Ch. 1. Voting theory

As a college student, the mathematician Kenneth Arrow studied both social science and mathematics. He later became known for applying mathematical reasoning to real-world problems that at first might seem quite unrelated to math. We begin by quoting his most celebrated discovery.

Arrow's Impossibility Theorem
No method for determining the result of an election involving three or more candidates is both democratic and *consistently* fair.

Now, a **theorem** is *a mathematically precise result that can be proven to be true*. For example, you may recall the *Pythagorean Theorem*[1] from high school. We will not "prove" Arrow's theorem, but after completing Chapter 1, you should be able to convince yourself that Arrow's theorem is in fact true.

1. Preference ballots and preference schedules

Ex.: Mathematics Appreciation Society.
Suppose an election is held for the President of the Math Appreciation Society at Tasmania State University, and the following 37 ballots are collected. Each voter has been asked to rank their choices of candidate Alice, Boris, Carmen, and Dave with the numbers 1st, 2nd, 3rd, and 4th. What's the best way to decide who won?

Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot
1st A	1st B	1st A	1st C	1st B	1st C	1st A	1st B	1st C
2nd B	2nd D	2nd B	2nd B	2nd D	2nd B	2nd B	2nd D	2nd B
3rd C	3rd C	3rd C	3rd D	3rd C	3rd D	3rd C	3rd C	3rd D
4th D	4th A	4th D	4th A	4th A	4th A	4th D	4th A	4th A

Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot
1st A	1st C	1st D	1st A	1st A	1st C	1st A	1st C	1st D
2nd B	2nd B	2nd C	2nd B	2nd B	2nd B	2nd B	2nd B	2nd C
3rd C	3rd D	3rd B	3rd C	3rd C	3rd D	3rd C	3rd D	3rd B
4th D	4th A	4th A	4th D	4th D	4th A	4th D	4th A	4th A

Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot
1st C	1st A	1st D	1st D	1st C	1st C	1st D	1st A	1st D
2nd B	2nd B	2nd C	2nd C	2nd B	2nd B	2nd C	2nd B	2nd C
3rd D	3rd C	3rd B	3rd B	3rd D	3rd D	3rd B	3rd C	3rd B
4th A	4th D	4th A	4th A	4th A	4th A	4th A	4th D	4th A

Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot	Ballot
1st C	1st A	1st D	1st B	1st A	1st C	1st A	1st A	1st D
2nd B	2nd B	2nd C	2nd D	2nd B	2nd D	2nd B	2nd B	2nd C
3rd D	3rd C	3rd B	3rd C	3rd C	3rd B	3rd C	3rd C	3rd B
4th A	4th D	4th A	4th A	4th D	4th A	4th D	4th D	4th A

Ballot
1st A
2nd B
3rd C
4th D

[1] If c is the length of the side opposite the right angle in a right triangle, and if a and b are the lengths of the other two sides of the triangle, then $a^2 + b^2 = c^2$. Your high school teachers probably did not prove this theorem, but on the Internet, the interested student can easily find hundreds of different arguments (or **proofs**) which demonstrate that the theorem is true for *all* right triangles. (Try "`proof pythagorean theorem`" on your favorite search engine.)

First, we organize the ballots.

There are only a handful of possible ways to fill out the ballot. If we stack identical ballots together, we see that there were 5 different ballots in the election.

Ballot	
1st	A
2nd	B
3rd	C
4th	D

14

Ballot	
1st	C
2nd	B
3rd	D
4th	A

10

Ballot	
1st	D
2nd	C
3rd	B
4th	A

8

Ballot	
1st	B
2nd	D
3rd	C
4th	A

4

Ballot	
1st	C
2nd	D
3rd	B
4th	A

1

The following table summarizes what we have done.

1st choice	A	C	D	B	C
2nd choice	B	B	C	D	D
3rd choice	C	D	B	C	B
4th choice	D	A	A	A	A
number of voters:	14	10	8	4	1

Such a table is called a **preference schedule** for the election. From now on, all elections will be given in the form of a preference schedule.

Before we decide who won this election, we define a few more vocabulary terms. A ballot in which the voters must rank their choices is called a **preference ballot**. A ballot in which ties are not allowed is called a **linear ballot**. (So the Math Appreciation Society ballots are *linear preference ballots*.) We will only consider linear preference ballots in this Chapter.

If a voter prefers A over B, and also prefers B over C, it naturally follows that this voter prefers A over C. This fact is called the **transitivity of voter preferences**. This seemingly trivial fact will be used throughout the Chapter. Whenever a candidate is eliminated from a preference ballot, we simply move up all the lower ranked candidates to fill the gap (p. 5, Fig. 1-4).

2. Plurality method—Majority Criterion—Condorcet Criterion—Insincere voting

One way to decide the result of the Math Appreciation Society election is to declare that the candidate with the greatest number of 1st place votes is the winner. We call this the **plurality method** for deciding the election. We say that the candidate who receives the most 1st place votes has received the **plurality**.

Discarding all 2nd, 3rd, and 4th place votes, we see that Alisha wins:

1st choice	A	C	D	B	C
number of voters:	14	10	8	4	1

Alisha: 14 first place votes
Boris: 4 first place votes
Carmen: 11 first place votes
Dave: 8 first place votes

Is there anything unfair about awarding the election to Alisha? Her detractors might notice that $10 + 8 + 4 + 1 = 23$ voters chose someone else as their first choice. Significantly more voters opposed Alisha than supported her!

If a candidate receives at least half of the 1st place votes, we call him or her the **majority candidate**. (Notice that a majority candidate always has the plurality of votes, i.e.[2] more than anyone else.)

[2] The abbreviation *i.e.* means "that is." The abbreviation *e.g.* means "for example." You will not be tested on these abbreviations of the Latin phrases *id est* and *exempli gratia*.

If there are only 2 candidates, and we use the plurality method to decide who won, the winner will necessarily be the majority candidate. But when there are 3 or more candidates, the plurality may be less than half, as has happened several times in U.S. presidential elections.

candidate in 1824 election	**percent of electoral votes won**
Andrew Jackson	37.9%
John Quincy Adams	32.2%
William Crawford	15.7%
Henry Clay	14.2%

A fundamental principle of a fair democratic election is that, *if* there is a majority candidate, then that candidate should be the winner.

Majority Criterion

If Candidate X has a majority of the first-place votes, then Candidate X should be the winner of the election.

This is the first of the four **Fairness Criteria** we shall study in this Chapter. In every election that is decided using the plurality method, the Majority Criterion holds true.

We say that a voting method **satisfies** a Fairness Criterion if the Criterion holds true for every election decided by that method. Clearly, the plurality method satisfies the Majority Criterion.

We say that a method **violates** a Fairness Criterion if it is possible to have an election in which the Criterion is false. Be careful with this definition!!! It does not mean that the Criterion *must* be false. It only says that it's *possible* to have an election in which it is false.

Explain in your words what it means for a voting method to violate the Majority Criterion.

Ex.: Marching band (1).
The marching band at Tasmania State University has been invited to perform at five different bowl games: the Rose Bowl, the Hula Bowl, the Fiesta Bowl, and the Orange Bowl. The following preference schedule shows the results of an election held among the 100 members of the band.

number of voters:	49	48	3
1st choice	R	H	F
2nd choice	H	S	H
3rd choice	F	O	S
4th choice	O	F	O
5th choice	S	R	R

Who wins by the plurality method? How many voters did *not* choose the winner as their first choice? Using common sense alone, who *ought* to be the winner?

Compare the winner to each of the other candidates in the following table in a *head-to-head* comparison, counting only 1st place votes.

H	R	51–49
H	F	
H	O	
H	S	100–0

A candidate that is preferred over each of the other candidates in a head-to-head race is called the **Condorcet candidate**. Note that not every election has a Condorcet candidate.

In 1785, the Marquis of Condorcet proposed the following Fairness Criterion.

Condorcet Criterion

If Candidate X is preferred over each other candidate in a head-to-head race, then Candidate X should be the winner of the election.

We may now summarize some of the advantages and disadvantages of the plurality method.

Pros:
- It is simple to decide who won
- Satisfies the Majority Criterion

Cons:
- Does not consider voter's 2nd, 3rd, etc., choices
- Results may easily be manipulated by insincere voters
- the Condorcet Criterion

Let's look at the second "con" listed above. Have you ever considered voting for a third-party candidate? If the candidate had little to no chance of winning, you might have wondered whether you would be "wasting your vote." Some voters will vote *against* their favorite candidate, and instead vote for a candidate who is more likely to win. We call this **insincere**

voting.[3] In a close election, relatively few insincere voters can dramatically change the results.

Ex.: Marching band (2).
Consider the marching band example. Three of the band members, the Dorsey triplets, prefer the Fiesta Bowl, but they realize that there is no chance that the Fiesta Bowl will win. Reasoning that they would be "wasting their vote," the Dorsey triplets cast their votes for the Hula Bowl instead.

Real preferences

number of voters:	49	48	3 Dorseys
1st choice	R	H	F
2nd choice	H	S	H
3rd choice	F	O	S
4th choice	O	F	O
5th choice	S	R	R

Votes cast

number of voters:	49	48	3 Dorseys
1st choice	R	H	$\mathbf{H}$
2nd choice	H	S	$\mathbf{F}$
3rd choice	F	O	S
4th choice	O	F	O
5th choice	S	R	R

The 3 Dorseys have completely changed the outcome of the election: now the Rose Bowl loses, and the Hula Bowl wins.

We see that, if the plurality method is used to decide the winner, voters are pressured to vote for one of only two candidates, since voters may conclude that, "All votes for anyone other than the second place are votes for the winner."[4] **Duverger's Law** states that the plurality method necessarily leads to a two-party system, given enough time. As you may recall, the following two elections were characterized by some as unfair insofar as a third-party candidate became a "spoiler."

1992 presidential election

candidate	**percent of popular vote**
Bill Clinton	43.0%
George H. W. Bush	37.4%
Ross Perot	18.9%

2000 presidential election

candidate	**percent of popular vote**
George W. Bush	47.9%
Al Gore	48.4%
Ralph Nader	2.7%
Pat Buchanan	0.4%

Can you devise a method for deciding the winner of an election which does <u>not</u> pressure voters to vote for one of the two leading candidates?

[3] A.k.a. **strategic** or **tactical voting**.
[4] Quote taken from Wikipedia, "Plurality voting system," `http://en.wikipedia.org/wiki/First-past-the-post`.

3. The Borda count method

Ex.: Sportswriters' polls
Sportswriters are regularly asked to rank teams in most major sports. Suppose a poll is conducted in which 7 writers are asked to rank 3 college basketball teams from best to worst, and the preference schedule is:

rank	**order**		
1st (best)	C	A	A
2nd	A	C	B
3rd (worst)	B	B	C
number of voters:	4	1	2

A: Gonzaga College, B: Clemson College, C: The Citadel

How should we decide the results of this poll?

We can assign each rank from 1st to 3rd a certain number of points, as follows.

1st	3 pts.
2nd	2 pts.
3rd	1 pts.

Notice that if there are N candidates or choices, then 1st place is worth N points, 2nd is worth $N-1$ points, etc., and last place is always worth 1 point.

The number of points will have to be multiplied by the number of ballots in each stack of identical ballots. Use a table to organize your work as you total up the points earned by each team, as shown.

1st (best)	$C : 4 \times 3 = 12$	$A : 1 \times 3 = 3$	$A : 2 \times 3 = 6$
2nd	$A : 4 \times 2 = 8$	$C : 1 \times 2 = 2$	$B : 2 \times 2 = 4$
3rd (worst)	$B : 4 \times 1 = 4$	$B : 1 \times 1 = 1$	$C : 2 \times 1 = 2$
number of voters:	4	1	2

A: Gonzaga College, B: Clemson College, C: The Citadel

Now tally up the points for each team:

$$A \text{ gets } 8+3+6 = 17 \text{ points.}$$
$$B \text{ gets } 4+1+4 = 9 \text{ points.}$$
$$C \text{ gets } 12+2+2 = 16 \text{ points.}$$

We call this method the **Borda count method**. The candidate or choice with the most points is called the **Borda winner**.

Now try one on your own!

Ex.: A school principal selection goes awry (p. 11, Ex. 1.6)
The four finalists for the job of school principal at Washington Elementary school are Mrs. Amaro, Mr. Burr, Mr. Castro, and Mrs. Dunbar. Each of the 11 school board members gets to rank the candidates, and the Borda winner gets the job. The preference schedule for this election is reproduced here:

rank	**order**		
1st choice	A:	B:	C:
2nd choice	B:	C:	D:
3rd choice	C:	D:	B:
4th choice	D:	A:	A:
number of voters:	6	2	3

(a.) Who gets the job?

(b.) Who is the majority candidate?

(c.) Who is the Condorcet candidate?

We see that the Borda count method

(c.) satisfies / violates the Majority Criterion.

(d.) satisfies / violates the Condorcet Criterion.

4. Plurality-with-elimination method

Some municipalities require that a candidate obtain a majority of the first-place votes to be elected. When there are three or more candidates, quite often there is no majority candidate.

A **run-off election** is typically held at this point: the last place candidate is eliminated from the ballot, and a new election is held.

The **plurality-with-elimination method** (a.k.a. **instant runoff voting**, the **Hare method**) is a more efficient way to implement the same process. This method has become somewhat of a trend in recent years.

Voters fill out a preference ballot so that they do not need to vote over and over. From the original preference schedule, we eliminate the candidates with the fewest first-place votes one at a time until one of them gets a majority. (How do we know this must eventually happen?)

Plurality-with-elimination method

- **Round 1**.
  Count the first-place votes for each candidate. If a candidate has a majority of first-place votes, then that candidate is the winner. Otherwise, eliminate the candidate (or candidates if there is a tie) with the fewest last-place votes.
- **Round 2**.
  Cross out the names of any candidates eliminated from the preference schedule, and recount the first-place votes. If a candidate has a majority of first-place votes, then that candidate is the winner. Otherwise, eliminate the candidate (or candidates if there is a tie) with the fewest last-place votes.
- **Round 3**.
  Repeat Round 2 until a winner is found.

Ex.: Homecoming Queen election (p. 34, #28)
Find the winner of the election under the plurality-with-elimination method.

rank	order									
1st choice	A	A	A	B	B	B	C	C	D	D
2nd choice	C	B	D	D	C	C	A	B	A	B
3rd choice	B	D	C	A	D	A	D	A	C	C
4th choice	D	C	B	C	A	D	B	D	B	A
number of voters:	153	102	55	202	108	20	110	160	175	155

The following example shows that the plurality-with-elimination method has serious, but quite subtle problems.

Ex.: A mess of Olympic proportions

The cities of Athens, Barcelona, and Calgary are competing to be the host city for the 2020 Olympics. A secret vote of the 29 members of the Executive Council of the International Olympic Committee is to be held.

Two days before the actual election, a *straw poll*[5] is held. Here is the preference schedule for the straw poll.

1st choice	A	B	C	A
2nd choice	B	C	A	C
3rd choice	C	A	B	B
number of voters:	7	8	10	4

Who wins the straw poll by the method of plurality-with-elimination?

When word gets out that Calgary is favored to win the election, the four delegates represented by the rightmost column of the straw poll's preference schedule decide to switch their votes and vote for Calgary first. Here is the preferences schedule for the actual election.

1st choice	A	B	C
2nd choice	B	C	A
3rd choice	C	A	B
number of voters:	7	8	14

Who wins the actual election by the method of plurality-with-elimination?

[5] A **straw poll** is an unofficial vote or poll indicating the trend of opinion on a candidate or issue.

Monotonicity Criterion
If Candidate X is the winner of an election and, in a re-election, the only changes in the ballot are changes that favor X and only X, then Candidate X should be the winner of the re-election.

The plurality-with-elimination method *violates* both the Monotonicity Criterion and the Condorcet Criterion.

5. The method of pairwise comparisons—Counting pairwise comparisons

So far all the voting methods we have seen violate the Condorcet Criterion. The next method we will study is the classic example of one that does not violate it.

The method of pairwise comparisons is like a round-robin tournament in which each candidate is matched head-to-head against each other candidate. Each head-to-head match is called a **pairwise comparison**.

For each pairwise comparison that Candidate X wins, Candidate X receives 1 point. If there is a tie, each candidate receives $1/2$ point. The winner by the **method of pairwise comparisons** (a.k.a. **Copeland's method**) is the candidate who receives the most points.

This method obviously satisfies the Condorcet Criterion. (Why?)

It also satisfies both the Majority Criterion and the Monotonicity Criterion.

Unfortunately, the method of pairwise comparisons does violate a fourth Fairness Criterion, which we will soon introduce. First, we look at an example.

Ex.: The NFL draft (p. 19, Ex. 1.12)
A new expansion team, the Los Angeles LAXers, has been added to the NFL, and hence gets the opportunity to choose first in the upcoming draft. The coaches and team executives narrow the list of candidates to five players: Allen, Byers, Castillo, Dixon, and Evans. By team rules, the choice must be made by the method of pairwise comparisons. Here is the preference schedule after the 22 voters (coaches, scouts, and executives) turn in their preference ballots.

1st choice	A	B	B	C	C	D	E
2nd choice	D	A	A	B	D	A	C
3rd choice	C	C	D	A	A	E	D
4th choice	B	D	E	D	B	C	B
5th choice	E	E	C	E	E	B	A
number of voters:	2	6	4	1	1	4	4

Who is the newest member of the team?

Now suppose Castillo is eliminated from the original preference schedule. We have the following new preference schedule.

1st choice	A	B	B	B	D	D	E
2nd choice	D	A	A	A	A	A	D
3rd choice	B	D	D	D	B	E	B
4th choice	E	E	E	E	E	B	A
number of voters:	2	6	4	1	1	4	4

Now who wins?

Independence-of-Irrelevant-Alternatives (IIA) Criterion

If Candidate X is the winner of an election
and, in a re-count, one of the non-winning candidates is removed from the ballots,
then Candidate X should be the winner of the re-election.

A practical problem with the method of pairwise comparisons is that, as the number of candidates grows, the number of pairwise comparisons explodes. For 5 candidates we have 10 pairwise comparisons. For 10 candidates we have 45 pairwise comparisons. For 100 candidates we have $4,950$ pairwise comparisons. Computing all these comparisons would not be fun.

We want to be able to *count* how many pairwise comparisons there are, given the number of candidates. We start by developing a mathematical formula that at first seems unrelated.

Ex.: Long sums made short (p. 22, Ex. 1.14)
What is the sum of the first 49 counting numbers?[6]

[6] The **counting numbers** (a.k.a. **natural numbers**) are the positive whole numbers: namely, the numbers $1, 2, 3, 4, \ldots$.

Do you believe that the order doesn't matter when we add a list of numbers? For example, $1+2+3=3+2+1$. Convince yourself that a list of numbers can be added in any order whatsoever, and the resulting sum is always the same.

To sum up the first 49 counting numbers, we will use a clever trick. Let S stand for the sum: that is,

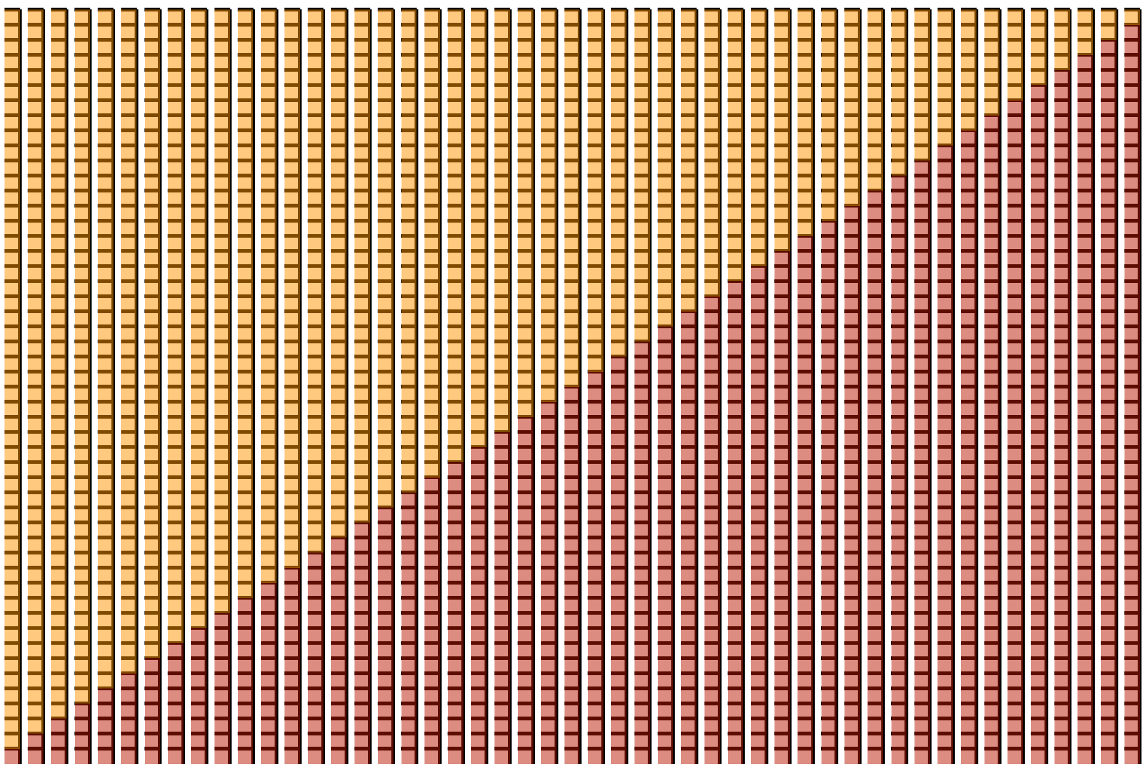
$$S = 1+2+3+\cdots+48+49.$$

Then

$$\begin{aligned} 2\times S = S+S &= (1+2+3+\cdots+48+49)+(49+48+\cdots+3+2+1) \\ &= (1+49)+(2+48)+(3+47)+\cdots+(48+2)+(49+1), \end{aligned}$$

since order doesn't matter when we add a list of numbers.

Now, it is obvious that each of the 49 sums of two numbers $(1+49)$, $(2+48)$, $(3+47)$, $\ldots$, $(49+1)$ is equal to 50. (A picture can make what is "obvious" a little easier to see! What does each vertical bar represent? What does the shading of the bars represent?)



The number of squares in this rectangle is

$$2S = 49\times 50,$$

where we have gotten the right hand side 49×50 by calculating the rectangle's area (width $\times$ height).

But what we want is S, so we divide this number by 2 to get what we're looking for:

$$S = \frac{49\times 50}{2} = \boxed{1225}.$$

In general,

The sum of the first L counting numbers is
$$1+2+3+\cdots+L = \frac{L\times(L+1)}{2}.$$

Now, how many pairwise comparisons are involved when there is an election with L candidates? To fix ideas, we will look at a numerical example.

Ex.: Counting pairwise comparisons. (p. 22, Ex. 1.15)
Consider an election with 10 candidates, A, B, C, D, E, F, G, H, I, and J.

Let's count all the pairwise comparisons.

- Compare A against the 9 candidates B, C, D, E, F, G, H, I, J.
- Compare B against the 8 candidates C, D, E, E, F, G, H, I, J.

  $\vdots$
- Compare H against the 2 candidates I, J.
- Compare I against the 1 candidate J.

There are a total of $\dfrac{9 \times (9+1)}{2} = \boxed{45}$ pairwise comparisons in an election with 10 candidates.

(Do you see why we used $L = 9$ and not $L = 10$?)

The number of pairwise comparisons
in an election between N candidates is $1 + 2 + 3 + \cdots + (N-1) = \dfrac{(N-1) \times N}{2}$.

6. Rankings—Recursive ranking

Each of the methods we have seen thus far can be extended to decide not only the winner, but 2nd place, 3rd place, and so on. We call the assignment of 2nd place, 3rd place, etc., a **ranking** of the candidates.

Recall that, for the *plurality method*, 1st place is the candidate with the most first-place votes For the *extended plurality method*, 2nd place is the candidate with the second most first-place votes, 3rd place is the candidate with the third most first-place votes, and so on.

For the *extended Borda count* and for *extended pairwise comparisons*, 2nd place is the candidate with the second most points, and so on.

For the *extended plurality-with-elimination method*, the candidates are ranked in the reverse order of elimination.

Another way to extend the methods we have seen is by repeatedly applying the method and removing the winner of each round. We call this **recursive ranking**.

For example, if A, B, C, and D compete in an election decided by Borda count, and A receives the most points, then A is ranked 1st place, and we remove A from the list of candidates, obtaining a new preference schedule. We then hold a second round using the new preference schedule. The winner of this second round is awarded 2nd place and removed, and the process repeats until every candidate has been ranked.

HW emendation: *SKIP problems #41, 45, 49 in Ch. 3.*

Ch. 3. Fair division

1. Fair division games

The essential parts of any "fair division game":

- S: the *set* of things, collectively called the **booty** (or **goods**, or **loot**), to be divided
- $P_1, P_2, P_3, \ldots$: the players who share the booty
- a **value system** for each player that determines how much any part of S is worth to that player personally

Assumptions about the players:

- **Rationality**: Each player seeks to maximize his or her share of S, and pursues this goal guided by reason alone.
- **Cooperation**: All players agree to play by the rules of whichever division game is chosen.
- **Privacy**: No player has any useful information about the other players' value systems.
- **Symmetry**: Players have equal rights in sharing S: that is, every player is entitled to a proportional share of S.

Suppose there are N players amongst which to divide the booty. The purpose of a fair division game is to divide S into N shares and assign a share to each player in such a way that every payer gets a "fair share." So what constitutes a fair share?

A share s of the booty S is called a **(proportional) fair share to player** P if s is worth at least $\frac{1}{N}$ of the total value of S *in the opinion of* P.

Ex.
Suppose that there are 4 players, and that the booty S consists of a 1972 Fender Stratocaster electric guitar together with an amplifier, a Steinway grand piano in poor condition, a 1983 Buick sedan together with an assortment of air fresheners, and a baseball card collection that includes a Mark McGwire card from his rookie year. Further suppose that to Paul, one of the four players, each of these items is worth a certain percentage of the booty's total value, as follows:

s_1:	guitar, amplifier	40%
s_2:	piano	10%
s_3:	sedan, air fresheners	20%
s_4:	baseball card collection	30%

Note that these the values of these four shares in *Paul's* mind: the sedan may be less valuable than the piano to someone else, but possibly Paul only knows how to play guitar, and thus has little use for a piano in rough shape.

Which of the four shares s_1, s_2, s_3, s_4 is a fair share to Paul?

