Numbers and Stuff

Shmuel Weinberger

Department of Mathematics
University of Chicago

shmuel@math.uchicago.edu

In memory of DJ Newman
my first Mathematics teacher
the author of many proofs from The Book

WARNING: This is a draft. It will change

Table of Contents

4. Algebra: Number Systems.
    1. A few elementary observations

    Appendix: Two recent theorems

    2. Commutative groups
    3. Subgroups, quotients, and isomorphism.

    Appendix: Some applications

    Appendix: Noncommutative groups and symmetry.

    4. Rings (Fund thm of arithmetic revisited, chinese remainder theorem, formal power series and combinatorics)
    5. Polynomials and roots(and the number of solutions in various rings. In particular some quadratic polynomials)

5. Topics in the theory of real numbers.
    1. Rational numbers.
    2. How to tell if an algebraic number is rational
    3. Transcendental numbers
    4. Real numbers, what are they really?
    5. Odd degree polynomials have real roots.
    7. Appendix: An arithmetic proof of the impossibility of angle trisection.

Introduction: What is Mathematics?

No doubt you have an ideas about that. After all, you studied a lot about numbers and how to manipulate them in elementary school. In high school, you might have learnt how to solve equations. Even geometry seemed to involve numbers via lengths and angles and areas (although after you think about it, you will see that a lot of the focus there really wasn't numerical). So, you might be tempted to think Math is the study of numbers.

It is true that mathematicians study numbers and a lot of our interest will be in different *types* of numbers, and their properties. But that doesn't really define the essence of mathematics.

Before giving you my answer, I will give you G.H. Hardy's answer. Hardy was a great English mathematician from the first half of the last century whose great love outside of mathematics was cricket. He wrote a very interesting and somewhat controversial essay, "A Mathematician's Apology" making the case for mathematics as a (cerebral and austere, to be sure) art: mathematical creations, for Hardy, should be judged for their beauty, not their utility.

I highly recommend reading the apology – although sexist and frankly elitist (with a whiff of tragic in the way that athletes who become coaches sometimes seem), it is elegantly written, thought provoking, and inspiring (in the ways that exhortations to greatness can be, when done well). It is available for free at http://www.math.ualberta.ca/mss/ .

> The fact is that there are few more 'popular' subjects than mathematics. Most people have some appreciation of mathematics, just as most people can enjoy a pleasant tune; and there are probably more people really interested in mathematics than in music. Appearances suggest the contrary, but there are easy explanations. Music can be used to stimulate mass emotion, while mathematics cannot; and musical incapacity is recognized (no doubt rightly) as mildly discreditable, whereas most people are so frightened of the name of mathematics that they are ready, quite unaffectedly, to exaggerate their own mathematical stupidity.
>
> A very little reflection is enough to expose the absurdity of the 'literary superstition'. There are masses of chess-players in every civilized country—in Russia, almost the whole educated population; and every chess-player can recognize and appreciate a 'beautiful' game or problem. Yet a chess problem is simply an exercise in pure mathematics (a game not entirely, since psychology also plays a part), and everyone who

> calls a problem 'beautiful' is applauding mathematical beauty, even if it is a beauty of a comparatively lowly kind. Chess problems are the hymn-tunes of mathematics.

He goes on a couple of pages later

> A chess problem is genuine mathematics, but it is in some way 'trivial' mathematics. However ingenious and intricate, however original and surprising the moves, there is something essential lacking. Chess problems are unimportant. The best mathematics is serious as well as beautiful—'important' if you like, but the word is very ambiguous, and 'serious' expresses what I mean much better.

I will now explain the two examples that Hardy gives, which capture the seriousness of math as well as its beauty, and then I'll return to a simple problem that explicates the way in which "a chess problem is genuine mathematics" and also captures another important theme about the spirit of mathematics.

The first example is that the theorem that there are an infinite number of *prime numbers*. A prime number is a whole number, such as 2 or 3 which aren't divisible by any other number than 1. They are distinguished from numbers, such as 4, 6 and 9 that are *composite* – i.e. composed of other numbers, 4 = 2*2, 6 =2*3, 9 =3*3. (We don't count 1 as a prime because it doesn't help in "breaking up" other numbers into factors.)

Theorem: There is no "last prime number".

In other words, you can start listing them 2,3,5,7, 11… but you'll never finish. After any particular prime, you'll find another. This theorem is found in Euclid's "Elements", and we will follow his discussion.

Notice that after the number 1, every number is either prime or breaks up into a product of smaller numbers, and those numbers break up into a product of primes. So they are all prime or divisible by a prime (i.e. can be divided without remainder into a prime number of pieces).

Imagine a list of all the primes there are. Now consider the number N which is 1 more than the product of all of these primes. N is not divisible by any of those primes. (If p is one of those primes, N is 1 more than a multiple of p, so cannot itself be a multiple of p.)

So either N itself is prime or it is divisible by some prime, which cannot be on the original list.

Which means that the list you wrote down containing what you thought were all the primes was missing something.

Later on, we will phrase this theorem as saying that the set of prime numbers is *infinite*.

Notice the proof doesn't tell us what the missing prime is. The logic of the proof is "proof by contradiction". We discover that something is true because the alternative is inconsistent. Like Sherlock Holmes said "When you have eliminated the impossible, whatever remains, however improbable, must be the truth."

The primes are fairly common to begin with, 2,3,5,and 7 are the primes below 10 (40%), but they become more and more sparse: there are only 25 less than 100 (25%), and 168 less than 1000 (16.8%); by 100,000 there are less than 10% primes, and at a billion it's around 5%. They become more rare, but they don't ever run out.

After all, we saw that them ever running out is impossible – otherwise there'd be a comprehensive list of all of them and we saw that no (finite) list (that can be multiplied together) could ever be comprehensive.

As Hardy said, "The proof is by reductio ad absurdum, and reductio ad absurdum, which Euclid loved so much, is one of a mathematician's finest weapons5. It is a far finer gambit than any chess gambit: a chess player may offer the sacrifice of a pawn or even a piece, but a mathematician offers the game."

Hardy's other example is also known to the Greeks. It is the irrationality of $\sqrt{2}$.

Theorem: $\sqrt{2}$ is an irrational number.

This means that $\sqrt{2}$ is not a fraction a/b where a and b are whole numbers. (We don't distinguish between proper and improper fractions: 3/2 is acceptable to us.)

This was very upsetting to the Pythagoreans (the followers of Pythagorus, the gentleman credited with the theorem about the length of the hypotenuse of a right triangle) and legend has it that they killed the person who leaked this theorem to the press.

Indeed, since $\sqrt{2}$ is the length of the diagonal of a unit square (or the hypotenuse of a unit equilateral right triangle), it should be some kind of number, but the Pythogoreans only knew about fractions and couldn't fathom any other kind of number. (Probably you think about $\sqrt{2}$ as a decimal 1.4142135… and know the … goes on forever. Later we will parse what kind of number such a thing is.)

Here is the proof. As before, we will argue by contradiction (reductio ad absurdum). We might as well assume that a and b don't have any common factor. (This is the process of reducing a fraction to its lowest form, where low means the size of the denominator.)

If $a/b = \sqrt{2}$, then $a^2=2b^2$. This means that a is even, since a product two numbers (here both being a) can be even only if at least one of them is. In other words, there is a c so that $a = 2c$. Then $(2c)^2=2b^2$, so $4c^2 = 2b^2$, and therefore $b^2=2c^2$ and b is even as well. In yet other words, a and b have 2 as a common factor.

For those who want a reminder about why the product of odd numbers is odd (it is not *just* an empirical fact) recall that odd numbers are those which leave a remainder of 1 when divided by 2. So we can think of an odd number as being of the form 2m+1. The product of two odd numbers

$$(2n+1)(2m+1) = 4nm +2n + 2m +1 = 2(2mn+n+m) + 1$$

leaves a remainder of 1 when divided by 2, so is odd.

A similar but more complicated argument shows that a product of two numbers is divisible by 3 only if (at least) one of them is divisible by 3. This enables a proof that $\sqrt{3}$ is irrational.

We can have a product of two numbers not divisible by 4 divisible by 4, and indeed $\sqrt{4}$ is not irrational. However, 6*2 = 12 is a product of two numbers not divisible by 12 which is itself divisible by 12, et $\sqrt{12}$ is irrational (it is $2\sqrt{3}$, which must be irrational, as $\sqrt{3}$ is). Explaining the pattern of these observations will occupy us a great deal in chapter 3. For now, we can certainly admire the simplicity of this observation, and the profoundness of its implications – it goes to the very meaning of the idea of a number.

<u>A problem</u>.

Going in a different direction than Hardy, let's discuss something much less serious – a problem on a chess board.

Here's the problem: Suppose that you have an 8 x 8 chess board, and you'd like to cover it using 2 x 1 dominoes. It's easy enough to do so. You can lay all of them horizontally or vertically. If we were pushing to the numerical, I might ask you how many ways are there to do this. There 65,536 ways of doing this just dividing the 8x8 board into 16 2x2 regions and deciding independently of one another whether to cover each region with a pair of vertical or a pair of horizontal dominoes. We will later see that there are over a million possibilities. (There are actually exactly 12,988,816 many ways of doing this.)

But suppose we remove just the bottom left square, then it is impossible. There will be 63 squares left, but each domino covers two spaces, so any time we place down dominoes, we must cover an even number of squares. So whatever we do, we will have to have at least one square left out.

Now, suppose we take out two squares, say the bottom left and the top right. Is it now possible?

There will be 62 squares left, an even number. So we can't just argue in terms of odd and even as we did before and conclude that it's impossible.

In fact, if we remove the bottom left and the top left then it is easy to do (and the number of squares is the same as if we had removed bottom left and top right).

Try it.

Here's my first try on the diagonal problem ; it doesn't work.

INSERT HERE A CHESSBOARD W/FAILED ATTEMPT

Indeed, we have **two** left over.

Before jumping to see the solution I'll present below, you should try different things. After trying to cover this area unsuccessfully a number of times, you might come to the conclusion that it's impossible. But, there are so many possibilities, how can you be sure? Maybe it is possible, but just hard to find, like in many other puzzles?

Let's try to do the same problem on a smaller checkerboard. **If you cannot solve a problem, look for a simpler (and hopefully related) problem you cannot solve. You might learn something by trying the simpler one first.**

Now 7 x 7 is smaller, but when we try this one, we see that odd/even is enough to solve this case. 6 x 6 is smaller and still even, but also still too big to understand.

Let's go to the extreme 2 x 2. Here when we remove those squares there are two left and we can see that it's impossible. We might not have a good idea why. But at least we now believe it **conceivable** that odd/even is not a strong enough tool to solve this kind of problem.

What happens when we try 4 x 4? Alas there still seem to be many possibilities and exhaustive search doesn't work. But, nevertheless we try a few to see if we see some pattern.

If you do this problem on a checkerboard, then maybe you'll be led to the solution in the coming section.

2. An Impossibility Result.

Whenever we are proud of a result, we call it a theorem. (There are other words such as (in their plural forms) propositions, lemmas (lemmata if you are fancy), and corollaries that are also used for statements that we prove --- and it is a matter of taste what one calls the result, but since this is the first statement we are making that wasn't known 2000 years ago, we must call it a theorem.)

Theorem. It is impossible to tile the 8 x 8 chessboard using 2 x 1 tiles so that only the bottom left and top right corners are uncovered.

Proof. Notice that we place a domino on the chessboard, not only is it the case that two squares are covered, but indeed, it is always one black and one white square.

Consequently, any part of the chessboard that can be covered by dominoes must have an equal number of white and black spaces.

Finally, observe that on the part of the chessboard that we are considering there are 32 white squares and 30 black, so no matter what we do, there must be at least two (extra) whites uncovered. QED

(QED is short for Latin meaning "the proof is done; it's time to celebrate.")

Before proceeding, let me clarify why what we did was Mathematics.

Firstly, the objects we were dealing with weren't really physical: they were *abstractions*. The only things about physical chessboards or dominoes that are relevant are properties about the relations of squares to one another. It isn't that we *experiment* on one chessboard and then *generalize* to another one, and *hypothesize* that the results still apply (and *apply for grants* to *test* this).

Secondly, although the problem made perfect sense on an unpainted chessboard, its solution made use of the usual coloring of the chessboard into black and white squares. This coloring can be thought of as an *additional structure* imposed on the problem. With this structure in place, we have an easy solution.

Thirdly, the structure, at least in this case, is formally unnecessary for solving the problem. It is possible to do an exhaustive search (and much easier for us thanks to computers than it once was). However, if one does it by computer, then one will be much less confident about the generalization from 8 x 8 boards to, say, 10000 x 10000 boards (which are beyond the computational capacity of current computers).

The mathematician cares about her problems --- but she also has a soul that loves the beauty in its solutions. She is always on the search for *deeper structure* that governs phenomena from hidden realms. The mathematician is the person who sees a problem like this and gets the idea that painting the 8x8 grid in a particular way can be used for analyzing domino problems.

Another thing that the mathematician will do when presented with the previous state of affairs is that she will find it irresistible to probe further and look for problems where the white/black trick doesn't work but there still seems to be a phenomenon.

Problem: Is it possible to tile a 10 x 10 checkerboard using 4 x 1 dominoes? Note that in a 10 x 10 checkerboard there are an equal number of black and white squares: 50 of each.

Every 4 x 1 domino covers an equal even number of black and white squares, So, it looks numerically possible,

Hmm. We can't eliminate it just using the black and white method.

We can try varying the problem. Instead of considering just the 10 x 10 board, we can consider an m x n board. Here m and n are just letters that stand for various numbers.

Note that the product mn must be divisible by 4, because each 4 x 1 domino covers four squares.

Is this enough? For m=1 it is. For m =1, the product mn = n and is divisible by 4 exactly when n is, and when n is, we can cover this "rectangle" by n/4 dominoes.

But, what about for m=2?

Then mn is divisible by 4 exactly when n is even. The first case of interest is the 2 x 2 case, but obviously no (4 x 1) domino fits in there.

Indeed if m = 2 then all of the dominoes must be put in horizontally, and therefore, n must be divisible by 4.

Thus we might be led to:

Conjecture: The m x n rectangle can be covered by 1 x 4 dominoes exactly if m or n (or both) is divisible by 4.

A conjecture is just something that we call attention to as a problem, phrased in a provocative way intended to prejudice which way the answer will be. In English, the word "conjecture" is often preceded by the adjective "mere" and is intended to connote speculation. Mathematicians often conjecture on the basis of intuitions and years of thought about a problem. But, conjectures are sometimes true and often false.

To show that a conjecture is true, one must verify it by proof, but to disprove it, it often suffices to just give a counterexample. For example, in this case, if we found a 6 x 18 rectangle that can be covered by 1 x 4 tiles, then we will have disproved the conjecture.

But you'll soon see that the conjecture is true and we won't find a way to tile the 6 x 18 board.

To verify (prove) our conjecture, we cannot merely make use of the structure that sufficed in the 2 x 1 domino problem we considered before. It does not seem strong enough. However, thinking as mathematicians, we wonder, what was it about 2 x 1 that enabled a black/white structure to be so useful. Maybe there's another structure that is appropriate to solving problems about 1 x 4 dominoes.

3 colors (say Red, White, and Blue) don't seem helpful, because always some color will be left out in each domino, and we don't seem to have control over which one is missed.

So, let's try using 4 colors. We can use whichever ones we want, but as a serious professional, I will call them A,B,C, and D. If I paint the rectangle as follows, then every placement of a domino contains exactly one A, one B, one C and one D.

ABCDABCDABCDABCDA….
BCDABCDABCDABCDAB….
CDABCDABCDABCDABC….
DABCDABCDABCDABCD….
ABCDABCDABCDABCDA….

At this point, we have enough structure to prove the conjecture, although maybe we don't have enough computational facility with the structure to just see how many each of A,B,C, and D we have in an m x n chessboard.

Let's check what happens for the 10 x 10 chessboard whose upper left corner is painted A. Of course, one can just do it and count: There are 25 A's 26 B's 25 C's and 24 D's. They are not all equal.

But, is there some way that we can do this with less calculating?

Of course, we can use the dominoes! Any time we cover an area by dominoes, that region has an equal number of A,B,C,D's.

We can remove such a region without changing the balance of the four colors.

Since a 4 x 10 can be covered by 10 4 x 1's, we can remove that region from our chessboard, leaving a 6 x 10 region.

And if it works once, it'll work twice (if only real life were like this), so we can cut down to 2 x 10.

Now let's remove 2 x 4 twice, to cut down our "chessboard" to a manageable 2 x 2. That is clearly colored as below:

AB
BC

So we clearly have one more A and C than D and two more B's. This explains the count we had made 'by hand" before.

Problem: Prove the conjecture now yourself by assembling all of our observations.

Related Problems: Which rectangles can be covered by 2 x 2 squares? By 3 x 3 squares? By 2 x 3 rectangles?

Problem: Suppose you want to tile 1 x n rectangles, but now you are allowed to use two kinds of dominoes, 1x2 and 1x3. (1) For which n is it possible? (2) What about if you can use 1x3 and 1x4 dominoes? (3) 1x3 and 1x5? (4) 1x4 and 1x6? (5) Can you discern a pattern?

