



CONTENTS

1	Cognitive Biases	1	3
1.1	Fundamental Attribution Error		3
1.2	Self-Serving Bias		4
1.3	In-group favoritism		4
1.4	The Bandwagon Effect and Groupthink		5
1.5	The Curse of Knowledge		5
1.6	False Consensus		6
1.7	The Spotlight Effect		7
1.8	The Dunning-Kruger Effect		8
1.9	Confirmation Bias		9
1.10	Survivorship bias		11
2	Sampling Data		13
2.1	Planning a Study		13
2.1.1	Vocabulary		13
2.1.2	Why Not Always Use a Census?		14
2.2	Planning and Conducting Surveys		16
2.2.1	Biased Sampling		16
2.3	Simple Random Sampling		18
2.4	Other Random Sampling Methods		21
2.5	Bias in Surveys		25
3	Data tables in SQL		29
3.1	Using SQL from Python		33

4	Advanced SQL and Database Management	35
4.1	Keys	35
4.2	Normalization	36
4.2.1	Functional Dependencies	37
4.2.2	1NF	37
4.2.3	2NF	38
4.2.4	3NF	40
4.3	Entity-Relationship Diagram	41
4.4	Relational Algebra Basics	42
4.5	Relating Two Tables	43
4.6	Advanced Queries	44
4.7	Exercises	44
5	Diodes, Rectifiers, and Photovoltaics	47
5.1	Silicon	47
5.2	Diodes	49
5.2.1	Rectifiers	52
5.2.2	Photovoltaics	53
6	How Cameras Work	55
6.1	The Light That Shines On the Cow	55
6.2	Light Hits the Cow	57
6.3	Pinhole camera	58
6.4	Lenses	59
6.5	Sensors	61
6.6	Aspect Ratio	62
6.6.1	Printing	63
7	How Eyes Work	65
7.1	Eye problems	67
7.1.1	Glaucoma	67
7.1.2	Cataracts	68
7.1.3	Nearsightedness, farsightedness, and astigmatism	68
7.2	Seeing colors	70
7.3	Pigments	72
A	Answers to Exercises	75
	Index	81

Cognitive Biases 1

In this section, we are going to take a look at research findings about *cognitive biases*. These are universal quirks found in the human thought process. Cognitive biases are a bit different from other kinds of biases, such as racial biases. Everyone, regardless of nationality, race, or gender is subject to these cognitive traps. You might be wondering, why do I need to learn about cognitive science in order to be an engineer? The most important tool we have as problem solvers is our own minds. We are going to be looking at ways that our minds can trip us up.

Our brains were shaped over millions of years by evolution. The resulting mind is an amazing and powerful tool, but it is not flawless. The human brain has tendencies (or biases) that nudge us toward bad judgment and poor decisions.

When someone first gave you a hammer they handed it over with a warning: "Do not hit your thumb!" No matter how careful you are with the hammer, at some point you will still hit your thumb. It is the same with cognitive biases. In the course of life, all of us will fall prey to these cognitive biases. Knowing about them is the first step in protecting ourselves.

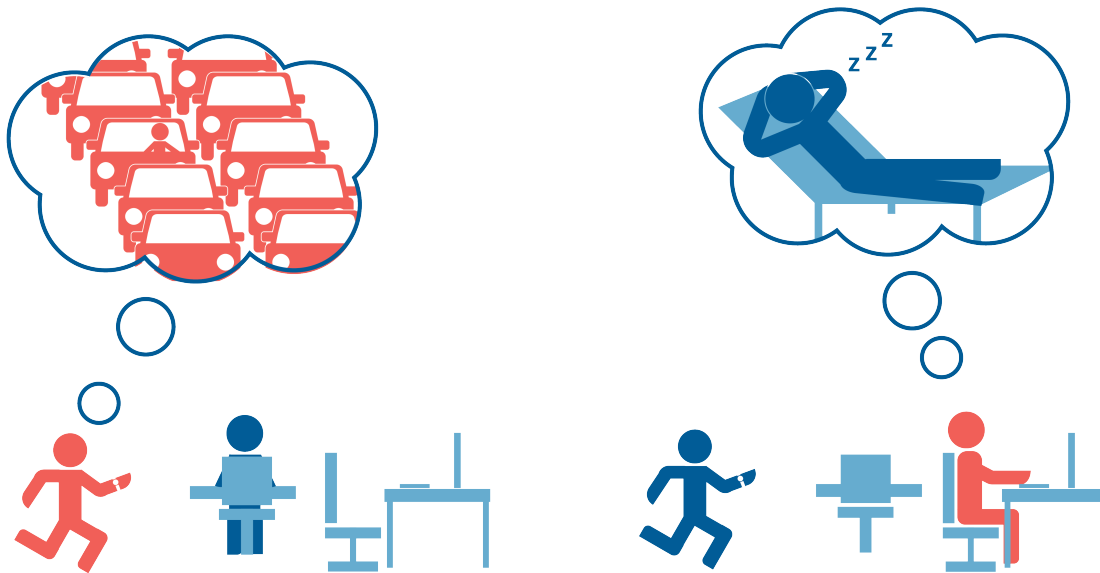
It would be irresponsible to teach you powerful ideas without also teaching you about the cognitive biases that follow them. There are about 50 that you should know about, but let us start with only a few.

1.1 Fundamental Attribution Error

You tend to attribute the mistakes of another person to their character, but attribute your own mistakes to the situation.

If someone asks you "Why were you late for work today?" you are likely to have an excuse, such as "I got stuck in a crazy traffic jam."

However, if you notice your coworker is late for work, you are likely to think "My coworker is lazy."



The solution? Cut people some slack. You probably do not know the whole story, so assume that their character is as strong as yours.

Or maybe you also need to hold yourself to a higher standard? Do you find yourself frequently rationalizing your bad judgment, lateness, or rudeness? This could be an opportunity for you to become a better person whose character is stronger regardless of the situation.

1.2 Self-Serving Bias

Self-serving bias is when you blame the situation for your failures, but attribute your successes to your strengths.

For example, when asked “Why did you lose the match?” you are likely to answer “The referee was not fair.” When you are asked “Why did you win the match?” you are likely to answer “Because I have been training for weeks, and I was very focused.”

This bias tends to make us feel better about ourselves, but it makes it difficult for us to be objective about our strengths and weaknesses.

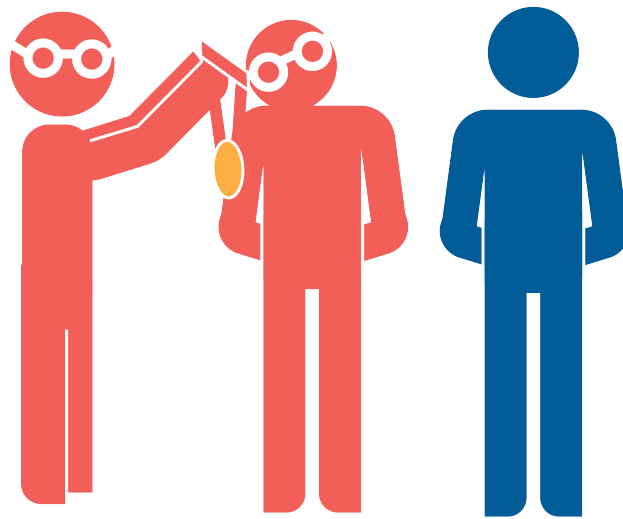
1.3 In-group favoritism

In-group favoritism: We tend to favor people who are in a group with us over people who are not in groups with us.

When asked “Who is the better goalie, Ted or John?”

If Ted is a Star Trek fan like you, you are likely to think he is also a good goalie.

As you might imagine, this unconscious tendency is the source of a lot of subtle discrimination based on race, gender, age, and religion. As we mentioned earlier, racial bias is not a cognitive bias, but one can still feed into the other.



1.4 The Bandwagon Effect and Groupthink

The bandwagon effect is our tendency to believe the same things that the people around us believe. This is how fads spread so quickly: one person buys in, and then the people they know have a strong tendency to buy in as well.

Groupthink is similar: To create harmony with the people around us, we go along with things we disagree with.

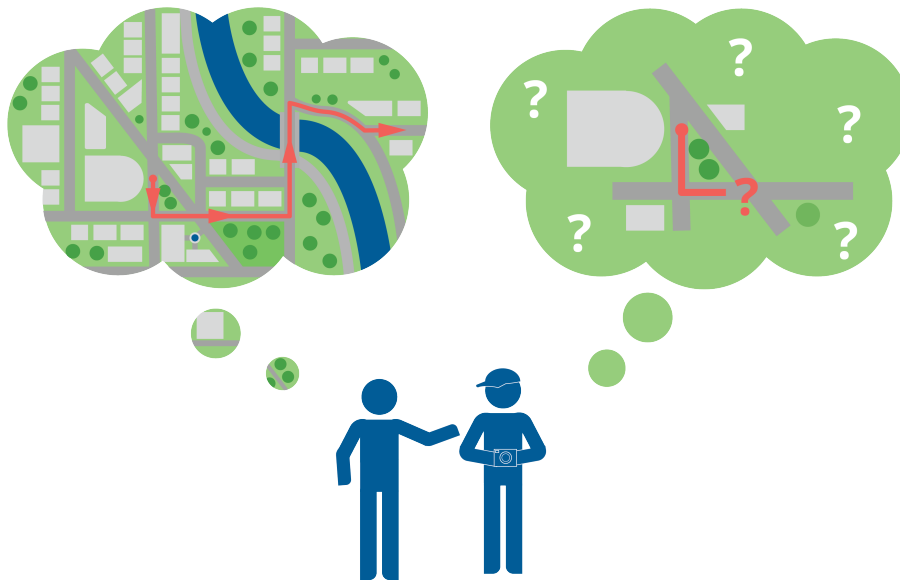
It takes a lot of perspective to recognize when those around us are wrong. And it takes even more courage to openly disagree with them.

1.5 The Curse of Knowledge

Once you know something, you tend to assume everyone else knows it too.

This is why teaching is sometimes difficult; a teacher will assume that everyone in the audience already knows the same things the teacher knows.

For example, imagine a local who has lived in a city for years giving directions to a tourist. The local has an in-depth understanding of the city, and gives overly quick and detailed instructions. The tourist politely smiles and nods, but stopped following after the local began listing unfamiliar street names and landmarks.



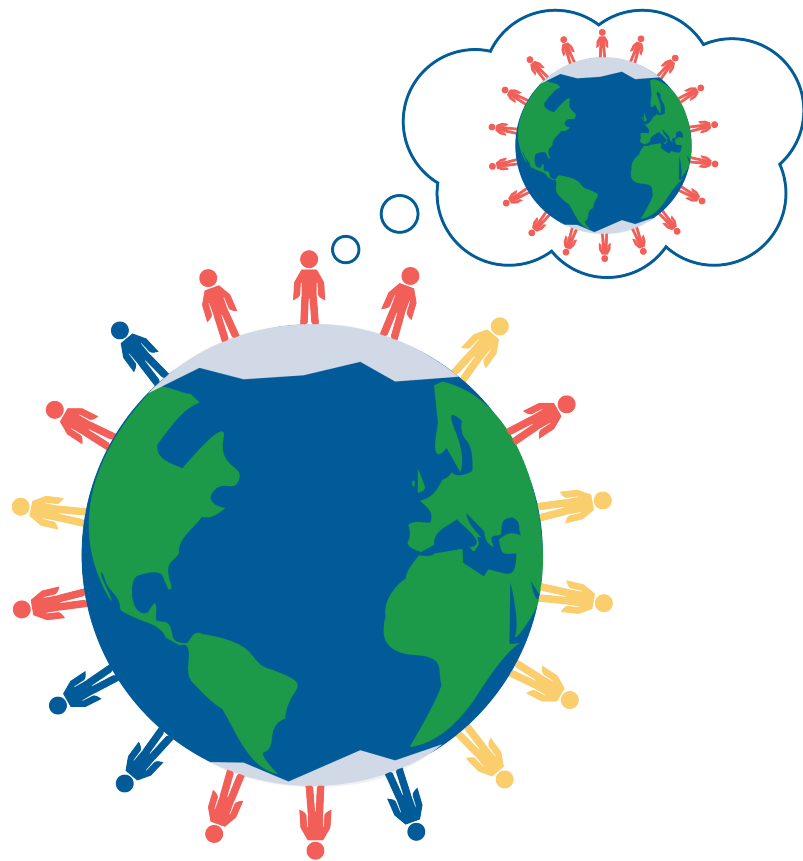
When we learn that a friend does not know something that we know, we are often very surprised. This surprise can sometimes manifest as hurtful behavior.

When we find a gap in our friend's knowledge, we try to remind ourselves that the friend certainly knows many things that we do not. We also try to imagine how it would feel if they teased us for our ignorance.

1.6 False Consensus

We tend to believe that more people agree with us than is actually the case. For example, if you are a member of a particular religion, you tend to overestimate the percentage of people in the world who are members of that religion.

When people vote in elections, they are often surprised when their preferred candidate loses. "Everyone, and I mean EVERYONE, voted for Smith!" they yell. "There must have been a mistake in counting the votes."

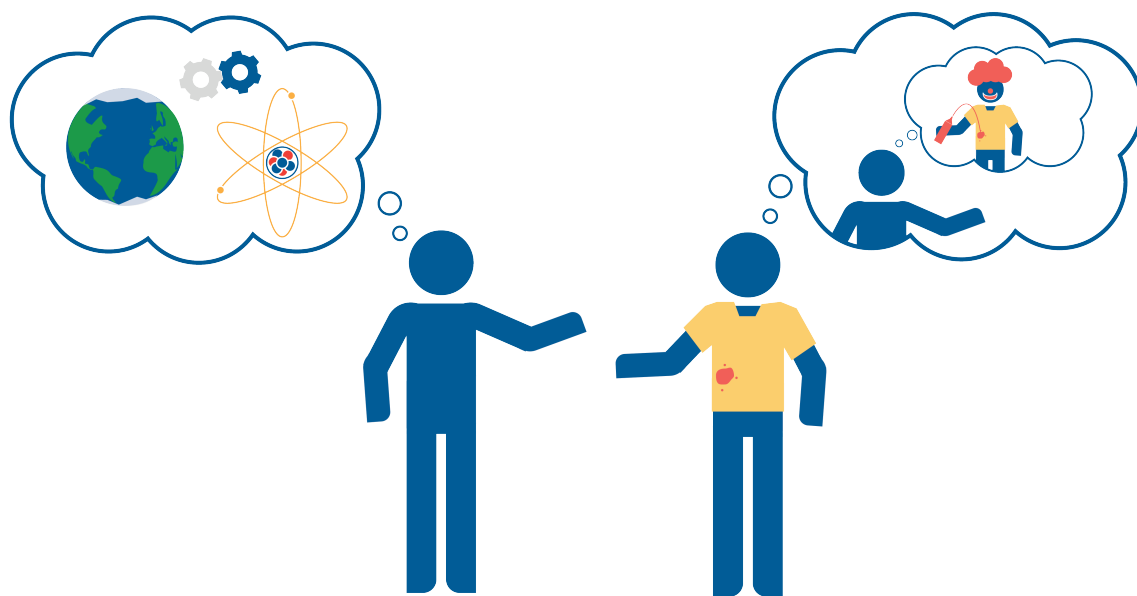


1.7 The Spotlight Effect

You tend to overestimate how much other people are paying attention to your behavior and appearance.

Think of six people that you talked to today. Can you even remember what shoes most of them were wearing? Do you care? Do you think any of them remember which shoes you wore today?

There is an old saying, “You would worry a lot less about how people think of you, if you realized how rarely they do.”