A **fair division method** is a set of rules that define how a game is played, where the game ends once each player has been assigned a share of the booty S.

Different fair division methods are used depending on what type of booty we have to divide among the players. In particular, depending on the nature of the set S, a fair division method is called **discrete**, **continuous**, or **mixed**.

- In a *discrete* fair division game, the booty S is made up of indivisible objects. E.g. a piano, a car, a diamond ring, a piece of candy.
- In a *continuous* fair division game, the booty S can be divided in infinitely many different ways, and a share can be made larger or smaller by any any small amount. E.g. a birthday cake, a pizza, the space in a storage unit, a plot of land.
- In a *mixed* fair division game, the booty S is made up of both discrete and continuous components.

Note that we make the standing assumption that pieces of candy are "indivisible" (although in theory we could cut a piece of candy into as small pieces as we like). The same goes for pieces of jewelry (although these could in theory be melted down).

2. Two players: The divider-chooser method

The **divider-chooser** fair division method can be used when there are two players, and S is continuous. If we use a cake as a metaphor for the booty, we can summarize the rules of this game as follows: *You cut, I choose*. In detail, Player One, called the **divider**, divides the cake in two. Then Player Two, the **chooser**, chooses one of the two pieces.

Ex.
See p. 88, Ex. 3.1: "Damian and Chloe divide a cheesecake."

It is always better to be the chooser than the divider, because the chooser *always* has an opportunity to choose a piece worth *more than* one-half of the total.

On the other hand, this method is fair, because the divider is guaranteed a piece worth exactly one-half of the total. (Why?)

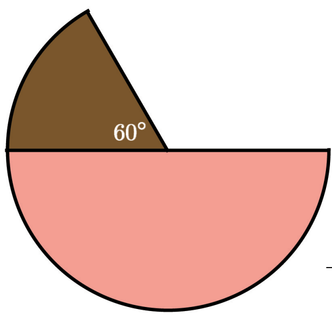
#14a (p. 114).
Raul and Karli are planning to divide the chocolate-strawberry mousse cake in the figure using the divider-chooser method. Raul values chocolate three times as much as he values strawberry. Karli values chocolate twice as much as she values strawberry.

If Raul is the divider, which of the following cuts are consistent with Raul's value system?

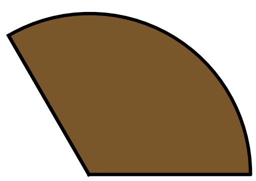
Question:
Is **cut (iii)** consistent with Raul's value system?

Solution:

piece	**comparative worth**



Chocolate:	$\frac{60^\circ}{360^\circ}$	$\times$	3	$=$	$\frac{180^\circ}{360^\circ}$	$=$	$\frac{1}{2}$
Strawberry:	$\frac{180^\circ}{360^\circ}$	$\times$	1	$=$	$\frac{180^\circ}{360^\circ}$	$=$	$\frac{1}{2}$
TOTAL WORTH:							**1**



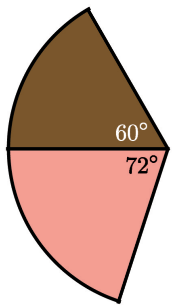
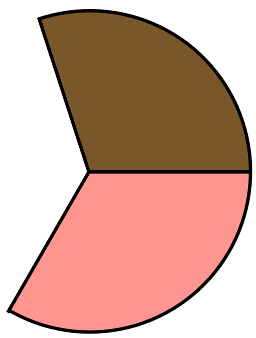
Chocolate:	$\frac{120^\circ}{360^\circ}$	$\times$	3	$=$	$\frac{360^\circ}{360^\circ}$	$=$	1
TOTAL WORTH:							**1**

Answer:
$\boxed{\text{Yes.}}$ The two pieces have the same comparative worth, according to Raul's value system.

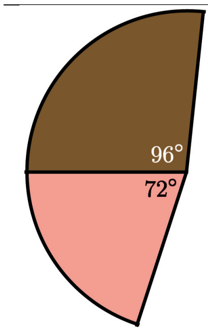
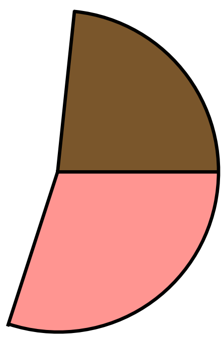
Notice the meaning of the columns in the chart below:

	proportion of booty		**value multiplier**				
Chocolate:	$\frac{60^\circ}{360^\circ}$	$\times$	3	$=$	$\frac{180^\circ}{360^\circ}$	$=$	$\frac{1}{2}$
Strawberry:	$\frac{180^\circ}{360^\circ}$	$\times$	1	$=$	$\frac{180^\circ}{360^\circ}$	$=$	$\frac{1}{2}$
TOTAL WORTH:							**1**

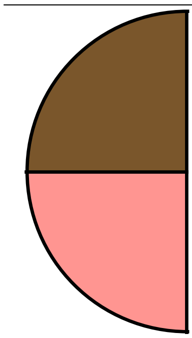
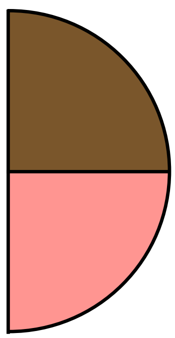
Question:
Is **cut (iv)** consistent with Raul's value system?

piece	**comparative worth**
	
	

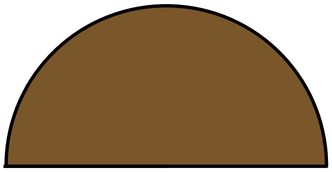
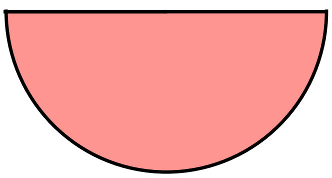
Question:
Is **cut (v)** consistent with Raul's value system?

piece	**comparative worth**
	
	

Question:
Is **cut (i)** consistent with Raul's value system?

piece	**comparative worth**
	
	

Question:
Is **cut (ii)** consistent with Raul's value system?

piece	**comparative worth**
	
	

3. Lone divider method

The divider-chooser method does not make sense if there are 3 players. However, in this case, the method can be *extended*, as the Polish mathematician Hugo Steinhaus discovered. Steinhaus's idea was later extended to any number N of players. We call this extension for N players the **lone-divider method**. Let's look at the lone-divider method for exactly 3 players.

- Randomly choose one of the players to be the divider. The other players will be choosers.
- The divider cuts the cake S into three pieces. The divider will get one of these pieces, but does not know which one, and therefore cuts the cake into three shares that he thinks are equal in value.
- The first chooser, C_1, declares which shares are fair shares to her (perhaps by writing this information down on a slip of paper). We call this declaration C_1's **bid**. Since C_1 is not guaranteed that she will get one of these shares, it is in her interest to write down *all* shares she considers fair, not only the best share.
- The second chooser, C_2, does the same thing. Note that C_1 and C_2 do not know each others' bids.
- Separate out all the pieces of S that *do not appear on either* bid list. We will call these U-**pieces** (for $u$nwanted). Note that, for both choosers, a U-piece is valued at less than $1/3$ of the total value.
- What's left is all the pieces of S that appear on *at least one* of the two bid lists. We call these C-**pieces** (for $c$hosen). Note that there is always at least one C-piece.

Case 1: *There are two or more C-pieces.*

See the examples (3.2 and 3.3 in the book, pp. 91-92) to follow.

Case 2: *There is only one C-piece.*

If the two choosers agree that one of the U-pieces is least desirable, give it to the divider.

If the choosers do not agree, pick a U-piece randomly, and give it to the divider.

There are now two pieces left, a U-piece and a C-piece. Combine these into one single piece.

Use the divider-chooser method to divide this new single piece into two shares for the two choosers.

Ex.
p. 91, Ex. 3.2: "Lone-divider with 3 players: Case 1, Version 1."

	s_1	s_2	s_3
Dale	$33\frac{1}{3}\%$	$33\frac{1}{3}\%$	$33\frac{1}{3}\%$
Cindy	35%	10%	55%
Cher	40%	25%	35%

Ex.
p. 92, Ex. 3.3: "Lone-divider with 3 players: Case 1, Version 2."

	s_1	s_2	s_3
Dale	$33\frac{1}{3}\%$	$33\frac{1}{3}\%$	$33\frac{1}{3}\%$
Cindy	30%	40%	30%
Cher	60%	15%	25%

Ex.
p. 92, Ex. 3.4: "Lone-divider with 3 players: Case 2."

	s_1	s_2	s_3
Dale	$33\frac{1}{3}\%$	$33\frac{1}{3}\%$	$33\frac{1}{3}\%$
Cindy	20%	30%	50%
Cher	10%	20%	70%

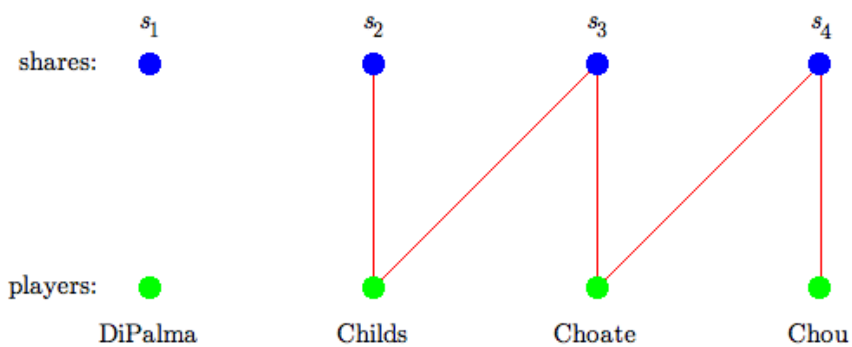
We will now look at some problems involving the lone-divider method which you will be expected to know how to solve. Many of our problems (involving the lone-divider method) fall into one of the following categories:

- Find a fair division (i.e. each player's fair share), given all N players' bid lists.
- Reconstruct the bid lists, given the values of each share to each player.
- Count the number of fair divisions, and explain how you figured out how many there are.

You may solve these problems in any way you like. We will demonstrate some effective ways to organize your work.

#22a (p. 116).
Four partners, DiPalma, Childs, Choate, and Chou are dividing a piece of land among themselves using the lone-divider method. Using a map, the divider DiPalma divides the property into three pieces s_1, s_2, s_3, s_4. When the choosers' bid lists are opened, Child's bid list is $\{s_2, s_3\}$, Choate's bid list is $\{s_3, s_4\}$, and Chou's bid list is $\{s_4\}$. Describe a fair division of the land.

Solution:



Answer:

#27a (p. 116).

Six players D, C_1, C_2, C_3, C_4, C_5 are dividing a cake among themselves using the lone-divider method. D cuts the cake into 6 slices s_1, s_2, s_3, s_4, s_5, s_6. When the choosers' bid lists are opened, C_1's bid list is $\{s_2, s_3, s_5\}$, C_2's bid list is $\{s_1, s_5, s_6\}$, C_3's bid list is $\{s_3, s_5, s_6\}$, C_4's bid list is $\{s_2, s_3\}$, and C_5's bid list is $\{s_3\}$. Describe a fair division of the cake.

Solution:

	s_1	s_2	s_3	s_4	s_5	s_6
C_1		○	○		⊗	
C_2	⊗				○	○
C_3			○		○	⊗
C_4		⊗	○			
C_5			⊗			
D	○	○	○	⊗	○	○

#29b (p. 116).

Four partners, Egan, Fine, Gong and Hart are dividing a piece of land valued at $\$480,000$ among themselves using the lone-divider method. Using a map, the divider divides the property into four pieces s_1, s_2, s_3, s_4. The following table shows the value of each of the piece in each partner's eyes, but some entries in the table are missing.

	s_1	s_2	s_3	s_4
E	80	85		195
F		100	135	120
G	120		120	
H	95	100		110

Numbers in thousands of dollars.

Describe the choosers' respective bid lists.

4. Lone-chooser method

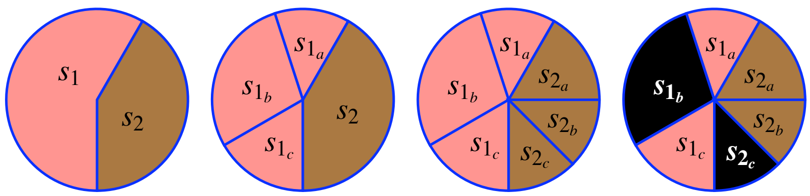
The lone-chooser method is another way to extend the divider-chooser method to three or more players. We first look at the case where the number of players is $N = 3$.

Lone-chooser method for $N = 3$ players.

- **(Setup.)** Randomly choose one of the players to be the *chooser*. The other players will be dividers.
- **(Division.)** Using the divider-chooser method, the $N - 1 = 2$ dividers

$$D_1 \text{ and } D_2$$

  divide the booty S among themselves into two fair shares. Let's say that D_1 ends up with s_1, and that D_2 gets s_2.
- **(Subdivision.)** Each divider splits his or her share into $N = 3$ subshares. Let's call the three subshares of s_1 by the names s_{1_a}, s_{1_b}, s_{1_c}.
- **(Selection.)** The chooser C now takes one of D_1's subshares, and one of D_2 subshares. These $N-1 = 2$ subshares make up C's final share. D_1 keeps the remaining subshares from s_1, and D_2 keeps the remaining subshares from s_2.



Notice that we used a *different* fair division method (namely, the divider-chooser method) to fairly divide the booty among the $N - 1 = 2$ dividers.

When $N = 4$, we can use the *lone-chooser method for* 3 *players* to fairly divide the booty among the $N - 1 = 3$ dividers.

In general, for any number $N \geq 4$ of players, if we know how to fairly divide the booty among $N - 1$ dividers *by any method whatsoever*, we can apply the following method to fairly divide the booty among N players.

Lone-chooser method for $N \geq 4$ players.

- **(Division.)** Using any method whatsoever, the $N - 1$ dividers

$$D_1, \ D_2, \ D_3, \ \ldots, \ D_{N-1}$$

  divide the booty S among themselves into $N - 1$ fair shares. By definition of a fair share, each player ends up having claimed a share they consider to be worth $\frac{1}{N-1}$ of the total value of S.
- **(Subdivision.)** Each divider splits his or her share into N subshares.
- **(Selection.)** The chooser C now takes one subshare from each divider's share. These $N - 1$ subshares make up C's final share. D_1 keeps the remaining $N - 1$ subshares from s_1, D_2 keeps the remaining $N - 1$ subshares from s_2, and so on.

Activity.
"More cake cutting," Instructor's Resource Manual, pp. 18-21.

Ch. **3**. *Fair division*, §§3.1–3.7 <u>except</u> 3.5.

A.	Shares, fair shares, fair divisions	1, 3, 5, 9
B.	Divider-chooser method	11, 13, 15
C.	Lone-divider method	19, 23, 29
D.	Lone-chooser method	31, 35, 39
E.	Last-diminisher method	41, 45, 49
F.	Method of sealed bids	51, 55, 57
G.	Method of markers	59, 61, 67
JOG.	Method of sealed bids	78

OMIT #41, 45, 49

6. Method of sealed bids

The **method of sealed bids** is a discrete fair division. It includes five steps.

1. (**Bidding.**) Each player makes a bid in dollars for each of the items in the booty, giving her honest opinion of the worth relative to her value system.
2. (**Allocation.**) Each item goes to the individual with the highest bid. If there is a tie, then some means such as a coin flip is used to randomly determine who obtains the share. Note that it is possible that one player may end up with most or all of the shares according to this method.
3. (**First Settlement.**) We determine whether each player owes or is owed money by the estate. To determine this we first calculate each player's *fair-dollar share* of the estate.

   A player's **fair-dollar share** is computed by adding that player's bids, and dividing the total by the number of players N. If the fair-dollar share is less than the fair share of the dollar worth of the estate, then the player is owed the difference. Otherwise, the player owes the difference (in cash) to the estate.

   The process of deciding how much each player owes or is owed is called **settling up**.
4. (**Division of Surplus.**) If there is a surplus of cash after settling up, then the amount is divided evenly among the players.
5. (**Final Settlement.**) The final settlement is determined by adding the surplus found in step 4 and the amount allocated in step 3.

The method of sealed bids works well if the following to conditions are met.

- Each player must have enough money to play the game. If a player has no means of obtaining cash or credit then they are at a disadvantage.
- Secondly, each player must accept money as a substitute for any particular item. This means that no player can value items as priceless.

Ex.

Suppose that 5 siblings (Daniel, Bob, Jeni, Karla, Ashlie, Tung) wish to divide fairly and equally their parents estate. Their parents leave them 4 items: a baseball card collection, a Model T Ford, the house, and a dog named Bubbles. We will use the method of sealed bids to fairly and evenly divide the estate among the 5 kids, even though we only have 4 items!

1. (**Bidding.**) Each player makes a bid in dollars for each of the items in the booty, giving her honest opinion of the worth relative to her value system.
2. (**Allocation.**) Each item goes to the individual with the highest bid.

	Daniel	Bob	Jeni	Karla	Tung
Card Collection	$200,000$	$180,000$	$10,000$	$100,000$	$105,250$
Model T	$10,000$	$300,000$	$20,000$	$162,000$	$105,250$
House	$100,000$	$125,000$	$300,000$	$133,000$	$105,250$
Bubbles	$290,000$	$296,000$	$210,000$	0	$105,250$
TOTAL	$600,000$	$601,000$	$540,000$	$395,000$	$421,000$
Fair Dollar Share	$120,000$	$120,200$	$108,000$	$79,000$	$84,200$

Numbers in thousands of dollars.

3. (**First Settlement.**) We determine whether each player owes or is owed money by the estate. To determine this we first calculate each player's *fair-dollar share* of the estate. A player's **fair-dollar share** is computed by adding that player's bids, and dividing the total by the number of players N. If the fair-dollar share is less than the fair share of the dollar worth of the estate, then the player is owed the difference. Otherwise, the player owes the difference (in cash) to the estate.

4. (**Division of Surplus**) If there is a surplus, then the amount is divided evenly among the players.

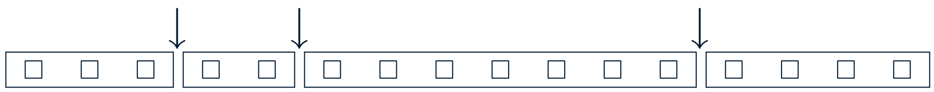
5. (**Final Settlement**) The final settlement is determined by adding the surplus found in step 4 and the amount allocated in step 3.

7. Method of sealed bids

The Method of Markers is a discrete fair division. The k items in the booty will be lined up in a random order (see p. 106). We call this lineup of items an **array** of items.

We'll say that the leftmost item is in Position 1, and that the rightmost position is in Position k, and similarly for the Positions 2, 3, 4, $\ldots$, $k-1$ in between.

Each of the N players places $N-1$ markers that represent their idea of a fair share. The $N-1$ markers divide the array into N portions (or **segments**).



3 markers partition an array into 4 segments.

Each person's bids are still considered secret! (For example, we might ask each player to write down their bids on a list, as we did in the lone-divider method.)

This method ensures that each player will end up with at least one of their bid segments, which means everyone obtains a fair share of the booty.

The method of sealed bids is carried out as follows.

1. **(Bidding.)** Each player writes down where she wants her $N-1$ markers.
2. **(Allocation.)** Scan the array from left to right. Stop when the first marker comes up. The player who placed that marker receives the first segment in her bid (i.e. all items to the left of her marker).

   Continue this process, scanning from left to right, stopping when a player's marker is reached, and giving the leftmost segment to that player.

   If two players have placed a marker in the same position, flip a coin or use some other means of randomly deciding who gets the leftmost segment.
3. **(Division of Leftovers.)** Any leftover pieces are distributed randomly, e.g. by drawing straws.[1]

What would you say are the advantages of this method?

What are some disadvantages?

[1] Or stones, etc.: we call this ancient method drawing lots, and it appears in the Christians' New Testament: "And they parted his clothing, and cast lots." To be precise, **drawing lots** means that each of N players blindly chooses one object out of a collection of $N-1$ like objects and 1 unlike object, e.g. stones (one of which is a different color) from a bag, or drinking straws (one of which is short).

Ch. 5. Graph theory

1. Sets and relations

Definition 1.

A **set**[1] is a collection of objects of any sort: people, numbers, books, outcomes of experiments, geometrical figures, etc. Thus we can speak of the set of all integers, or the set of all oceans, or the set of all possible sums when two dice are rolled and the number of dots on the uppermost faces are added, or the set consisting of the residents of the city of Washington, D.C.

A set must be **well-defined**, by which we mean that, for any object whatsoever, the question, "Does this object belong to the collection?" has an unambiguous "yes" or "no" answer. It is not necessary that we personally have the knowledge required to decide which answer is correct. We must know only that, of the answers "yes" and "no," exactly one is correct.

Let us also agree that no object in a set is counted twice. That is, the objects are **distinct**. It follows that, when listing the objects in a set, we do not repeat an object after it is once recorded. For example, the set of letters in the word "banana" is not a set containing six letters, but rather the three *distinct* letters b, a, and n.

One way to write a set is to list all its objects (which are called its **members** or **elements**) one by one; this way of writing a set is called **the roster method**. We use curly braces $\{$ and $\}$ to indicate that we are talking about the *set* of things in the list. So, for example, the symbols

$$\{1, 2, 3\}$$

mean, "The set consisting of 1, 2, and 3," and the symbols

$$\{\text{Tom}, \text{Jill}, \text{Janet}\}$$

mean, "The set consisting of Tom, Jill, and Janet."

[1] This paragraph and the next two are adapted from Samuel Goldberg, Probability: An introduction (Dover: 1986). Note that our "definition" of a set is is not logically satisfactory, as the author emphasizes. The idea of a "set" is so deep in the foundations of mathematics that to define it in terms of more basic concepts would be beyond the scope of this course.

Two sets can be multiplied. The result is called the *Cartesian product*.

Definition 2. The **Cartesian product** of two sets A and B is the set of all pairs with the first element in A and the second element in B.

The Cartesian product can be visualized as a grid, with the members of A as the columns, and the members of B as the rows.

Example.
Define $A = \{$yellow, red$\}$.
Define $B = \{$car, bus, truck$\}$.

	yellow	red
car	yellow car	red car
bus	yellow bus	red bus
truck	yellow truck	red truck

So

$$A \times B = \{\text{yellow car}, \text{red car}, \text{yellow bus}, \text{red bus}, \text{yellow truck}, \text{red truck}\}.$$

Question.
If two sets A and B each contain only finitely many elements, how can we find the number of elements in the set $A \times B$?

Example.
Define

$$A = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}.$$

Find $A \times A$.

When we multiply a set A by itself, the result is the set of all pairs of members in A. This is the most important definition in this section:

Definition 3. A **relation** on a set A is a subset of the set $A \times A$. In other words, a relation is a set of pairs of members of A. We say that a and b are **related** if the pair ab is in the relation.

For our purposes, the order in a pair does not matter. For example, ♂♀ and ♀♂ will be considered the same pair, and the same goes for ⚀⚅ and ⚅⚀.

Example.
Let A be the set $\{\text{Austin}, \text{Dallas}, \text{El Paso}, \text{Houston}, \text{San Antonio}\}$ of cities whose airports are served by Southeast Airlines. Let R be the set of pairs of cities joined by a straight line on the following map; suppose these lines represent which cities are connected by direct flights on Southeast Airlines.

Texas Map - ©2006 Destination360

Then

$$R = \{\text{Dallas-Houston}, \text{Houston-San Antonio}, \text{San Antonio-Austin}, \text{San Antonio-El Paso}, \text{San Antonio-Dallas}\}$$

is an example of a relation on the set A.

2. Graphs

Definition 4. A **simple graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is two sets:

- a set $\mathcal{V}$ of objects called **vertices**, and
- a relation $\mathcal{E}$ on $\mathcal{V}$.

The set $\mathcal{V}$ may be any set whatsoever. We call $\mathcal{V}$ the **vertex set** of the graph. The members of $\mathcal{V}$ are the **vertices**, and the pairwise relationships are the **edges**: *X is related to Y if and only if XY is an edge*. In this case, we call X and Y the **endpoints** of the edge XY. We call $\mathcal{E}$ the **edge set**.

Notice that, according to this definition of a graph, edges may not be repeated.

To give a graph, we often use a picture, called a **presentation** or **rendering of the graph**. For example, the dots and lines drawn on the above map of Texas gives a graph $\hat{\mathcal{G}}$ whose vertex set is the set of all cities labeled on the map, and whose edge set is the set R. In this example, "*X is connected to Y by a direct flight if and only if XY is an edge*."

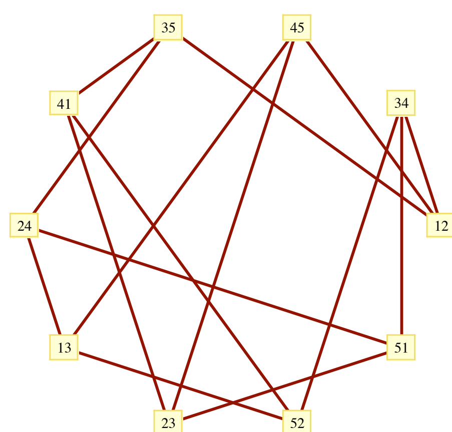
Definition 5. A vertex which is not the endpoint of any edge is called an **isolated vertex**. (Give an example of an isolated vertex in $\hat{\mathcal{G}}$.)

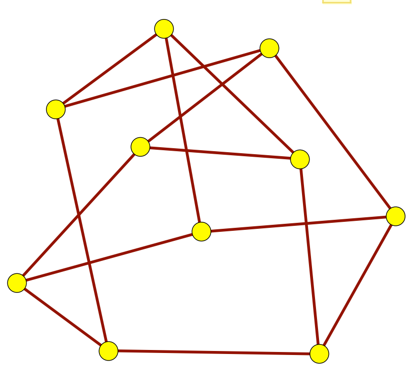
However, a graph should not be confused with its presentation. Different pictures can represent the same graph.

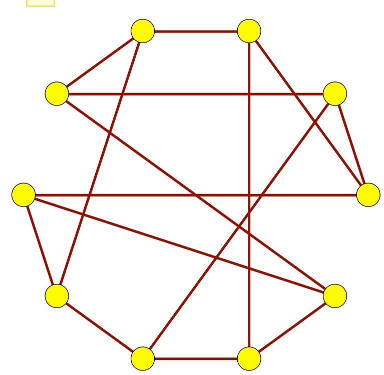
Example.
Let $\mathcal{V} = \{1, 2, 3, 4\}$. Let $\mathcal{E} = \{12, 23, 34\}$. Draw two different presentations of $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

Example.
Three different presentations of the *Petersen graph* are shown below. This graph can be defined as follows:

> Let $\mathcal{V}$ be the set of pairs of distinct numbers $1, 2, 3, 4, 5$ (for example, 12, but not 11 or 22). Let $\mathcal{E}$ be the relation of pairs of pairs that have no numbers in common (for example, the vertex 12 is related to the vertex 34, but 23 is not related either to 24 or to 13). Then the **Petersen graph** is the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.