3. Impossibility results? Impossible!

People commonly say things like "It is impossible to prove a negative" (although obviously they could never prove such a thing)[1] or, in a more cautionary vein, "while many were saying that human flight is impossible, the Wright brothers went out and built a plane[2]".

Well, we proved a negative. No one can cover the chessboards that we proved were impossible to be covered.

The irrationality of $\sqrt{2}$ was also a negative. We showed that there were no solutions to a particular equation ($a^2 = 2b^2$) in integers.

We haven't accomplished the impossible, It really wasn't impossible. Indeed we showed you can sometimes prove a negative by going ahead a proving a negative.

In fact, we will prove many negatives as we continue our explorations.

Later on, with due deference to my friends from Missouri[3], we will also demolish the prejudice that the only way that one can be sure that something exists is by producing it.

---

[1] Listen to the Beatles: "There's nothing you can do that can't be done (OK) nothing you can see that can't be sung (OK) ….There's nothing you can know that isn't known (Huh?!?)" All you need is love.

[2] It's not really true that no one besides the Wright brothers believed that flight was possible. For some interesting history, see

[3] The "Show Me" state.

Words.

1. The Need for Precision.

August 1972

Dear Mom,

If you heard that there was a fire in our camp, you must be worried. I want to let you know that no one died and the number of missing kids is small – less than five, I'm sure. Every search party was successful, according to its own standards, and none missed more than a kid or two. I am very proud of my counselor.

I also want to say that camp is great and I am having a good time, but any camp that loses even a few kids is not the kind of place that one can feel safe in.

Please send me money!

Love

Shmuel

Now, there was no fire in the camp --- yet, not one sentence in this letter home was false. (My mother was not fooled; she knew that there was some kind of trick but sent me some spending money anyway.)

Before we really get to business, it is important to understand how mathematicians speak. First of all, we care a lot about "truth", the opposite of which is "falsity". We will do our best to make unambiguous, yet meaningful, sentences, that are either true or false. Nothing can be both true and false[4]. If a sentence is both, it has two meanings: a true one and a false one.

Let me begin by putting together sentences and seeing how they combine.

---

[4] This is true of declarative sentences, but not of exhortations. Look both ways before crossing. Be bold; if you are too careful, you will miss your opportunity. I tell my students to "Ignore what everyone else did on this problem – it's unsolved and needs a new idea." But will also tell them "Don't be so arrogant. A lot of smart people worked on this and didn't solve it --- you would do well by studying what they did and see where they made progress and where they were stymied."

NOT: Not A is true means that A is false. To say that "This book is not interesting" is exactly to deny that "This book is interesting". When asked, have you stopped cheating, an ever honest person has no choice but to respond "No". Alas, people can be come

AND. When we say that A and B are true, I mean that both A is true and B is true.
To say that dinner was delicious and not fattening, means that dinner was both delicious and it was not fattening.

You have to watch out when people say that Goebbels was a good Nazi. It looks like they mean to say he was good and he was a Nazi. But, what they mean was that he was good at Naziness.

OR. When we say that A or B is true, then we mean that A is true or B is true. It could be that both are true. So, for instance if A is true, then no matter what B is, A or B is true.

In common conversation, people sometimes use "or" as "exclusive or". We'd feel that someone who said "Shape up or ship out! " and fired an employee who then "shaped up" was being unfair. It is implicit that the speaker was indicating A or B but not both. WE NEVER MEAN EXCLUSIVE OR UNLESS WE SAY SO EXPLICITLY.

Sorry for all those capitals. I just needed to be emphatic.

If I say that "Being the best candidate will get you elected", am I committed to accepting the sentence that "Being the best candidate or the biggest buffoon will get you elected"?

No, not really. What went wrong? Let's look at "implication"

IMPLIES (or if … then …) We say that A implies B, if whenever A is true, B is true. Consider "If you give your teacher an apple every day, then you will pass the course." I can only complain if I don't pass having given my teacher an apple every day. Of course, I could pass if I learn the material as well and not give the teacher any fruit.

It is sometimes useful to write down a "truth table" to figure out what it would mean for a complicated sentence to be true. Here it is easier to give some examples rather than a definition.

A	B	Not A	A and B	A or B	A implies B

		-A	A&B	A∨B	A → B
T	T	F	T	T	T
T	F	F	F	T	F
F	T	T	F	T	T
F	F	T	F	F	T

On the top line we wrote in words the operations, and on the second line, we wrote a symbolic form of the sentence above. So, when you are being cross examined by this scary policeman on the road and he asks, "Did you drink before driving or did you not?", you produce a truth table: A = I drank before driving and –A = I did not drink then. The policeman seems to be asking me A ∨ –A. I produce a table like the following

A	-A	A ∨ –A
T	F	T
F	T	T

And, therefore, you must answer "True, sir[5]."

Of course, what you mean, and I am sure the defense will point this out, that it is true that you either drank or did not drink before driving. It might have been wiser, then, to answer a different question (or not answer any questions without a lawyer's advice, because answering tricky questions in stressful circumstances is probably wise to be avoided).

(One of the basic rules of practical politics is that if you are asked a question you don't want to answer, answer a question that you do.)

Let us now think about the "being the best candidate or biggest buffoon" example. (You should imagine we are discussing the situation in a good functional democracy.)

Let A = "You are the best candidate". And let B = "You are the biggest buffoon". Does A → A ∨ B?

A	B	A∨B	"A → A ∨ B"

[5] Policemen seem to like being called sir, in my experience.

T	T	T	T
T	F	T	T
F	T	T	T
F	F	F	T

So, no matter what, A does imply A$\vee$B. Let's now think about the proposition that (A $\rightarrow$ C) $\rightarrow$ (A$\vee$B$\rightarrow$C). Is that necessarily true?

Here, C is the statement that "You will be elected". We believe that A $\rightarrow$ C, i.e. that if you are the best candidate, you will be elected.

A	B	C	(A $\rightarrow$ C)	A$\vee$B	(A$\vee$B$\rightarrow$C)	(A $\rightarrow$ C) $\rightarrow$ (A$\vee$B$\rightarrow$C)
T	T	T	T	T	T	T
T	T	F	F	T	F	T
T	F	T	T	T	T	T
T	F	F	F	T	F	T
F	T	T	T	T	T	T
F	T	F	T	T	F	F
F	F	T	T	F	T	T
F	F	F	T	F	T	T

In listening to spoken English, we are not always so good in hearing parentheses. People might mean (A $\rightarrow$ C) $\vee$ (B $\rightarrow$ C) but say (A$\vee$B$\rightarrow$C).

By the way, notice how complicated the table got. The more clauses that can have truth values, the larger the table. Because there are now 3 clauses, represented by the letters, A,B and C, so we have 8 possible truth assignments.

In practice, one has to simply get used to parsing sentences very carefully to know what they mean and do not mean. When you're stuck, you can produce a truth table and check to be absolutely sure you haven't neglected any possibilities.

Another useful relationship between statements that mathematicians love is

IF AND ONLY IF. A$\leftrightarrow$C means that A is true exactly if C is true. So the truth table looks like:

A	C	A↔C
T	T	T
T	F	F
F	T	F
F	F	T

In an ideal world, with the previous interpretations of what A and C mean, I would be very happy if it were the case that A↔ C.

Exercise: Convince yourself that (A↔B) ↔ (A→B) & (B→A). Express this in words.

Here is a shockingly useful and sometimes confusing tautology[6].

Proposition: (A→B) ↔ (-B→ -A)

We will first check this by truth tables (there are four cases, because there are two independent parts A and B) and then we'll think about what this means.

A	B	-A	-B	(A→B)	(-B→ -A)	(A→B) ↔ (-B→ -A)
T	T	F	F	T	T	T
T	F	F	T	F	F	T
F	T	T	F	T	T	T
F	F	T	T	T	T	T

The proposition says that saying that A implies B is exactly the same as saying that not B implies that not A is true i.e. that A is false. In other words, to say that A implies B is exactly the same as saying that if B does not hold, it must be that A is false. For if A were true, B would have to be.

This proposition is the basis for a technique called proof by contradiction that we will illustrate in the next section.

Another important relation is that

[6] A tautology is something that is always true, no matter what.

$$-(A \vee B) \leftrightarrow (-A)\&(-B).$$

"Not" switches around "ands" and "ors". To say that I am not (both) gorgeous and smart means that I am not gorgeous or I am not smart. I only need fail to have one of these characteristics to fail having them both (if you see what I mean).

You might now want to reread my letter to Mom and check that despite my camp being perfectly safe and boring, the letter was completely true.

2. Quantifiers and Fun and Games.

It is important to observe how "not" interacts with "every". When I deny that "All Americans are open, honest, and friendly" what actually happens?

Denying "open, honest, and friendly" is asserting "un-open or dishonest or unfriendly". We already discussed that. But the "All Americans" becomes "There is some American".

The denial of "All Americans are open, honest, and friendly" is actually "There is some American who is un-open, dishonest, or unfriendly". Just finding one American who is not open or who is dishonest or not friendly will be enough to show the falsity of ""All Americans are open, honest, and friendly." Unfortunately, this is not too hard to find.

And, of course the denial of "There exists" is either of the two equivalent statements "There does not exist…" or "For all things, it is not the case that…"

For example, to deny that "There is a free lunch" means either of the two entirely equivalent statements: "There is no free lunch" or "All lunches are unfree."[7]

Now, for notation: We will write "$\forall x$" to say "For all x" and we write "$\exists y$" to say "There exists a y."

So $-(\forall x\ P)$ is "It is not the case that P is true of all x". It is equivalent to $\exists y\ -P$, that is "There is a y so that P is not true of y."

[7] Will you join me in working to free the lunches?

Now a reward for swallowing these abstractions and playing with them.

Theorem Every game between two players that is played sequentially (i.e. the players take turns) and ends in a finite amount of time has a strategy for one of the other players.

Here's an example: I call it "amateur's chess". It is played the way chess is, between White and Black, except that one of the players, say White, is an amateur. In amateur chess, if the amateur draws the game by usual rules – he wins. I think it is reasonable – I know that when I draw a good chess player, I am quite proud and think of myself as a winner and she thinks of herself as a loser.

So, we will now see that amateur chess has a strategy: it can be written down in a book, which you can memorize and then following it, win.

This means "amateur chess" is an unfair game. One of the players, either white or black, but obviously not both has a strategy that he can follow and guarantee a win.

I don't know the strategy, and I don't even know who is the guaranteed winner. It could theoretically be that Black has an advantage in chess, maybe because White has to commit to a strategy first --- or for whatever reason --- and that, as a result, even the amateur advantage of winning in the case of a draw is not enough to tip the game. It's also possible that "amateur white" has the strategy, and then no amount of brilliance will save the expert. I don't know and I don't expect to ever find out. The amount of calculation involved is too great.

To firm up our understanding of what a strategy is, let's imagine the incredibly boring sequential "Rock, Paper, Scissors". There are two players, Paper beats Rock, Scissors beat Paper, and Rock beats Scissors[8].

[8] To break ties, we can make the house rule that youngest wins in a tie. But that is only necessary if we would want to invoke our theorem.

Now, in our set-up of the type of games we permit, we are committed to playing this sequentially[9]. You go first. There's nothing you can do to save yourself. If you play rock, I'll play paper. If you play paper, I'll play scissors. My strategy, of course, involves looking at what you do --- and then I react to it in the obvious way.

Let's write in symbols what it would mean for player I to have a strategy in a game. And, let's assume that it always ends after four turns.

Player I moves first and picks and x. Then II will pick some y, where y is an allowable move. I cannot know what it is, so he must be able to react with some z (depending on what x and y were). This choice of z must be so good that whatever w II chooses won't make a difference and player I will then win.

Let's write this in symbols:

(Player I strategy) $\exists x \forall y \exists z \forall w$ I wins

This means, there is an x (I's first move) so that no matter what y chooses to respond with, I can find a z so that I wins after II makes whatever (futile) response it tries.

If the game takes longer, we will need a longer sentence to express what it would mean for I to have a strategy. But to say a game is finite means that after some number of turns, the game must certainly be over.

You might want to think through for yourself why amateur chess has this property. (Recall that in ordinary chess if the chess board is in exactly the same position it had been in twice earlier, then the game is called a draw.)

We can now prove the theorem. Suppose that Player I does not have a strategy. Then

-(Player I strategy) $- (\exists x \forall y \exists z \forall w \ldots$ I wins)
equivalently $\forall x \exists y \forall z \exists w \ldots -$ (I wins)

[9] Games where the players take their turns simultaneously or don't have full information of what the others have done are much more interesting. In that case, one can't guarantee a win. We will have to delay a discussion of this to the future, when we discuss probability and randomness.

Note – (I wins) (i.e. not I wins) is the same thing as saying that II wins, by our assumption that our games have no draws. So we can read the second "math line" above as saying that "No matter what x I plays, there is a response y that II can choose, so that whatever z I replies with can be thwarted by II's choosing an appropriate w giving II the win.".

In other words, II has a strategy!

We have proven the theorem.

And, notice that we proved the theorem without actually producing a strategy! It is a non-constructive existence proof!

Remark: For usual games, where at each point each play has only finitely many options for its next turn, one can, in principle try to examine all possibilities and work backwards to come up with a strategy for one or the other player. In practice for a game like chess or go, there are way too many for even computers to accomplish such a search (although they have gotten really good at playing such games.)

However, there are games where each player can have infinitely many possible moves, and the theorem still applies to it.

Here is a challenging example.

The game is played with three piles of pebbles. In each turn you can take a pebble from the leftmost pile and put as many pebbles as you want into any of the other piles. Once a pile is empty, there will be one fewer pile and you still must remove a pebble from the new leftmost pile.

The player who removes the last pebble wins.

Suppose the piles start with with 2, 2 and 1 pebbles. Then player 1 will remove 1 from pile 1 and might add 325,192 to the second pile and the new situation will then be three piles with 1 in the leftmost, 325,194 in the middle, and 1 in the rightmost. Player II might respond by taking 1 out of the leftmost and putting 5 into the rightmost, then there will be only two piles left, the left containing 325,194 and the right containing 6. And then play I goes, removing 1 from the left, leaving 325,193 there but might top off the right pile to have 23 trillion pebbles in it.

I recommend you think about why this game cannot last forever – it must terminate. (But it can take a long time….) Even if both players (let's assume they are immortal) cooperate in trying to keep the game going on for a long time…they cannot make it last forever. Each player can have infinitely many choices of possibilities (while there is more than one pile), so exhaustive search cannot find the strategy for you. You might enjoy thinking about why this game has strategy and try to discover what it is. (And for the alternative version where the one who picks the last pebble loses). They are not hard – even if they can't be discovered by brute force.

Exercises.

1. Which of the following statements are tautogies?
    a. $A \rightarrow AVB$
    b. $(A \rightarrow -A) \rightarrow -A$
    c. $(A\&B) \rightarrow (C \rightarrow B)$
    d. $((AVB) \rightarrow C) \rightarrow (-C \rightarrow -B)$

2. If X & Y is a tautology, where X and Y are complicated statements, why are both X and Y tautologies.

3. Give an example of a tautology of the form XVY, where neither X nor Y are tautologies.

4. Why is $X \rightarrow Y$ a tautology whenever Y is?

5. Prove that Amateur's chess always ends.

6. Here's a game played on a 1 x n strip. Players take turns putting down pieces on the board. You can put your piece on any unoccupied spot that is not next to a spot occupied by your opponent. If you cannot put your piece down you lose.

Try it for several values of n and try to develop a strategy.
What happens if you play the same game on two 1 x n strips (and each turn you have a choice of where, on either strip, you put your piece)?

Counting

Counting can be complicated. Our first major goal is to show that, in theory, if you count something twice, you should get the same answer both times. (This is a perennial problem in elections. In practice, recounts always get different answers, hopefully close though, when there are a lot of things to count.)

Establishing this will require careful analysis and precise definitions, like what it means to count something, or even what kinds of somethings can one count? One of our first milestones will be a proof that when you count the objects in your bag once, and then put them all back in, and take them out in a different order, you still get the same number when you are done. While this seems to us too obvious for words, it is not really, and it will take a bit of careful thought to get there.

Along the way, we will prove a theorem that cuts to the core of the meaning of rationality of the decisions of a committee or a society (the Arrow impossibility theorem).

Then we will move onto actually counting different things, like, say, how many ways one can tile a 2 x 15 strip of chessboard using 2x1 dominoes. We will see why there are more than a million tilings of the chessboard by 2 x 1 tiles.

This chapter closes with two remarkable results, one of the nineteenth and one of the twentieth century. We will prove Cantor's theorem that there is more than one size of infinity (indeed: there are infinitely many infinities, and no largest). And, then, we will explain, and, following Turing, we will sketch a proof of Godel's theorem that not all true mathematical facts can be proved on the basis of finitely many axioms.

1. Sets.

A set is a collection of things. It is standard to denote them using "braces". {blue pandas} is the set of blue pandas. I am an element of that set if I am a blue panda, and not otherwise.

{1,2,4} contains three elements, namely 1, 2, and 4. {Edward Teller, The Father of the Hydrogen Bomb} contains one element, namely Edward Teller, because he is the father of the hydrogen bomb. Similarly, {men} = {mortal men} because all men are mortal.

{} contains no elements, because there is nothing inside it. It is called the empty set and is often denoted by ∅.

