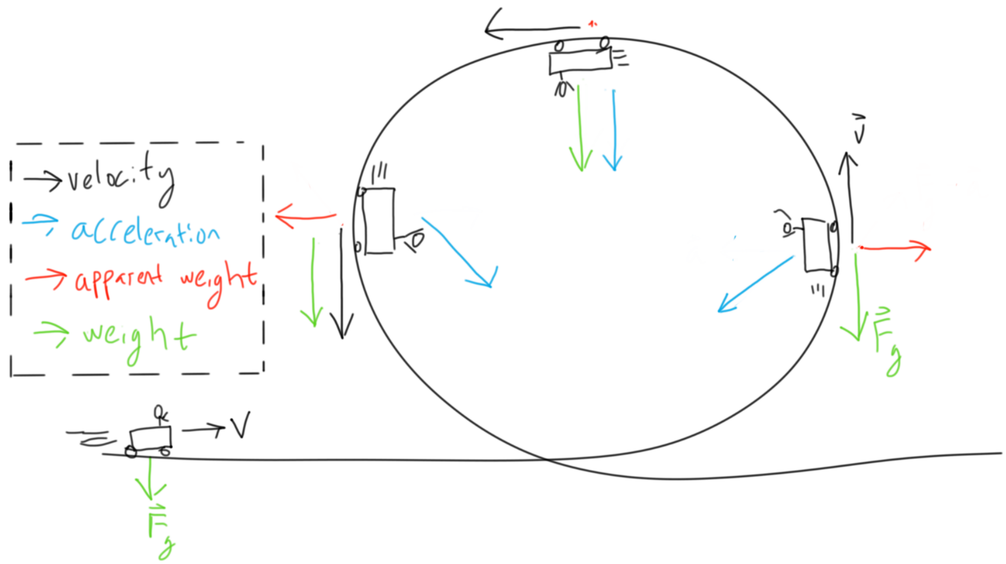

Figure 4.4: **A rollercoaster on a loop-the-loop** Illustration of the instantaneous velocity, instantaneous acceleration, force of gravity, and apparent weight. As the car starts going around the loop its acceleration is pointed down and to the left (not *directly* towards the center, because the motion is not at a uniform speed).

physical principles, which they do by designing the shapes of loops so that the riders always feel glued to their seats (for instance).

Pause and reflect! Why don't you fall out of a rollercoaster at the top of a loop-the-loop? How does your answer also explain why you can swing a bucket of water over your head without getting wet? Or control a basketball as you dribble it?

Pause and reflect! By thinking about the centripetal force involved in uniform circular motion, can you guess the *minimum speed* needed when you enter a loop-the-loop in order to make it all the way around.

Having thought about rotational motion, angular momentum, and centripetal forces over the last two chapters, here's a link to an artistic demonstration of these concepts in action!

Chapter 5

Mechanical objects, Part 2!

Many of the solid machines and objects in the world around us seem quite complicated, but in fact most of them behave according to the physical principles we have already encountered! In this chapter, we will continue to think about forces, energies, and conservation laws in a few more contexts. We'll think about why some objects tip over easily and others don't in the context of ultimately thinking about riding a bicycle, and we'll learn how planets and satellites orbit around celestial bodies, and how rocket ships escape from the Earth!

5.1 Bicycles

5.1.1 Static stability – when do object fall over?

Before we talk about bicycles, let's talk about *static stability*, which refers to an object's ability to be in stable equilibrium when at rest. Just being in equilibrium means there are no forces or torques acting on an object, but being in stable equilibrium means if the object is pushed a bit away from it's stable equilibrium point, it will experience *restoring forces or torques* that will return it to its stable equilibrium position. For example, if you put a marble in an empty bowl it will be statically stable: give it a little push, and it will roll back down the bowl to where it started. If you flip the bowl over, perhaps you can precariously balance

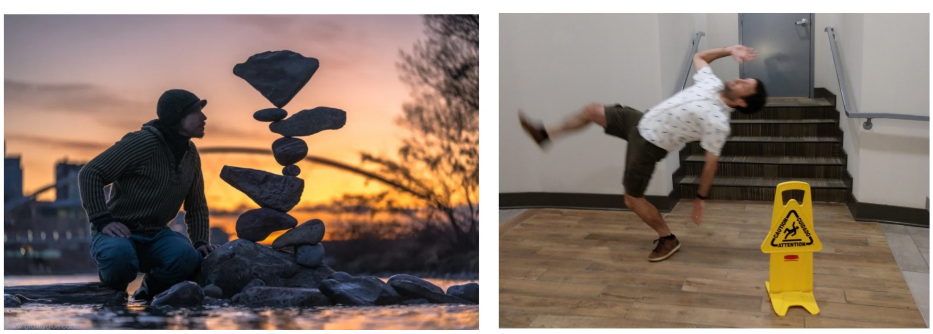

Figure 5.1: **Static stability** (Left) The artist Michael Grab playing with balance and gravity to create beautiful, stable sculptures with nothing but oddly-shaped rocks and the force of gravity. (Right) Your professor, about to fall in a particularly ironic location?

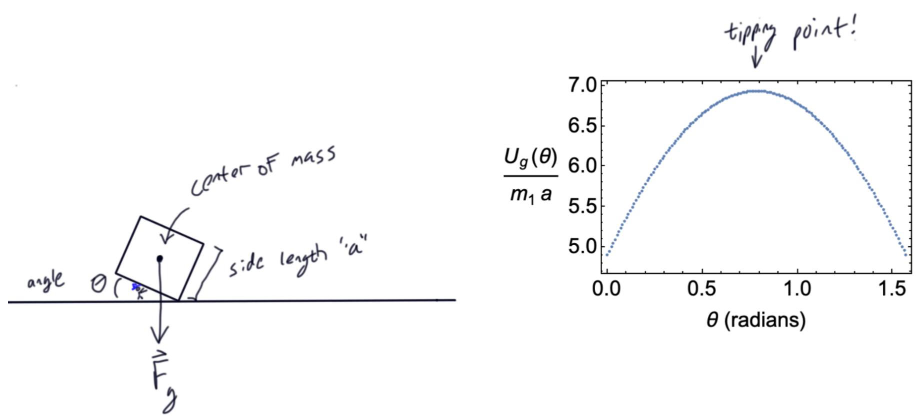

Figure 5.2: **Tipping over a square block** (Left) We imagine trying to tip over a square block. Gravity is pulling straight down on the block's center of gravity, and as we lift a corner of the block up by an angle θ the center of gravity rises – the block is gaining gravitational potential energy. If we let go, that potential energy would be converted into kinetic energy and the block would return to a flat position. (Right) Which way does the block fall? If we think of the total potential energy of the block (here, just from gravity), the block will fall in the direction that brings it "downhill" if we imagine the potential energy as a surface!

the marble on top, but it will be statically *unstable*: give it a little push and it will roll away. For many of the objects around us we have an intuitive sense of this, but sometimes whether something is in stable equilibrium is less obvious, as in Fig. 5.1.

Rather than relying on our intuition, what physical principles can we use to know if an object is statically stable or not? One principle relies on a lot of details: you can think about an object at rest and then figure out all of the forces and torques that will act on it if you give the object a little nudge. More generally, though, we can think about the *total potential energy* of an object as it is moved around. For example, let's look at the block being rotated in Fig. 5.2. In this example, we've imaging rotating a block and asking when it will tip over and when it will return to its original position. Here we just need to think about the gravitational potential energy of the block: At first, tilting the block a little bit raises its center of gravity, and that increases the gravitational potential energy. If you let go at this point, the block will try to reduce its potential energy as quickly as possible, and it will return to its original position. Tip it too far, though, and it will be faster to fall away from its original position.

It is, in fact, a very general principle that *an object will feel a force in whatever direction reduces its total potential energy the most*! So, if it is more complicated to look in detail at all of the forces and torques and ask if they are restoring or not, we can just look at all of the forms of potential energy that might be relevant. For gravity-based stability, we can ask if an object's potential rises when the object is nudged: if so, the object is statically stable and will return to equilibrium after being nudged; if not, it won't!

For the stability of objects due to influences with gravity, this general principle leads to a handy rule of thumb. Let's think about all of the points at which an object is resting on a surface – the two wheels of a bicycle, the four wheels of a car, etc. – and call the shape formed by all of those contact points the *base of support* of the object. Our rule of thumb tells us that an object will be in stable rotational equilibrium (it won't tip over when nudged) if a vertical line passing through the object's center of gravity goes through the *interior* of the base of support. This rule of thumb follows from geometry: if the vertical line through an object's center of gravity goes through the base of support of the object, then tilting the object will *raise* the object's center of gravity and, hence, increase its gravitational potential energy. On the other hand, if the vertical line is outside the object's base of support – or if it right on the edge and the center of gravity is above the edge – then the object is out of luck: it is either in the process of toppling over or it is in unstable equilibrium, and the slightest push will knock it over

5.1.2 Dynamic stability – what keeps a bicycle upright while it's moving?

So, how do bicycles stay upright while we ride them? Their base of support is a single edge (between the points of contact each wheel makes with the ground), so *at best* a bicycle and rider could be in an unstable static equilibrium by carefully putting the center of gravity of the bike plus the rider above that edge[1]. And yet: a moving bicycle is quite stable. Similarly, trying to balance a coin on edge is not a particularly fun way to pass the time during a pandemic, but a *rolling* coin stay upright for as long as it's rolling fast enough. What's going on? It turns out that while a bicycle and a rolling coin lack static stability, they both possess *dynamic stability* – a form of stability in which the object's own motion helps create restoring forces! Let's see what accounts for the dynamic stability of a bicycle.

First, because a bicycle wheel spins, we know it has angular momentum, and will continue spinning at the same angular speed and about the same axis unless it is acted on by torques. That's nice, but it's hardly enough to keep a wheel upright. The existing angular momentum of the wheel, however, can create a self-steering effect known as *gyroscopic precession*, which occurs when an object experiences a torque which is perpendicular to its angular momentum. When a bicycle is perfectly upright, the support force acting on the wheel is straight up – directly towards the wheel's center of mass – and so *the ground exerts no torque on the wheel*. When the wheel starts to lean in one direction, though, the upward support force no longer points directly towards the wheel's center of mass, and so there *is* a torque, and that torque is perpendicular to the axis about which the wheel is spinning[2]. The wheel "precesses," changing its direction in the same direction it is moving. As the wheel rolls in that direction, it tends to bring the center of mass of the wheel back above where it is touching the ground – a form of dynamic stabilization! The strength of this gyroscopic precession effect depends on how much angular momentum the object – does the center of mass come back above the contact point before the object topples even farther over? – but this effect explains why a coin can stay upright as it rolls. As long as it's rolling fast enough, gyroscopic precession will stabilize it; eventually friction slows it down to the point where it topples over[3].

[1] It's even worse for a unicycle, whose base of support is a single point!

[2] it might help to draw a picture and use the right-hand rule to see this!

[3] This precessing phenomena, by the way, also comes into play with spinning tops, and helps explain why they wobble more and more just before they fall over.

Gyroscopic stabilization is good enough for rolling coins; it turns out that bicycles have one additional trick up their sleeve. Bicycles are shaped very cleverly, so that when a bicycle is leaning in one direction, the gravitational potential energy can be *decreased quickly* by steering the front wheel in the direction the bicycle is leaning! Once again, by steering in the direction of the lean and by moving fast enough, the bike is able to bring its center of mass back above the edge forming its base of support. Here's a nice youtube video summarizing some of these ideas quite nicely, and separately, here's a short article talking about how gyroscopic stabilization is not sufficient – certainly not if you want to ride your bike in a particular direction rather than wherever the wobbles take you! In the youtube video, notice how when the handle bar is locked the bike stays upright – like a coin rolling – but only for a very short time!

Why do we lean into turns?

The combination of effects above – self-steering of the wheels due to gravitational potentials and the self-steering due to gyroscopic precession – helps us understand why bicycles can be dynamically stable while we're riding them in a straight line. But what about when we want to turn? How does that work, and why do we lean as we go through turns? Actually, now that you think of it, maybe you realize it's not only when you're on a bike do you lean through turns – you do it on bicycles and motorcycles and skis and water skies and, for that matter, even if you're just running through a tight turn barefoot!

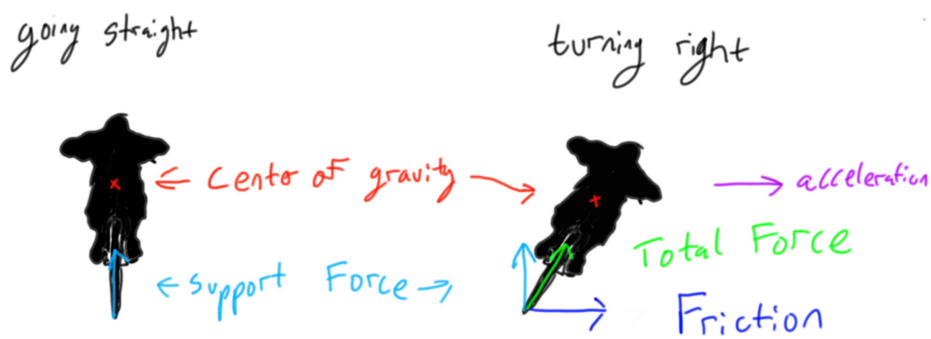

Figure 5.3: **Leaning into a turn** (Left) While moving forward and thanks to the dynamic stability of the bicycle, the center of gravity of the combined bicycle and rider stays about the base of support's edge. (Right) When going through a turn, the rider must accelerate towards the center of a circle that matches their turning path. Friction provides the force needed to do this, but it also exerts a torque on the rider (since there is a lever arm between where friction acts and the center of gravity of the bike-rider combination). To balance this torque, the rider leans into the turn: now the support force from the ground can provide a counter-torque, stopping the rider from flipping over.

The key observation is that as you go through a turn you know you must be accelerating *towards the center of a circle* with an arc that matches the path you're trying to take, and if there's an acceleration there must be a force! On a bicycle, it is the force of friction that supplies the center-seeking force, but the force of friction is acting at some distance away from your center of mass, and it's pointing not towards your center of mass but towards the center of that circle. Thus, you are experiencing a torque, and if you're not careful that torque will flip you over[4]! To counteract this torque, we make use of what we learned above: as a wheel *leans* the support force still points up. By finding the right angle of the lean, we can make the *combined* force of friction and the support force point in the direction of our center of mass, negating this problematic torque. This is illustrated in Fig. 5.3.

5.1.3 Choosing the right gear

Briefly, above we discussed why bicycles are able to stay upright when they're moving, and how to go through a turn by leaning into it... What about all of those gears we can switch between? The drivetrain of a typical modern bicycle is illustrated in Fig. 5.4. When bicycles were first invented it was the *front* wheel that the cyclist powered: the pedals were connected to cranks that directly turned the wheel, so the force from the cyclist's foot was converted directly into a torque on the wheel. This was well and good for a little while, but people soon got tired of pedaling very fast on flat ground or when going downhill (when it is easy to spin even that large wheel) and that absolutely grinding when going uphill[5].

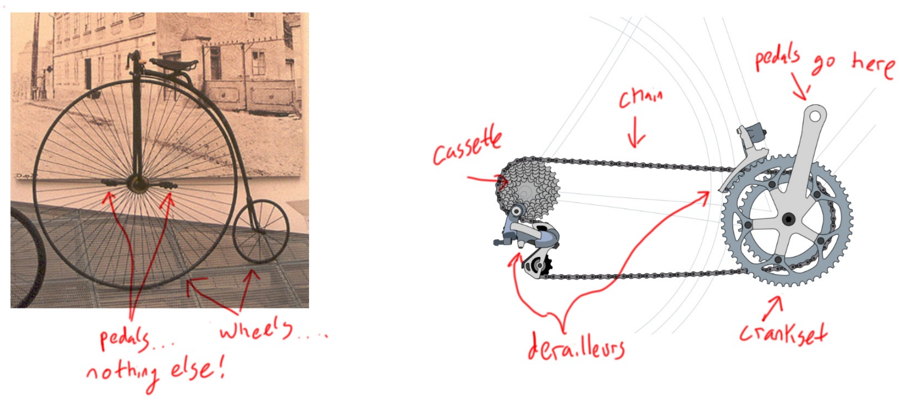

Figure 5.4: **Comparison of bicycle power systems** (Left) A penny-farthing, perhaps the first thing to be called a "bicycle." The pedals directly turn the ridiculously big front wheel, and on flat ground penny-farthings can go pretty fast: each time your feet drive the pedals around the bike moves forward by the circumference of the big wheel. This one-to-one ratio of pedal revolutions to front-wheel-revolutions meant it was very hard to go uphill! (Right) A modern bicycle drivetrain: A chain connects one of a few gears on the front wheel's crankset to one of the gears on the back wheel's cassette, and derailleurs allow the rider to change which gears the chain connects. By adjusting the ratio of the sizes of the connected gears, the cyclist can choose how far the bike will move forward with each complete revolution of the pedals.

It's something of a quirk of physiology – or perhaps it is an evolved trait – that humans turn out to be pretty good at outputting certain levels of power under various conditions. When it comes to applying force while moving your feet in a circle, it turns out that humans

[4]This is why, for instance, a car going through a tight turn at high speeds can flip over.

[5]I imagine people also got tired of falling from such a relatively great height.

get a maximal power output by sticking within a narrow range of cadences (the rate at which the pedals are driven around) and a exerting a "medium-hard" push on the pedals as they do so. Over long distances, this choice turns out to be much more human-power-efficient than, say, applying a light force with an extremely high cadence or a huge force at a very low cadence. Penny-farthing wheel sizes were basically optimized to have this power efficiency if you were cycling over flat ground, but going uphill or downhill? Out of luck.

Modern bikes use indirect drive systems, where your pedaling turns a gear on the front wheel, and that gear drives the motion of the rear wheel by having a chain run over a gear on both the crankset and the cassette. By *shifting gears* the rider can choose their level of mechanical advantage: to move the bike a specific distance forward, the rider can choose to exert large forces over fewer revolutions of the pedals or small forces over more revolutions of the pedals. By shifting gears as the rider traverses hills, the rider can try to stay in the sweet spot of human physiology, always exerting (close to) the same medium force with a medium cadence.

5.2 Satellites and planets

5.2.1 Orbiting around

In the last homework we saw that satellites flying around the earth could be thought of as undergoing (roughly) uniform circular motion where the force of gravity was supplying the centripetal force needed to keep changing the velocity vector of the satellite: just like a child exerting a force with her arms to stop from flying off a playground carousel, the attractive force of gravity between a satellite and the earth exerted a center-seeking force that kept the satellite going around and around the world.

In essence, a satellite is *free falling* around the Earth, but it is moving so fast that as it falls towards the Earth the curvature of the Earth makes the ground bend away from the falling object. One of Newton's thought experiments that illustrates this is shown in Fig. 5.5. He imagined a powerful cannon that was able to bring a cannonball up to an incredible velocity. The fast the cannonball left the cannon, the farther it would travel, until at some speed the curvature of the Earth bent away from the cannonball sufficiently rapidly that the cannonball made it all the way back to where it started. If we neglect air resistance (and the possibility of the cannonball crashing into something!), once set up in flight this cannonball would keep circling the Earth forever!

We call the path that an object takes as it free falls around a celestial object (a planet, a star, etc.) its *orbit*, and the length of time to complete one orbit the *orbital period*. The orbital period of a satellite or a planet ultimately depends on how far away it is from the object it is orbiting around. Why? Because gravity is the only thing supplying a centripetal force, and the strength of gravity depends on how far two objects are!

Near the surface of the earth, we've typically said that the magnitude of an objects weight is mg, where $g = 9.8\ \mathtt{m/s}^2$, and that the gravitational potential energy associated with changing the height of an object by h is $U_g = mgh$. These are just *approximations*, though – typically quite good approximations, but approximations nonetheless. They serve us well when thinking about behavior in a narrow range of altitudes near the surface of the

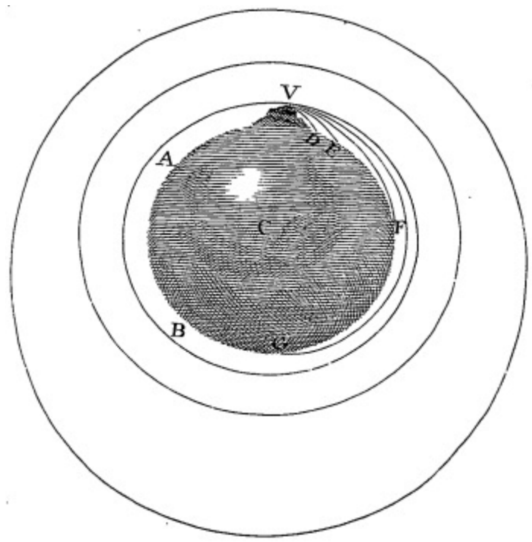

Figure 5.5: **Newton's Cannon** A cannonball, after being fired, moves with a velocity in the direction it was fired and then starts falling towards the center of the Earth. If fired with a sufficiently high velocity, the cannonball moves so far around the Earth that the curvature of the Earth matters, and with enough initial speed the cannonball will make it all the way around the Earth. From Newton's *A treatise of the system of the world*, 1728

Earth, but for satellites and spacecraft, we need to do better!

One of the things Newton discovered is called the *universal law of gravitation*, which says the magnitude of the force due to gravity between two objects is

$$F_g = \frac{Gm_1 m_2}{r^2},$$

where m_1 and m_2 are the masses of the two objects, r is the distance between them, and the gravitational constant $G = 6.6720 \times 10^{-11}$ $\texttt{N m}^2/\texttt{kg}^2$ is a *fundamental constant of nature.* That's the magnitude of the force vector, what about its direction? The force due to gravity acting on object 1 points from it towards object 2, and (by Newton's 3rd Law) the force acting on object 2 points towards object 1. This expression describes the force of gravity between any two objects: you and the earth, an apple and an orange in the same fruit bowl, your face and your cell phone as you hold them 18 inches apart. The distance you have to use in this equation is the *distance between the centers of mass of the object*, so for an object near the surface of the earth, r is the radius of the Earth (about 6371 kilometers). So, where does our old "mg" expression come from? Well, if we take the radius of the Earth to be 6371 kilometers and the mass of the Earth to be 5.972×10^{24} $\texttt{kg}$ and plug it into the formula above we get

$$F_g = (6.6720 \times 10^{-11} \ \texttt{N m}^2/\texttt{kg}^2) \times (5.972 \times 10^{24} \ \texttt{kg}) \frac{m_1}{(6371 \ \texttt{km})^2} \approx m_1 \times (9.82 \ \texttt{N/kg}).$$

Now, with the aid of the above equation, it's clear that the farther two objects are (and hence, the bigger r is), the smaller the force due to gravity is. But we saw in the last chapter that for uniform circular motion the centripetal force was directly related to the square of the *angular velocity* and to the *radius*. For a circular orbit we know $\vec{F}_{centripetal}/m_1 = \omega^2 r$; if the centripetal force is coming from gravity, we could rearrange this expression (notice how the mass of the orbiting object doesn't matter!) to relate the angular velocity and the radius of the orbit:

$$\omega^2 = \frac{Gm_e}{r^3}.$$

Intuitively: just above the surface of the Earth, a satellite needs to move screamingly fast in the horizontal direction to not hit the ground. For higher orbits, the satellite needs to travel more slowly so that it falls in a graceful circle. At some point (about 36000 `km` above the surface of the earth, the orbital period reaches 24 hours, and the satellite is said to be *geosynchronous* – since the orbital period of the satellite matches the Earth's rotation, an observer on the ground will see the satellite come back to the same point in the sky once every 24 hours. If a geosynchronous satellite is, additionally, traveling East around the equator, it is *geostationary*: to an observer on the ground the satellite appears to stay in the same place in the sky all the time!

5.2.2 Kepler's Laws

It turns out, though, that not all orbits are simple circles (something you might remember from the fact that the Earth and the Sun are not the same distance apart all year, and something you also might have guessed after looking at Fig. 5.5). The astronomer Johannes Kepler (working off extremely-accurate-for-his-time data from Tycho Brahe[6]) published three laws describing planetary motion in 1609 and 1619 – one of the major achievements of Newton's work was the fact that he could show Kepler's Laws followed as a consequence of his own three laws of motion together with his law of universal gravitation! What did Kepler find?

Kepler's First Law states that planets move in *elliptical* orbits, where one of the foci of the ellipse is the sun. You might remember from geometry that an ellipse is not just any random oval, it's a generalization of a circle. Just like a circle is the set of all points equally distant from the center of the circle, an ellipse is the set of all points such that the sum of the squares of the distance from two foci are the same (as in Fig. 5.6). A circle is just an ellipse where the two foci are in the same spot. For planets, one of the foci of their elliptical orbits is the sun, and the other focus is at some point in space.

Kepler's Second Law states that, just as planets don't have to orbit in a simple circle, they don't have to do so at a constant speed! We recognize this as a statement of the *conservation of angular momentum* – the force of gravity from the sun on a planet points directly towards the sun, and so it does not exert any torque. Thus, the planet's angular

[6] Quite a character, by the way. Among other things, he once lost his nose in a duel – the cause of which was an argument over who was a better mathematician! – and wore a (silver, gold, or brass) prosthetic nose for the rest of his life. He also kept a pet elk, until "unluckily the elk one day walked up the stairs into a room, where it drank so much strong beer, that it lost its footing when going down the stairs again and broke its leg, and died in consequence"

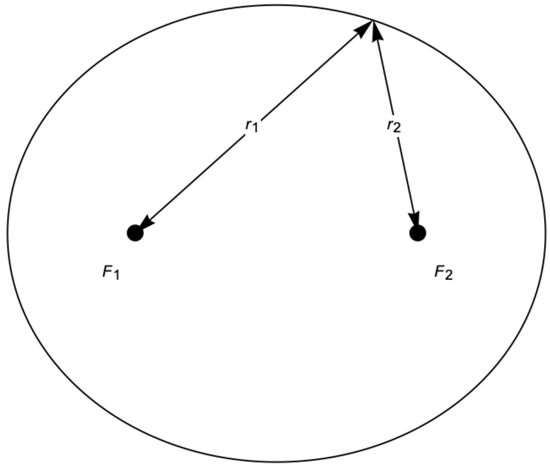

Figure 5.6: **An ellipse** Is defined by two *foci* (F_1 and F_2 in the figure) and a magnitude, c. The ellipse is the set of all points with $r_1^2 + r_2^2 = c^2$.

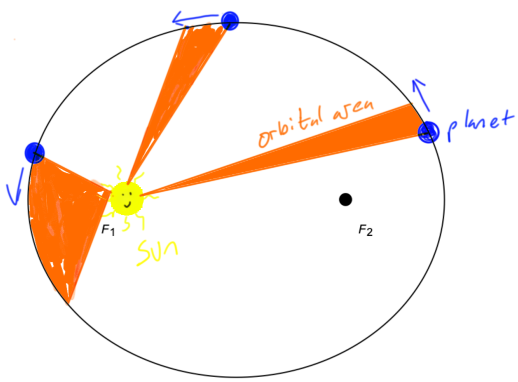

Figure 5.7: **Kepler's Second Law of planetary motion** An orbiting body will sweep out an equal amount of "orbital area" in equal amounts of time.

momentum is constant as it orbits around the sun – just like the figure skater, this means when a planet is closer to the sun it must be orbiting faster! *Kepler's second law* is a precise statement of this: if you imagine a line connecting the sun to a planet, that line sweeps out *equal areas in equal times.* This is illustrated in Fig. 5.7.

Kepler's Third Law is an observation about the orbital period of an orbiting object. It states that *the square of the orbital period is proportional to the cube of its average distance from one of the foci of the ellipse.* Notice how similar this is to the relationship we already worked out for orbits which are circular! The angular speed, ω, is *inversely related* to the period of rotation[7], and for a uniform circular motion due to gravity we said that $\omega^2 = \frac{Gm_e}{r^3}$.

[7]Why is this? Think about what ω means, and think about units!

5.3 Rocket ships

Launching people into outerspace – what a totally bonkers idea! Also, an incredibly impressive sight: here's a clip of a shuttle launch, illustrating the basic principles.

5.3.1 Blasting off...

At their heart, rocket ships are actually extremely simple! In our everyday lives we think about generating motion by *pushing off* against stuff: we push against the ground with our feet to walk and run, we can push a wall while sitting in a roller chair to slide across the floor, and so on. Well, that's one way to move around, but it would hardly work in outer space: there's nothing to push off against! So, what do we do instead?

Do you remember learning (in Chapter 3 of these notes) about using conservation of momentum to escape the frozen, frictionless lake? In that case, even with nothing to push off against, we could exploit the conservation of momentum and pick up a velocity in one direction by throwing our hat in the opposite direction. Rocket ships are doing the same thing, except instead of throwing a hat, they are throwing superheated, supersonic gas molecules behind them! There are lots of engineering details – how do you design the shape of a rocket nozzle so that the gas molecules are moving as fast as possible and in a direction directly away from the direction you want to move in – but we already have the physics principles to understand the broad idea. Chemical rocket fuel is burned in a small volume of the engine, and exhaust gas is forced out through the nozzle at an extremely high velocity – as fast as 10000 `mph` during a shuttle launch!