Exercise.
How many different graphs exist having vertex set A, B, C?

How many different graphs exist having vertex set A, B?

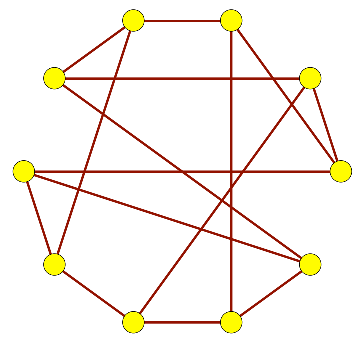
How many different graphs exist having vertex set A?

Problem.
How many different graphs exist having vertex set A, B, C, D, E, F, G?

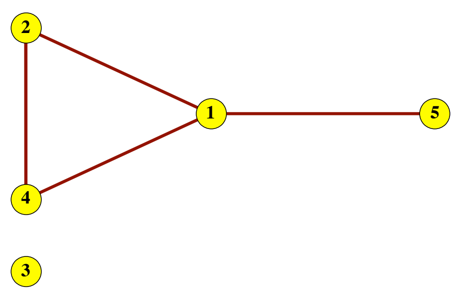
3. Degree

Definition 6. The **degree of a vertex** is the number of edges having that vertex as an endpoint, where loops are counted twice. If a vertex has odd degree, we call it an **odd vertex**. Otherwise, we call it an **even vertex**.

Example.
In the Petersen graph, each vertex has degree 3.



Example.
Find the degree of each vertex in the following graph.



Theorem 1: The Handshaking Theorem for Simple Graphs.
Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph. The sum of the degrees of the vertices in $\mathcal{V}$ is twice the number of edges in $\mathcal{E}$.

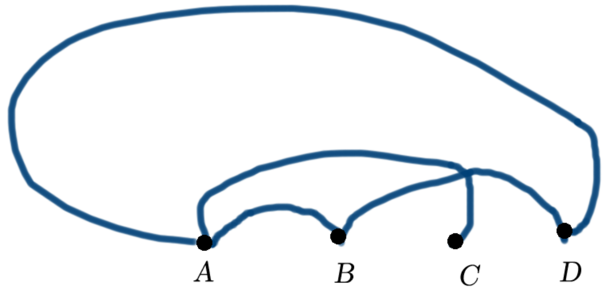
A **loop** is an edge both of whose endpoints are the same. Provide an argument that explains why this theorem is true for any simple loopless graph.

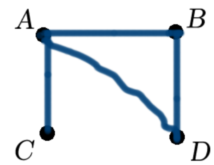
If we allow graphs to have *loops*—and we do—-does this change the theorem? What if we allow multiple edges between the same two vertices?

Corollary.
A simple graph has an even number of vertices of odd degree.

Why does this follow from the Handshaking Theorem? Does it matter whether we allow loops?

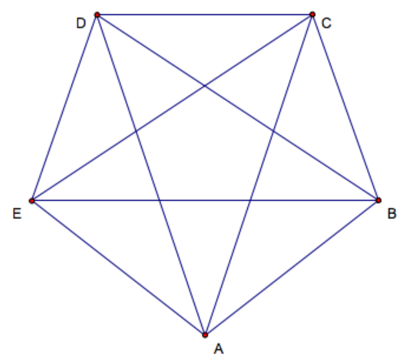
Ex. 1. Are the following two presentations the same graph?





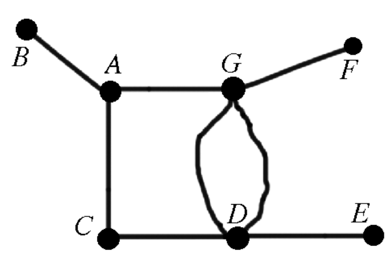
Ex. 2.

(a.) Give the vertex set $\mathcal{V}$ and the edge set $\mathcal{E}$ of the graph $G = (\mathcal{V}, \mathcal{E})$ rendered below.



(b.) Give the vertex set $\mathcal{V}$ and the edge set $\mathcal{E}$ of the graph $G = (\mathcal{V}, \mathcal{E})$ rendered below.

(c.) Give the vertex set $\mathcal{V}$ and the edge set $\mathcal{E}$ of the graph $G = (\mathcal{V}, \mathcal{E})$ rendered below.



Ex. 3. For each of the following descriptions of a graph, either draw a presentation of such a graph, or explain why no such graph is possible. (*Hint*: All but one are possible.)

(a.) A connected graph $G = (\mathcal{V}, \mathcal{E})$ with 8 vertices such that each vertex has degree 3.

(b.) A disconnected graph $G = (\mathcal{V}, \mathcal{E})$ with 8 vertices such that each vertex has degree 3.

(c.) A graph $G = (\mathcal{V}, \mathcal{E})$ with 8 vertices such that each vertex has degree 1.

(d.) A graph $G = (\mathcal{V}, \mathcal{E})$ with 15 vertices such that each vertex has degree 5.

4. Adjacency

Definition 7. We say that two vertices X and Y are **adjacent vertices** if XY is in the edge set.

Exercise.
Draw a graph whose vertex set is $\mathcal{V} = \{1, 2, 3, 4\}$, such that each vertex is adjacent to every other. Then do the same thing with the vertex set $\mathcal{V}' = \{1, 2, 3\}$.

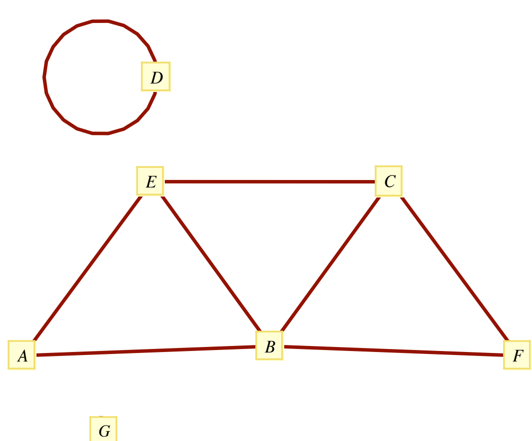
Definition 8. The **complete graph on n vertices** is the graph whose vertex set $\mathcal{V}$ consists of n elements, and whose edge set contains every pair of vertices in $\mathcal{V}$.

When a graph has many edges, it would be tedious to write out which vertices are adjacent to each other. A better way to organize this information is by using an *adjacency matrix*.

In our class, an **adjacency matrix** will be a table whose rows are labeled by the vertices in the graph, and whose rows are also labeled by the vertices in the graph. (If this reminds you of the Cartesian product, it is no coincidence.)

We place a 1 or 0 in each square: 1 if the vertices are adjacent, and 0 if not.

Exercise.
Find the adjacency matrix of the following graph (which is in "three pieces").



Exercise.
How many edges are there in a complete graph on n vertices?

Hint: Have we done anything like this before, in an earlier unit? Draw a few graphs for small values of n (say, 3, 4, 5 vertices) and see if you can write a general formula that predicts the number of edges for any value of n.

Problem.
How many graphs having a vertex set $\{A, B, C, D\}$ are possible?

Hint: Use your answer to the previous exercise to get started. Don't draw all the possible graphs, it's too tedious! (Don't write out all the possible adjacency matrices, either—but thinking about the adjacency matrix may help you get started.)

Resolution: How many simple graphs on n labeled vertices?

We have seen that every simple[2] graph can be represented by an adjacency matrix.

We also saw that many of the entries in the adjacency matrix can safely be discarded, because they are redundant. Since the adjacency matrix is symmetric, the only entries we need are those entries on the diagonal and above it. After throwing out the other entries (below and to the left of the diagonal), we can still reconstruct the graph.

The number of non-redundant entries in the adjacency matrix for a simple graph on n vertices was (we saw last class):

$$\frac{n \times (n+1)}{2}.$$

But this did not tell us how many graphs there are on n labeled vertices. It only told us how many entries we need to keep, when we look at a simple graph's adjacency matrix.

Today we will answer the question, "How many different simple graphs on n labeled vertices are possible?" by answering the question, "How many different adjacency matrices for a simple graph on n vertices are possible?"

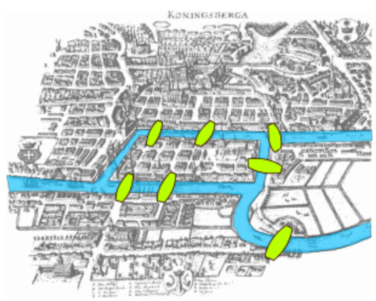
[2]Recall that a **simple** graph is a graph *without* multiple edges.

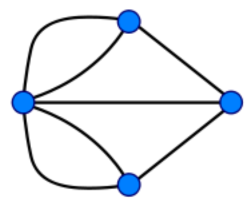
5. Paths and circuits

Definition 9. A **path**[3] is a sequence of vertices with the property that each vertex in the sequence is adjacent to the next. Alternately, a path can be thought of as the sequence of edges joining such a sequence of vertices. *Each edge must appear only once in a path.* The number of edges in a path is called the **length of the path**.

Definition 10. A **circuit** is a path which begins and ends at the same vertex.

Example.
The map on p. 160 shows the bridges of Königsberg, Germany as they appeared to the mathematician Leonhard Euler (pronounced "oiler") when he visited the city in the 1730s. We reproduce it here in miniature, at left. A simpler, if less picturesque, version appears at right.

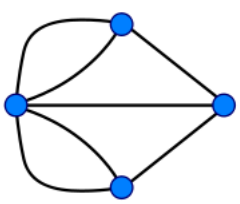
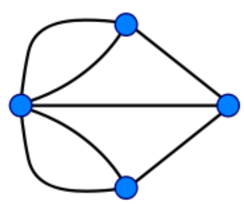
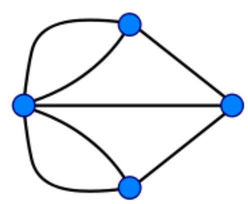




This is the most famous graph in all of graph theory, because it was the first. The first article about graph theory was published by Euler in 1736, and it was about this graph.

Lots of math—geometry, algebra, the theory of numbers—is old. Not graph theory. It wasn't considered "serious math" until as late as the 1960s. Today it is considered one of the most exciting branches of math, at least in part because it is so marketable. Graph theory can give answers not only to problems involving transportation networks (air traffic, highway planning) and communication networks (telephony, Internet, etc., not to mention the power grid), but also to problems of management and business (scheduling a large number of students in a large number of classes, increasing efficiency of workflow, designing committees).

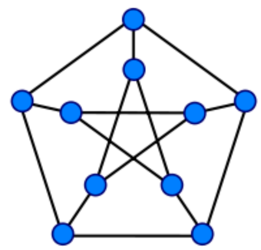
Exercise.
Find a path of length 5 in the graph of Konigsberg's bridges. Then find a path of length 6. Is there a circuit of length 6? Is there a path of length 7? Is there a path of length 8?

[3] Tannenbaum's definition of a "path" is not universally accepted. Other authors say a "simple path" (Rosen, Discrete Math and its Applications) or a "trail" (West, Introduction to Graph Theory.

Definition 11. A **unicursal drawing** is a tracing of a presentation without lifting the pencil or retracing any of the edges.

We will develop methods for determining whether a graph has a unicursal tracing, even when the graph is quite complicated. Does the Petersen graph have one? (Here is another presentation of the Petersen graph:)

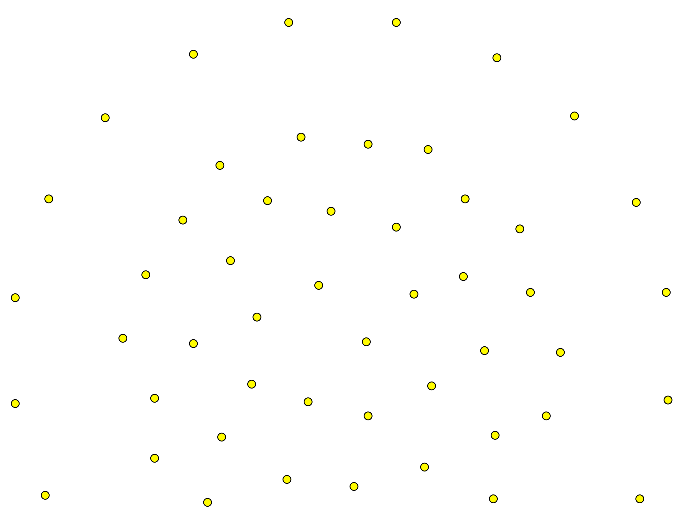


Definition 12. A graph is called **connected** if there is a path between every pair of distinct vertices in the graph.

Definition 13. An **Euler circuit** is a circuit that passes through every edge of the graph.

Theorem 2. A connected graph has an Euler circuit if and only if each of its vertices have even degree.

Exercise.
Can you convince yourself that, if a connected graph has an Euler circuit, then each of its vertices indeed must have even degree? Try sketching a few graphs using the swath of random dots provided.



A *routing problem* is concerned with finding ways to route the delivery of goods and/or services to an assortment of destinations.

For example (see p. 163 for a picture), if a security guard needs to patrol every block[4] in the neighborhood, he doesn't want to walk down the same block twice. If in addition he needs to pick up his car wherever he left it at the beginning of his patrol, we have the following routing problem:

- Is there a circuit that covers every block?
- If some blocks must be covered more than once, what is an *optimal* route that covers every block?

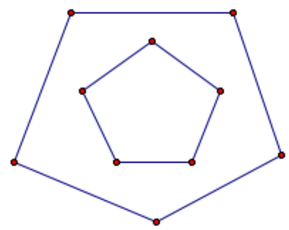
The former is an example of the **Existence Question** for a routing problem. The latter is an example of the **Optimization Question**. Theorem 2 completely answers the existence question for the security guard's routing problem. In the sections to follow, we will discuss methods to solve the optimization question.

[4]Note that Tannenbaum uses the word "block" to mean one side of a square block: that is, there are four blocks on each side of a square plot of land boxed in by streets.

6. Euler's theorems

Recall that an **Euler path** is a path that passes through every edge in the graph. An *Euler circuit* is an Euler path that begins and ends at the same vertex.

Look at the graphs on page 188. Which graphs have Euler paths? Which have Euler circuits? We can tell at a glance that some graphs certainly do <u>not</u> have an Euler path (or an Euler circuit):



—but for the more complicated graphs, it would be nice to have an easy <u>test</u> for whether or not there is an Euler path (or Euler circuit).

vocabulary term	*example*
Routing problem	*Find a walk in a given graph which crosses every edge, and which minimizes the number of edges crossed more than once.*
Existence question for a routing problem	*Does there exist an Euler circuit (i.e. no re-crossings) in the given graph?*
Optimization question for a routing problem	*What is the fewest number of re-crossings Euler circuit in the given graph?*
Negative solution	*There is no Euler circuit in the given graph.*
Constructive solution	*There is an Euler circuit in the given graph, and here it is!*
Non-constructive solution	*There is some Euler circuit in the given graph, somewhere.*

Theorem 3. Euler's circuit theorem.
A connected graph has an Euler circuit if every vertex is even.

What kind of solution to the routing problem of finding an Euler circuit does this theorem provide?

How can we use this theorem to establish that a given graph does <u>not</u> have an Euler circuit?
...So Euler's circuit theorem <u>also</u> provides a negative solution to the problem for certain graphs.

Theorem 3'. Euler's circuit theorem: Negative solution.
A connected graph has no Euler circuit if

A similar result can be used to establish the existence or non-existence of Euler paths, but there is one detail to be addressed.

Proposition.
If a vertex in an Euler path has degree 1, then the path either begins or ends at that vertex.

Why?

Theorem 4. Euler's path theorem.
A connected graph with exactly two odd vertices has at least one Euler path.
Moreover, every such path starts at one of the odd vertices and ends at the other.

We can now be certain that our failure to find an Euler path or an Euler circuit in the Bridges of Königsberg graph was due not to any lack of cleverness on our part. Euler's path theorem[5] is extraordinarily powerful—no matter *how* complicated the graph is, it provides an easily obtained solution. Using Euler's Path Theorem on the Bridges of Königsberg, which is a rather uncomplicated graph, is a little like using an atom bomb to swat a fly.

[5] In addition to the circuit theorem and path theorem, Euler also proved both the Handshaking Theorem and its corollary in his famous 1736 article, which can easily be found reprinted in English.

7. Fleury's algorithm

Euler's path[6] theorem solves the *existence question for the problem of finding an Euler path*. But it says nothing about how to actually *find* the Euler path: the theorem only says, *there is some Euler path, somewhere*.

For a small graph, trial-and-error is a pretty good way to find an Euler path. First, use Euler's theorems to establish whether there is such a path or not—you don't want to waste your time looking for something which the theorems say do not exist! Then start drawing paths. Better use a pencil!

However, when the graph is huge—as any real-world graph of practical significance is likely to be—it is out of the question to use trial-and-error. For example, the graph that represents all the power stations in the nation's electric grid contains thousands of vertices, joined by hundreds of thousands, perhaps millions of edges. Trial-and-error is simply not practically feasible, given that we have better things to do with our time than sketching a few quadrillion paths until we find one that works.

A little *piece* of another graph which one would not want to use trial-and-error on is shown below (you may have seen it at a certain sandwich shop).

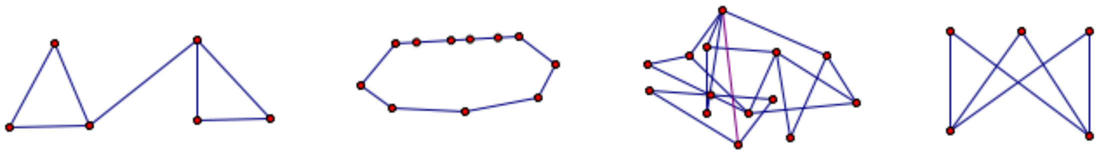
FIG. Subways of Lower Manhattan and Downtown Brooklyn

Notice that if one subway station is closed or disabled, we have a different graph (because the vertex set changes). If this station was a vertex of high degree, it is quite possible that whatever Euler paths we had before the station closure can no longer be traced. This implies that, every time a station is temporarily closed, the problem of finding an Euler path must be solved all over again from scratch.

[6] Throughout this paragraph, the word "path" may be replaced in all instances by the word "circuit."

We would like to have a systematic procedure to follow, which will tell us at the end whether there is or is not an Euler path. This is provided by Fleury's algorithm.

Definition 14. A **cut-edge** (or **bridge**) is an edge in a connected graph whose deletion yields a disconnected graph.



Which of the above graphs have cut-edges?

Fleury's Algorithm

- **Preliminaries.** Verify that Euler's theorems do in fact guarantee the existence of an Euler circuit (or path), otherwise we are wasting our time.
- **Setup.** Draw two copies of the graph side-by-side. Label the left graph FUTURE, and the right graph PAST.
- **Start.** Choose a starting vertex. If we are looking for an Euler circuit, we can start anywhere. If we are looking for an Euler path, we must start at an odd vertex.
- **Walk.** On the FUTURE graph, choose an edge to walk down, and erase it. If you have a choice, don't choose a cut-edge. However, if you have no other choice, take it.
- **Finish.** Repeat the previous step until you can't travel anymore. When you can't travel any more, the circuit (or path) is complete.

HW emendation: *SKIP problems #41, 43 in Ch.* **5**.

Ch. 6. The traveling salesman problem—REV 2

1. Hamilton paths and Hamilton circuits

Definition 15. A **Hamilton path** in a graph $G = (\mathcal{V}, \mathcal{E})$ is a path that visits every vertex in $\mathcal{V}$ exactly once.

Definition 16. A **Hamilton circuit** in a graph $G = (\mathcal{V}, \mathcal{E})$ is a path that visits every vertex in $\mathcal{V}$ exactly once, and then ends at the vertex at which the path began.

Ex. Suppose that a graph $G = (\mathcal{V}, \mathcal{E})$ has an Euler circuit. Which of the following statements is true?

(a.) G must have an Euler path.

(b.) G must *not* have an Euler path.

(c.) G may or may not have an Euler path.

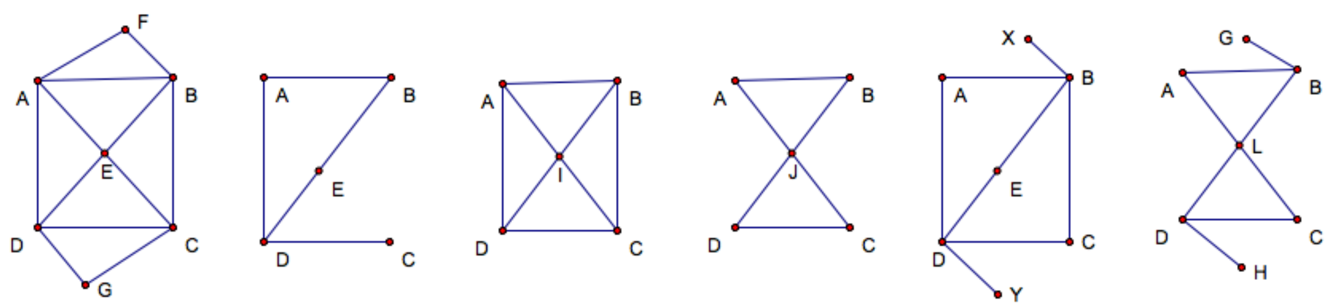
Ex. Suppose that a graph $G = (\mathcal{V}, \mathcal{E})$ has a Hamilton circuit. Which of the following statements is true?

(a.) G must have a Hamilton path.

(b.) G must *not* have a Hamilton path.

(c.) G may or may not have a Hamilton path.

Ex. Verify and extend your answers to the previous two exercises by determining which of the following graphs have Euler circuits, Euler paths, Hamilton circuits, and Hamilton paths.



graph →	1	2	3	4	5	6
Has an Euler path?						
Has an Euler circuit?						
Has a Hamilton path?						
Has a Hamilton circuit?						

We have easy tests for determining whether *any* given graph has an Euler circuit or an Euler path. It is natural to think that there must be such a test for Hamilton circuits and paths—but there isn't.

However, some success has been had in proving that any graph *satisfying a given property* must have a Hamilton circuit (or path). We call such a property a **sufficient condition (for the existence of a Hamilton circuit or path)**.

Theorem 5. Dirac's theorem.
A connected graph on N vertices such that *every vertex has degree at least* $\frac{N}{2}$ must have a Hamilton circuit.

Ex. *Prove* that K_N has a Hamilton circuit for every $N > 2$. (In this context, the word "*prove*" means: "*use known facts to make a bulletproof argument that K_N has a Hamilton circuit, no matter what N is.*")

Proof.
Since K_N is complete, every vertex in K_N has degree $N-1$. It is enough to prove that $N-1$ is greater than or equal to $\frac{N}{2}$, because if this is so, then it follows by Dirac's theorem that K_N is guaranteed to have a Hamilton circuit.

Working backwards, we find that the last inequality (1) written below follows from the (given) fact that $N > 2$.

$$N > 2. \tag{5}$$

$$(2N) - N > (N+2) - N. \tag{4}$$

$$2N > N + 2 \tag{3}$$

$$N > \frac{N}{2} + 1. \tag{2}$$

$$N - 1 > \frac{N}{2}. \tag{1}$$

2. The number of Hamilton circuits in a complete graph

Recall from Chapter 1 that the complete[1] graph K_N on N vertices has $\frac{N(N-1)}{2}$ edges. Of course, we didn't call it "the complete graph K_N" in Chapter 1: we called it an "arrow diagram," and we used it to keep track of all the pairwise comparisons between N candidates.

We can show that K_N has $\frac{N(N-1)}{2}$ edges more easily now, i.e. without recourse to pictures of shaded boxes and additional formulas, because we have developed the language of graph theory.

Ex. I am thinking of a positive number N. Without knowing what N is, say what the degree of a vertex in K_N is. Use your answer to determine the sum of all the degrees in K_N. Then use the Handshaking Theorem (Theorem 1) to find the number of edges in K_N.

Proof.
The degree of a vertex in K_N is $N-1$, because each vertex is adjacent to every *other* vertex by definition of a "*complete*" graph.
Since there are N vertices in K_N, again by definition of K_N (as "the complete graph *on N vertices*"), it follows that the sum the degrees of all the vertices in K_N is

$$N(N-1).$$

Let the letter e stand for the number of edges in K_N. By the Handshaking Theorem, the sum of the degrees of all the vertices in K_N is double the number e of edges in the graph, so

$$2e = N(N-1).$$

Therefore, the number of edges in G is half the sum of the degrees, which we have shown to be $N(N-1)$. That is,

$$e = \frac{N(N-1)}{2}.$$

[1] See your Chapter 5 notes for a discussion of complete graphs.

It is in fact quite obvious from looking at K_2, K_3, K_4, and K_5 that for small[2] values of $N \geq 3$, the complete graph K_N must have a Hamilton circuit: we can see it! In cases like these, when it is apparent that the graph has a Hamilton circuit, a more interesting question is: *How many different Hamilton circuits are there?*

We are all too lazy to count the number of Hamilton circuits in K_N when N is not tiny. The table on page 203 shows that a complete graph has over 3 million different Hamilton circuits when the graph has a measly 11 vertices.

```
<< Combinatorica`
Partition[Table[ShowGraph[CompleteGraph[n]], {n, 1, 20}], 5] // TableForm
```

Out[20]//TableForm=

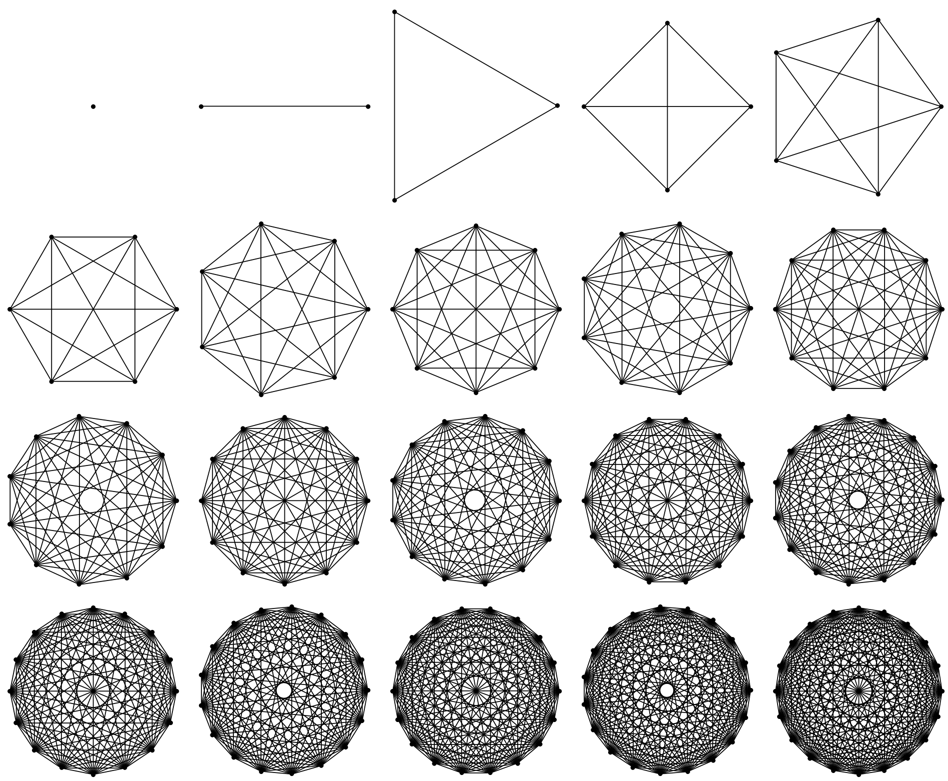


FIGURE 1. Complete graphs, drawn by Wolfram Research's `Mathematica` software (available on all ACC Learning Lab computers). The two lines of code at the top of the figure suffice to generate the picture shown.