We have already used two methods of describing sets: listing its elements or giving a property that its elements share. It is not always easy to tell whether two sets are the same. For instance one doesn't know if A = {inhabited planets in the solar system in the year 3001} is the same as B = {the Earth} or ∅ or, perhaps, C = {Mars} (the most likely?).

A favorite among mathematicians is **N** = {1,2,3,4,….} of natural numbers. For many purposes **Z** = {0, ±1, ±2, ±3…}, the set of all integers (positive, negative, and zero) is even more important.

I will use the notation |n| for the set {1,2,3…n}, so |5| = {1,2,3,4,5}

Let's discuss a few more examples. {Shmuel Weinberger} is a completely different set that {cells in Shmuel Weinberger's body} because the first set has only one thing in it, and the latter has many. { {1,2} } is different from {1,2}. { {1,2} } is a set with one element in it, while {1,2} is a set with two elements in it.

Indeed { {1} } is different from {1} because the latter has one element, which is a *number*, while the one element of the former set is a *set*.

Notice that the above "strange examples" involved sets contained other sets as elements of them. You can think of a set as a bag, and it contains things in it. It might contain other bags inside itself. It knows that it contains those bags, but it doesn't know

about the contents of those bags: 1 is an element of {1}, but not of { {1} }. It is an element of {1, {1}}.

Definitions: A is a ***subset*** of B if every element of A is an element of B. This is denoted by A⊂ B. If A is not a subset of B, we write A ⊄ B. We use the notation a ∈ A when we want to say that a is an element of A, and I bet you can guess what a ∉ A means.

Think carefully about the following.

Examples:

{1,2} ⊂ {1,2,3}. 1 ∈ {1,2,3}
{Whales} ⊂ {Mammals}
{Elsa Lanchester} ⊂ {Brides of Frankenstein}[10].
Elsa Lanchester ∈ {Brides of Frankenstein}
N ⊂ **Z**.

Nonexamples:

{1,2,3} ⊄ {2,3,4,5}
{1,2} ∈ { {1,2} } {1,2} ⊄ { {1,2} }. 1 ∉ { {1, 2} }
{Whales} ⊄ {Fish}
{Madonna} ⊄ {Brides of Frankenstein}[11]
Z ⊄ **N.**

It is possible for A ⊄ B and B ⊄ A. Neither of {Elsa Lanchester} and {2, 3, 4, 5} is a subset of the other.

(When my daughter said "I am 5" she did not mean that she is an element of {2, 3, 4, 5}[12].)

2. Operations on Sets.

[10] This example (and the next) involves not distinguishing actors from the roles they play.

[11] As of this writing (2008).

[12] In the first draft of this section, this sentence was written in present tense.

There are various things we can do to sets besides describe them and their interrelationships.

From A, we can form A^c, the complement of A. $x \in A^c$ if and only if $x \notin A$. x lies in the complement of A if and only if x is *not* an element of A. (The expression x lies in X is synonymous with $x \in X$. Do you like the use of upper and lower case letters to indicate different things: although annoying at first, it's actually very helpful in keeping track of where some element lives as reasoning becomes complicated.)

Given two sets, A and B, we can form $A \cup B$. An element lies in $A \cup B$ if (and only if) it lies in A *or* it lies in B. This is called the *union* of A and B.

We can also form $A \cap B$, the *intersection* of A and B, and it consists of the elements that lie in both A and B. If $A \cap B = \emptyset$ we say that A and B are *disjoint*, and will sometimes refer to $A \cup B$ as a *disjoint union.*

If we were to try to imitate the logical operations, we would make a symbol to correspond to $\rightarrow$. We shall not bother to: it is less important, and we can make do with the clunkier $A^c \cup B$ that it would be equivalent to. Much more useful is $A \cap B^c$ which is denoted A-B. It is called the *relative complement of B in A*, or sometimes just the *complement of B in A*, or even "*A minus B*".

Exercises:

1. Why is $A^{cc} = A$?
2. Show that $(A \cap B)^c = A^c \cup B^c$.
3. What is $(A \cap B)^c$?
4. Given three sets, A, B, and C, show that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
5. Given three sets E,F, and G, using the operations we discussed, how many new sets can you construct? Can you give an example that shows that all of your sets are the different.

Another, rather different way of combining two sets, that is probably less familiar is called the *product construction*, or less clumsily, the *product of A and B*, denoted by A x B.

It consists of pairs one is an element of A and the second is an element of B.

Let us give some examples:

If A = {Men} and B = {Women}, then A x B consists of pairs each consisting of a man and a woman. If A has 4 elements and W has 3 (as on Gilligan's island) then A x B has 12.

In the pairs that we are talking about, it is critical to remember that we are insisting on *ordered* pairs, so that the first is an element of A and the second is an element of B. After all, if Pat and Sam happen to both (be names of) elements of both A and B, then we will get (Pat, Pat), (Pat, Sam), (Sam, Pat) and (Sam, Sam) as being four different elements of A x B.

We will use this (a,b) notation when writing elements of products. The product is useful when we want to define relationships between elements of different sets. Or even in the same set. In ancient Rome, we might be interested in the subset $S \subset P \times P$ where we consider P = {people living in Rome) and S = { (s,t) | s is a slave of t}. For relationships like this, the meaning of (you, me) $\in$ S is that you are my slave, while (me, you) $\in$ S would be the completely different idea of me being your slave, and I think we would all agree that it would not be a good idea to confuse these ideas.

Given a set, A we can also form *P*(A), the *power set* of A. It is the set of all subsets of A. For instance if A = {red knight, black rook}, the *P*(A) = { $\emptyset$, {red knight}, {black rook}, A} So *P*(A) contains 4 elements (each of which is a set). (How many subsets does it have?)

Perhaps you will be surprised by the fact that $P(\emptyset)$ is not empty. But, now that I mention it, no doubt you realize that this is obvious. It contains one element: what is it?

Exercises:
1. If the set S has n elements in it, how many elements does its power set have?
2. How many elements are there in $\emptyset$ x **N**?

Digression: Equivalence relations.

Another important way to get new sets out of old is via the idea of an equivalence relation on a set S. An equivalence relation E can be described as a subset of SxS that has the following three properties. (In an appendix[13] we will consider another important type of relation on a set.)

(1) $(s,s) \in E$ for every s in S.	(reflexitivity)
(2) If $(s,t)) \in E$, then so is (t,s)	(symmetry)
(3) If $(s,t) \in E$ and $(t,u) \in E$, then we also have $(s,u) \in E$.	(transitivity)

E is a way of dividing S up into pieces that consists of elements that are similar to each other in some important way.

Let's give some examples and non-examples.

1. S = any set, and $E \subset SxS$ is the diagonal = $\{(s,s) \mid s \in S\}$. Then if (s,t) , s=t. All of the conditions are obvious.

2. S = {jars of paint in my closet}. E will consist of pairs so that both jars have the same color paint. Again, the axioms are obvious: if jar 1 and jar 2 have the same color paint (grue) and jar2 and jar 3 have the same color paint (which must also be grue), then jars 1 and 3 have the same color paint[14].

3. Let S = **Z** and let us say $(s,t) \in E$ if and only if s-t is less than 10. This is **not** an equivalence relation, since, it is neither transitive nor transitive. $(0, 1000) \in E$, (why?) but $(1000, 0) \notin E$. Also (12,6) and (6, 0) are both in E, but $(12, 0) \notin E$.

---

[13] Digressions are discussions that can be skipped temporarily. Appendices can be safely skipped forever.

[14] It is worth noting that empirically this example might not be correct. If by "having the same color" we mean that the eye cannot distinguish these, no doubt it is possible to have two quite similar colors that can just barely be distinguished, but that something right "between" them cannot be distinguished by either.

4. On the other hand, if S = **Z** again and we say (s,t) ∈ E if and only if s-t is even, then E **is** an equivalence relation. For any s, s-s is even, if s-t is even, then t-s = -(s-t) is even (the negative of an even number is even) and if s-t and t-u are both even, then so is s-u = (s-t) + (t-u) (the sum of two even number is even).

Common notations are aEb or a~b if (a,b) ∈ E. We will usually use the latter, unless for some reason we are considering more than one equivalence relation at the same time. So, in example 4, s ~ t means that s and t have the same parity, i.e. that either s and t are both odd or s and t are both even.

Equivalence relations are essentially the same things are *partitions* of the set S into a collection of subsets. Let us make this precise.

Definition: A partition of S is a family $\mathscr{T}$ of subsets. We can think of $\mathscr{T} \subset P(S)$. We assume that every element of S is an element of some element of $\mathscr{T}$. (We could say the union of all of the elements of $\mathscr{T}$ is A --- except that, strictly speaking, we never discussed what the union of an infinite number of sets would be.)

Moreover, we assume that the different elements of $\mathscr{T}$ are **disjoint**. That is if A,B ∈ $\mathscr{T}$, then either A = B or A ∩ B = ∅.

Here's how to go between the two notions.

Given an equivalence relation, we consider for each s, [s] = {t ∈ S, so that t ~ s}. [s] is called the *equivalence class of s*.

Certainly the union of the equivalence classes is S. Why are they disjoint? Suppose that u ∈ [s] and u ∈ [t], then I want to show that [s] = [t]. If u ∈ [s], then u ~ s by definition, and s ~ u (by reflexitivity); as u ~t, s~t by transitivity. If s ~ t then any element of [s] is an element of [t] (and vice versa) by the same argument.

On the other hand, if we have a partition of S, then I will say that s ~t if s and t are both elements of the same element of $\mathscr{T}$. (You should check that this works: any partition gives an equivalence relation.)

From an equivalence relation E, the sets of equivalence classes, S/E is an important abstract object that we will often have to study.

Examples.

1. In example 1 above, S/E = S. In #2, S/E = {colors of paint that I have in my closet}. In #4, S/E = {parities of integers} = {odd, even}

2. Here's substantial example. Let S = **Z**x(**Z** – {0}). Let us say that (a,b) ~ (c,d) if ad=bc. Reflexitivity and Symmetry are obvious, but Transitivity is not. So let us check it.

Suppose that (a,b) ~ (c,d) and (c,d) ~ (e,f). Then

$$ad = bc$$
$$cf = de$$

We can multiply the first equation by f to get

$$adf = bcf$$

and we can multiply the second by b to get

$$bcf = bde$$

combining these, we get

$$adf = bde$$

so

$$d(af-be) = 0.$$

Since $d \neq 0$, it follows that $(af-be) = 0$, so (a,b) ~ (e,f).

What are the equivalence classes? A little experimentation might help [(1,1)] = {(s,t) | 1s = t1} i.e. [(1,1)] = {(s,s)}. Similarly [(2,1)] = {(s,t) | s = 2t}. etc.

The equivalence classes correspond to having the same ratio. So we can think of **Z**x(**Z** – {0})/~ as the set of ratio of integers (where the bottom is not 0). In other words, it is **Q**, the rational numbers, and we have built in the fact that 1/2 = 2/4 = 9/18.

Exercises:

1. Let S = {1,2,3}. How many different equivalence relations can you define?

2. Suppose S is a set and E and F are both equivalence relations on S. Say that (s,t) ∈ EF if sEt and sFt. Show that this defines an equivalence relation.

On the other hand, give an example where EVF, defined by (s,t) ∈ EVF if sEt or sFt, is *not* an equivalence relation. (Which of the properties can you contradict.

Appendix: **Orders and Arrow's impossibility theorem**.

We now discuss a completely different type of relation that a set might have: an order. The theory of orders has a certain richness, which we will only scratch.

The material in this appendix isn't used elsewhere in these notes. On the other hand, it is a lovely combination of axiomatics with careful reasoning, and it would be a shame to not make it available to you at this point.

Definition. An order on a set S is a subset $O \subset SxS$ that satisfies the following conditions. We will write $s \leq t$ if $(s,t) \in O$.

For all s, $s \leq s$.
If $s \leq t$ and $t \leq s$, then $s = t$.
If $a \leq b$, and $b \leq c$, then $a \leq c$.

The basic example to keep in mind (although it has some special features that not all interesting orders have) is S is a set of number and ≤ means "less than or equal to".

Another example might be S = {words} and $s \leq t$ means that s comes before t in the dictionary. Notice that here we order words by their first letter, and then use the second letter to break ties, the third for all ties after the second letter rule fails, and so on. The same trick can be used to order High School students seeking admittance to a university: $s \leq t$ means that t has a higher math score than s, but if equal, break the tie using verbal, and if those are equal, compare social security numbers.

Notice that we have not demanded that any pair of elements be comparable. If S is a set with an ordering, SxS can be ordered using OxO. In our numerical example, we would then have in this order $(2,3) \leq (3,4)$ (because both coordinates satisfy the one variable ≤-condition) but (1,2) and (2,1) can't be compared at all, because each is larger in a different coordinate.

Some people refer to orders with the properties that we have called attention to "partial orders", and use the word orders only for "*total orders*", that is, orders that have the property than any two elements can be prepared. People sometimes refer to sets that have been given a partial order *posets*, --- I have never heard anyone refer to these as "osets".

Another important example of an order (i.e. partial order) is $P(S)$, the set of subsets of S, where $A \leq B$ will mean that $A \subset B$.

Exercises.
1. For which S is $P(S)$ with $\subset$ a total order?
2. How many total orders are there on {1, 2, 3, 4} (We will do this problem later in the chapter)?
3. Is $P(S)$ wth $\supset$ an ordered set? (That is, declare $A \leq B$ if $A \supset B$, that is, if $B \subset A$.)
4. Let $\mathbf{O} \subset P(SxS)$ be the set of subsets that correspond to partial orderings, ordered using $\subset$. Let $T \in \mathbf{O}$ be a total order. Suppose that $T \leq U$, show that $T = U$.

An ordered set S is *well-ordered*, if every subset has a least element. A subset $A \subset S$ has a least element, if there is some element $a \in A$ so that any element b of A satisfies $a \leq b$.

For example, **N** is well-ordered by $\leq$ (less than or equal to), but **Z** is not: the set of negative integers has no least element. Well orderings on sets arise in understanding infinite sets and also in understanding *induction* – a subject we will return to when we count.

Now let us turn to economics, and discuss the problem of social choice.

First let us imagine a society that is trying to elect a president from among two candidates. Our society contains n citizens (all of whom have an opinion, and express it at the election). We might choose to adopt the following rule that "pick the winner according to who gets the majority". This will work when n is odd, but when n is even we might use a tie breaker rule of "and then go alphabetically" (favoring Bush over Gore and McCain over Obama).

One can argue that this is unfair. It is unfair to the alphabetically challenged candidates. That doesn't bother us, after all, they are not voters. What is more problematic is that it is unfair to their supporters, who are members of the society --- but there is a very weak sense in which it is fair: had they supported an alphabetically superior candidate, then they would have reaped this "unfair advantage".
ction

A more problematic solution to the problem would be to exclude the weakest person in society (when there is an even number) (I will assume there is just one – say the youngest, a seniority rule.) and then use majority rule.

We can say that it's alphabetical except when overturned by a 2/3 majority.

Another thing to do is just to have a kingmaker (aka dictator) whose opinion is the only one considered.

ALL of these procedures do the following:

> They take the individual choices (here of an element of a set of two choices for president) of the members of society and combine them into a choice made on behalf of all of society.
>
> They satisfy unanimity: if everyone agrees, then that is what is chosen by society.

We are interested in solving the social choice problem, that is, the problem of aggregating individual choices into a choice for a society. The problem comes when we put in more conditions.

For instance, suppose we demand:

> (RF) (Radical Fairness) All members of society vote by secret ballot. Their votes are recorded, but who votes which way is not. On the other hand, we insist on fairness to the candidates in the sense that if A beats B in one election, and then we run it again with everyone switching their vote, then B must beat A.
>
> (M) (Monotonicity) If A beats B in an election, and then it's redone with some of the citizens changing their vote from B to A (and none from A to B), then A stil beats A.

The first is a fairness condition we would want in an ideal world, and the second is one of "rationality". It would be irrational (would it also be unfair?) for an increase of support for A to make him lose a won election.

Theorem. Majority rule is the only radically fair monotone social choice function.

Alas, it doesn't work in an electorate with an even number of citizens. (On the other hand, how likely would it be that the electorate would split exactly evenly[15]?) So problems of social choice are not always solvable, but in this case it is easy to understand the reason.

Problem: Prove this theorem.

Hint: Argue that secret ballot means that the question of whether A wins or not only depends on how many votes A gets. Monotonicity means that there is a threshold: once he gets at least a certain number X, then he wins. Finally, using RF again, now applied to the candidates, conclude that X = n/2.

It also means that we can't find an election process satisfying all of these conditions when the society has an even number of members – because there is no way to handle ties. (Introducing randomness, e.g. flipping a coin, isn't allowed because we asked that the aggregation procedure tell us who the winner is, and half the time with the given output it will puck candidate A and half the time candidate B.)

Things become much more interesting when we have three candidates.

If all we were doing was picking a president, then we could again go with taking the candidate with the most votes. This would satisfy both (RF)[16] and (M). But suppose that our society consists of three subgroups whose populations and choices are listed as below:

---

[15] I do remember an episode of Popeye where that actually happened, with Popeye and Bluto having to court Olive Oyle for the deciding vote.

[16] You might want to think about how to phrase the aspect of radical fairness that applies to candidates in this circumstance. (There are several choices, but they don't have much of an impact on our discussion in the text.)