5.3.2 ... into space!

As the rocket pushes the exhaust gas behind it, the gas pushes forward on the rocket, giving it an acceleration. When that force from the gas on the rocket is enough to overcome the rocket's own weight, it lifts off the ground and starts picking up more and more speed. Because the rocket can continuously accelerate (by exhausting more and more gas behind it), as long as it can keep firing the rocket engines it will eventually make it far, far away from the Earth.

What if we wanted to launch into space not by firing rocket engines, but just by building a really[8] big slingshot – or something equivalent – that would give us some initial velocity so fast that, like Newton's cannonball, we could either start orbiting the Earth or leave the Earth completely behind us? How fast would we need to launch ourselves to pull off this trick?

Here we are asking about the *escape velocity*, the velocity an object needs in order to escape from Earth's gravity. And it turns out that we can figure out this escape velocity just by thinking about the conservation of energy! However, just like Newton's universal law of gravitation is a more accurate expression for the force of gravity than the approximation, mg, that we usually use near the surface of the Earth, once again we're going to need a more accurate expression for gravitational potential energy. The potential energy stored in

[8] *really*

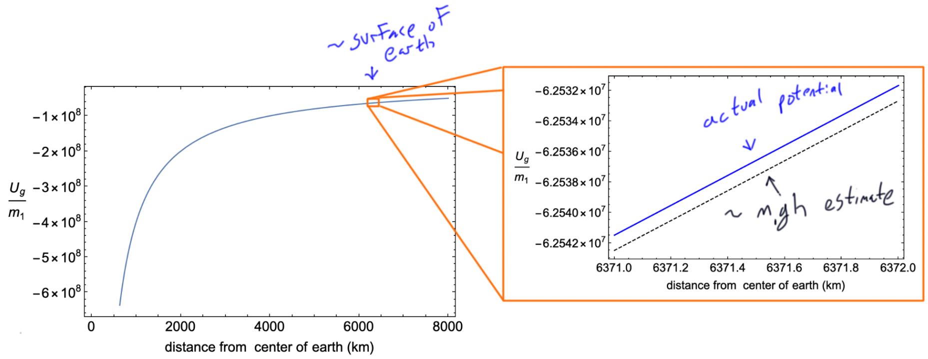

Figure 5.8: **Gravitational potential energy** (Left) A plot related to the potential energy stored between an object with mass m_1 and the Earth, as a function of how far the object is from the Earth's center. The whole potential is clearly not just a straight line! (Right) Zooming in very close to the range of the Earth's surface, we see that we can *approximate* the potential by a straight line, which I've drawn with a dotted line whose slope is g.

the gravitational attraction between two objects is actually

$$U_g = \frac{-Gm_1m_2}{r},$$

where we've got the same quantities that appeared in the force law, but now there's only one power of r in the denominator (and there's a minus sign)[9]. You might think that this expression looks pretty different from the "mgh" approximation we usually use. Figure 5.8 shows why, in a narrow range of distances close to the Earth's surface, approximating the complicated potential by a straight line works pretty well!

How can we make use of this? Well, on the surface of the Earth we're starting with a certain amount of gravitational potential energy, and we want to figure out how much *kinetic energy* we would need to start with in order to get so far away from the Earth that we would no longer have any gravitational potential energy! Ignoring air resistance and other such complications, we are basically asking what initial velocity has as much kinetic energy as our starting gravitational potential energy.

A little bit more explicitly, let's think about it as follows: at the moment of launch we initially have a certain amount (say, $\frac{1}{2}m_1v_e^2$) of kinetic energy and a certain amount of gravitational potential energy, $U_g = \frac{-Gm_1m_2}{r}$. We want to know what value of v_e is *just barely* enough to escape from the Earth; this break-even point would be when in our final state we no longer have *any* left over kinetic energy, but we also have zero gravitational potential energy. As an equation:

$$\frac{1}{2}m_1v_e^2 = \frac{-Gm_1m_2}{r} \Rightarrow v_e = \sqrt{\frac{2m_2G}{r}},$$

where (from the way I've written this) m_2 is the mass of the planet we're escaping from and r is the distance from that planet's center of gravity that we're starting from.

[9] Again, if you've taken calculus you should notice that there seems to be a relationship between the force and the derivative of the potential!

Chapter 6

Fluids!

In the last few chapters we've focused on solid objects – ones that keep their shape, resist deformations, and so on. But much of the world around us is a *fluid*[1]: air and water come to mind, maybe you've noticed them? Unlike solid objects, fluids don't hold a well-defined shape, and so (for instance) it's harder to exert forces on them directly, and they move in sometimes unusual ways. They still obey the laws of motion and conservation that we've encountered so far, but thinking about how those laws apply sometimes gets harder when objects are not rigid masses. We'll start thinking about these fun systems in this chaper. We'll think about why balloons and boats float in air and water, we'll think about why all fo the air molecules in the atmosphere aren't just pulled down to the surface of the Earth, and we'll think about how we can do work to move fluids around.

6.1 Gases

Gravity acts on the molecules and atoms in a gas in the same way it does everything else, so how is it that a helium-filled ballon (or a hot air balloon) can float? A balloon filled with helium has mass, and so by the universal law of gravitation the Earth pulls balloons down, yet if we let go of a helium balloon they float away. What's going on, and what forces haven't we accounted for here?

To understand what the physical forces are holding balloons aloft, we have to start by thinking about gases. The air around us has mass, but what makes a gas different from the sorts of solid objects we've met in this class so far is that gases have neither a fixed shape nor size, and the same number of molecules of air can be shaped, molded, or shoved into all sorts of different volumes: a dozen Aluminum A80 scuba cylinders can hold the same amount of air as is in a typical dorm room, for instance!

Behind the malleability of gases is they fact that at the microscopic level they look very different from a solid. In a solid, the atoms and molecules of the material sit in particular places with very particular arrangements, as is particularly obvious when looking at the arrangement of atoms in a crystal. Gases, in contrast, are chaos: they are made up of atoms and molecules that simply fly around in every direction. They collide with each other, and

[1]A term that covers both liquids and gases

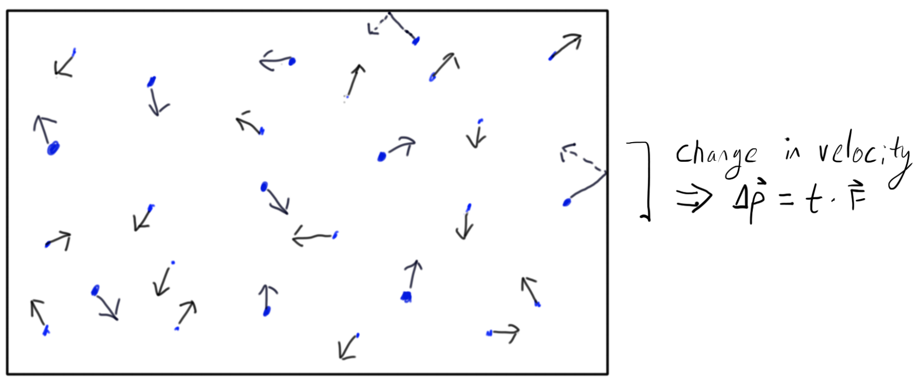

Figure 6.1: **Pressure** The pressure a gas exerts on the walls of its container comes from the – a force

with the walls of their containers, with no set arrangements between where any two particles in a gas will be at any time.

Why is it that gases can fill up the available space? Why doesn't all of the air settle down right on the surface of the Earth due to the force of gravity? Ultimately, it's because of the thermal energy that a gas has! In a gas, thermal energy is associated with the disordered motion of the particles composing the gas: a hotter gas is one in which the particles are flying, spinning, and vibrating faster than a colder gas composed of the same substance. *Temperature* is a measure of the average "internal" kinetic energy of the particles in a gas – here "internal" means the undirected kinetic energy that is not going into overall motion of all of the particles in the same direction, but is composed of the particles internally moving around.

The individual particles that make up gases are extremely small: an O_2 molecule in the air might only be about 2×10^{-10} m, with a mass of just 2.66×10^{-26} kg. But because their mass is so small, they also move around at enormously quick speeds: at room temperature individual particles in the air might be moving at 500 m/s! Because they collide and bounce so frequently, they don't spend moving consistently in the same direction before they bounce off of something, and this random, collision-filled existence tends to spread particles in a gas as far apart as possible.

6.1.1 Air pressure and the ideal gas law

Let's think about what's going on inside a bike tire, a tube filled up with air at a higher pressure than outside the tire. Inside the tube what is going on? There are some air particles flying around at high speed, colliding with each other and with the inside wall of the tire. Every time a tiny particle collides with the tire, it bounces off: *the particle experiences a change in its momentum*, and we know that means it experienced a force over some collision time. The tire is exerting a force on the particles as they change their momentum, and by Newton's 3rd Law the particles exert an equal and opposite force on the wall of the tires! This is illustrated in Fig. 6.1.

Because there are a lot of particles inside, and they're flying at such high speeds, these collisions happen frequently. Because the gas particles aren't going in any particular direction, there isn't a coherent net force on the tire, there is just *an average amount of force exerted on the tire wall over the surface area of the inside of the tube.* This is what we mean by a pressure:

$$\text{Pressure} = \frac{\text{force}}{\text{surface area}}.$$

The SI unit of pressure is 1 $\mathtt{N/m}^2$, which is also called a *pascal*[2] ($\mathtt{Pa}$). For reference, 1 pascal is a pretty small pressure: a stack of 13 pieces of paper sitting on a table is exerting about a pascal of pressure; the air around us (at see level) has a pressure of about 10^5 $\mathtt{Pa}$.

For a gas exerting a pressure on its container, the pressure is produced by all of these tiny, frequent collisions between the gas particles and the container wall. This already gives us an intuition for how the pressure will vary as we change things about the gas! The more frequent the collisions are, or the greater the changes in momentum for each collision, the greater the pressure. One way to increase the collision frequency is the *put more gas in the container.* If you double the number of air particles in the bike tire, you'll (roughly) double the number of times a gas particle bounces off of the tire wall.

The other way to increase the collision frequency is to make the particles *move faster*! Suppose we quadruple the average kinetic energy of the gas particles: this increases the *speed* of the particles by a factor of 2, which means that not only do the collisions with the wall happen twice as often[3], but the average force exerted during a collision will also be a factor of 2 greater. Combined: quadrupling the internal kinetic energy quadruples the pressure.

We said about that temperature is a way of measuring the internal kinetic energy of a gas, but we have to be careful with what units we use here! In particular, we need to use *absolute* temperature scales, ones where a temperature of zero means the particles in the gas aren't moving at all! For this, I'm afraid, Celsius and Fahrenheit just won't cut it! The SI absolute temperature scale is the Kelvin ($\mathtt{K}$), which is defined so that changing temperature by 1 degree Kelvin is the same as changing temperature by 1 degree Celsius, but $0\ \mathtt{K} = -273.15°$Celsius. Now we write down what we learned so far as a proportional relationship:

$$\text{pressure} \propto \text{density} \cdot \text{absolute temperature}.$$

Do you see why we need an absolute temperature in the above expression? At negative degrees Fahrenheit and Celsius, gases still exert *positive* pressures, so we need to use a temperature scale that starts at zero and only goes up!

We can make the above expression an equation, not a proportionality, by introducing the *Boltzmann constant*,

$$k_B = 1.381 \times 10^{-23} \frac{\mathtt{J}}{\mathtt{K}},$$

the constant of proportionality we need. With it in hand, we have found the *ideal gas law* – perhaps you saw it in a high school chemistry class:

$$p = k_B \frac{N}{V} T,$$

[2]Named after the French mathematician / child prodigy Blaise Pascal. Among other things, Blaise worked as a tax collector, and is considered one of the co-inventors of mechanical calculators.

[3]do you see why this is so? It might help to draw a picture!

where N is the number of gas particles and V is the volume they're occupying[4]. It's called the *ideal* gas law because it refers to an idealized approximation to the way real gases behave, but it turns out to be pretty good! It explains why driving down the highway makes your tire's pressure go up, or why pressure drops inside of footballs on a cold day.

6.2 Floating and buoyancy

Knowing about pressure, we can finally understand why balloons can rise into the air and why objects can float in fluids. If an object is immersed in a fluid – a gas or a liquid – the particles in that fluid constantly hit the surface of the object, exerting a pressure on its surface. The total forces involved in those collisions can sometimes be fairly large, but they often cancel each other out (the air pressure inside a soccer ball doesn't spontaneously cause the soccer ball to start flying off in some direction!).

In fact, in the absence of gravity all of these forces *would* tend to cancel out, and the pressure of a stationary fluid would be uniform throughout. In the presence of gravity, though, there is a *gradient in pressure*: the pressure increases as you go down into the fluid. You may have experienced this when swimming to the bottom of a deep pool: at the bottom you feel the weight of the entire vertical stack of water above you, whereas just below the surface barely any amount of water is pushing down on you.

So, when an object is in a fluid, the fluid below the object has a *higher pressure* than the fluid above the object, and thus there is an overall upward force that will act on the object – this is called *buoyancy* or the buoyant force.

How big is the buoyant force? Let's start with a thought experiment, where we imagine filling a bag which has a volume of 1 cubic meter completely with water – no air bubbles! – tying a rope to it, and lowering it into a pool. Using standard numbers for the density of water, the weight of this bag will be about 9777 N; when it's completely submerged, though, how much force do we need to exert on the rope to keep the bag in place? None! Just as still water doesn't experience any net forces from being suspended in other bits of still water, our bag of water will just float neutrally at whatever depth we lower it to. Apparently the upward buoyant force is also 9777 N.

Let's keep thinking about this (see Fig. 6.2: What if we replaced the water in the bag with something that was twice as dense? We've got a bag with a weight of 19554 N, and it still occupies 1 cubic meter of space. What will happen when we lower it into the pool? It's not like the water molecules in the pool know anything about what's inside the bag: they exert exactly the same forces on the bag as they did in the first case! Thus, the buoyant force will be the same (9777 N), and if we don't want the bag to sink we need to supply an additional 9777 N in the upward direction to balance the full weight of the bag.

By this reasoning if we had a bag that weighed 19542 N but occupied 2 cubic meters, it would not require any upward additional force to hold in place – the buoyant force would perfectly counteract the total weight. What if we had a bag that weighed 19542 N but was 3 cubic meters? Now the buoyant force would be *greater* than the object's weight, and since

[4]You may have also see this instead as $PV = nRT$, where n is not the number of particles but the number of moles, and R is a different constant: the equations are the same, and its just a matter of using Avogadro's constant to convert between them.

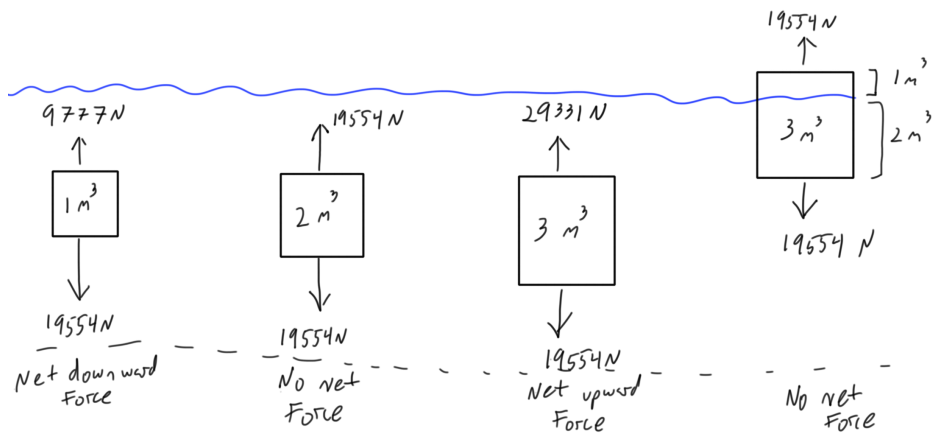

Figure 6.2: **Buoyancy** The buoyant force acting upward on an object is equal to the weight of the fluid displaced by the object. Here we consider three objects that weigh the same amount (19554 `N`) but have different sizes, and I'm rounding so that 1 cubic meter of water has a weight of ≈ 9777 `N`.

the buoyant force (upward) would be greater than the weight (downward), the object would float to the surface! How far above the waterline would the object float? It would keep rising until the weight and the buoyant force balanced, so 1 cubic meter would have to poke out above the surface.

What we've just discovered is called *Archimedes' Principle*, the idea that an object wholly or partially immersed in a fluid is acted on by an upward buoyant force equal to the weight of the fluid that object displaces. Such an object experiences (at least) two forces: a downward force due to gravity acting on its own mass, and an upward buoyant force related to the *volume* of fluid it displaces. So, will an object float or sink (in air, or in water, or in any other fluid)? It depends on the balance between these two forces! If an object displaces a volume of fluid whose weight is greater than the weight of the object itself, it will float; if not, it will sink.

Historical interlude: Archimedes' principle in connected to the story of his famous "Eureka" epiphany. Archimedes had been tasked with determining whether the King of Syracuse's crown was pure gold, or if it had been debased with silver – it was easy enough to figure out the weight of the crown, but knowing whether it was pure gold required knowing the *density* of the crown, and that meant figuring out the volume of a pretty irregular shape. According to legend, Archimedes realized while taking a bath that the level of water in the bath rose when he got in, and he realized that the *amount* that the water rose must be the same as the volume of his own body. Thus, submerging the crown in a totally full container would spill an amount of water equal to the volume of the crown. This was, apparently, so exciting that Archimedes ran naked through the streets shouting "eureka[5]!"

So, why can a steel boat float in water, given that steel has a greater density than water? Because the shape of the boat's hull can displace a volume of water whose weight matches that of the boat before it is completely submerged! Why can a helium balloon float in air?

[5]Greek *εύρηκα*, "I have found [it]!"

Because if you fill a helium balloon to the same pressure as the surrounding air, the total weight of the helium atoms will be less than the equivalent weight of air particles it would have taken to fill that balloon (One helium atom weighs about 14% as much as the average air particle, so 1 $\mathtt{m}^3$ of helium weighs about 14% as much as 1 $\mathtt{m}^3$ of air at sea level.)[6]

What about hot-air balloons, which are filled not with helium but with, well... hot air? Air has a density of only about $1.2\mathtt{kg/m}^3$, so not that many objects float naturally in it. A bag filled with absolutely nothing – a vacuum – would do the trick, but such a bag would rapidly collapse! The air outside the bag would exert an inward pressure – 100000 N at sea level! – and there would be nothing to exert a counterbalancing outward pressure. Implicit in the helium balloon example above, we want to fill our floating bag with something that weighs less than the air around it but can still exert an outward pressure on the bag to keep it from crumpling up.

How do we make a gas that has the same pressure as the air around us (so the balloon won't be crushed) but is less dense (so the balloon will float)? The ideal gas law points the way – we heat the air up! The hotter particles inside the balloon are moving faster, so we don't need as many of them to create the same pressure[7]. Fewer particles inside the balloon, though, means the density of air inside the balloon will be less – the average density of the hot air balloon is less than the density of the colder air outside, and so the upward buoyant force acting on the balloon can be more than the weight of the balloon (and the air inside of it, and the passengers) pulling it down!

6.3 Liquids and water pressure

In the next chapter we'll be thinking a lot more about fluids in motion – and objects moving through fluids – but for now let's take a moment to talk about some of the essentials of how pressure works in liquids. You may have wondered about why the water pressure in your house/dorm/class building seems to vary depending on which floor you were on, or why the water towers you sometimes see when you drive around hold water up so high, or (for that matter), what's going on when you use a straw to take a sip of your beverage of choice.

6.3.1 Pressure inside a liquid

In thinking about those questions, we need to think more about water pressure; this is important because water – just like everything else! – accelerates in response to forces, and if there are no forces acting on it then it won't go anywhere. Let's start out by thinking about water filling up a horizontal pipe (horizontal so that we don't have to think about gravity for

[6]It might not be obvious that when the balloon is filled to a specific pressure the *number of particles* inside the balloon will be the same. But at a given temperature, each gas particle has the same average internal energy regardless of its own mass – lighter particles move faster at some temperature in exactly the right amount so that the pressures they exert are the same as a heavier-but-slower particle of something else at that temperature.

[7]Notice: the fact that the pressure inside and outside the balloon can be the same – when the balloon is floating at a particular height and experiencing no net force in the upward or downward direction – means that there is no overall tendency for air to try to get in or out of the balloon. Hence, no real need to seal the balloon off!

the moment). If the pressure inside the pipe is uniform everywhere, then all of the different little bits of water will not feel any net force, and so the water will not accelerate. If there is an unbalanced pressure, though, the water will begin flowing from regions of high pressure to regions of low pressure.

Just like a solid object experiencing a force, this is not an instantaneous process, but one which follows the consequences of Newton's Laws: Water at rest tends to stay at rest, and in response to net forces (which can come from an unbalanced pressure) it will *accelerate* in the direction of the net pressure imbalance. This is how city plumbing works: as long as you have some devices which can control and impose complicated patterns of high- and low-pressure regions in different parts of different connected pipes, you can orchestrate the flow of water inside those pipes as you like.

How do we create pressure imbalances in a liquid, and how can we change the pressure in a liquid? The easiest way is simply to try to compress it! One of the things that separates a liquid from a gas is that a certain mass of a given liquid has a definite volume – it does not have a predefined *shape*, but the amount of space it takes up is some definite amount. One consequence of this is that water is basically *incompressible*: as you vary the pressure in the water it experiences essentially no change in the volume of space it takes up.

So, as a thought experiment let's imagine we have a thin plastic soda bottle filled completely to the brim with water – no air bubbles trapped inside, for convenience – and sealed with the cap on tight. When the bottle is just sitting there, the pressure of the water in the bottle is the same as the pressure of the atmosphere outside the bottle[8].

What happens if we squeeze the bottle? Well, we are applying a force on the bottle, and the bottle in turn applies a force on the water inside of it. But since the bottle is completely full and sealed, and since water is incompressible, the bottle doesn't get crushed, and instead the pressure *inside* the bottle rises while we are squeezing it. ***Pascal's principle*** says that in this case the pressure changes uniformly throughout the soda bottle; more generally it says that a change in pressure of an enclosed incompressible liquid is transmitted to every part of the fluid and to all of the surfaces the fluid is in contact with. This means that squeezing the sides of the bottle increases the pressure of the water inside the bottle everywhere equally, and with enough increase in pressure the bottle will eventually break (probably by the plastic threads of the screw top being deformed).

Pascal's principle underlies, for instance, the way hydraulic lifts work (see Fig. 6.3): a given force applied to a small surface area creates a given change in the pressure of the fluid ($p = F/A$, remember). This change in pressure is communicated everywhere throughout the fluid, and can act against (say) a piston with a much larger surface area. Equality of pressure means a much larger *force* can be communicated from the smaller to the larger piston.

If we had squeezed on the soda bottle with an open cap, the pressure in the water would increase and it would begin accelerating towards where the pressure is lower – in this case, up and out of the bottle. Let's think about operating a water pump: when we apply a force to a water pump, that force is equal to the pressure in the water times the surface area of the piston. In the process of of this, the displacement of the water is the volume of water

[8]How do we know this? Well, the gas particles in the air are constantly bombarding the plastic bottle, but we do not see the bottle move or get crushed. The water inside must be pushing from the inside of the bottle outward with the same pressure!

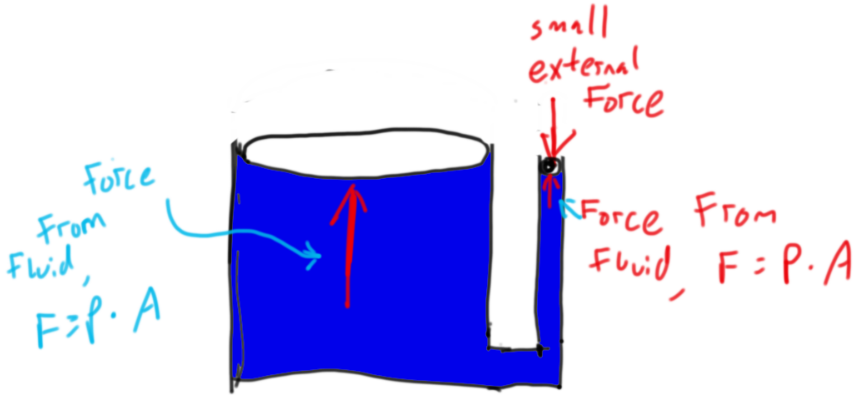

Figure 6.3: **Hydraulic lifts** Pascal's principle lets us use liquids as another type of simple machine: a pressurized liquid exerts a force on pistons proportional to their surface area. A small force exerted on a small piston raises the pressure in the fluid, leading to a larger force on the large piston.

pumped divided by the surface area of the piston. Combining this, the work done to to pump water is

$$\text{work} = \text{force} \cdot \text{displacement} = \text{pressure} \cdot \text{volume}.$$

6.3.2 Water seeks its own level

In the presence of gravity, the pressure in a fluid increases with depth (as we said while talking about buoyancy!): as you dive down into a pool, your body is feeling the weight of a greater and greater mass of water above you, spread out over the surface area of your body. This is true of the air in the atmosphere, too, but since water is much denser than air the effect of this are much more noticeable on a human scale.

A common (?) saying is that "water seeks its own level" – a principle that explains how siphons work: when a connected mass of water has multiple free surfaces (exposed to the air), and when those free surfaces are at the same atmospheric pressure, there are no pressure imbalances in the fluid and the water will accelerate to lower its gravitational potential energy. See Fig. 6.4, looking at a"u"-shaped pipe filled with water. What this means is that if the water has one free surface which is *higher* than another free surface, the water can reduce its total gravitational potential energy by reducing its average height, with some water from the highest area filling in the lowest area. The natural flows associated with this process have been used were known at least as far back as ancient Egyptian and Indus-valley civilizations.

Let's close this section by thinking about how drinking straws work. When we suck on a straw to bring water from a glass to our mouths, what are we actually doing? Are we exerting an attractive force on the liquid? Or is there some other force pushing water up the straw? What we are actually doing is sealing off part of a connected system of water: the water in the glass outside the straw is exposed to the atmosphere, and is at atmospheric

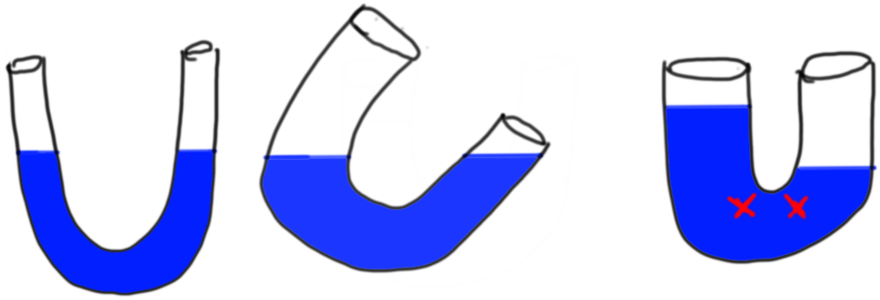

Figure 6.4: **Water seeking it's own level** (Left and center) In a U-shaped tube, the water will adjust so that the water level is the same on both sides, even as the tube is tilted. (Right) Suppose water didn't "seek its own level." Consider the two parts of water marked with x's. If the water is not moving, the pressures at these positions must be the same (otherwise the water would flow towards the region of lower pressure). But we know that water increases with depth, and there is more water above the x on the left. Hence: in the situation illustrated the pressure in the fluid is *not* uniform, and the water will flow until the heights of the sides are equalized.

pressure. Before we start drinking there is also some air in the straw (above the liquid), which we then suck out! With no (or very little air) above it, the liquid in the straw feels less pressure pushing it down, but outside the straw the liquid is still at atmospheric pressure. In the presence of this pressure imbalance, the liquid starts rising up the straw.

The fact that pressure imbalances are how straws work mean there is a maximum height of a straw that can work! Do you see the argument? As water rises up a straw, the weight of the column of water starts to compensate for the lack of pressure from the air. Since atmospheric pressure is about 10^5 Pa, and since the density of water is about 997 kg/m^3, a straw longer than 10 m just won't work!

Chapter 7

Fluids and motion!

In the last chapter we talked about fluids at rest, and how gradients in pressure would lead to the flow of fluid towards regions of low pressure (or lead to buoyant forces acting on an object in that fluid). But we avoiding discussing what's going on when the fluid actually is in motion! In part this is because moving fluids can be really quite complicated; in this chapter, though, we'll try to get a flavor of the physical principles underlying fluid motion.

7.1 Flowing water

7.1.1 Viscosity

When we watch water flow – down a river, out of a faucet, as we pour a cup of coffee – the first thing we notice is *how* the water flows. Compare how easily water flows to, say, honey, which flows slowly and sluggishly; what's going on? As we tip a jar (or a bear) of honey upside down, the force of gravity is pulling the honey towards the Earth just as it does everything, so what is opposing this downward motion and making honey move so slowly?