1.8 The Dunning-Kruger Effect

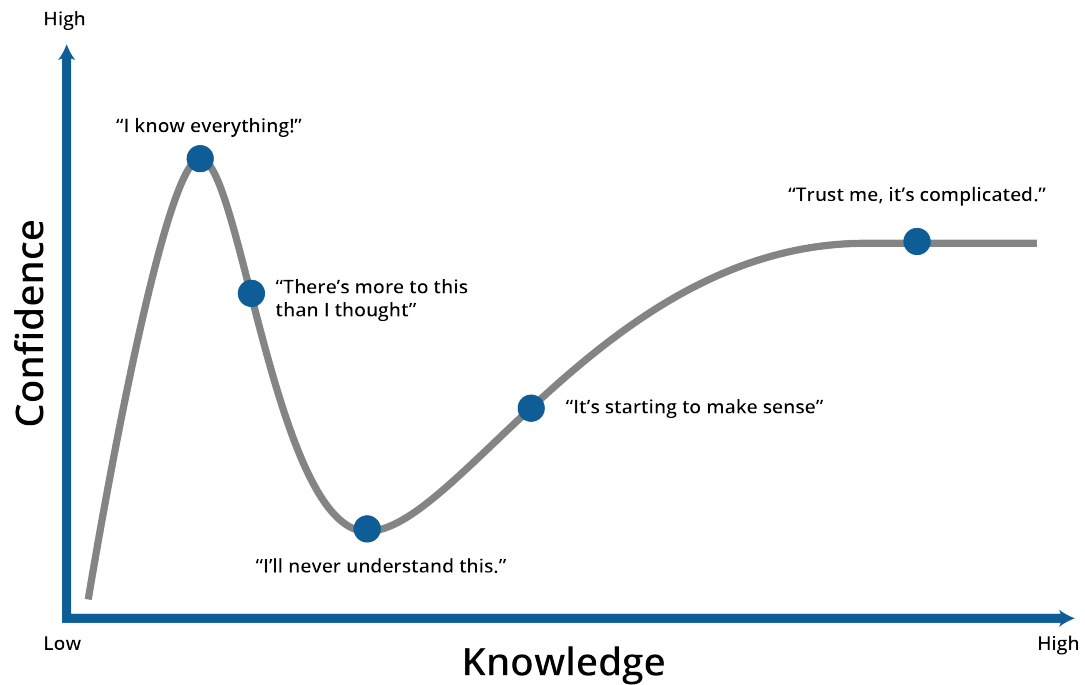
People who know very little about a subject tend to overestimate how much they know. In the original 1999 study by David Dunning and Justin Kruger, the students who scored worst on tests of logic and grammar guessed that they had done better than most of their classmates. The skills you need to do something well are often the same skills you need to notice that you are doing it badly.

When a person does not know all the nuance and context in which a question is asked, the question seems simple. Thus, the person tends to be confident in their answer. As they learn more about the complexity of the space in which the question lives, they often realize the answer is not nearly so obvious.

For example, a lot of people will confidently proclaim “Taxes are too high! We need to lower taxes.” An economist who has studied government budgets, deficits, history, and monetary policy, might say something like “Maybe taxes *are* too high. Or maybe they are too low. Or maybe we are taxing the wrong things. It is a complex question.”

When we are talking with people about a particular topic, we do our best to defer to the person in the conversation who we think has the most knowledge in the area. If we disagree with the person, we try to figure out why our opinions are different.

Similarly, you should assume that any opinion that is voiced in an internet discussion is, at best, wildly over-simplified. If you really care about the subject, read a book by a respected expert. Yes, a whole book — there are few interesting topics that can be legitimately explained in under 100 pages.

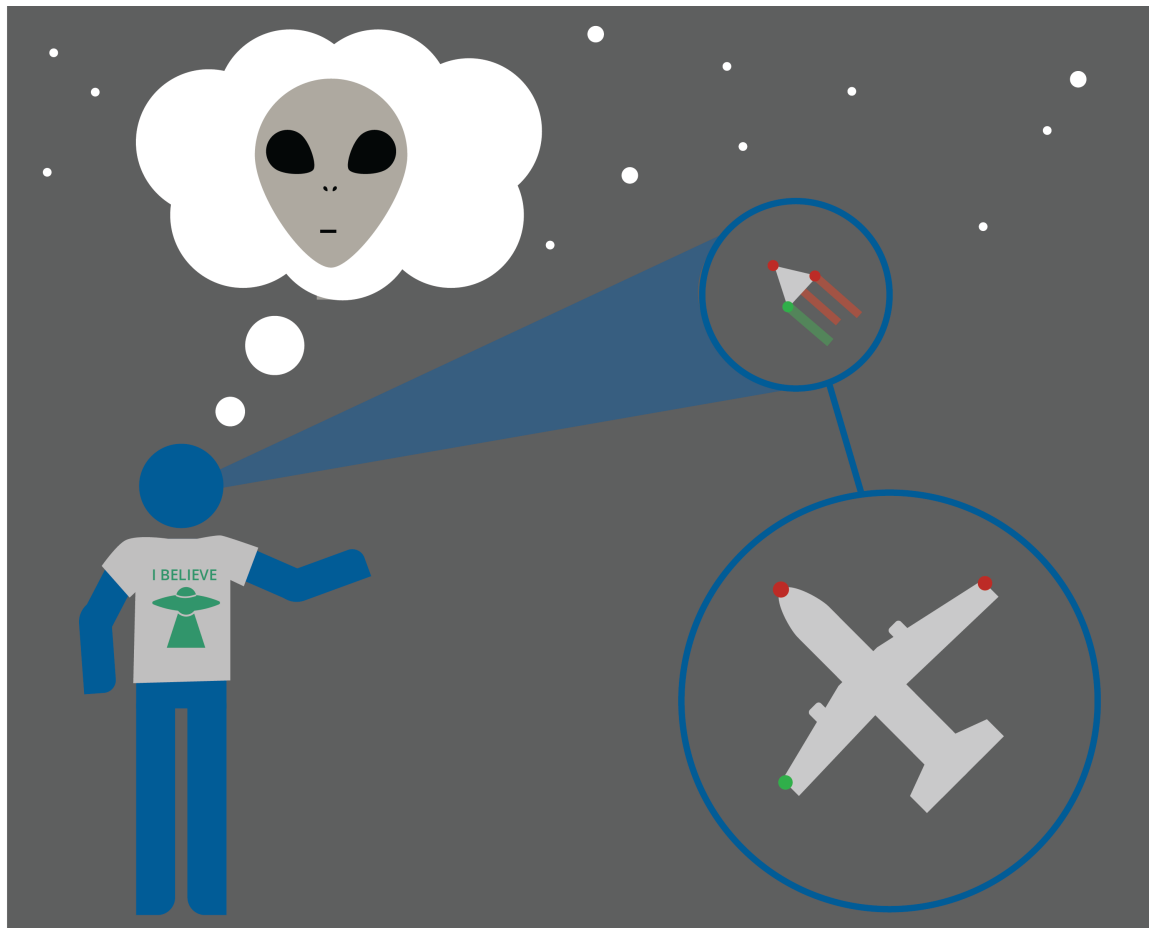


1.9 Confirmation Bias

You tend to find and remember information that supports beliefs you already have. You also tend to avoid and dismiss information that contradicts your beliefs.

If you believe that intelligent creatures have visited from other planets, you will tend to look for data to support your beliefs. You will remember every blurry photo of a strange light in the sky, and forget the hundreds of sightings that turned out to be airplanes, weather balloons, and the planet Venus.

Confirmation bias is one reason why people do not change their beliefs more often.



Confirmation bias wrecks many, many studies. The person doing the study often has a hypothesis that they believe and very much want to prove true. It is very tempting to discard data that does not support the hypothesis. A person may even throw all the data away and experiment again and again until they get the result they want.

When you design an experiment, you must describe it explicitly before you start. You must tell someone: “If the hypothesis I love is incorrect, the results will look like this. If the hypothesis I love is correct, the results will look like that. And if the results look any other way, I have neither proved nor disproved the hypothesis.”

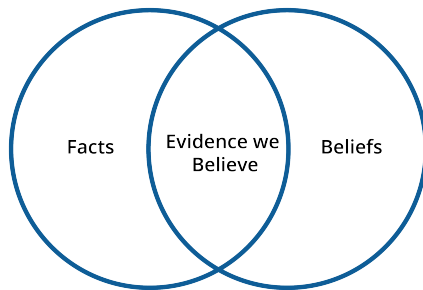
Once the experiment is underway, you must not change the plan and you must not discard any data.

This is scientific integrity. You should demand it from yourself, and you should expect it from others.

Watch a TED Talk and Learn More About Confirmation Bias: What shapes our perceptions (and misperceptions) about science? In an eye-opening talk, meteorologist J. Marshall Shepherd explains how confirmation bias, the Dunning-Kruger effect and cog-

nitive dissonance impact what we think we know – and shares ideas for how we can replace them with something much more powerful: knowledge.

https://www.ted.com/talks/j_marshall_shepherd_3_kinds_of_bias_that_shape_your_worldview



1.10 Survivorship bias

You will pay more attention to those that survived a process than those who failed.

After looking at several old houses, you might say “In the 1880s, they built great houses.” However, you have not seen the houses that were built in the 1880s and did not survive. Which houses tended to survive for a long time? Only the great houses – you are basing your opinion on a very skewed sample.



Worked example. A student says, “Everyone in my class agrees with the new rule, so most students in the school must like it.” Which bias is at work?

Solution. False consensus: the student assumes that their own circle's view is shared more widely than it may be. The class is also a small sample that may not represent the school.

Exercise 1 Name the Bias

A manager thinks the team's plan works because the three people who talk most in meetings all support it. Which bias might affect this conclusion?

Working Space

Answer on Page 75

Exercise 2 Name the Bias, Again

Name the bias in each situation.

Working Space

1. After acing a test, Maya says, "I am great at chemistry." After failing the next one, she says, "That test was unfair."
2. A driver cuts you off, and you think, "What a jerk."
3. A news story says, "These five college dropouts became billionaires, so you do not need college to succeed."
4. You spill ketchup on your shirt and spend the whole party sure that everyone is staring at it.

Answer on Page 75

Sampling Data

2.1 Planning a Study

Every day, decisions get made based on data. A city council votes on a new transit route based on ridership surveys. A school principal adjusts the lunch schedule based on student feedback. A pharmaceutical company decides whether to release a new drug based on clinical trial results. In every case, the quality of the decision depends entirely on the quality of the data behind it.

A study plan states what question you are answering, how you will gather information, and whom you will gather it from.

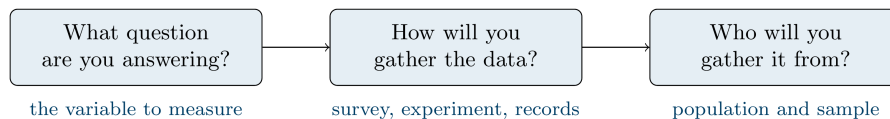


Figure 2.1: A study plan answers three questions: what you want to know, how you will collect data, and whom you will collect it from.

2.1.1 Vocabulary

population The entire group of individuals or items that you want to learn about. The population is defined by the question you are asking.

frame Also called a sampling frame. A list of all the members of the population from which a sample will be drawn. The frame and the population often differ slightly. A school's enrollment database, a city's voter registration list, and a company's customer records are all examples of frames.

sample The part of the population that is actually studied. You collect data from the sample and use it to draw conclusions about the population.

sample survey The process of collecting information from a sample. The information gathered is then used to make inferences about population parameters.

census The process of collecting information from every member of the population rather than just a sample. A census gives complete information, but it comes at a cost.

Worked example. A survey interviews 200 of the 1,000 households in a town. What

fraction of the population is in the sample? Is this a census?

Solution. $\frac{200}{1000} = 20\%$ of the households. It is not a census, because 800 households were not asked.

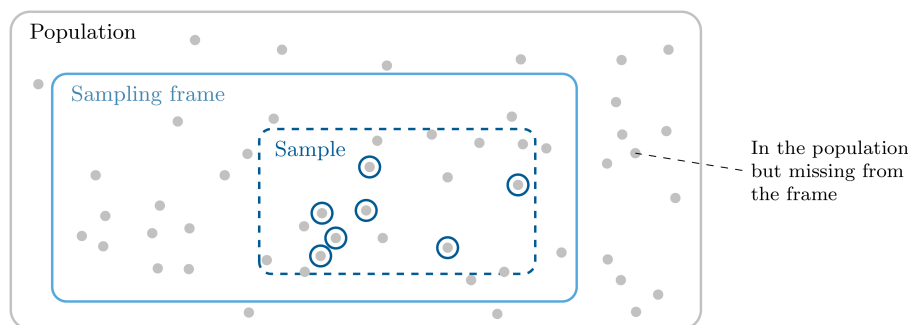


Figure 2.2: The sample is drawn from the sampling frame, which is a working list of the population. Members missing from the frame can never be sampled.

Exercise 3 Identifying the Population and Sample

A city's parks department wants to understand how often residents use public parks. Staff randomly select 200 households from the city's water billing records and conduct phone interviews.

Working Space

1. What is the population?
2. What is the frame?
3. What is the sample?
4. Is this a census? Explain.

Answer on Page 75

2.1.2 Why Not Always Use a Census?

A census measures everyone in the population. For a small group, such as a football team, that is easy.

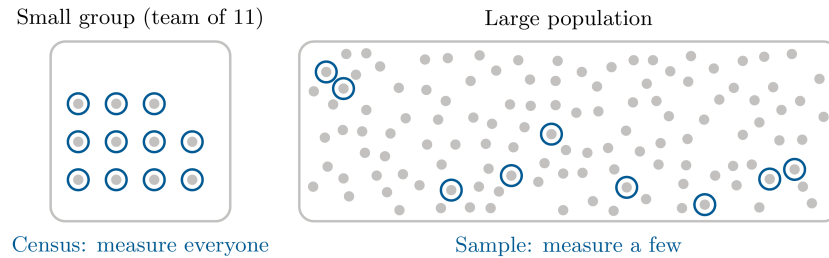


Figure 2.3: A census measures every member of a small group. For a large population, a sample of members is measured instead.

For most populations, a census is impractical:

- The population is too large. A company that wants to know what all households watch cannot contact every one.
- The measurement destroys the item. If testing a battery drains it, checking every battery would leave none to sell.
- The information goes out of date. If a survey of everyone takes months, opinions may have changed by the time it ends.

Exercise 4 Census or Sample?

For each situation, decide whether a census is feasible or whether a sample should be used. Briefly explain your reasoning.

Working Space

Situation	Census or Sample? Explain
A school librarian wants to know which of her 8 student volunteers prefers morning or afternoon shifts.	
A battery manufacturer wants to test the lifespan of batteries by running them until they die.	
A government agency wants to estimate the average commute time of workers across a country of 40 million employed adults.	
A restaurant owner wants to know the dietary restrictions of the 22 guests attending a private event.	

Answer on Page 75

2.2 Planning and Conducting Surveys

The way you choose people determines whether the sample represents the population or leaves parts of it out.

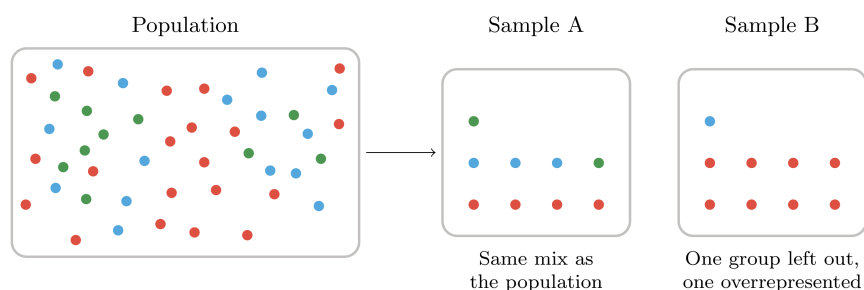


Figure 2.4: Sample A keeps the same mix of groups as the population. Sample B leaves one group out and overrepresents another.

2.2.1 Biased Sampling

A sample is *biased* if it favors some people or answers over others. Bias does not go away just because the sample is bigger.

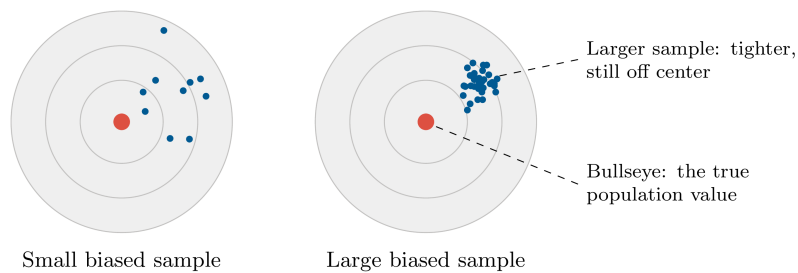


Figure 2.5: A larger biased sample clusters more tightly, but its center stays away from the true population value.

Each of the following methods selects people without chance, so each tends to produce a biased sample.

With *judgmental sampling*, one person picks whoever they believe looks representative. A school inspector who visits only the classrooms she already rates highly is using judgmental sampling.

A *convenience sample* uses whoever is easiest to reach. A journalist who interviews only people at the bus stop outside his office will miss commuters who drive, cycle, or travel at other hours.

When people decide for themselves whether to answer, the result is a *voluntary response sample*. An online poll on a news site attracts mostly visitors with strong opinions, while people who care less rarely respond.

Researchers using *quota sampling* recruit people until a preset count for each group, such as an age band or gender, is filled. If a population is 60% female and 40% male, a quota sample of 100 stops recruiting women at 60 and men at 40. The group mix matches the population, but whoever is approached first fills each quota.

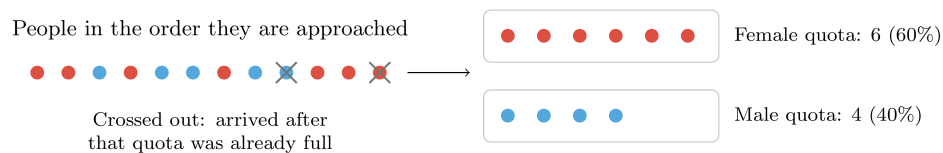


Figure 2.6: Quota sampling fills each group to a fixed number in the order people are approached. Selection within each group is not random.