Party 1 (29%)	Party 2 (31%)	Revolutionary (40%)
C<B<A	C<A<B	A<B<C

In the above scenario, C wins with most of the vote. However, both of the "business as usual parties" would prefer their opponent's candidate to C. Certainly a very strong case could be made that B should be the president. 60% prefer him to the winning candidate (and 71% prefer him to the other losing candidate). Moreover, 60% of people prefer A (the other losing candidate[17]) to C.

To avoid this problem, it seems that we should ask each voter, not just for their top choice, but their whole list of preferences, i.e. for their total order on the set of candidates. Then we can try to aggregate those choices to figure out a whole total ordering: useful, say, in replacing the president after an impeachment or resignation or so.

The problem we are trying to get around is phrased by:

> (IIA, Independence of irrelevant alternatives). For any candidates A and B, if A <B at the end of an election, and some subset of the citizenry changes their vote regarding where C lies in their rankings, still A < B.

The question of whether A is preferred to B should not depend on who else is out there.

In 1950, Kenneth Arrow[18] published the following astonishing theorem in a paper entitled "A difficulty in the concept of social welfare":

Theorem: The only way to aggregate total orders among three candidates satisfying IIA and unanimity is dictatorship.

[17] not a monicker that

[18] Born 1921, winner of the 1971 Nobel Prize in Economics.

In other words, if every member of society makes his preferences entirely freely, and we want to combine these in a way that satisfies unanimity[19] (if everyone agrees, we do as they agree to do) and IIA (the question of whether A beats B only depends on what people think about A and B not about "irrelevant alternatives") then the system – no matter how it is described to you – is a dictatorship. It picks out one member of the citizenry whose vote, independent of how everyone else votes, determines the election.

This theorem shows (at least) that what we mean by rationality for individual decisions does not serve as an infallible guide to what we can expect from a society. It might also mean that we don't really know what we mean when talk about ideas like the "interests of a group" if there is more than one person in that group[20].

[19] I only know of one voting system which violates unanimity: To slightly oversimplify, the Sanhedrin would acquit in capital cases in situations of unanimous vote to convict on the principle that such a person could not have gotten a fair trial.

[20] Indeed, many models of the human mind involve a number of different modules that compete and interact with each other, as, classically in imagining that "justice" and "mercy" or "fear" and greed" argue their case mentally. More prosaically, the various "voters" can correspond to various priorities that we want to take into account in making a decision among several alternatives. Minsky's metaphor for this is a "society of mind" and he argues energetically for the empirical correctness of such a view. (You might also find various of Steven Pinker's books interesting in this regard.) It is then not all shocking that some types of irrationality might be necessary for human preferences. I am certain that, when house-hunting, I was victim to non-transitivity in my pairwise subjective rankings of the desirability of various houses. This is entirely to be expected because when looking for a house, one uses many different criteria in making the comparisons.

To prove the theorem, there are two essential steps. The first is to identify which citizen is the dictator and then to prove that he is indeed a dictator. We will follow a beautiful argument of John Geanakopolos (of Yale University).

Suppose that we have an aggregation procedure satisfying our desiderata. Our first claim is that if each citizen places B either at the top or the bottom of their choice orders, then so must society (even if some put B on top and some on bottom).

For, if B were placed in the middle, with C on top and A on bottom (with no loss of generality – otherwise repeat the following argument with A and C interchanged), we can change all A's to be above C without changing their rankings with respect to B (B is always top or bottom for each individual). Thus, B will stay where it was, in the middle – above A and below C, despite unanimity demanding that A end up above C.

(INSERT A FIGURE HERE?)

Now let us consider what happens when we change the votes in the election from an election where B comes out on bottom, one at a time, till we finally reach one where B is on top. There is a first citizen where the outcome changed. That citizen, we claim, is a dictator[21]. (Indeed, if so, she is *the* dictator, for obviously, there can be at most one.) Let us call her Citizen X.

We will reveal her dictatorial powers a bit at a time.

First, we claim that she is an AC dictator. In other words, what he decides about the relative position of A and C determines the societal judgment.

Remark: This is enough to show that (RA) is impossible, since an AC dictator will be a dictator for BC and AB, i.e. would be a dictator, if we had symmetry among outcomes

[21] If you already believed Arrow's theorem, it would be obvious that this citizen must be the dictator. So this is a reasonable strategy to employ if somehow you had suspected the theorem to be true. (Despite all the reasoning in this appendix, nothing we do explains how to guess Arrow's theorem.)

(i.e. fairness to candidates). But dictatorship isn't fair to voters. This argument, though correct, would lead us to the conclusion that all we need to do is decide how to prioritize fairness to voters and fairness to candidates. Arrow's theorem shows that that is not the issue.

Let's consider any allocation of preferences among all the citizens regarding A versus C, subject to the first X-1 of them having B topmost and the last n-X -1 having B bottom-most, but X's decision being the one we used to identify her. By IIA, B is on the bottom if that is X's will, and B is on the top, if that is X's will.

Suppose that on this choice, then, that X has A<C. Then changing X's preference to be A <B < C, it must be the final outcome has A<B (C is irrelevant to the

A<B decision) and also B < C. By transitivity we must have A < C. i.e. X's choice determined the final choice on the AC decision.

(Insert figure here of X's choices)

But, Citizen X is not satisfied to be an AC dictator. We claim that she is indeed a complete dictator. So, we must show that she is also an AB dictator and a BC dictator. We just show AB; BC is no different.

By repeating the previous argument using C's as the extremes of our orders, we can identify a Citizen Y who is an AB dictator. Our claim is then that Citizen X = Citizen Y. (Does this seem like a conspiracy to you?)

The point is that in this election having either AC or AB dictatorial power is so strong, that it precludes anyone else from having the other.

Suppose this were not true. Then let's consider an election where the preferences for X and Y are as follows (and assume also that whatever everyone prefers, they all agree that A < C):

X (the BC dictator) prefers A < C < B
Y (the AB dictator) prefers B < A < C

Because of X, at the end we have C < B, and because of Y, B < A. So the dictatorships force the final conclusion to be C < B < A. However, the two dictators – and everyone else --- agree that A < C.

This shows an inconsistency in the axioms, assuming that X were not the same as Y. Thus X is the same as Y, and she is both the AB and AC dictator. Indeed she is a dictator.

Remark: Arrow's theorem also holds if there are more than 3 candidates, but we won't discuss how to modify the above reasoning.

2. Functions.

As we count the elements of a set, A, we are relating the elements of A to a set of integers. This type of relationship is encoded in a device called a function.

If A and B are sets, a *function* is an assignment to each element a in A, some element in B. We denote this by writing f: A → B.

Here are some examples:

A = B = {Human beings}. F: A → B assigns to each a their mother. This seems like a function, although we have to deal with some complicated cases.

Adam and Eve: they had no parents. So F would not assign a value to these elements. If there were a first human, in humanity's evolutionary history, then we would have a

similar problem from that perspective, as well (although we might be able to solve this by making B larger, e.g. {primates}).

With technology, the "birth mother" can be different than the "biological mother" or "genetic mother". Both can be different than the "legal mother"; the last might not exist. To make a function, one must *specify* what the output should be.

Specifying "birth mother" and insisting that A = {Human beings other than Adam and Eve} and B = {Human beings}, then we do have a function F: A → B.

2. Let A = $P(\mathbf{N}) - \{\emptyset\}$ then elements are nonempty subsets of the natural numbers. We can consider the function f(S) = smallest element of S ∈ **N.** So f({345, 28, 1192}) = 28.

Since a set of natural numbers always has a smallest element, and only one such element, this function f makes sense. If we had not excluded the empty subset of **N**, then f would not be a function, because we would not have a value for f(∅).
Also, if we'd replaced **N** by **Z**, we wouldn't have a function, because, say the set of negative odd numbers {-1, -3, -5,…..} has no smallest element.

3. If A is any set, then $id_A : A \to A$ defined by $id_A(a)=a$ is a function, called the *identity function*. If A is a subset of B, then the function $i_{A,B}: A \to B$ also defined by the formula $i_{A,B}(a) = a$, is a function (where do we use the assumption that $A \subset B$?) called the *inclusion function of A in B.*

If A is any set, there if a function g: A → {Mathematicians}, where for any a∈A, g(a) = Shmuel Weinberger. This is an egotistic example of a *constant function*.

Notice that if A is any set, there is exactly one function h: ∅ → A.

If A = **Z** x **Z,** then $f(x,y) = x^2 + y^2 - xy + 1$ defines a function f: A → **N**, although this is perhaps not obvious. It is obvious that if defines a function g: A → **Z**. (You might want to spend some time thinking about why this function actually does define a natural number, rather than just an integer. Hint: note it suffices to check this for x and y nonnegative, and that $x^2 + y^2 - 2xy = (x-y)^2$ is always nonnegative.)

If A = **Z**x(**Z** – {0}), then g(s,t) = s/t ∈ **Q** defines a function.

For later purposes we will need a few ways of combining functions.

Restriction. If f: A → X is a function, and B⊂A, then $f|_B$, the *restriction of f to B*, is defined by $f|_B(b) = f(b)$ for any b∈B. This function is "defined by the same formula as f", but it is not f. $f|_B$ is not defined on the elements of A-B!

Composition. If we have functions f: A → B and g: B → C, then we can form the composition of g and f, denoted by gf: A → C and is defined by the formula:
gf(a) = g(f(a)).

Notice that f(a) is an element of B, and that we know what g does to any element of B, so g(f(a)) makes sense, and is an element of C.

Let's do some examples, just to be sure that we are in agreement as to what everything means.

Let f: **Z**x**Z** → **N** be given by $f(x,y) = x^2 + y^2 – xy + 1$. If D = { (t,t) | t is an integer}. D is the "diagonal" subset of **Z**x**Z**. Then $f|_D(t,t) = t^2 + 1$. If g(x,y)=x-y, then $g|_D$ is a constant.

Now suppose f: {Humans besides Adam and Eve} → {Humans} be the "genetic mother" function, and let g:{Humans} → {Humans} be the function defined as follows:
g(m) = m's oldest daughter, if m has a daughter
m, if m does not.

Notice that g is indeed a function (assuming that one cannot give birth to twin girls simultaneously). gf(x) is a function that assigns to any human other than Adam or Eve (note the "or") x's mother, if x has no sister or x's oldest sister, if x has one.

III) Suppose f: **Z** → **Z** is defined by $f(z) = z^2$. Then $ff(z) = z^4$ and $fff(z) = z^8$.

(Sometimes we write f^n for f composed with itself n times --- but you can imagine that this notation does at times cause trouble. After all, if you are used to algebra, you might look at equations like

$$f(z) = z^2$$

and

$$f^2(z) = z^4$$

and conclude z = 0 or f = z = 1, (as the first equation gives z(z-f) = 0, so z = 0 or z =f, but the next equation then implies that $z^3 = z^4$, which only has the solutions z = 0 or z = 1) which is complete nonsense.

Our f is a function, not a number! We will use this notation sometimes, but we hope that you are forewarned to be careful.)

Unions. Example II had another type of method for defining a function. Suppose that $A = B \cup C$. If we already have functions $g: B \to X$ and $h: C \to X$, we can **try** to combine these into a function $f: A \to X$ by means of the rule:
f(a) = g(a) if a is an element of B

f(a) = h(a) if a is an element of C.

The only problem is that if a is in **both** B and C, we might get two different "answers" for what f(a) should be. So, we only get a well defined function if g and h satisfy the *compatibility condition*:

$$g|_{B\cap C} = h|_{B\cap C}\,.$$

Sometimes we write $f = g \cup h$.

Note: We could have used the formula above to try to define $g \cup h$ but change the second line to be f(a) = h(a) if $a \in C\text{-}B$. (Why does this obviate the need for a compatibility condition?). The main trouble is that we would then not have the pretty formula $g \cup h = h \cup g$. (Why not? And, why does it hold with the definition we have given?) This would be especially confusing if we wanted to take the union of many different functions.

Question: If we had 5 different subsets of A, and we used the "wrong" definition of union, how many "unions" of functions could you produce? (Later on, this will be a much easier question.)

3. Types of functions

We have seen several functions and ways of combining them. Now, I would like to call attention to certain classes of functions that arise frequently.

Definition: A function f: A → B is *injective*, (or is an *injection*), also known as *one to one (1-1)* if there are no two distinct elements of A that go to the same element of B. In other words,

$$f(a) = f(a') \rightarrow a = a'.$$

Examples:

1. $f(x) = x^2$ as a function $\mathbf{N} \rightarrow \mathbf{Z}$ is injective, but as a function $\mathbf{Z} \rightarrow \mathbf{N} \cup \{0\}$ it is not, since f(1) = f(-1), but $1 \neq -1$.

2. Let M = {married people}, and let s: M → M assign to m his/her spouse. For this to be a function, each m must have one spouse, so for this to be an example, we need to be in a monogamous society. In that case, s is *injective*.

Exercise: $ss = id_M$. Show that this formula implies that s is 1-1.

3. Let h: $\mathbf{N} \times \mathbf{N} \rightarrow \mathbf{N}$ be defined by $h(a,b) = (2a-1)2^b$.[22] This function is 1-1. Let's think about this.

To say that a function f is 1-1, it means that by looking at f(s) it is somehow possible to recover s (at least in theory….as there is nothing else other than s that is assigned this value by f).

So suppose we have a number n in the image of h. We want to recover (a,b). We can find b as by counting how many times we can divide by 2 until we get to an odd number.

[22] Recall the exponential function 2^b is 2 times itself b times. So $2^3 = 8$ and $2^{11} = 2048$.

Having recovered b in this way, we also have an odd number (the result of having divided by 2^b). If we add 1 to that odd number , we get an even number that we can then divide by 2: that will give us a.

A related concept is *onto* or *surjectivity*.

Definition: A function f: A → B is *surjective*, (or is a *surjection*), also known as *onto* (as in, "f maps A onto B") if for each b ∈ B we can find an element of A, a, so that f(a) = b.

Examples:

1. None of the squaring examples are onto. 2 is not the square of any integer.
2. The function s is onto: (assuming monogamy, so that s is a function) every married person is someone's spouse.

Exercise: Again, deduce surjectivity from the equation $s^2 = id_M$.

3. The function h is also not onto; no odd number is h(anything). However, if we consider g(a,b) = h(a,b)/2 = = $(2a-1)2^{b-1}$, then we get a surjection, as I will leave to you to check.
4. Suppose f: **N** x **N** → **N** is defined by f(n, m) = m, then f is onto but not 1-1.
5. Suppose g:**N** x **N** → **Z** is defined by g(n, m) = m, then g is not onto.

Notice that the only difference between examples 4 and 5 are where we think of the function as taking values. When we think about a function, it is not just a formula. It is *a way of assigning elements of a set B to elements of the set A*. **A and B are part of the very meaning of the function.** Otherwise, concepts like injectivity and surjectivity would not make sense.

Definition: If f is a function f: A → B, then *the image of f*, denoted im(f) is the subset of B defined by

$$Im(f) = \{b \in B \mid \text{for some } a \in A, f(a) = b\}.$$

So, f is onto is the same as saying that B = im(f).

In example 1, the image is "the set of squares". In example 2, M = im(s). In example 3, g has the even numbers as image, and in example 5, im(g) = **N** ⊂ **Z**.

It is not always easy to figure out what im(f) is for interesting functions.

Theorem: Suppose that A is not the empty set. (a) A function f: A → B is 1-1 if and only if there is a function g: B → A so that $gf = id_A$.
(b) A function f: A → B is onto if and only if there is a function g: B → A so that $fg = id_B$.

We will prove part (a), leaving part (b) to you[23]. First, suppose that one can find such a function g. Now suppose that f(s) = f(s'). Then

$$\begin{aligned} s &= gf(s) \text{ (why?)} \\ &= gf(s') \text{ (as } f(s) = f(s')) \\ &= s'. \end{aligned}$$

Now, suppose that f is injective, then we want to find a function g that will serve us well. Suppose that b ∈ Im(f), that means that b = f(a). In that case, we want to set g(b) =a, so that gf(a) = a. (Notice that for this to make sense, b can only be f(1 thing) – otherwise this formula would not define just one value for g(b).)

For b ∉ Im(f), it doesn't really matter what we define g(b) to be. Just pick one element of A, say t, and define g(b) = t for all b ∉Im(f). (Here is where we use the fact that A is not empty.)

In short, we define g by

$$\begin{aligned} g(b) &= a \text{ if } b \in Im(f) \text{ and } b = f(a) \\ &= t \text{ if } b \notin im(f) \end{aligned}$$

Note that $gf = id_A$. QED

Our final definition of this section is:

---

[23] This is a somewhat dirty trick as I am eliding a logical point. One needs to invoke the "axiom of choice" at some point in the construction of g. Whenever we define a function by saying "this set of choices is nonempty, so just pick one of the elements of that set" but we haven't specified the mechanism for picking the element, we are using the axiom of choice. The axiom of choice asserts that this is legitimate. Here for each a, you will pick some b that f sends to a.

Definition: f: A → B is a *bijection* means that f is both injective and surjective. It is 1-1 and onto. Such an f is sometimes called a *1-1 correspondence*.