The answer is that all moving fluids experience frictional forces *within* the fluid itself! You can think of these frictional forces as coming from layers of fluid trying to slide over and past other layers of fluid, and in the context of fluid flows these kinds of frictional forces exerted by a fluid on itself are called **viscous forces**. Like other frictional forces, they oppose relative motion, and the measure of how much a fluid resists relative motion within the fluid is called the *viscosity*. The SI units of viscosity are "Pascal-seconds" ($\texttt{Pa}\cdot\texttt{s}$): air at room temperature has a value of about $2\times10^{-5}\texttt{Pa}\cdot\texttt{s}$, water at room temperature has a value of about $1\times10^{-3}\texttt{Pa}\cdot\texttt{s}$, and honey at room temperature has a value of about $10^{3}\texttt{Pa}\cdot\texttt{s}$.

I keep saying "room temperature" there because viscosity depends on temperature – something you know if you've ever heated up honey before pouring it. Viscous forces ultimately comes from the way atoms and molecules break and form weak bonds as they move past each other, and particles with more thermal energy can break these bonds more easily. It's pretty hard to predict the details of how the viscosity of liquids changes with temperature, and the formulas can get pretty complicated, so we're not going to present them here. Instead, let's focus on the physical picture of what's going on when, say, you try to tip honey out of a jar: the honey molecules that are actually touching the jar can get stuck there by chemical attractions, and so they remain pretty stationary. Honey that isn't even touching

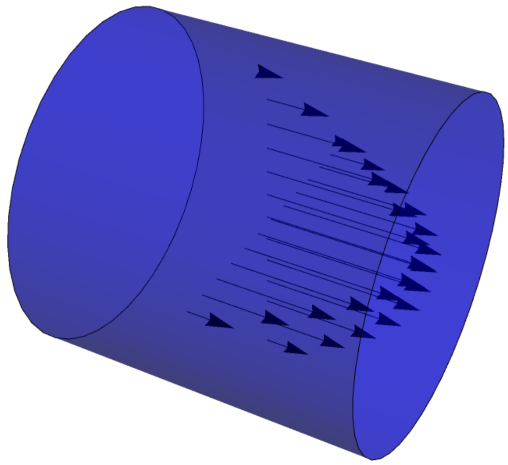

Figure 7.1: **Velocity inside a straight hose is not uniform!** When a fluid is flowing smoothly through a pipe, viscous forces act within the fluid and oppose relative motion. As a result, fluid will flow fastest at the center of the hose and slowest near the wall of the hose. This is indicated by the arrows above.

the jar then has to slide past those stationary molecules, and the viscous forces – the internal frictional forces acting between the different layers of molecules that are now trying to have relative motion – acts to oppose that motion.

The same thing happens in water when you try to tip it out of a jar, or move it through a pipe, or through a garden hose. It's just that the much lower viscosity of the water means there are a lot less of these viscous forces getting in the way.

7.2 Flow through a hose

7.2.1 Flow through a straight hose

So, let's first take a look at a simple scenario: water flowing through a straight, horizontal hose. The water molecules that are actually touching the hose tend to have chemical interactions with the hose and are held almost stationary. This motionless layer exerts viscous forces on the layer just next to it, slowing that layer down a bit. That somewhat slower layer does the same thing to the next layer, and so on: the water at the center of the hose will be moving faster than all of the other layers of water, as in Fig. 7.1.

Because of all of these viscous forces, instead of flowing freely real water requires a consistent gradient in pressure to keep it moving along in a hose or in a pipe. Just like doing work against the force of friction when sliding a solid object along the ground, this means it requires *work* to move the fluid around (as we hinted at in the last chapter), and some of that work is "wasted," going into frictional (viscous) heating! Even if we don't always notice it in the same way we do when we rapidly rub solid objects together, as a fluid flows it heats up!

There is an important difference between viscous friction and ordinary sliding friction: for sliding friction the magnitude of the friction force doesn't really depend much on the relative velocities of the objects – as long as two surfaces are moving even a little bit relative to each other, the sliding friction force is what it is. In contrast, viscous forces can become much larger as the relative velocities within the fluid becomes larger and larger. That's important to keep in mind.

We can summarize the way a fluid flows through a straight pipe or hose by *Poiseuille's Law*, which says that the flow rate of fluid flowing through a straight pipe is:

$$\frac{\Delta V}{\Delta t} = \frac{\pi}{128}\frac{\Delta p}{L}\frac{D^4}{\eta}.$$

Let's go step by step through this equation and talk about where these terms come from! On the left hand side "$\Delta V/\Delta t$" is the change in volume in fluid divided by the change in time – we can think of this as a measure of the flow rate of the fluid: imaging holding a fictitious bucket collecting all of the water flowing out of a pipe, and asking how much volume of water gets added to the bucket over time. On the right hand side, we first see $\pi/128$, a numerical constant that Poiseuille figured out needed to be there to make the equation right. Next we have:

1. The flow rate is proportional to the pressure difference between the start and end of the pipe, Δp. The pressure gradient determines how hard the water is being pushed, so it perhaps makes sense that the harder we're pushing the water the more flow there is.

2. The flow rate is *inversely* proportional to the length of the hose, L: the longer the hose the more opportunities viscous forces have to slow it down.

3. The flow rate is *inversely* proportional to the viscosity, η. That makes sense, since viscosity is what is opposing relative motion in the first place!

4. Finally, the flow rate is proportional to the *fourth power of the diameter* of the hose, D. The fact that it's the fourth power might be surprising. But if you double the diameter of a hose you (a) quadruple the amount of space for the fluid to occupy *and* (b) the fluid at the center of the hose can travel four times as fast. Combined, doubling the hose diameter lets you flow sixteen times as much fluid in the same amount of time for the same pressure difference.

Just to briefly dwell on that last point, hose diameter to the fourth power is a pretty strong effect! It means that if you're starting with the same water pressure at the beginning of your garden hose, increasing the hose diameter by $\approx 20\%$ will basically *double* the amount of water your hose pumps out!

7.2.2 Dynamic pressure variations: going around the bend

One of the things that makes fluid flow complicated is that in addition to static variations in pressure in the fluid – cause by someone pushing on it, or squeezing the sides of a hose, for instance – the motion of the fluid itself can lead to variations in the pressure within the fluid!

So, imagine we have a pipe with a ninety-degree bend in it (and for this section, let's even ignore the effects of viscosity, just to make our life a little easier!). As the water is approaching the bend it is flowing at some velocity, and as it exits the bend it returns to this velocity, but what about as it goes around the bend? As the water starts bending around, it's velocities and pressures start to vary! We can think of this in terms of centripetal acceleration if we like: to change the direction of the velocity of the water, the pipe bend has to exert a force on the water, and the fluid will develop an internal gradient of pressure! On the *outside* of the bend the pressure is the *highest*, and on the *inside* of the bend the pressure is the *lowest*.

But what do we know about fluid flows when there is a gradient of pressure? That the fluid will try to accelerate towards the regions of low pressure! So, at the same time that there are these gradients in pressure, the *speed* of the fluid changes around the bend. Close to the inside of the bend the fluid *speeds up*, and on the outside of the bend the fluid *slows down*. Sometimes it helps to remember these effects by inventing a fictitious "pressure potential energy" for a fluid flow; just as when you throw an object up you are (temporarily) exchanging kinetic energy for gravitational potential energy (and when the object comes back down it is exchanging gravitational potential energy for kinetic energy), fluid velocities slow down as they increase their "pressure potential energy" and speed up as they reduce it.

7.3 Turbulent vs smooth flow

So far we have been thinking about smooth, neat, orderly flow of fluids as they move from place to place. This smooth flow is called *laminar flow*, and it refers to fluids flowing in nice smooth layers: if you were to put two drops of different food color into different parts of a laminar flow, you would see two streaks of color following the "stream lines" of the flow, neither mixing with each other nor getting far away from each other. Laminar flow tends to results when viscosity is dominating the flow, since viscosity tries to stop nearby parts of a fluid from moving very different from each other.

But we all know that flow can be more chaotic than that: when we see rapids in a river, or a stream of water from a faucet roiling around inside a glass, we see *turbulent* flows. Nearby bits of fluid in a turbulent flow will soon be mixed and separated and mixed again with other bits of fluid, and it's all quite noisy and chaotic. Whereas laminar flow is often dominated by viscosity, turbulent flow is typically dominated by inertia: the momentum of each little parcel of fluid carries it around, and viscous forces are not enough to make it change direction or slow down substantially.

Importantly, flows can transition back and forth between laminar and turbulent regimes: as water flows down a river, there might be sections of calm, smoothly flowing water along with section of crushing, turbulent whitewater... followed by more sections of smooth, laminar

Figure 7.2: **A plume of hot air rising from a candle** Close to the candle flame, a plume of hot air rises smoothly (in *laminar* flow) up; a short distance later the variations in pressure and velocity in the air becomes quite chaotic, leading to *turbulent* motion. Image by Gary Settles, using "Schlieren" imaging to detect variations in the temperature and pressure of the air.

flows! Figure 7.2 shows a photograph of the hot air rising from a candle going from laminar to turbulent.

In the rest of the chapter, as we think about how balls fly through the air and how planes stay aloft, we're going to need to think about when fluid flows *around an object* are turbulent or laminar. This is going to depend on a lot of features of whatever system we'll be considering, such as the viscosity of the fluid (high viscosity? more likely to be laminar), the speed of the fluid (high speed? more inertia, and more likely to be turbulent), the density of the fluid (high density? the more inertia a volume of fluid has, and thus more likely to be turbulent), and the size of the object the fluid is flowing around (large object? less likely that viscous forces will keep the fluid flowing smoothly all the way around the object).

How can we combine all of those features into a single number that lets us figure out whether flow will be laminar or turbulent? In fact, at the beginning of this course we met the *Reynolds number*[1],

$$\mathrm{Re} = \frac{\text{density} \cdot \text{object size} \cdot \text{flow speed}}{\text{viscosity}} = \frac{\rho L v}{\eta}.$$

In this equation the "object size," L, refers to a characteristic linear size of the object (like

[1]The Reynolds number, by the way, was introduced by the physicist / mathematician George Stokes in 1851, but named after Osborne Reynolds by yet another physicist, Arnold Sommerfeld, in 1908

the radius of a sphere, or the side edge of a cube). Reynolds found that when this number exceeds about 2300 you have turbulent flow, and when it is below that number you have laminar flow. You can see this simply by stirring a pot of water with a spoon! Take a spoon (which has a width of maybe 2.5 centimeters), and move it through the water at, say, 5 centimeters per second (2 inches per second). The Reynolds number will be

$$\mathrm{Re} \approx \frac{(977\ \mathtt{kg/m^3}) \cdot (0.025\ \mathtt{m}) \cdot (0.05\ \mathtt{m/s})}{10^{-3}\ \mathtt{Pa \cdot s}} \approx 1250,$$

and the flow will be smooth and laminar. Now move the spoon faster, say at 50 centimeters per second: the Reynolds number will be

$$\mathrm{Re} \approx \frac{(977\ \mathtt{kg/m^3}) \cdot (0.025\ \mathtt{m}) \cdot (0.5\ \mathtt{m/s})}{10^{-3}\ \mathtt{Pa \cdot s}} \approx 12500,$$

and the flow will be turbulent! This also gets us used to thinking about different frames of reference: move an object through a fluid can be thought of in the same way as flowing the fluid past the object!

7.4 Aerodynamics

Sports provides one of the most direct ways that many of us see the fun effects of objects moving through fluid – In soccer, how does a striker take a free kick that can curve around a whole wall of defenders? In tennis, why to top-spin shots seem to dive down after crossing the net? In baseball, how do knuckleballs work? In this qualitative section, we'll explore some of the key concepts explaining all of these phenomena!

7.4.1 Slow-moving stuff and laminar airflow

Any object moving through a fluid – a ball flying through the air, or a kayak cutting through the water – experiences fluid-dynamic forces: forces exerted on the object because of its motion relative to the fluid. When the fluid is a gas, these are called *aerodynamic* forces[2] The most important of these are *drag forces* that point "down-wind" of the object and *lift forces* which can point sideways, up, or down (as we'll see soon). In this section I'll stick to talking about objects moving through the air, but many of the concepts have a hydrodynamic analog.

All of these various forces get more and more complicated as the relative speed of the object and the air gets bigger, so we start out by considering a ball moving *slowly* through the air. Furthermore, we'll choose a *frame of reference* moving along with the ball – in this frame of reference the ball looks stationary with the air is moving past it. If the relative motion is so little that the air flow stays laminar, the streams of air cleanly separate and come back together in front of and behind the ball as it travels, and the ball leaves a trail of air behind it – called the *wake* – which is smooth.

As depicted in Fig. 7.3, even in this smooth flow there are still variations in pressure and speed of the air around the ball! The air has to bend around the ball, and bends in fluid

[2] And when the fluid is liquid they're called *hydrodynamic* forces.

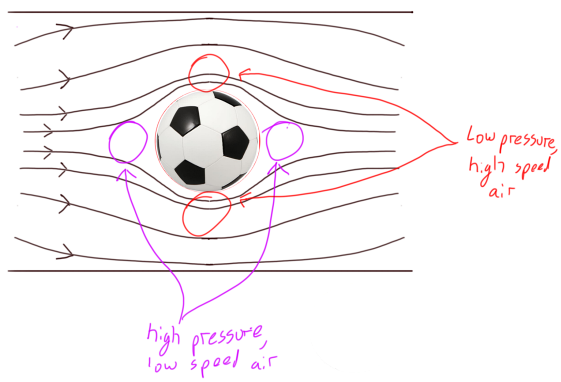

Figure 7.3: **Air flowing slowly past a soccer ball** When air flows slowly over an object it can maintain *laminar* flow. Here, air is shown flowing left-to-right. As air flows directly towards the ball it increases its pressure and slows down, and air immediately on the other side shows the same features. Air around the ball (on the top, bottom, and sides) is at a lower pressure and the air moves at higher speeds.

flow must involve pressure differences (remember: fluid will accelerate towards lower-pressure regions), and these pressure variations occur in the air close to the surface of the ball itself. Air moving directly towards the ball gets bent away, so there must be *higher-pressure*, and slower, air immediately in front of the ball. But instead of getting deflected away from the ball, the air flowing past it bends in towards the ball: this means the pressure near the sides of the ball must be *less than* the pressure of the rest of the air.

This whole chain of logic also works as we follow the air around the sides to the back: the air past the half-way point of the ball starts bending away from the ball again, so there must be another region of higher-pressure air *behind* the ball, just as there was in front of it!

In this simple case, we see that the airflow around the ball is extremely symmetrical: there are higher-pressure regions in front of and behind the ball, and all around the top, bottom, and sides there are lower-pressure regions. All of these pressure forces cancel each other out, and the ball actually doesn't experience any force due to pressure. In this case, the only aerodynamic force acting on the ball is the *viscous drag*. This type of motion is, overall, pretty easy to think about (that is, nothing weird or unexpected really happens), but only the slowest of motions of an object through a fluid actually have this nice, purely laminar character.

7.4.2 Faster-moving stuff and turbulent airflow

Above, with a slow-moving ball, we saw smooth laminar flow of air around the ball, and the result was a very symmetrical pattern of pressure around the ball. As a result, the ball experienced viscous drag (slowing it down), but no other aerodynamic forces. When the Reynolds number gets high enough, as we know, the fluid flow is not laminar but turbulent, and things start to get more interesting.

If we look back at Fig. 7.3, we see that as the air moves towards the back of the ball it sees an area of increased pressure (which is what cause the air to bend away from the ball and back into the laminar stream lines). At high Reynolds number the viscous forces

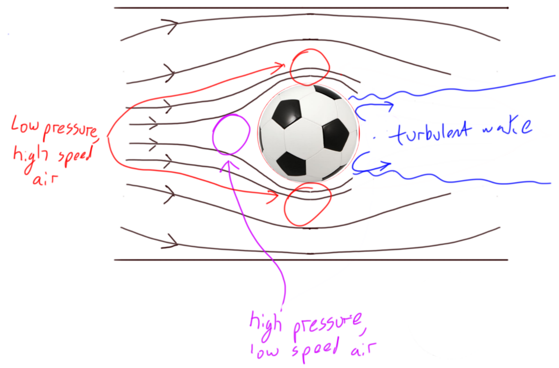

Figure 7.4: **Air flowing slowly past a faster-moving soccer ball** When air flows quickly over an object, with a Reynolds number between $2 \times 10^{-}10^{5}$, the boundary layer stalls part-way around the ball, leaving a large turbulent wake behind it. The asymmetry in the pressure creates a large pressure drag that slows the ball down.

between the layers of air being bent and those right next to the ball (the so-called *boundary layer*) aren't strong enough to guide the boundary layer into the region of high pressure.

Eventually, the boundary layer of air comes to a complete stop in the face of this high-pressure region; increase the Reynolds number even more and it actually *reverses* (as seen in Fig. 7.4. This radically disrupts the flow of air around the ball, leaving a large wake of turbulence behind the ball; in that turbulent wake of air the overall pressure is roughly atmospheric, and the pattern of pressure is no longer symmetrical. The regime we're talking about here are when the Reynolds number is between $2 \times 10^3 - 10^5$: high enough to not be laminar but not into the next regime of very turbulent motion we'll look at in a second. To be concrete, let's put numbers on this in the context of a regulation soccer ball, which has a radius of just about 11 cm. Using numbers for the density and viscosity of air at sea level, a soccer ball moving at 1 foot per second (roughly a third of a meter per second) has a Reynolds number of about 2000, whereas a soccer ball flying through the air at 33 miles per hour (15 meters per second) has a Reynolds number of about 10^5. So: laminar flow only happens for *really* paltry motion by the standards of sports!

What are the consequences of this non-symmetrical pattern of pressures? There is a region of high pressure in front of the ball, and no corresponding region of high pressure behind it to balance it out. The result is a large *pressure drag* on the ball. You can also think of this using ideas about momentum from earlier in the class: some of the momentum of the ball is being transferred to the turbulent motion of the air particles the ball is leaving in its wake. How big an effect is pressure drag? Roughly speaking, the pressure drag is typically proportional to the cross-sectional area of the turbulent wake (which is typically, for moderate speeds considered in this section, roughly the cross-sectional area of the ball itself), and it proportional to the *square* of the speed of the ball through the air:

$$\text{pressure drag} \propto (\text{cross-sectional area}) \cdot (\text{speed})^2.$$

Even more turbulent airflow around an object

To recap, at low Reynolds numbers the whole flow of air around a ball is laminar, and even the wake behind the ball is just a small wake of smoothly flowing air. In this case, the main drag force is *viscous*. At intermediate Reynolds numbers ($2 \times 10^{-}10^{5}$), the boundary layer of air around most of the ball is still laminar, but it can't make it all the way around the ball. The result is a large turbulent wake, which creates a large *pressure drag* that slows the ball down.

At even higher Reynolds numbers, the boundary layer *itself* becomes turbulent. The result is that fast moving balls leave behind *smaller*, more focused turbulent wakes behind them than when the same ball is traveling more slowly. This effect is not small!

Let's put some numbers in: for a regulation baseball the transition point of Re $\approx 10^5$ occurs when the ball is traveling at about 100 miles per hour: faster than (most) pitches but slower than when a baseball leave the bat. The large pressure drag might take a fastball pitched at 90 miles per hour and slow it down closer to 80 miles per hour in the 60-foot distance between the pitcher and the batter. In contrast, after being struck the baseball has a speed so large that the boundary layer of air around it becomes turbulent, and it experiences relatively less pressure drag as a result. If not for this effect – that is, if turbulent boundary layers of air didn't lead to less pressure drag, every potential home run with quickly stall out and plop down somewhere in the outfield.

This same effect, by the way, explains why there are dimples on golf balls! One way to create a turbulent boundary layer that helps reduce the pressure drag on a ball is simply by moving the ball fast enough, but another is to make the surface of the ball itself bumpier! The little dimples on a golf ball make it harder for the air flows to stay laminar: this actually increases the total viscous drag on the golf ball, but that increase is more than compensated by the potential for much less pressure drag after the golf ball has been hit.

7.4.3 Curveball!

So far we have looked at pressures in front of and behind a ball moving at various velocities, but in the other directions (above, below, and to the sides) the pressures have been symmetric. These balanced pressure forces meant that the ball didn't start moving in any direction, but we've all seen a pitcher throw a curve ball, or a soccer play curve a free kick – how does that work? We perhaps have an intuition that spin is involved, but if a ball is spinning which way does it curve? Towards the spin? Away from it? In a different directional altogether?

Curveballs make use of another aerodynamic force, *lift*, where an object deflects air asymmetrically, or creates asymmetric pressures around it, in such a way to generate forces in a direction different from simply opposing the direction of the object's motion. I think it's pretty cool – it's not immediately obvious how we get air to push sideways on an object moving through it, after all!

Curveballs operate by two primary mechanisms. The first of these mechanism works regardless of whether the air flow is turbulent or laminar, and it's called the *Magnus force*[3]. Even in laminar flow, the idea behind the Magnus force is that a spinning object helps move

[3] After H. G. Magnus, a rare scientist better known as a teacher than for his research, who nevertheless had something named after him.

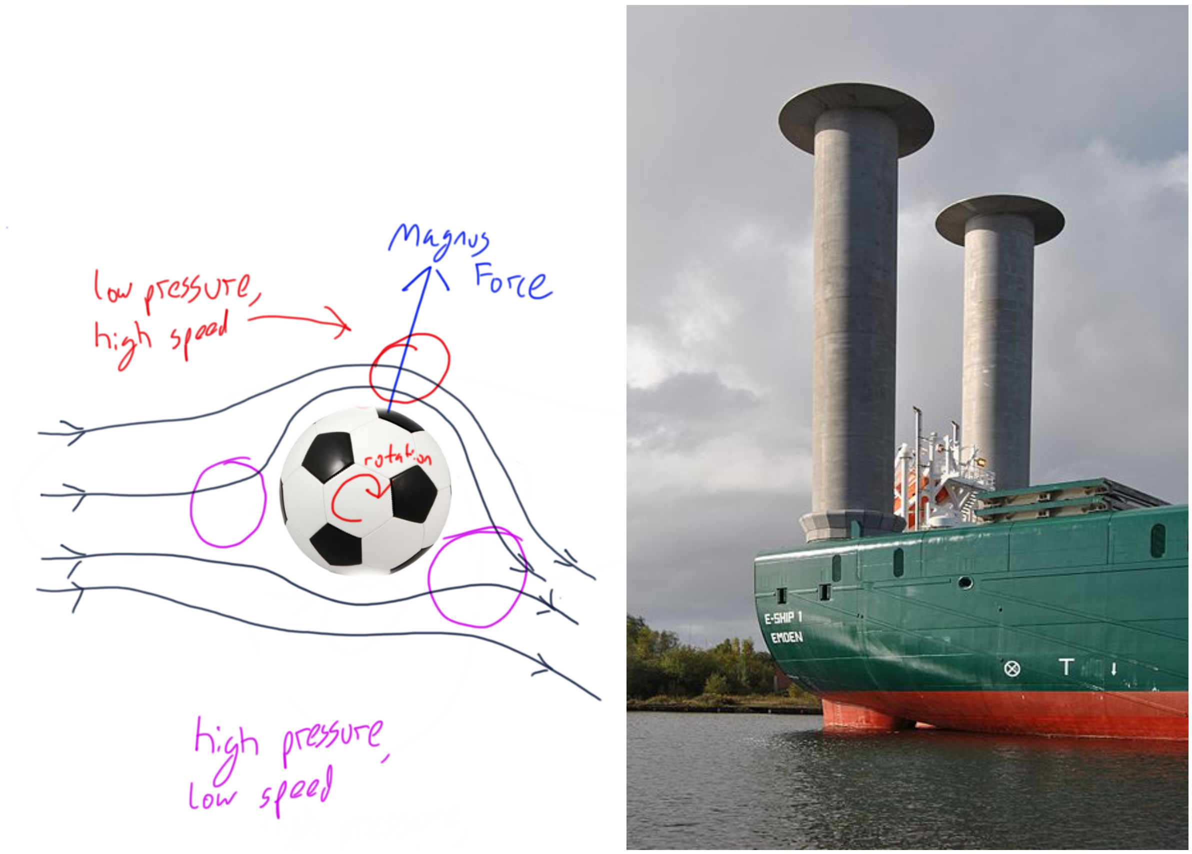

Figure 7.5: **The Magnus force in action** (Left) A rotating object moving through a fluid makes creates imbalances in pressure, leading to a Magnus force on that objects due to the pressure asymmetry. This effect is present in both laminar and turbulent flowing fluids (Right) Bizarrely, this same principle can be used to make a very funny-looking "sail boat!" Rather than having sails, one can put giant rotating cylinders that stick up above the ship. When wind blows past the rotating cylinders, the resulting Magnus force can be used to propel the "rotor boat"!

some of the boundary layer of air around with it. This means that the flow pattern around a moving, spinning ball will be asymmetric (see Fig. 7.5). If we think about the paths the air takes as it flows around the ball, in the "long way" the air must really bend around the ball, and much of the necessary change in direction involves turning in towards the ball – just as in our examples above, this corresponds to a region of lower pressure. In contrast, the air following the "short way" around the ball spends most of its path turning away from the ball, and this corresponds to regions of higher pressure. Put together, this asymmetry leads to a net force pointing towards the low-pressure region!

While the Magnus force is the only "lift" force that operates in laminar flow, when the flow is turbulent there is another aerodynamic force that can contribute to the curve. Remember that at large Reynolds numbers the ball is leaving a turbulent wake behind it; a rapidly spinning ball can *deflect* this wake! When the ball isn't spinning the wake "detaches" from the boundary layer symmetrically, but when the ball spins the wake doesn't detach evenly on all sides: it detaches a little layer for air moving with the spin and a little sooner for air moving against the spin. This leads to a deflection of the wake of turbulent air behind the ball, and this "wake deflection force" makes the ball curve *in the same direction as the Magnus force.*

So, both of these aerodynamic forces tend to push the ball, making it curve "in the direction of the spin:" if you are looking top-down on a ball and it is spinning clockwise it will curve to the right, and if it's spinning counterclockwise it will curve to the left. But you could also impart top-spin and back spin: top-spin creates aerodynamic forces that make the ball curve *downward* (keeping your very fast-moving top-spin forehand shot inside the tennis court) and back-spin creates aerodynamic forces that make the ball curve *upward* (typically not actually making it lift up – these forces tend to be weak relative to the acceleration due to gravity for sports balls – but making it seem to "hang" in the air more than you would expect).

In case you're curious: "knuckleballs" are actually characterized by their *lack* of spin as they move through the air. For a baseball, this starts to highlight the importance of the seams of the baseball, as they can trigger transitions to turbulent flow for small parcels of air around them. These small regions of turbulence lead to a bunch of erratic deflections acting on the ball – hard for a pitcher to control, but also hard for a batter to predict!

Chapter 8

Heat and and insulation!

Having discussed fluids, in this chapter we're going to turn towards the physical behavior of *heat*! We started giving clues about the nature of heat and temperature in the last chapter, when we talked about the temperature of a gas being related to the "internal kinetic energy" of the gas molecules, but what is heat? How does it flow from one object to another? *When* does it flow from one object to another?

We'll discuss these questions in this chapter. As a fun historical note (and one of the reasons it is particularly amusing to follow chapters on fluids with chapters on heat), starting in the late 18th century scientists believed that "heat" was a type of fluid: a self-repelling invisible gas called "caloric" that could flow in and out of pores in objects and which carried heat with it. There was much debate about whether "cold" was a separate fluid (called "frigoric"), or whether it was simply the absence of caloric in a substance. The self-repelling nature of caloric helped scientists of the day explain many observations[1], although a more modern understanding of heat (and how it is definitely not an actual substance!) superseded it with Lord Kelvin's publication, "On the Dynamical Theory of Heat[2]."

8.1 Stoves

It feels like a distant memory now that I live in Atlanta, but I grew up in a much colder part of the country and the importance of efficiently heating a home (or the pleasure of standing in front of a roaring fire) has been a part of my life since I was born[3]. But what is actually going on when you're sitting in front of a fire? What is producing the light we see and the heat we feel, and why do we feel the heat of the fire even without touching it[4] and even though the air in between us and the fire might still be cool?

[1]For instance: why does my hot cup of coffee cool down and lose heat to the air in the room? Because, the followers of caloric theory would say, the self-repelling nature of caloric made it want to flow from regions of dense caloric (the hot coffee) to regions of less dense caloric (the colder air).

[2]Sounds like a page-turner, no?

[3]Family legend has it that when I was born it was -40°, although now that the internet lets us look up historical weather records easily that's easily disproved...It was, nevertheless, a very cold winter's day!

[4]And please: do not touch fire.

8.1.1 Thermal energy in solids (aka, "stuff on fire")

We've encountered several examples of thermal energy in substance, and of heat being generated: the thermal energy of a gas was associated with the internal kinetic energy of the gas particles flying through the air at high speeds (even though the gas might not have any overall motion of its center of mass), and the sliding friction of a bookcase across a floor generated heat as it made us continue to exert forces to continue to displace the already-moving object.