For populations that are hard to reach, *snowball sampling* starts with a few qualifying participants and asks each of them to refer others who also qualify. Researchers use it to study hidden groups, such as people experiencing homelessness or people with a rare disease, but the sample follows social networks rather than chance. *Snowball sampling* is

beyond the scope of the AP Statistics exam.

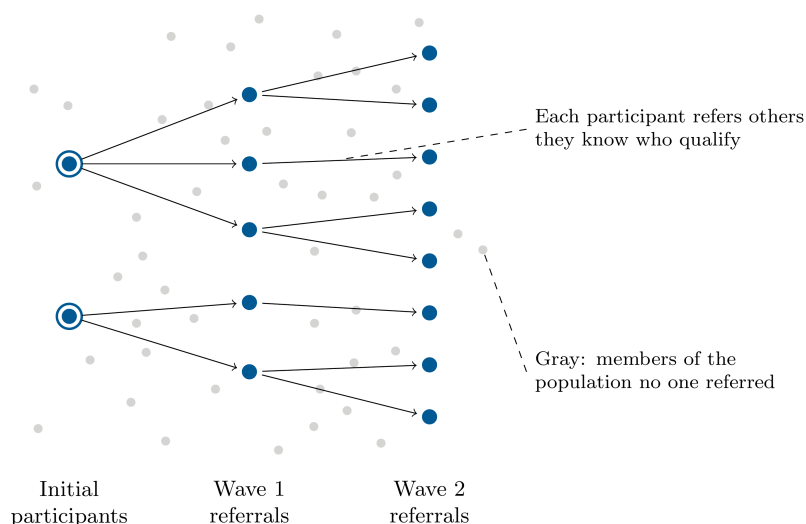


Figure 2.7: Snowball sampling starts with a few initial participants. Each wave of referrals recruits the next, so the sample grows through personal connections.

All five methods leave selection to judgment or convenience. Random selection gives every member a fair chance of being chosen and usually produces more trustworthy results.

2.3 Simple Random Sampling

A *simple random sample* (SRS) is a sample chosen so every group of the same size has the same chance of being picked. An SRS can be drawn in two ways:

- In *sampling with replacement*, each chosen member returns to the pool and could be picked again.
- Each member can be chosen only once in *sampling without replacement*.

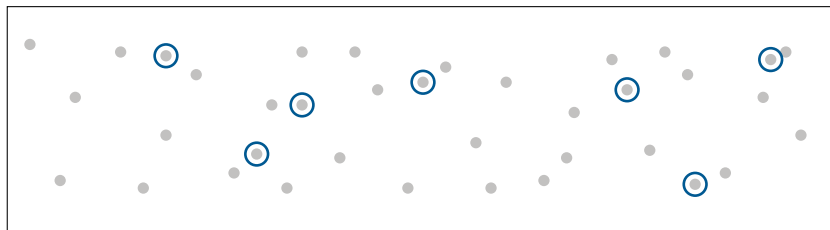


Figure 2.8: In simple random sampling, every individual in the population has an equal chance of being selected, and there is no grouping or pattern to who is chosen.

Chance, rather than convenience or judgment, must decide who is selected. Three ways to do this:

- Assign every member of the population a number. Use a random number table, entering at any random point and reading numbers sequentially. Select members whose numbers appear, skipping values outside your range and any repeats, until you have the sample size you need.
- Use a calculator. On a TI-84, $\text{MATH} \rightarrow \text{PRB} \rightarrow 5:\text{randInt}($ generates random integers. The command $\text{randInt}(1, 500, 30)$ produces 30 random integers between 1 and 500. Ignore repeats.
- Use a physical randomization device. Write each member's identifier on an identical slip of paper, mix thoroughly in a container, and draw without looking until the sample is complete.

Exercise 5 **Selecting an SRS**

A regional transit authority wants to survey 25 of its 300 bus drivers about workplace safety conditions.

Working Space

1. Describe a complete procedure for selecting an SRS of 25 drivers using a random number table.
2. A supervisor suggests selecting the 25 drivers who are easiest to reach during the next shift change. What type of sample is this, and what bias might it introduce?
3. A colleague suggests selecting 5 drivers at random from each of the authority's 5 depots. Is this an SRS? Explain.

Answer on Page 76

Computing in Python

In this exercise you will load real world population data published by the World Bank,

treat all 265 countries and territories as your population, draw a simple random sample, and compare your sample mean to the true population mean.

1. Import the necessary libraries.

```
1 import pandas as pd
```

2. Load the dataset directly from the web and filter to the year 2020.

```
1 url =  
  ↳ "https://raw.githubusercontent.com/datasets/population/main/data/population.csv"  
2 df = pd.read_csv(url)  
3  
4 pop2020 = df[df["Year"] == 2020][["Country Name",  
  ↳ "Value"]].reset_index(drop=True)  
5 print(pop2020.head())  
6 print(f"Total entries: {len(pop2020)}")
```

3. Calculate the true population mean. This is the parameter you are trying to estimate.

```
1 true_mean = pop2020["Value"].mean()  
2 print(f"True population mean: {true_mean:,.0f}")
```

4. Draw an SRS of 20 countries and compute the sample mean.

```
1 sample = pop2020.sample(n=20)  
2 sample_mean = sample["Value"].mean()  
3 print(f"Sample mean (n=20): {sample_mean:,.0f}")
```

Exercise 6 Sampling Real Data from the Web

Working Space

1. Run step 4 five times and record the sample mean each time. How close is each to the true mean? Are they all the same?
2. Repeat question 1 with a sample size of 80. How do the results compare? What does this tell you about the relationship between sample size and

sampling error?

Answer on Page 76

2.4 Other Random Sampling Methods

An SRS can be expensive or imprecise for some populations. Three other random methods are common, and all three appear regularly on the AP exam.

In *systematic sampling*, you list the population, choose a random start, and then pick every k th member. To sample 40 students from 800, start at a random number between 1 and 20, then take every 20th student. The method needs a complete list of the population.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Random start: 7
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	Then every 20th
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	7, 27, 47, ...

Figure 2.9: Systematic sampling with a random start of 7 and $k = 20$ selects members 7, 27, 47, and so on.

For *stratified random sampling*, you split the population into groups called strata whose members share a characteristic, such as students in the same grade or patients in the same ward. You then take a random sample from each stratum. Every group is guaranteed a place in the sample, which matters whenever groups need to be compared.

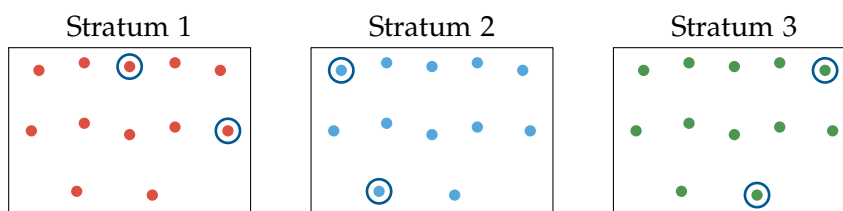


Figure 2.10: Each stratum is a single color, since members within a stratum share a characteristic. A small random sample (circled) is drawn from each stratum.

With *cluster sampling*, you split the population into groups called clusters, each with roughly the same mix of members as the whole population. You then randomly select some clusters and study everyone inside them. This suits populations spread over a wide

area, where reaching everyone is expensive. A company surveying customers in 12 cities might randomly select 8 of them and survey every customer in each.

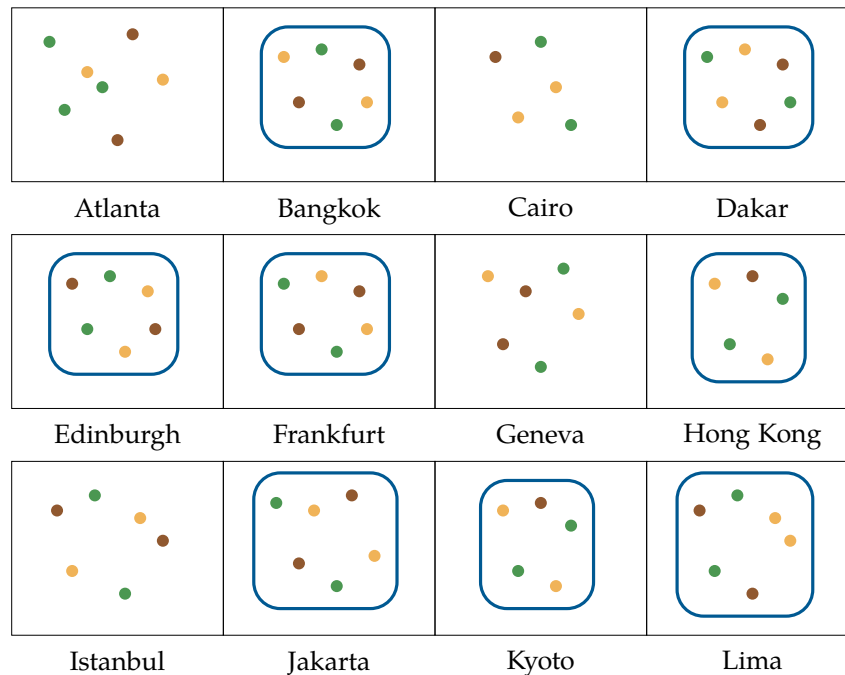


Figure 2.11: Each dot represents a customer and each color represents a different customer segment. Atlanta, Cairo, Geneva, and Istanbul were not selected. The other 8 cities were randomly selected as clusters, and every customer within each selected city becomes part of the sample.

Strata and clusters are both groups, but they play opposite roles:

It is important to understand the difference between strata and clusters. *Strata* are groups that are different from each other but similar within, and you sample from all of them. *Clusters* are groups that are similar to the whole population, and you sample only some of them but take everyone inside those.

Exercise 7 Identifying Sampling Methods

Identify the sampling method used in each scenario: SRS, systematic, stratified, or cluster.

Working Space

Scenario	Method
A supermarket chain selects 10 of its 80 stores at random and surveys every employee in those stores.	
A university randomly selects 50 students from each of its four faculties to participate in a wellbeing survey.	
A water utility selects a random starting meter on its numbered list and then tests every 25th meter for contamination.	
A film festival puts all 900 registered attendee names in a database and uses software to randomly select 90.	

Answer on Page 76

Computing in Python

Apply three sampling methods to the same dataset and compare what each one produces. The dataset contains 30 students across three grade levels with their test scores.

Create the dataset.

```

1  import pandas as pd
2  import numpy as np
3
4  np.random.seed(42)
5
6  data = {
7      "student_id": range(1, 31),
8      "grade": ["9th"]*10 + ["10th"]*10 + ["11th"]*10,
9      "score": (
10         list(np.random.randint(60, 80, 10)) +
11         list(np.random.randint(70, 90, 10)) +
12         list(np.random.randint(80, 100, 10))
13     )
14  }
15
16  df = pd.DataFrame(data)
17  print(df)

```

Compute the true population mean.

```

1  true_mean = df["score"].mean()

```

```
2 print(f"True mean score: {true_mean:.1f}")
```

1. Simple random sample. Draw an SRS of 9 students.

```
1 srs = df.sample(n=9)
2 print(f"SRS mean: {srs['score'].mean():.1f}")
3 print(srs[["student_id", "grade", "score"]])
```

2. Stratified random sample. Draw 3 students from each grade level.

```
1 stratified = df.groupby("grade").sample(n=3)
2 print(f"Stratified mean: {stratified['score'].mean():.1f}")
3 print(stratified[["student_id", "grade", "score"]])
```

3. Cluster sample. Randomly select one entire grade level as the cluster.

```
1 chosen_grade = pd.Series(df["grade"].unique()).sample(n=1).values[0]
2 cluster = df[df["grade"] == chosen_grade]
3 print(f"Cluster selected: {chosen_grade}")
4 print(f"Cluster mean: {cluster['score'].mean():.1f}")
```

Exercise 8 Comparing Sampling Methods

Run each of the three samples above three times.

Working Space

1. Record each sample mean in the table.

Method	Run 1	Run 2	Run 3	Most stable?
SRS				
Stratified				
Cluster				
True mean	79.7			

2. Does the SRS always represent all three grades?

3. Which method gives the most stable estimate? Why does the stratified

sample tend to stay closer to the true mean than the SRS?

4. Why does the cluster mean vary the most?

Answer on Page 77

2.5 Bias in Surveys

A random sample can still produce misleading results, because bias can enter after selection through how questions are asked and who responds.

The chance difference between a sample result and the population value is called *sampling error*. Different samples give different answers, and this variation shrinks as the sample grows. Bias behaves differently: it pushes results in one direction, and a larger sample does not reduce it.

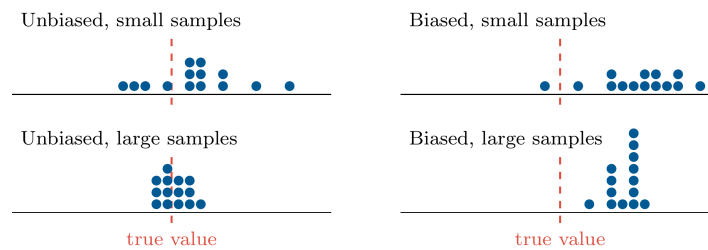


Figure 2.12: Each dot is the estimate from one sample. Sampling error shrinks as the sample size grows. Bias moves every estimate away from the true value, whatever the sample size.

Four types of bias are common in surveys.

response bias Respondents give answers that differ from their true beliefs or behaviors, often because the topic is sensitive or an interviewer creates social pressure. Food-preparation workers asked about hand-washing by a hygiene inspector will likely overstate how often they wash, because they know the answer the inspector wants.

nonresponse bias People selected for the sample cannot be reached or decline to take part, and they differ systematically from those who respond. Engaged parents return a mailed school-satisfaction survey at higher rates than disengaged parents, so the

results overstate satisfaction.

undercoverage bias Part of the population is missing from the sampling frame. A survey about internet access run through online banner ads cannot reach people without internet access, the group most relevant to the question.

wording effect bias The phrasing of a question shifts the answer. Compare these two versions:

- “Do you support stricter emissions standards for vehicles?”
- “Do you support stricter emissions standards for vehicles, even if it means higher car prices and fewer model choices?”

Attaching costs to the second version makes “no” more likely, even among people who support the policy.

Exercise 9 Diagnosing Bias

Identify the most likely type of bias in each scenario and explain why it applies.

Working Space

1. A streaming platform sends a satisfaction survey to all subscribers but receives responses from only 8% of them. The platform reports that 91% of subscribers are satisfied.
2. A survey on recycling habits is conducted by going door-to-door in affluent neighborhoods only, because those streets are easier and safer to walk.
3. An interviewer asks employees: “Our new open-plan office is designed to encourage collaboration. How much has it improved your productivity?”
4. A survey on alcohol consumption is conducted at a community health fair by a team wearing hospital uniforms. Respondents significantly underreport their weekly intake.

Answer on Page 78

Exercise 10 **Sample Fraction**

A survey of 25 drivers is taken from a list of 300 drivers. What percent of the drivers are in the sample?

Working Space

Answer on Page 78

Data tables in SQL

Every time you log in to a website, check your grades online, or buy something with a card, a program somewhere asks a database a question, probably in a language called SQL. It has been the standard way to talk to data for fifty years, and it is still one of the most useful skills a programmer can have.

Most organizations keep their data as tables inside a *relational database management system*, rather than in CSV files or spreadsheets. Developers talk to those systems using a language called SQL (“Structured Query Language”).

Some relational database managers are pricey products you may have heard of before, such as Oracle or Microsoft SQL Server. Some are free, such as PostgreSQL or MySQL. These are server software that client programs talk to over a company’s network.

There is also a small, free database, called `sqlite`, that keeps all its tables in a single file. It runs inside nearly every phone and web browser on the planet. We can use SQL to create, edit, and browse those tables. `sqlite` is free, fast, and very easy to install. We will use `sqlite` instead of a networked database management system.

If you look in your digital resources, you will find a file called `bikes.db`. We created this file using `sqlite`, and now you will use `sqlite` to access it.

In the terminal, get to the directory where `bikes.db` lives. To open the `sqlite` tool on that file:

```
1 > sqlite3 bikes.db
```

(If your system complains that there is no `sqlite3` tool, you need to install `sqlite`. See this website: <https://sqlite.org/>)

Please follow along: type each command shown here into the terminal and see what happens.

We mostly run SQL commands in this tool, but there are a few non-SQL commands that all start with a period. To see the tables and their columns, you can run `.schema`:

```
1 sqlite> .schema
2 CREATE TABLE bike (bike_id int PRIMARY KEY, brand text, size int,
```

```
3      purchase_price real, purchase_date date, status text);
```

That is the SQL command that we used to create the bike table. You can see all the columns and their types.