Examples:

1. Let f: {1,2,3} → {Moe, Larry, Curly} be defined by f(1) = Moe, f(2) = Larry, f(3) = Curly is a bijection.

2. The function $g(a,b) = (2a-1)2^{b-1}$: **N**x**N** → **N** considered above is a bijection.
3. The h: **N** → {positive even integers} given h(x) = 2x is a bijection.
3': The function j: {positive even integers} → **N** given by j(x) = x/2 (since x is even, this makes sense) is a bijection.

The situation of examples 3 and 3' is very common. If f: A → B is a bijection, then using the theorem above we can find a g: B → A so that gf = id_A and fg=id_B. If we have such a pair of functions, then they demonstrate that both f and g are bijections.

The following theorem summarizes some major points about injections, surjections, and bijections.

Theorem: Suppose that f: A → B and g: B → C are functions, and gf: A → C is their composition. Then,

a. If f and g are 1-1 then so is gf.
b. If f and g are onto, then so is gf.
c. If gf is 1-1, then so is f.
d. If gf is onto, then so is g.

Exercises: (Parts of #3 are best approached after section 4).
1. Suppose that f: A → B and g: B → C are functions, and gf: A → C is their composition. Suppose further than f is onto and gf is an isomorphism, show that f and g are isomorphisms.
2. (a) Give an example where gf is surjective, but f is not. (b) Give an example where gf is an injection, but g is not.

3. (a) Show that if f: A → A is an isomorphism, then there is exactly one isomorphism, which we will call f^{-1} so that $ff^{-1} = f^{-1}f = id_A$. (b) Show that $(f^{-1})^k = (f^k)^{-1}$. (c) Define $f^0 = id_A$. Prove, that for all f, and all integers $f^a f^b = f^{a+b}$. (d) Prove that $(f^k)^l = f^{kl}$.

4. The Fundamental Theorem of Caveman Mathematics.

One day Ogg and Ogg[24] were considering the question:

"Is there an injection from {Mastodon, Bison, Fire, Rock} into {Mastodon, Bison, Rock}?" Their path-breaking conversation and the subsequent paper they wrote are classic and we will quote here excerpts:

Ogg: I think there is an injection.
Ogg: Cannot be. There is no Fire in B.
Ogg: But, look. I don't need the Fire in B. I can send Bison to Rock and Fire to Bison.
Ogg: And Mastodon?
Ogg: Mastodon to Mastodon!
Ogg: And Rock?
Ogg: Rock to Rock of course!
Ogg: But it isn't injective. Ogg sent Bison to Rock!
Ogg: Try something else, then. Say Rock to Rock and Bison to Bison.
Ogg: Mastodon to Mastodon?
Ogg: No. Then we have problem with Fire. Fire to Mastodon.
Ogg: Then where we send Mastodon?
Ogg: To Fire!
Ogg: But then, not injective again.
Ogg: Aaaargh. There must be a better way.
Ogg: There are 81 possible functions A → B. We can be here many days if we try them all[25].

[24] Back then, there only was one name, Ogg. This is a very old theorem.

[25] It is amazing that Ogg was able to count the number of functions from A to B at this early stage in the development. Genius always amazes.

At some point, Ogg comes up with the idea of counting. He says, let's assign integers to each element of the set, in order. So 1→Mastodon, 2 → Bison, 3 →Fire, and 4→Rock. Then we can say that A has 4 elements in it. (Then repeating the process to B, we say that B has 3 elements, and one is left proving the theorem that a set with 4 elements cannot be mapped injectively into one with 3 elements.)

But Ogg objects! How do you know that whichever order you take the elements of A out in, you get the same answer?

After a while, Ogg comes up with a good argument – which we will present below. Before presenting it, I must explain the principle of mathematical induction.

Induction. The principle of induction says this. Suppose that you want to prove that something is true for all natural numbers. Then it suffices to prove

(1) That it is true for n=1

(2) If it is true for all smaller integers, i.e. all $m \leq n-1$, then it is true for n.

After all, if it were false for some n, there would be a smallest n for which it fails. By (1) that isn't n=1, and by (2) the smallest failure can't be the number n. (For it to fail for n, if must fail for some $m \leq n-1$.)

Here is a simple example of a proof by induction:

Theorem: $1 + 3 + 5 + \ldots.. + (2k-1) = k^2$. The sum of the first k odd natural number is the k-th square.

Proof. It is true for k=1. The first odd number is 1 which is the first square.

It is also true for k =2, as $1+3 = 4 = 2^2$, but this is irrelevant to our plan. Nor would checking it for k =3,4,5,6,… 10000, bring us any closer to our goal --- they would check that the theorem is true in many cases; it might serve as *empirical evidence* that we should believe it. But, it is not a proof.

Let's suppose that the theorem was verified through the first n odd numbers, and we want to study the sum of the first n+1. Then

$1 + 3 + 5 + \ldots.. + (2n-1) + (2n+1)$

$= [1 + 3 + 5 + \ldots.. + (2n-1)\ ] + 2n+1$

$= n^2 + 2n+ 1$ (Here we are using the fact that the theorem has been verified for n)

$= (n+1)^2$.

This shows that if it is true for n, it is true for n+1. Since it is true for n=1, we get the result for all n.

This kind of argument is the most common form of induction: (2) is only applied to deducing n from n-1. The following example was discovered by Brooks Weinberger.

The Game of Split. The game is played between two players. Your turn begins with you being presented with two piles of stones. In your move, you throw away one pile, and split the other pile into two piles, which you present to your opponent.

It you ever get presented with two piles with one stone each, you lose. (Since you will not be able to do your job of splitting after throwing away a pile.)

If not, you won't lose that round – you'll always be able to make some move. However, as we saw in the last chapter there is a strategy in any game like this, and "Split" has the following easy strategy.

Strategy: Keep a pile with an even number of stones, throwing away the other, and present your opponent with two odd piles. If you get only odd piles, you will lose – unless the opponent makes a mistake. So begin making a lot of noise, smoking a cigarette, or discussing something irrelevant. Suggest playing a different game.

The strategy proves itself by the right induction.

We say (a,b) is a *losing position* if nothing you do, can guarantee victory. We say (a,b) is a *winning position* if it can play a move that wins. Notice that if (a,b) wins then one of a or b has the property that (a,x) or (b,x) is winning for any x. After all, if you throw away b and then split a somehow, and then guarantee a victory, then after throwing away x you can do the same with a, and still have a victory.

So, we really need to identify “winning numbers”. 1 is not. ((1,1) loses.) Therefore 2 is a winner, as you can split a 2 pile into 2 ones. You’d never want to give your opponent a 2.

Therefore, 3 is not a winner, because it splits into 1 and 2, and 2 is winning for the opponent. So you never want to split a 3.

Indeed, no odd number k can win, as it splits it into an even number and an odd (both smaller than k) – so the opponent gets a smaller even number, already observed to win. Moreover, any even number wins, by splitting the even pile and giving your opponent two odd piles.

Exercise: Write the informal argument above as a formal induction. You will need to separate the cases of (a,b) where (at least) one is even, and where both are odd.

XXX Exercise: Prove the following sums by induction:
Sum(n) (n^2) =
(2^n) =
(x^n) =

(Due to Brooks Weinberger 2005[26]) Analyze and prove the strategies for (a) unequal split: you must split the pile you choose into two unequal piles, (so two is no longer a winning number) and (b) three split: here you split one of your two piles, and then throw out one of your three piles.

[26] If you get stuck, you can look at the paper “On split-like games” at http://www.math.uchicago.edu/~shmuel/3sp.pdf

Now we can return to the remarkable paper of Ogg and Ogg:

On the theory of counting
By Ogg and Ogg[27]

The goal of this paper is to establish a theory of counting that can be used to explain lists of people who live in a cave, groceries, animal populations and many other sets.

The idea is this. We will associate a natural number to each of these sets. It goes like this. We say 1 and then remove an item from the set. Then we say 2 and remove another item. Then we say 3 and remove another element. If this process stops, we say that our set is *finite*. The last number we said when we were able to remove an element is called *the number of elements in the set*. If this doesn't stop, then we say the set is *infinite.* **N** is an example of an infinite set.

We do not know if there are any examples that occur in nature. (If A were empty, we will say that "*A has 0 elements*" – which is just a fancy way of say A has no elements.)

Let us make the above process a bit more precise. Notice that we can think of our method of counting the elements of A as a function from {1,2,3….n} → A, where the number k goes to the k-th element removed from A. This map is injective, since each element is removed after it is counted, so we cannot ever map another integer to that element.

The map is surjective because, if some element were not in the image then we could remove it and announce n+1.

(**N** is infinite because for any function g:{1…n} →**N**, the number g(1)+g(2) +….+g(n)+1 is larger than any g(i), and therefore does not lie in the image of g. In other

[27] Partially supported by a generous grant of the Cave Mathematical Foundation.

words, no function from any {1,2….n} →**N** can be surjective.) We can therefore make our idea of "counting" precise using the following:

Definition: Let n∈N. Recall that we denote {1,2,….n} ⊂**N** by |n|. We say that "*A has n elements*" if there is a bijection f: |n| → A. (We say "A *has at least n elements* if there is an injection j: |n| → A.)

The following result seems to us surprising:

Theorem 1: If A has n element and A has m elements, then n=m.

This means that if we have a finite set, and we count {Mastodon, Bison, Bird} by saying Mastodon 1, then Bison 2, and Bird 3 we get 3. But another Ogg might say Bison 1, Bird 2, and Mastodon 3. And he'll get 3. But Ogg did a completely different counting process than we did; he still gets the same answer as if by Fate.

We have done experiments with piles of rocks, and usually when there were twelve or more rocks, and we counted it three or four times, at least once we got a different answer than the other times. But when we carefully painted symbols on the rocks and wrote down our (ac)countings of the pile, we found that the counts did coincide. We assume that with large piles, people will miscount by skipping something or counting something more than once.

(With people it is very hard to count, because everyone is named Ogg. Maybe better to find a new way to give labels to people. But maybe not – why would you ever want to count people?)

Because of the theorem, we can write #A = n if there is a bijection between A and |n|.

Theorem 2: (a) There is a bijection between finite sets A and B iff #A = #B. (b) There is a surjection A → B iff #B ≤ #A. (c) There is an injection A → B iff #A ≤ #B.

Theorem 3: (a) If A is finite, then so is any subset. (b) If A and B are finite, so is A ∪ B. Moreover #(A ∪ B) = #A + #B - #(A∩B).

We will prove theorem 1 and leave theorem 2 to our readers. Then we will prove theorem 3. Together, these theorems give a good description of what finite sets look like. We wonder whether there is anything like Theorem 2a for infinite sets[28].

We prove theorem 1 by induction. Clearly, if A is in a bijective correspondence with the empty set, it is empty, so the theorem is clear if #A = 0.

Suppose that for all k <n, we know that if |k| can be put into bijection with a set A, and so can |l|, then k = l. Now we want to study what happens when $A \leftrightarrow |n|$.

Lemma 1: If there is a bijection $f: |n| \to A$ and $g: |m| \to A$, then there is a bijection $h: |m| \to |n|$

This lemma lets us replace the study of sets, by a study of numbers (notice this is possible because to each number we have already assigned a set).

Proof: $f^{-1}g$. QED

Lemma 2: For any k < n there is bijection $f_k: |n| \to |n|$ so that $f_k(k) = n$ and $f_k(n) = k$.
Proof: Define f by
$f_k(a) = a$ unless a = k,n
$f_k(k) = n$
and $f_k(n) = k$.
The three possibilities are mutually exclusive, so this does define a function. Notice that

$f_k^2 = id_{|n|}$, so f_k is a bijection. QED

Lemma 3: If there is a bijection $h: |m| \to |n|$, then there is one called h' with h'(m) = n.
Proof: Suppose h(m) = k <n, then let $h' = f_k h$. Then
$h'(m) = f_k(h(m)) = f_k(k) = n$. QED

[28] Editor's note: This remark was remarkably prescient, only seriously approached in the 19th century. We will later, following Cantor, show that not all infinite sets have the same size, i.e. can be put into a bijection with each other.

Lemma 4: If h: |m| → |n| is a bijection with h(m) = n, then h restricted to |m-1| has image in |n-1|. This restriction establishes a bijection between |m-1| and |n-1|.

Proof: Suppose k < m. Since h(m) = n and n is the largest element of |n|, as h is injective h(k) < n. The restriction of an injection is an injection, so we only need check that this restriction is a surjection. Let u ∈ |n-1|. u = f(a) for some a ∈ |m|. Since u ≠ f(m) (=n) it lies in the image of |m-1|. QED

The theorem follows, because, by induction n-1 = m-1, so n = m.

Let us now prove theorem 3. Our proof of theorem 1 showed that for any S, either S is finite or there is an injection j:**N** → S. However, if #A = n, restricting j to |n+1| (and composing with the inclusion of S in A) gives a contradiction to theorem 2c.

For 3b let us first assume that A and B are disjoint, i.e. that A ∩B = ∅. Then the claim is that #(A ∪ B) = #A +#B. Suppose #A = n and #B = m, so that f: |n| → A and g: |m|→ B are bijections. Then we claim (and you should prove) that h:|n+m| → A ∪ B defined by:

$$h(k) = f(k) \in A, \quad \text{if } k \le n$$
$$= g(k-n) \in B, \quad \text{if } n < k \le n+m$$

is a bijection.

The general case follows from the following observations.

$$A = A - B \cup (A \cap B)$$
$$A \cup B = (A-B) \cup B.$$

And both unions are disjoint unions (the pieces are dsjoint). The first line implies that #A = #(A-B) + #(A∩B), or #(A-B) = #A - #(A∩B). So the second line gives #(A U B) = #(A-B) + #B = = #A -#(A∩B) + #B, as was to be proved. QED.

Theorem 4. If A and B are finite, then so is AxB. #AxB = #A #B.

Proof: We will prove this by induction on #B. If #B = 0, then AxB ↔ ∅ and both side of the formula are 0. Suppose, now, that B is non-empty, so it contains an element we denote by b. B = (B-{b}) ∪ {b} is a disjoint union.

We now have

$$A \times B = A \times (B-\{b\}) \cup A \times \{b\}$$

So by theorem 3

$$\#(A \times B) = \# (A \times (B-\{b\})) + \#(A \times \{b\})$$

Since $\#(B - \{b\}) = \#B - 1$ (why?) and $A \times \{b\} \leftrightarrow A$, using the induction hypothesis that the product formula is correct for sets with smaller numbers of elements, we identify the right hand side with

$$= \#A(\#B - 1) + \#A$$
$$= \#A\#B.$$

Problems:

Prove theorem 2. Notice that it includes the pigeonhole principle. If $f: A \to B$ is a function and $\#B = \#A$, then for some b in B, there are at least two elements of A, say a and a'. for which $f(a) = f(a')$.

Show that if A is a union of k-sets $A_1, \ldots A_k$ and $\#A > k$, then some A_i has at least two elements. More generally, if $\#A > kr$, then some A_i has at least r elements.

(Inclusion-Exclusion principle.) Suppose A,B and C are finite sets, then show that the following formula is correct:

$$\#(A \cup B \cup C) = \#A + \#B + \#C - \#(A \cap B) - \#(A \cap C) - \#(B \cap C) + \#(A \cap B \cap C).$$

What is the formula for $\#(A \cup B \cup C \cup D)$?

Suppose that if in a class of 45, 25 students speak English, 20 speak French, and 10 speak Spanish. Of the English speakers, 4 speak French and 6 speak Spanish. Show that everyone who speaks both French and Spanish also speaks English. Need there be such a person? How many can there be?

5. Methods of Counting.

The few theorems we already have, together with induction are quite strong tools for counting various finite sets. We will call theorem 4 the "multiplication principle", and we have theorem 3 (together with problem 3 above) the "inclusion exclusion principle".

We will give some examples.

Ensembles: I own 3 jackets, 5 pants, 6 shirts, and 10 pairs of socks. How many possible ensembles can I make, given that I care not at all about aesthetics or matching or the like (the socks come paired, because even I would not wear two left socks).

Let's model the problem as follows: J = the set of my jackets, P = the set of my pants, S = the set of my shirts and Z = the set of my pairs of socks. My ensemble is merely an element of the product J×P×S×Z. This finite set has 3*5*6*10 elements, by the multiplication principle, so I have 900 choices of what to wear every day. No wonder it takes me so long to get dressed in the morning.

Ensembles Revisited: After a trip to Greenwich Village, I discovered that socks don't have to match! I realized, it is much better to put all 20 socks into a basket, and each day I can pick two of them. Now how many possibilities are there?

The only change in the analysis will be the set Z which now should refer to the set of pairs of socks. How many pairs are there?

I have 20 choices of what to put on my left foot, and having made this choice, I only have 19 choices for my right sock. So multiplication principle now says that Z is replaces by a set with 20*19 = 380 elements. (And my flexibility of haberdashery has gone through the roof.)

But, we have to be careful. This is not (yet!) a valid application of the multiplication principle because we are not considering a product of sets Z×Z' where Z' = {possible second socks}, because for each first sock, we have a different set of possibilities for what the second sock will be.

Let me give you two ways around this problem.

The simpler, but much less flexible, way is to say that our choices are Z×Z – the diagonal. I can take any sock for my left foot and any sock for my right, but I will never choose the same for both. By multiplication and inclusion-exclusion, the total number will be 20*20 – 20 = 380 (no coincidence).

The better method is to make the multiplication principle more general, more flexible. The set-up we will use is like this. We have a set E and a function $f:E \to S$ to another set. (For us E = {pairs of socks} and the function assigns to each pair the sock that goes on my left foot. S = {socks}).