[2] Of course the graph K_0 on 0 vertices has no Hamilton circuit. Is there any problem finding a Hamilton circuit in K_1? What goes wrong in K_2?

Even when $N = 6$, counting all the Hamilton circuits would be brutally tedious.

Out[97]=

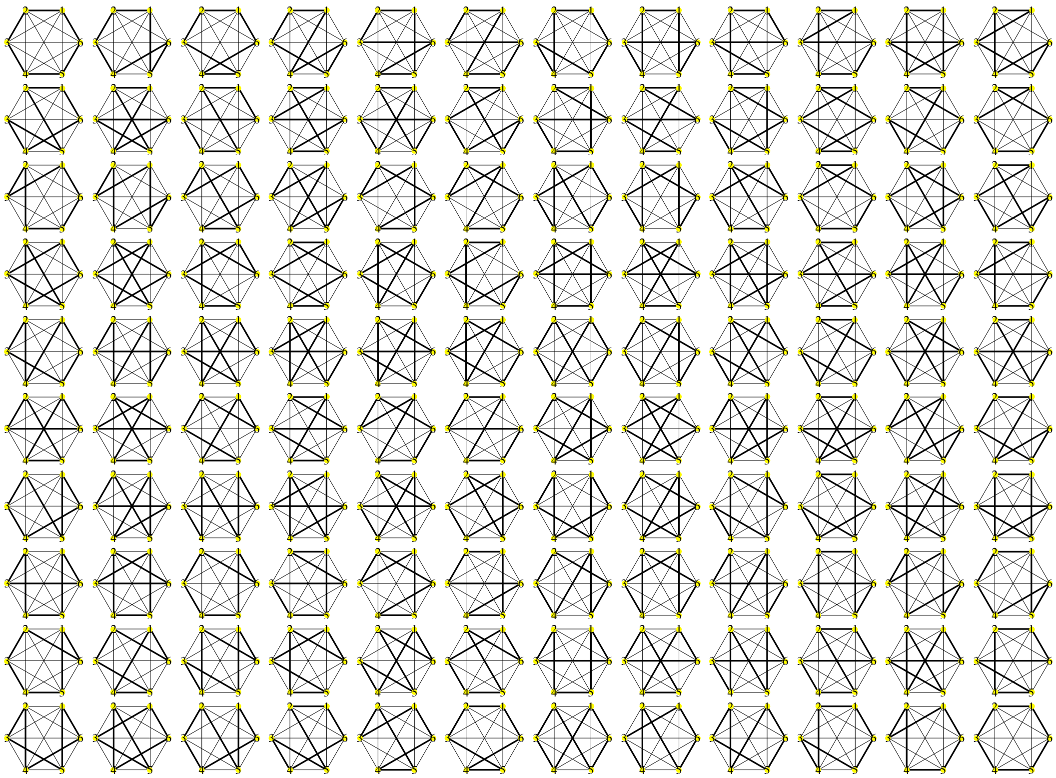


FIGURE 2. `Mathematica` again. Code (much uglier this time) appears in the footnote.[3]

Now, in the previous table of Hamilton circuits (FIGURE 2), we see that the first and last circuit appear to be identical. The same thing happens when we ask the computer to find all the Hamilton circuits in K_5:

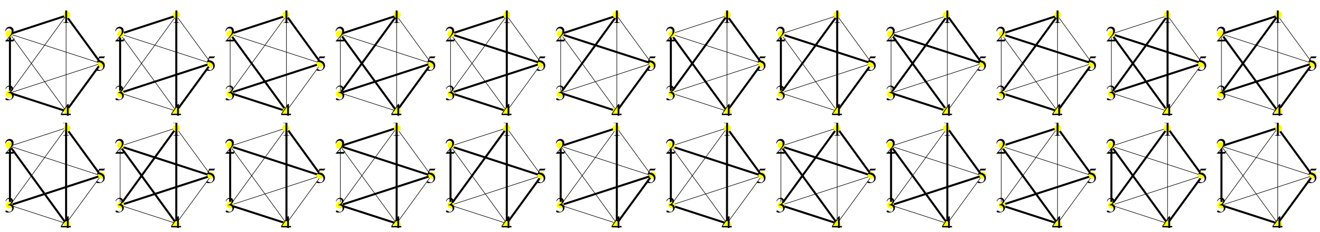


FIGURE 3. Hamilton circuits in K_5.

What is going on here? Why does the computer appear to be counting the same circuit twice?

[3] `ShowAllHamiltonianCycles[NN_, columnwidth_] := Module[H = HamiltonianCycle[CompleteGraph[NN], All], ShowGraphArray[Partition[Table[Highlight[SetGraphOptions[CompleteGraph[NN], VertexColor→Yellow, VertexStyle→Disk[Large]], Partition[H[[k]], 2, 1]], k, 1, (NN - 1)!], columnwidth], VertexNumber→True, VertexNumberPosition→Center]]; ShowAllHamiltonianCycles[6, 12]`

The reason is that the circuits $1, 2, 3, 4, 5, 1$ and $1, 5, 4, 3, 2, 1$ are *different* circuits, in so far as the edges come in a different order—

$$\begin{array}{rl} \textit{leftmost circuit in Fig. 3:} & 1, 12, 23, 34, 45, 51, 1 \\ \textit{rightmost circuit in Fig. 3:} & 1, 15, 54, 43, 32, 21, 1 \end{array}$$

—but the two circuits "*look*" the same if we ignore the order and write both paths as a set[4] of edges:

$$\{12, 23, 34, 45, 51\} = \{15, 54, 43, 32, 21\}.$$

That is, the left hand side and the right hand side are exactly the same set.

We will stipulate that two circuits that consist of the same edges, traversed in a different order, are to be considered two different circuits.

By contrast, we say that it doesn't matter where you start a Hamilton circuit. That is, using K_4 to visualize the situation,

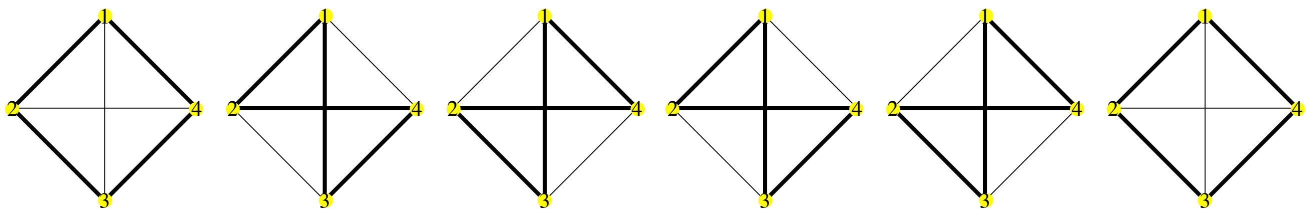


Figure 4. Hamilton circuits in K_4.

we will say that the circuits

$$\begin{array}{c} 1, 12, 23, 34, 41, 1 \\ 2, 23, 34, 41, 12, 2 \\ 3, 34, 41, 12, 23, 3 \\ 4, 41, 12, 23, 34, 4 \end{array}$$

are all considered to be the same circuit.

We will call the vertex at which a circuit begins and ends the **reference point**, as the book does. But observe that our notation for a circuit is slightly different than Tannenbaum's! We specify the *edges*, where Tannenbaum gives only the *vertices* in the circuit.

When we are counting Hamilton circuits, we must be sure to specify a reference point. We then list each Hamilton circuit just once—because, for example, once we have counted (let's say) $1, 12, 23, 34, 41, 1$, we are guaranteed by our use of 1 as a reference point not to count any of the (identical) circuits $2, 23, 34, 41, 12, 2$, etc.

Be consistent about your reference point. All the circuits you count must begin and end at the *same* reference point, which we choose *once and for all* (that is, for the duration of such a counting problem).

[4] Recall that the order in which members are listed does not matter in a set, and that the order in which endpoints X and Y are written in an edge $XY = YX$ does not matter in a graph whose adjacency matrix is symmetric.

Ex. How many distinct Hamilton circuits are there in K_N?

Hint: Pick a reference point. How many choices are there for the second point in a Hamilton circuit starting at the reference point you picked? Now, for each of those choices of the second point in a Hamilton circuit, how many choices are left for the third point? So how many ways are there to pick the first 2 vertices in a Hamilton circuit, given a reference point fixed once and for all? Once you have answered the latter question, you should be ready to extend your reasoning to picking all N vertices.

Proof.
There are N ways to choose the reference point, $N-1$ ways to choose the second point, $N-2$ to choose the third, and so on. Thus there are

$$N \times (N-1) \times (N-2) \times \cdots \times 3 \times 2 \times 1$$

Hamilton circuits altogether.
However, these are not distinct. In fact, each individual Hamilton circuit has been counted N times, since it does not matter which vertex we write as the reference point.
Therefore, there are only

$$\big[N \times (N-1) \times (N-2) \times \cdots \times 3 \times 2 \times 1\big] \div N = (N-1) \times (N-2) \times \cdots \times 3 \times 2 \times 1$$

<u>distinct</u> Hamilton circuits.

The **factorial** $k!$ of a natural[5] number k is the number

$$k! = k \times (k-1) \times (k-2) \times \cdots \times 3 \times 2 \times 1.$$

Your calculator[6] will have an $\boxed{x!}$ or $\boxed{n!}$ key which you can use to compute factorials.

Ex. How many distinct Hamilton circuits are there in K_N?

Ex. Compute.

(a.) $15!$

(b.) $30!$

(c.) $\dfrac{30!}{15!}$

Ex. Given that $10! = 3,628,000$, find $9!$ without a calculator.

[5]A **natural number** is a positive integer: that is, one of the numbers 1, 2, 3, 4, 5, $\ldots$.
[6]See syllabus for calculator requirements.

3. Traveling salesman problems

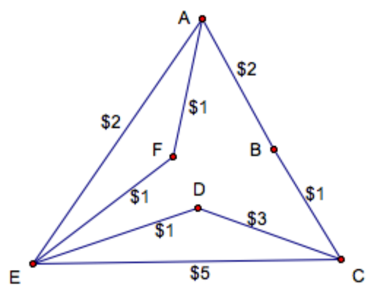
Examples 6.5 and 6.6 (pages 204-205) suggest that the metaphor of a traveling salesman can be applied to problems faced by NASA engineers. We may see one or two problems that involve a salesman, but **the traveling salesman problem** in general means that the following optimization question is to be answered:

Is there a least expensive (in time, money, etc.) circuit which visits every vertex in a given (complete) graph?

In this chapter, we will often be looking at complete graphs, so the *existence question* has already been answered (see above). If we find a Hamilton circuit which is cheaper than any other Hamilton circuit, we call it an **optimal route** (or **solution**)

Definition 17. A **weighted graph** is a graph each of whose edges is labeled with a number, called the **weight** of that edge. If $G = (\mathcal{V}, \mathcal{E})$ is a weighted graph, and XY is an edge in $\mathcal{E}$, we write $w(XY)$ for the weight of XY.

Definition 18. The **(total) weight of a walk** in a weighted graph is the total of the weights of every edge in the walk.

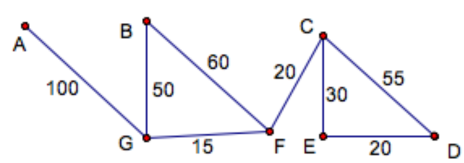


For example, for the above graph, we write

$$w(A, AE, EC, CA, AE, E) = w(AE) + w(EC) + w(CB) + w(BC) = \$9.$$

Since paths and circuits are walks, we can compute their weights, too.

Ex. List all the AD-walks, and compute the total weight of each. Which one has the lowest weight?

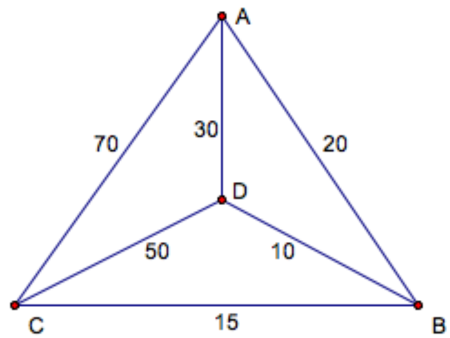


$A, AG, GB, BF, FC, CE, ED, D$	$100 + 50 + 60 + 20 + 30 + 20 = 280$
A, AG, GF, FC, CE, ED, D	$100 + 15 + 20 + 30 + 20 = 185$
A, AG, GB, BF, FC, CD, D	$100 + 50 + 60 + 20 + 55 = 285$
A, AG, GF, FC, CD, D	$100 + 15 + 20 + 55 = 190$

The walk with the lowest weight is $\boxed{A, AG, GF, FC, CE, ED, D}$.

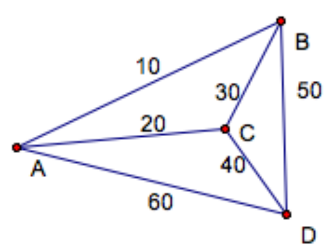
STRATEGIES FOR FINDING A HAMILTON CIRCUIT IN A WEIGHTED GRAPH

Strategy 1: Exhaustive search. List all possible Hamilton circuits. Calculate the total weight of each circuit. The optimal route is the circuit with the least total weight.

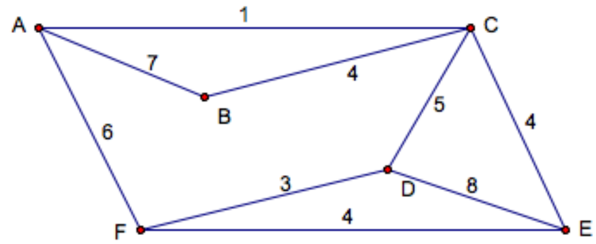


We call two circuits *mirror image circuits* if they contain the same vertices in exactly the reverse order. For example, $ABCDE$ and $AEDCB$ are mirror images.

Ex. Use a *tree diagram* to list all the Hamilton circuits in the following graph. Which Hamilton circuits are mirror image circuits? Compute the total weight of each of the Hamilton circuits you found.

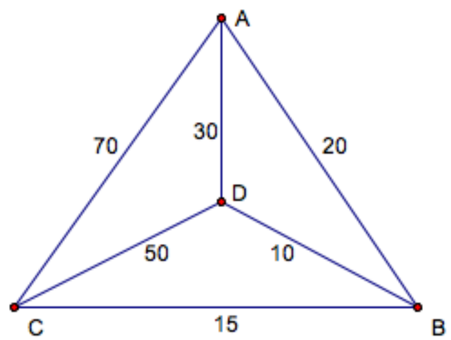


Ex. Find the optimal Hamilton circuit by exhaustive search. Start by drawing a tree diagram that shows all the Hamilton circuits in the following graph.

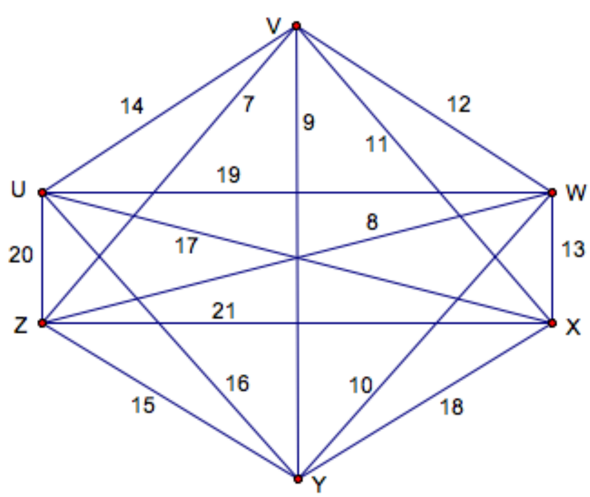


Strategy 2: Cheapest neighbor. (This method only works for complete graphs.) Pick a reference point. From there, travel to a neighboring vertex along the least expensive available edge. From each new vertex, travel along the least expensive available edge among those leading to a neighboring vertex which has not yet been visited. When every vertex has been visited, return to the reference point.

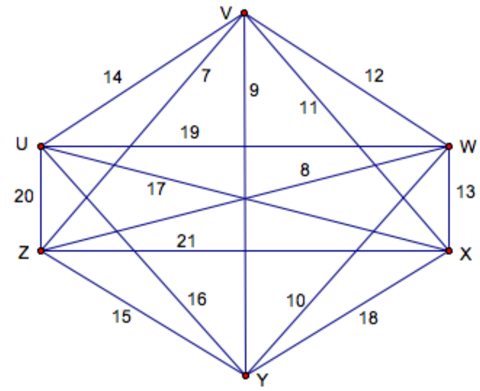
Strategy 2 is <u>not</u> guaranteed to discover the optimal route—but it is typically much faster than exhaustive search.



Ex. Find the cheapest-neighbor circuit for starting vertex A.



Ex. Find the cheapest-neighbor circuit for starting vertex U. Find the cheapest-neighbor circuit for starting vertex V, for starting vertex W, for starting vertex X, Y, and Z. Which of the six cheapest-neighbor circuits is cheapest?



ALGORITHMS FOR FINDING A HAMILTON CIRCUIT IN A WEIGHTED GRAPH

Brute-force algorithm

1. Make a list

2. Calculate

3. Report

Nearest-neighbor algorithm

1. Designate a starting vertex (or use the given starting vertex).

2. From the starting vertex,

3. For each successive vertex,

4. Repeat until

5. From the last vertex, travel directly to the starting vertex.

Repetitive nearest-neighbor algorithm

1. Designate a starting vertex (or use the given starting vertex).

2. Carry out the nearest-neighbor algorithm for each

3. Re-write each nearest-neighbor circuit as a circuit that starts and ends at the designated starting vertex.

An algorithm can be *optimal* and/or *efficient*.

An algorithm is called **optimal** if it always finds an optimal solution when implemented correctly.

An algorithm is called **efficient** if the amount of computational effort to implement the algorithm grows in some reasonable proportion when the size of the input to the problem increases.

The brute force algorithm is optimal, but terribly inefficient. (See p. 211.) What about the nearest-neighbor algorithm? What about the repetitive nearest-neighbor algorithm?

Ch. 7. Networks

1. Trees

Definition 19. A **tree** is a connected graph with no circuits.

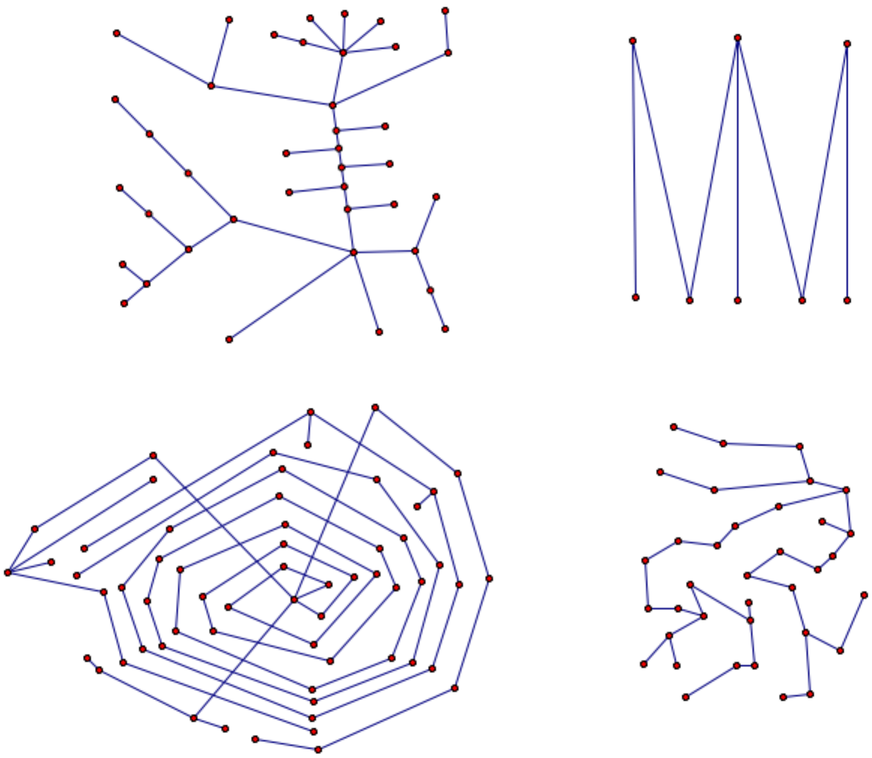
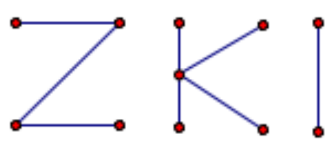


FIGURE. Four trees.

Ex. Each of the letters of the Roman alphabet (A, B, C, . . . , Z) determines a graph, once we add vertices wherever two or more (possibly curved) line segments meet, and wherever a line segment ends.

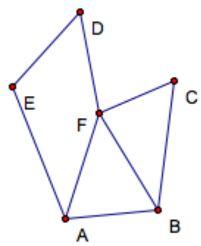


Which capital letters in the Roman alphabet determine trees (give at least 3 examples)? Which do *not* determine trees (give at least 3 examples)?

2. Spanning subgraphs

Definition 20. A **network** is a connected graph.

Definition 21. A **subgraph** of a graph $G = (\mathcal{V}, \mathcal{E})$ is a graph $G' = (\mathcal{V}', \mathcal{E}')$ such that $\mathcal{E}'$ is a subset of $\mathcal{E}$. That is, we call G' a subgraph of G if every edge of G' is an edge of G. We say that G' **spans** G (or that G' is a **spanning subgraph**) if every vertex of G is a vertex of G': that is, if $\mathcal{V}$ is a subset of $\mathcal{V}'$.



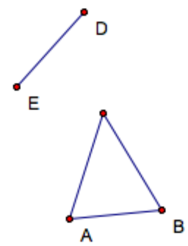


FIGURE. A graph G (at left) and a subgraph G' (at right) which does not span G.

Practical problems involving networks often involve finding a spanning subgraph G' of an existing network G.

Ex. 7-1. The following graph represents the road network between 7 mining towns deep in the heart of the Amazon jungle. A telephone company needs to provide Internet service to the 7 towns, and the only practical option is to bury fiber-optic cable along the already existing roads connecting the towns. In the graph below, the cost of putting a fiber-optic cable along each road is given as a weight in millions of dollars.

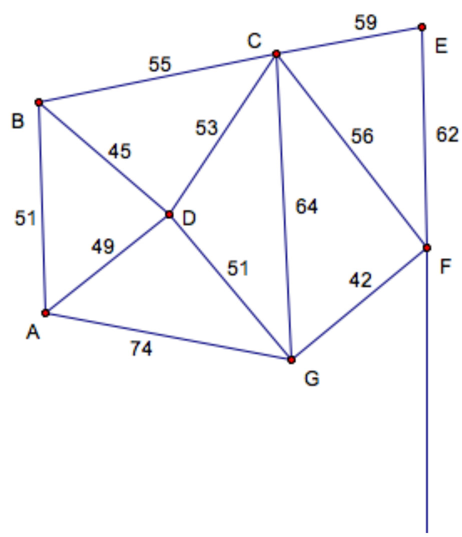


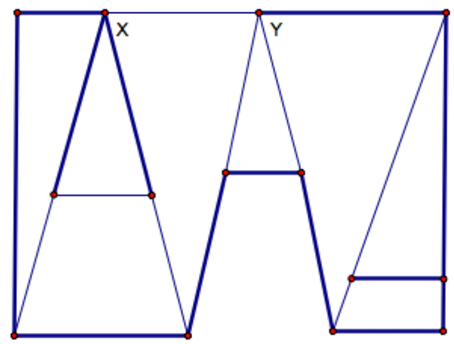
FIGURE 7-1.

The telephone company needs to find a network which
(i.) utilizes the existing network of roads,
(ii.) connects all the towns, and
(iii.) has the least cost.

Translating these requirements to the language of graph theory, we get
(i.) The network must be a *subgraph* of the existing road network.
(ii.) The network must *span* the existing road network.
(iii.) The network must be an *optimal solution* to the problem: that is, the total weight of the network must be as small as possible.

Since we are looking for an optimal solution, it follows that the spanning subgraph we are looking for must have no circuits.

Why not? (Consider the subgraph, shown in thick edges, of the following graph.)



We can now give a formal definition that summarizes the requirements of the telephone company (see Figure 7-1, previous page).

Definition 22. A **tree** is a connected graph (that is, a network) with no circuits.

Definition 23. If a subgraph G' of a graph G is a tree, we call G' a **spanning tree** for G if G' spans G (that is, if all the vertices of G are vertices of G').

Definition 24. Among all the spanning trees for a given graph G, the **minimal spanning tree** for G is the one with least total weight.