You want to see all the rows of data in that table?

```
1  sqlite> select * from bike;
2  4997391|GT|57|269.61|2009-05-03|rented
3  5429447|Cannondale|50|215.91|2002-02-17|broken
4  5019171|Trek|58|251.17|1985-07-11|rented
5  3000288|Cannondale|57|211.08|1993-01-05|broken
6  880965|GT|52|281.75|1995-08-02|available
7  ...
```

You will see 1000 rows of data!

SQL keywords and table names are not case-sensitive, so you can also write it like this:

```
1  sqlite> SELECT * FROM BIKE;
```

Often, you will see SQL with just the SQL keywords in all caps:

```
1  sqlite> SELECT * FROM bike;
```

(The values inside quotes *are* case-sensitive: 'GT' and 'gt' are different brands to SQL.) The semicolon marks the end of a command; it tells sqlite that you are done typing and it should run.

SQL lets you choose which columns you would like to see. The asterisk used above signifies all columns, and the bike_id, brand only gets the bike's id and brand from the table:

```
1  sqlite> SELECT bike_id, brand FROM bike;
2  4997391|GT
3  5429447|Cannondale
4  5019171|Trek
5  3000288|Cannondale
6  ...
```

Using WHERE, SQL lets you use conditions to decide which rows you would like to see, and can be combined with the common operators AND, OR, and NOT:

```

1  sqlite> SELECT * FROM bike WHERE purchase_date > '2009-01-01' AND brand = 'GT';
2  4997391|GT|57|269.61|2009-05-03|rented
3  326774|GT|56|165.0|2009-06-27|available
4  264933|GT|52|302.43|2009-07-09|available
5  5931243|GT|55|173.56|2009-11-26|rented
6  4819848|GT|51|221.71|2009-12-11|rented
7  9347713|GT|52|232.32|2009-06-13|available
8  3019205|GT|58|262.94|2009-08-22|available

```

Using DISTINCT, SQL lets you get just one copy of each value:

```

1  sqlite> SELECT DISTINCT status FROM bike;
2  rented
3  broken
4  available
5
6  Busted
7  Flat tire
8  good
9  out
10 Rented

```

Look at that list. The blank line is a bike with no status at all (SQL calls a missing value NULL). And someone has been typing whatever they felt like: Busted, Flat tire, Rented with a capital R. Real data is messy.

You can fix it by editing rows. To change every Busted to broken, use an UPDATE statement with a SET:

```

1  sqlite> UPDATE bike SET status='broken' WHERE status='Busted';
2  sqlite> SELECT DISTINCT status FROM bike;
3  rented
4  broken
5  available
6  Flat tire
7  good
8  out
9  Rented

```

You can insert new rows:

```

1  sqlite> INSERT INTO bike (bike_id, brand, size, purchase_price, purchase_date,
2  ↵ status)
3  ...> VALUES (1, 'GT', 53, 123.45, '2020-11-13', 'available');
4  sqlite> SELECT * FROM bike WHERE bike_id = 1;

```

```
4 | 1|GT|53|123.45|2020-11-13|available
```

The `bike_id` must be *unique*: it is the primary key, and `sqlite` will refuse a second bike with the same ID.

You can delete rows:

```
1 | sqlite> DELETE FROM bike WHERE bike_id = 1;  
2 | sqlite> SELECT * FROM bike WHERE bike_id = 1;
```

Be careful: `UPDATE` and `DELETE` without a `WHERE` clause change or delete *every* row in the table, and there is no undo.

To get out of `sqlite`, type `.exit`.

Exercise 11 Clean up the statuses

Write `UPDATE` statements that change Flat tire to broken and Rented to rented. Then run `SELECT DISTINCT status FROM bike;` to check your work. Which strange values are left?

Working Space

Answer on Page 78

Exercise 12 SQL Query

Execute an SQL query that returns the `bike_id` (no other columns) of every Trek bike that cost more than \$330.

Working Space

Answer on Page 78

3.1 Using SQL from Python

The people behind sqlite created a library for Python that lets you execute SQL and fetch the results from inside a python program.

Let us create a simple program that fetches and displays the bike ID and purchase date of every Trek bike that cost more than \$330.

Create a file called `report.py`:

```
1 import sqlite3 as db
2
3 con = db.connect('bikes.db')
4 cur = con.cursor()
5
6 cur.execute("SELECT bike_id, purchase_date FROM bike WHERE purchase_price > 330 AND
7 ↪ brand='Trek'")
8 rows = cur.fetchall()
9
10 for row in rows:
11     print(f"Bike {row[0]}, purchased {row[1]}")
12
13 con.close()
```

When you execute it, you should see:

```
1 > python3 report.py
2 Bike 4128046, purchased 2007-08-06
3 Bike 7117808, purchased 1995-03-12
4 Bike 7176903, purchased 1986-07-03
5 Bike 827899, purchased 2009-03-14
6 Bike 363983, purchased 1970-08-16
```


Advanced SQL and Database Management

If you recall from our basic SQL chapters, we have the ability to create sets of rows and columns that represent data. We can represent a set of 1000 bikes, each with a respective price, status, brand, and id, in the form of rows (you may hear the term *tuples*) and columns (attributes). In this chapter, we will build out a larger database, with more than one table, but, here is the twist: we will find a way to link them together, so that data can be searched across tables.

4.1 Keys

In the experimentation from our first chapter, you may have seen the term *primary key*. An attribute (or set of attributes) that can uniquely identify rows within a table are called a primary key. That may have just sounded like a set of buzzwords, so let us take a step back:

Take for example the following table:

StudentID	Email	SSN	FirstName	LastName	EnrollmentDate
1001	alice@univ.edu	123-45-6789	Alice	Smith	2024-01-15
1002	bob@univ.edu	987-65-4321	Bob	Johnson	2024-01-15
1003	carol@univ.edu	555-55-5555	Carol	Williams	2024-02-01

Table 4.1: STUDENT table demonstrating key types.

We can identify columns or combinations of columns that could be used to uniquely identify each row. We can identify pretty much infinite *super keys*, which are any set of attributes that uniquely identifies a row. What sets these apart is that they may include extraneous attributes that are not required for uniqueness. For example, in the STUDENT table, both {StudentID} and {StudentID, Email} are super keys, but the latter includes an unnecessary attribute for uniqueness. Keys that contain more than one attribute are called *composite keys*, and are represented in sequence with parentheses such as (StudentID, Email).

A *candidate key* is a minimal super key, meaning it has no extraneous attributes. In our STUDENT table, {StudentID} and {SSN} are candidate keys because they uniquely iden-

tify each row and do not contain unnecessary attributes.

From the candidate keys, a *primary key* is selected. By convention, the primary key is underlined. The primary key is the main candidate key chosen to uniquely identify each row in the table. In our STUDENT table, {StudentID} could be chosen as the primary key. Then, {SSN} would be considered an *alternate key*, which is a candidate key not selected as the primary key. The reason we select {StudentID} over {SSN} is that {SSN} has real-world representation, which we call a *natural key*. We do not want to conflate real world attributes with database management tools such as primary keys.

A real world example: What if we were using an email as the primary key? Then, if the email schema changes due to the business being bought out or renamed (this happened to one of our authors!), we are stuck! A change like <firstinitial><lastname>@company.com to <firstname>.<lastname>@company2.com would break the database.

This is why we introduce *surrogate keys*. Surrogate keys are keys with no real world meaning, often stable identifiers that do not change. This includes auto-incrementing integers, Universally Unique Identifiers (*UUIDs*), or Hex Codes that do not have the opportunity to change. In short, we prefer surrogate keys for primary keys to ensure stability and avoid issues with real-world attribute changes.

In a few sections we will explore how foreign keys allow us to link tables together using these primary keys.

4.2 Normalization

With all of these key types, our tables can get pretty large. When this happens, we will want to normalize our database to ensure efficiency and reduce **redundancy**.

We have a process called *normalization*, which involves organizing, reordering, and restructuring the columns and tables of a database to minimize redundancy and dependency. This can help

- improve performance by reducing the amount of redundant data stored, as databases reach high volumes of data.
- prevent anomalies, which are inconsistencies or errors that can occur when redundant data is updated, inserted, or deleted.
- maintain data integrity

Ultimately, there is one primary goal of normalization:

The table can only express one version of truth.

4.2.1 Functional Dependencies

Normalization rests on a single formal idea: the *functional dependency*. We say an attribute (or set of attributes) Y is functionally dependent on attribute X , written $X \rightarrow Y$, if knowing the value of X tells us exactly one corresponding value of Y . In other words, X *determines* Y .

For example, in our STUDENT table from earlier, $\{\text{StudentID}\} \rightarrow \{\text{Email}\}$, since each StudentID corresponds to exactly one Email. This is exactly why $\{\text{StudentID}\}$ can serve as a key: a candidate key is a minimal set of attributes that functionally determines every other attribute in the table.

Two particular kinds of functional dependency will help us reason through the normal forms:

- A *partial dependency* occurs when a non-key attribute depends on only part of a composite primary key, rather than the whole key.
- A *transitive dependency* occurs when a non-key attribute depends on another non-key attribute, rather than depending directly on the primary key.

Removing partial dependencies brings a table to 2NF, and removing transitive dependencies brings a table to 3NF.

Let us go through an example. In this book, we will only focus on the first three normal forms: 1NF, 2NF, and 3NF. There are videos in your digital resources that go deeper into the higher normal forms.

Below is table representing an intentionally poorly designed database representing lab equipment at a university laboratory.

Table 4.2: Equipment Checkout Log

StudentID	EquipmentID	CheckoutDate	StudentName	StudentEmail	StudentYear	EquipmentName	EquipmentLocation	LocationManager
1001	E42, E55	2026-09-20	Alex	alex@univ.edu	Junior	3D Printer, Oscilloscope	Lab B, Lab A	Maria, David
1002	E42	2026-09-21	Jordan	jordan@univ.edu	Senior	3D Printer	Lab B	Maria

1

4.2.2 1NF

A database is in first normal form (1NF) if it meets the following criteria:

¹Notice how small the text had to get to render the entire table within the page width, this is also a clear indication that something is off!

- Each column contains atomic values (no repeating groups or arrays).
- Each row is unique (no duplicate rows).

Notice how the table violates the first normal form (1NF) because it contains repeating groups and multiple values in a single column. In 1NF, each column should contain atomic values, and each row should be unique. Invalid attribute values include:

- Multiple EquipmentIDs in a single cell ({E42, E55})
- Multiple EquipmentNames in a single cell ({3D Printer, Oscilloscope})
- Multiple EquipmentLocations in a single cell ({Lab B, Lab A})
- Multiple LocationManagers in a single cell ({Maria, David})

To remedy this, we make a second row for student Alex: Now our table conforms to the

Table 4.3: Equipment Checkout Log [1NF]

StudentID	EquipmentID	CheckoutDate	StudentName	StudentEmail	StudentYear	EquipmentName	EquipmentLocation	LocationManager
1001	E42	2026-09-20	Alex	alex@univ.edu	Junior	3D Printer	Lab B	Maria
1002	E42	2026-09-21	Jordan	jordan@univ.edu	Senior	3D Printer	Lab B	Maria
1001	E55	2026-09-22	Alex	alex@univ.edu	Junior	Oscilloscope	Lab A	David

first normal form (1NF), with each column containing atomic values and each row being unique.

4.2.3 2NF

1NF was quite easy, now we move on to the second normal form (2NF), which requires the following:

- The table is in 1NF.
- No partial dependencies (no non-key attribute should depend on a part of a composite primary key).
- All non-key attributes are fully functionally dependent on the primary key.

For demonstration purposes, lets add a fourth entry into our table:

Table 4.4: Equipment Checkout Log [Pre-2NF]

StudentID	EquipmentID	CheckoutDate	StudentName	StudentEmail	StudentYear	EquipmentName	EquipmentLocation	LocationManager
1001	E42	2026-09-20	Alex	alex@univ.edu	Junior	3D Printer	Lab B	Maria
1002	E42	2026-09-21	Jordan	jordan@univ.edu	Senior	3D Printer	Lab B	Maria
1001	E55	2026-09-22	Alex	alex@univ.edu	Junior	Oscilloscope	Lab A	David
1001	E74	2026-09-23	Alex	alex@univ.edu	Junior	CNC Machine	Lab C	Larry

We can establish that the primary key of this table is the combination of {StudentID, EquipmentID, CheckoutDate}, as each student can check out multiple pieces of equipment, and each piece of equipment can be checked out by multiple students.

However, we have lots of repeated data in here:

- Student information (name, email, year) is repeated for each piece of equipment they check out.
- Equipment information (name, location, location manager) is repeated for each student who checks it out.

This redundancy indicates that our table is not in 2NF, as non-key attributes are partially dependent on the composite primary key. To achieve 2NF, we need to separate the student and equipment information into their own tables, leaving only the checkout details in the original table.

The new schemas will look like this:

Student (StudentID, StudentName, StudentEmail, StudentYear)
 Equipment (EquipmentID, EquipmentName, EquipmentLocation, LocationManager)
 Checkout (StudentID, EquipmentID, CheckoutDate)

<u>StudentID</u>	StudentName	StudentEmail	StudentYear
1001	Alex	alex@univ.edu	Junior
1002	Jordan	jordan@univ.edu	Senior

Table 4.5: Student Table [2NF]

<u>EquipmentID</u>	EquipmentName	EquipmentLocation	LocationManager
E42	3D Printer	Lab B	Maria
E55	Oscilloscope	Lab A	David
E74	CNC Machine	Lab C	Larry

Table 4.6: Equipment Table [2NF]

<u>StudentID</u>	<u>EquipmentID</u>	<u>CheckoutDate</u>
1001	E42	2026-09-20
1002	E42	2026-09-21
1001	E55	2026-09-22
1001	E74	2026-09-23

Table 4.7: Checkout Table [2NF]

Notice how each table now has a single, clear primary key: {StudentID} for Student, {EquipmentID} for Equipment, and {StudentID, EquipmentID} for Checkout. Every non-key attribute is now fully dependent on its table's key, with no partial dependencies remaining. 2NF often splits up a table into multiple tables to remove partial dependencies and serves as a stepping stone to 3NF, which further eliminates transitive dependencies.

4.2.4 3NF

To transform a table from 2NF to 3NF, we need to meet the following conditions:

- The table must already be in 2NF (and, consequently, in 1NF).
- There should be no *transitive dependencies*, meaning non-key attributes should not depend on other non-key attributes.

Looking at our table, we see a transitive dependency:

EquipmentID $\rightarrow$ EquipmentLocation $\rightarrow$ LocationManager

These are all stored in the same table, which creates a transitive dependency that violates 3NF.

We now must make a new table to remove this transitive dependency, ensuring that each non-key attribute depends only on the primary key of its table.

Equipment (EquipmentID, EquipmentName, EquipmentLocation)
 Location (EquipmentLocation, LocationManager)

<u>EquipmentID</u>	EquipmentName	EquipmentLocation
E42	3D Printer	Lab B
E55	Oscilloscope	Lab A
E74	CNC Machine	Lab C

Table 4.8: Equipment Table [3NF]

<u>EquipmentLocation</u>	LocationManager
Lab B	Maria
Lab A	David
Lab C	Larry

Table 4.9: Location Table [3NF]

4.3 Entity-Relationship Diagram

Now that we have normalized our tables, we can create an *entity-relationship diagram* to visually represent the relationships between the tables. Often, this process may even clear up further FDs or issues with the database design.

In an ER diagram, each table is a box, and a line between two boxes is a relationship. The marks where a line meets a box tell you *how many* rows of that table can be involved. This style of drawing is called *crow's-foot notation*, after the three-pronged mark for “many”:

Mark at the end of the line	Meaning
two bars	exactly one
a circle and a bar	zero or one
a bar and a crow's foot	one or many
a circle and a crow's foot	zero or many

The circle means “zero is allowed,” the bar means “at least one,” and the crow's foot means “more than one is allowed.” Read each end of the line separately: the mark next to a box says how many rows of *that* box go with one row of the box at the other end.

Using our normalized tables, we can create an ER diagram to visualize the relationships between them.

We can say that our EquipmentLocation requires having a LocationManager, and (for the sake of our database), each LocationManager is responsible for managing a unique location. Therefore, we can say that LocationManager and EquipmentLocation have a *one-to-one relationship*.

Figure 4.1a shows this relationship as an ER diagram. Both ends of the line have two bars: each location has exactly one manager, and each manager has exactly one location.

How about when it comes to a Lab Location and Lab Equipment? Well, a Location that is recently built could have zero pieces of equipment associated with it, while an existing location might have one or more pieces of equipment. However, a given piece of equipment will always be associated with exactly one location, even if there are multiple of that piece of equipment in other locations (think: multiple 3D printers in different labs).

This scenario represents a *zero-to-many relationship* between the Location and Equipment tables (see Figure 4.1b). The end next to Location has two bars (each piece of equipment is in exactly one location), and the end next to Equipment has a circle and a crow's foot (a location has zero or many pieces of equipment).