For each $s \in S$, we can consider the subset of E, $f^{-1}(s) = \{e \in E | f(e) = s\}$.

First of all, we have not assumed that f is surjective, so f^{-1} is not definable as a function $S \to E$. (Indeed, to be well defined, f would have to be bijective.) If s is not in the image of f, $f^{-1}(s) = \emptyset$.

So, this f^{-1}notion is different than the one we discussed before[29]. On the other hand, when f is bijective, then the old notion makes sense, and this one more or less agrees with the other one. The new $f^{-1}(s)$ is the set that contains just one element, namely the old $f^{-1}(s)$.

It is convenient to have both notions, and the notation is reasonable for both notions --- I hope it is not too confusing. If it is ever necessary to deal with both at the same time, then it will be necessary to introduce a temporary notation for the new notion. After all, can we really abide a formula like:

$$f^{-1}(s) = \{f^{-1}(s)\}$$

that the above discussion would suggest?

Now, let us notice that if $s \neq t$, then $f^{-1}(s)$.and $f^{-1}(t)$ are disjoint. (Where would an element of $f^{-1}(s) \cap f^{-1}(t)$ go under f?) As a result, E is a disjoint union of these subsets, and repeated application of the union formula (= inclusion exclusion, when there is nothing to exclude: all intersections are empty), we obtain:

$$\#E = \sum_{s \in S} \#(f^{-1}(s)).$$

where the large means that we are taking a sum of many terms, here one term for each element of S. In our situation, the $f^{-1}(s)$ are essentially[30] the possible socks that can fill out a pair with the given one. So that each of these subsets has the same number of elements, namely 19. As a result, in the special case when $\#(f^{-1}(s))$ is the same for each s, we obtain that

$$\#E = \#S\#(f^{-1}(s))$$

---

[29] We can think of f^{-1} as a function $f^{-1}: S \to P(E)$.

[30] What logical lapse am I hiding with the word "essentially", and why can I get away with it?

where the second factor is the common value of #($f^{-1}(s)$).

The above formula is so useful that we will give it a name, too; let's call it the *generalized multiplication principle*. We will see it in action soon enough.

Ensembles Revisted Again: If I were packing an ensemble for a trip, then I would have overcounted the size of the decision I have to make. After all, my pair consisting of the pink sock and the yellow sock is two pairs when it comes to wearing the pair – which foot should I put the yellow sock on? – and only one pair when it comes to packing.

The set of *unordered pairs* of socks, in other words, of pairs of socks where the order doesn't matter, or in yet other words, the set of subsets of Z that have 2 elements in them, has exactly 20*19/2 = 190 elements.

If you don't yet see it, let us consider the function

$$h: \{\text{ordered pairs of socks}\} \rightarrow \{\text{unordered pairs}\}$$

It sends a pair (a,b) to the set {a, b}. Since $a \neq b$, $\#h^{-1}(\{a,b\}) = 2$ as

$$h^{-1}(\{a,b\}) = \{ (a,b), (b,a) \}.$$

And finally #{ordered pairs of socks} = 2*{unordered pairs} by the generalized multiplicative principle.

Getting married in the old days Traditionally, marriage was defined as a union of one man and one woman. No man wass allowed to have more than one wife, and no women may have more than one husband. Suppose that we have m men and m women and we wish to know how many ways there are to pair them up.

The multiplication principle makes easy work of this. There are m possibilities for the first women's choice, m-1 for the second, m-2 for the third, …..till we get 1 for the last. So the total is

$$1*2*3*4….*(m-1)*m = m!$$

The m with an exclamation point is **defined** by this equation. (I **do not mean** to say enthusiastically that 1*2*3…*m = m.) Indeed, m! is usually much bigger than m. Here are the first few factorials.

m =	1,	2,	3,	4,	5,	6,	7,	8,	9,	10,
m!=	1,	2,	6,	24,	120,	720,	5040,	40320,	362,880,	3,628,880

100! has 157 digits (and starts with a 9 – what does it end with?).

Notice that the traditional marriages is (yes this is the right grammar!) the same as the bijections from {men} to {women}. So the set of bijections of |n| to itself has n! elements. It is the same as the number of (total) orderings on a set with n elements.

Getting married in 2021 America: Nowadays, by law marriage is still monogamous, but can be any union of two individuals. Let's list the people alphabetically. There are 2m of them. The first person can be married to any of 2m-1 others. The next unmarried person (it could be person number 2 if number 1 didn't marry number 2, but it might be number 3) has a choice of 2m-3 others. And so on. The final answer is therefore:

$$(2m-1)*(2m-3)*\ldots..*1 = (2m)!/2^m m!$$

(We leave checking the formula to you.) So if m=5, there are 120 possibilities traditionally and 945 possibilities now.

Astronauts: There is a class of 100 at Starfleet academy, and they are looking for 3 good beings to send on a mission. How many choices are there? The number of ordered triples is 100*99*98, but each unordered triple has 6 (=3!) orderings. So the number of possible crews is 161,700.

In general if we are interested in the number of subsets with r elements that can be taken from a set with m elements, it is:

$$\frac{m(m-1)\ldots.(m-r)}{r!}$$

Problem: Show that this quantity = m!/r!(m-r)!. Notice that this quantity is symmetric in r and m-r. In other words, if one replaces r by m-r, one gets the same number. So there are an equal number of subsets of |1000| with 1 element (there are 1000 of these!) and with 999 elements. Give a direct explanation of this by producing a bijection for any S, between {A | A $\subset$ S and #A = r} and {B | | B $\subset$ S and #(S-B) = r}.

This quantity, C(m, k), is called a *binomial coefficient*. C(m,k) is often read as *m choose k*, meaning the number of ways of choosing k from m things.

The reason it is called a binomial coefficient is the following:

Binomial Theorem:

$$(x+y)^m = \sum C(m,k)x^k y^{m-k}$$

We interpret C(m,0) = 1. The sum is from k=0 to m.

Let us verify this for m =2. Multiply out

$$\begin{aligned}(x+y)^2 &= (x+y)(x+y)\\ &= xx + xy + yx + yy\\ &= x^2 + 2xy + y^2.\end{aligned}$$

The middle line is the key. One gets many terms. The all are strings of x's and y's of length m – one letter from each term. To get an x^k, the string must have exactly k x's, so we are interesting in the number of ways of choose the k places in this string of length m in which to put an x.

A special case of the Binomial theorem has a nice interpretation. Suppose x =y =1, then the theorem says that if we sum over k, $\Sigma C(m,k) = 2^m$.

If S is a set with m elements, then C(m,k) is the number of subsets with exactly k elements. Thus 2^m is the number of all subsets, i.e. $\#P(S) = 2^m$ ($= 2^{\#S}$).

We can see this another way. Observe a bijection

$$X: P(S) \rightarrow \{\text{functions } f: S \rightarrow \{0,1\}\}$$

defined as follows. Suppose $A \in P(S)$, let X(A) be the following function, $X(A): S \rightarrow \{0,1\}$.

$$\begin{aligned}X(A)(s) &= 1 \text{ if } s \in A\\ &= 0 \text{ if } s \notin A.\end{aligned}$$

Exercise: Check that this is a bijection for any set S.

How many such functions are there? Suppose S = |m|. We have two choices for what f(1) is, and then two choices for f(2), and so on. The multiplication principle show there are 2*2*…*2 (m times) = 2^m possibilities.

If (S is finite and) we consider the function n: $P(S) \to \mathbf{Z}$ assigns to each subset the number of its elements. Its image is {0,1,….#S}. Then the generalized multiplication principle boils down to (#P(S) =) $2^m = \Sigma C(m,k)$ (=#$n^{-1}(k)$).

It is sometimes useful to organize binomial coefficients into Pascal's triangle:

1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1

etc.

Each line starts and ends in a 1, and each number is the sum of the two number above and immediately to its right and left. In other words

$$C(m,k) = C(m-1,k-1) + C(m-1,k).$$

One can prove this from the binomial formula, as follows (for m =3):

$$\begin{aligned}(x+y)^3 &= C(3,0)y^3 + C(3,1)xy^2 + C(3,2)\,x^2 y + c(3,3)x^3.\\ &= (x+y)^2(x+y)\\ &= (C(2,0)y^2 + C(2,1)xy + C(2,2)x^2)(x+y)\\ &= (C(2,0)y^2 + C(2,1)xy + C(2,2)x^2)(y+x)\\ &= C(2,0)y^3+C(2,0)y^2x+C(2,1)y^2x\\ &\qquad +C(2,2)x^2y+C(2,2)x^3\\ &= C(2,0)y^3 + (C(2,0)+C(2,1))y^2x\\ &\qquad +(C(2,1)+C(2,2))yx^2 + C(2,2)x^3.\end{aligned}$$

Comparing the coefficients arising in the first and last lines gives the equation defining Pascal's triangle.

Exercise: Another way to verify this relationship is to use the formula.

$$C(m,k) = m!/k!(m-k)!$$

Prove that $C(m,k) = C(m-1,k-1) + C(m-1,k)$.

Taking a stroll in Manhattan.

Manhattan is essentially a grid. Let's suppose we want to walk directly from 8th Avenue and 96th Street to 5th Avenue and 116th Street. How many ways are there to do this? We will only walk North and East, i.e. no detours.

The trip will take a total of (8-5) + (116-96) blocks = 23. Of this set of 23 one block strolls we have to pick 3 occasions to go east. So the number is $C(23, 3) = 23*22*21/6 = 1771$ possible routes.

Notice that we can count the number of routes differently. There will be a last time that we are on 6th Avenue, and once we've left, our route is completely determined. As a result, we obtain that the number of routes is:

$$C(2, 2) + C(3,2) + \ldots + C(22, 2)$$

In the first summand we count the ways that go straight from 8th Avenue to 6th. The second is the number of ways, if you are allowed to go North 1 block before getting there, and so on, till you consider the last one where you leave 6th Avenue on 116th St.

The upshot of (the obvious generalization of) this discussion is:

$$C(2, 2) + C(3,2) + \ldots + C(r, 2) = C(r+1, 3)$$

And indeed,

$$C(k, k) + C(k+1,k) + \ldots + C(r, k) = C(r+1, k+1).$$

What formulae do these relations boil down to for k = 0,1,2,3?

Dominoes.

We will not develop enough technique to count all of the 2x1 tilings of the 8x8 chessboard, but we already have enough to count some special classes of them.

Divide the board into a 4x4 grid each of which consists of a 2x2 mini-board. So we have 16 mini-boards. Every mini-board has two tilings: one vertical and one horizontal.

Consequently, we can count the tilings where dominoes are always placed down in pairs: there are $2^{16} = 65536$ of them.

Soon we will see that there are many more than this.

Fibonacci and dominoes.

Pascal's triangle is defined by a *recursion relation*. Ignoring the formula for C(m,k), we can think of the formula C(m,k) = C(m-1,k-1) + C(m-1,k) as saying that we can inductively determine the C(m,k) if we know them for smaller integers. You have to start somewhere, so you start with C(0,i) = 0 if $i \neq 0$ and C(0,0) = 1.

Lots of other sequences can be described by recursion relations. For instance, $a_n = 2^n$ can be defined by $a_n = 2a_{n-1}$ and $a_0 = 1$. (The recursion formula only describes the sequence once one has the initial condition $a_0 = 1$.)

In many cases, we can a posteriori find a formula for the sequence, as we did for the binomial coefficients and for the powers of two, but there is no guarantee. Of course, once you've guessed a formula, it is natural to try to prove it by induction (the opposite of one of the exercises above).

A famous sequence defined by recursion is the Fibonacci sequence that was one of the earliest examples I am aware of of naïve armchair modeling in mathematical biology. He imagined that rabbits are born, reproduce twice and then die. Ignoring sex and other complications, the population looks like this:

Day 1: 1 baby rabbit
Day 2: 1 adult and 1 baby.
Day 3: 1 adult and 2 babies.
Day 4: 2 adults and 3 babies.

And so on. Let a_n = #adults on day n, and F_n = #babies on day n.

Then $a_n = F_{n-1}$: every adult was a baby yesterday. (If it were an adult yesterday, it would be a rabbit ghost today.)

$F_n = a_{n-1} + F_{n-1}$.

Or, in other words,

$F_n = F_{n-2} + F_{n-1}$.

This recursion sequence together with $F_1 = 1$ and $F_2 = 1$ determines the Fibonacci numbers (= # babies). Here are the first few Fibonacci numbers.

1,1,2,3,5,8,13,21,34,55,89,….

Later on we will discuss the relation between Fibonacci numbers and the Golden Mean and some of their interesting number theoretic properties.

Exercise: Show that F_n is even if and only if n is a multiple of 3. For which n does F_n end in a 0?

Now I would like to point out a connection between F_n and a tiling problem. Suppose we have a 2xn strip that we would like cover by 2x1 dominoes. We can divide the set of tilings into 2 disjoint subsets: the tilings whose rightmost pair is covered by a single vertical tile, or by the right sides of 2 horizontal tiles.

Instantaneously inventing obvious notation, we can say that

$$T_n = V_n \cup H_n$$

However now notice two key bijections:

$$V_n \leftrightarrow T_{n-1}$$

(take a tiling that ends vertically and remove the last tile) and

$$H_n \leftrightarrow T_{n-2}.$$

Hence, # T_n satisfies the Fibonacci recursion, except that $\#T_1 = 1$ and $\#T_2 = 2$. So we obtain the formula (why?) # $T_n = F_{n+1}$.

So the number of ways to tile a 2X8 strip is $F_9 = 34$ ways.

The number of tilings of the chessboard that nicely fit into four rows of 2x8 are 34^4 = 1,336,336. The number that do so either vertically or horizontally is 2*1336336-65536 = 2607136. Vertical or Horizontal doubled the number, but then, in inclusion-exclusion, we must subtract those tilings that are both vertical and horizontal. There are 65536 of those, as we have seen above.

Of course, there are many more, as we have only calculated a different special class. The following gives you access to some more of these, but we will stop here.

Problem: Develop a recursion relation for tiling a 3xn strip. (Note: n must be even). Then consider dividing the chessboard into 2 3x8 strips and a 2x8. What estimate do you get. You can improve your final answer by a factor, by doing this subdivision in more than one way.

XXX 6. Infinite Sets.

It is now time to turn our attention to infinite sets. They really are different than finite sets!

For instance, if S is finite, and $T \subset S$ is a subset with the "same number of elements" i.e. for which a bijection exists $S \leftrightarrow T$, then S = T. (Why? Consider S = (S-T) ∪ T.)

But $\mathbf{N} \leftrightarrow 2\mathbf{N}$, the natural numbers are in a 1-1 correspondence with the even natural numbers (send k to 2k), but there's an infinite amount left over as well. Similarly $\mathbf{Z} \leftrightarrow \mathbf{N}$ (send z to 2z if z is positive, and to 1-2z, if it is not).

As poetry[31], "The infinite can recede into itself and in no way be diminished."

Theorem: S is infinite if and only if it has a proper subset T so that $S \leftrightarrow T$.

(Recall that a proper subset is a subset that is not the whole set.) We already observed the "if" direction. For the reverse direction, note that in the Oggian theory of counting, we found that for any infinite set S there was an injection $j: \mathbf{N} \rightarrow S$. Let us write, S = j(**N**) ∪ (S-j(**N**)). Then, letting T = j(2**N**) ∪ (S-j(**N**)), T is a proper subset (with infinite complement in S) and $T \leftrightarrow S$ (use the identity on S-j(**N**), and the bijection $\mathbf{N} \leftrightarrow 2\mathbf{N}$ on the image of j).

Here is an important definition:

[31] At least, I think it's poetry.

Definition: A set S is *countable* if there is an injection $i:S \to \mathbf{N}$.

Obviously, all finite sets are countable, and now we have **Z** added to our list. The following is obvious:

Lemma: Any subset of a countable set is countable.

The following theorem greatly expands our list:

Theorem: A countable union of countable sets is countable.

Proof: List our sets S_1 S_2, S_3,, …., whose union is S, and let i_1, i_2, i_3,…. be their respective injections into **N**. Now consider the function $I: S \to \mathbf{N}$, defined by

$I(s) = 2^k(2i_k(s) -1)$ if $s \in S_k$ and s does not lie in any S_j with $j<k$.

The $2i_k(s) -1$ changes the injections on the subsets to have images in the odd numbers. We then are multiplying by a power of 2 so their images now lie in disjoint sets: S_1 goes to numbers that 2*odd, S_2 goes to numbers that 4*odd, and so on. As a result, when we combine injections we still get an injection. (The business about "$s \in S_k$ and s does not lie in any S_j with $j<k$" is to make sure that our function is well-defined: we only want to define the function on each element one time.)

Corollary: The rational numbers **Q** form a countable set.

Proof: The rational numbers are just the set of fractions. Fractions all have denominators. So **Q** is the union of fractions with denominator 1 (aka **Z**) and those with denominator 2 (not necessarily in lowest form), denominator 3, etc. Fractions with denominator $k \leftrightarrow \mathbf{Z}$ by sending a/k to a. Now the theorem applies.