Thermal energy is, more generally, the portion of the *internal energy* of an object or substance that is associated with temperature. The can be contrasted with energy associated with external forces: when we lift an object up into the air we increase its gravitational potential energy, but this is energy associated with an external force (in this case, the force due to gravity). When we wave a baseball bat back and forth we are exerting forces to give it kinetic energy, but this kinetic energy does not affect the internal energy of the bat.

I say "portion of internal energy" above because *most* internal energy of an object is *not* associated with temperature: the atoms and molecules of a substance store an enormous amount of both chemical potential energy (in the form of chemical bonds that keep molecules together) and nuclear potential energy (keeping the atoms themselves together)! It turns out that the chemical bonds keeping molecules together act kind of like springs: the atoms in a molecule have a certain distance away from each other they like to sit – bring them closer and they feel a repulsive force, bring them farther apart and they feel an attractive force. Thermal energy in solids is made up of the internal kinetic energy of these *atoms oscillating back and forth*, both within a single molecules and in the way that atoms on different molecules will want to be arranged in particular patterns with respect to each other.

In a solid the atoms want to sit in particular positions relative to each other, but when they're oscillating and vibrating more and more due to their thermal energy they tend to be slightly farther apart from each other compared to when they have less thermal energy (that's because the forces an atom experience aren't "symmetrical" – the repulsive force when atoms are brought too close is stronger than the attractive force when they're too far apart). Thus, as an object's temperature rises the object gets a little bit bigger in all directions. This is sometimes described by the *coefficient of thermal expansion* of a material, which measures the fractional change in an objects volume per change in temperature. As an example, iron has a coefficient of thermal expansion of about 3.6×10^{-5} per degree Kelvin. What does this mean? It means for every degree you heat a chunk of iron, the volume of that chunk grows to 1.000036 of the volume it was before (equivalently, if you heat a chunk of iron up by 50 K, its volume will increase by about 0.18%).

So, similar to the internal kinetic energy that goes into "temperature" in a fluid, in a solid we have this disordered dance of particles vibrating around contributing to the thermal energy of, say, a burning log. When two objects have different amounts of this internal kinetic energy, the jiggling motion at the surface of one tends to disproportionally bump the molecules and atoms of the other object, and the disordered internal kinetic energy of the "hotter" object tends to flow towards the cooler object.

We'll think about the three primary mechanisms of heat flow in just a moment, but first, you might wonder: what is actually going on when a piece of wood is on fire? The answer is that when a piece of wood is *not* on fire the molecules in the wood are storing a large amount of energy as chemical potential energy, and much of this is in the way that carbon and hydrogen atoms are bound together in the cellulose[5] in the wood. The thing is, though, that carbon and hydrogen can form much tighter chemical bonds not with each other but with the oxygen in the air around us. Because the carbons and hydrogens are bound together, though, it's hard for them to spontaneously come apart and find a nearby oxygen molecules.

That's where matches come in! The heat from a match is enough to break some of the existing chemical bonds, and the hydrocarbons break up, fusing with the oxygen to become a combination of water and carbon dioxide. Because the new chemical bonds are much tighter, the hydrocarbons release a lot of what they were formerly storing as chemical potential energy as *heat*: disordered thermal energy that flows into nearby atoms and molecules. This heat is enough to trigger the breakup of nearby hydrocarbons, and the whole process can sustain itself, throwing off heat as more and more cellulose burns up (until it's all gone, of course).

8.1.2 Warming an object up

Before getting into the details of how heat can move from one object to another, let's make sure we can characterize how the temperature of an object *changes* when heat flows into or out of it! When we add some amount of thermal energy to a substance, how much does it change the temperature of that substance? The *heat capacity* of an object refers to how much thermal energy you need to transfer to an object in order to raise its temperature by 1 degree. This depends both on the material something is made of, and also how much of that material the object has! If you have two differently-sized cast iron skillets, it takes more thermal energy to raise the temperature of the big on than the small one.

Thus, it is sometimes easier to characterize materials by their *specific heat*:

$$\text{specific heat} = \frac{\text{heat capacity}}{\text{mass}}.$$

Specific heat has dimensions of energy per kilogram per temperature, and the SI units for it are $\mathtt{J/(kg{\cdot}K)}$, the "joule per kilogram per Kelvin. Metals tend to have *relatively small* specific heats, which corresponds to being easier to heat up. For instance, copper has a specific heat of 386 $\mathtt{J/(kg \cdot K)}$, whereas water has a specific heat[6] of about 4190 $\mathtt{J/(kg \cdot K)}$, and air at constant (atmospheric) pressure has a specific heat of about 1001 $\mathtt{J/(kg \cdot K)}$.

Let's think about that last number in the context of heating up a room. The volume of a dorm room at Emory might be, say, 30 $\mathtt{m}^3$, so if you want to raise the temperature of your room by a single degree you need to heat up all of the air inside (about 37.6 $\mathtt{kg}$) of it by one degree, which the specific heat tells us takes takes about 38000 $\mathtt{J}$ of energy. If it's a cold winter's day in Atlanta and you open your window (who knows why you would do such a thing, but letting in a blast of cold air in the process), the cost of raising the temperature in the room back from $-5°$C up to $20°$C after you close the window is almost a million Joules! Heating costs can add up quick!

[5]Unlike what I said in the lecture, the actual chemical formula for cellulose is a repeating unit of $(C_6H_10O_5)$

[6]Notice how this number looks suspiciously similar to the conversion factor between Joules and food calories!

8.2 How heat flows

Coming back to our wood stove, or our burning piece of firewood: the burning log (or the glowing embers) are really hot – hotter than the air around us – how does the heat actually move from one place to another?

8.2.1 Conduction of heat

Conduction is the flow of heat through an object, or through to objects that are physically in contact. The idea here is that the *heat* moves from hotter to colder, even though all of the atoms and molecules involved stay in place! Consider putting an iron poker into a roaring fire. The fire (with its fast-moving particles) bombards the iron atoms at the surface of the poker, transferring tiny amounts of kinetic energy every time there is a collision. Eventually the iron atoms in the tip of the poker have picked up enough internal kinetic energy from this bombardment that they are vibrating quite a bit about their *equilibrium positions*. Now, the atoms in the hot tip of the poker are vibrating and oscillating more than the atoms they are sitting next to, and they start push against the less-oscillating atoms nearby. Every time they do this, they are doing tiny, microscopic amounts of work on the nearby atoms, and those atoms start vibrating more and more, too.

In this way, atom-by-atom, heat is *conducted* through the solid object. Metals are particularly thermally conductive because they have an additional channel by which they can conduct heat. In a metal, there are *mobile electrons* that are not bound to particular atoms, but can instead travel pretty freely throughout the metal – this is why metals conduct electricity so well, actually! Well, these free-traveling electrons can pick up some energy from randomly oscillating atoms, but instead of transmitting that thermal energy *only* to a neighboring atom, they can travel great distances within the metal before depositing thermal energy. For the most part, substances that conduct electricity well also conduct heat really well[7]

If you're cooking on a stove, conduction of heat through the frying pan to the food it is directly in contact with, and then through the food itself, can be a great way of moving heat around. But if you're trying to, say, heat up a room with a fire, most of the room is emphatically *not* in direct contact with the fire[8].

8.2.2 Convection of heat

Convection of heat, our next mechanism of heat transfer, refers to heat that is transported by a moving fluid. Consider a wood-stove in a room: the hot metal of the wood stove comes into contact with air (a fluid!), and heats up that close-by air. Because hot air is less dense than cool air, it is lifted by the buoyant force upwards towards the ceiling. Cooler air from somewhere else in the room has to flow towards the stove to replace the rising hot air, and a circulating convection current can eventually form. In this way, heat from the stove is carried

[7] With a few exceptions: because of the details of their chemical bonds and the unusual lattice structure of their atoms, diamonds are surprisingly good conductors of heat even though they're terrible conductors of electricity. This fact is sometimes used by jewelers to distinguish real diamonds from fake!

[8] If it is, Emory's emergency fire safety number is 404-727-6111

by the moving air currents to other parts of the room. On its own, most of the heat convected in this way goes towards the ceiling, but if you add a ceiling fan to blow the still-warm air downwards you can nicely heat up the whole room!

This might seem a bit... bland? But guess what – we've just explained where wind comes from!! When the sunlight shines down on the Earth, it heats up the ground. It also heats up the water, but because of the reflectivity of water and the way water interacts with sunlight, the water gets warmed up by less than the land. The ground warms up the air above it, that warmer air starts to rise, and it is replaced by the relatively cooler air that is above the water. Wind, then, is just a bonkers, giant set of convection currents that come about because of solar heating, with surface winds blowing towards warmer spots that are creating more areas of rising air. The precise details are complicated and messy, but that's the underlying physics!

One thing I kind of glossed over in that paragraph, though: how is the sunlight warming up the Earth in the first place?

8.2.3 Radiation

In addition to bumping into each other as atoms vibrate and wiggle around, they can also both emit and absorb *electromagnetic radiation*. We'll talk more about this radiation at the end of the semester, but for now we can hold onto the idea that it is composed of any time of electromagnetic wave, which includes radio waves, microwaves, visible light, ultraviolet light, and x-rays – those are all basically the same physical type of thing, distinguished by a property called wavelength. More on that in a few chapters!

So, objects can emit this EM radiation, and they can also absorb it, and that gives a new mechanism for transferring heat *across space*, without objects being connected (either from being directly in contact, or from a moving fluid between them). The type of EM waves emitted by an object depend on its temperature: cold objects emit radio waves, microwaves, and some infrared light; a hot object emits all of these plus visible or even ultraviolet light! When a hot coal starts to glow red that's a sign that it is hot enough to start emitting EM radiation in the visible part of the spectrum of EM radiation. The amount of radiation emitted also depends on the temperature: hot objects through off more EM radiation in addition to throwing off a broader spectrum of radiation. The amount of radiation *absorbed* depends on many things, too: shiny metals tend to reflect rather than absorb incoming EM radiation, whereas black objects are really good at absorbing incoming radiation.

Our eyes can only see a small portion of the electromagnetic spectrum[9], but even objects that aren't visibly glowing are constantly emitting and absorbing electromagnetic radiation. When two objects of different temperatures share a "line of sight" they are each being bombarded by the thermal radiation from the other. But since the hotter object is throwing off more radiation, on average the heat tends to flow from the hotter object to the colder one.

As you stand in front of a fire, a quick way to tell if the heat you feel on your face is coming from convection (with the fire heating the air which then flows towards you) or

[9] Different animals can see different, sometimes much broader portions of this spectrum! Here's an article that gives a few examples of how other animals see "color." If that article is the first time you've hear about mantis shrimp, prepare to go down a fascinating rabbit hole as you learn just how bizarre they are!

radiation (with the fire pelting you with EM radiation), is to put your hand in front of your face: if your face suddenly feels cooler it's because less of this radiative heat is reaching your face.

8.3 Insulation

The flip side to talking about heat and how heat moves from one object to another is asking, "*how can we stop heat from flowing between objects?*" When we want to keep a mug of hot coffee hot, what can we do to slow down the flow of heat away from the coffee? What are the limits on how well we can insulate ourself with our clothing during the winter, or our houses, or (for that matter) the Earth?

It turns out to be impossible to *completely* isolate an object, so that it doesn't gain or lose any heat whatsoever to the objects or the environment around it, but in the following sections we'll look individually at how the three different mechanisms of heat flow can be slowed down. In the process we'll understand a bit more about how those different mechanisms behave in the first place. A few general words about insulation, first, though.

A first point is that in some senses, losing energy as thermal energy is an unavoidable waste[10] – as we run a field our body converts some stored food energy into kinetic energy, but loses a lot of it to heat. We burn gas trying to drive our cars from point A to point B, and as we cruise on the highway a lot of the energy we're expending is getting turned into heat as we fight against pressure drag slowing the car down.

Any time heat leaves an object that needs to stay at some temperature in order to function – like say, every warm-blooded mammal! – energy needs to be expended just to replace the heat that was lost. For instance, what about a human, and the energy from food we need just to maintain all of the biological processes in our body? You probably know that a person needs in the ballpark of 2000 food calories per day, which is about 83 food calories every hour: converting that to SI units,

$$83\ \texttt{kCal/hr} \approx 96\ \texttt{W}.$$

That is, even when not doing anything, a person is constantly generating about 100 W, warming up the room around them like a space heater, and it turns out that human physiology is much better at regulating how it releases this heat into the environment than it is at regulating the amount of thermal energy it generates in the first place. Hence: the human body stays very close to 37°C all the time, and in hot weather your body tries to encourage heat loss (e.g., by evaporative cooling of extra sweat production) and in cold weather the body tries its best to generate a little extra heat (by shivering, enhancing the base level of muscle motion, in an effort to maintain body temperature when losing a lot of heat to cold surroundings).

But the easiest way to stay warm is by putting on extra clothing. That, and other strategies at providing thermal insulation, is what we'll talk about next.

[10] At least from a human perspective, where we're trying to use energy to accomplish various aims – I doubt the universe itself assigns a value judgement to the way energy flows between different forms.

8.3.1 Limiting thermal conduction

As we learned above, heat flows through a material by conduction whenever one part of the material is hotter than another part of it (or if it is in contact with a different object at a different temperature). How fast does heat flow by this mechanism, though?

The answer to that question first depends on the material itself – heat simply flows better through some materials (like metal) than it does through others (like, say, glass). We characterize this by the *thermal conductivity* of a material, which has SI units of watts per meter per kelvin ($\mathtt{W/(m \cdot K)}$). Some typical numbers for thermal conductivity would be around 310 $\mathtt{W/(m \cdot K)}$ for gold, 50 $\mathtt{W/(m \cdot K)}$ for steel, 0.8 $\mathtt{W/(m \cdot K)}$ for window glass, 0.21 $\mathtt{W/(m \cdot K)}$ for human skin, 0.17 $\mathtt{W/(m \cdot K)}$ for oak wood, and 0.025 $\mathtt{W/(m \cdot K)}$ for air.

As you can see, that's a very wide range! What else does conductive heat flow depend on? It also depends difference in temperature between the hot and cold parts of an object (or objects), the area of contact, and the separation between the hot and cold parts. So: if you're standing in socks on a cold floor 10°C colder than you are vs on a very cold floor 40°C colder than you are, you'll lose heat four times as fast on the very cold floor. If you stand on one foot rather than both feet, you'll cut your heat loss in half[11]. And if you make your sock twice as thick you'll cut your heat loss in half. Putting all of these factors together:

$$\text{heat flow} = \frac{\text{thermal conductivity} \cdot \text{temperature difference} \cdot \text{area}}{\text{separation}},$$

or[12]

$$H = \frac{k \cdot \Delta T \cdot A}{d}.$$

8.3.2 Limiting convective heat loss

The human body does a surprisingly good job at minimizing conductive heat loss[13] But ultimately, the human body is warmer than the air around it (just about everywhere on Earth, anyway), and so it will lose heat to the air it's in contact with.

However: air itself is a very poor conductor of heat (check out it's quite low value of conductivity above!), and so if you're staying still you actually only really heat up a thin layer of air around you. If you could get that air to stay still, you wouldn't lose very much heat at all via conduction to the air, because the temperature of the air near your body would be almost the same as the temperature of your body. Of course, we know that warmer air rises, and so this air warmed by your body naturally rises away from you or, even worse, perhaps you're outside and a wind blows air not-yet-warmed by your body towards you,

[11] At least, until your leg gets tired, I suppose

[12] Again, you'll start to note that as we meet more and more physical relationships, we have to start using the same letter to stand for different physical quantities. It's an important reason to learn the *ideas behind the formulas* rather than just memorize the formulas themselves!

[13] It does this by a few very neat tricks. Both skin and fat (stored just under the skin) are poor conductors of heat, so the heat from the body's core takes a long time to get to the skin. Our body shapes are pretty compact (within the limits set by needing arms and legs to interact with the environment, we're a lot closer to the shape of a sphere – very good at minimizing surface area! – than to a flat pancake). And the body even tries to minimize the temperature difference between the skin and the air by having our hands and feet be a bit cooler than our core temperature. Clever.

replacing the air you just warmed up! This is why the outdoors feels colder on a windy day – your body loses heat more rapidly in the face of wind – and people have tried to calibrate "wind chill" scales to account for this.

So, what to do to combat this convective heat loss – the heat lost to the air our body warmed up by conduction as the newly-warm air moves away? Evolutionarily, mammals have covered themselves with hair or fur. The dense tangles of fibers that compose, for instance, a flemish giant rabbit's fur act to *trap* pockets of air next to the rabbit's skin. This air gets warmed up by conduction, and it only escapes away from the rabbit's body very slowly, with very little flow of air. So, the rabbit's warm skin is surrounded by a jumble of fur and trapped air: both of those materials conduct heat poorly, and the rabbit stays warm in the winter.

You have, perhaps, noticed that humans do not have so much fur or hair: instead, we can wear clothes. The real (insulating) function of clothes is that just like a rabbit's fur they can trap air close to our bodies, reducing the convection of warm air away from us. The same idea is used with wetsuits when swimming in cold water: the material of the wetsuit first absorbs and then traps a layer of water against the swimming, and once the swimmer's body heat has warmed up that layer of water, the wetsuit's design keeps that same warm water from diffusing away.

8.3.3 Limiting radiative heat loss

As mentioned above, objects are constantly emitting and absorbing electromagnetic radiation. The amount of thermal energy an object radiates away per unit time (that is, the thermal *power* it radiates away) can be expressed by the Stefan-Boltzmann[14] law:

$$P = e \cdot \sigma \cdot T^4 \cdot A,$$

or, in words, the thermal power radiated by an object is equal to its emissivity (e) multiplied by the Stefan-Boltzmann constant[15] (σ) multiplied by the fourth power of the absolute temperature (T^4) multiplied by the surface area of the object (A). The constant is

$$\sigma = 5.67 \times 10^{-8}\ \mathtt{J}/(\mathtt{s} \cdot \mathtt{m}^2 \cdot \mathtt{K}^4),$$

and the emissivity is a property of the surface of the object. It's values range between 0 and 1: an object with $e = 1$ absorbs all light that hits it and emits light as efficiently as possible. Because we perceive objects that absorb all incoming light as the color black[16], these ideal thermal radiating surfaces are called "ideal blackbodies." Human skin is not so great at absorbing or emitting thermal radiation in the visible part of the EM spectrum, but it is surprisingly good at doing so in the infra-red part of the spectrum (as, in fact, are most non-metallic objects!), and the emissivity of human skin is about $e = 0.97$.

[14]We already met one of Ludwig Boltzmann through his constant in the ideal gas law. Josef Stefan was an Slovenian physicist of the same era.

[15]Whose value in SI units is secretly easy to remember: "five point six seven times ten to the minus eight!"

[16]Here's a link to a youtube video about "Vantablack," a recently invented surface coating which absorbs very close to 100% of incoming visible light. Looking at it, like staring into the void, is spooky.

Notice that T^4 in the Stefan-Boltzmann law: that's a power of absolute temperature that grows very quickly! That explains why radiative heat transfer can be so noticeable, as we feel the radiated heat from the fire on our face, or from the sun hitting us with its rays! Let's plug in some numbers for a typical human, whose skin temperature is about 306 $\mathtt{K}$ and which might have a surface area of about 1.6 $\mathtt{m}^2$:

$$P_{human} \approx 0.97 \cdot \left(5.67 \times 10^{-8} \ \frac{\mathtt{J}}{\mathtt{s} \cdot \mathtt{m}^2 \cdot \mathtt{K}^4}\right) \cdot (306 \ \mathtt{K})^4 \cdot (1.6 \ \mathtt{m}^2) \approx 772 \ \mathtt{W}.$$

So, if you were to suddenly find yourself in the vacuum of space, even if there was nothing touching you to conduct heat away, you would be radiating thermal energy at almost 800 $\mathtt{W}$ – at that rate, you'd quickly get quite cold, indeed! The first thing that helps us from freezing to death here on Earth is that yes, we're emitting a lot of EM radiation, but we're also *absorbing the radiation* emitted from other nearby objects that are at similar temperatures!

Although there's not much we can do to prevent ourselves from radiating this thermal power, insulation can still help us keep it in! A heavy coat next to our body will absorb a lot of the heat our body radiates, and as long as the coat is a poor conductor (i.e., a good insulator) the part of the coat exposed to the outside air will not be at as a high a temperature and the coat won't radiate so much thermal energy away to the environment.

Another thing we can play with is that *emissivity* factor in the Stefan-Boltzmann law. Just as an "ideal blackbody" is the perfect emitter or absorber of EM radiation, an perfectly white[17] (or shiny) object would *reflect* all of the light that hit it – it would not absorb any thermal radiation, and it also wouldn't emit any!

Shiny metals tend to have very low emissivities in all relevant parts of the EM spectrum, so if you wrap a hot object in a shiny metal you can radically reduce the object's effective thermal emissivity. This is the principle behind wrapping, say, hot food in aluminum foil: even though aluminum conducts heat from the food extremely well, it *reflects the radiation from the food back at the food*! This is also why emergency rescue blankets are made out of metallic cloth: wrapping yourself in such a blanket (mostly) stops you from exchanging thermal radiation with the rest of the world around you.

In contrast, transparent materials *also* have extremely low emissivities (quite close to zero, at least in the part of the spectrum they are transparent to), but whereas a shiny material reflects incoming radiation, a transparent material simply lets it through. You might have noticed when sunlight streams through a window onto a patch of the floor: the beam of light can heat up the floor even as the window stays cool!

[17] This is another place where the fact that we can only see part of the EM spectrum deceives our intuition a bit: white clothing still actually does a great job of emitting and absorbing infrared light.

Chapter 9

Thermodynamics!

In the last chapter we learned about the three primary mechanisms by which heat flows, and that when left to its own devices heat will naturally move from hotter to colder objects. The technology that powered the industrial revolution was all based on the idea that we can *actively manipulate* the transfer of heat: a steam engine could not only move heat from a heat source to a cold reservoir, but do useful work in the process! Similarly – and now that I live in Atlanta *extremely importantly* – by doing enough work, an air-conditioner can *reverse* the natural flow of heat, transferring heat from colder object to hotter objects!

We've gotten used to the idea that thermal energy is another form of energy, and that energy can be converted from one form to another, but what are the rules of this game? What determines how heat can flow not only naturally (as in the last chapter), but when we're also doing or receiving work from that flow of heat? This subject – governing the "movement" of heat, is called thermodynamics[1], and is one of the true cornerstones of modern science!

In this chapter, we'll first meet the laws of thermodynamics (just as we met the laws of motion in Chapter 2); we'll then see what those laws tell us about everything from the engines in our cars to the air conditioners keeping us comfy.

9.1 The laws of thermodynamics!

Without an air-conditioner to keep our homes / apartments / dorms comfortable in the Southern summers, the air inside would be uncomfortably hot (and humid): heat from the outside would naturally flow into our rooms, and it wouldn't stop until the inside was the same temperature as the outside. Gross.

So, air conditioners work (as we'll see in an upcoming section) by actively pumping heat out of your house, by why is that necessary? Aren't there other things we could do to cool our house down? For instance, you might be thinking, why not

[1] "Thermodynamics is a funny subject. The first time you go through it, you don't understand it at all. The second time you go through it, you think you understand it, except for one or two small points. The third time you go through it, you know you don't understand it, but by that time you are used to it, so it doesn't bother you any more." – Arnold Sommerfeld, As quoted in: J.Muller, Physical Chemistry in Depth (Springer Science and Business Media, 1992)

1. Let heat flow directly from my house to my neighbor's[2] house?
2. Destroy some of the thermal energy in my house?
3. Convert some of the thermal energy in my house to something more useful, like electrical energy?

Unfortunately[3], the laws governing thermal energy rule out all of these possibilities. Let's learn why!

9.1.1 The "zeroth" law of thermodynamics

Historically, the people used to talk about the three laws of thermodynamics, but realized there was an assumption they were making that seemed so obvious they didn't think to write it down, and it relates to our first scheme for cooling down our house above.

The problem with that scheme is the following: our house is in *thermal equilibrium* with the outside air, which means that when our house is the same temperature as the outdoors, no heat flows between the outside and the inside of our house. Our neighbor's house (in the absence of air conditioning) is also in thermal equilibrium with the outside air, and we have

> **The Zeroth Law of Thermodynamics says** If object A is in thermal equilibrium with object B, and if object B is in thermal equilibrium with object C, then objects A and C are also in thermal equilibrium.

In other words, thermal equilibrium is a *transitive* relationship. That might feel obvious, but there are plenty! of non-transitive relations that[4] we[5] know[6] about[7].

But thermal equilibrium is, indeed, a transitive relationship, and that means there is some quantity we can measure – some universal scale we can use – to judge the degree to which things are or are not in thermal equilibrium with each other. This is why we have a basis for meaningfully talking about "temperature" as a judge of the thermal energy of an object: when two objects are at the same temperature, even if you give heat the opportunity it will not spontaneously flow from one to the other.

So, sadly, our first scheme won't work: our neighbor's house is the same temperature as ours, which means our houses are in thermal equilibrium, and we can't just hope that heat will spontaneously flow from our house to their's.

---

[2] Sorry, Dave!

[3] Or, fortunately for Dave

[4] Like "*like*": I like pecan pie, my arch-rival likes pecan pie, I don't like my arch-rival

[5] Like "rock-paper-scissors:" rock beats scissors, and scissors beats paper, but rock does not beat paper.

[6] Like the food chain: A heron eats a small fish, and a small fish eats plankton, but a heron does not eat plankton.

[7] Like restaurant preferences: On a given day when faced with 1-1 choices I might prefer to go out for sushi rather than Grandma's noodles, and Grandma's noodles rather than chaat, while also preferring to go out for chaat rather than sushi. Choosing a restaurant can be hard!

9.1.2 The first law of thermodynamics

You were (probably? hopefully?) skeptical of the second scheme "destroy some thermal energy" from the beginning – we learned that energy is conserved, after all! So, no luck for us when it comes to cooling our home by "destroying" thermal energy, our only choice will be to either move it somewhere else or convert it into another form.

> **The first law of thermodynamics** says that *heat is a form of energy*, and if you consider a stationary object, the change in that object's internal energy is equal to the heat added to the object minus the work done by the object on its environment.

So, if we heat up an object its internal energy increases, and if we do work *on* an object its internal energy increases. In symbols we'll often use Q for heat and W for work, so

$$\Delta U = Q - W.$$

Note that we have written this so that when the *object* does work W is positive, and when work is done *on* the object W is negative.

The first law, in some ways, just makes explicit that heat is a form of energy, and so we can move energy around either by doing work or transferring heat. But we can't just destroy energy.

9.1.3 The second law of thermodynamics

Our third scheme does seem quite so harebrained as the first two, at least at first glance: sure, we can't destroy energy, but we've already seen that we can change energy from one form to another, so why not just find a way to transform thermal energy into some more useful form of energy?

Sadly, things are not so easy. Thermal energy is encoded in the disorderly, erratic motion of atoms within a substance, and converting that into neat, ordered, useful forms of energy is challenging and rarely happens spontaneously. When we accidentally knock over a wine glass we see it shatter into a disordered collection of tiny shards, but why don't we ever see the shards just spontaneously reassemble themselves? There's nothing in Newton's Laws that gives us the answer to this – play a video of two billiard balls colliding in reverse, and it still looks like a system governed by Newton's Laws!

In fact, such processes, whereby random motion leads to an ordered outcome is possible. If you waited long enough it could happen that the wine glass would suddenly reassemble, or that a burned piece of firewood would have all of those chemical reactions reversed and the log would be back to its original self... but you would have to wait much, *much*[8] longer than the age of the universe before you saw that happen. A safe word for such fantastically unlikely events is "impossible"!

The idea here is that turning order into disorder is easy, but *recovering* order from a disordered state is essentially impossible – at least, not without doing some work! Even in the best of circumstances, all we can say is that the disorder of an isolated system never decreases.

[8] **much**

In physics there are precise ways to quantify just how disordered something is, which we call a system's *entropy*[9]. Unlike energy, which is conserved, the *entropy of a thermally isolated system* ***never decreases***.

The caveat above is important: if a system (a set of objects) can exchange heat with another system, there's no guarantees about what happens to the entropy of just one of the systems.

With all of that, what is the second law of thermodynamics? It turns out there are many different equivalent ways of expressing the content of this law! They all sound, at first, quite different from each other, and it was a major achievement to show that each "version" of the second law is equivalent to each other version[10]. Here are some samples:

The Second Law of Thermodynamics says

> *Kelvin's statement*: No process is possible whose sole result is the complete conversion of heat to work ("There are no ideal engines")
>
> *Clausius' statement*: No process is possible whose sole result is the transfer of heat from cold to hot ("There are no ideal refrigerators")
>
> *Entropy version*: The total entropy of an isolated system cannot decrease over time, and is constant if and only if all processes are reversible[11].