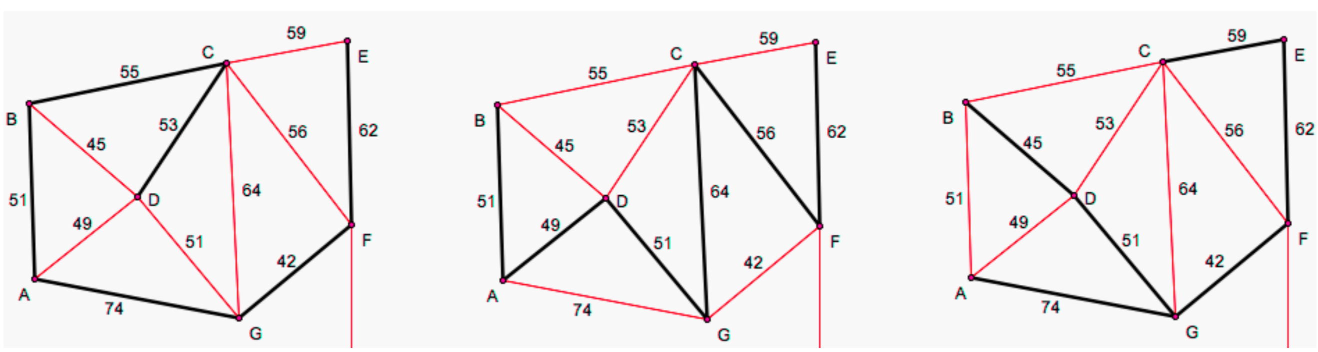


FIGURE. Three different spanning trees (thick lines) of the graph given in Figure 7-1 (shown here as thin lines). The total weight of each spanning tree is computed by adding the weights of every edge in the tree.

Are any of the three graphs shown above <u>minimal</u> spanning trees for the road network?
Why or why not: how would you justify your answer?

3. Properties of trees

Some graphs are "more" connected than others. The complete graph K_N is of course the "most" connected.

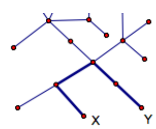
Trees occupy an important niche between disconnected graphs and graphs which are *overconnected*. A tree is special in so far as it is *barely connected*: this means several things.

Definition 25. A graph G is called k-**connected** if between any two vertices X and Y of G there are at least k distinct XY-paths in G. A k-connected graph is called **overconnected** if $k \geq 2$. A graph G is called **barely connected** if between any two vertices X and Y there is one and only one XY-path in G. *(You need not memorize the definitions in this paragraph.)*

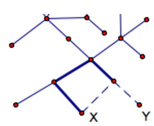
We now provide *proofs* of a few properties which all trees satisfy. (Recall that a **proof** is an airtight argument that a mathematical proposition is true without exception, e.g. for *every* tree.)

Theorem 6. Every tree is barely connected. That is, given any two vertices X and Y of any tree G, there is one and only one XY-path in G.

Proof.
Let X and Y be two vertices in a tree G.
Since a tree is by definition connected, there is at least one XY-path in G (by definition of connectedness).



Either there is a second, distinct XY-path in G, or there is not. We will show that it is impossible that there be a second XY-path in G.
Assume there is a second XY-path (shown as a dashed line) in G. Then there must be a circuit somewhere in G: specifically, in the subgraph of G made up of the two XY-paths:

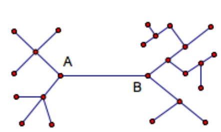


But since G is a tree, G cannot possibly have a circuit (by definition of a tree).
When we assumed that there is a second XY-path, we were able to argue that G has a circuit. Since we know that it is impossible for G to have a circuit, we conclude that our assumption was incorrect: that is, we conclude that there is no second XY-path. And this is what was to be shown. □

Definition 26. A **cut-edge** is an edge in a connected graph whose deletion disconnects the graph.

Corollary to Theorem 6. Every edge in a tree is a cut-edge.

Proof.
Let G be a tree. Then (by definition of a tree) there are no circuits in G.
Assume that there is an edge AB of G such that AB is *not* a cut-edge. (We will show that this assumption leads to an impossibility.)
By definition of a cut-edge, deleting AB does *not* result in a disconnected graph.



In particular, there must be a path from A to B which does not pass through AB. *Complete the proof by arguing that something impossible happens.*

□

Theorem 7. If there is one and only one path joining any two vertices of a graph G, then G is a tree.

Proof.
Let G be a graph such that there is one and only one path in G joining any two given vertices of G.

A tree is a connected graph with no circuits. We need to show that
(i.) G is connected, and
(ii.) G has no circuits.
By definition, G is connected because every two vertices of G are joined by a path: this proves (i).

Prove (ii) *by answering the question:*

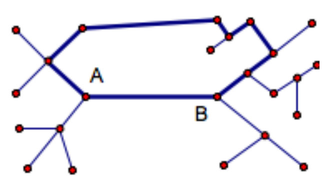
Why does G have no circuits?

(Hint: Assume G does have a circuit, and explain what goes wrong!)

□

Theorem 8. If every edge in a graph G is a cut-edge, then G is a tree.

Proof.
Let G be a connected graph such that every edge is a cut-edge.
We need to show that G has no circuits.
Assume that G has a circuit, and let AB be an edge in the circuit.



Complete the proof by arguing that something impossible happens.

□

Theorem 9. A tree on N vertices has $N-1$ edges.

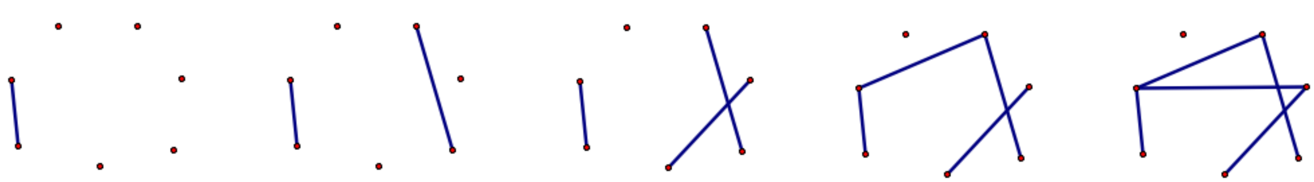


FIGURE. Adding edges one at a time, being careful not to form a circuit, we see that $7-1$ edges suffice to join 7 vertices. (This figure is not a proof for all *N, but it demonstrates Theorem 9 in the particular case when $N=7$.)*

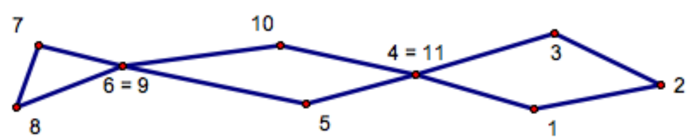
Don't spend too much time deciphering the proofs of the Lemma and the Theorem on this page. However, you *should* try to convince yourself that the Lemma and the Theorem are in fact true! (Take a moment with the Lemma before reading its proof, and see if you can convince yourself. Justifying Theorem 10 to oneself requires a bit more imagination.)

Lemma to Theorem 10. A circuit $1, 2, 3, \ldots, n-1, n, 1$ has length n.

Proof.
A path joining 2 vertices has 1 edge, and every time we add 1 vertex to those joined by the path, we add 1 more edge:

In general, a path joining n vertices has exactly $n-1$ edges.

Let $1, 12, 23, 34, 45, \ldots, (n-2)(n-1), (n-1)n, n1, 1$ be a circuit joining n vertices (some of which may be repeated).



The path $1, 12, \ldots, (n-1)n, n$ from vertex 1 to vertex n has length $n-1$.
Adding the edge $n1$ to complete the circuit, we now have n edges. $\square$

The proof of Theorem 10 is markedly more difficult than any of the proofs we have seen so far, and you are not responsible for understanding the nitty gritty details of the argument.

Theorem 10. If a network G with N vertices has $N-1$ edges, then G is a tree.

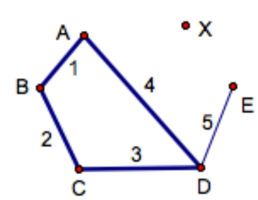
Proof.
We will show that:

If there is a circuit in a graph G with N vertices and $N-1$ edges, then G is disconnected.

It will then follow that:

If G is a connected graph with N vertices and $N-1$ edges, then G has no circuits.

Let G be a graph with N vertices and $N-1$ edges.
Let k be the length of the longest circuit (shown in thick edges) in G.



There are only $N-1$ edges in G so, by the Lemma, since a circuit joining k vertices has length k, we have $k \leq N-1$.
Then there are $N-k \geq 1$ vertices not on the longest circuit in G.
Since there are only $N-1$ edges in G, it follows that (*) there is a vertex X of G such that there is no path from X to the circuit. Thus G is disconnected.

To see this (*), observe that if each of the $N-k$ vertices were attached by a path of length ≥ 1 to some vertex in the longest circuit, this would add at least $(N-k) \times 1$ edges to the k edges we already know about in the longest circuit. We would then have k edges (in the circuit) plus at least $(N-k) \times 1$ edges (in paths attached to the circuit), which is a total of at least $N = k + (N-k) > N-1$ edges. Therefore, not all of the $N-k$ vertices not on the circuit can be attached by paths to the circuit: there simply aren't enough edges in the graph.

Therefore, if G is a network with N vertices and $N-1$ edges, then G is a tree. $\square$

Ch. 11. Symmetry

Textbook: Peter Tannenbaum, *Excursions in Mathematics*, 6th edition.

1. First definitions for geometry. Rigid motions.

Reading: Section 11.1

Perhaps the most important textbook ever written is Euclid's Elements, a work of geometry written in ancient Greece. We don't know how much of Euclid's book was his own original invention, but we do know that many of the results in his Elements had appeared previously, both in Greece and in other cultures.[1] The book's originality is mainly due to its style. Every fact which is presented is immediately followed by a mathematical "proof" (or argument) written in a perfectly meticulous manner: each sentence in the proof follows immediately from the preceding sentences, with no illogical jumps[2] or missing steps left for the reader to fill in.

The Elements begins with a problem which always accompanies the first few steps a newborn science must take: we might call it the *Problem of First Definitions*. Geometry, like any science, requires that there be *precise* definitions of the things to be studied. But when we set out to write our definitions, we have no precise definitions to which we can refer, so precisely how can we speak? You will have to decide for yourself whether Euclid's solution to the Problem of First Definitions was a happy one.

Euclid's Elements begins with the following definitions.

Definition 1. A *point* is that which has no part.

Definition 2. A *line* is breadthless length.

Definition 3. The *ends of a line* are points.

Definition 4. A *straight line* is a line which lies evenly with the points on itself.

Definition 5. A *surface* is that which has length and breadth only.

Definition 6. The *edges of a surface* are lines.

Definition 7. A *plane surface* is a surface which lies evenly with the straight lines on itself.

Are these definitions perfectly clear? We will not give our own formal definition of "line" and "plane," which for us will be *undefined terms*. However, once we have "defined" the first few (imprecisely defined) terms, we must ensure that *every definition from then on is perfectly precise*. In this way, the validity of theorems later to be proven is guaranteed, so long as we accept the first few definitions. Mathematics is sometimes called the *hypothetical science* for this reason—it doesn't *absolutely* prove anything. But it can prove quite a lot if we all agree to certain hypothetical definitions.[3]

The next definition—our first "real" definition—introduces the topic we will be studying throughout this chapter.

A rule for moving each point in the plane to a new position is called a **plane transformation**.

In other words, a plane transformation—call it $\mathcal{M}$—moves each point in the plane from its starting position P to its ending position P'. In symbols, we write this fact as

$$\mathcal{M}(P) = P' \qquad \text{or} \qquad \mathcal{M} : P \mapsto P'$$

but you are not required to use or know this notation (called *function notation*).

---

[1] We know most about China, India, and the Arab world, where mathematical discovery and innovation often flourished in the millenia before and during the European Dark Ages.

[2] Well, almost none. Modern mathematicians are able to point out a few serious logical gaffes on Euclid's part, and after reading this page, you should be, too.

[3] Another fundamental mathematical idea which may seem extraordinarily difficult to define is a *number*, since the definition has to be broad enough to include irrational numbers like π and complex numbers like $\sqrt{-1}$, which are far removed from our experience and common-sense intuitions. Modern mathematicians have indeed given a precise definition of what it means to be a number, but as we all know, it's perfectly possible to do arithmetic without being able to say with complete precision what a number *is*.

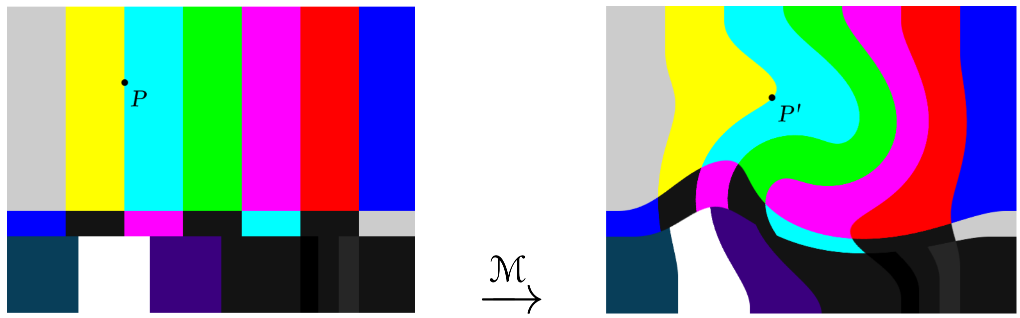


FIGURE. A plane transformation $\mathcal{M}$ which moves the point at starting position P to the ending position P'.

We can define a plane transformation by specifying the following rule:

If any figure appears drawn in the plane, pick it up and move it.
Move every figure that appears in exactly the same way.

A plane transformation defined by such a rule *does not alter the size or shape of any figure*. This is (obviously) because picking up an object and moving it somewhere else in the plane does not change its shape.[4]

Definition. A plane transformation which does not change the shape or size of any plane figure is called a **rigid motion** (or **isometry**).

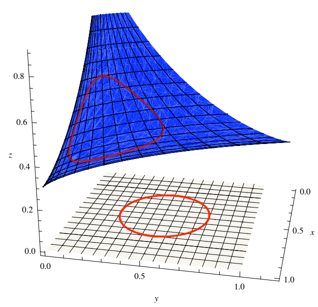


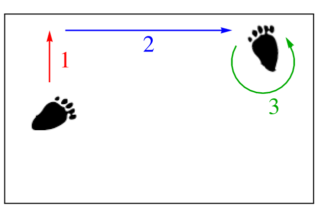
FIGURE.

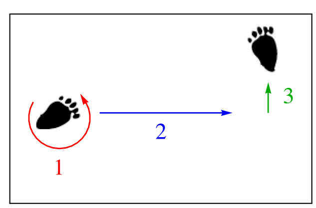
A *rigid motion*, which is a *plane* transformation, does not change the shape of the circle.

But moving the circle around on the curved surface above the plane—a portion of a shape called an astroidial ellipsoid—definitely does!

In particular, a rigid motion cannot involve any ripping, stretching, shrinking, or bending (compare the plane transformation at the top of this page).

Suppose you and I each have a plane in front of us, upon which lies a piece of paper in the shape of a footprint. Say I move the footprint on my plane as shown in the figure below at left, and you move the footprint on your plane as shown below at right. The footprint ends up in the same place, even though we did not move it in the same way. When this happens, we call the two rigid motions *equivalent*.



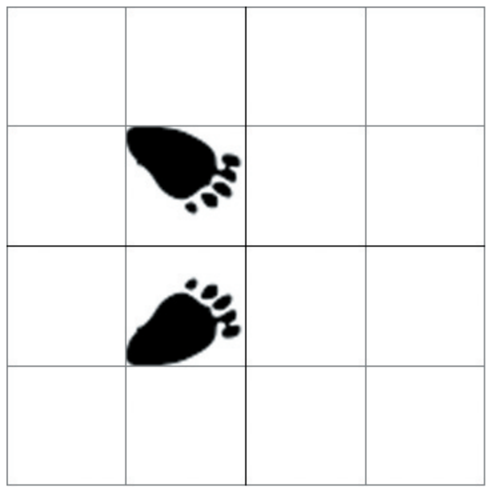


[4]This is *not* true for spaces other than planes: planes are *flat*. The blue surface in the figure in the middle of this page is an example of a space that is not flat. Our universe is another example. The three-dimensional space we see is curved, in the sense that the curvature of space changes in the vicinity of a massive object (picture a bowling ball resting on a bedsheet held aloft by a circle of people). In particular, a black hole changes the shape of anything near it, provided we can all agree on what we mean by the word "shape"...

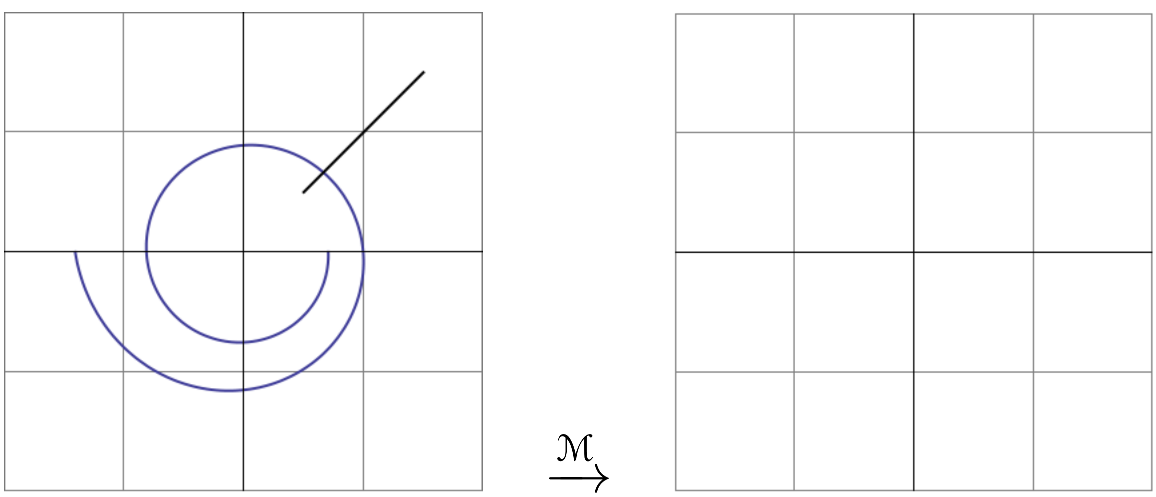
Definition. Two rigid motions are called **equivalent rigid motions** if they both move a given figure in starting position A to the same ending position B.

Our definition implies that we can completely characterize a rigid motion by specifying the starting position and ending position of an object being moved. What happens in between—that is, the particular details of the motion, shown as arrows in the figure at the bottom of the previous page—is unimportant for our purposes.

Ex. 1.
The following image defines a rigid motion, call it $\mathcal{M}$.

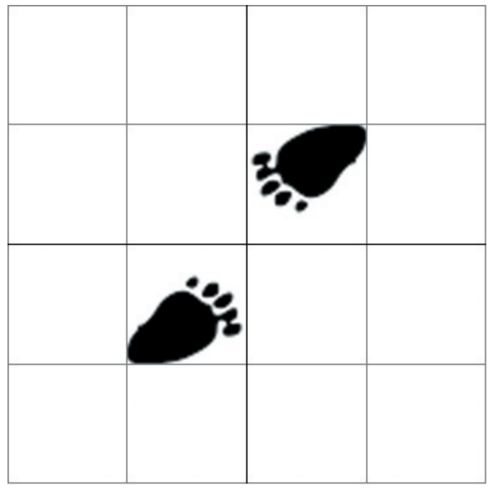


Determine what $\mathcal{M}$ does to the plane shown at left by drawing the ending position of each figure on the plane shown at right. (The grid lines are provided for reference only.)

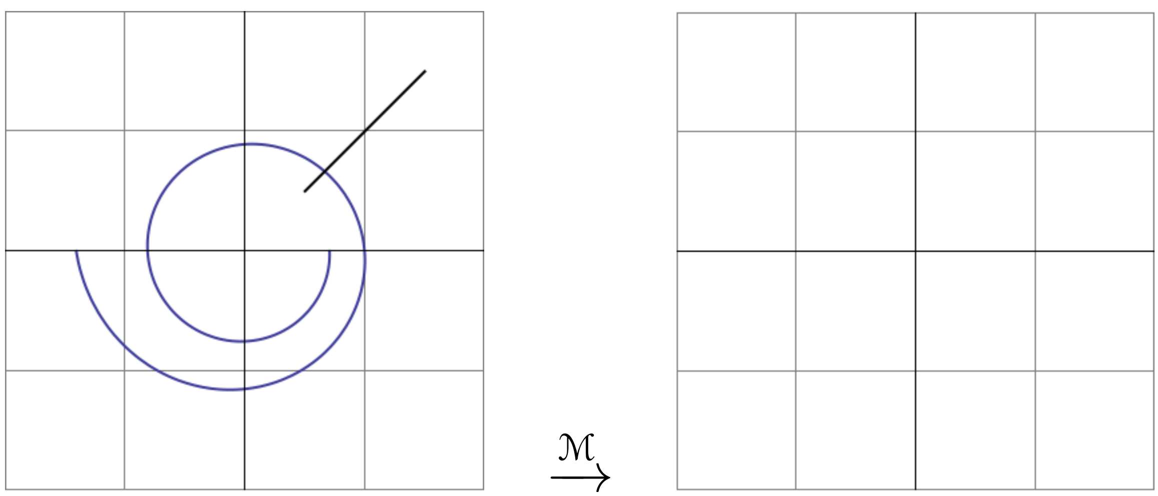


How would you describe this transformation in words (that is, without using pictures)?

Ex. 2.
The following image defines a rigid motion, call it $\mathcal{M}$.

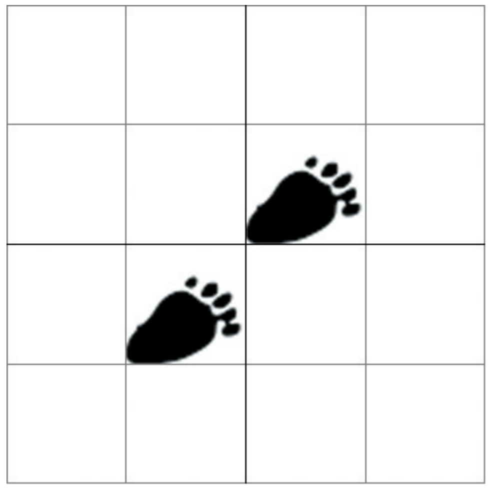


Determine what $\mathcal{M}$ does to the plane shown at left by drawing the ending position of each figure on the plane shown at right. (The grid lines are provided for reference only.)

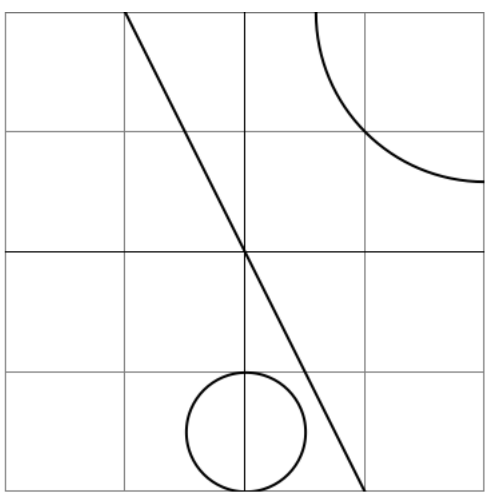


How would you describe this transformation in words (that is, without using pictures)?

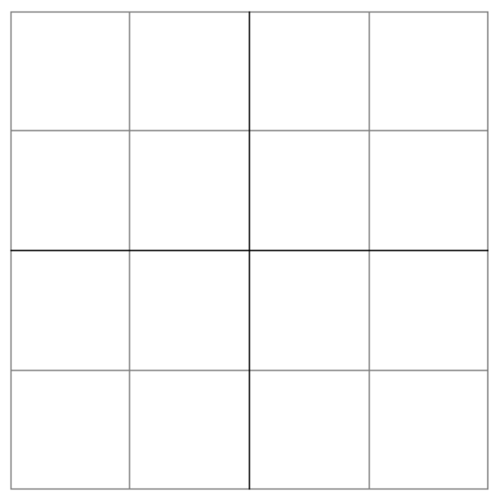
Ex. 3.
The following image defines a rigid motion, call it $\mathcal{M}$.



Determine what $\mathcal{M}$ does to the plane shown at left by drawing the ending position of each figure on the plane shown at right. (The grid lines are provided for reference only.)

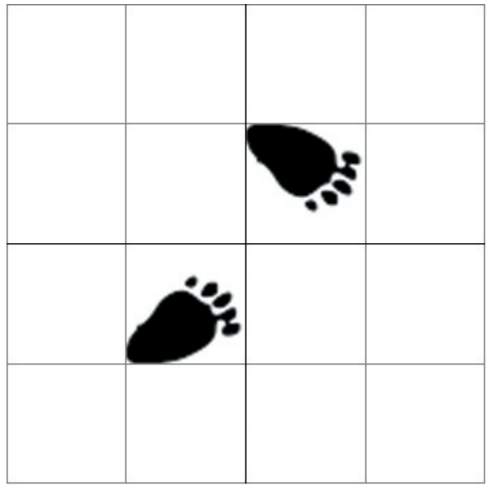


$\xrightarrow{\mathcal{M}}$

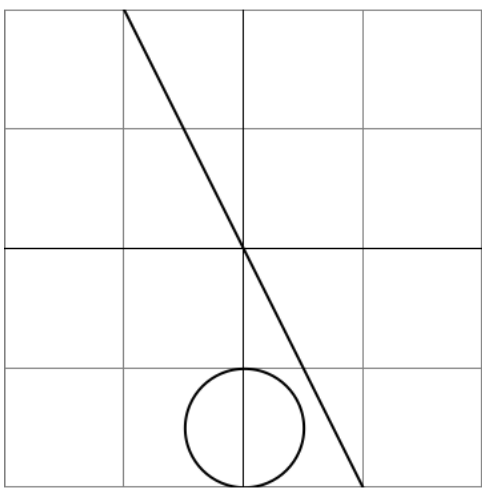


How would you describe this transformation in words (that is, without using pictures)?

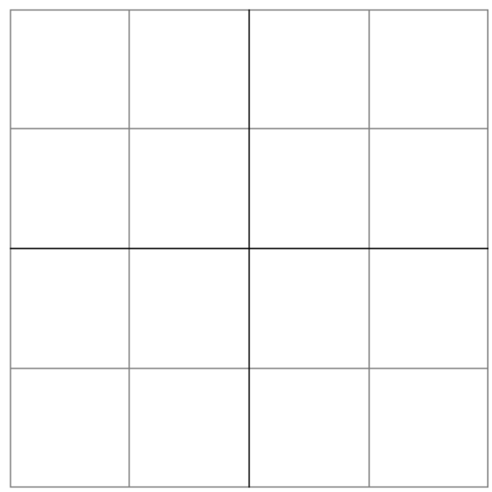
Ex. 4.
The following image defines a rigid motion, call it $\mathcal{M}$.