Theorem: An infinite countable set can be put in bijective correspondence with **N.**

Proof: Let $i: S \to \mathbf{N}$ be an injection. Define $j(1) = s_1$ where $i(s_1)$ = minimum integer $\in$ Im(i). Suppose we have inductively defined j(i) for i<k, then define j(k) (recursively) by setting $j(k) = s_k$ where $i(s_k)$ = minimum integer $Im(i|_{S-Im\ j(|k-1|)})$.

In other words we order S by the size of its images in **N** and then send **N** to S in order, 1 to the smallest, 2 to the next smallest, and so on.

If S is infinite, this will be defined for all of **N**. (If S is finite, then at some point the s_k that you are asked to find will be an element of the empty set!)

A little thought shows that this function is a bijection.

Exercise: Complete the proof that the constructed function is indeed a bijection. (Why is it surjective?)

Exercise: Show that the following are equivalent for non-empty sets:

(1) S is countable

(2) S is finite or bijective to **N**

` (3) The is a surjection t:**N** → S.

At this point, the reader must be wondering (unless they remember I've tipped my hand earlier) whether this is it. Are all sets bijective to **N** or |n| for some n? Are all sets countable?

Theorem (Georg Cantor, 1845-1918) *P*(**N**) is not countable.

Before proving this, let me remind you of a larger set of numbers than **Q**, the real numbers **R**. We will have a lot to say about this set later. Real numbers are called rational when they are elements of **Q** and *irrational* when they are not. It is reported that the person who leaked the sacriligous fact that $\sqrt{2}$ is irrational[32] was killed by the Pythagoreans[33].

Real numbers are often described by their decimal expansions: they are things like 312.12345678910111213 14….. where the … indicate that they keep on going. In

---

[32] We will study this type of question in the next chapter. We will be able to prove that $\sqrt{2} + \sqrt{3}$ is irrational – if you know the proof for $\sqrt{2}$, you might enjoy this as a challenge.

[33] Maimonides wrote that the diagonal of a square is both a number and is not: he was not being mystical. He meant that there are two conceptions of number: one based on quotients of integers, what we call rational numbers, and the other is based on lengths of lines that can be constructed by some method. Then the length of the diagonal of the unit square is $\sqrt{2}$ by the Pythagorean theorem is a length that is not rational, a number which is not.

elementary school you learnt how to do arithmetic with these numbers. There was just one funny rule: if the end was all 9's, you could replace it by changing the previous digit by 1 and putting in all 0's. It is the custom to drop an infinite number of 0's at the end. So:

.3999999999… = .400000000… = .4

Theorem: **R** is not countable.

We will use the word *uncountable* for not countable in the sequel.

We will prove both theorems simultaneously by showing that a special subset, $B \subset \mathbf{R}$, of real numbers, whose decimal expansion starts at the decimal point and consists entirely of 0's and 1's is uncountable.

Any set with an uncountable subset is uncountable (why?), and $B \leftrightarrow P(\mathbf{N})$, because to a *B*-real number you can assign the subset of **N** corresponding to where there are 1's in the decimal expansion. (This is a bijection. What does 0 correspond to? What does 1/9?)

The proof that *B* is uncountable goes by what is called the diagonal method.

Suppose that we had a surjection $s:\mathbf{N} \to B$.

Consider the real number r whose k-th digit is 1 if the kth digit of s(k) is 0 and whose k-th digit is 0 if the k-th digit of s(k) is 1.

Note that $r \in B$. We claim that r is not in the image of s. To see why, and why this is called the diagonal method, let us draw a picture.

Suppose the following is the first few decimals of some putative enumeration of *B.*

.11111111111111111111111111
.10010101010000001111111000
.11010101100011011010101010
.00000011111101101010110110
.01100101010010101011110110

etc. Now choose the diagonal choices of 0's & 1's. That is .10001…. Here the nth digit of the number is the n-th digit of the n-th decimal in our enumeration. No problem with this number. But, now, let's reverse the roles of 0 and 1. The number r we produced is .01110….

The claim is that this is not the first: it has the wrong first digit. It can't be the second: it has the wrong second digit. It can't the nth: it has the wrong nth digit. So r is not in the image!

This proves Cantor's theorems: indeed, it is much easier than the fundamental theorem of caveman mathematics although the result is much less obvious!

The proof is completely remarkable. We just showed that there is one real number not on the list. The truth is that an uncountable number of reals must be missed. If only a counter number of them would be missed, then **R** would be a union of two countable sets and therefore countable!

By finding one missing element of **R**, Cantor shows that more must be missed than can ever be counted.

Is there more than merely uncountable? The following shows that there is a set larger than **R**, and with a bit more work, you can play the same game to get infinitely many infinities. That is, infinitely many infinite sets, no two of which are in bijective correspondence. (When you then ask how big an infinitity is the infinitude of infinitudes, you are well on your way to becoming a set theorist.)

Theorem: For any set S, there is no surjection $S \to P(S)$; equivalently there is no injection $P(S) \to S$.

The proof is based on the same idea as we have tried, just we can't write is as concretely. Let $f: S \to P(S)$ be a function. We will now define a subset C using f, that will not lie in the image.

Let $C = \{s \in S \mid s \notin f(s)\}$, Note f(s) is a subset of S for each little s. We will put s in C exactly if s is not in f(s).

Now we check that C is not = f(t) for any t. Suppose it were, then would $t \hat{I} C$? Well, if $t \in f(t)$, it (t) would not be (in C), but then surely $C \neq f(t)$. If $t \notin f(t)$, then, by definition, $t \in f(t)$, and again $C \neq f(t)$. So C is not in the image, and there is no surjection.

Appendix[34]. The limits of computation (and of proof): Godel, Turing, and Rice[35].

Now I'd like to shift gears entirely to another topic, one that also has venerable roots in antiquity, but in this case, the revolutionary progress took place in the 20th century. I am talking about logic, and, for the purpose of this lecture, only of the original ideas of Godel and Turing.

Godel's theorem (which can be summarized as asserting that there are unprovable statements about any sufficiently complicated system that can be finitely expressed) caused an early revolution in mathematics in the early 20th century. It is widely misquoted, like Heisenberg's uncertainty principle and Einstein's theory of relativity. Ironically, these supreme accomplishments of the human mind are often taken to inspire the futility of striving to make hard-fast judgments.

We will return to Godel in a few moments, after saying a word or two about Turing. Living as we do, in a time when computers are on every desk and in almost

[34] The main ideas of this appendix can be understood, I think, at this point. An appreciation for the significance of the axiomatic method – the method that is directly limited by Godel's work, might come only from the next chapter. Also, we will use a few facts from the theory of prime numbers and factorization that the reader is likely to have seen, and that will be proven in the following chapter, as well.

[35] This appendix is adapted from the first chapter of my earlier book "Computers, Rigidity, and Moduli: The large scale fractal geometry of Riemannian moduli spaces" Princeton University Press 2004, that I cannot recommend strongly enough. (See LIAR: Lexicon of Intentionally Ambiguous Recommendations.)

every pocket, it is much easier for us to apprehend his basic ideas, without getting involved in a morass of detail.

Consider a computer that reads instructions written in some language, say, English. We could convert every English instruction into a number by the following artifice:

A → 2
B → 3
C → 5

and so on, assigning to each letter a prime number[36], and then continuing through all the punctuation marks. This uses up the first 30 primes (say). Now let's start over again and assign to A the 31st prime, (and the 61st and 91st, and so on) and to B the 32nd one (and the 62nd, the 92nd and so on), C the 33rd and on and on.

Given a sentence or even a book, one can now multiply together all of the prime numbers associated to all the letters and punctuation marks and get a (very big) number. The first letter will be assigned a prime from the first time the letters are listed, the second letter/space/punctuation takes a prime from the second list, and so on.

Not every number comes from a sentence: 1 and 4 don't because there're no combination of letters that produce these numbers, but many other numbers don't come up because they arise from gibberish (non-words) or ungrammatical constructions.

Still, an English speaker will (by definition) be able to recognize exactly which numbers correspond to sentences, etc. This is called the **arithmeticization** of English.

By the way, English isn't really a good language for our purposes, nor is any natural language. We have trouble with ambiguous words and sentences. The point of

[36] Prime numbers are natural numbers only divisible by themselves and 1. There is an infinite supply of them, and when you multiply different prime numbers in different ways, you always get something different (unique factorization of integers). All of these statements will be proved in the next chapter.

"computer languages" is that, at least with regard to texts instructing us to do things, one can make a language where each grammatical sentence can be unambiguously decoded.

Part of the technical work in Godel's proof of incompleteness was making sure that arithmetic could be written in such a fashion. Turing did the same thing with his "machines" and their arithmeticization – these were intellectual antecedents for the physical computers and their programs that, so few years after their introduction, now seem indispensable for day to day life.

If we wanted to, we could even go further and list all the integers that correspond to sentences, arguments, or the instructions of a computer program in numerical order. For instance, the smallest words in our ordering of English are "a", "I", and "ah", which corresponded to 2, 23, and 2*(77 = 39th prime), but are now 1,2, and 3 respectively.

In short, we can now discuss notions like the "first word", "second sentence", and, what's really important in the sequel, the "nth computer program" (or the "mth proof").

For convenience we will only deal with computer programs that operate on positive integers. Arithmeticization, in theory exactly as above[37], lets such machines

[37] But, in practice, with much more finesse.

do things like word processing, control factories, and interact with humans in video games.

Let's consider a few computer programs and non-programs.

Program 1:
Input an integer k.
Add one to k, and call this k.
Output k

This program just adds one to an integer.

Program 2
1. Input an integer k
2. Add one to k and call this k.
3. If the result of line 2 is less than 100 go back to line 2
Output the answer.

This program is a bit more complicated, and I found it necessary to number the lines. What it does takes a drop longer to decode. If you input an integer less than 100 it outputs 100, and from 100 on, it adds one.

Here's a last one (for now):

Program 3:
Input an integer k.
Multiply k by 3 and call the result k.
Go back to line 2.

This program is quite loopy! For every integer the machine just sits there and never outputs anything. (I'd say that it's not doing anything, which is right for all practical purposes, but, from its own point of view, it's busy working.) Many other programs

would be just as useless as Program 3 (for instance if line 2 were replaced by line 2 of Program 2).

We say that that program 3 does not **halt** for any input i.

Now consider

Program 2'
Input an integer k
Add one to k and call the result k.
If the result of line 2 is more than 100 go back to line 2
Output the answer.

This program halts for k<100, and does not halt for k>99.

We finally come to Turing's theorem:

Theorem (Turing, 1912-1954): There is no computer program (or algorithm) that will decide whether program a will halt on input b.

To be a drop more pedantic, there is no program which will output 0 on input 2^a3^b if program a halts on input b, and will output 1 on 2^a3^b if program a does not halt on input b. (We don't care what the program would do to integers divisible by other primes.)

We'll get to the proof in a moment, but it's worth seeing that this implies.

Godel's Theorem (Kurt Godel, 1906-1978): There are arithmetic propositions about the integers that can neither be proved nor disproved within any finite axiom system.

You have to be careful to tease through the issues hidden in Godel's statement.

With an infinite system of axioms, I[38] could list all truths about the integers as axioms about them, and then just look to see whether my proposition is on the list!

On the other hand, proof here need not necessarily be what we normally think of as "mathematical proof"; it has just be something which:

works (i.e. proved statements are in fact "true"),
can be checked, (i.e. if I write down a proof, everyone else (e.g. any proof checking computer) should agree with me that the proof is valid), and
proofs can be written down as a series of "English" statements.

The second condition can be paraphrased as saying that "proof" should not be a matter of opinion. (In the third condition, we just mean that there's a "good" language that expresses the proof, again, not really a natural language.) In fact, we suppose that there is an algorithm that decides whether or not a series of sentences comprises a correct proof.

Mathematical proof has these properties. (But, in theory, there could be other types of "reasoning" could also be approached by similar techniques.)

Godel's theorem is <u>not</u> asserting that "any sufficiently complicated system cannot understand itself". It is not about "understanding" at all.

What Godel's theorem does say is absolutely astounding. It tells us that a completely well defined mathematical system, like the positive integers, cannot be completely analyzed by mathematical tools. The axioms might determine the system up to isomorphism[39], yet still not be strong enough to enable us, by the method of "proof", to determine all the facts we'd like to know about the system.

The reason, so to speak, is that set of truths is so complicated, the issue of truth so subtle, that we cannot ever conclusively discover all of the truths.

---

[38] Theoretically, assuming I had access to all the truths about integers through some superhuman intuitive faculty.

[39] The meaning of this statement will be clearer after the next chapter.

Godel's theorem was first proved by encoding a classical Greek paradox (the liar's paradox) into the proof system, and has subsequently been proved by encoding different paradoxes. However, the Turing proof I am about to sketch is somehow a bit more concrete. An example of the kind of statement that cannot be proved or disproved (*in general*) is that "program a does not halt on input b". For some a's and b's this can be proved or disproved. We gave examples above.

The proof of Godel's theorem from Turing's is conceptually quite simple. It's the old story of having a room full of monkeys typing Shakespeare.

` Suppose that for every (a,b) there were a proof that either a halts on b or that a does not. Let's define a program of the sort that Turing ruled out:

input a, and b.
let k=0
replace k by k+1
check if "proof k" is a proof that a halts on b; if so, output 0 and stop.
check if "proof k" is a proof that a does not halt on b; if so, output 1 and stop.
Go back to step 3.

Under our assumption that Godel's theorem is false, i.e. there will be some l>0 so that "proof l" works either in line 4 or line 5. (Here we used properties 3 and 2 of our requirements of proofs: property 3, so we could talk about "proof k", and property 2, so

we could check what the proof is proving!) Now, since condition 1 is correct, we'll have decided whether or not a really halts on b. QED.

The proof of Turing's theorem is not that hard either.

Suppose we had a program P that decided the halting problem. Now let's consider a new program:

Program T.
Input k
Invoke P to decide whether program k halts on input k.
If k does not halt on input k, output 1 and stop.
If k does halt on input k, compute the effect of program k on input k, add 1, output this number and stop.

Program T is a program assuming that P is. Therefore, it can be written out in excruciating detail, and then assigned a number t. Now, I ask, what does T do on input t? Certainly it halts; T halts on all inputs. The program will stop at line 4, but the output will be, by definition, T(t)+1. Since T(t) = T(t)+1 is impossible, P could not have, in fact, existed. QED.

This argument is clearly a form of Cantor's diagonal argument, that we used in the previous section to see that the real numbers form an uncountable set. It's remarkable that the same argument that grants us the power to begin analyzing the infinite also tells us that there are limits about what we could hope to know about finite integers!

Exercise: Prove Rice's theorem that asserts the following: If P is any property of a function[40] (e.g. being constant or never halting) that is true of some programs, but not of others. Then there is no algorithm that will tell you, in general, whether a given k corresponds to a Turing machine with this property or not.

[40] For the current purposes, the definition of a function must be expanded to all "partially defined functions", as the computer program might not halt, and therefore give a value, at some inputs.

In some sense, nothing about a function can be universally inferred from a description of the function.

We will not discuss it here, but there is a hierarchy of how difficult these impossible decision problems are. For instance, the question of whether a Turing machine halts is genuinely no easier or harder than the question of whether it halts on some input – but the question of whether it halts on all inputs is definitely harder. This direction is studied by the subject of "Computability theory"[41]

[41] Also known (for historical reasons) as recursive function theory.

8.

9. Algebra: Number Systems.

Our goal in this chapter is to gain a deeper understanding of the fundamental number systems of the integers, **Z**, and the real numbers, **R**, that is, the numbers that are describable by decimals.

Understanding is often achieved by going on a quest (as anyone who has watched a Hollywood movie knows). We will understand the familiar much better after being acquainted with less familiar; we also sometimes discover that foreign shores have their own beauty unrelated to our parochial needs. ""Remember what Bilbo used to say: It's a dangerous business, Frodo, going out your door. You step onto the road, and if you don't keep your feet, there's no knowing where you might be swept off to.""

The numbers we are used to are most notable in terms of the operations that we do to them. In other words, we are more interested in the systems than the numbers. I won't tell you why 7 is the most mystical number or why 13 is unlucky. (I will explain why the decimal expansion for 1/3 repeats right away but for ¼ it stops and for 1/6 it repeats only after a bit of noise --- but that explanation will be systemic, not idiosyncratic.)

The most common things we do with numbers is add, subtract, multiply and divide them. Occasionally we compare them: discuss less than or greater than.

To understand these things, we will consider them one at a time. We will abstract properties of these operations (and relations) and see what necessarily follows from them – and see interesting examples of what does not.

1. A Few Elementary Observations.

Although we will study divisibility systematically later, we might as well make some simple observations to begin with. Even these will serve us well in the sections to follow.

If a and b are integers, then we say a|b, *a divides b*, if there is some integer so that b = ac.

The following theorem is very useful:

Theorem: (1) If a|b, then a|bb' for any b'. (2) If a|b and a|c, then a|(b+c).

Proof: If ak = b, then akb' = bb'. if b =ab' and c =ac', then b+c = ab'+ac' = a(b'+c'). QED

We can paraphrase (1) as saying that if a divides b, and b divides c, then a divides c. (Transitivity for divisibility.) (2) says that the multiples of any given integer is "closed under addition". For instance, being even is the same thing as the multiples of 2. (2) in this case is the statement that the sum of any two even numbers is even.

Definition: An integer n is a *prime number,* if $n \neq \pm 1$ whenever n =ab, then either $a = \pm 1$ or $b = \pm 1$.