A student can check out many pieces of equipment, and a piece of equipment can be checked out by many students (see Figure 4.1c). This is the relationship behind the whole Checkout table from the 2NF split. We did this so that we could accurately track which

student has checked out which piece (or pieces of equipment).

A relational database cannot store a many-to-many relationship directly, so we resolve it with a linking table. In the figure, the dashed line shows the many-to-many relationship between Student and Equipment, with a circle and a crow's foot at each end. Below it, the Checkout table turns it into two one-to-many relationships: each checkout belongs to exactly one student and exactly one piece of equipment, while a student, or a piece of equipment, can appear in zero or many checkouts.

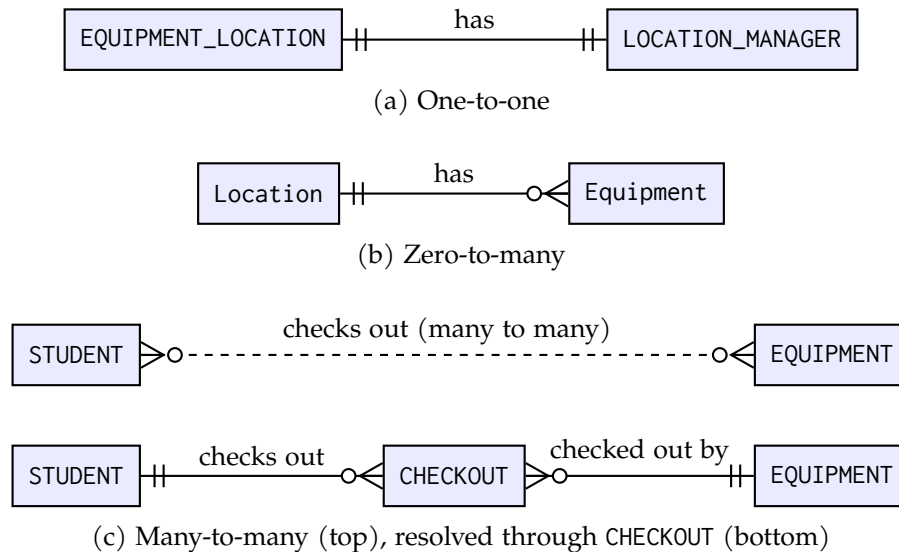


Figure 4.1: ER diagrams for the three relationship types in the lab equipment database. Two bars mean “exactly one”; a circle with a crow’s foot means “zero or many.”

4.4 Relational Algebra Basics

We have previously talked about set theory and binary operations such as *and*, *or*, and *not*. We can also apply those to data tables using *relational algebra*.

To begin, let us use our normalized equipment table (4.8) to demonstrate some basic relational algebra operations. We are going to fill it with some more data, however, to demonstrate the operations more clearly (see Table 4.10).

EquipmentID	EquipmentName	EquipmentLocation
E42	3D Printer	Lab B
E55	Oscilloscope	Lab A
E74	CNC Machine	Lab C

Table 4.10: Equipment Table [3NF]

Relational Algebra	SQL
π	SELECT
$\times$	, or CROSS JOIN
$\bowtie_{\theta}$	INNER JOIN <table> ON <condition>
σ	WHERE
τ	ORDER BY
$\cup$	UNION
$-$	EXCEPT
$\cap$	INTERSECT

4.5 Relating Two Tables

Now let us say we have two tables, Equipment and Location, and we want to relate them using a *foreign key*. The Equipment table has a column EquipmentLocation that references the primary key EquipmentLocation in the Location table. This allows us to link each piece of equipment to its corresponding location.

Let us first create the Equipment table with a foreign key referencing the Location table.

```

1 CREATE TABLE Location (
2     EquipmentLocation VARCHAR(50) PRIMARY KEY,
3     LocationManager VARCHAR(50) UNIQUE
4 );
5
6 CREATE TABLE Equipment (
7     EquipmentID VARCHAR(50) PRIMARY KEY,
8     EquipmentName VARCHAR(50),
9     EquipmentLocation VARCHAR(50) NOT NULL, -- the not null constraint ensures that
10     ↪ every piece of equipment must have a location
11     FOREIGN KEY (EquipmentLocation) REFERENCES Location(EquipmentLocation)
12 );

```

We *constrain* the values in the EquipmentLocation column of the Equipment table to only those that exist in the Location table, ensuring referential integrity between the two tables.

Relational algebra is a notation for operations on tables. The symbol π keeps only the listed columns, and the symbol $\subseteq$ means that every value on the left appears on the right. Using relational algebra, we can say that

$$\pi_{\text{EquipmentLocation}}(\text{Equipment}) \subseteq \pi_{\text{EquipmentLocation}}(\text{Location})$$

All values in the EquipmentLocation column of the Equipment table must exist in the

EquipmentLocation column of the Location table. However, it does not guarantee that every location in the Location table has corresponding equipment in the Equipment table. For example, here are the two tables again, with altered data

EquipmentLocation	LocationManager
Lab B	Maria
Lab A	David
Lab C	Larry

Table 4.11: Location Table

EquipmentID	EquipmentName	EquipmentLocation
E42	3D Printer	Lab B
E55	Oscilloscope	Lab A

Table 4.12: Equipment Table

Table constraints do not break when foreign key values are missing in the referencing table. For example, even though “Lab C” exists in the Location table, it does not appear in the Equipment table, and this is perfectly valid. However, if we tried to insert a row into the Equipment table with an EquipmentLocation value that does not exist in the Location table (such as “Lab D”), the database would reject it due to the foreign key constraint.

4.6 Advanced Queries

Request: Which labs have no equipment?

Algebra: $\pi_{\text{EquipmentLocation}}(\text{Location}) - \pi_{\text{EquipmentLocation}}(\text{Equipment})$

SQL:

```
SELECT EquipmentLocation FROM Location
EXCEPT
SELECT EquipmentLocation FROM Equipment;
```

4.7 Exercises

Exercise 13 Normalize to 3NF

A table Orders(OrderID, CustomerID, CustomerCity) has the functional dependencies $\text{OrderID} \rightarrow \text{CustomerID}$ and $\text{CustomerID} \rightarrow \text{CustomerCity}$. Decompose it into tables that are in 3NF.

Working Space

Answer on Page 79

Exercise 14 Translate to SQL

Using the Equipment table above, write an SQL query that returns the names of the equipment in Lab A.

Working Space

Answer on Page 79

Diodes, Rectifiers, and Photovoltaics

The phone in your pocket contains billions of tiny switches, all carved from a material that is neither a good conductor nor a good insulator: a *semiconductor*. The simplest semiconductor device is the *diode*, a one-way valve for current. Understand the diode, and you are most of the way to understanding solar cells, LEDs, transistors, and the chips inside every computer. (The Digital Logic chapter turns transistors into a circuit that adds.)

We are going to focus on semiconductors made of silicon. You can make semiconductors out of other materials, but silicon has proved to be very versatile and cost-effective.

5.1 Silicon

A silicon atom has 14 protons and 14 electrons. The inner 10 electrons are arranged in two complete and stable shells. The outer four electrons are in an incomplete shell, which allows them to form bonds with other atoms. These outer electrons are called *valence electrons*, and they are the ones that form bonds. A complete outer shell would have eight electrons.

Atoms can share electrons to get complete shells. This is called *covalent bonding*. When atoms share electrons, they form a *covalent bond*. The electrons are shared between the atoms, and the atoms are held together by the shared electrons.

In a silicon crystal, each silicon atom shares its outer four electrons with four neighboring atoms. With its four electrons and an electron borrowed from each of its four neighbors, each silicon atom has eight electrons in its outer shell. This is a complete shell, and the atoms are stable. With every electron locked into a bond, almost none are free to move, so pure silicon conducts electricity poorly.

Phosphorus has five valence electrons. If you put a phosphorus atom into silicon crystal, it shares four of its electrons with its neighbors, but it has one electron left over after the complete shell of eight is built. That extra electron is free to wander through the crystal and carry current. The more phosphorus you add, the better the silicon conducts. Adding small amounts of another element on purpose is called *doping*.

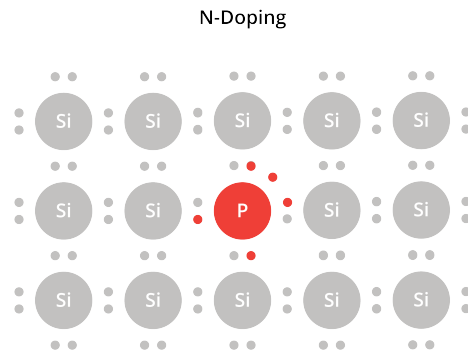


Figure 5.1: Phosphorus atom in a silicon crystal

Because this kind of doping frees up some electrons, which are negatively charged, we call this *n-doping*.

Boron has only three valence electrons. If you put a boron atom into silicon crystal, it shares three of its electrons with its neighbors, which leaves a "hole" that would like to be filled by an electron. We call this *p-doping*.

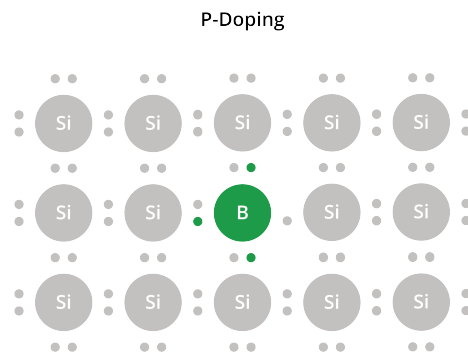


Figure 5.2: Boron atom in a silicon crystal

Both n-doped and p-doped silicon are conductors of electricity. In n-doped silicon, we say the loose electrons are the *carriers* of the current. In p-doped silicon, we say the holes are the carriers.

Doped silicon is still almost entirely silicon. Often only about one atom in a million is replaced with boron or phosphorus, and that is enough to change its conductivity

enormously.

5.2 Diodes

A diode is a semiconductor device that allows current to flow in one direction only.

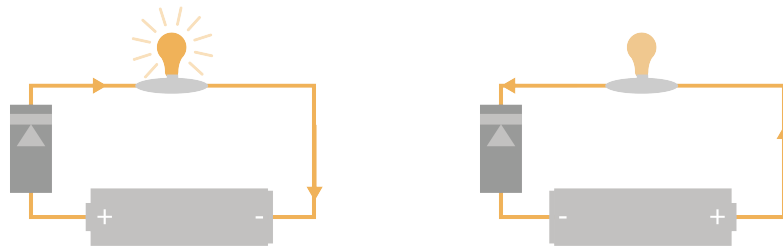


Figure 5.3: Current flows in one direction through the diode, not in the other direction.

Diodes can be made with silicon that has been n-doped and p-doped. A diode looks like this:



Figure 5.4: What Diodes Look Like

We say that current flows from the positive terminal to the negative terminal of a battery. However, remember that the electrons are actually moving from the negative terminal to the positive terminal. This is a common source of confusion in any discussion of diodes.

One end of the diode is called the *anode*. The anode is p-doped. The other end is called the *cathode*. The cathode is n-doped.

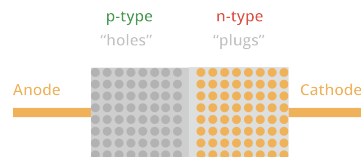


Figure 5.5: A p-n Diode

The region close to where the anode and cathode meet is called the *depletion region*. In this space, the spare electrons in the cathode (the n-doped material) migrate over to fill the holes in anode (the p-doped material). This leaves the anode side slightly negative (extra electrons) and the cathode side slightly positive (missing electrons), creating an electric field across the junction. For silicon, the voltage across this region is about 0.7 volts. It acts like a small hill that pushes back on any more electrons trying to cross from the cathode to the anode.

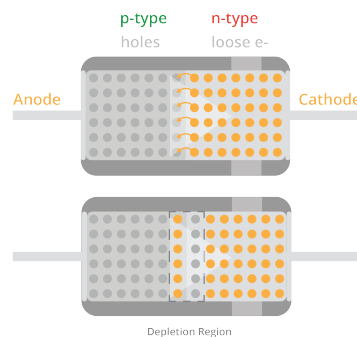


Figure 5.6: The depletion region in a diode

What happens if we overpower that 0.7 volts? Push electrons into the cathode with, say, 2 volts. This shrinks the depletion region, and the diode becomes a conductor. This is called *forward bias*. For example, if the negative end of a battery is connected to the cathode of a diode and the positive end is connected to the anode, current will flow through the diode.

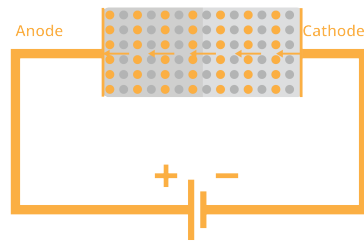


Figure 5.7: Electrons flow through the diode

If we reverse the battery (*reverse bias*), the depletion region grows, and the diode becomes an insulator.

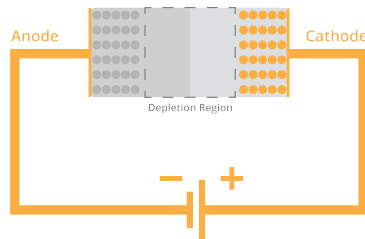


Figure 5.8: Reverse the electric field? No flow.

Here is the symbol for a diode:

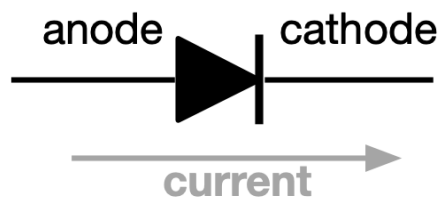


Figure 5.9: Diode Symbol

Conventional current flows in the direction of the arrow (electrons flow the other way). It will not flow in the opposite direction.

Light-emitting diodes. When an electron crosses the junction of a forward-biased diode

and falls into a hole, it gives up energy. In ordinary silicon diodes, that energy becomes heat. In an *LED*, made from materials like gallium arsenide, it comes out as a photon. The energy drop sets the photon's energy, and so its color: about 1.8 eV for red, 2.6 eV for blue. That is why LEDs need different voltages for different colors.

Worked example: current through a silicon diode. A diode that is conducting drops about 0.7 V. Put it in series with a 1 k Ω resistor and a 5 V supply. The resistor sees $5 - 0.7 = 4.3$ V, so the current is $I = \frac{4.3}{1000} = 4.3$ mA.

5.2.1 Rectifiers

One of the really useful things we can do with a diode is to convert alternating current (AC) into direct current (DC). This is called rectification. Remember that AC pushes and pulls the electrons back and forth in the wire. Rectification ensures that the electrons only move in one direction.

A *half-wave rectifier* uses a single diode to ensure that the current flows in one direction through the load.



Figure 5.10: Half-wave Rectifier

Current flows in the direction of the arrow on the diode. If the voltage is reversed, the diode will block the current from moving.

A *full-wave rectifier* uses four diodes to swap the direction of the current through the load when necessary.

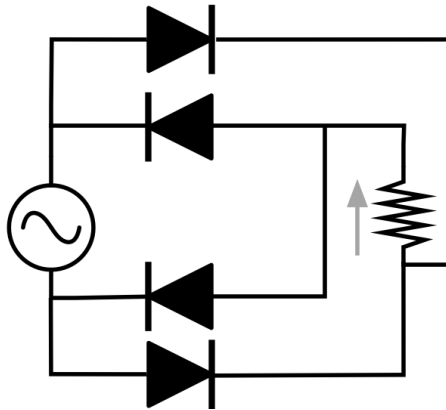


Figure 5.11: Full-wave Rectifier

5.2.2 Photovoltaics

A solar cell (or *photovoltaic* cell) is an LED run backward. It is a large, thin diode with a p-n junction just under the surface, where sunlight can reach it. When a photon with enough energy (more than about 1.1 eV for silicon) is absorbed near the junction, it knocks an electron loose, leaving a hole behind.

Now the built-in field of the depletion region takes over. It pushes the freed electron toward the n side (the cathode) and the hole toward the p side (the anode). Electrons pile up on the n side, and the cell becomes a small battery. The electrons cannot cross back through the junction against the field, so they take the long way: out through the wires, through the load, and back to the p side, where they fill the holes.

Each silicon cell makes only about half a volt, so a solar panel connects dozens of cells in series.

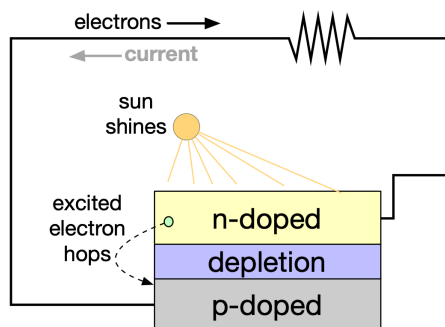


Figure 5.12: Solar Cell

Exercise 15 How many cells?