Determine what $\mathcal{M}$ does to the plane shown at left by drawing the ending position of each figure on the plane shown at right. (The grid lines are provided for reference only.)



$\xrightarrow{\mathcal{M}}$



How would you describe this transformation in words (that is, without using pictures)?

It turns out that *every rigid motion* is one of the four rigid motions we have just seen.

basic rigid motion	**informal description**
reflection	flipping across a line
rotation	pivot around a point
translation	sliding along a line
glide reflection	slide, then flip across the line of sliding

These four rigid motions will be called the **basic rigid motions of the plane**.

We will examine each of the four basic rigid notions in more detail. Each of the four will be given a symbol: F_ℓ for reflections, $R_{P,\theta}$ for rotations, T_v for translations, and $G_{v,\ell}$ for glide reflections. These symbols will be explained as we go.

2. Fixed points. Reflections. Composition of rigid motions.

Reading: Section 11.2

When a figure S is moved by a rigid motion $\mathcal{M}$, the figure in its new position is called the **image of S under $\mathcal{M}$**.

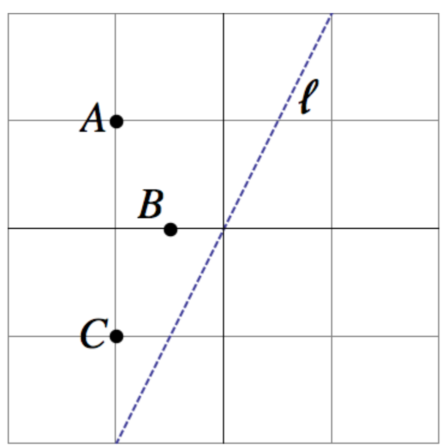
In particular, if $\mathcal{M}$ moves a point in position P to a new position P', we call P' the image of P under $\mathcal{M}$.

A **reflection** is a rigid motion in which every figure is moved so that its image is the "mirror image" of its starting position.

This definition begs the question, *What do we mean by "mirror image"?* To have a mirror image, one must have a mirror. For a reflection, this "mirror" is a line, a line we will usually call ℓ.

The symbol F_ℓ will denote reflection in a given line ℓ.

Ex. 5a.
Find the images under the reflection F_ℓ of the three points A, B, and C.



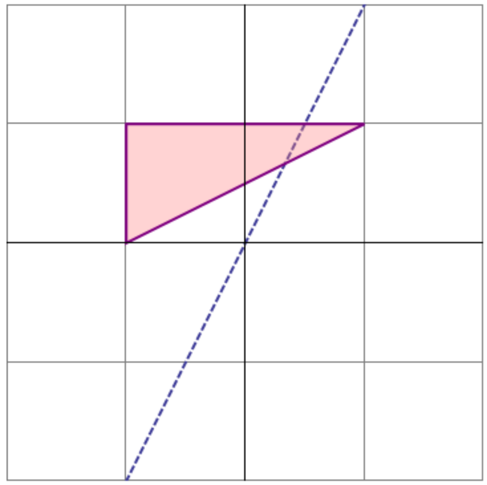
Ex. 5b.
A point whose image under a rigid motion $\mathcal{M}$ is the same as its starting position is called a **fixed point of $\mathcal{M}$**.

Find a fixed point for the reflection F_ℓ. Label it in the above picture, say as P.

How many fixed points does F_ℓ have?

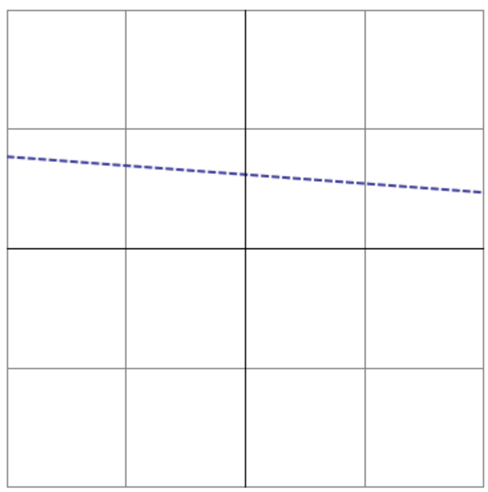
Ex. 6.

If ℓ is the dashed line in the picture below, what is the image under F_ℓ of the triangle?



Ex. 7a.

Let m be the dashed line in the picture below. Draw a triangle whose image under F_m is exactly the same as its starting position.



Ex. 7b.

Identify the fixed points of F_m.

A rigid motion is called **improper** (or **opposite**) if it reverses orientation. Otherwise, we call the rigid motion **proper** (or **direct**).

Ex. 8a.
The symbol $\circ$ is called **composition**. We write $\mathcal{M} \circ \mathcal{N}$ to mean, *Do* $\mathcal{N}$ *first, then* $\mathcal{M}$. (Note that the order is right-to-left).

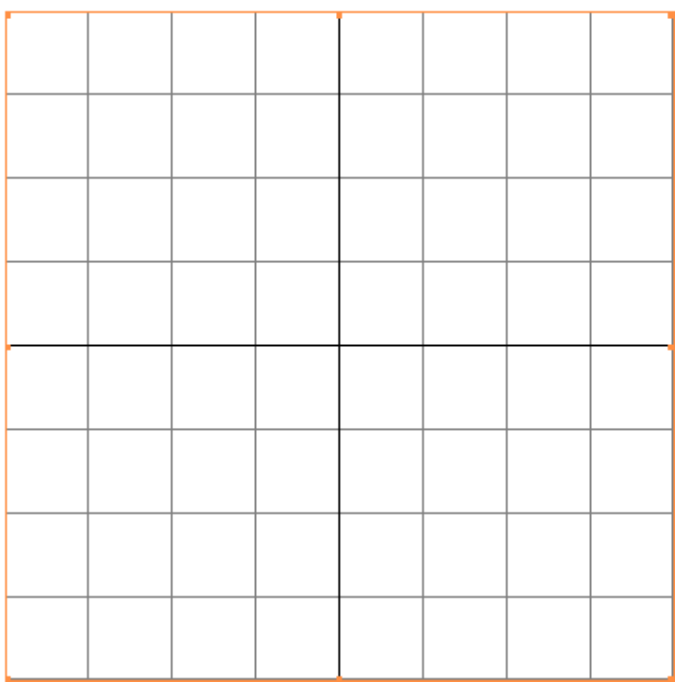
In particular, $F_m \circ F_\ell$ means, *Reflect across the line* ℓ *first, and then reflect across the line* m.

If k, ℓ, and m are three lines (possibly not all distinct), which of the following rigid motions are proper, and which are improper?

1. F_k
2. $F_\ell \circ F_k$
3. $F_m \circ F_\ell \circ F_k$

Ex. 8b.
Let ℓ be a line. Describe the effect of the rigid motion $F_\ell \circ F_\ell$ in words.

Ex. 8c.
Draw two (distinct) parallel lines on the grid provided, and label them ℓ and m. Describe the effect of the rigid motion $F_m \circ F_\ell$ in words.



3. Angle measures. Rotations. The identity motion.

Reading: Section 11.3

To carry out a reflection, we need only one piece of information: we need to know what the line ℓ of reflection is.

What do we need to know if we want to carry out a *rotation*? We need two things: a *point* P around which to rotate, and an *angle* with measure θ (given in degrees or radians, but we will always use degrees).

The point P is called the **rotocenter** (or the **center of the rotation**).

We really ought to call θ the **measure of the angle of rotation**, but we will usually be sloppy and omit the phrase "*the measure of*."

One way or the other, we *must* indicate to the reader which way an angle opens. Tannenbaum does this in words, by saying *clockwise* or *counter-clockwise*.

We will usually use the standard practice of mathematicians, which is as follows. A *positive angle measure* indicates that the angle opens *counter-clockwise*. A negative angle measure indicates an angle that opens clockwise.

The symbol $R_{P,\theta}$ will denote the rotation with rotocenter P, and angle of rotation θ.

The diagram below shows some angle measures. Each angle measure refers to the angle formed where the labeled line segment meets the half-line which begins at the center of the circle, and "ends" at the arrowhead labeled x.

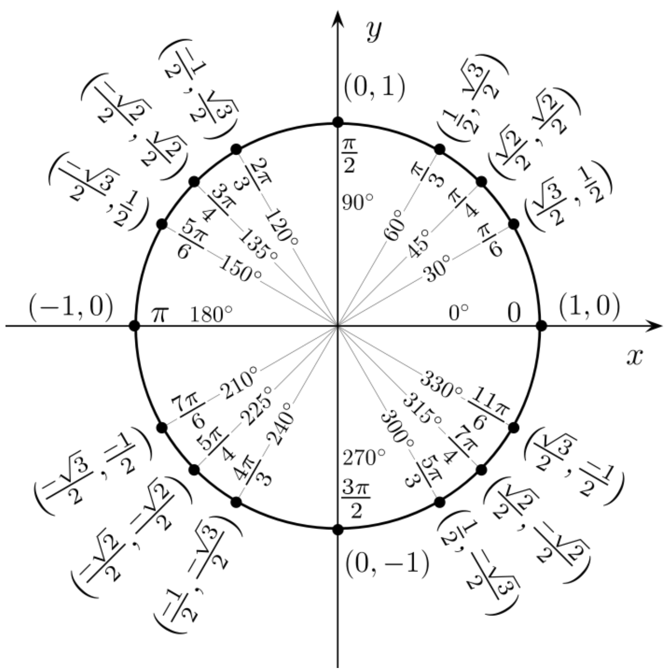
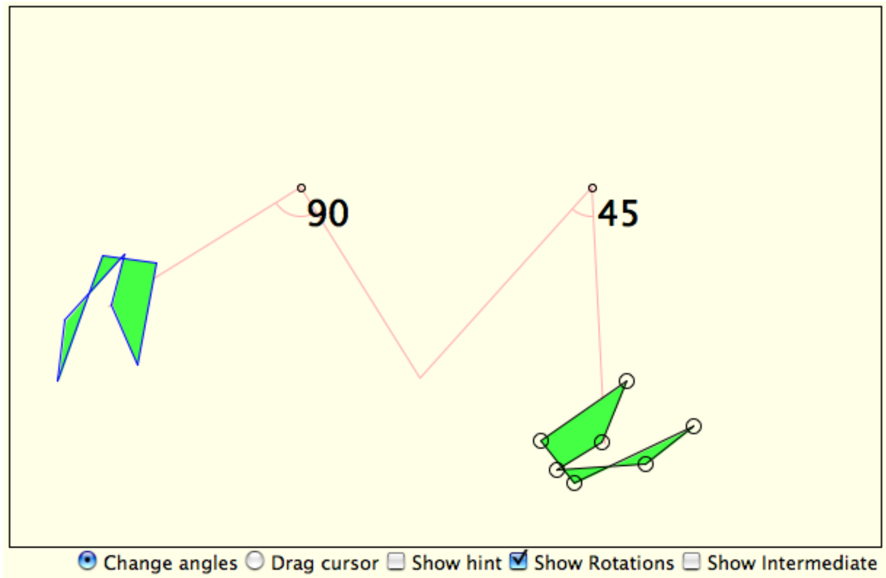


FIGURE. Angle measures appear on the inside of the circle.
We will not use radians, which appear here as fractional multiples of π.
All information outside the circle, which is used in other math classes, may also be disregarded.
You are responsible for understanding angle measures given in degrees (°).

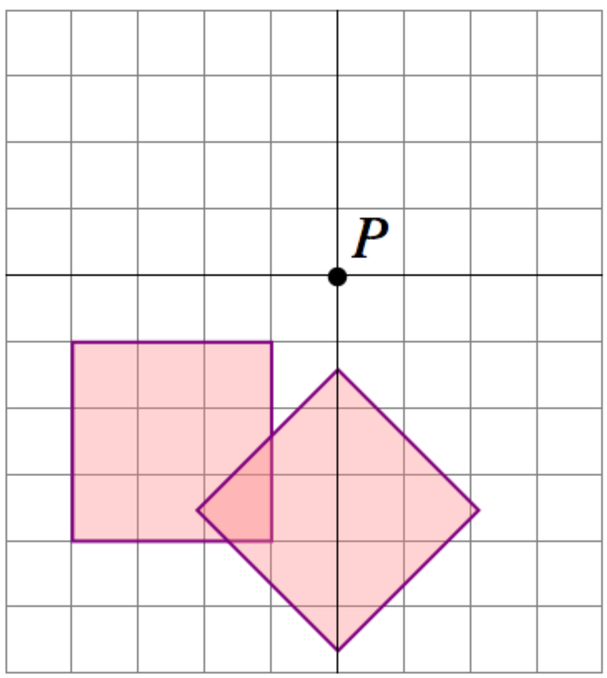
Ex. 9.
Are rotations all proper? What about the composition of two rotations? Does your answer to the last question change if

we replace the word “two” with the word “three”? How about the composition of more than three rotations?

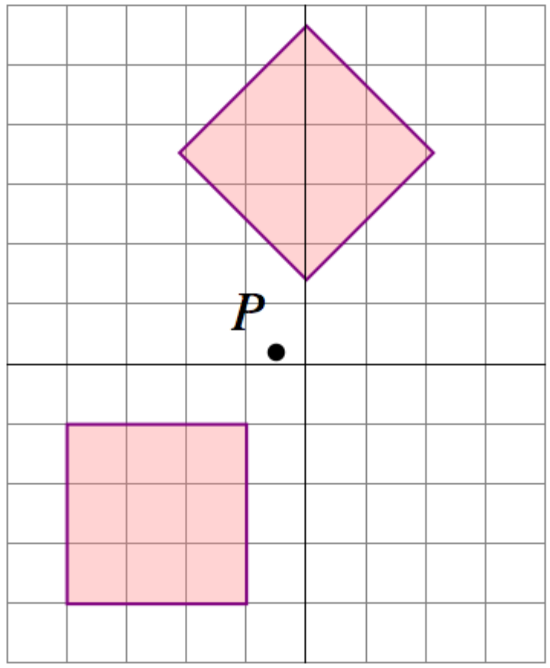
The following screenshot of an applet at `http://www.cut-the-knot.org/Curriculum/Geometry/RotationTransform.shtml` shows the effect of doing one rotation, followed by a second rotation around a different rotocenter.



Ex. 10a.
A rotation $R_{P,\theta}$ is shown in the picture below. Find two different possible values for the angle measure θ.



Ex. 10b.
A rotation $R_{P,\theta}$ is shown in the picture below. Find two different possible values for the angle measure θ.



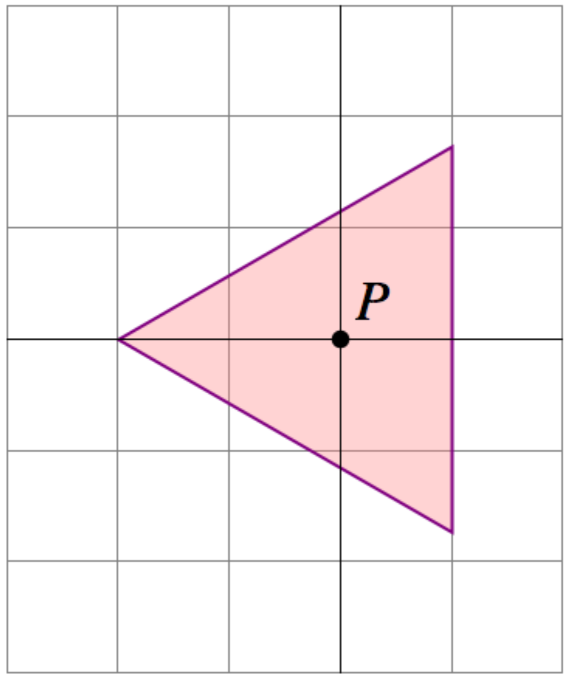
The **identity motion** is the rigid motion defined by doing nothing. That is, we don't move any points at all.

Now, a rotation $R_{P,360^\circ}$ of 360° around any given point P moves every point exactly back to where it started. Thus every point in the plane is a fixed point of $R_{P,360^\circ}$.

Ex. 11a.
Suppose P is a point in the plane. Find an angle measure θ such that $R_{P,45^\circ} \circ R_{P,\theta}$ is equivalent to the identity motion. Then find a different value of θ which also works.

Ex. 11b.
Suppose P is a point in the plane. Find an angle measure θ such that $R_{P,45^\circ} \circ R_{P,\theta}$ is equivalent to $R_{P,180^\circ}$.

Ex. 12.
For how many different rotations with rotocenter P is the image of the triangle exactly the same as the starting position of the triangle?



Ex. 13.
Suppose P is a point in the plane. How many fixed points does $R_{P,\theta}$ have? What are they? Does it matter what the value of θ is?

If we are given a point P and its image P' under some unknown rotation $R_{X,\theta}$, can we identify the rotocenter X? (Say no!) In fact, any point on the perpendicular bisector of line segment $\overline{PP'}$ might be the rotocenter, so long as we don't already know what the angle of rotation θ is. (See Figure 11-8, p. 379.)

If we are given two points P and Q, and their images P' and Q', we can determine the rotocenter as follows.

1. Find the perpendicular bisector of $\overline{PP'}$.
2. Find the perpendicular bisector of $\overline{QQ'}$.
3. If the two perpendicular bisectors are different, then the rotocenter is the point at which the two perpendicular bisectors intersect. (See Figures 11-8a and 11-8b.)
4. If the two perpendicular bisectors are the same, then the rotocenter is the point at which the half-lines $\overrightarrow{PQ}$ and $\overrightarrow{P'Q'}$ meet. (See Figure 11-8c.)

4. Translations. Inverse motions.

Reading: Section 11.4

What information do we need in order to carry out a translation?

A **vector** (usually represented graphically by an arrow) is a __________________ and a __________________.

Vectors are most often defined by joining two points—a **foot** (or **initial point**) A, and a **head** (or **terminal point**) B—with an arrow. The length of the arrow and its direction characterizes the displacement of B relative to A: in other words, how much one should move the point A to "carry" it to the point B, and in what direction.

If v is a vector defined by an arrow *from* A *to* B, we use the symbol

$$-v$$

to mean, "the vector in the opposite direction (from B to A) with the same length as v."

Ex. 14.
Draw a vector which moves a point A by (roughly) 1 inch, in the northeast direction. Then draw two more vectors which describe exactly the same motion.

A **translation** is the rigid motion defined by sliding each object by the length of a vector, in the direction of that vector.

The symbol T_v will denote translation by a given vector v.

Ex. 15a.
I am thinking of a vector v. How many fixed points does the translation T_v have?

Ex. 15b.
A translation is proper or improper?

Suppose $\mathcal{M}$ is a rigid motion. Then the **inverse motion** for $\mathcal{M}$ is a rigid motion $\mathcal{N}$ whose composition (in either order) with $\mathcal{M}$ is the identity motion.

In symbols, the statement that "$\mathcal{N}$ is the inverse motion for $\mathcal{M}$" means that both

$$\mathcal{M} \circ \mathcal{N} \qquad \text{and} \qquad \mathcal{N} \circ \mathcal{M}$$

are equivalent to the identity motion.

Ex. 16.
I am thinking of a vector v. Find a vector w such that $T_w \circ T_v$ is equivalent to the identity motion.

In general,

$$T_w \circ T_v =$$

The symbol $\mathcal{M}^{-1}$ will denote the inverse motion for a given rigid motion $\mathcal{M}$.

In general,

$$\left(T_v\right)^{-1} =$$

Ex. 17.
Find a pair of rigid motions $\mathcal{M}$ and $\mathcal{N}$ such that

$$\mathcal{M} \circ \mathcal{N} \qquad \text{and} \qquad \mathcal{N} \circ \mathcal{M}$$

are <u>not</u> equivalent.

5. Glide reflections. Commuting rigid motions.

Reading: Section 11.5

We say that $\mathcal{M}$ and $\mathcal{N}$ **commute**[5] **(with each other)** if

$$\mathcal{M} \circ \mathcal{N} \qquad \text{and} \qquad \mathcal{N} \circ \mathcal{M}$$

are equivalent.

Ex. 18.
Find a pair of rigid motions $\mathcal{M}$ and $\mathcal{N}$ such that

$$\mathcal{M} \circ \mathcal{N} \qquad \text{and} \qquad \mathcal{N} \circ \mathcal{M}$$

are equivalent. Verify that $\mathcal{M} \circ \mathcal{N}$ and $\mathcal{N} \circ \mathcal{M}$ are equivalent by drawing a triangle T, and the image of T under each of the two "compound" (or "laundry list") transformations $\mathcal{M} \circ \mathcal{N}$ and $\mathcal{N} \circ \mathcal{M}$, showing each intermediate position of the triangle under the transformations with dashed or colored lines.

(You may want to doodle on a piece of scratch paper before filling in your answer here.)

[5]The word may remind you of a certain property of addition and multiplication: namely, the two **commutative properties** $a + b = b + a$ and $a \cdot b = b \cdot a$. For "ordinary" numbers, these two properties always hold. But for other types of mathematical objects, they may not. For example, addition of matrices $A + B$ is commutative, but multiplication is not, since it is not always true that $AB = BA$ when A and B are two $n \times n$ matrices. Returning to the present context, certainly *composition is not commutative*. That is, if $\mathcal{M}$ and $\mathcal{N}$ are two rigid motions, then it need not be true that the two "compound" rigid motions $\mathcal{M} \circ \mathcal{N}$ and $\mathcal{N} \circ \mathcal{M}$ are equivalent. (You are not responsible for understanding all that has been said in this footnote, but you should be able to convince yourself that the sentence before the current one is a fact.)

Proposition.
True or false:

Any translation T_v commutes with any reflection F_ℓ.

(If true, explain why. If false, give an example that demonstrates why not.)

A **glide reflection** is the composition of a translation T_v and a reflection F_ℓ, provided that the vector v is parallel to the line ℓ.

Ex. 19a.
I am thinking of a glide reflection $G_{v,\ell}$. How many fixed points does $G_{v,\ell}$ have? Does changing the relationship between v and ℓ change your answer?

(Show your work by picking a suitable v and ℓ of your choice, and then drawing a figure T and its image under $G_{v,\ell}$. Draw the intermediate position with dashed or colored lines.)

Ex. 19b.
Do there exist proper glide reflections? To put the same question in other words: Are all glide reflections improper?

Ex. 20.
If we are given a point P and its image P' under an unknown glide reflection $G_{v,\ell}$, is that enough information for us to carry out the glide reflection for other figures?

Ex. 21.
Find the inverse motion for $G_{v,\ell}$, where v is a given (but unknown) vector and ℓ is a given (but unknown) line parallel to v.

6. Symmetry as a rigid motion

Reading: Section 11.6

A rigid motion that moves a geometric figure back onto itself is called a **symmetry of the figure**.

Ex. 22.
Find all symmetries of a square.

Ex. 23.
Find all symmetries of the propeller shown in the figure.

FIGURE. The Hamilton Standard 54H60 propeller.

We say that two objects have the same **symmetry type** if they have exactly the same symmetries.

For example, the symmetry type of the square is called D_4 (or **dihedral**-4). (The 4 stands for four reflections and four rotations.)

Ex. 24.
Find all symmetries of the propeller shown in the figure. Its symmetry type is called Z_5 (or **cyclic**-5).

FIGURE. A propeller on display at the Singapore Changi International Airport.

Ex. 25.
Find all symmetries of the shape shown in the figure.

FIGURE.

Ex. 26.
Find all symmetries of the shape shown in the figure. Its symmetry type is called D_1 (or **dihedral**-1).

FIGURE.

Ex. 27.
The playing card called "six of spades" is shown below. Two different versions of the same playing card are shown. Find all the symmetries of the six of spades. Its symmetry type is called Z_1 (or **cyclic**-1).

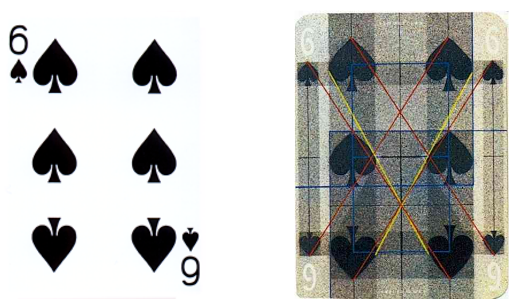


FIGURE. Two versions of the same playing card.

Ex. 28.
Find all symmetries of the figure. Its symmetry type is called

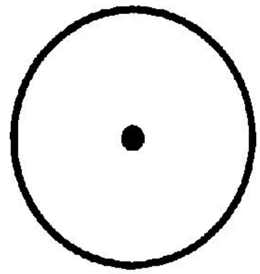


FIGURE. A familiar shape.

Summarizing, we have seen three broad categories of symmetry type:

- **Cyclic** symmetry types Z_N

- **Finite dihedral** symmetry types D_N

- The **infinite dihedral** symmetry type D_∞

Ch. 12. Geometry of fractal shapes

1. What is a fractal?

Reading: Section 12.1

A **fractal** is a geometric figure in which more and more features appear as the magnification increases.

The hallmark of a fractal is that, whenever we zoom in on a piece of it, we discover new features at the smaller scale which resemble features found at the larger scale. This phenomenon is called **self-similarity**.

- In a fractal that is **approximately self-similar**, features at a smaller scale resemble, but are not identical to, features found at the larger scales.

  Here are some examples of approximately self-similar fractal shapes:

Romanesco broccoli

Michael F. Barnsley, "Wattle"
http://www.superfractals.com/

A woodburn fractal

- In a fractal that is **exactly self-similar**, the same patterns are exactly repeated as we zoom in further and further.

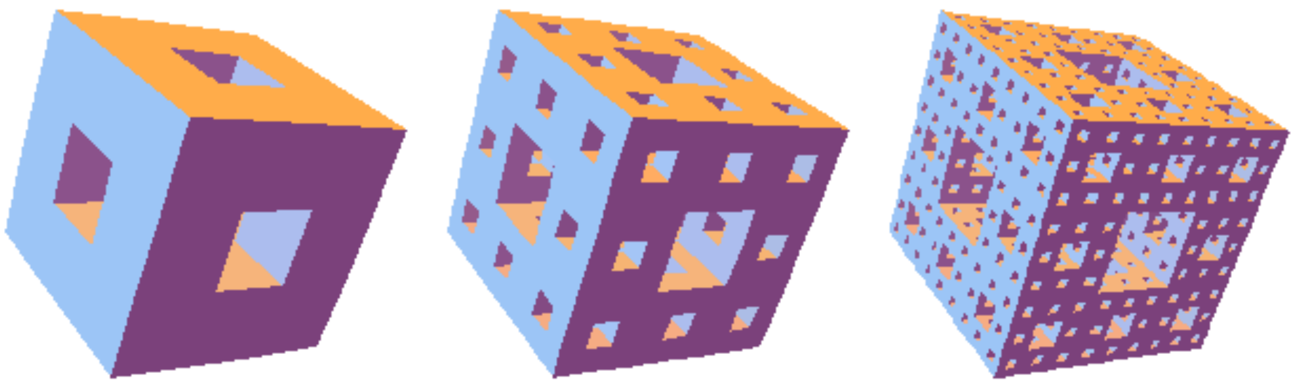
  (None of the above three images are "exactly" self-similar.)