In this class we won't dive into the details of why these statements are all, ultimately, the same, but I cannot stress enough how fundamental and important the second law of thermodynamics is! Here's a quote from C. P. Snow, a British writer and chemist born at the beginning of the 20th century, that tries to express how fundamental this law is in the sciences:

> "A good many times I have been present at gatherings of people who, by the standards of the traditional culture, are thought highly educated and who have with considerable gusto been expressing their incredulity at the illiteracy of scientists. Once or twice I have been provoked and have asked the company how many of them could describe the Second Law of Thermodynamics. The response was cold: it was also negative. Yet I was asking something which is the scientific equivalent of: Have you read a work of Shakespeare's?" – C. P. Snow, Rede

[9] "But as I hold it to be better to borrow terms for important magnitudes from the ancient languages, so that they may be adopted unchanged in all modern languages, I propose to call the magnitude S the *entropy* of the body, from the Greek word $\tau\rho o\pi\eta$, transformation. I have intentionally formed the word *entropy* so as to be as similar as possible to the word *energy*; for the two magnitudes to be denoted by these words are so nearly allied in their physical meanings, that a certain similarity in designation appears to be desirable." – Clausius, Ninth Memoir, *On several convenient forms of the fundamental equations of the mechanical theory of heat*. For all that talk of borrowing terms for important magnitudes from ancient languages, note that Clausius tried to name the unit of entropy "the Clausius," a calorie per degree Celsius... that unit did not stick around!

[10] Scientists like Kelvin and Clausius showed they were equivalent by proving that starting from one version of the second law you could prove any other version of it!

[11] We won't get into what "reversible" means in this context, but its the thermodynamics version of thinking about frictionless situations for objects in motion – it's a condition that makes some of our analyses easier, it's rarely completely appropriate in the real world, but some systems are pretty close to being reversible or letting us neglect the effects of friction

> Lecture of 1959, as recorded in *The two cultures*, Cambridge University Press, 2012.

So, this law says that we can't perfectly turn heat into useful work, but it does allow us to turn some amount of heat into useful work, or move heat from a hotter to a colder place if we simultaneously do work – as long as the total entropy doesn't decrease. On its own, heat flowing from a hot to a cold object increases the total entropy (which is another way of talking about why heat naturally flows in that direction). Another possible strategy to cool down your house would be to, say, pump some cold water into your house, let that cold water absorb some of the thermal energy in your house (cooling it down), and then dump that warmer water outside your house: this, too, increases the total amount of entropy in the house-water-outside system, you've had to do work (to move the water around), and your house has gotten a bit cooler. We'll look at this in more detail soon.

9.1.4 The third law of thermodynamics

Before we get to air conditioners, engines, and automobiles, let's at least mention the third of the three laws of thermodynamics.

> **The Third Law of Thermodynamics says** that the entropy of a system approaches a constant value as the temperature approaches 0 `K`.

If not knowing the second law of thermodynamics is like a student of English literature not knowing Shakespeare, being unaware of the third law is perhaps more like never having read Marlowe. The third law is a bit different in character from the other laws of thermodynamics, and is more sensitive to the details of the theories governing what happens at extremely low temperatures (like quantum mechanics), but it does have one very important consequence: *the unattainability of absolute zero*! The chemist Walther Nernst[12] helped show that, as a consequence of the third law, *it is impossible for any process to reduce the entropy of a system to its absolute-zero value in a finite number of steps.* In other words: if you want to cool a system down to absolute zero, you have to cool it down *infinitely slowly* – and we don't have time for that!

9.2 Engines and automobiles!

The last section was a little bit more conceptual and, perhaps, abstract compared to the rest of these notes. So, let's see some concrete examples of using these laws of thermodynamics in the context of *engines*: devices which use the flow of heat to do usable work!

9.2.1 Work, heat, and temperature in a piston

The first law of thermodynamics says that heat is a type of energy, and that means that we can alter the internal energy of a system not only by doing mechanical work on it but also

[12] One of the handful of people to have the dubious distinction of both winning a Nobel Prize winner and being listed as a war criminal.

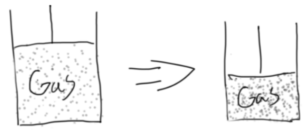

Figure 9.1: **Doing mechanical work on a gas**

by arranging for heat to flow into or out of the system. Consider the cartoon of a gas-filled piston shown in Fig. 9.1, how can we increase the temperature of the molecules (increasing their internal kinetic energy)? The most obvious answer from our everyday lives is to use a transfer of heat – say, by lighting a fire near the piston. Another way, though, is to do mechanical work by *compressing the gas*! At the level of gas particles, you can think about this like a giant tennis racket imparting momentum to the atoms and molecules via collisions.

This process can go either way: we can do work on the piston to compress the chamber of gas (putting energy into the system) *or* we can release a lock keeping the piston in place which lets the gas to work on the piston as it expands! The way these concepts are all linked is by the first law, $\Delta U = Q - W$, where each of the terms in that equation can be positive or negative. I should note that there are different "sign conventions" for work[13], but as long as you have the concepts clear it should be easy to use either sign convention.

Not only can the process of doing work go either way, but the consequences and workings of this whole process can sometimes be subtle. For example, suppose I ask, "When you do work to compress the gas in a piston, does the temperature in the gas go up, or down, or does it stay the same?" From our discussion above you might think the answer is that it always goes up, but not so! It depends on *how* we compress the gas.

1. **Case 1**: If you compress the gas *very quickly* there is no time for there to be any substantial heat transfer, which means $Q \approx 0$. By the first law, that means the internal energy of the gas goes up, so the temperature increases. A thermal process in which no heat gets transferred, like this, is called an *adiabatic* process.

2. **Case 2**: If you compress the gas *extremely slowly*, then as you compress the gas there is time for any excess thermal energy above the gas' original temperature to flow out of the gas into the piston and then into the surroundings. Here, $Q < 0$ as heat flows out of the gas, and it matches the work done on the gas, $|Q| = |W|$. Thus the internal energy of the gas doesn't change, and so the temperature stays the same. A thermal process in which the temeperature is constant, like this, is called an *isothermal* process.

And, of course, you can have thermal processes which fall in between these two limiting cases. But the analysis is (usually) simpler when we stick to these idealized versions of doing work and letting heat flow, and we can build up our intuition for thermal proceses in general.

[13] Are we thinking from the perspective of the object, or from the rest of the environment? Answering this question tells you whether the work, W, is positive when the environment does work on the object or when the object does work on the environment.

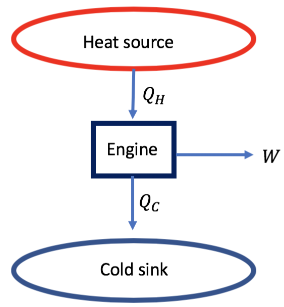

Figure 9.2: **An idealized "engine,"** a device which moves a certain amount of heat from a hot source at temperature T_H to a cold place at temperature T_C, and in the process does some work!

9.2.2 Abstract engines

Schematically, engines are devices that do what you see in Fig. 9.2: their net effect is to move some amount of heat, Q_H, from a hot place, dump some amount of heat, Q_C, in a colder place, while doing some work, W, in the process.

The *efficiency* of an engine is a measure of how well the engine is able to extract useful work during this process, and it is defined as

$$\eta = \frac{W}{Q_H} = \frac{Q_H - Q_C}{Q_H} < 1.$$

The efficiency cannot be greater than 1, because that would violate the conservation of energy, and Kelvin's statement of the 2nd law of thermodynamics says that the efficiency cannot *equal* 1, either[14]. Having $\eta = 0$ is certainly allowed – that's just letting heat naturally flow from a hotter to a colder place, but who would pay for an engine that does zero work?

You might wonder, what sets the limit on just how efficient an engine could possibly be? Sadi Carnot[15] deduced that the answer (which is ultimately a consequence of the second law of thermodynamics, but Carnot did not know that!) depended *entirely* on the temperatures of the hot and cold objects the engine was working between! Carnot's theorem says that the maximum efficiency of *any* engine is

$$\eta_{max} = \frac{T_H - T_C}{T_H}.$$

9.2.3 An ideal gas engine

Let's make this discussion a bit more concrete by considering an ideal gas ($PV = k_B NT$) inside a piston, and see how we can manipulate this system to make an engine! Fig. 9.3 shows a *pressure-Volume cycle* for the gas in the piston. What we're drawing at every point

[14] From the equation, $\eta = 1$ means that $Q_C = 0$, but then the engine would be a process whose sole effect was the conversion of heat to work! Having $\eta < 1$ gets around Kelvin's statement, since the process now involves both the conversion of heat to work *and* the dumping of some amount of heat.

[15] A 19th century French scientist and engineer, named after a 13th century Iranian poet

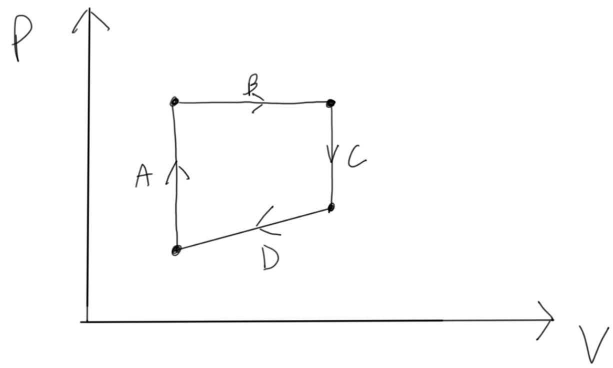

Figure 9.3: **Putting a gas through a power cycle to do some work!**

in this diagram is the *current state of the system* – how much volume is the gas taking up, and what is the pressure inside the gas? From that information, assuming the gas is ideal, we can work out the temperature of the gas, should we want to.

What we've actually drawn is a *cycle* that we're putting the piston through, a loop in this pressure-volume diagram that we'll have the piston repeat over and over. In the figure I've labeled four different segments of the cycle; let's think about each of them!

(A) In this segment the pressure increases and the volume stays the same. From the ideal gas law we know the temperature *must* increase, and since the volume is the same – there is no compression or expansion of the gas – no work has been done, $W = 0$. There has been no work, but the temperature (i.e., the internal energy) has increased, so *heat has flowed into the piston*. In other words: we lit a fire under a piston in the locked position.

(B) In this segment the volume increases and the pressure stays the same. Again, from the ideal gas law the temperature must increase, so $\Delta U > 0$, and since the gas is expanding the gas is pushing against the piston, doing (positive) work. But $\Delta U = Q - W$, and if ΔU is positive and W is positive, that must mean Q is also positive. In other words, the piston is doing work while the heat is still flowing into the piston: the piston has been unlocked, but is still over a fire.

(C) Now we reverse the logic of part (A): the pressure decreases while the volume stays constant ($W = 0$), so the temperature of the gas must also decrease. Since the internal energy decreased by no work was done, heat must have flowed out of the piston. In other words: we locked the piston position and exposed it to something cold (an ice bath, or cold air, etc).

(D) In the last leg of the cycle, the pressure and volume are *both* decreasing, so clearly the temperature decreases, too. Now W is a negative number, so the only way to make sure $\Delta U < 0$ to match the decrease in temperature is to have Q also be negative. In other words, heat is still flowing out of the gas while it compresses: we have unlocked the piston while still exposing it to something cold.

The above is just one of may cycles we could have drawn, but it illustrates the general principle:

1. **work**: On segments (A) and (C) no work is done. On segment (B) the piston does work on the environment, and on segment (D) the environment does work on the piston.

2. **heat**: Heat flows *into* the piston on segments (A) and (B), and flows out of the piston on segments (C) and (D).

3. **internal energy**: The internal energy of the gas in the piston (its temperature) *increases* on segments (A) and (B), and decreases on segments (C) and (D).

This list gives us the signs of these three physical quantities, but also important are the magnitudes of each of them. We wont calculate them here, but some very important points. First, the total change in the internal energy of the system after a complete cycle is *zero*! Second, the engine takes in more total heat in segments (A) and (B) than it loses in segments(C) and (D). Third, the engine does more work on the environment in segment (B) than the environment does on it in segment (D).

After all of this, what do we have? A device which absorbs a total amount of heat, $Q_{hot} = Q_A + Q_B$, when it's over a fire, releases a total amount of heat, $Q_{cold} = Q_C + Q_D$, when exposed to something cold, and does a total amount of work, $W = W_B + W_D$, during the whole cycle. That's the definition of an engine!

9.2.4 More realistic engines

The cycle we drew above is a heavily idealized version of the cycle used in early steam engines[16], for instance the version invented by Thomas Newcomen in 1712. In that steam engine, segments (A) and (B) of our path basically occurred simultaneously, as a piston valve was opened to let the piston fill up with steam from a coal boiler (Q is absorbed, the temperature and volume of the gas in the piston increases, and the piston pushed the handle of a water pump). Segments (C) and (D) also basically happened at the same time: the piston valve was closed, and the outside of the piston was sprayed with cold water (so Q flowed out of the gas in the piston, the volume decreased, and the piston position returned to its original spot to start the cycle over again).

In a 2-stroke internal combustion engine (patented in 1881, and still used, I believe, in things like chainsaws and outboard motors on small boats), a spark plug ignites a small explosion of gasoline, which rapidly pushes a piston down. This simultaneously drives a crankshaft, expels hot exhaust, and compresses a reservoir of fuel to bring the next load of fuel into position. As the piston cools down (just from the cold air around it, for instance), the piston comes back up, which compresses the fuel near the spark plug, ready to be lit on fire again. These combustion engines convert chemical energy (via explosion) to create the "heat reservoir" that allows the engine to work, rather than being directly exposed to a reservoir which is continuously hot.

The 4-stroke internal combustion engine is basically the same idea, but is separates out the fuel intake and the exhaust expulsion into separate stages. This turns out to make a

[16] For some helpful animations, see this youtube video, for instance.

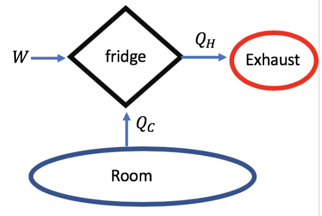

Figure 9.4: **An idealized "refrigerator,"** a device which pumps heat from a cold place to a hot place, an operation only permitted if work is done on the fridge!

more efficient engine, but only one of the four stages of the engine operation actually does the useful work. Thus, combining four such cylinders with staggered times of operation is a popular method to give a relatively smooth output of continuous usable work!

9.3 Air conditioners!

A *refrigerator* (or "heat pump") can be thought of as an engine running backwards! It's a device which uses (rather than outputs) work in order to take some amount of heat from a cold place and move it to a hot place. An schematic is shown in Fig. 9.4, where it should be clear that just by reversing the directions the arrows point (which, in our equations is equivalent to *changing the sign of each term*) makes the schematic look just like our picture of an engine.

Similar to the efficiency of an engine, the *performance* of a heat pump is defined as the ratio of the amount of heat it can pump relative to the amount of work it needs to do so:

$$\omega = \frac{Q_C}{W} = \frac{Q_C}{Q_H - Q_C}.$$

Since a heat pump is an engine in reverse, it shouldn't surprise you to learn that the limit on performance of even the most perfect heat pump is *determined by the difference in temperatures* between the hot and cold regions the heat pump is operating between:

$$\omega_{max} = \frac{T_C}{T_H - T_C}.$$

Just as there are many different cycles that could be used to power an engine, there are many different ways to design a refrigerator. All of them, though, make use of a thermodynamic cycle in which work is used to pump heat against the direction it would naturally flow.

9.3.1 Phase changes from solids, liquids, and gases

Rather than just walking through our ideal-gas engine example backwards, I want to describe the kind of refrigerator that is in our kitchens. But to do that, I have to briefly talk about phase changes! I've already introduced the idea that the differences in how we interact with different *phases* of the same substance – say: ice, water, and steam – ultimately come from the differences in the microscopic arrangement of the water molecules. Ice is a solid, with water molecules in a symmetrical crystal; water is a liquid, where the molecules are disordered but occupying a definite volume; steam is a gas, where the molecules are disordered and filling the available space.

When steam molecules are independently flying around, they have a fair amount of chemical potential energy. They can release some of this energy by bonding together into a liquid, but these bonds are weak, and if the molecules have too much internal kinetic energy, the bonds will break apart. But, when the temperature is low enough, the bonds hold the liquid together. Even still: the bonds are weak enough that water molecules are always transiently breaking and reforming bonds as they flow past each other: this is why liquid water can adapt its shape.

The water molecules can – if the temperature is low enough so that the resulting configurations are stable – release even more chemical potential energy by bonding together a bit more stiffly in the symmetrical arrangements we call ice. Most substances, when they transition from a liquid to a solid, become denser – water happens to be a rare exception, and that's why ice floats!

Melting

When an ice cube is sitting in a glass of water, what is happening? At first, heat from the water (and surrounding air) flows into the ice cube, and the ice cube's temperature starts to rise (straight out of a typical freeze, the temperature of the ice cube is probably about 0° F (or −18° C). This continues until the ice reaches its melting temperature – 32° F (or 0° C), at which point it begins to...melt!

At this point, the ice does not get any warmer; instead, the thermal energy going into the ice goes towards breaking those stiff chemical bonds keeping the ice together. The heat used to transform a certain mass of a solid into a liquid without changing the temperature of the solid is called the *latent heat of fusion*, and ice has a very large latent heat of fusion! Even without changing the temperature *at all*, it takes about $330,000$ J of thermal energy to convert 1 kg of ice at 0° C to 1 kg of water at 0° C. Compare that to what we learned about the specific heat of water (4190 $\mathtt{J}/(\mathtt{kg}\cdot\mathtt{K})$): the same amount of energy that would melt but not change the temperature of a block of ice could raise the temperature of the same mass of water by 78.6° C! That is, it takes about as much energy to melt ice as it does to start with the same mass of water at room temperature and boil it!

This latent heat goes both ways: it takes thermal energy to melt ice into water at zero Celsius, and when water freezes into ice at zero Celsius it releases thermal energy as it freezes – this is the energy you must continue to remove from the water at 0° C to get it to freeze into ice at at 0° C.

A similar mechanism happens at at 100° C, but now when thinking about the coexistence of water and steam. Water molecules are *really* hard to separate, and the *latent heat of vaporization* for water is even bigger than the latent heat of fusion – something like $2,300,000$ J are needed to turn 1 kg of water at 100° C to 1 kg of steam at 100° C. This, by the way, is why *evaporative cooling* can be so effective: the body perspires, and then a large amount of thermal energy can go into turning that perspiration into vaporization, leaving you cooler.

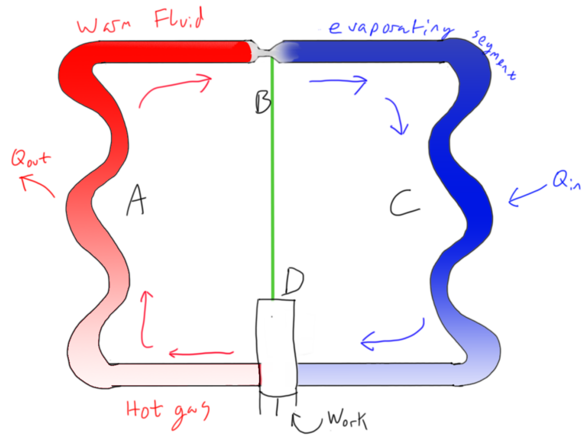

Figure 9.5: **A schematic of a vapor-compression refrigeration unit,** the type of cooling device keeping the food in your kitchen cold! Two parts of the device (hot and cold regions) are separated by some insulating layer, shown in green.

9.3.2 Vapor-compression refrigerator

Let's walk through a particular design, a *vapor-compression refrigeration system*, which is schematically illustrated in Fig. 9.5. If you want to see one in real life, take a look at a standard kitchen fridge!

In such a fridge a refrigerant fluid – carefully chosen for its thermal properties and temperatures and pressures at which is changes from a liquid to a gas – is put through a cycle. Let's walk through the key components of a vapor-compression cycle, in the order illustrated.

(A) The *condenser* is typically a winding coil, and at this part of the cycle the fluid inside these coils is relatively hot, and in the gaseous state. The large surface areas of these winding coils lets them efficiently *radiate* heat away, dumping heat into the room. As the gas inside these coils cools it *condenses* into a liquid phase.

(B) A relatively warm, high pressure liquid flows from the condenser to the *expansion valve*, also known as a constrictor. This valve is basically a narrow constriction, which dramatically reduces the pressure in the fluid.

(C) Because of lower pressure, the warmish fluid is suddenly above the temperature at which it will evaporate[17]. Heat from the already-cool air in the cold compartment flows into the fluid in *the evaporator*, another set of coiling tubes, to provide the energy needed for the liquid to actually turn into a gas[18].

[17] We haven't discussed why this is the case, but it's true: at higher pressures liquids will boil at lower temperatures! That's why, for instance, water boils at a lower temperature in Denver than in Atlanta, and why recipe times need to be adjusted accordingly!

[18] This is the principle of evaporative cooling, the human body's primary mechanism to prevent overheating.

(D) After winding through the evaporator, the refrigerant liquid is now a gas. True to its name, *the compressor*, well, compresses this gas. This is the stage where we need to put work *into* the cycle: we feed (typically) electrical energy into the compressor so that the compressor does work on the incoming low-density gas, jamming it into a much smaller volume and, in the process, increasing it's temperature. At this point, the cycle starts anew.

Chapter 10

Mechanical Waves!

In this chapter we'll start thinking more about *waves*, which can be repetitive motions of an object or disturbances in a medium like air or water. We'll start out by revisiting the "simple harmonic oscillators" we first met in the context of springs, and see how the repetitive motion of simple oscillators help us keep track of time. That will equip us to think about sound waves, which are disturbances in (for instance) the air around us, and some musical instruments that can produce them. We'll close the chapter by talking about ocean waves and the tides.

10.1 Clocks

10.1.1 Harmonic oscillators again

Back in the chapter on "Mechanical Objects" we learned about motions governed by Hooke's Law, $\vec{F} = -k\vec{x}$, and how that led to simple harmonic motion. For instance, a mass attached to a spring was pulled so that the spring stretched away from its equilibrium rest length would accelerate if it were let go (as the elastic potential energy stored in the spring was converted to kinetic energy of the mass), reaching a maximum velocity when it got to the position corresponding to the rest length, slowing down as the spring compressed (as the mass' kinetic energy was converted back into elastic potential energy), and then accelerating back towards where the mass started. In the absence of friction or drag, this process would go on and on, forming a regular cycle of the mass oscillating back and forth.

Something we didn't talk about: *how long* does each of these cycles take? If we stretch a spring and let it go, how long do we have to wait for the mass to accelerate, slow back down as it compresses the spring, accelerate again, and reach the position from which we let it go? What *could* the answer to that depend on? Surely the mass of the object matters, since mass tells us how resistant the object is to changes in its velocity. Similarly, the spring constant characterizing the stiffness of the spring, k, should matter: it helps tell us the magnitude of the force on the object, and we know that forces are crucial for determining how an object changes its velocity.

The only other physical quantity we have in this problem is the maximum distortion of the spring – how far back we pulled the mass in the first place before we let it go! The furthest displacement from equilibrium is called the **amplitude** of oscillation, and a

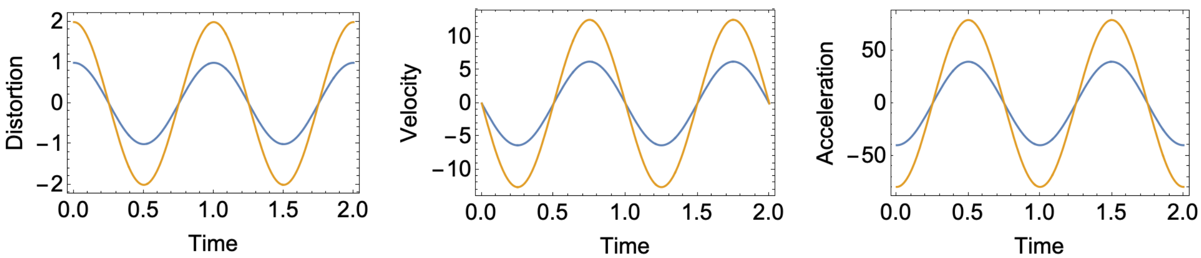

Figure 10.1: **Distortion, velocity, and acceleration of an oscillator** If a mass on a spring is pulled to two different initial distortions and then let go, the distortion of the spring, the velocity of the mass, and the acceleration of the mass all trace out interrelated paths over the course of a the oscillator's cycles.

remarkable feature of simple harmonic oscillators is that the length of time of each of their cycles is *independent of the amplitude*! This is ultimately a consequence of the very special types of restoring forces they experience, always proportional to the current displacement from equilibrium. Oscillators that have large amplitudes have to travel farther with each cycle, but they also experience larger maximum restoring forces. The mathematics of this is precisely balanced so that the larger forces move the object through this greater distance in *exactly the same amount of time.*

This means that the **period** of the oscillation – the length of time to complete one cycle – depends only on the object's mass and the spring's stiffness. Dimensional analysis can help us out here: a period has dimensions of time ($[T] = \mathbb{T}$) , mass has dimensions of mass ($[m] = \mathbb{M}$), and $[k] = \mathbb{M}\mathbb{T}^{-2}$. How can we combine k and m to get something with dimensions of time? Apparently we need to take the ratio of m to k, and then take a square root! There is also a numerical prefactor to get precisely the right answer, which is that the period T is:

$$T = 2\pi\sqrt{\frac{m}{k}},$$

The **frequency** of an oscillation is the number of cycles that happen per unit time:

$$f = \frac{1}{T} = \frac{1}{2\pi}\sqrt{\frac{k}{m}},$$

which has dimensions of inverse time. The SI unit is the Hertz[1] , 1 $\texttt{Hz} = 1\ \texttt{s}^{-1}$. So, an oscillator that takes 4 seconds to get back to where it started has a period of 4 s and a frequency of 0.25 Hz. When an oboe plays an "A" note at 440 Hz to tune an orchestra, something about the air (we'll see what later in this chapter!) is oscillating with a period of $1/440\ \texttt{s} \approx 0.0023\ \texttt{s}$.

[1] Named after the physicist Heinrich Hertz – among other things he conducted incredible experiments with radio waves, but failed to even imagine that radio waves could be used for communication.

10.1.2 Resonance and natural frequencies

We've just learned that simple harmonic oscillators have what we might call a *natural frequency*, a particular frequency associated with motion about their stable equilibrium configuration. All sorts of things around us have such natural frequencies – when you bump into a rocking chair, it will rock back and forth at some particular frequency determined by how it is shaped; the strings on a guitar vibrate back and forth at particular frequencies when plucked – and in most of these systems *friction cannot be ignored*, and the oscillations will eventually die away. To keep the oscillations going, we often rely on **resonances**, in which the amplitude of an oscillation can be increased by periodically applying a force *at a frequency close to the natural frequency of the oscillation*.

If you've ever played on a swing, you know exactly how this works! Maybe you want to push your friend to get them swinging, but at first they're just sitting in the swing. So, you start giving them pushes at the same point in every swing cycle (say, at the lowest point of the swing when they are going at their fastest speed forward; alternately, at the back of the swing when they temporarily pause as their velocity changes direction). By timing your pushes to the natural frequency of the swing, you pump more and more total energy into the system in such a way that (a) reinforces the natural oscillatory motion and (b) can compensate for energy lost due to friction or drag.

This idea of applying forces with a frequency that matches (or nearly matches) a system's natural frequency lets us sustain mechanical oscillations, like in a spring or swing system. It also explains how rubbing the top of a wine glass can get it to sing (or how singing at a wine glass can make it break!), and how a bow can make a violin string ring out[2].

10.1.3 Clocks

It also, conveniently, helps us build clocks! Timekeeping devices have been around for a long time, based on tracking things in the natural world. Sundials are a classic example: by watching the shadow cast by a stationary object move around, one can know what fraction of the day has elapsed since sunrise (not particularly helpful on a cloudy day, or at night, but it's something!). Shorter periods of time could be tracked by filling hourglasses with an amount of sand that takes (approximately) a fixed amount of time to run out when flipped over, but measuring the inflow or outflow of water from special containers (Again, quite helpful, but who wants to be in charge of constantly flipping an hour glass over and over, or to be constantly refilling a bucket of water just to track the time?).

Water-based clocks were basically the state-of-the-art in clock technology until pendulum-based clocks came around[3] were an important innovation in the time-keeping space. When a pendulum is not swaying back and forth its center of mass is directly below the pivot point – that's the pendulum's equilibrium position. If the pendulum is pulled to one side, notice that gravity (pointing straight down) does not act in the same direction as the rod (or string)

[2]Check out the animation in the "Helmholtz motion" section of this page!

[3]Inspired by observations of Galileo and invented by the dutch physicist Christiaan Huygens, pendulum clocks were basically the most accurate way of keeping time from the mid-1600s to the 1930s. Pendulum clocks were sufficiently accurate that by the late 1600s a "minute hand" was added to keep track of divisions of time within an hour.

of the pendulum is able to apply tension. There is a restoring force due to gravity, then, that causes the pendulum to accelerate towards its equilibrium position, and as long as the amplitude of oscillation is not too big, the magnitude of this restoring force is proportional to the displacement away from the pendulum's equilibrium position.