Each solar cell in a panel produces 0.5 V. How many cells must be connected in series to make a panel that produces 30 V?

Working Space

Answer on Page 79

Exercise 16 A Larger Supply

A 12 V supply, a silicon diode with a 0.7 V drop, and a 2 k Ω resistor are in series. Find the current.

Working Space

Answer on Page 79

How Cameras Work

It is a sunny day, and you are standing in a field a few meters from a cow. You pull out your phone and snap a picture. In a few thousandths of a second, light that left the Sun eight minutes ago bounces off the cow, squeezes through a lens smaller than a pea, and lands on a chip with 24 million tiny light meters. Let us follow that light the whole way.

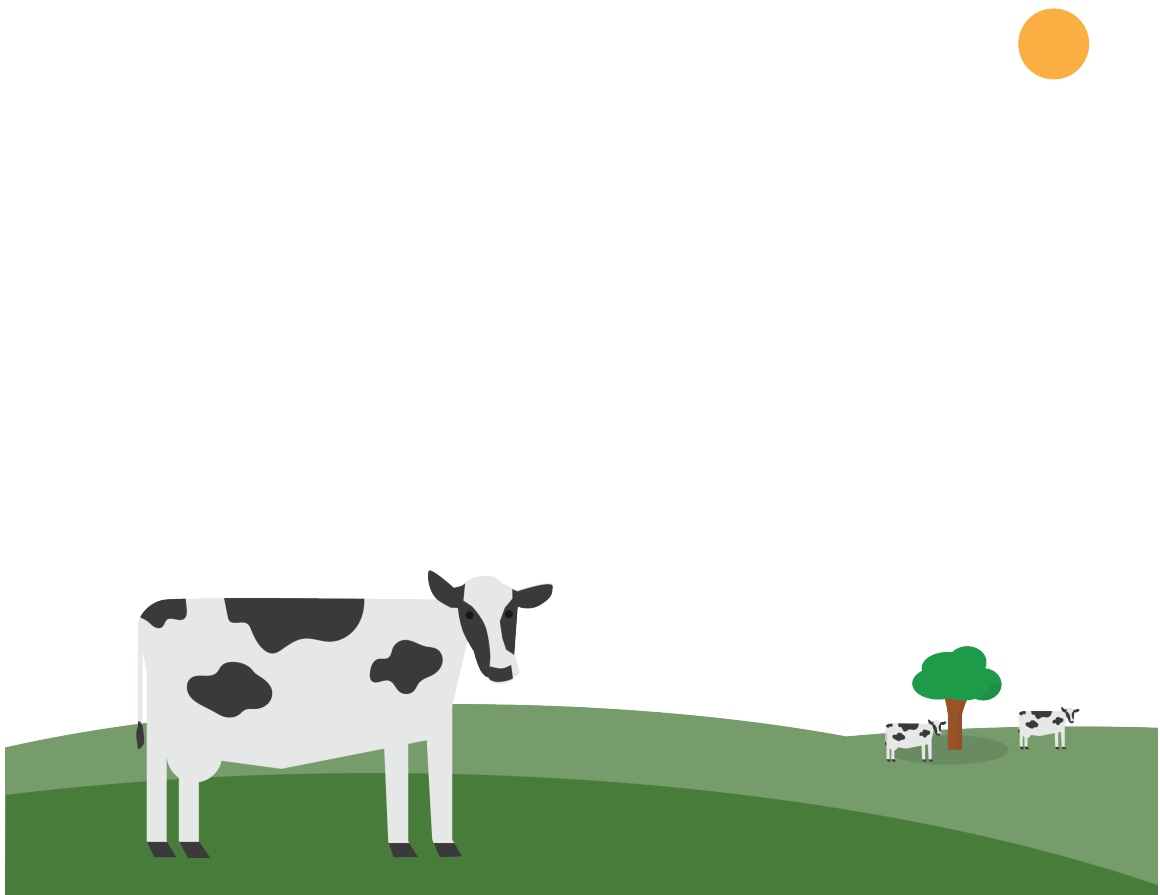


Figure 6.1: A cow in a field.

6.1 The Light That Shines On the Cow

The sun is a sphere of hot gas. About 70% of the gas is hydrogen. About 28% is helium. There is also a little carbon, nitrogen, and oxygen.

Gradually, the sun is converting hydrogen into helium through a process known as “nuclear fusion” (which we will be discussing more in a future chapter). A large amount of heat is created in this process. This heat makes the gases glow.

How does heat make things glow? The heat pushes the electrons into higher orbitals. When they come back down to a lower orbital, they release a photon, which travels away at the speed of light. (Recall from the photons chapter that light is a stream of these little packets of energy.)

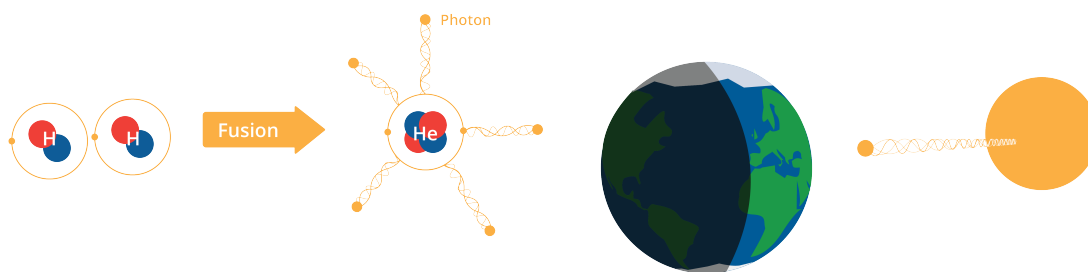


Figure 6.2: Photons from the Sun travel to the cow.

Heat is not the only way to push the electrons into a higher orbital. For example, a fluorescent lightbulb is filled with gas. When we pass electricity through the gas, its electrons are moved to a higher orbital. When they fall back to a lower orbital, light is created.

What is the frequency of the wave that the photon travels on? Depending on what orbital it falls from and how far it falls, the photon created has different amounts of energy. The amount of energy determines the frequency of the electromagnetic wave, and ultimately, its color!

The Sun’s hot gas contains many kinds of atoms, each with many different orbitals, and the atoms are jostling constantly. Thus, the light coming from the sun is made up of electromagnetic waves of many different frequencies.

We can see some of these frequencies as different colors, but some are invisible to humans, such as ultraviolet and infrared.

6.2 Light Hits the Cow

When these photons from the sun hit the cow, the hide and hairs of the cow will absorb some of the photons. These photons will become heat and make the cow feel warm. Some of the photons will not be absorbed – they will leave the cow. When you say “I see the cow,” what you are really saying is “I see some photons that were not absorbed by the cow.”

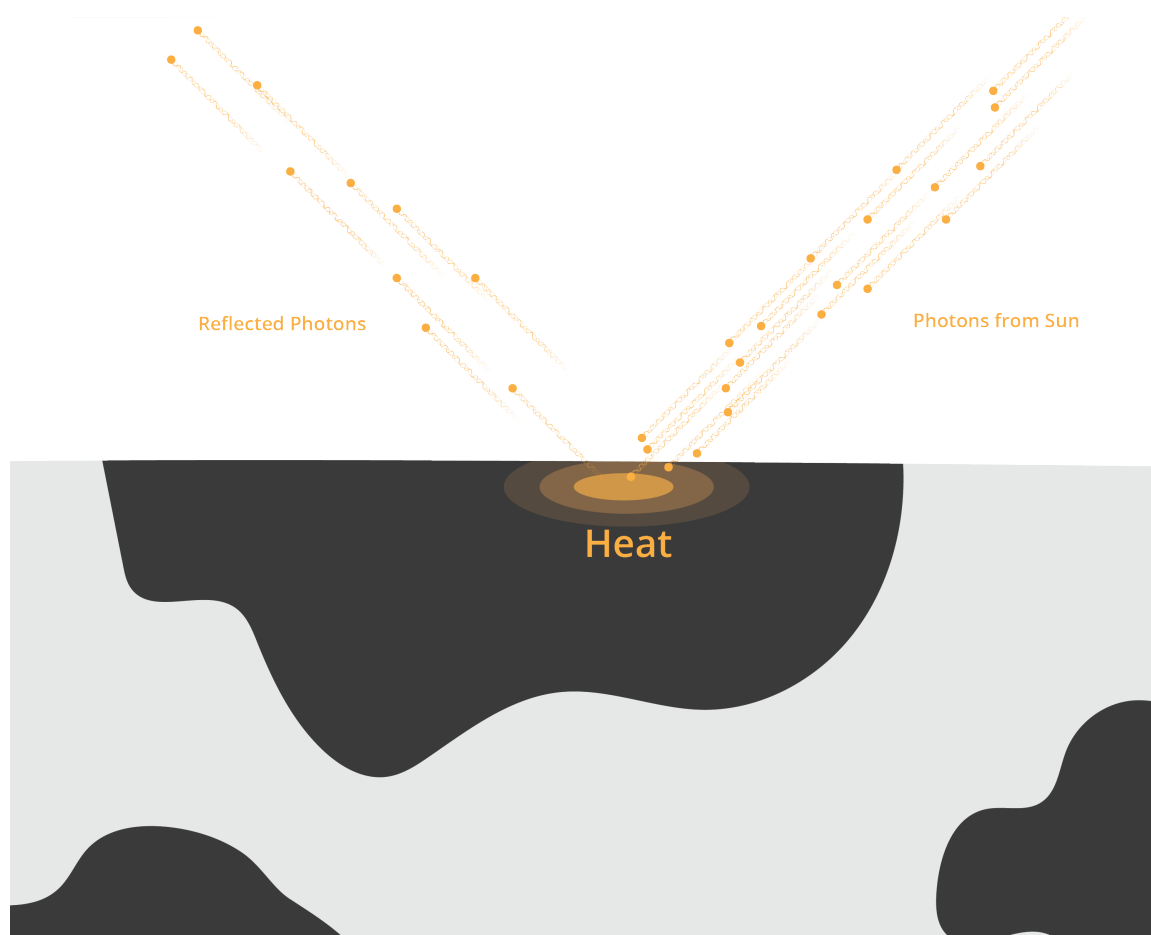


Figure 6.3: Photons hitting the cow create a heated spot, where some are absorbed.

Different materials absorb different amounts of each wavelength. A plant, for example, absorbs a large percentage of all blue and red photons that hit it, but it absorbs only a small percentage of the green photons that hit it. Thus, we say “That plant is green.”

White things absorb very small percentages of photons of any visible wavelength. Black things absorb very *large* percentages of photons of any visible wavelength. That is why, on a hot summer day, a black car with black seats and interior will heat up on the inside much hotter than a white car.

Before we go on, let us review: The sun creates photons that travel as electromagnetic waves of assorted wavelengths to the cow. Many of those photons are absorbed, but some are not. Some of those photons that are not absorbed go into the lens of our camera.

6.3 Pinhole camera

The simplest cameras have no lenses. They are just a box. The box has a tiny hole that allows photons to enter. The side of the box opposite the hole is flat and covered with film or some other photo-sensitive material.

The photons entering the box continue in the same direction they were going when they passed through the hole. Thus, the photons that entered from high hit the back wall at a low point. The photons that came from the left hit the back wall on the right. This is how the image is projected onto the back wall, rotated 180 degrees; what was up is down, what was on the left is on the right.

You can make one with a shoebox, a pin, and a sheet of wax paper for the back wall. Point it at a bright window, and you will see a dim, upside-down picture of the outdoors.

Worked example. A 2 m tall object stands 10 m from a pinhole camera whose back wall is 0.05 m behind the pinhole. How tall is the image?

Solution. The rays through the pinhole form two similar triangles, so the image is $\frac{0.05}{10}$ as tall as the object: $\frac{0.05}{10}(2) = 0.01$ m, or 1 cm.

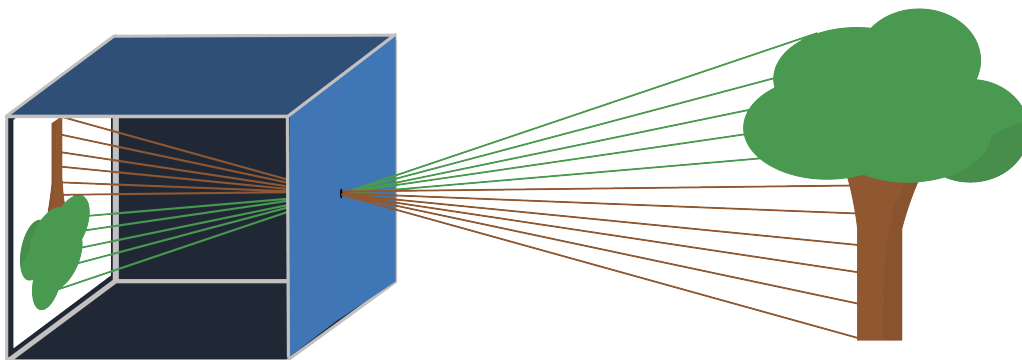


Figure 6.4: A pinhole camera flips the image it “sees” by 180°.

Exercise 17 **Height of the image**

Let us say that the pinhole is exactly the same height as the shoulder of the cow, and that the shoulder is directly above one hoof. This means the pinhole, the shoulder, and the hoof form a right triangle.

Now, let us say that the camera is being held perpendicular to the ground. The pinhole, the image of the shoulder, and the image of the hoof on the back wall of the camera now also form a right triangle.

These two triangles are similar.

The shoulder is 2 meters from the hoof. The cow is standing 3 meters from the camera. The distance from the pinhole to the back wall of the camera is 3 cm. How tall is the image of the cow on the back wall of the camera?

Working Space

Answer on Page 79

6.4 Lenses

Now, a quick review: A photon leaves the sun in some random direction. It travels 150 million km from the sun and hits a cow. It is not absorbed by the cow, and heads off in a new direction. It passes through the pinhole and hits the back wall of the camera. That seems incredibly improbable, right?

It is improbable. A pinhole lets in very little light, so pinhole images are dim, especially at dusk. You could make the hole bigger, but then light from each point of the cow lands on a whole blob of the back wall, and the picture blurs. The solution is a *lens*: a big opening that still makes a sharp image. Focus a lens on a wall with a dot drawn on it. All the photons from that dot that hit the lens, no matter which part of the lens they pass through, get bent to the same spot on the back wall of the camera. A lens gathers light

like a big hole but makes a sharp picture like a pinhole.

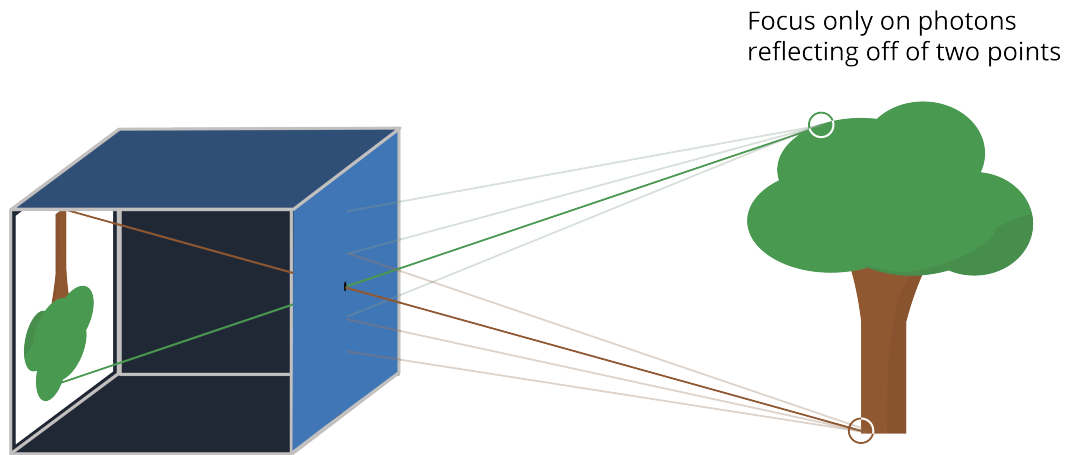


Figure 6.5: A lens bends all the light from one point of the object to one point of the image.

Note that the image still gets flipped. Light rays that enter the lens parallel to its axis all pass through a single point, the *focal point*.

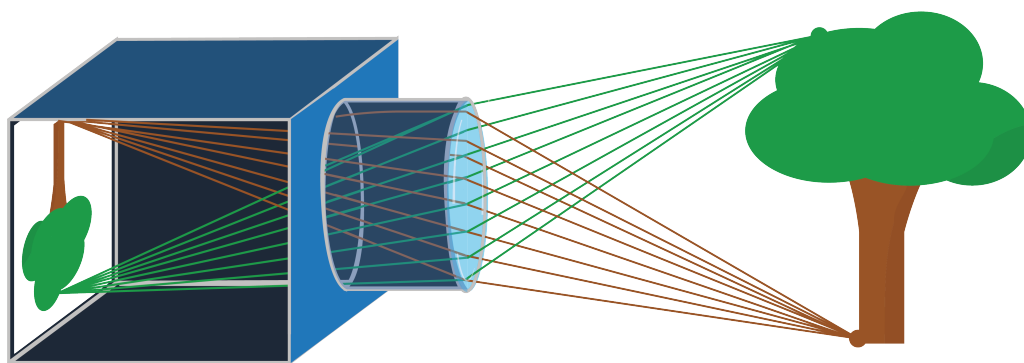


Figure 6.6: A lens still flips the image, but introduces a focal point.