We can describe how to build an *exactly* self-similar fractal by giving a list of instructions, and instructing the builder to repeat the instructions indefinitely.

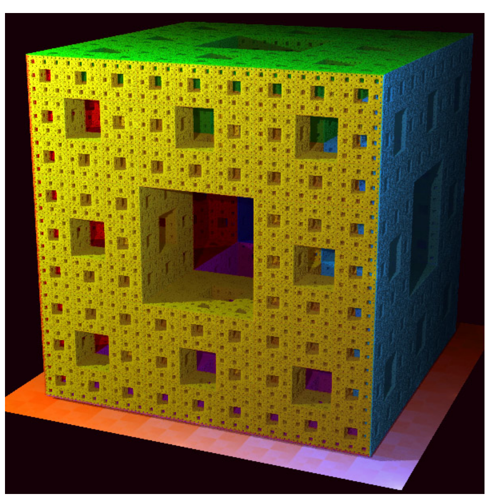
A list of clear and precise instructions for carrying out a task is called an **algorithm**, and each repetition of the algorithm is called an **iteration**, so we will call a list like those that follow an **iterated algorithm**.

Ex.: Iterated algorithm for constructing the Menger sponge.

Step 1.	Begin with a cube, each of whose sides measures 1 unit in length.
Step 2.	Divide the cube into $3 \times 3 \times 3 = 27$ identical "sub-cubes" by dividing each side into three identical squares.
Step 3.	Remove the center sub-cube and the 6 sub-cubes adjacent to it.
Step 4.	For each sub-cube, repeat steps 2 through 4.



Notice that Step 4 says to repeat some of the steps—including Step 4 itself! Because of this, the algorithm can never be completed. However, as the figure below suggests, computers can draw a picture that provides detail at a smaller scale than any scale large enough to be seen by the naked eye:



Ex.: Iterated algorithm for constructing the Sierpiński gasket.

Step 1.	Begin with a triangle, each of whose sides measures 1 unit in length.
Step 2.	Divide the triangle into 4 identical "sub-triangles" by dividing each of the original triangle's sides into two identical line segments, and joining their endpoints as shown.
Step 3.	Remove the center sub-triangle.
Step 4.	For each sub-triangle, repeat steps 2 through 4.

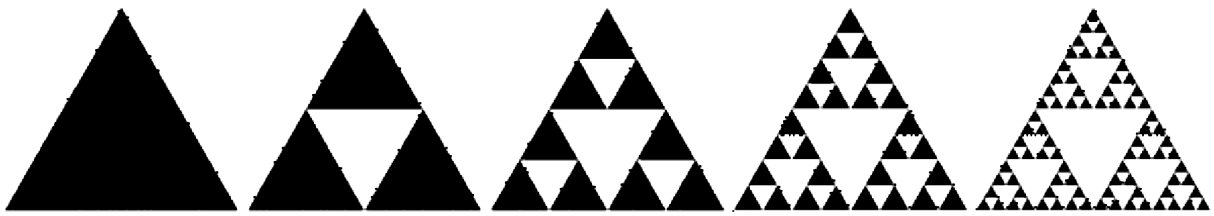


FIGURE. *From left to right, the Sierpiński gasket at stage* $n = 1, 2, 3, 4, 5$

Ex.: Iterated algorithm for constructing the Cantor middle-thirds set.

Step 1.	Begin with a line segment measuring 1 unit in length.
Step 2.	Divide the line segment into 3 identical "sub-segments."
Step 3.	Remove the center sub-segment.
Step 4.	For each sub-segment, repeat steps 2 through 4.

FIGURE. *From top to bottom, the Cantor set at stage* $n = 1, 2, 3, 4, 5, 6, 7$

In each of the three exactly self-similar fractal shapes we have just seen, many curious things would happen if it were possible to repeat the procedure "infinitely many times."[1]

It's been said of the Cantor set that, in the end, "*all that is left is dust*." To put it in slightly less poetic terms: *After infinitely many deletions have been carried out, the original line segment is so riddled with holes that there are no segments left intact.* A similar fact is true for the Menger sponge and the Sierpiński gasket. There is no (3-dimensional) "space" left in the sponge, nor is there any (2-dimensional space) left in the gasket. On the other hand, not *every* point is deleted—infinitely many points still remain[2] in the Cantor set at stage $n = \infty$.

Incidentally, Georg Cantor suffered a nervous breakdown, and after studying the set he discovered, decades of graduate students have nearly followed suit. You will *not* be tested on your understanding of the previous paragraph or the footnotes in this document.

[1] Infinity ∞ is a very slippery idea to think about. It's *not a number*: the symbol ∞ represents a quantity larger than every number on the number line. It is by no means obvious what it means to tell someone to "*Iterate the algorithm infinitely many times*," and the math used to justify the "existence" of the shape produced by infinitely many iterations (thus producing a thing called a "*direct limit of a sequence of spaces*") is far, far beyond the scope of this class.

[2] It is a somewhat mind-bending fact that *there are points still remaining in the Cantor set at stage* $n = \infty$ *which are not the endpoints of any sub-segment produced in any previous stage* $m < \infty$.

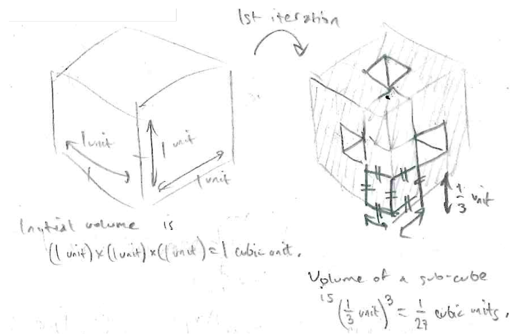
2. Measuring an exactly self-similar fractal after n iterations

Returning to the Menger sponge, we see that the shape sits in three-dimensional space. It makes sense to ask what its **volume** is: that's one way to measure the Menger sponge.

To answer this question, we will calculate the volume of the sponge at the 2nd stage, then the 3rd iteration, etc., and see if we can spot a pattern.

(The following are two suitable test questions.)

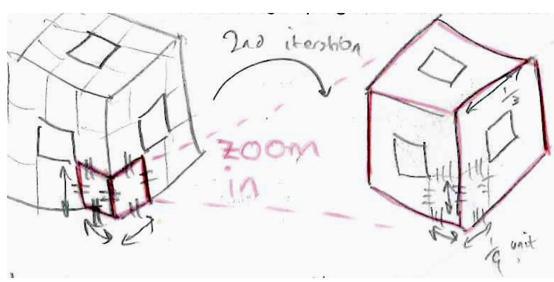
Ex. What is the volume of the Menger sponge after the 1st iteration? (This means the *second* stage: that is, after doing the algorithm *once*.)



Delete 7 sub-cubes. The remaining volume equals

$$
\begin{aligned}
\begin{pmatrix}\text{initial}\\ \text{volume}\end{pmatrix} - \begin{pmatrix}\text{deleted}\\ \text{volume}\end{pmatrix} &= 1 \text{ cubic unit} - \left(7 \times (\text{volume of one sub-cube})\right)\\
&= 1 - \left(7 \times \left(\tfrac{1}{3}\right)^3\right) \text{ cubic units}\\
&= \tfrac{27}{27} - \tfrac{7}{27} \text{ cubic units}\\
&= \tfrac{20}{27} \text{ cubic units.}
\end{aligned}
$$

Ex. What is the volume of the Menger sponge after the 2nd iteration? After the 3rd?



For each of the 20 sub-cubes in the original volume, delete 7 sub-cubes.

$$
\begin{aligned}
\begin{pmatrix}\text{initial}\\ \text{volume}\end{pmatrix} - \begin{pmatrix}\text{deleted}\\ \text{volume}\end{pmatrix} &= \tfrac{20}{27} \text{ cubic units} - 20 \times 7 \times (\text{volume of one sub-sub-cube})\\
&= \tfrac{20}{27} - 140 \times \tfrac{1}{729} \text{ cubic units}\\
&=
\end{aligned}
$$

	Initial volume	Deleted volume	Remaining volume
Stage 1 (1st iteration)	1	$7 \times \left(\frac{1}{3}\right)^3$	$1 - 7 \times \left(\frac{1}{3}\right)^3 = \frac{27}{27} - \frac{7}{27} = \frac{20}{27}$
Stage 2	$\frac{20}{27}$	$20 \times 7 \times \left(\frac{1}{9}\right)^3$	
Stage 3	$\frac{400}{729}$		

With each new iteration, what happens to the volume of the Menger sponge? It can in fact be shown that the volume tends toward 0. That is, if we had an infinite amount of time to build this thing, we would eventually end up with a shape that has a volume of 0: no "space" would be left inside the sponge.

Let's look at the Sierpiński gasket, and see if we can measure its *area*. (It makes sense to talk about its area, as opposed to its volume, because the Sierpiński gasket sits in two-dimensional space.)

(The following problem models the first several homework problems in this Chapter.)

Ex. Find the area of the Sierpiński gasket after 1 iteration, after 2 iterations, and after 3 iterations. For simplicity, suppose that the initial triangle has area 1.

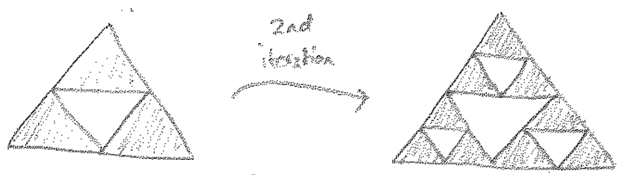
4 sub-triangles

$A = (\text{initial area})$.

Suppose $A = 1$ square unit

New area is $A - \frac{1}{4}A = 1 - \frac{1}{4}$ square units

$= \frac{3}{4}$ square units.



Each sub-triangle has area

$\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$ square units.

Initial area $= \frac{3}{4}$ square units.

New area is $\begin{pmatrix}\text{initial}\\ \text{area}\end{pmatrix} - \begin{pmatrix}\text{deleted}\\ \text{area}\end{pmatrix}$

$= \frac{3}{4} - 3 \times \frac{1}{16}$ square units

$= \frac{9}{16}$ square units.

Each sub-sub-triangle has area

$\left(\frac{1}{4}\right)^3 = \frac{1}{64}$ square units.

Delete 3 sub-sub-triangles from each of the 3 sub-triangles.

New area is $\begin{pmatrix}\text{initial}\\ \text{area}\end{pmatrix} - \begin{pmatrix}\text{deleted}\\ \text{area}\end{pmatrix}$

$= \frac{9}{16} - 3 \times 3 \times \frac{1}{64}$ square units

$= \frac{36}{64} - \frac{9}{64} = \frac{27}{64}$ square units.

Finally, we return to the Cantor set. What is the *total length* of all the segments which make up the Cantor set at stage n?

Ex. Find the total length of the Cantor set after each of the first five iterations.

(The following question goes a little bit beyond what you are expected to do on a test.)

Ex. Write a formula that gives the total length of the Cantor set after n iterations, no matter what n is. (*Hint*: Generalize the answers of the previous question by creating a formula that works for the first five iterations.)

Sections 12.2 and 12.3 will NOT be covered in this class. You may skip them in your reading.

3. Complex arithmetic

Reading: None. This section of the handout provides a review of complex numbers. It should be completed before proceeding to Section 12.4 in the textbook.

Algebraically,

$$i = \sqrt{-1}.$$

That is, when we do arithmetic with numbers like i, we stipulate to the axiom that

$$i^2 = -1.$$

In a sense, that's all i is: it's the number which, when you multiply by itself, you get -1: that is, $i \times i = -1$. However, this purely algebraic definition leaves us cold. It would be nice to have some geometric intuition, so that this strange "number" $i = \sqrt{-1}$ might be visualized.

Geometrically, multiplying a vector v by -1 yields the vector with the same length, but in the opposite direction. (Recall that we defined *vectors* in Ch. 11, Symmetry.) If we draw a point (call it the **origin**) and agree to draw *all* vectors with their feet at this point, we see that multiplying a vector v by -1 rotates v by $180°$ around the origin.

In symbols,

$$R_{\text{origin},180°} = \begin{pmatrix}\text{multiply} \\ \text{by } -1\end{pmatrix}.$$

We now ask: *How do you do half of a rotation by* $180°$? Geometrically, the answer is obvious: just rotate $90°$ around the origin. Doing this yields $R_{\text{origin},90°}$, and we know that doing this twice yields $R_{\text{origin},180°}$.

Algebraically, we begin with the fact that $R_{\text{origin},180°}$ is multiplication by -1. We now invent a number, call it i, and declare that multiplying by this number i *twice* is the same as multiplying by -1 (once). In symbols,

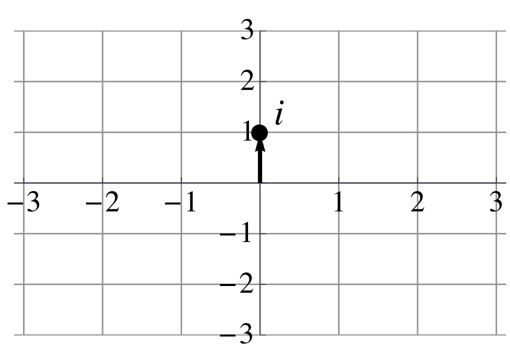
$$\begin{aligned} i^2 \times v &= (i \times i) \times v \\ &= -1 \times v \\ &= -v. \end{aligned}$$

That is, multiplying by i is half of a rotation by $180°$.

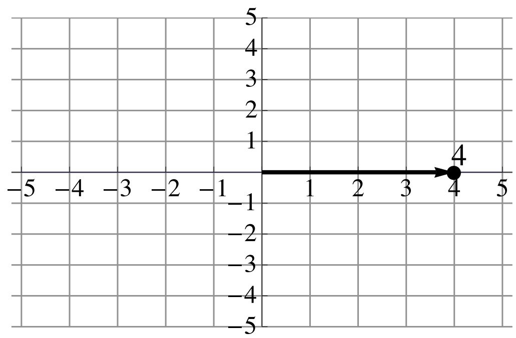
We have now explained *why* $i^2 = -1$. (Notice that $(-i)^2 = -1$ also.) But the question remains: *Where is* i*, really?* Can we "point" to it?

The answer is, Yes! *Every complex number, including* $i = \sqrt{-1}$*, is a point in the plane.* If you remember the analytic geometry you studied in high school algebra, you may regard i simply as the point $(0, 1)$ in the xy-plane. If you don't remember it, you have a slight advantage in that you can focus on the interpretation we will use: namely,

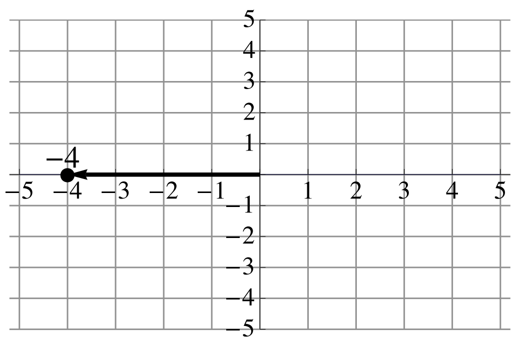
i is "the point you end up at when you travel 1 unit up from the origin."

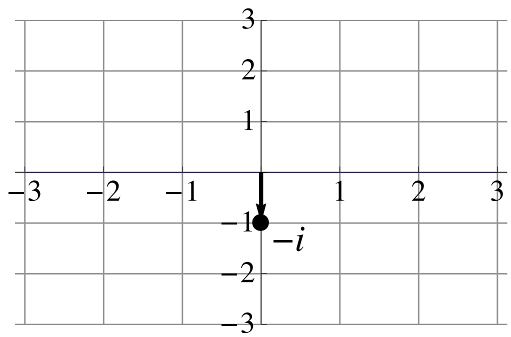


Similarly, we think of a positive number x as "the point you end up at when you travel x units to the *right* from the origin."[3]

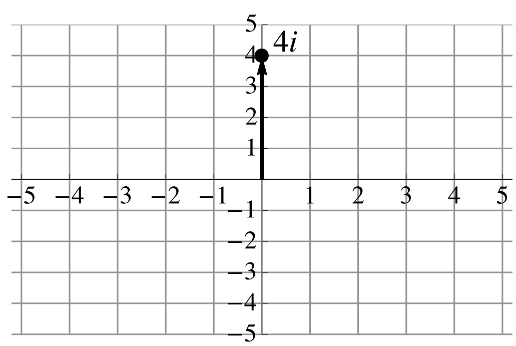


We see that multiplying by -1 rotates the given vector $180°$ around the origin,





and similarly, $i \times 4$ is the point you get when you rotate $v = 4$ *half* as much (that is, by $90°$).



You should now be able to draw a picture that shows where $-4i$ is on your own.

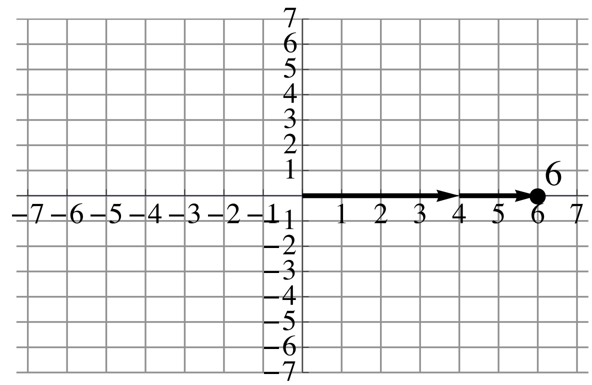
[3] Notice that we can describe each *point in the plane* in terms of a *vector with its foot at the origin*. Indeed, this is one way to define complex numbers: as such vectors.

We now want to define addition of these *points*, which we can think of as vectors with their feet at the origin. What should addition

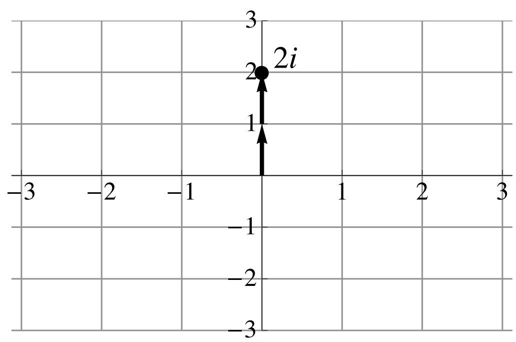
$$u + v$$

of two points u and v mean?

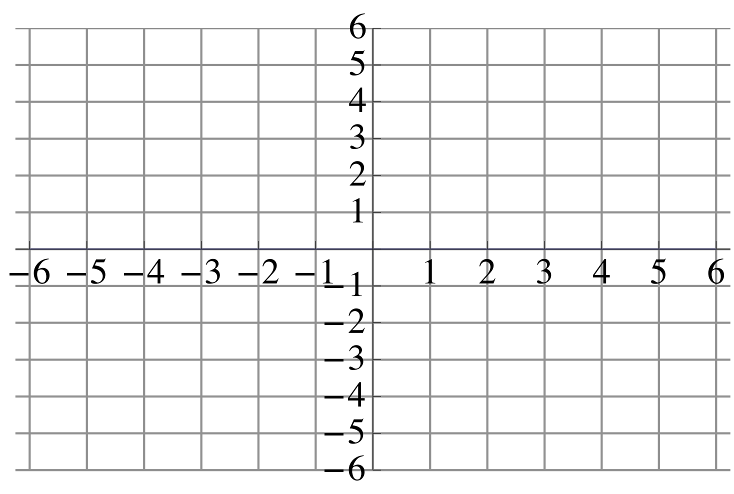
We certainly want our new "point addition" to be the same as ordinary addition when we add ordinary numbers like 3, 4, -7, and 0. We define addition by declaring that $u + v$ is the point you end up at when you start at the origin, follow the vector for u, and from there, follow the vector for v. For example, the addition $4 + 2$ can be visualized as follows.



Similarly, it makes sense that $i + i = 2i$ should be the vector "Go 2 units up from the origin," since i is the vector "Go 1 unit up from the origin."



Where is $-4 + 2i$?



In general, a number of the form

$$x + yi$$

(called a **complex number**) can be drawn as the vector, "Go x units to the right, and y units up," if x and y are real numbers, provided that we remember that if x is negative, we must instead go *left*, and if y is negative, we must instead go *down*.

We will call x the **horizontal coordinate**, and y the **vertical coordinate**[4] of the complex number $x + yi$.

To add two complex numbers, just add the horizontal coordinates to get the new horizontal coordinate, and then add the vertical coordinates to get the new vertical coordinate. For example,

Ex. Add $1 - i$ and $-6 + 3i$.

Ex. Describe the geometric effect of adding two vectors. (You might answer in the form, "Start at ______________ ...and end up at ______________.")

To multiply two complex numbers is a little more complicated. Let's say we are asked to multiply

$$(1 - 2i) \times (3 + i).$$

You may use ordinary algebra here (e.g.,the so-called F.O.I.L. method), or you may use a grid to organize your work. The grid is called the **area model for multiplication**, because each little box is filled in with the "area" of its sides.

	3	$+i$
1	3	$+i$
$-2i$	$-6i$	$-2i^2$

$$(1 - 2i) \times (3 + i) = 3 + i - 6i - 2i^2 = 3 - 5i - 2i^2 = 5 - 5i.$$

[4]The real numbers x and y are typically called the **real part** and the **imaginary part** of $z = x + yi$.

Ex. Demonstrate that multiplying by i rotates a given complex number $90°$ by calculating

$$(3-2i) \times i,$$

and demonstrate that multiplying by -1 rotates a given complex number $180°$ by calculating

$$(3-2i) \times (-1).$$

5. Film: Arthur C. Clarke's Colours of Infinity

A link to this video on YouTube appears on our D2L website in the News section.

6. Mandelbrot sequences

Reading: Section 12.4

We have seen that the formula

$$Z = z^2 + C$$

generates the Mandelbrot set. But how, exactly?

Recall that the picture of the Mandelbrot set we typically see assigns a color to each point in the plane. We will only use two colors: black, and colored (the particular shade doesn't matter, for our purposes).

To decide which color to give a certain point, we take that point as the **seed**: this means that we feed it into the above equation, and then iterate the equation many times. We then determine whether the numbers we get are either (*a*) getting further and further away from the origin (that is, **escaping**), (*b*) getting closer and closer (that is, **attracted**) to the origin, or (*c*) none of the above. If the numbers we get are escaping, the original number (the *seed*) is not in the Mandelbrot set.

To demonstrate the idea, we will take for a seed $C = -1$. This is our first term, so we call it s_1, and we begin filling out a table.

term	value of s_n
$n=1$	$s_1 = -1$

What's the next term? The formula discussed in Colours of Infinity says that we have to square and add C, so that's what we do:

$$(s_1)^2 = (-1)^2 = 1,$$
$$\text{so } (s_1)^2 + C = 1 + (-1) = 0.$$

Our result was 0, so we add this as the second term—call it s_2—to the table.

term	value of s_n
$n=1$	$s_1 = -1$
$n=2$	$s_2 = 0$

We might be tempted to stop here and say that, if we start with $C = -1$, we get closer to 0. But before we draw this conclusion, let's see what happens to the next term.

To get the 3rd term, we again use the formula $Z = z^2 + C$, squaring the previous term and adding C. For z, we use the previous term $s_2 = 0$, and squaring this gives 0. Now we add $C = -1$ to get the next term:

$$s_3 = (s_2)^2 + C = (0)^2 + (-1) = -1.$$

Adding this to the table, we see that we are *not* getting closer and closer to 0.

term	**value of s_n**
$n = 1$	$s_1 = -1$
$n = 2$	$s_2 = 0$
$n = 3$	$s_3 = -1$

In fact, if we continue "iterating"—that is, squaring and adding $C = -1$ to get the next term, again and again—we will quickly discover a simple pattern. The sequence of numbers alternates between two numbers, always in the same pattern: $-1, 0, -1, 0, -1, 0, \ldots$

Let's define a few vocabulary terms, so we can speak precisely about these ideas.

Choose a complex number C in the plane to be tested. The sequence of (infinitely many) numbers

$$s_1, s_2, s_3, \ldots,$$

is called the **Mandelbrot sequence with seed** C if the first term is $s_1 = C$, and every subsequent term s_N is obtained by applying the Mandelbrot formula

$$s_N = \left(s_{N-1}\right)^2 + C$$

(that is, to get the new term s_N, square the previous term, and add C to the result).

To draw the Mandelbrot set as a colored image, we test each point C in the plane by using it as the seed for a Mandelbrot sequence.

- If the Mandelbrot sequence gets further and further from the origin ($0 = 0 + 0i$), we call the sequence *escaping*, and the test point C is <u>not</u> in the Mandelbrot set. The points that are not in the Mandelbrot set are usually colored with a bright or a dull color depending on how *quickly* the sequence gets further and further from the origin.
- Otherwise, the test point C <u>is</u> in the Mandelbrot set, and we color it black.

Ex. Determine whether or not the Mandelbrot sequence with seed $C = 1$ is escaping.

Ex. Determine whether or not the Mandelbrot sequence with seed $C = -0.125 + 0.75i$ is escaping.

Ch. 15. Probability

1. Notation for sets

A **set** is a collection of objects of any type whatsoever: people, numbers, books, outcomes of experiments, geometrical figures, etc. Thus we can speak of the set of all integers, or the set of all oceans, or the set of all possible sums when two dice are rolled and the number of dots on the uppermost faces are added, or the set consisting of the cities of Austin and San Marcos and all their residents.

The objects in a set are called its **members**. We will often write a set using the following conventional notation.

List all the members of the set, and then surround the list with curly braces $\{\ \}$.

For example

$$\{1, 2, 3, 4, 5\}$$

is the set of the first five counting numbers.

If we want to give the set a name, we write (*symbol*) $= \{\dots\}$. For example, the following sentence says, "*The set* A *is the set of numbers* $1, 2, 3, 4,$ *and* 5."

$$A = \{1, 2, 3, 4, 5\}.$$

Compare the set A to the following set:

$$Z = \{1, 3, 5\}.$$

Clearly every member of Z is also a member of A. When this is true of two sets, we say that Z is a **subset** of A, and we write this in symbols as $Z \subset A$.