In other words, you see, *a pendulum is basically a simple harmonic oscillator*! Again we can think about the *period* of a pendulum as a competition between the mass hanging at the bottom of the pendulum (giving a resistance to accelerations) and the restoring force, but here's you'll notice that the strength of the restoring force also depends on the mass of the pendulum. Thus for an idealized pendulum, the period of oscillation doesn't even depend on how much mass it has, it only depends on the *length* of the pendulum and the acceleration due to gravity:

$$T_{pendulum} = 2\pi\sqrt{\frac{l}{g}}.$$

Short pendulums swing back and forth more frequently than a longer one[4], and by choosing pendulums of particular lengths we can easily start constructing clocks.

Since pendulum clocks are designed to operate for long amounts of time, there's plenty of time for friction (about the pivot point) and drag (of the pendulum bob swinging through the air to slowly sap energy out of the system. Hanging weights inside pendulum given tiny little pushes to the pendulums, and are of course designed to give *resonant* pushes, timed to the natural swing of the pendulum.

More modern time-keeping devices use the same principles, working off the natural frequencies of various objects and devices. Mechanical watches, for instance, make use of an oscillator composed of a balance wheel (which looks kind of like a tiny bicycle wheel) connected to a *torsional spring*, that is, a tiny coil spring that exerts torques on the balance ring when it is rotated away from its equilibrium position. These have the major advantage over pendulum clocks in that they are designed so that *gravity does not exert a torque on the balance ring*, so they don't need to be on level ground[5] in order to operate! Electronic watches make use of small vibrating crystals (often quartz crystals), that naturally oscillate at particular frequencies. Quartz is a substance called a *piezoelectric*, which means that if you apply electric fields to them they experience distortions, and if they are distorted they generate electric fields. This means that a small quartz crystal can form the core of electronic clocks that use electric currents to both detect the minuscule vibrations of the crystal and also supply the power to keep those vibrations going.

[4]We can quickly do the math, here: at the surface of the Earth a pendulum of length 0.25 `m` has a period of about 1 second, and a pendulum of length 1.0 `m` has a period of about 2 seconds. This basically determines the internal gears and mechanisms of wall and floor clocks.

[5]The history of clocks, it turns out, is intimately connected to the history of navigation on the open ocean. Knowing that the Earth is tilted on its axis, a sailor could quite accurately determine his or her latitude by using a sextant to measure how high in the sky the sun appears at high noon. But how to determine how far east or west you are was a real issue (leading to shipwrecks, and lost boats, and inefficient sailing routes...so much so that the British government offered a massive prize called the Longitude rewards – worth millions of today's dollars – to whoever could come up with a working system that sufficiently precisely determined longitude. Where do clocks come into this? Because if you set a clock to noon at high noon in some reference location, you can then tell your current longitude based on when high noon occurs in your current location *relative to what your clock thinks noon should be*!

10.2 Musical instruments

When we pluck a guitar string, or hit a piano key, or play an oboe, what physically is going on? What is the sound that we're hearing, and why do different instruments *sound different* even when they are playing the same "note"? For that matter, how is it that we can store the sound of an entire symphony (or EDM, or k-pop band, or whatever else you're listening to) in the grooves of a record or in a sequence of zeros and ones on your computer?

10.2.1 Vibrations, oscillations, harmonics

To start with, let's think about a vibrating guitar string. As shown in Fig. 10.2, when a string is under tension and it is pulled away from a straight line into a curve, the tension in the string results in a restoring force trying to pull the string back to a straight line. The simplest mode of oscillation just has the middle of the string displacing the most and then vibrating back and forth: this simplest mode is called the "fundamental" vibrational mode, and the frequency of the fundamental mode of a string depends on several things: the length of the string L, the "linear density" of the string[6] μ, and the tension F_t:

$$f = \frac{1}{2L}\sqrt{\frac{F_t}{\mu}}.$$

Dimensionally, do you see that the right hand side has the units of `Hz` that we're looking for?

However, as also shown in Fig. 10.2, this fundamental mode is not the only way that a stretched string can oscillate back and forth! The string can also vibrate back and forth as if it were two strings of half the length, which is called the "second harmonic mode." Since the tension in the string and its linear mass density is the same, the frequency of the second harmonic is as if the length of the string was half as long – i.e., the frequency is twice as large. Similarly, the fourth harmonic mode has four little segments of the string oscillating up and down, each acting like a string one-fourth of the original length and vibrating at a frequency four times higher.

The existence of these harmonic modes is why different instruments sound different, because the same string can be simultaneously vibrating in more than one mode at the same time! The total shape of the string can be a *superposition* of the modes composing it – a fun word that basically means the sum of the shapes! When a guitarist plucks a string the string begins vibrating in a particular combination of both the fundamental and the higher-order harmonic modes, and as the note rings out each of these different modes decay away (due to friction, and drag) at different rates.

The particular mixture of modes that each different instrument produces is called the timbre of an instrument, and it's how our ears can tell apart the sound produced by a guitar and a violin[7], or even by the same violin with the bow pulled across the strings at different positions.

[6] how much mass per unit length the material composing the string has

[7] Notice that the process of drawing a bow across a violin string is another example of resonant energy transfer. The bow hairs stick to the strings and then slip across them before sticking again – over and over – and the frequency of the stick-slipping matches the frequency of the oscillating string. Just like pushing your

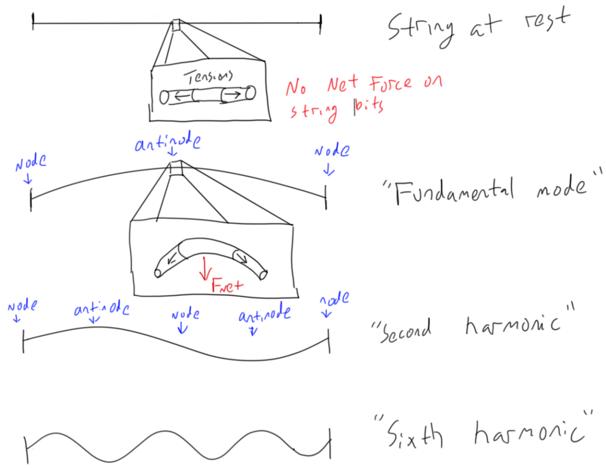

Figure 10.2: **Vibrational modes of a string under tension** A string stretched between two points has many modes in can vibrate in. Each of these different modes acts like an oscillator: when the string is curved the tension in the string acts as a restoring force, as seen in the zoom in on fundamental vibrational mode. Higher-order "harmonic" modes can also vibrate on the same string, corresponding to different numbers of *nodes* (where the string is at the same spot it would be if it weren't vibrating) and *anti-nodes* (where the string moves up and down the most). This wikipedia page has a nice animation of these vibrating modes, if you'd like to see them!

10.2.2 Other musical instruments

Stringed instruments all work in this way: By using strings of different thicknesses and under different tensions, each string can be *tuned* so that it has a different fundamental frequency. An instrumentalist can alter this fundamental frequency by holding the string down against the instrument (say, a guitar fret) – this effectively shortening the string, reducing the longest wavelength and hence increasing the fundamental frequency. However, air is pretty good at "getting out of the way" of a small string vibrating back and forth, and if you've ever plucked a random string you were holding taut you know that it doesn't sound anything like a nice guitar. The stringed instruments are designed so that the vibrations of the strings get turned into oscillations of the entire body of the instrument. The vibrating surface of the instrument that compresses and rarefies the air near it, and that's exactly what a sound wave is!

Most other musical instruments work not by creating a standing wave on a string but by directly vibrating a column of air: this is what pipe organs and flutes and clarinets, and trumpets (etc.) all do. In this case the fundamental frequency of the instrument is related to the total length of the column of air in the instrument. The very short piccolo can produce only high-frequency vibrations, whereas the massive, ceiling-scraping pipes in one of those pipe organs in old cathedrals can produce rumbling low-frequency vibrations.

So, instruments have collections of fundamental frequencies they can produce, and the balance of harmonics produced at the same time is what determines the unique overall

friend on a swing with pushes timed to the swing's natural frequency, this imparts energy into the string to allow it to keep ringing out its note.

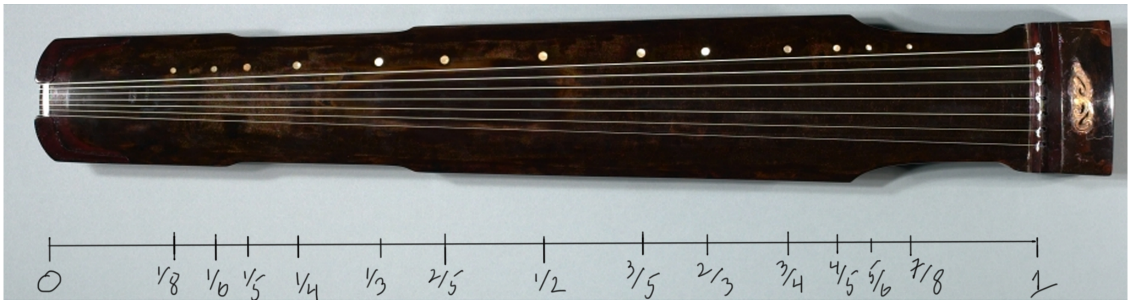

Figure 10.3: **A guqin,** a traditional stringed Chinese instrument. Markings on the instrument measure out simple integer ratio lengths of the strings. Image from Lingfeng Shenyun, 2006

sound of different instruments even as they play the same fundamental pitch. Most musical traditions around the world are based on the idea that humans seem to find pitches arranged in particular ratios to be pleasing. For instance, the current[8] Western scale is essentially based on a note called A_4, or "middle A," a note whose pitch has been standardized to 440 `Hz`. An *octave* refers to a note whose frequency is a multiple of two, so one octave above A_4 is the note A_5, at 880 `Hz`. Other powerful ratios include 3/2 (what musicians call a *fifth*[9]) – for instance the note E_4 sounding at 660 `Hz` – and 5/4 (what musicians call a *third*[10]) – for instance the note C_4 sounding at 550 `Hz`. The importance of these simple ratios is not restricted only to Western musical traditions. For instance, Fig. 10.3 shows a stringed Chinese instrument, the guqin, whose history goes back at least three thousand years. Clearly visible are markings positioned at integer ratios of the strings' length!

These nice, even ratios of frequencies are deeply ingrained in how we perceive musical sounds from noise[11]. This is why percussion instruments are sometimes in a different category than winds and strings. Drums, for instance, have vibrating surfaces rather than vibrating strings, and it turns out that the harmonics of a vibrating surface are much more complicated than the harmonics of a vibrating linear system. By "more complicated," what I really mean is that the harmonics are *not* simple integer ratios of the fundamental! They are also much better at damping out oscillations, and that combination is why we perceive most drum hits as brief atonal *thwaps* rather than specific pitches. Other percussion instruments, though, involve striking objects – say, the wooden bars of a xylophone – that have been carefully shaped so that their natural frequencies of vibration correspond to specific pitches.

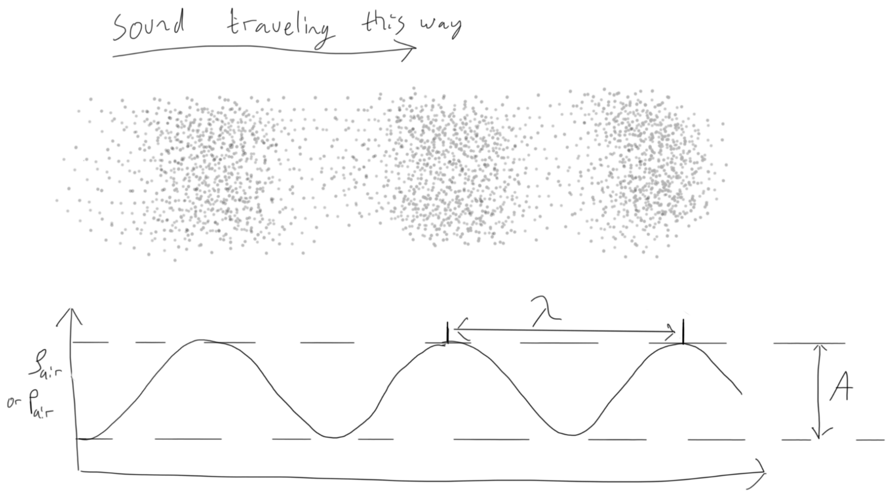

Figure 10.4: **A sound wave!** Sound is a longitudinal wave composed of variations in the density (or pressure) of the fluid traveling in some direction. The *amplitude* of the wave is the largest difference in the waveform from a peak to a trough, and and the *wavelength* is the distance in between two repeated parts of the wave. The sketch is... not perfect.

10.2.3 Sound waves

So, we have all of these instruments vibrating and oscillating; the sound that actually moves from an instrument to our ears does so as a *traveling wave* of compressed and rarified air, as in Fig. 10.4. There is an alternating pattern of these high-density and low-density regions[12] traveling at some speed and in some direction – hence, a *wave velocity* – and the physical separation between consecutive peaks of these waves is again the wavelength. The speed at which a wave travels can be written as an equation:

$$\text{wave speed} = \text{wavelength} \cdot \text{frequency}.$$

Does that mean that if we play a note with a lower frequency it travels at a different speed?

No! The speed of sound through a fluid is determined by the properties of the fluid, *not* the properties of the oscillation. The speed of sound in air[13] is about 345 $\mathtt{m/s}$ at sea level and 20° C. Just like the standing waves on a string had a frequency inversely proportional

[8] Post "well-tempering" debates

[9] Examples of an interval of a fifth include the opening trumpets in Also sprach Zarathustra, or the interval between how Billy Ray Cyrus and Lil Nas X sing "Take" and "Horse" in Old Town Road

[10] For instance, the interval between "Take" and "My" in the remix of Old Town Road

[11] In the traditional liberal arts education, the core *trivium* (Grammar, Rhetoric, and Logic) was followed by the *quadrivium* (arithmetic, geometry, music, and astronomy). The quadrivium is essentially a Pythagorean system of study: arithmetic and geometry were viewed as the studies of numbers and shapes "at rest," and music and astronomy were numbers and shapes in relation to each other and "in motion." Very poetic.

[12] which are actually pretty minute variations compared to atmospheric pressure! We can detect sound waves whose amplitudes are barely a millionth away from atmospheric pressure!

[13] For comparison, the speed of sound in water is closer to 1500 $\mathtt{m/s}$. Test this out the next time you're at a pool or a lake with your friends! If you hold your head sidewise half underwater, so that one ear is in the

to the length of the string (or the distance between nodes, for the harmonics), sound waves have this inverse relationship between how far apart the wave crests are and how frequently the patterns of density are oscillating.

By the way, thinking about sound as a traveling wave moving with some velocity helps explain the *doppler shift*! Have you ever noticed that the sound of an oncoming car is higher than when the car drives past you and starts moving away? Because of the relative motion of the car, your ears are encountering the crests and troughs of the sound waves at a different rate compared to the rate at which they were produced. If the source and the listener are *approaching* each other, the listener will encounter more peaks per unit time compared to if the source and listener are stationary (or receding from each other) – in effect, it's *as if* the listener is hearing a higher frequency!

We'll close this section by thinking about the "loudness" of sound, intuitively it makes sense that it depends on the amplitude of these waves of compressed and rarified air vibrating our ear drums, but more precisely we could say that our ears are measuring the **intensity** of the sound waves, which depend both on the *energy* per unit time being emitted by a source of sound and the *distance* between the source and the listener.

If you imagine someone standing in the middle of an opening field clapping, you can imagine that sound waves are traveling outward from that clap in all directions: you could hear the clap from in front of, behind, to the side, and above the person generating the sound. Thus, the energy of the sound is expanding outward in a basically spherical pattern. What is the intensity of the sound we hear, if we are standing some distance away? It's

$$I = \frac{\text{energy/time}}{\text{Area of the sound wave}} = \frac{\text{Power}}{4\pi r^2},$$

where in the final expression I've used the surface area of a sphere for convenience. On the one hand, if we wanted to quantify "loudness" we could just compute intensity values and give the answer.

On the other hand: there is an *enormous* range of sound loudnesses that humans can detect. A commonly quoted value for the softest sound undamaged human ears can detect is about $I_0 \approx 10^{-12}$ W/m^2, and we start to experience pain when intensities get up to around $I_{pain} \approx 1$ W/m^2. When confronted with these measurements that have to be able to account for many orders of magnitude, we tend to use "logarithmic scales," like the Richter scale for earthquakes. In this case we often define the loudness of sound in decibels, dB, defined as

$$I = 10^{dB/10} I_0.$$

What does this mean? Something at 0 dB has an intensity of $I = 10^0 I_0 = I_0$, i.e., it is barely at the edge of what we can perceive. Something at 10 dB is 10 times louder than that. Similarly, something at 80 dB is 100 times louder than something at 60 dB, and so on. For reference, a "normal[14] conversation with someone a meter away is probably exposing you to sound in the 40 - 60 dB level.

water and one is in the air, sound from an event will reach your underwater ear *faster* than the ear sticking out in the air. So, if your friend makes a splash 5 meters away from you, you'll hear it in the water about 0.01 s before you hear it in the air – since the temporal resolution of our ears is about 8 milliseconds, you should just be able to hear the two different sounds!

[14] Non-election-related

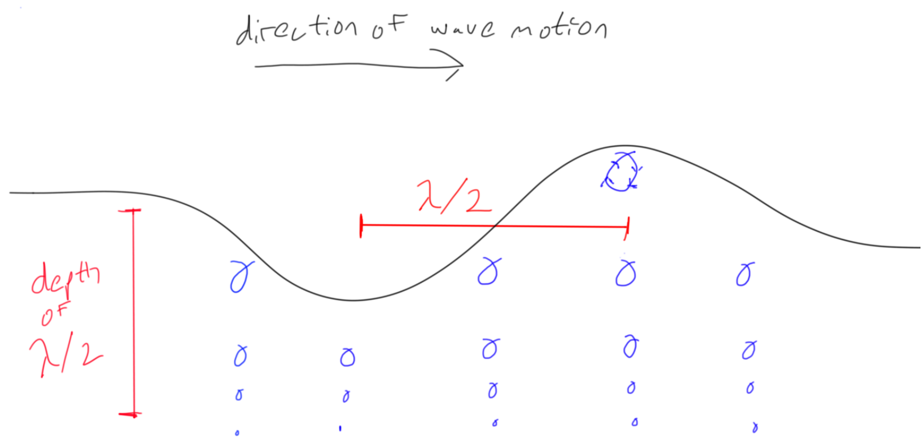

Figure 10.5: **Anatomy of a wave** A traveling wave on water. At the surface the wave peaks and dips smoothly above the natural level of the water with some amplitude, with the crest and the trough separated by half of the wavelength, λ, of the wave. As this shape travels in some direction across the surface, the actual water molecules do not travel neatly in the same direction. Instead, as the wave crest passes, the actual water molecules move around in little almost-closed circles (visible if you track the progress of a floating bird or bottle as it slowly rides the waves). The size of these circular motions is largest at the surface, and basically die out by the time you look $\lambda/2$ underneath the natural level of the water.

10.3 Water waves

Of course, when we think of "waves" the first thing our mind probably turns to are waves on the surface of the water[15]. Water waves are, in fact, much more complicated than the simple vibrations on a string or oscillations of sound waves we looked at above, but a lot of the same principles do help us understand what is going on.

Figure 10.5 gives shows the basic anatomy of a traveling wave on the surface of the water[16]. We recognize a lot of features of the waves we looked at earlier: there are wave speeds and wavelengths and frequencies and amplitudes, but unlike sound waves, where $s = \lambda f$, waves traveling across the surface of the water have a speed *which depends on their wavelength,* and sometimes also the depth of the water.

10.3.1 Capillary waves

Before we get to the kind of big waves you might see on the ocean, there are also small *capillary waves* that we see all the time as little ripples on the surface of water (on water as

[15] "*Now, the next waves of interest, that are easily seen by everyone and which are usually used as an example of waves in elementary courses, are water waves. As we shall soon see, they are the worst possible example, because they are in no respects like sound and light; they have all the complications that waves can have.*" – The Feynman Lectures on Physics, Chapter 51 Section 4

[16] In these notes I'll ignore "seiches," which are standing waves that can sometimes develop in lakes, channels, bays, and so on (and can sometimes be quite large!)

wind blows lightly over it, or in the concentric waves that form as you skip a stone across a lake, or drip a drop of water into a glass already mostly full. These small waves – which have wavelengths no more than a few centimeters – are due to the *surface tension* at the interface between the water and the air. This surface tension, which comes from the fact that water molecules like to stick together, is why you can fill a glass of water slightly more than full and watch it form a small stable dome above the top level of the glass without it spilling.

In this case, surface tension acts as a restoring force when the surface of the water is disturbed: the water (energetically) wants to stay as flat as possible. Just like the sound and vibrational waves discussed earlier (and the other waves we're about to discuss), these capillary waves can combine, overlap, add together, and so on. You can see the way the waves passing over and through each other combine by experimenting with some water yourself (or by looking at the infinite number of images on the internet); just as the earlier cases waves on water can superpose. When two crests coincide they lead to a peak that's even bigger, when a wave and a trough of equal heights coincide they cancel out for the time they're on top of each other.

10.3.2 Gravity waves

Surface tension is the stabilizing restoring force that gives us the little ripples on the surface of the water, but larger waves are instead stabilized by *gravity*. We learned in the chapter on Fluids that "water seeks its own level," so if there are large shapes of water that crest above and dip below the natural height of the water then there will be a restoring force as the entire water system tries to reduce its gravitational potential energy. As the wind sweeps over the ocean this can lead to large waves whose wavelengths can be anywhere from[17] 60 to 600 $\mathtt{m}$. For these waves the frequency (or, equivalently, the period) is determined by the forces generating the gravity waves, and the speed of the wave depends on this frequency. In deep water (again, where the depth of the water is more than half the wavelength of the wave), the wave speed is

$$s = \sqrt{\frac{g\lambda}{2\pi}} = \frac{gT}{2\pi},$$

from which you can see the relationship between wavelength and period for these deep water waves. To throw some actual (somewhat typical) numbers in here, a wave with $\lambda = 200\ \mathtt{m}$ would travel across the surface of the ocean at 17.7 $\mathtt{m/s}$ and has a period of about 11.3 $\mathtt{s}$.

The fact that the speed of the wave can depend on the depth of the water comes about because of how those little circles of moving water extend to about half a wavelength below the surface of the water. In *deep water* the speed of water waves is given by the equation above, but in *shallow water*, where the depth is less than about $\lambda/20$ the speed of the waves changes to $s \approx \sqrt{gh}$, where h is the depth of the water. In between depths of $\lambda/2$ and $\lambda/20$? The math is more complicated and there aren't of these simple formulas to use.

You can see that in shallow water the speed of waves stops depending on the wavelength of the wave (although *how shallow* the water has to be before this happens depends on

[17]You might have noticed the large gap in between the "few centimeters" of capillary waves and the "tens of meters" of gravity waves. In these notes I'm keeping things simple by calling gravity waves those where the surface tension of water really doesn't matter. In reality, for waves of intermediate wavelength both surface tension and gravity are important.

the wavelength). As a deep-water wave approaches land (where the floor of the gets closer and closer to the surface), both the speed and wavelength of the waves decreases and the amplitude of the waves increases (this is what happens in that complicated intermediate depth regime). When the depth gets into the "shallow water" regime, the wave forms peaks and breaks as it moves. The character of the wave breaking depends on how quickly the depth is changing: if the changes are gradual, the waves kind of roll and bubble as the break over a long distance. If the depth changes quickly, though, half of a wave crest won't have time to form, and the wave will break in a way perfect for taking cooling surfing photos.

Wind-generate gravity waves tend to be in the range of $60\ \mathtt{m} < \lambda < 600\ \mathtt{m}$, traveling at tens of meters per second. Enormously longer-wavelength waves can be generated by the sudden shifting of the Earths tectonic plates. These seismic ocean waves can have wavelengths of *hundreds of kilometers*; with such a large wavelength these seismic waves are *always* in the "shallow water" regime where their speed is related to the depth of the ocean. These waves travel are not only extremely dangerous (carrying potentially devastating amounts of energy with them), but they can travel *very* fast. From the formula above, the deeper the water the faster such a wave moves: the average depth of the oceans might be about 3.5 km, and hence a tsunami can cross oceans at ≈ 200 m/s (over 400 miles per hour). Basically, these seismic waves can cross the ocean as fast as a jet plane.

10.3.3 The tides

The cataclysmic waves generated by seismic activity can be pretty big – $\lambda = 200$ km is certainly nothing to sneeze at! – but the largest water "waves" on the planet are actually the globe-spanning crests and troughs of oceanic water that we call the tides (whose wavelength is half of the circumference of the entire Earth, about 20000 km).

First, for those who might never have visited the shore, what are tides? Well, if you go to an ocean shore you'll notice that in addition to the waves coming in to the beach, there is an overall cycle to how far up the beach the water reaches. Every six hours and fifteen minutes the water level rises until "high tide" is reached, and then the water level falls for another six hours and fifteen minutes until "low tide" is reached, and the cycle repeats.

Next, what *causes* the tides? Remarkably, the dominant source is the gravitational pull of the moon! We know that the strength of the gravitational attraction between two objects depends on distance, but we usually think of the strength of gravity as being uniform and constant (for instance, with us always carrying around the same g for all of our "at the surface of the Earth" calculations). However, the Earth is big enough – relative to how far away the moon is – that the side of the Earth nearest the moon and the side of the Earth farthest from the moon experience slightly different amounts of gravitational pull from the moon.

If the Earth was made out of jello, or some other really deformable substance, we would see this effect in the shape of the Earth itself, which would look more like an egg than a sphere[18]. But the Earth is a pretty solid object, so the small variations in the moon's gravitational attraction don't bend it out of shape very much at all.

The water on the surface of the Earth, though? That can very easily adapt its shape to

[18] Of course, if the Earth were made out of jello...lots of things in life would be quite different.

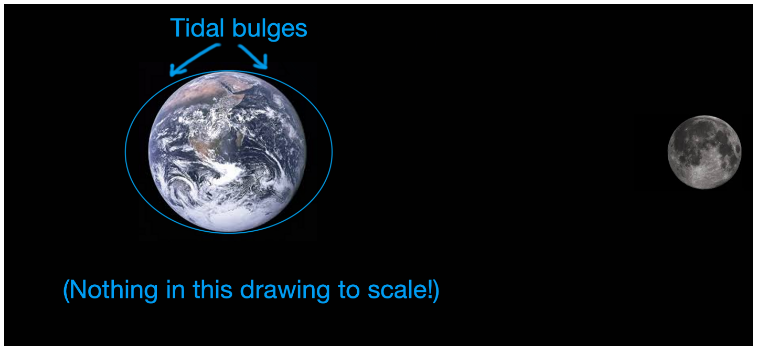

Figure 10.6: **The Moon's influence on the tides** The moon's gravitational field creates two tidal bulges in the Earth's oceans. As the Earth rotates once per day and the moon revolves around the Earth once per lunar month (about 29.5 days), these tidal bulges sweep across the Earth with a period of about 12 hours and 24 minutes.

reflect the gravitational fields it's experiencing. Here's what happens: the side of the Earth close to the moon feels a slightly stronger gravitational pull from the moon, so water on that side of the Earth bulges out slightly towards the moon: if you like, the water on that side is trying to accelerate towards the moon faster than the rest of the Earth is. At the same time, the water on the far side of the Earth feels a weaker gravitational pull from the moon, so the Earth is accelerating towards the moon faster than the water on the far side of the Earth is: the water on the far side of the moon *also bulges out away from the Earth*! This is illustrated in Fig. 10.6: at any moment on the surface of the Earth there are two bands of high tides and two bands of low tides flowing around the Earth with a period of about 12 and a half hours (the time from one high tide to the next).

The gravitational field of the sun *also* creates tidal forces here on Earth. On the one hand, the sun is a lot more massive than the moon, but it's also a lot farther away. The net effect is that the Sun's tidal influence is roughly half that of the moon. Two times per month the sun, Earth, and moon are all lined up, and the tidal forces from the sun and moon *reinforce* each other, leading to extra-high and extra-low tides, called spring tides. Two other times per month the sun, Earth, and moon make a right triangular shape, so the "high tide" from the sun overlaps with the "low tide" from the moon. The moon wins this fight, but on these two days the tides are less strong than usual, and are called neap tides.

Chapter 11

Light!

In this chapter we're going to take a whirlwind tour of some of the most important types of waves: *electromagnetic waves*! We talked a bit about them in the context of hot objects giving off electromagnetic radiation back in our discussion of thermodynamics, and in this chapter we'll really focus in on it. We'll start by talking more about what it *is*, and then talk about how it *behaves*[1]!

11.1 What is light?

11.1.1 Historical interlude

The experience of eyes detecting light has been – with the exception of the blind – common since much longer than the dawn of our species. But how do we actually *see* light? What is the nature of light, and what is the process by which our body detects it and sends some signal to the brain? In modern times we have gotten so used to the idea that light is some something which bounces off stuff in the outside world, enters our eyes, and gets detected by the specialized biological structures has been taught so well that it can be difficult to remember that this is *far* from obvious. Ancient Greeks[2] thought that a type of fire shot out of our eyes, "feeling" out for things in the world around us and making sight possible. Euclid[3] wasn't so sure of the "fire shooting out of our eyes" stuff, because he was pretty sure that light moved at a particular velocity; this made it quite the mystery that we could open our eyes at night and *instantly* see the stars (which were, presumably, quite far away!).