The distance from the lens to its focal point is called the lens's *focal length*. Telephoto lenses, that let you take big pictures of things that are far away, have long focal lengths. Wide-angle lenses have short focal lengths.

6.5 Sensors

The camera on your phone has a sensor on the back wall of the camera. The sensor is broken up into tiny rectangular regions called pixels. When you say a sensor is 6000 by 4000 pixels (most common ratio for photography), we are saying the sensor is a grid of 24,000,000 pixels: 6000 pixels wide and 4000 pixels tall.

Each pixel is a tiny patch of silicon, a cousin of the solar cell from the diodes chapter. When a photon with enough energy (recall $E = hf$) strikes it, the photon frees an electron, and the pixel collects it. The minimum energy for silicon, called its *band gap*, is about 1.1 eV. A photon with less energy passes right through without freeing anything. During the exposure, each pixel counts up its electrons: more photons mean more electrons, which mean a brighter pixel. Then the camera measures the charge in each pixel as a voltage and turns it into a number.

But a pixel cannot tell what color a photon was. So a mosaic of tiny color filters sits on top of the sensor, one filter per pixel, in a checkerboard called a *Bayer pattern*. Half the pixels are under green filters, a quarter under red, and a quarter under blue. (There are twice as many green ones because our eyes are most sensitive to green.) The camera then estimates all three colors at every pixel by looking at its neighbors. So if your camera has a *resolution* of 6000×4000 , the finished image has 24,000,000 pixels, each with three numbers: the intensity of red, green, and blue light, for a total of 72,000,000 numbers. We call these numbers “RGB” for Red, Green, and Blue. The RGB values range from 0 – 255 for each channel. We will talk more about this when hexadecimal is introduced.

Exercise 18 Camera Pixels

A digital camera sensor is made of silicon, with a band gap of 1.1 electronvolts. If the sensor is exposed to monochromatic light with a wavelength of $\lambda = 1200$ nm, what does the resulting image look like? Why?

Hint: $E = \frac{1240}{\lambda}$ where λ is in nanometers and E is in electronvolts (eV).

Working Space

Answer on Page 79

6.6 Aspect Ratio

When you are buying a monitor, TV, or even smartphone, you may hear the term **aspect ratio**. The aspect ratio is a set of two numbers representing the ratio between the width and the height. Conventionally, the width comes first, separated by a colon, and then the height:

width : height

Some of the most common aspect ratios are:

- 16 : 9 (known as Widescreen)
- 17 : 9
- 9 : 16
- 4 : 3 (known as Fullscreen)
- 4 : 5
- 3 : 2
- 1 : 1 (perfect square, like on a smartwatch)

As we talked about above, the aspect ratio does not describe the actual size of the screen or image. Instead, it describes the shape. A movie theatre screen can have the same aspect ratio as your monitor, even if the screen is 360 inches and the monitor is 24 inches.

Aspect ratio describes the proportional relationship between width and height. Resolution describes the total number of pixels used to represent an image.

For example, both of the following resolutions have a 16 : 9 aspect ratio:

1280×720

1920×1080

Mathematically, the aspect ratio is a dimensionless quantity. Since both width and height are measured in the same units (such as centimeters, inches, or pixels), the units cancel when forming the ratio. This means aspect ratio describes proportion rather than physical measurement. Most often, we determine the aspect ratio of a digital display by comparing the number of horizontal pixels to vertical pixels.

For example:

$$1920 \times 1080 \Rightarrow \frac{1920}{1080} = \frac{16}{9} = 16 : 9$$

Digital cameras and smartphone cameras capture images using a rectangular sensor. The shape of this sensor determines the default aspect ratio of the image. For example:

- Many digital cameras use a 3 : 2 sensor.
- Micro Four Thirds cameras use 4 : 3.
- Most video recording uses 16 : 9.

Important: when you change the aspect ratio setting on a camera, the device does not change the sensor itself. Instead, it crops part of the image to match the selected ratio. Cropping reduces the number of recorded pixels, which may reduce file size. However, actual file size also depends on compression methods used by the device.

Human vision covers a wider angle side to side than up and down, so wide shapes like 16 : 9 feel natural and immersive. That is why TVs and movie screens are wide. But phones are usually held upright, so short videos made for phones use the tall 9 : 16 shape, which is just 16 : 9 turned on its side.

6.6.1 Printing

You can imagine this may cause a problem for printing out photos. Most printing paper is 8.5 inches $\times$ 11 inches or 8 inches $\times$ 10 inches. This causes an issue, as that fits a 8.5 : 11 $\iff$ 17 : 22 or 4 : 5 aspect ratio, respectively. We combat this by using very standardized printing sizes (in inches):

- 6 $\times$ 4 $\rightarrow$ 3 : 2,
- 7 $\times$ 5 $\rightarrow$ 7 : 5,
- 10 $\times$ 8 $\rightarrow$ 5 : 4,
- and, rarely, 8.5 $\times$ 11 $\rightarrow$ 1.29 : 1

Going to any framing or printing store will give you a very standardized set of sizes, or some multiple of these.

Most smartphones take photos at 4 : 3 or 3 : 2. A 3 : 2 photo fits a 6 $\times$ 4 inch print exactly; a 4 : 3 photo fits an 8 $\times$ 6 inch print.

Exercise 19 Will it fit?

A phone photo is 4032×3024 pixels. What is its aspect ratio? If you print it as a 6×4 inch photo, how much of the image must be cropped?

Working Space

Answer on Page 80

Exercise 20 Taller Person

A 1.8 m person stands 20 m from the same camera. How tall is the image on the sensor?

Working Space

Answer on Page 80

How Eyes Work

You have been carrying two cameras around your whole life. They focus automatically, adjust to sunlight a billion times brighter than starlight, and fix themselves when they get scratched. In this chapter, we will see how they work, what goes wrong with them, and why your TV only needs three colors to fool them.

Dr. Craig Blackwell has made a great video on the mechanics of the eye: <https://youtu.be/Z8asc2SfFHM>

Mechanically, your eye works a lot like the camera from the last chapter. The eye is a sphere with two lenses on the front. The outer lens is called the *cornea*, and it does most of the bending of light. The second lens is simply called “the lens.” Tiny muscles squeeze or relax the lens to change its shape, which fine-tunes the focus for near or far objects. (A camera does this by moving its lens instead.) Between the two lenses is an aperture that opens wide when there is very little light, and closes very small when there is bright light. The opening is called the *pupil* and the tissue that forms the pupil is called the *iris*. When people talk about the color of your eyes, they are talking about the color of your iris. The blackness at the center of your iris is your pupil.

There are two types of photoreceptor cells in your retina: rods and cones. The rods are more sensitive; in very dark conditions, most of our vision is provided by the rods. The cones are used when there is plenty of light, and they let us see colors.

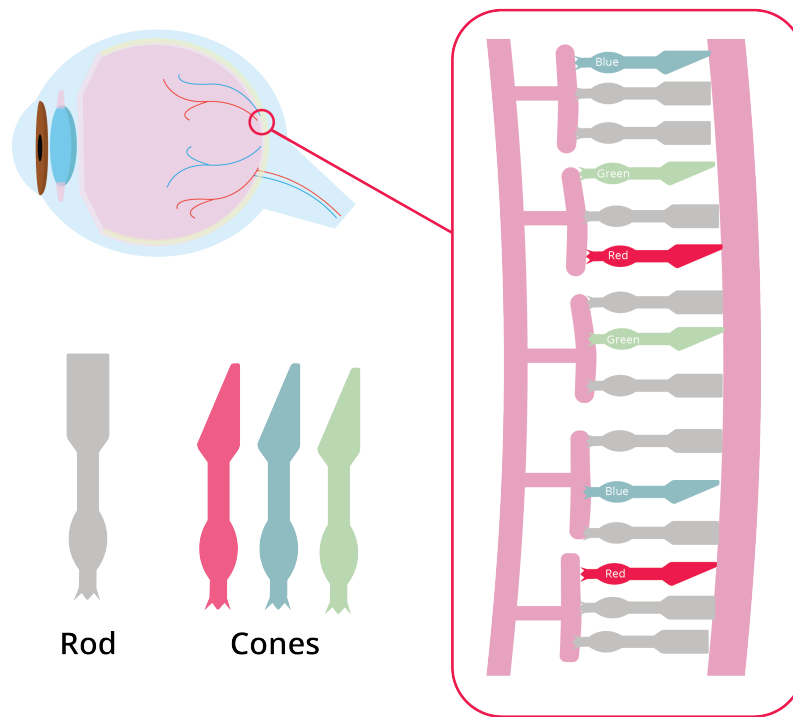


Figure 7.1: Rods and cones, the light-sensing cells of the retina.

The white part around the outside of the eyeball? That is called the *sclera*.

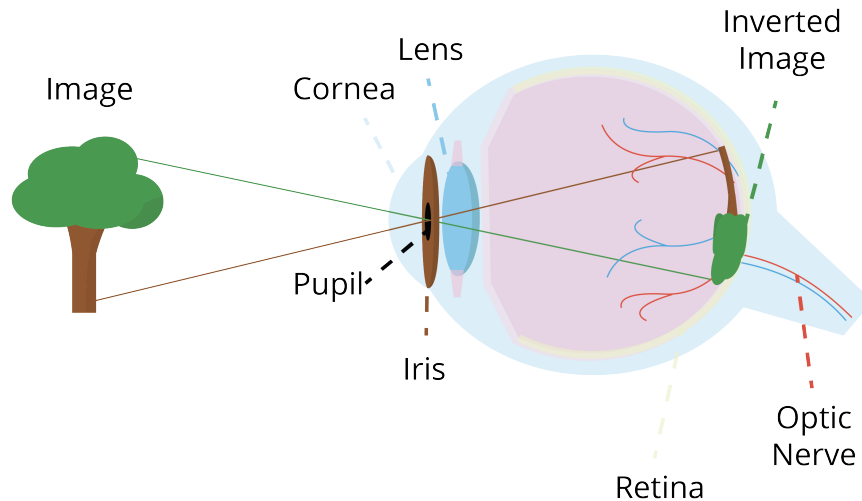


Figure 7.2: The eye works just like a camera's lens.

The walls of the eye are lined inside with the *retina*, which has sensors that pick up the light and send impulses down the optic nerve to your brain.

Just like a camera, the image is flipped upside down when it lands on the back of the eye (see figure 7.2). Your brain flips it back.

Where the optic nerve leaves the eye, there are no light sensors at all. Every eye has this *blind spot*, but you never notice it, because your brain fills in the gap with whatever is around it.

7.1 Eye problems

Now that you know the mechanics of the eye, let us go over a few things that commonly go wrong with the eye.

7.1.1 Glaucoma

The space between your cornea and lens is filled with a fluid called *aqueous humor*. To feed the cells of the cornea and lens, the aqueous humor carries oxygen and nutrients

like blood would, but unlike blood, it is transparent so you can see. Aqueous humor is constantly being pumped into and out of that chamber. If aqueous humor has trouble exiting, the pressure builds up and can damage the eye. This is known as *glaucoma*. See Figure 7.3.

7.1.2 Cataracts

The lens should be clear. As a person ages, the proteins in the lens break down and clump together, becoming opaque. This can also be accelerated by diabetes, too much exposure to sunlight, obesity, and high blood pressure. From the outside, the eye will look cloudy. This is called a *cataract*, and it makes it difficult for the person to see.

This problem can be corrected, however. The person's cloudy lens is removed and replaced with a clear, manufactured lens.

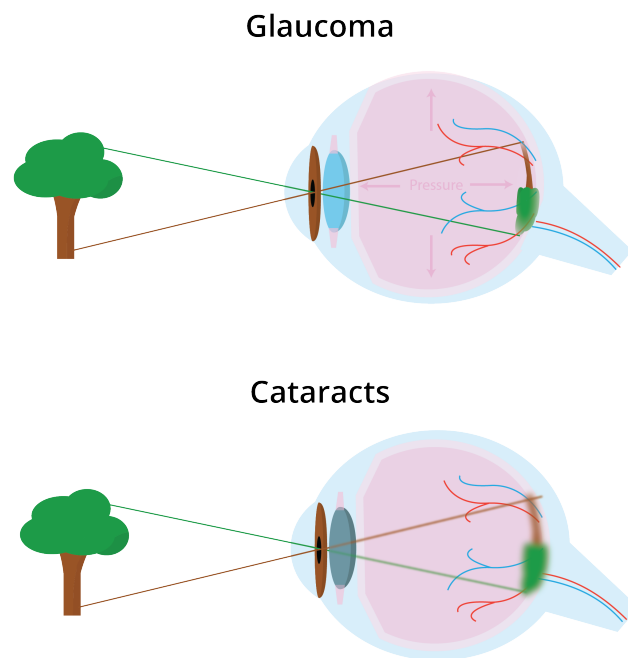


Figure 7.3: Cataracts and Glaucoma represented in the eye cross section.

7.1.3 Nearsightedness, farsightedness, and astigmatism

If you are in a dark room and a tiny LED is turned on, the photons from that LED can pass through your cornea in many different places. If your eye is focusing on that light correctly, all the photons should meet up at the same place on the retina.

Figure 7.4 shows what happens when they do not.

If the lenses are bending the light too much, the photons meet up before they hit the retina and get smeared a bit across it. To the person, the LED would appear blurry. The eye is said to be *nearsighted* or *myopic*. This signals that near objects appear correctly, but farther objects appear blurry.

If the lenses are not bending it enough, the photons would meet up behind the retina. Once again, they get smeared a bit across the retina and the LED looks blurry to the person. The eye is said to be *farsighted* or *hyperopic*. The objects which are distant can appear clearly, while closer objects are heavily blurred.

Your lenses are supposed to bend the photons the same amount vertically and horizontally. If one dimension is focused, but the other is myopic or hyperopic, the eye is said to have *astigmatism*.

Myopia, hyperopia, and astigmatism can be corrected with glasses or contact lenses. A nearsighted eye bends light too much, so its glasses use *diverging* lenses that spread the light a little before it enters the eye. A farsighted eye needs *converging* lenses for extra bending. Doctors can also do surgical corrections, usually by reshaping the cornea with a laser.

Eye doctors describe lenses by their *power* in *diopters*: $P = \frac{1}{f}$, with f in meters. Converging lenses have positive power, and diverging lenses have negative power.

Worked example. A pair of reading glasses has converging lenses with a focal length of 0.5 m. What is their power?

Solution. $P = \frac{1}{0.5} = +2$ diopters. A nearsighted person's prescription might read -2.5 , meaning diverging lenses with $f = -0.4$ m.

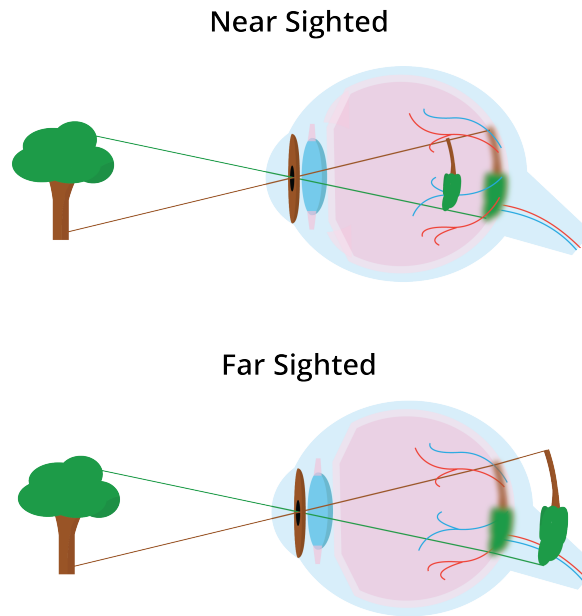


Figure 7.4: Nearsightedness versus farsightedness.

7.2 Seeing colors

TED-Ed has made a good video on how we see color. Watch it here: https://youtu.be/l8_fZPHasdo

When a rainbow forms, you are seeing different wavelengths separating from each other. In the rainbow:

- Red is about 650 nm.
- Orange is about 600 nm.
- Yellow is about 580 nm.
- Green is about 550 nm.
- Cyan is about 500 nm.
- Blue is about 450 nm.
- Violet is about 400 nm.