If we have two sets, call them X and Y, the **union** of the two sets is the set consisting of all members of X together with all members of Y. For example, if

$$X = \{1, 2, 3\}$$

and

$$Y = \{2, 4, 6\},$$

then

$$X \cup Y = \{1, 2, 3, 4, 6\}.$$

Notice that we do not repeat the number 2, although it shows up twice, once in X and once in Y: *in a set, we do not allow members to be repeated.*

(The symbol $\cup$, pronounced "*cup*," means "*the union of the two sets.*")

The set consisting of all objects that are members *both* of X and of Y is called the **intersection** (symbol: $\cap$, pronounced "*cap*") of the two sets:

$$X \cap Y = \{2\}.$$

Notice that it's possible for the intersection to be empty! If this happens, we say the sets are **disjoint** (or **mutually exclusive**). For example, find a pair of sets among the three sets X, Y, and Z (defined as above) which are disjoint.

When a set has no members, we write it by writing the list of members as usual (there aren't any, so it's a very short list!), and then surrounding the list with curly braces. We get

$$\{\ \}$$

a set with no members, and we call it the **empty set**.

You do not need to memorize the symbols $\subset$, $\cup$, and $\cap$. However, incorrect usage of curly braces $\{\ \}$ and the equals symbol $=$ is frowned upon.

2. Random experiments and sample spaces

The mathematical theory of probability describes and analyzes the behavior of an experiment whose outcome does not follow any discernible pattern. Probability theory is a special case of set theory.

A **random experiment** is an activity whose outcome cannot be predicted ahead of time.

Some examples of a random experiment:

- Do I win at blackjack if the values of cards in my hand add up to 18?
- If I fill a one-liter jug with river water, is the liter polluted?
- If a customer sues our company, do we win in court?
- If I flip a coin, which side lands face up?

The **sample space** of an experiment, which we denote by the letter $\mathcal{S}$ (cursive S), is the set of all possible outcomes of a given experiment, each of which can occur exactly one way.

For emphasis, we formally define an **outcome** as an outcome of a random experiment which can occur in only one way.

Ex.

Experiment:	I flip a penny and a nickel, and see which side landed face up on each coin.
Outcomes:	penny H and nickel H, penny H and nickel T, penny T and nickel H, penny T and nickel T
Sample space:	{penny H/nickel H, penny H/nickel T, penny T/nickel H, penny T/nickel T}

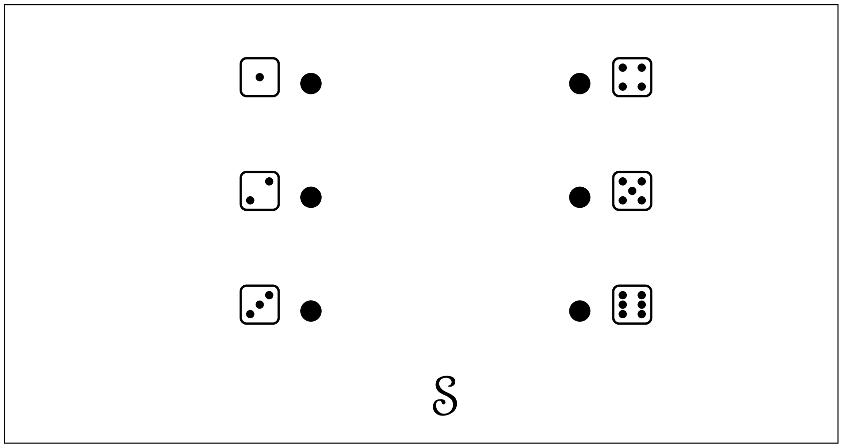
Notice in particular that, the "event" that "*one coin lands heads up and the other lands tails up*" is not an "outcome," since this event can happen in more than one way.

If E is a finite set, we write $|E|$ for the number of members of E. This number is called the **cardinality** of the set E.

Ex.

Experiment: I roll an ordinary,[1] fair[2] die.

Sample space: The sample space $\mathcal{S}$ is represented by the following "set diagram":



Number of outcomes in the sample space: $|\mathcal{S}| = 6$.

Ex.

Experiment: The Lakers trail the Jazz in a basketball game by a single point in the final second of overtime, when a player on the Lakers is fouled. The fouled player goes to the line to shoot two free throws, worth 1 point each. Which team wins the game, or is it a tie?

Sample space:

Number of outcomes in the sample space:

[1] By an **ordinary die**, we mean one with 6 sides labeled as follows: ⚀, ⚁, ⚂, ⚃, ⚄, ⚅

[2] A **fair die** is one which is equally likely to land on any of its sides, as opposed to an **unfair** or **loaded die**.

When a random experiment is repeated, each repetition of the experiment is called a **trial**.

A **Bernoulli experiment** is a random experiment which can have exactly two outcomes. We usually call its two outcomes **success** and **failure**.

The description of a Bernoulli experiment can always be rephrased as a yes-or-no question.

Some examples of a Bernoulli experiment:

- If our company sells a customer a one-year life insurance policy, do we have to pay out? (*success* = pay out, *failure* = the customer survives the year)
- If a mammogram comes back positive, does the patient really have cancer? (*success* = patient has cancer, *failure* = mammogram was a false positive)
- If I flip a coin, which side lands face up? (Rephrase this experiment as a yes-or-no question. Which outcome should you call a success?)

Ex.

Experiment: A hockey player shoots two free shots, but she only gets to take the second shot if she makes the first shot. Which shots does she make?

Sample space:

Number of outcomes in the sample space:

Ex.

Experiment: A gambler repeatedly plays roulette, doubling his bet each time, until he loses.
Sample space:

Number of outcomes in the sample space:

Ex.

Experiment: What numbers come up on two fair dice when rolled?

In the following table, each box represents a possible outcome of the two dice, one of which has been colored black. Thus the table represents the sample space: each empty box is an individual outcome.

Sample space:

	⚀	⚁	⚂	⚃	⚄	⚅
⚀						
⚁						
⚂						
⚃						
⚄						
⚅						

Note that we treat the two dice as *distinguishable* (e.g., one with white dots and one with black dots), so that the two outcomes ⚀⚂ and ⚀⚂ are different outcomes.

This method for visualizing the sample space extends in an obvious way to the experiment, "*Roll one die*":

⚀	⚁	⚂	⚃	⚄	⚅

We see that, when we roll two dice, there are   outcomes in the sample space, whereas when we roll one die, there are   outcomes in the sample space.

Ex.

How many outcomes will be in the sample space for the following experiment?

What numbers come up on four fair dice when rolled?

Ex.

How many outcomes will be in the sample space for the following experiment?

What comes up when nine coins are flipped?

Ex.

Dolores is a young saleswoman planning her next business trip. She packs three different pairs of shoes, four skirts, six blouses, and two jackets. Assuming that all these items of clothing match, how many different outfits can Dolores make?

To solve problems like this one, we use the **multiplication rule**, which states:

When a process is carried out one stage at a time,
the number of ways it can be done
is found by multiplying the number of ways each of the stages can be done.

Ex.

Five candidates are running for office in a club. The winner will become the President of the club, the runner-up will become Vice President, and the third-place candidate will become the Secretary.

The sample space of this election consists of how many different outcomes?

Ex.

Dolores packs 3 pairs of shoes, 4 skirts, 3 pairs of slacks, 6 blouses, 3 turtlenecks, and 2 jackets. As before, assume that the items are color-coordinated so that everything goes with everything else.

We define an "outfit" as consisting of a choice of shoes, a choice of "lower wear" (a skirt or a pair of slacks), a choice of "upper wear" (a blouse, or a turtleneck, or both), and a choice of whether or not to wear a jacket.

How many such outfits can Dolores make?

3. Counting for non-beginners: Permutations and combinations

For some experiments, the multiplication rule tells us how many outcomes are in the sample space. In fact, the multiplication rule is just one of many techniques for *counting*.

Of course none of us are complete beginners at counting. One technique which all of us have certainly mastered is the **brute force**, or **roster technique for counting**: to count the members of a set, list each member of the set exactly once (or draw a set diagram in which each member appears exactly once), and count the members by hand. This technique is always an option for us when we are faced with counting problems.

However, for the types of counting problems we will encounter, brute force is both tedious and error-prone. Both disadvantages stem from the fact that it is often quite hard to list (for example) all the outcomes of an experiment, unless one has a systematic way of listing them (such as a *tree diagram*).

We have already seen a second technique for counting: the *multiplication rule* tells how to count the number of ways a process with many stages can be completed, provided that we are given the number of ways to complete each stage.

The next example requires a third technique.

Ex.

Baskin-Robbins offers 31 flavors of ice cream. We will use the name **true double** to mean two scoops of ice cream *of two different flavors*. Count the number of true doubles available at Baskin-Robbins.

Now, the multiplication rule gives $31 \times 30 = 930$, but this is <u>not</u> correct. Explain why.

What is the actual number of true doubles available at Baskin-Robbins?

We take the example one step further.

Ex.

What is the actual number of true *triples* available at Baskin-Robbins?

We will answer this question by first answering two questions:

(A) How many choices are there for the first scoop, the second scoop, and the third scoop?

(B) How many choices of all three scoops result in the same bowl of ice cream?

(A)

Use the multiplication rule to count the number of ways to carry out the process:

(1) choose a flavor for the first scoop,

(2) choose a different flavor for the second scoop, and

(3) a third flavor (different from the first two) for the third scoop.

(B)

How many ways are there to put three flavors X, Y and Z in an order?

Use your answers to (A) and (B) to count the number of true triples available at Baskin-Robbins.
(There is a strong analogy here to the process of counting *distinct* Hamilton circuits!)

Ex.

How many ways are there to form a 3-person committee from a group of 6 people?

Ex.

A poker player is dealt five cards from a 52 card deck, one face down and four face up. How many different outcomes are possible?

Ex.

How many passwords of between 3 and 5 characters are possible using the characters $@$, $\$$, $*$, $\#$, and $\%$, assuming you are allowed to repeat a character more than once?

Ex.

Ten chairs are arranged in a circle. Count the number of different ways to seat 10 people in the 10 chairs.

Ex.

A college student must choose 3 out of the following 6 classes when registering for next semester's classes.

algebra	English	history
Spanish	chemistry	gym

Assuming that the college offers exactly one section of each class, count how many different schedules are possible.

We now summarize two of the different counting techniques we have developed over the course of working the above example.

An **arrangement of r slots** is a choice of r things from a set consisting of n different objects.

A **permutation** is an ordered arrangement. A **combination** is an unordered arrangement.

The number ${}_nP_r$ stands for,

the number of ways to fill r slots, in order, choosing from n different objects.

The number ${}_nC_r$ stands for,

the number of ways to fill r slots, without regard to order, choosing from n different objects.

Formulas for the numbers ${}_nC_r$ and ${}_nP_r$ appear in the book on page 519, but you are encouraged to instead memorize how to compute ${}_nP_r$, and then use the formula

$${}_nC_r = \frac{\text{(number of ordered arrangements of } r \text{ choices from } n \text{ objects)}}{\text{(number of different ways to put } r \text{ things in an order)}} = \frac{{}_nP_r}{r!}$$

when necessary.

What counting technique is used to compute ${}_nP_r$, the number of ways to choose r things in order from a pool of n objects?

Ex.

In the Massachusetts State Lottery, a ticketholder chooses 5 numbers from 1 through 55. The winning numbers are given in order, e.g. $17 - 29 - 36 - 53 - 55$. Count the number of ways ...

1. ... to choose 6 numbers.
2. ... to choose 6 numbers *in order*.

4. Pascal's triangle.

The figure below is known as **Pascal's triangle**. To find ${}_nC_r$, count down to the nth row (when counting, start at 0 as shown!). The rth hexagon in that row (again, start counting at 0) is the value of ${}_nC_r$.

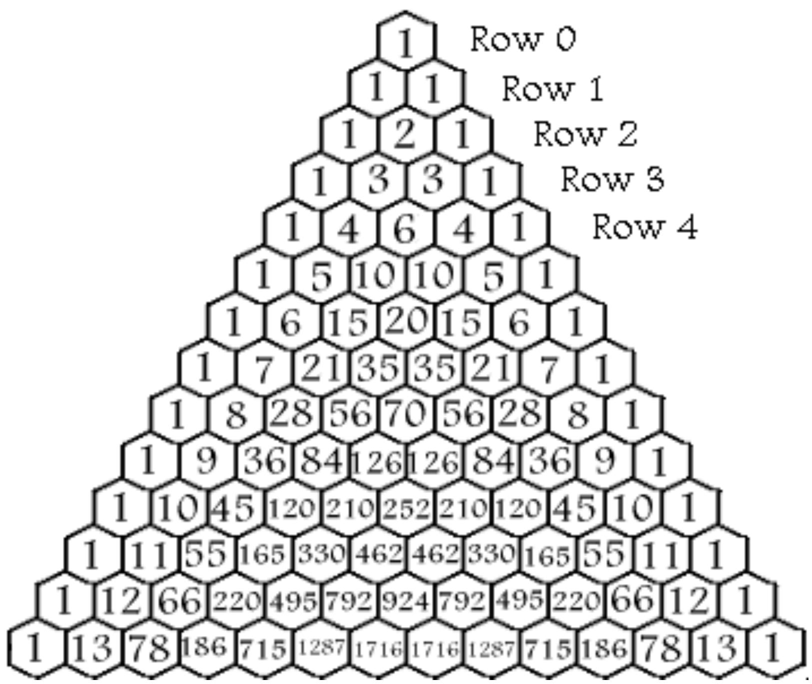
Ex.

$${}_3C_2 = \frac{3!}{2!\ 1!} = 3 = \frac{3!}{1!\ 2!} = {}_3C_1.$$

$${}_4C_2 = \frac{4!}{2!\ 2!} = \frac{4 \times 3 \times 2 \times 1}{(2 \times 1)(2 \times 1)} = \frac{24}{4} = 6.$$

Ex.

Given the fact that ${}_nC_r = {}_nP_r \div r!$, how can you use Pascal's triangle to find ${}_nP_r$ quickly? (By "quickly," I mean, "using one of the four basic arithmetic operations at most r times.")



If time allows, your instructor will give a detailed demonstration of how to construct Pascal's triangle. If not, consider the following verbal instructions.

The top row is called Row 0. It contains just one entry, a 1. The second row, Row 1, contains two hexagonal positions, and each is filled in with 1.

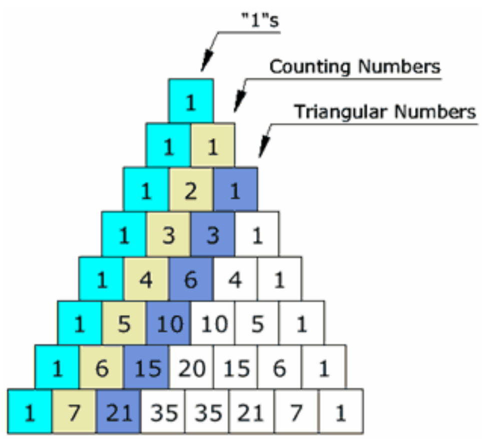
To fill out the next row, fill in the first and last hexagons with a 1, and for each of the other hexagons in the row, add the two entries immediately above. For example, in the middle hexagon on row 2, we add the 1 above and to the left of it, and the 1 above and to the right of it, to get 2. The process described in this paragraph may be repeated indefinitely.

The following is taken from the following website.

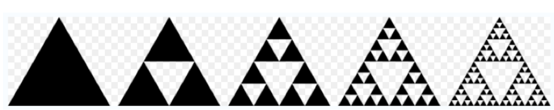
`http://people.bath.ac.uk/sjb37/patterns.html`

Many, many patterns emerge in Pascal's triangle. As shown, the number 1 is repeated along the leftmost diagonal of the triangle, the *counting numbers* run along the second-leftmost diagonal, and the *triangular numbers* appear on the diagonal after that.

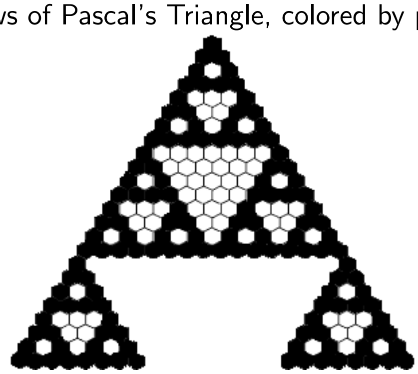
> "There are two definitions of triangular numbers, one informal one (but which explains the name) and one which is more formal. I will state the formal definition:
>
> A **triangular number** is a number obtained by adding all positive integers less than or equal to a given positive integer n. The triangular number
>
> $$T_n = n + (n-1) + (n-2) + (n-3) + \cdots + 2 + 1$$
>
> is therefore an additive analogue of the factorial
>
> $$n! = n \times (n-1) \times (n-2) \times (n-3) \times \cdots \times 2 \times 1."$$



As we write out more and more rows of Pascal's Triangle, coloring odd numbers black, and even numbers white, we get an increasingly more detailed picture of the fractal known as Sierpiński's Triangle.



Above: The first few iterations of Sierpiński's Triangle.
Below: The first several rows of Pascal's Triangle, colored by parity (i.e. even-or-oddness).



5. Probability spaces.

Like several of the branches of mathematics we have explored during this semester, probability theory is a relatively young science. The theory wasn't stated cohesively until the 1930s, when the mathematician A. N. Kolmogorov developed a formal system which for the first time made precise what exactly a "probability" is. We will not delve into Kolmogorov's work—to do so is beyond the scope of an undergraduate course—but we will explain some of the most important ideas incorporated in the formal system used nowadays for probability theory.

Suppose $\mathcal{S}$ and $\mathcal{T}$ are two sets. A **function** $\mathrm{P} : \mathcal{S} \to \mathcal{T}$ (pronounced: "a function P from S to T") is a rule which assigns to each member of S exactly one member of T. The rule may be specified by a verbal description, a table, or by an equation (if the sets involve numbers, which need not be the case).

For example, the following table defines a function which assigns a price to each food product sold at McDonald's. Here, S is the set of food products, and T is the set of prices.

food product X	**price** $\mathrm{P}(X)$
bacon cheeseburger	$\$2.79$
vanilla shake	$\$1.39$
apple pie	$\$0.99$
double cheeseburger	$\$0.99$

Our notation for functions is as follows. To say the following sentence in symbols,

"The price assigned to apple pie by the function P is $\$2.79$."

we write simply

$$\mathrm{P}(\text{apple pie}) = \$2.79,$$

placing the member X of $\mathcal{S}$ in parentheses, and writing the number assigned to X on the right hand side of the equals symbol.

Let $\mathcal{S}$ be the sample space for a random experiment. Each subset of $\mathcal{S}$ is called an **event**.

A **probability assignment** is a rule which assigns a number $\Pr(E)$ between 0 and 1 (where each of 0 and 1 is allowed) to each event E in the sample space.

To put it more tersely: a *probability assignment* is a function from the set of all events to the set of real numbers no less than 0 and no more than 1.

To keep all these vocabulary words straight, consider the following table, which lists equivalent ideas on the same row.

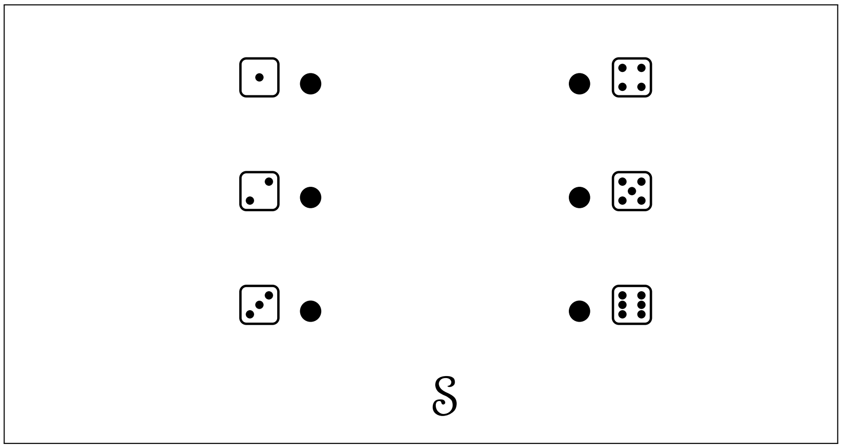
set theory	**probability theory**
universal set $\mathcal{S}$	sample space $\mathcal{S}$
subset of $\mathcal{S}$	event
member of $\mathcal{S}$	outcome

Recall that an outcome *can only happen in one way*—see the first example in section **2** of this document. The primary use of "events" is to collect all the different ways (i.e. outcomes) a thing can happen (e.g. "both coins land heads up").

Ex.
Suppose that one ordinary die is rolled. Identify the event E consisting of all outcomes in which an even number comes up. . .

(i) . . . by drawing a circle in the set diagram provided, and

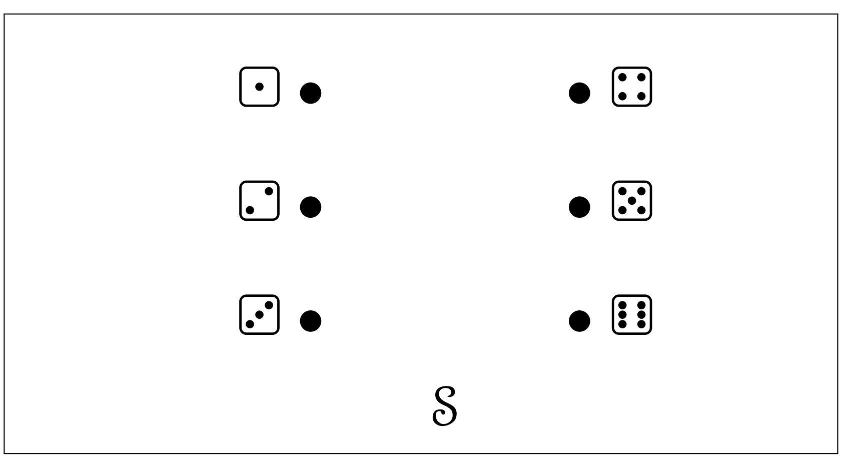
(ii) . . . by writing out the set using symbols only (that is, using correct set notation) in the space below the set diagram.



Ex.
Suppose that one ordinary die is rolled. Identify the event F consisting of all outcomes in which the number ⚂ does *not* come up. . .

(i) . . . by drawing a circle in the set diagram provided, and

(ii) . . . by writing out the set using symbols only (that is, using correct set notation) in the space below the set diagram.



Now, if the experiment is rolling an ordinary, *fair* die, then each side of the die is equally likely to come up. It follows[3] that the probabilities of rolling each side are given by the following table.

event E	**probability** $\Pr(E)$
{⚀}	$1/6$
{⚁}	$1/6$
{⚂}	$1/6$
{⚃}	$1/6$
{⚄}	$1/6$
{⚅}	$1/6$

A **probability space** is two things:

- . . . a *sample space* $\mathcal{S}$, and
- . . . a probability assignment, i.e. a rule which assigns to each subset E of $\mathcal{S}$ a number called the **probability** $\Pr(E)$ **of event** E, which must be no less than $0 = 0\%$ and no more than $1 = 100\%$.

Okay, but how do we find the probability $\Pr(E)$ for a given event? We define the **probability of an event** E to be the sum of the probabilities of each outcome in E.

We see that, for the experiment of rolling one ordinary, fair die, the sample space is just the set $\mathcal{S} =$ {⚀, ⚁, ⚂, ⚃, ⚄, ⚅}, and the probability assignment determined by the above table.

Ex. In the probability space described above, what is the probability of the event {⚁, ⚄}.

[3]We will not study the "axioms" of probability theory, but it is from these axioms that the conclusion in the footnoted sentence follows. We explain: One important axiom is that the probability of the sample space $\mathcal{S}$ itself is $1 = 100\%$. If there are exactly 6 outcomes in the sample space, and each is equally likely, it must be the case that each outcome has probability $1/6 = 16\frac{2}{3}\%$, because an outcome can happen in only one way, and because of another axiom, namely, that the sum of the probabilities of each individual outcome must total $1 = 100\%$.

Instructions: The experiment for exercises #1–#5 is rolling two 6-sided dice. For each event, first shade the boxes which correspond to the given event. Then find the probability of the given event. *Note*: "One" means "at least one."

#1.
$A =$ set of outcomes in which the total number of dots on the sides facing up is 10.

	⚀	⚁	⚂	⚃	⚄	⚅
⚀						
⚁						
⚂						
⚃						
⚄						
⚅						

$\Pr(A) =$

#2.
$B =$ set of outcomes in which (at least) one die comes up 3.

	⚀	⚁	⚂	⚃	⚄	⚅
⚀						
⚁						
⚂						
⚃						
⚄						
⚅						

$\Pr(B) =$

#3.
$C =$ set of outcomes in which one die comes up 3, and one die comes up even.

	⚀	⚁	⚂	⚃	⚄	⚅
⚀						
⚁						
⚂						
⚃						
⚄						
⚅						

$\Pr(C) =$

#4.
$D =$ set of outcomes in which one die comes up odd, and one die comes up as a number divisible by 3.

	⚀	⚁	⚂	⚃	⚄	⚅
⚀						
⚁						
⚂						
⚃						
⚄						
⚅						

$\Pr(D) =$

#5.
$G =$ set of outcomes in which neither die comes up 3.

	⚀	⚁	⚂	⚃	⚄	⚅
⚀						
⚁						
⚂						
⚃						
⚄						
⚅						

$\Pr(G) =$

The situation is quite different when the outcomes of an experiment are *not* equally likely.

For example, Steve Nash is exceptionally good at shooting free throws. Over the course of his career, he has successfully made 90% of his free throws. Here, the sample space for the experiment of taking one free throw is

$$\mathcal{S} = \{\text{succeeds to make the free throw}, \text{fails to make the free throw}\},$$

but the probability of success is *not* $1/2$. (What *is* the probability of success? What is the probability of failure?)

Ex.
Consider the following experiment:

Amy buys a single $\$1$ *lottery ticket.*

Suppose that the probabilities of each possible outcome are given by the following table.

event E	**probability** $\Pr(E)$
win one free ticket	$1/6$
win $\$10$	$1/50$
win $\$1,000$	$1/5,000$
win $\$1,000,000$	$1/5,000,000$

(a) What is the probability that Amy wins more than $\$300$?

(b) What is the probability that Amy does not win anything? (Give your answer as a decimal.)

The examples we have seen suggest the following definition and formula.

The set B consisting of all members of $\mathcal{S}$ *except* the members of a given event A is called the **complement** of A. We say that A and B are **complementary**.

If A is an event, and B is the complement of A, then $\Pr(B) = 1 - \Pr(A)$.