Something of a persistent thorn in the side of anyone trying to understand light for most of human history was the following: in many cases, light seems to behave like a tiny atomic packet of energy, as if there's some "particle" of light that can bounce off of objects, be absorbed by other objects, and so on. On the other hand, in many *other* cases the behavior of light only seems to make sense if it is a wave of some sort: it bends around corners like a wave does, it can constructively and destructively interfere with other sources of light like a wave, and the way it bends when moving from one transparent medium to another (say,

[1]And how, in a pinch, you can guess how deep in the ocean you are by looking at the color of a nearby lobster.

[2]Empedocles, 5th century BCE

[3]The "founder of geometry," living between the 4th and 3rd centuries BCE

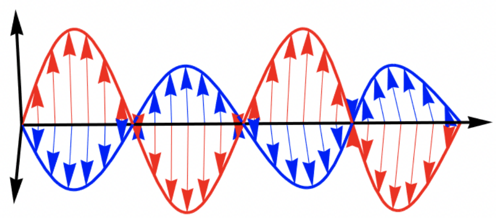

Figure 11.1: **An electromagnetic wave,** traveling to the right. The electric field (Red) and magnetic field (Blue) at each point along the direction of motion are at right angles to each other, and oscillate as the wave moves.

from air into water) makes the most sense if you think of light as a wave. We won't get into the details, but what was missing from most of human history is an understanding of *quantum mechanics* – quantum mechanics gives a precise mathematical explanation for why light sometimes "acts like" a particle and sometimes "acts like" a wave. It's not that it's somehow both some of the time, it's that it's a different *type* of entity (which we might call a "wavefunction"), neither particle nor wave.

In this class, to explain a lot of the most interesting ways we interact with light, I'm not going to talk about quantum mechanics, and instead I'll mostly talk about light as a very special *type* of wave, an "electromagnetic wave."

11.1.2 Electromagnetic radiation

All electromagnetic (EM) waves consist of a changing *electric* and *magnetic*[4] field traveling through space (by "through space" I mean the two waves are changing their positions – EM waves can indeed move through outer space, or air, or water, or any other transparent material). We haven't talked much about electricity and magnetism in this class, but electric and magnetic fields are physical, vector quantities that can permeate the world around us, exerting forces on electrical and magnetic objects that they interact with. Maybe you've held a magnet in your hand and used it to pick up a paper clip, or find a stud in the walls of your house: a magnetic field surrounds the magnet, ready to interact with magnetic materials close enough by. At every position in the universe we can talk about the value of the electric and magnetic field: it's a vector related to the forces that an electric charge or a magnetic object would feel if they were located at that point.

Figure 11.1 tries to show what an EM wave "looks" like. Like other traveling waves, it is moving in a particular direction (here, to the "right"), and it is composed of *something* oscillating back and forth. In this case, the thing oscillating back and forth is the *vector value of the electric and magnetic fields* at various positions along the direction the wave is traveling.[5]

[4]Hence the name

[5]An example might help: radio waves are a type of EM wave, and radio antennas are basically long metal rods. As a radio wave passes through the rod, the changing electric and magnetic fields make the electrons in the metal rod slosh back and forth; this gets converted into an electrical signal your radio can play back to you as sound

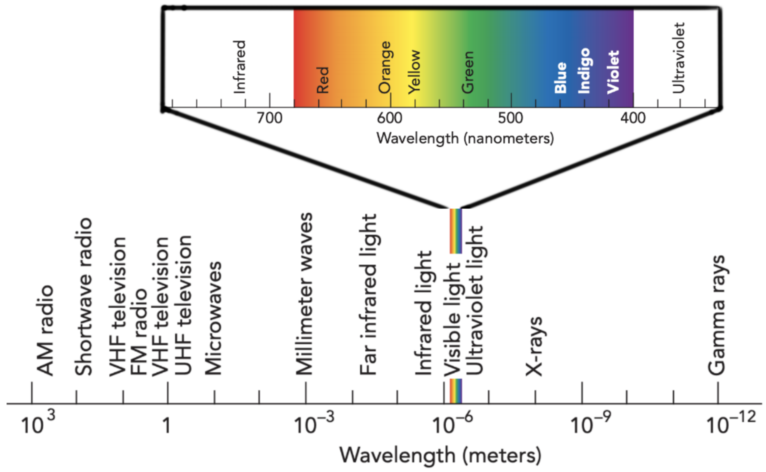

Figure 11.2: **The electromagnetic spectrum,** with different EM waves specified by their wavelength. Adapted from Figs. 7.3.2 and 7.3.3 of *How Things Work, The Physics of Everyday Life, 5th Edition*, by Louis A Bloomfield.

All electromagnetic waves are this same kind of thing: an oscillating couplet of electric and magnetic fields dancing through space in a very particular way. In the last chapter we met the idea that a traveling wave has a speed which is equal to the product of the wavelength of the wave multiplied by its frequency. Electromagnetic waves have an extremely special property: *the speed of an electromagnetic wave in a vacuum is a universal physical constant*, and no object with mass can travel faster than EM waves. The "speed of light" is a fundamental speed limit of physical objects in our universe, and the many interesting relationships that come from having such an absolute speed limit formed the core of Einstein's Theory of Special Relativity. In a vacuum the speed of these waves is

$$c = \lambda\nu = 299792458\texttt{m/s},$$

Where λ is traditionally used for the wavelength of an EM wave (just like in the last chapter) and ν is traditionally used for its frequency. As we'll see later on, when EM waves move through materials (like air, or water) they slow down, but in a vacuum they all travel at the same rate. The way we talk about different EM waves is by talking about their wavelength (or their frequency – given their speed in a vacuum if I tell you one you know the other!), and EM waves can in principle have any wavelength whatsoever, from wavelengths shorter than the size of an atom to wavelengths longer than the size of our solar system.

Figure 11.2 shows a representation of a wide portion of the *electromagnetic spectrum*, with the common names we give to EM waves of various wavelengths. Of note, you'll see the entire range of human perception that we call visible *visible light* is crammed into just a tiny portion of the entire spectrum, corresponding to electromagnetic waves with wavelengths between about 6.8×10^{-7} m (red light) to about 4.0×10^{-7} m (violet light) – ultimately these wavelengths are related to something about the size of the structures that our eyes use to detect incoming light. By thinking about our discussion of resonance in the last chapter, can you think of what the size of something to do with our eyes' photoreceptors should be to best detect visible light?

Before moving on, it's important to remember: electromagnetic radiation is *not a wave*, it's a different type of thing! Although EM radiation sometimes behaves like a wave, sometimes it also behaves like a particle (called a *photon*) that has no mass whatsoever but carries a small packet of energy with it. A crucial feature of EM radiation is that each little packet of it carries around a very specific amount of energy, determined entirely by the frequency of the EM wave (or, equivalently, by the wavelength). The amount of energy carried by a photon is

$$E_{photon} = h\nu = \frac{hc}{\lambda},$$

where "Planck's constant", $h = 6.626 \times 10^{-34}$ $\mathtt{J \cdot s}$, set the proportional relationship between energy and frequency. It is, you'll notice, a tiny number! Even though the frequencies of visible light are very large, they only carry tiny amounts of energy in each little packet. For instance, green light might have a wavelength of around 500×10^{-9} $\mathtt{m}$, which would correspond to a frequency of about 6×10^{14} $\mathtt{Hz}$. This is still an energy of only $E = 3.9 \times 10^{-19}$ $\mathtt{J}$. Since lightbulbs are rated in tens or hundreds of Joules per second, you can infer that lightbulbs are emitting enormous numbers of these little packets of light every second!

11.2 Mirrors!

Because the wavelengths of visible light are so small (tiny fractions of a single meter!), and because the energy that each tiny packet of light carries with it is so small, we can talk about the *behavior* of visible light using an approximate theory that goes by the name of "geometric optics." In this picture, when visible light is traveling through air (or through the vacuum of space, or through water, or through glass) it *moves in a straight line*, and interesting things only happen when a ray of light strikes an object, or moves between different types of medium (for instance, a beam of light going from air into water). Furthermore, in this picture, light rays doesn't meaningfully *interfere* with each other: when we're standing in a well-lit room, rays of light are bouncing all over the place, criss-crossing each others' paths constantly, and yet the rays of light that make their way into our eyes are not affected by the rays of light it might have passed through. All of these approximations are not-quite-right, but for everyday visible light the degree to which these approximations are incorrect is minuscule.

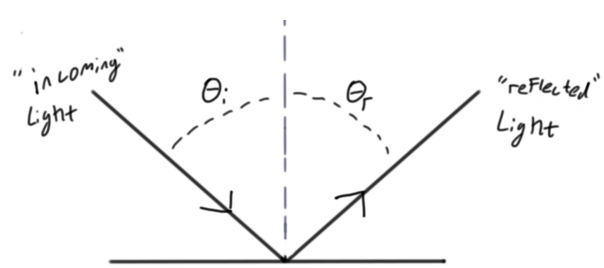

Figure 11.3: **Reflection of light off of a smooth surface:** the "angle of incidence" equals the "angle of reflection".

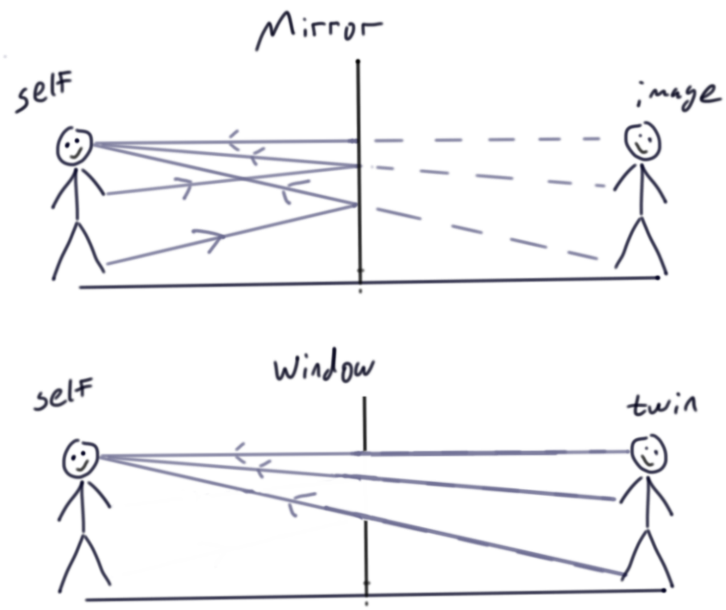

Figure 11.4: **A mirror and a window** Light rays reflect at equal angles when striking a flat mirror. When we see ourself in the mirror, our brain thinks it's seeing rays of light that originated on the "other side" of the mirror, similar to if we were peering through a window at our twin.

So, to start out we have light going in straight lines between to points (say, the framed photo on my desk and my eye), and now we want to ask what happens when light hits various materials (say, a ray of light that is emitted by the light bulb on the ceiling and hits the photo on my desk). The simplest object is a *flat mirror*. When a ray of light strikes a mirror, it does not continue traveling in a straight line, but rather the direction of travel changes. How does it change? Simple: the light hitting the mirror changes so that the angles are equal. That is, the angle between the incoming beam and the mirror is *the same* as the angle between the outgoing (or "reflected" beam and the mirror. This is shown in Fig. 11.3, and in symbols it is often written

$$\theta_i = \theta_r.$$

Our brains aren't pre-wired to process the actual paths of light rays, so when we see an object in a mirror our brains interpret the image as if rays of light actually originated behind the mirror. By carefully tracing out where the light hitting our eye must have *actually* originated from, as in Fig. 11.4, we can understand what we are actually seeing in a mirror. Based on this discussion, can you sketch out some of the paths that different rays of light must be taking if you're staring not at a flat mirror but at one of those curved "fun house" mirrors?

Mirrors are simple enough, and light bouncing off any smooth surface basically behaves in the same way – the angle of incoming and outgoing light is the same. Mirrors are special in that they are smooth even when viewed at the microscopic level (similar, actually, to very still water)! So, beams of light come in, and beams of light bounce off; this is called *specular reflection*. Most surfaces that look pretty flat to us are actually pretty rough at the microscopic level: light continues to obey the equation above for incident and reflected light, but the effective orientation of the surface changes rapidly from one spot to the next. For specular reflection, bundles of nearby light stay nice and near each other, but for *diffuse* reflection this is not the case. Rather than being able to see a copy of our face when we stare at a painted wall, the light going from our wall to our eyes started in all kinds of uncorrelated places.

Speaking of "painted" – most material objects will reflect some light that strikes them and *absorb* some of the light that strikes them, and the process of how this happens (which rays of light get reflected and which get absorbed) depends on the material and the *wavelength* of the light. So, an object that appears, say, blue is reflecting most of the blue-wavelength light coming at it and absorbing most of the every-other-color-wavelength light coming at it.

11.3 Rainbows!

Light bouncing off of stuff is pretty straightforward, but light does something very interesting when it goes from one medium into another, say, from air into water. You may have noticed that if you put a straw in a half-full glass of water it appears bent, or if you're looking down at a rock in a shallow lake and naively try to grab at where it appears you might miss. Why does light bend like that, and how much does it bend[6]?

11.3.1 Refraction of light

To help us understand why light bends as it goes from air to water, let's take a strange detour to Fig. 11.5, where we consider a lifeguard that suddenly has to save a swimmer. Any given lifeguard will be able to run over the sand faster than they can swim through the water, so when a lifeguard hears a cry for help what should they do? In the moment, most of them probably take a straight-line path over the beach, plunge into the water, and swim out to save the swimmer. However: since they can run faster than they can swim, a lifeguard could get to the swimmer *faster* by not taking a straight-line path, but rather by taking a path composed of two straight lines – one on the beach, and one through the water – that bend by a certain amount at the beach-ocean interface. What angle should the two straight lines bend at to *minimize the time it takes for the lifeguard*? It depends on the lifeguard's speed over land, speed in the water, and the angle of the straight-line path.

We might call this idea the "Principle of Least Time[7]: the idea that sometime what you are trying to do is minimize the amount of time it takes you to do something, instead of just minimizing the distance you have to travel to do it. It turns out – for reasons intimately connected with the proper quantum mechanical description of electromagnetic waves – that

[6]Fun historical aside: this question stumped the ancient Greek philosophers, and it is one of the few examples I know of where the ancient Greeks (Claudius Ptolemy, in this case) actually listed *experimental data* (in this case, a table of measurements of a beam of light striking the surface of water at some angle and being deflected by some angle). Remember that footnote from the first chapter on the Laws of Motion, where Aristotle believed that women had fewer teeth than men? Recalling, also, that Aristotle had a wife and could count should help emphasize how little the ancient Greeks thought experimental data should influence the intellectual process of building a theory to understand the world

[7]And, in the context of light, what's sometimes called "Fermat's Principle."

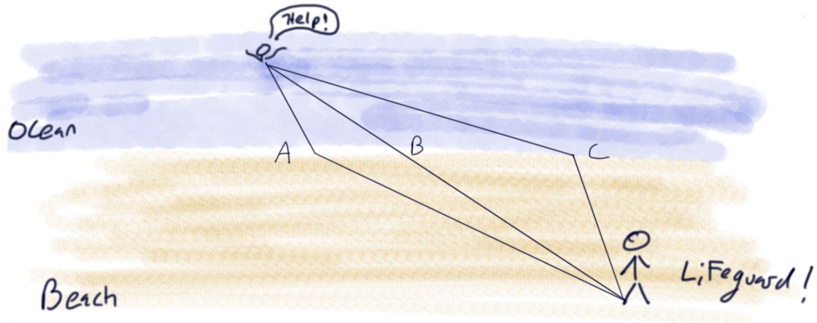

Figure 11.5: **Rescue at the beach!** A lifeguard is standing on the beach and hears a cry for help! The lifeguard can run over the sand much faster than she can swim; what path should she take to save the floundering swimmer? Path A! Notice, also, that how different path A should be from path B (the straight line path) depends on the angle between the lifeguard and the swimmer, and on how different the lifeguard's speed is on land and in water.

light obeys a principle of least time – the path that light follows is the one that minimizes not the distance it travels but the time it spends traveling.

This becomes interesting when you think about the case of mirrors above: we wrote down a rule that light obeys there, $\theta_i = \theta_r$. This mathematical equation can be derived by assuming *either* that light takes the path of least distance *or* that light takes the path of least time[8]. But if light moves from, say, air to water, and *if* the speed of light is different in those two substances, the "shortest path" description of light's journey and the "fastest time" description no longer give the same predictions! It's like our lifeguard on the beach: if the lifeguard minimizes the distance she takes the straight-line path, and if she minimizes the time she will take a path that goes farther in the substance she moves quickest in.

It turns out that light does, in fact, move at different speeds through different substances! In a vacuum the speed of light is a fundamental constant, but when moving through stuff light moves a bit slower. For instance, in water light travels about 1.33 times slower than it does in a vacuum, and we can quantify this by defining the *index of refraction* for a material as the ratio of the speed of light in a vacuum to the speed of light in a material: $n = c/v$. For water $n \approx 1.33$, for air $n \approx 1.0003$, for window glass $n \approx 1.52$, for diamond $n \approx 2.42$, and so on.

If we know the different speeds that light travels in different materials, we can figure out what path the light will take to minimize the time in getting between any two points. We can quantify this by writing down an equation for what we see in Fig. 11.6, where an "incident" ray of light moves from one medium to another and experiences a change in the angle of its

[8]That is, the shortest and quickest path for light to go from your foot to your eye is, well, a straight line from your foot to your eye. But if you add the condition that the light has to hit the mirror first, then the shortest and quickest path is the one where the light bounce off the mirror at the same angle it approaches the mirror.

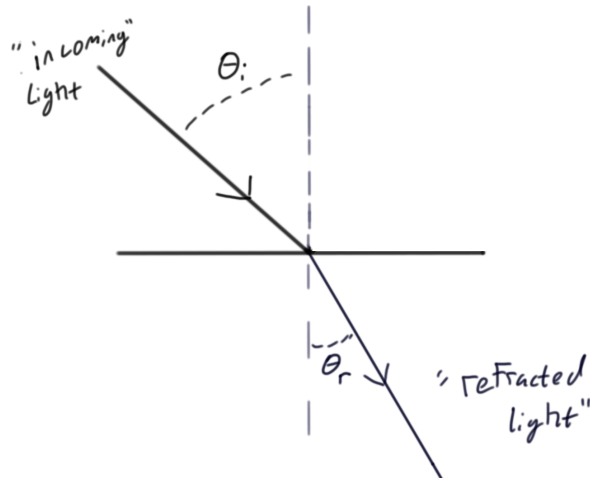

Figure 11.6: **Refraction of light** When a ray of light moves from one medium into another (say, from air into water) it *refracts*, turning away from it's original course.

trajectory. Mathematically, the relationship between the angles is captured by Snell's Law[9]

$$n_1 \sin\theta_i = n_2 \sin\theta_r,$$

the product of the index of refraction of the material light is moving through initially times the sine of the incident angle is equal to the index of refraction of the material light is moving into times the sine of the angle of the "refracted" light.

11.3.2 Chromatic aberration

Reflection and refraction – as we're about to see – are major reasons that we can look up into the skies on certain days and see beautiful rainbows; those processes control the how incoming sunlight bounces around when it meets a raindrop! But to explain the brilliant bands of color, I need to tell you one more little fact about the speed of light: in vacuum, it's always the same, no matter what wavelength of EM radiation we're talking about. When light is passing through a material, though (like air or water), the speed of an EM wave depends *just a little bit* on its wavelength. This is called the *dispersion* of light.

But we know from Snell's Law that the amount light bends depends on variations in the index of refraction of the materials its moving through, and the dispersion of light means that *different colors of light will bend by slightly different amounts* as they go from one medium to the next! This is why, for instance, a prism can split "white light" (composed, in fact, of every color of light moving together in a group) into a, well, prism of colors, as in Fig. 11.7. This phenomena is sometimes called "chromatic aberration," because back in

[9]Named after a Dutch mathematician, and from the name of the law you might think this mathematician's name was "Snell." Wrong! He was born "Willebrord Snel van Royen," and he seems to have gone by "Willebrord Snellius," both of which were apparently too complicated for English-speaking people of his day. As an aside, Snell worked out Snell's Law in the early 1600s, and that's more or less when the phenomenon of refraction was quantitatively understood by European scientists. The Persian mathematician Ibn Sahl clearly knew and used what we now call "Snell's Law" all the way back in 984.

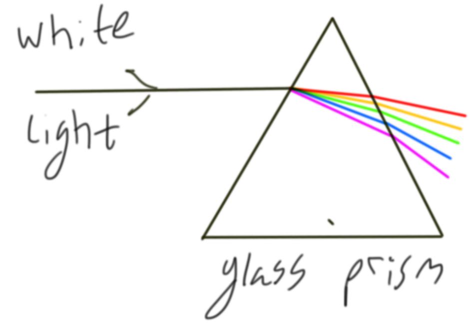

Figure 11.7: **A prism** Light of different wavelengths moves at very slightly different speeds through a material, so it refracts by different amounts! This lets a prism separate out different colors of light.

the Renaissance it really got in the way of making good lenses: a lens that perfectly focused light of one color at some distance would be slightly wrong at focusing light of a different color.

11.3.3 Sun and rain

Knowing about reflection, refraction, and chromatic aberration, we can now understand what we're looking at when we see a rainbow. First, let's think about what happens when a beam of incoming white sunlight hits a spherical droplet. Some of the light will reflect off of the droplet and stay in the air, but some of it will enter the water and *refract*: the precise angle it refracts, though, depends on its color. The chromatically separated wavelengths of light will hit the back of the water droplet at slightly different places, and once again some of the light will reflect and stay in the water droplet, and some will pass back out into the air behind the droplet. This process keeps going: some of the light inside the droplet reflects and some refracts every time the light meets the surface of the droplet. For one internal reflection, violet light will ultimately leave the droplet at a very slightly different angle (relative to the original incoming light) than red light will – a difference of just a few degrees. This whole scenario is illustrated in Fig. 11.8.

Finally, if you are standing with the sun behind you and if the sunlight can hit many droplets of rain in front of you, the above phenomena gives you a chance to see a rainbow. Compared to the position of the sun, there will be an entire circle of positions where the sun and the rain and your eye makes a 40° angle, and you will see an arc of purple. Likewise, there will be an entire circle of positions where the sun, drops of rain, and eye makes a 42° angle, and you will see an arc of red. This is schematically illustrated in Fig. 11.9 In between you will see all of the other colors of the rainbow, although how well you can make them out depends on your vision and atmospheric conditions. Of course, most of the time rainbows are just semi-circles – the ground kind of gets in the way of you seeing a full circular rainbow – but if you're in a plane (or skydiving!!) and conditions are just right be on the lookout!

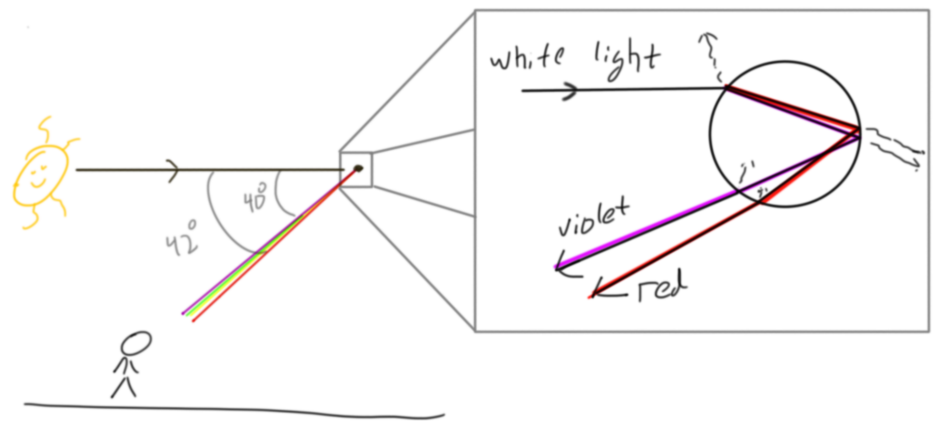

Figure 11.8: **Light reflecting and refracting around in a raindrop** When white light from the sun meets a raindrop, the different wavelengths of light take slightly different paths as the light reflects off of the surface of the droplet and refracts into and out of the water from the air. Violet light will make an outgoing angle of about 40°, and red light will make an outgoing angle of about 42°, with the incoming ray of white light.

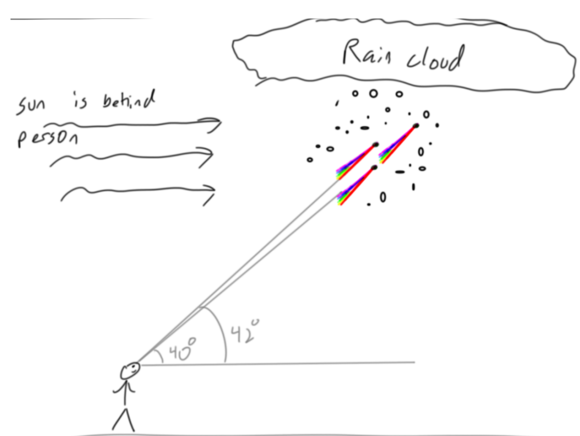

Figure 11.9: **Seeing a rainbow!** Because different colors of light will ultimately leave a water droplet at different angles, when an observer has their back to the sun and is looking at raindrops they will see (semi)-circular arcs of colors across the sky. The center of the rainbow will be directly opposite the sun itself.

11.4 The sky is blue!

We'll now explain why the sky is blue – perhaps one of the most elemental observations in the entire course of human history, but one for which we really understood the "why" of it shortly after the turn of the 20th century. To get started, we'll learn about what happens when EM radiation interacts with something small.

In the section on mirrors we saw what happens when an packet of electromagnetic radiation hits a large object: some of it gets reflected, some of it might get absorbed, and (if the material is transparent), some of it might pass into the material, bending according to the differences in indices of refraction. What happens, though, when a packet of electromagnetic radiation hits a small particle? Let's take a particular *dimensionless parameter* that will help characterize what we mean by small. We won't even give this number a particular name, but we'll think about a spherical particle of radius r being hit by a bit of EM radiation of wavelength λ, and consider the ratio

$$x = \frac{2\pi r}{\lambda}.$$

When $x \gg 1$, that is, when the object is very big compared to the wavelength of EM radiation we're thinking about, things are relatively simple: this is the regime of our everyday experience. When $x \approx 1$ the interaction of light with the object can be quite interesting[10]. What we'll focus on in this section, though, is called *Rayleigh scattering*, which occurs when $x \ll 1$: that is, when the object the light is hitting is much smaller than the wavelength of the light.

11.4.1 Tyndall and Rayleigh

Initial clues about, deep down, why the sky is blue come from some experiments conducted by John Tyndall[11] in 1859. Tyndall was conducting experiments in which he suspended tiny microscopic particles in clear fluids, finding that bluer light scattered more strongly than redder light when shining through the suspension. That is, redder (longer wavelength) light would be more likely to pass right through the fluid, whereas bluer (shorter wavelength) light would be more more likely to interact with something in the fluid and come out in a totally different direction. The difference was not huge, but it was noticeable. This phenomenon was studied in great detail later by Lord Rayleigh in the late 19th century, whose careful theoretical work showed that Tyndall's data could be explained by electromagnetic waves, and that his equations could apply not only to little particles, but also to individual atoms and molecules!

I don't want to present the full equations for how much light gets scattered, but the core of the result is this: when light of wavelength λ is moving through a material made up of spheres of radius r, the *scattering cross section* σ_s is proportional to six powers of the radius

[10]The is called the "Mie scattering" regime, and aerosols and bits of dust floating in the atmosphere have the right size to have $x \sim 1$ for wavelengths in the visible part of the spectrum.

[11]An Irish physicist who, among other things, provided some of the foundational work in our understanding of the greenhouse effect.

divided by four powers of the wavelength:

$$\sigma_s \propto \frac{r^6}{\lambda^4}.$$

This scattering cross section helps tell us how much light makes it through a material and how much gets scattered:

$$\text{fraction of light scattered} = \text{distance} \cdot \frac{\text{Number of scatterers}}{\text{Volume}} \cdot \sigma_s.$$

That is, if you send in a bunch of light of a particular wavelength, for each meter it travels a fraction of it will get scattered determined by the number of particles per unit volume (the number density of potential scattering particles in the way) multiplied by the scattering cross section.

11.4.2 Einstein

From Tyndall and Rayleigh's work we learned the major reason the sky is blue. The above equation tells us that the fraction of light scattered is proportional to λ^{-4}. Since blue light (say, with $\lambda = 400 \times 10^{-9}$ m) has a smaller wavelength than red light (say, with $\lambda = 700 \times 10^{-9}$ m), blue light will be scattered more by a factor of about $(700/400)^4 \approx 9$. We'll come back to that in a second. This earlier work could help explain light scattering in the sky, but nobody really knew how to actually *calculate* the scattering cross section, and so there was a constant debate: was the sky blue because tiny water droplets and bits of dust in the air were scattering light of different wavelengths differently, or was the sky blue because the *atoms and molecules of the air itself* were scattering light of different wavelengths differently?