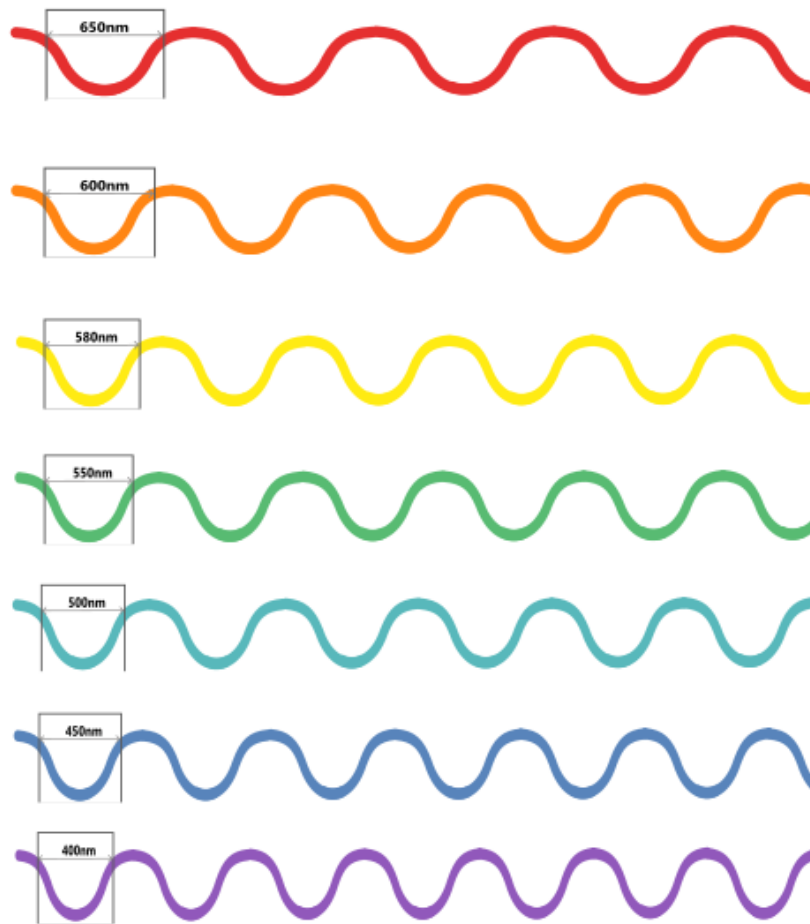


Figure 7.5: All colors have different wavelengths.

If you shine a light with a wavelength of 580 nm on a white piece of paper, you will see yellow.

However, if you shine two lights with wavelengths of 650 nm (red) and 550 nm (green), you will also see yellow.

Why? Our ears can hear two different frequencies at the same time. Why cannot our eyes see two colors in the same place?

As mentioned above, the cone photoreceptors in our eyes let us see colors. There are three kinds of cones. Each responds to a wide range of wavelengths, but each is most sensitive near a different one:

- “Red” (long-wavelength) cones: most sensitive near 560 nm, and still responding out

to about 700 nm.

- “Green” (medium-wavelength) cones: most sensitive near 530 nm.
- “Blue” (short-wavelength) cones: most sensitive near 420 nm.

The red and green cones overlap a lot, which is why light between them excites both.

When a wavelength of 580 nm hits your retina, it excites the red and green receptors, and your brain interprets that mix as yellow.

Similarly, when light that contains both 650 nm and 550 nm waves hits your retina, it excites the red and green receptors, and your brain interprets that mix as yellow.

You cannot tell the difference!

Now we know why the sensors on the camera are RGB. The camera is recording the scene as closely as necessary to fool your eye.

A TV or a color computer monitor only has three colors of pixels: red, green, and blue. By controlling the mix of them, it creates the sensation of millions of colors. (With 256 levels for each of red, green, and blue, a screen can make $256^3 \approx 16.7$ million colors.)

7.3 Pigments

A color printer works in the opposite fashion. Instead of radiating colors, it puts pigments on the paper that absorb certain frequencies. A pigment that absorbs only wavelengths near 650 nm (red) will appear to your eye as cyan. This makes sense, because the sensation of cyan is created when your blue and green receptors are activated.

Thus, pigment colors come in:

- Cyan: absorbs wavelengths around red
- Magenta: absorbs wavelengths around green
- Yellow: absorbs wavelengths around blue

If you buy ink for a color printer, you know there is typically a fourth ink: black. If you put cyan, magenta, and yellow pigments on paper, the mix will not absorb all the visible spectrum in a consistent manner. Our eyes are pretty sensitive to this, so we would see brown. This is why we add black ink to get pretty grays and blacks.

We call this approach to color CMYK (as opposed to RGB). If an artist is creating an image to be viewed on a screen, they will typically make an RGB image. If they are creating an

image to be printed using pigments, they typically create a CMYK image. (Most of us do not care so much — we just let the computer do conversions between the two color spaces for us.)

Exercise 21 **Mixing light**

A screen turns on its red and blue pixels at full brightness and leaves green off. What color do you see? What if a printer puts magenta and yellow ink on white paper?

Working Space

Answer on Page 80

Exercise 22 **Stronger Lens**

A corrective lens has focal length 0.25 m. What is its power in diopters?

Working Space

Answer on Page 80

Answers to Exercises

Answer to Exercise 1 (on page 12)

Groupthink and the bandwagon effect: quieter team members may go along with the loudest voices to keep the peace, so the manager hears agreement that may not be real. The three loudest voices are also a small, unrepresentative sample of the team.

Answer to Exercise 2 (on page 12)

1. Self-serving bias.
2. Fundamental attribution error. (Maybe the driver is rushing to the hospital.)
3. Survivorship bias. We hear about the dropouts who became billionaires, not the millions who did not.
4. The spotlight effect.

Answer to Exercise 3 (on page 14)

1. The population is all households in the city.
2. The frame is the city's water billing records.
3. The sample is the 200 households selected and interviewed.
4. No. A census would require interviewing every household in the city. Only 200 were selected here.

Answer to Exercise 4 (on page 15)

- Library volunteers: Census. With only 8 volunteers, asking all of them takes almost no effort.

- Battery lifespan: Sample. Testing can destroy the batteries, so a census would leave nothing to sell.
- Commute times: Sample. The population of 40 million is far too large for a full census to be practical.
- Dietary restrictions: Census. With only 22 guests, collecting complete information is straightforward and important for planning.

Answer to Exercise 5 (on page 19)

1. Number the 300 drivers from 001 to 300. Enter a random number table at any point and read three-digit numbers. Select any driver whose number falls between 001 and 300, skipping values from 301 to 999 and any numbers that have already been selected, until 25 drivers have been chosen. Alternatively, use `randInt(1,300,25)` on a TI-84, ignoring repeats.
2. This is a convenience sample. Drivers who are easy to reach at shift changes may work particular routes or shifts that differ systematically from others. For example, they may have different safety exposure than drivers on overnight or rural routes.
3. No. An SRS requires that every possible group of 25 drivers from all 300 has an equal chance of being chosen. This procedure cannot produce a sample where all 25 come from the same depot, so not all groups of 25 are possible. This is a stratified sample, not an SRS.

Answer to Exercise 6 (on page 20)

1. Answers will vary. The five sample means differ from each other and from the true mean, because each SRS contains a different set of entries. A sample that happens to include one or two very large entries lands far above the true mean.
2. With $n = 80$, the sample means cluster more closely around the true mean. Sampling error shrinks as the sample size grows.

Answer to Exercise 7 (on page 22)

- Supermarket: Cluster sampling. The stores are clusters, which means entire clusters are included once selected.
- University: Stratified random sampling. The four faculties are strata. An SRS is taken

from each.

- Water utility: Systematic sampling. Random start followed by every 25th unit.
- Film festival: Simple random sample. Every possible group of 90 attendees is equally likely.

Answer to Exercise 8 (on page 24)

1. Answers will vary. Stratified means stay within a few points of 79.7. Each cluster mean is exactly 70.3, 76.5, or 92.4, depending on the grade selected.
2. No. An SRS of 9 can oversample one grade and undersample another. Because 11th graders score much higher than 9th graders, an SRS heavy in 11th graders overestimates the true mean.
3. The stratified sample. Every grade is always represented, and scores vary less within a grade than across the whole school, so the estimate stays close to the true mean from run to run.
4. The cluster sample is a single grade, so its mean depends entirely on which grade is chosen. The grades differ in score, so the cluster mean can land far from 79.7.

More detail. Your runs will differ, because each run draws new random samples. The answers below describe what the code produces. For the data set created with seed 42, the true mean is 79.7, and the grade means are 70.3 (9th), 76.5 (10th), and 92.4 (11th).

1. **SRS.** An SRS of 9 students usually includes all three grades, but not always. For a given grade, the chance that none of its 10 students is chosen is about 1.2%, so the chance that some grade is missing is about 3.5%.
2. **Stratified.** Every grade contributes three students, so the sample mean is usually close to 79.7. Over many repeated draws, the stratified mean has a standard deviation of about 1.6 points, compared with about 3.0 points for the SRS.
3. **Cluster.** The cluster mean is the mean of the one grade chosen, so it is about 70.3 for 9th grade, 76.5 for 10th, or 92.4 for 11th. Over many repeated draws, its standard deviation is about 9.3 points.
4. **Table.** Your runs will vary, but expect the stratified means to fall within about 3 points of 79.7, the SRS means within about 5 points, and the cluster means at one of the three grade means. The cluster method is the least stable.
5. **Why.** The stratified sample is more stable because it always includes every grade, so differences between grades cannot shift the mean much. The cluster sample is the least stable because the whole sample comes from a single grade, and the grade means differ

by about 22 points.

Answer to Exercise 9 (on page 26)

1. Nonresponse bias. The 92% who did not respond are unlikely to be satisfied in the same proportions as those who did. Highly satisfied and highly dissatisfied subscribers are both more motivated to respond than those with neutral feelings, but the pattern of non-response is unknown and the reported figure cannot be trusted.
2. Undercoverage bias. Residents of lower-income neighborhoods are excluded from the frame entirely. Recycling behavior may differ systematically by neighborhood type, so the results cannot generalize to the full population.
3. Wording effect bias. The question assumes the office has improved productivity and asks only “how much.” Respondents are not given a neutral opportunity to say productivity has not changed or has declined.
4. Response bias. The formal appearance of the interviewers creates social pressure to give a health-conscious answer. Respondents adjust their reported behavior to match what they believe the interviewers expect.

Answer to Exercise 10 (on page 27)

$$\frac{25}{300} \approx 8.3\%.$$

Answer to Exercise 11 (on page 32)

```
1 UPDATE bike SET status='broken' WHERE status='Flat tire';  
2 UPDATE bike SET status='rented' WHERE status='Rented';
```

After these, the distinct values are rented, broken, available, NULL (shown as a blank line), good, and out. Nobody knows what good or out were supposed to mean, so you would have to ask someone at the shop before fixing those.

Answer to Exercise 12 (on page 32)

```
1 SELECT bike_id FROM bike WHERE purchase_price > 330 AND brand='Trek'
```

Answer to Exercise 13 (on page 44)

The city depends on the customer, not on the order, so it is a transitive dependency. Split the table into Orders(OrderID, CustomerID) and Customers(CustomerID, CustomerCity). Each non-key column now depends only on the key of its own table.

Answer to Exercise 14 (on page 45)

```
1 SELECT EquipmentName FROM Equipment WHERE EquipmentLocation = 'Lab A';
```

The result is Oscilloscope, the equipment with ID E55.

Answer to Exercise 15 (on page 54)

$\frac{30}{0.5} = 60$ cells, which is a very common panel size.

Answer to Exercise 16 (on page 54)

$I = \frac{12-0.7}{2000} = 5.65 \times 10^{-3}$ A, or 5.65 mA.

Answer to Exercise 17 (on page 59)

The two triangles are similar, so $\frac{x}{3 \text{ cm}} = \frac{2 \text{ m}}{3 \text{ m}}$, and $x = 2$ cm. The image of the cow is 2 cm tall (and upside down).

Answer to Exercise 18 (on page 61)

The pixels will remain dark, or black, because the energy of each incoming photon is less than the band gap of silicon. Calculating $E = \frac{1240}{1200} = 1.033$ eV, which is less than the work function of 1.1 eV.

Answer to Exercise 19 (on page 64)

$\frac{4032}{3024} = \frac{4}{3}$, so it is 4 : 3. To fit 3 : 2, keep the full width of 4032 pixels and cut the height to $\frac{2}{3}(4032) = 2688$ pixels. That crops $3024 - 2688 = 336$ pixel rows, about 11% of the image.

Answer to Exercise 20 (on page 64)

$$\frac{0.05}{20}(1.8) = 0.0045 \text{ m, or } 4.5 \text{ mm.}$$

Answer to Exercise 21 (on page 73)

Red plus blue light looks magenta. Magenta ink absorbs green, and yellow ink absorbs blue, so only red light is reflected: the paper looks red.

Answer to Exercise 22 (on page 73)

$$P = \frac{1}{0.25} = 4 \text{ diopters.}$$



INDEX

- alternate key, 36
- anode, 49
- aqueous humor, 67
- aspect ratio, 62
- astigmatism, 69
- band gap, 61
- bandwagon effect, 5
- Bayer pattern, 61
- biased, 16
- blind spot, 67
- boron, 48
- camera, 55
- camera
 - lens, 59
 - lens
 - focal length, 61
 - pinhole, 58
- candidate key, 35
- carriers, 48
- cataract, 68
- cathode, 49
- census, 13
- charge carriers, 48
- cluster sampling, 21
- Clusters, 22
- clusters, 22
- cmyk, 72
- cognitive bias, 3
- cognitive biases, 3
- color, 70
- composite keys, 35
- confirmation bias, 9
- convenience sample, 17
- cornea, 65
- covalent bond, 47
- covalent bonding, 47
- crow's-foot notation, 41
- curse of knowledge, 5
- DELETE, 32
- depletion region, 50
- determines, 37
- diode, 47
 - symbol, 51
- diopters, 69
- DISTINCT, 31
- doping, 47
- Dunning-Kruger effect, 8
- entity-relationship diagram, 41
- ER diagram, *see* entity-relationship diagram
- eye, 65
 - astigmatism, 68
 - cataract, 68
 - cornea, 65
 - farsightedness, 68
 - glaucoma, 67
 - hyperopia, 69
 - iris, 65

- myopia, 68
- nearsightedness, 68
- pupil, 65
- retina, 67
- sclera, 66
- false consensus, 6
- farsighted, 69
- focal length, 61
- focal point, 60
- foreign key, 43
- forward bias, 50
- frame, 13
- full-wave rectifier, 52
- functional dependency, 37
- fundamental attribution error, 3
- glaucoma, 68
- Groupthink, 5
- groupthink, 5
- half-wave rectifier, 52
- hole (semiconductor), 48
- hyperopic, 69
- image sensor, 61
- In-group favoritism, 4
- in-group favoritism, 4
- INSERT, 31
- iris, 65
- judgmental sampling, 17
- LED, 52
- lens, 59
- myopic, 69
- n-doped, 48
- n-doping, 48
- natural key, 36
- nearsighted, 69
- nonresponse bias, 25
- normalization, 36
- one-to-one relationship, 41
- p-doped, 48
- p-doping, 48
- p-n junction, 53
- partial dependency, 37
- phosphorus, 47
- photon
 - colors of, 57
 - wavelength, 57
- photovoltaic, 53
- photovoltaics, 53
- pigment, 72
- pixel, 61
- population, 13
- population parameter, 13
- power, 69
- primary key, 35, 36
- printing photos, 63
- pupil, 65
- quota sampling, 17
- rainbow, 70
- random number table, 19
- rectifier, 52
- relational algebra, 42
- relational database, 29
- relational database management system,
 - 29
- resolution, 61
- response bias, 25
- retina, 67
- reverse bias, 51
- rgb, 61, 71
- rods and cones, 65
- sample, 13
- sample survey, 13
- sampling error, 25
- sampling with replacement, 18
- sampling without replacement, 18
- sclera, 66
- Self-serving bias, 4
- self-serving bias, 4
- semiconductor, 47
- silicon, 47
- simple random sample, 18
- snowball sampling, 17
- solar cells, 53
- spotlight effect, 7

- SQL, [29](#)
- SQLite, [29](#)
- sqlite3 (Python library), [33](#)
- Strata, [22](#)
- strata, [22](#)
- stratified random sampling, [21](#)
- super key, [35](#)
- super keys, [35](#)
- surrogate key, [36](#)
- surrogate keys, [36](#)
- survivorship bias, [11](#)
- systematic sampling, [21](#)

- telephoto lens, [61](#)
- The bandwagon effect, [5](#)
- TI-84
 - randInt, [19](#)
- transitive dependencies, [40](#)
- transitive dependency, [37](#)

- undercoverage bias, [26](#)
- UPDATE, [31](#)
- UUIDs, [36](#)

- valence electron, [47](#)
- valence electrons, [47](#)
- voluntary response sample, [17](#)

- WHERE, [30](#)
- wide-angle lens, [61](#)
- wording effect bias, [26](#)

- zero-to-many relationship, [41](#)