This puzzle was finally settled in 1910, when Albert Einstein and Ludwig Hopf published a paper[12] that actually calculated a detailed formula for the scattering of light by molecules in the air, and that detailed formula agreed with what people could experimentally measure.

Here are some actual numbers, in case you're curious! Most of the air in the atmosphere is composed of nitrogen, and the scattering cross section of nitrogen when it is hit with green light of $\lambda = 532 \times 10^{-9}$ m is about $\sigma_s \approx 5.1 \times 10^{-31}$ m^2. At atmospheric pressure there are about 2×10^{25} nitrogen molecules per cubic meter, so we see that for every meter that green light travels through the air about 1/100000 is scattered away. Not a huge amount, but enough to make a difference!

11.4.3 What the sky looks like

So, on a bright sunny day, with the sun beaming light of all different wavelengths at the Earth, we look up into the sky and see a bunch of things. If we look at a patch of sky where the sun *isn't*, what we're seeing are little packets of light that *could not have come in a straight line from the sun*! Instead we see light that had to have bounced off of something on its journey into our eyeballs, and since shorter wavelengths are more likely to be scattered

[12] "Statistische Untersuchung der Bewegung eines Resonators in einem Strahlungsfeld"

off of the molecules in the air, we see a beautiful blue[13] sky! If we look at the sun itself[14] we see more of the wavelengths of light that *didn't* scatter away, giving the sun itself a yellowish tinge (composed of the red / yellow / orange part of the spectrum).

We look over at a fluffy cloud, and it is a brilliant white because the droplets of water inside the cloud do an excellent job of scattering light of all colors around, so the "white" sunlight – a combination of all of the colors – all gets scattered into our eyes pretty well.

As the sun is setting (or rising), any light moving in a straight line from it has to travel farther through the atmosphere to get to our eyes than when the sun is overhead. Because of this extra distance, even more of the green, blue, violet, and indigo light gets scattered away; the remaining un-scattered light appears quite red. Polluted air tends, sadly, to make for more beautiful sunsets, as small particles help scatter and can make the skies a more vibrant shade of red. Over the ocean there are some extra salt particles floating in the air, which are quite good scatterers and can lend the evening sky a lovely orange hue.

[13] You might wonder: "Why not a purple sky? After all – violet has an even shorter wavelength than blue does, so it scatters even more!" That's very true, and there are two main reasons why the sky looks blue rather than violet: First, even though the sun emits light of all frequencies, it doesn't emit equal amounts of all of them. Actually, the most likely light to be emitted by the sun is near $\lambda \approx 500 \times 10^{-9}$ m, which is in between green and blue. So, first, there is less purple light to be scattered than there is blue light to be scattered in the first place. Secondly, *human eyes* detect color with specialized receptor cells, and they do better and worse jobs of detecting various wavelength of light.

[14] Don't, by the way!

Chapter 12

Random walks!

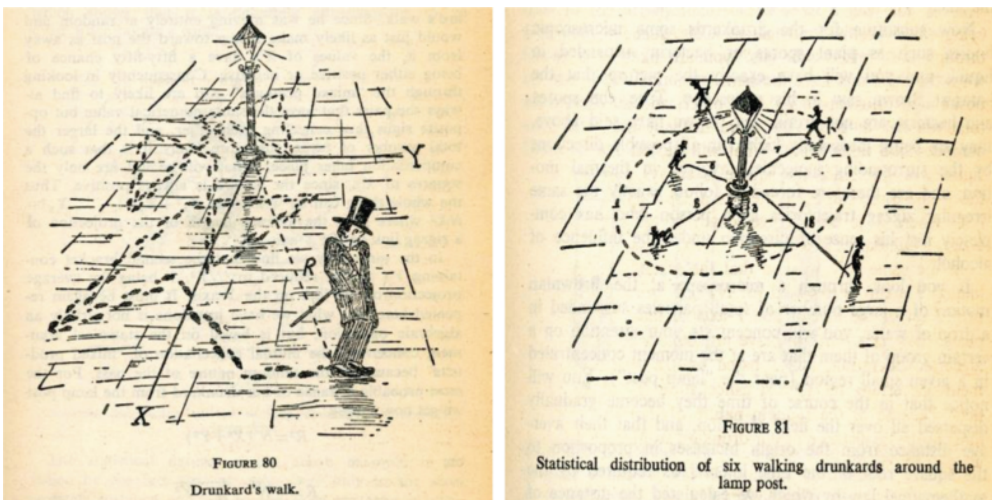

Figure 12.1: **The Drunkard's Walk** (Left) A rather-too-tipsy gentleman in an oversized top hat takes shuffling steps away from a light post, turning in a random direction every so often. After some time he is located $\vec{R}$ away from where he started. (Right) If this random process is repeated more than once, the drunkard would end at different vector displacements from his starting point each time. Figure reproduced without permission from "One, Two, Three – Infinity: Facts and Speculations of Science" by George Gamow (1947).

In 1905, back before googling an answer was possible, Karl Pearson[1] sent a note titled "*The Problem of the Random Walk*" to the scientific journal *Nature*:

> Can any of your readers refer me to a work wherein I should find a solution of the following problem, or failing the knowledge of any existing solution provide me with an original one? I should be extremely grateful for aid in the matter. A man starts from a point O and walks I yards in a straight line; he then turns through

[1]A brilliant English mathematician, often considered the founder of mathematical statistics (as well as being the actual founder of the worlds first statistics department, at University College, London). He was also a believer in social Darwinism and a proponent of eugenics (and even believed that his grotesque political beliefs were justified by his work in statistics).

any angle whatever and walks another I yards in a second straight line. He repeats this process n times. I require the probability that after these n stretches he is at a distance between r and $r + Sr$ from his starting point, O. The problem is one of considerable interest...

This problem[2] is, indeed, one of *considerable* interest! Colorfully depicted in Fig. 12.1, where it was called a "drunkard's walk," these random walks – the idea of a process that results from many random steps in random directions – comes up in a bewildering array of superficially disconnected phenomena, not only in physics but in biology, chemistry, economics, mathematics, etc etc. In this short chapter we'll briefly meet random walks in the context of repeatedly flipping a coin; we'll then see how this idea comes up in Fermi problems and face masks.

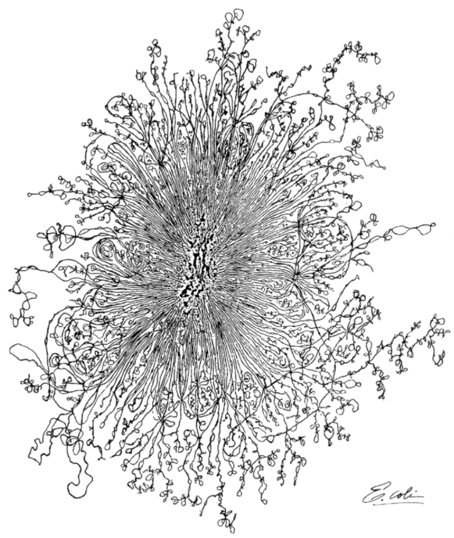

Figure 12.2: **Genomic DNA of a ruptured E. coli** Genomic DNA spilling out of a ruptured bacturium! The total length of DNA in an E. coli is roughly 1000 times longer than the bacterium itself, and if the cell wall is ruptured it all spills out. Zooming in on the path the DNA takes at the edges of this image reveals a random-walk-type structure! Electron micrograph reproduced from the Schmidt lab at Pittsburgh

12.1 Random walks

Let's start our discussion of random walks by playing a game in which I will repeatedly flip a coin. Every time it comes up heads I'll take a step forward, and every time it comes up tails I'll take a step backwards. Assuming the coin is a fair one, each flip has a fifty-fifty chance of coming up either heads or tails, and the game will be to find out how far away

[2]Incidentally, Lord Rayleigh (the same one we met in the last chapter) responded to Pearson, pointing out that this problem was mathematically equivalent to one he had partially solved in an 1880 paper titled "One the Problem of Random Vibrations, and of Random Flights in one, two, or three Dimensions"

from my starting point I get after a certain number of flips of the coin[3]. This is something of a simplified version of the random walk introduced above, but it has the essential features of what we want to understand.

Before analyzing how this game plays out, I want to emphasize something about what we expect as this game is being played: we expect that after I take some number of steps I almost certainly won't be as far away from my starting point as if I just pick a direction and take the same number of steps in that direction! The random process of flipping the coin will give me a back-and-forth motion, and it will take me much less far than if I just had some velocity and moved with that velocity for a certain amount of time.

Let's now look at our random forwards-and-backwards walk, governed by the whims of the coin we're flipping. Let's let x equal our position forward or backward relative to our starting place, and let's let angled brackets stand for "take the average." So, for instance, $\langle x\rangle$ is equal to our average position, if we were to play this game with a fixed number of coin flips many many times.

If we decide to only flip the coin once, the results of the game are pretty clear: half of the time I'll end up a step ahead ($x = 1$) and half of the time I'll end up a step behind ($x = -1$) where I started. Notice that the average of these positions is $\langle x\rangle = 0$, but that doesn't mean that I typically end up where I started! It's just that I'm equally likely to end up a positive or negative amount from where I started. We can account for this "equally likely to be forwards as backwards" by averaging the square of our position: $(+1)^2 = (-1)^2$, and $\langle x^2\rangle = 1$.

Okay, what if I decide to flip the coin twice before stopping? The equally likely outcomes are HH, HT, TH, and TT, so half the time I end up exactly where I started ($x = 0$), a quarter of the time I end up two steps ahead ($x = 2$), and a quarter of the time I end up two steps behind ($x = -2$) where I started. Again, my average *position* is zero – I was equally likely to go forwards as backwards – but the average of x^2 now is

$$\frac{1}{4}\left((-2)^2 + 0 + 0 + (+2)^2\right) = 2.$$

What about if we stop after three coins? There are eight possible coin-flip combinations: HHH, HHT, HTH, HTT, THH, THT, TTH, TTT. A short little bit of arithmetic should convince you that the average of x is again zero, but the average of x^2 is three. We could keep explicitly enumerating all possible outcomes of ever-increasing numbers of coin-flips, but perhaps you already see the pattern! If we flip the coin n times, our average position is always zero, and the average of x^2 (which is standing in as a measure of how far away from the start we end up) is equal to n. In Fig. 12.3 I give the *probability distributions* associated with the chances of ending up at different positions for various numbers of coin flips – as the number of coin flips gets very large perhaps you recognize the famous *Gaussian* distribution (also called the "normal" distribution or "bell-curve" distribution) in the pattern of final positions.

The key result for these *unbiased random walks* where we take n steps either forward or backwards is that $\langle x\rangle = 0$ and $\langle x^2\rangle = n$. Compare this to a different[4] game where no matter

[3]Well, I never said it would be a *fun* game...

[4]and particularly boring

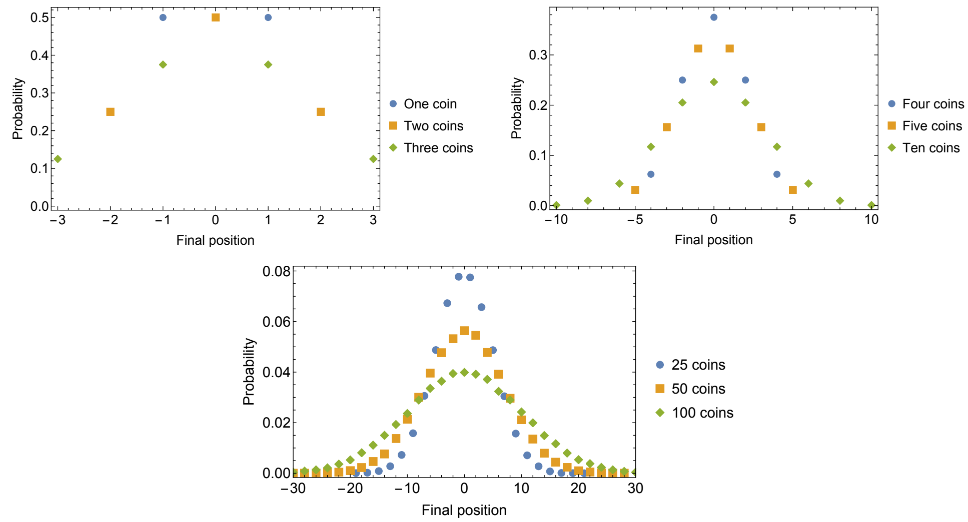

Figure 12.3: **The coin-flip game** You play a game where you start in a particular location and then flip a coin a certain number of times. Every time it lands on heads you take a step forward, and when it lands on tails you take a step backwards. For each graph, the x-axis shows possible final positions, and the y-axis shows the probability of being in that final position. Each set of points shows the *distribution of results* if you play the game using a particular number of coin flips. For many coin flips (bottom panel), the distribution looks unmistakably like a Gaussian distribution.

what the result of the coin flip is we always take a step forwards: in that game – which is like moving with a constant velocity, with no randomness at all! – we have $\langle x \rangle = n$ and $\langle x^2 \rangle = n^2$.

12.2 Fermi estimation revisited

It might not look like it at first glance, but our scheme for arriving at order-of-magnitude estimates is a bit like a random walk in disguise! Let's see how that's the case, and how it helps us understand why Fermi estimation is so handy.

The key is that we are trying to estimate the *order of magnitude* of the answer to our problem – let's write the correct answer as 10^y, and we're trying to get a good estimate for y. Our first step is to break the whole problem into a bunch of sub-problems that we'll then multiply together: maybe we break down the big question in 4 subquestions; each of those subquestions has a correct value – let's label them 10^{x_1}, 10^{x_2}, 10^{x_3}, and 10^{x_4} – and if we knew each of those correct values we would correctly estimate

$$10^{x_1} \times 10^{x_2} \times 10^{x_3} \times 10^{x_4} = 10^{x_1+x_2+x_3+x_4} = 10^y.$$

Typically, though, even after we break something into a sub-problem, we arrive at a subproblem we don't know the value of – we're force to try to make an informed estimate! That is, say, we don't actually know what x_1 is, so we make a guess for it. Our "geometric mean" rule for guessing corresponded to thinking of an overestimate (say, 10^a) and and underestimate (say, 10^b), and guessing that $x_1 = (1+b)/2$.

Whether that's the rule we use or not, we can think of this process like a random walk! Each time we guess at one of the exponents, x_i, for the answer to a subproblem, it's like we are taking a random step *away from the true answer*! Just like we added plus and minus ones in our coin-flip game, here we're adding the guesses to our subproblem to find a final answer. We set up our Fermi problem so that our final answer will be a sum of the exponents we guess, and so our final answer will be some typical "distance" away from the right answer. The random walk here is a random walk of *errors*, in a sense. But if we can break our problem down into steps, *and* if when we break down our problem into subproblems we have a tighter range for each of our sub-estimates, our "random walk" will not lead us as far away from the true answer as we might if we just guessed from the beginning!

For example, suppose you start with a problem and at first your best underestimate and overestimate are three orders of magnitude apart. You could guess from there, but you shouldn't be surprised if you are sometimes three orders of magnitude wrong! Now (to keep things simple), suppose you can break that big problem into 10 subproblems, each of which you think you are only underestimating or overestimating by a factor of 2. In the worst case you will end up overestimating *all* of the subproblems by a factor of 2 (or *underestimating* them all by that factor of 2), and you will be $2^{10} \approx 1000$ off from the right answer – just as bad as wildly guessing originally.

But, in the more typical case, your underestimates and your overestimates will help cancel each other out – just like with a random walk of 10 coin flips you expect to be standing $\sqrt{10}$ steps away[5], we expect our guess to be within $2^{\sqrt{10}} \approx 8.95$ of the right answer. That's within an order of magnitude, and we feel good about ourselves.

For real Fermi problems it's a bit more complicated than this simple analysis: we typically have different amounts of uncertainty about each of our sub-estimates (rather than knowing that all of them are within a factor of two)! Sometimes we get a subproblem we *know* the answer to, and other times we get a subproblem that still has a broad range of uncertainty. This kind of generalization of our simple coin-flip random walk – where we take steps of different lengths in our walk – is a bit more complicated, but the key ideas are the same: the random under- and over-estimates help cancel each other out, and our errors grow more slowly than "linearly" in the number of subproblems.

12.3 Masks and COVID-19

12.3.1 Particles, big and small

At any moment there are lots of *airborne particles* – little solids, or droplets of liquid – suspended in the air around us, and these particles come in an incredible variety of typical sizes[6]. Figure 12.4 shows a small collection of "typical sizes" of things floating around, and

[5] Remember, for the 10-coin-flip game $\langle x^2 \rangle = 10$, but x^2 has dimensions of $\mathbb{L}^2$, "distance squared," so we take the square root to get a "typical distance away from the start" measurement.

[6] In this section rather than working in meters we'll be working in *micrometers*, $1\ \mu\mathtt{m} = 10^{-6}\ \mathtt{m}$

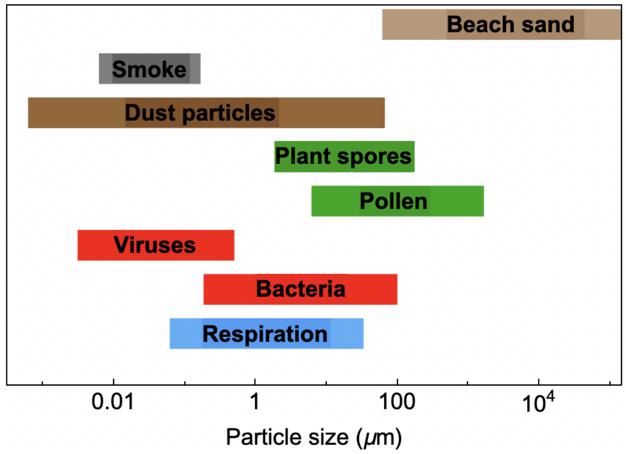

Figure 12.4: **Typical sizes of different things** Different types of things can come in very different sizes. The plot above shows the range of characteristic sizes of things in microns ($1\ \mu\mathtt{m} = 10^{-6}\ \mathtt{m}$). Data are from The Engineering Toolbox and Soft Matter Science and the COVID-19 Pandemic

an important thing we'll discover in this chapter is that the size of an airborne particle makes an enormous difference in the *way that particle moves around*.

For instance, "large" particles, which for the purposes of this section mean airborne particles bigger than $100\ \mu\mathtt{m}$ (a tenth of a millimeter, or about 4 thousandths of an inch) basically move in straight lines, and typically just fall down. This includes particles like snow flakes, drops of rain, and household dust. Strong winds can blow these particles around, but they tend to settle down to the Earth pretty quickly.

That's basically our experience of the behavior of most objects in the everyday world – we're used to interacting with things that are on our own size scale, and Newton's Law's tell us how those kinds of objects move around (maybe with a bit of help from the chapters on fluids, when we have to think about different levels of drag, on so on). What could be different?

To help figure out what could be different, let's do a rough little calculation. Suppose I consider a spherical droplet of water at rest in the air – let's ignore gravity for a quick second – and I imagine an oxygen molecule in the air colliding with the droplet. I want to know if it's possible that after this collision the droplet has picked up a velocity that has it moving about its own size every second. I hope it makes intuitive sense that such a size *exists* – oxygen molecules are moving at several hundred meters per second in the air, so if the droplet was only the size of a single water molecule a collision would send it flying off *much faster* than its own size per second, and for really big drops of water a collision with a single oxygen molecule would do basically nothing at all! It's not a question of whether this special-sized droplet *exists*; it's a question of *what size* such a droplet is!

Let's estimate the size of such a droplet by thinking about conservation of momentum. First, in the chapter on Fluids we met approximate numbers for the mass of an oxygen molecule ($m \approx 5.3 \times 10^{-26}\ \mathtt{kg}$) and its typical speed in the air ($v \approx 500\ \mathtt{m/s}$), so we already know the a typical magnitude for the momentum of such an oxygen molecule to be around $2.66 \times 10^{-23}\ \mathtt{kg \cdot m/s}$. Let's not worry too much about the details of the collision (was it elastic? inelastic? what was the coefficient of restitution?), but just ball-park the numbers by assuming that the water droplet gets about this much of an impulse. So, we want to equation that number with mv for the droplet, and we want to solve for the radius of the droplet assuming that $v \approx r/(1\ \mathtt{s})$. Thus:

$$2.66 \times 10^{-23}\ \mathtt{kg \cdot m/s} = mv = (\rho_w V_w)\, v = \left(977\frac{\mathtt{kg}}{\mathtt{m}^3}\right)\left(\frac{4}{3}\pi r^3\right)\frac{r}{1\ \mathtt{s}},$$

$$\Rightarrow r \approx 2.8 \times 10^{-7}\ \mathtt{m}.$$

What we've just learned is that droplets of water that have a radius of about 300 $\mathtt{nm}$ = 0.3 $\mu\mathtt{m}$ will get kicked around pretty easily just by being hit by the oxygen molecules whizzing around the air! The droplet will be bombarded by a bunch of such collisions, and each one is enough to make it start moving by its own size per second. That is: *small droplets of water take random walks in the air*!

So, "large" airborne particles (bigger than 100 $\mu\mathtt{m}$) are carried by their own inertia, falling out of the sky in basically straight lines pretty quickly under the influence of gravity. "Very small" airborne particles (smaller than about .1 $\mu\mathtt{m}$) take random walks through the air, getting big enough kicks to keep them aloft and randomly moving about for *extremely* long times. "Medium" airborne particles (say, between 1 $\mu\mathtt{m}$ and 100 $\mu\mathtt{m}$ – things like pollen, plant spores, respiratory droplets from coughs) fall slowly out of the air, and can be easily carried by wind relatively large distances. Smaller than that (viruses, bacteria, smoke particles, the small end of respiratory water droplets), and particles start to be affected more and more strongly by collisions with air molecules. They can take days (or years!) to settle out of the air, and if they make their way up into some of the more turbulent layers of the atmosphere they might be suspended more or less indefinitely.

12.3.2 A mask is not a sieve

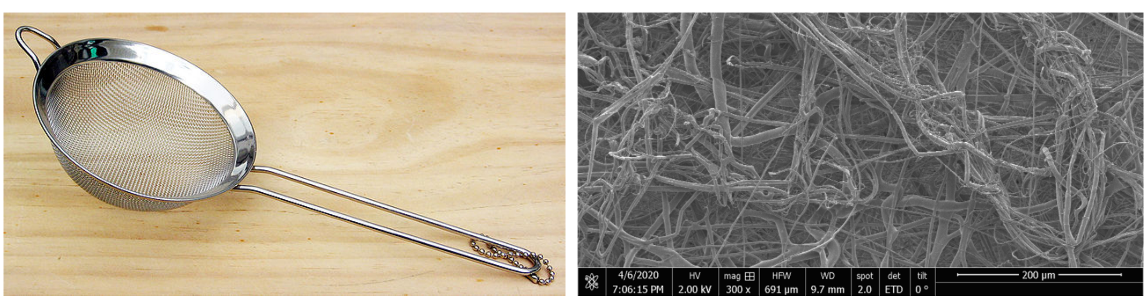

Figure 12.5: **A mask is not a sieve** (Left) A sieve uses a mesh to filter out everything bigger than the size of the holes in the mesh; anything smaller falls through. Image taken by Donovan Govan, CC BY-SA 3.0 (Right) A scanning electron microscopy image zooming in on a commercially available N95 mask (scale bar in the lower right corner). Note the many layers of criss-crossing fibers, each layer having "large" holes in it. Image by the Bandi lab at the Okinawa Institute of Science and Technology.

With so many different sizes of particles, perhaps you start to appreciate the interesting design challenge that exists when it comes to making a mask: you need the mask to stop big particles, small particles, and everything in between, while *simultaneously* allowing enough air to get through so that the person wearing the mask can actually breathe. This is made all the more challenging by the fact that the particles the mask is trying to stop *move differently* depending on how big they are.

This brings us to our first important point about face masks: *a mask is not a sieve*! Sieves – like the mesh strainer you might find in your kitchen, pictured in Fig. 12.5 – work by having holes of some particular size. Anything smaller than those holes can pass through, and anything larger than those holes will get stuck. But we just argued that masks need to stop both large *and* small things, and it wouldn't do much good to just let the small things through! It also is incredibly challenging to make a sieve with holes so small that it will catch tiny things (like smoke particles, or aerosolized viruses) while simultaneously allowing someone to breathe in and out easily.

So if a mask isn't a sieve, what is it? A better metaphor for masks is a spider-web! The design goal of a mask isn't to filter things by catching everything bigger than the holes in the mesh, the design goal is to get particles of all different sizes to *touch one of the fibers and get stuck on it*. The "sticky" part of the problem turns out to be not so hard: for very small objects something called the *van der Waals*[7] force means that at the small size scales we are thinking about, *everything is sticky*! The real trick, then, is to make sure that particles you want to filter out come into contact with the fibers composing your mask.

12.3.3 Random walks and impact capture

So, masks (and particle filters in general) tend to be made up of stacked layers of fibers; a representative zoomed-in view of an N95 mask is shown in Fig. 12.5. The relatively large openings between the fibers means that it is relatively easy to breathe through them, and *large* particles are easy to capture: trying to fly straight through the mask, they are essentially certain to eventually bump into one of the fibers and get trapped.

Perhaps paradoxically, *very small* particles are also easy to capture! If a mask was a sieve that would be completely. wrong: big particles would get stuck, and small particles would go right through. But actually, the fact that the trajectory of small particles is like a random walk means that they spend a long time wandering around the space near the mask and in the mask, and have an extremely high probability of randomly bumping into one of the fibers before they wander through the thickness of the mask. Basically, any simple cloth mask will do a very good job at filtering out these two ranges of particle size

Figure 12.6 shows a representative plot of the *efficiency* of a typical filter at blocking out particles of various sizes. Typical fiber filters are nearly perfect for both very large and very small particles, and it is *medium sized* small particles that present the biggest challenge! The particle size where any particular filter does the worst job is called the "most penetrating particle size" (MPPS), and unfortunately the typical MPPS is around a few tenths of a micron – the perfect size for lots of bad things to get through a mask! This is the range of some of the aerosolized droplets that you breathe out into the air all the time, or expel more forcefully when you cough or sneeze. It's also a droplet size that can hang in the air for a

[7] Caused by small shifts in where electrons are located around the nucleus when different atoms and molecules get close to each other

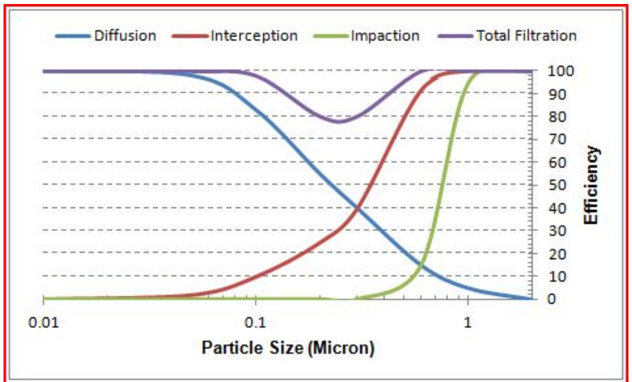

Diagram 1: Total filtration curve with individual mechanism by particle size

Figure 12.6: **Masks do better and worse jobs at filtering out particles of different size** Representative (but not universal!) image showing how efficient a particular HEPA filter is at filtering out particles of different sizes. Notice that the x-axis is on a logarithmic rather than linear scale. Image used without permission from a blog about industrial vacuum filters.

medium length of time, rather than just floating quickly to the ground, and be easily carried by air currents through the labyrinthine mesh of a face mask.

To deal with this pernicious size of particle, high-quality masks electrically charge the fibers that they are composed of. You perhaps have experience with this from rubbing two fabrics against each other in a low-humidity environment, typically in the winter, and having them build up a static charge[8]. The DNA inside a virus has a slight negative charge to it, so if one positively charges the fibers of a mask they will *attract* some of these bad particles[9].

How good a job a mask does at filtering out these "medium small" particles determines the rating of the mask, hence: an N95 mask filters out 95% of these difficult-to-capture particles[10]. Surgical and medical masks are mostly used in the context of water-based droplets (carrying harmful viruses inside the droplets, for instance), but in some industrial applications people are concerned about tiny *oil* droplets (an important here: some industrial oils can remove the electric charge from the fibers, making the whole mask much less good). A mask given a rating of "N" is not resistant to oil, "R" is somewhat resistant to oil, and "P" is oil-proof. With all of that, there are 9 categories of high-quality mask recognized and tested by the National Institute for Occupational Safety and Health: the first letter can be N, R, or P, and the number is tested to be 95, 99, or 100 (masks that do worse than filtering 95% are just not worth testing, I guess).

[8]Hence why taking clothes out of the dryer can be a mildly shocking affair

[9]Even neutrally charged particles will develop an imbalance in where their electrons are in the presence of a static electric field, and be attracted to the fibers!

[10]Under very specific conditions, including a specific airflow rate, a specific relative humidity, a specific density of incoming particles, etc.