The reason we exclude ±1 from being prime is for convenience later. Essentially, it is because we want to use primes for decomposing all other integers, and ±1 do not decompose other numbers.

Theorem: Every integer n>1 is a product of prime numbers.

Proof: If n =2 this is clear. Suppose, n is not prime, then n = ab where neither a nor b is ±1. Now, by induction, both a and b are products of primes, and the theorem is proved.

Note: This argument makes sense for any "semi-group" of integers, that is a set S for which if $s,s' \in S$, then $ss' \in S$. We can then discuss S-primes, as being the elements of S that are not decomposable within S.

Exercise: Verify that any element of a "semigroup of integers" S is a product S-primes.

Here is a strange example for S = {even integers}. Notice that 6 is an "even prime", as is indeed twice any odd number. (Why?). Now 6*10 = 2*30 describes a number as a product of different "primes", none of which divide each other.

Later we will prove the all important:

Fundamental Theorem of Arithmetic: There is one and only one decomposition of an integer as a product of primes, up to factors of ±1 and the order of the factors.

So, ignoring signs, one can factor 30 = 6*5 = 2*3*5, or 30 = 3*10 = 3*2*5. The intermediate factorizations are not into primes; the final ones are – and differ only in their ordering.

Here are the first few primes: 2,3,5,7,11,13,17,19,23,29, 31, 37….

There are many primes:

Theorem (Euclid, 300 B.C.): There are infinitely many prime numbers.

Proof: Suppose not, then consider N = the product of all of them. N+1 is not prime, since it is larger than any of the primes. Let p|(N+1) be a prime factor. p|N, as well. Hence (by the theorem above) p| ((N+1) – N) i.e. p|1 and we have a contradiction. QED.

But, they are rarer, the further out you go. 40% of the numbers up to 10 are prime. 25% of the numbers less than 100 are, 17% under 1000, and the percentage goes lower and lower the further out you go. Aside from 2 and 5, no prime "ends" with a 0,2,4,5,6, or 8 so 60% are eliminated right there.

Prime numbers are important throughout mathematics, and in many areas where mathematics is applied. Some of this will be apparent from the discussion to follow. They have been studied for millennia and much has been learnt about them. In an appendix, I will discuss some discoveries made in this millennium.

However, much is not known about the primes. Perhaps the easiest to state problem regarding primes that has withstood consistent onslaught is the *prime twin conjecture*. It asserts (and we have heuristic and numerical "evidence" but no proof) that there are infinitely many primes p, so that p+2 are both prime. {3,5}, {5,7}, {11,13}, {17,19} are all prime twins.

Problem: Let's say that {p, p+2, p+4} is a *prime triplet* if all three numbers are prime. Show that {3,5,7} is the only prime triplet.

One good ancient way to generate the prime numbers is the "sieve of Erastosthenes". It works by observing that any non-prime >1 is divisible by some smaller prime.

One starts by listing the numbers from 2 to infinity (although the practice is to stop earlier than that). 2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17, 18, 19, 20, 21, 22, 23, 24,25….

At the first step, we bolden 2, a prime, and then remove every second number (i.e. the multiples of 2). **2**,3,X,5, X,7, X 9, X 11, X,13, X,15, X 17, X, 19, X, 21, X, 23, X,25….

Now we notice 3 hasn't been crossed out. So we bolden it and cross out every third number thereafter (multiples of 3). **2**,**3**,X,5, X,7, X, Y, X 11, X,13, X,Y, X 17, X, 19, X, Y, X, 23, X,25…. (There is no need to use a different letter when you cross out the multiples of 3. We do it to foreshadow the coming exercise.)

Then we get 5 and the series becomes:). **2**,**3**,X,**5**, X, 7, X, Y, X 11, X,13, X,Y, X 17, X, 19, X, Y, X, 23, X, Z….

Problem: Observe that the first time a prime p crosses out a new number is at p^2. So to check whether a number k is prime, it suffices to check that it has no divisor >1 and ≤ $\sqrt{k}$

This part now has only numbers that will be boldened, and we get (remove crossed off numbers) **2,3,5,7,11,13,17,19,23**. Notice that each prime we put in our set of primes eliminates infinitely many new numbers (its multiples). As a result, it might not be surprising that their percentage dwindles as we "go to infinity".

Appendix: Two recent theorems.

In the past few years, there were two really remarkable advances on old and basic problems involving the primes. The first is due to Agrawal, Kayal, and Saxena. It says the following:

Theorem: There is an algorithm to test whether a number is prime that is polynomial in the length of the integer.

The length of an integer is how long it is when "written out"[42]. The length of 123456789 is 9, while the length of 123 is 3 and 12345678910 is 11.

The content of the theorem is this: To check for primality just by the methods we discussed, for a number k we have to do about $\sqrt{k}$ steps to look for divisors. So for 123456789 one would be looking forward to around 11,111 things to check. But as the length is 9, the theorem says that the bound better determined by the length than the number itself.

(Actually, the polynomial is currently of degree 6, and $9^6 = 531{,}441$ so I have cheated you. Still, for really large numbers this is a major advance.)

Later on in this chapter, we will discuss the Euclidean algorithm that finds the greatest common factor of two numbers, a challenge that also naively should involve the size of the numbers, but only involves the length of the smaller of the two.

The second theorem I would like to discuss concerns *arithmetic progressions*. An arithmetic progress is a sequence of integers, where any consecutive pair have the same difference.

1, 3, 5, 7, 9, 11, 13, 15….
54, 60, 66, 72, 78, 84, ….
7,17,27,37,47,57,….

are all arithmetic progressions. We will also be interested in finite arithmetic progressions, i.e. where we stop after a while.

1,4,9,16,25,36,….

[42] This idea can be expressed in terms of logarithms, but it is not necessary for our purposes to do so.

Is not. There are three obvious questions to ask about arithmetic progressions and primes. The first is: which arithmetic progressions contain primes? And was proven in the 19th century.

Theorem (Dirichlet, 1805-1859) An arithmetic progression (of positive integers) {a, a+d, a+2d, a+3d….} has infinitely many primes if and only a and d[43] have no common factor.

So, of the examples, above the first = {odd numbers} has infinitely many primes as does {10n-3} (the third). The second does not, because every element is a multiple of 3 [44]. (According to a theorem of Fermat, there is no arithmetic progression of length 4 consisting entirely of squares: 1, 25, 49 is an example of a progression of length 3.)

The next problem is whether there is an arithmetic progression that consists entirely of primes?

Problem: Show that any arithmetic progression of positive primes with difference d has length at most 2d-1.

Recently, by an intricate argument building on work of many other mathematicians from disparate areas, the following was shown:

Theorem (Green and Tao): There are arbitrarily long arithmetic progressions that consist entirely of primes.

Problem: Show that there are arbitrarily long progressions of composite numbers, with any given difference d.

For extra credit, construct arbitrarily long progressions consisting of integers that have no factor in common.

Is there such a progression of infinite length? (How long an arithmetic progression can you find containing the number 2? the number 11?)

2. **Commutative Groups.**

[43] So a denotes its first member and d the common difference.

[44] We could have also given the more obvious reason that they are all even.

Our first aim is to abstract some of the properties of addition of whole (or real) numbers and discover things about that system.

Definition: A commutative group is a set S, with a function called +: SxS → S. (Such a function is called a *binary operation*.) We will write a+b instead of +(a,b). We will assume that + satisfies the following conditions.

a) (identity) There is an element of S, called 0, so that a+0 = 0+a = a for any a.

b) (inverses) For any b, there is an element called –b, so b+(-b) = -b +b = 0.

c) (associativity) For any a,b, and c, a+(b+c) = (a+b) + c

d) (commutativity) For any a and b, a+b = b+a.

These are all well known properties of addition in our usual settings. They apply to usual addition if S = **Z, Q**, **R**, or **C**[45]. If S = **N**, then axiom b fails: almost no elements have inverses.

Here's a different one: Let S = **Q**-{0}, nonzero rational numbers, and for the binary operation we'll use multiplication. (I am loathe to write + = * because it just violates deep seated cultural habits.) The "identity element" that we called "0" in the axiom system is the number we call 1. The meaning of "-b" in this system is 1/b.

If we hadn't removed 0 from **Q**, the inverse axiom would have failed. 1/0 is meaningless.

Yet another example can be made artificially. Let me pick the underlying set to be {agree, disagree}. We define * their "interaction" by:

*	agree	disagree
agree	agree	disagree
disagree	disagree	agree

This is clearly commutative. (When you make an "addition table", commutativity is expressed by symmetry with respect to the diagonal.) The identity "0" is given by agree.

Associativity involves checking 8 equations. A typical verification would go like:

[45] **C** denotes the complex numbers. We will review them in section 4 below.

(agree*disagree)*disagree = disagree*disagree = agree
agree*(disagree*disagree) = agree*agree = agree.

Hence: (agree*disagree)*disagree = agree*(disagree*disagree).

Notice that since we do insist on inverses in a commutative group, I never have to define the other operation: subtraction. We define a-b = a + (-b). We add the additive inverse of b to a.

To give you a feel for the axiomatic method, let us prove:

Theorem 1: In any commutative group, any element a has only one additive inverse, and moreover -(-a) = a.

Proof. Suppose a+c = 0 = a+b, then c = 0 + c = (b+a) + c = b +(a+c) = b+0 = b. So c = b and we win the first part. For the second, note that a + (-a) = 0 which means that a is *an inverse* to –a so it is what we mean by –(-a).

Perhaps associativity is less familiar to you than commutativity. In practice, we will occasionally (notably in the next section, when we discuss symmetry) be willing to give up on commutativity, but never will we give up on associativity.

Another completely different example is given by **S** which consists of “angles”. Angles are like real numbers, except that 360° = 0°. (**S** stands for circle, the way I spell.)

The associative law takes a little thought: later we will give some general principles that make its verification straightforward. For now one has to worry about whether it matters on a case-by-case basis in an expression like (a+b) + c and a + (b+c) which of the various sums add up to 360 or more.

S has a funny property that we are not used to from our other examples. 180° + 180° = 0°. It is possible to have s+s = 0 without s = 0 in a group.

(By the way the element -1 is the multiplicative group of non-zero rationals also has the same property: **S** is different, 120°+120°+120° = 0°, but there is no element t like this in the rational numbers, a non-unit so that t+t+t = 0.)

S also has an interpretation in terms of a clock: we often count time using a clock, and the time is told by what angle the hand makes the midnight point. Just for hours, the minute hand is defined by the rule that 1 minute = 6°. For the hour hand (which only tells us about "half a day") 1 hour = 30°. You can tell minutes using the hour hand, if you have good enough vision, by 1 minute 1/2°.

If we just look at the hour times on **S** then we can label them 1,2,3,….12 (=0). Then addition is the same, except we cast away (multiples of) 12. Here we have a commutative group, usually denoted by $\mathbf{Z_{12}}$, with 12 elements in it.

At least the aspects of addition that we have so far discussed do not require an infinite set. In the problems, you will see, by contrast, that having in addition an ordering that behaves well with respect to addition, does force infinitude (or triviality).

By the way, there is an even simpler example of a finite commutative group: S = {0}. You have no choice about how to define + or what the identity is. It's a dumb example, and we call it the trivial group, but it is an example.

Definition: Let S be a commutative group, and let n be an integer. If n = 0, then we define ns = 0. If n>0, inductively define ns = s + (n-1)s. If n<0, define ns = -n(-s).

Problem: Prove (by induction) that for any abelian group (S, +), and all integers n and m, The formula (n+m)s = ns + ms is valid.

Theorem: If S, + is a finite abelian group, then for each s∈S, we can find an integer n>0, so that ns = 0.

In $\mathbf{Z_{12}}$, this number is different for different s: we can take n=3 for s = 4 and n = 12 for s = 5. In fact, n = 12 works for any element of $\mathbf{Z_{12}}$. The smallest (positive) n that works for an element s is called its *order*. We will denote it *ord(g)*.

Proof of theorem: Suppose #S = k. Then among the elements s, 2s, 3s,….(k+1)s there must be some coincidence (by the pigeonhole principle). One can't put k+1 **different** things in a set with k elements.

Suppose ps = qs with p>q. Then ps-qs = 0, but that means (p-q)s = 0, so we can set n = p-q. QED

We have proven more: we have shown that the order of each element is at most #S. Later we will prove even more: the order of each element divides #S, we will defer this to the next section.

Problems:

Definition: An ordered abelian group (G, +) consists of an abelian group with a total order, <, so that if a<b, then for any c, a+c < b+c.

1. Show that if 0 < a, then –a < 0.
2. Show that if G is an ordered group, then if ng = 0 for some n≠0, then g = 0.
3. Show that **Z** has *two* orderings. As does **Q**.

Here are some more examples of abelian groups.

For any integer, we can form $\mathbf{Z}_n$. We can think of it as {s∈**S** | ns = 0}. We can think of it as being the multiples of 12/n hours (shall we call it a *nour*?) n-nours = 0 nours. Maybe, better would be to call it 1 nour. 1+1+….+1 n times = 0.

So, for each integer n, there is a commutative group with exactly n elements.

Here is another construction. If A and B are commutative groups, then AxB can be thought of a group, by defining 0 as the pair (0.0), and addition by (a,b) + (a',b') = (a+a', b+b'). If A has n elements and B has m, then AxB has nm elements.

3. **Subgroups, quotients, and isomorphism.**

If A is a commutative group, and B⊂A is a nonempty subset with the property that b-b' ∈ B whenever, b,b' ∈B. Then we observe:

(1) 0 ∈ B

(2) b ∈ B if and only if –b∈ B.

(3) If b and b' ∈ B, then b+b' ∈ B.

Such a subset is called a *subgroup*. It is a group in its own right, and a subset of A (and the meaning of + is the same for the two sets). Notice that the multiples of any given integer n, n**Z**, is a subgroup of **Z**. The way we defined it in the last section, $\mathbf{Z}_n$ is a subroup of **S**.

Theorem: If B and C are subgroups of A, then so is B∩C.

Problem: What is 2**Z** ∩ 3**Z**? What is 4**Z** ∩ 6**Z**?
Problem: Show that 2**Z** ∪ 3**Z** is not a subgroup.
Problem: If G is a commutative group, show that nG = {g∈G | ∃h so that g = nh} is a subgroup of G.
Problem: If G is a commutative group, show that n-tor(G) = {g∈G | ng = 0} is a subgroup.
Problem: If G is a commutative group, show that tor(G) = {g∈G | ∃n ∈ **N**, so that ng = 0} is a subgroup. Hint: If ng = 0 and mh = 0 then nm(g-h) = 0 (Why?).
Problem: What is tor(**S**)? What is tor(**Q** –{0),*)? What is tor(**R)**?
Problem: Similiarly, what is n**S**, n(**Q** –{0)), n**R**? (These are not at all equally easy! Depending on what you mean by an answer, the case of **Q**-(0) can be quite difficult at this point.)

Theorem: If G is a commutative group, and A ⊂ G, then there is a "smallest" subgroup containing A; we will denote it by 〈A〉. Thus, A ⊂ 〈A〉, the latter is a subgroup, and if A⊂B and B is a subgroup, then 〈A〉 ⊂ B. We call this subgroup the *subgroup (of G) generated by A*.

Proof: G is a subgroup of G containing A, so there is at least one such subgroup. Consider the intersection of all subgroups of G containing A. That will be a subgroup of G containing A, and it will be included in any subgroup that contains A. QED

Example: If A = {g} contains just one element, then the subgroup generated by g is {ng | n ∈ **Z**.}. (Note that this might be a finite set. Indeed, # 〈A〉 = ord(g). (Why?)) Clearly, any group containing g must contain all of these elements, and further this set is a group under a addition, so it is the group generated by g.

Now we can give an important way of getting new groups from old: the quotient construction.

Definition/Theorem/Notation: Suppose A is a commutative group and B is a subgroup of A, then we define an equivalence relation on A by a~a' if a-a' $\in$ B. Let us denote the equivalence class containing an element a, by [a]. We define [a] + [a'] to be [a+a'] . This is a group with identity = [0]. We will denote this group by A/B.

This is a theorem, because we have to check that everything makes sense. So let us do so. First suppose that a~a' and b~b', then we want to see that a+b ~ a'+b' (so that + "repsects" the equivalence relation ~). But,

a+b – (a'+b') = (a-a') + (b-b')

which is a sum of two elements of B, and hence lies in B. Therefore (a+b) ~ (a'+b'). That [0] is an identity element is obvious. The inverse –[a] is clearly given by [-a]. Associativity follows from associativity for A. QED

Let think about **Z**/12**Z.** It has elements [0], [1], [2]….[11]. It has [12] as well; but [12] = [0].

Indeed, **Z**/12**Z** "feels" exactly like $\mathbf{Z}_{12}$ the group of hours on the clock --- except that we write [k] on the left and k o'clock on the right. Note that **Z**/2**Z** is similarly reminiscent of the "Agreement group" of the last section, where 0 corresponds to agree and 1 to disagree. Both of these have the same feel as {odds, evens} where 1 corresponds to odds and 0 to evens.

Let us make precise this idea of groups corresponding to each other.

Definition: If G and H are commutative groups,, we say that G and H are *isomorphic* if there is a bijection f: G → H so that

(1) f(0) = 0

(2) f(g+g') = f(g)+g(g')

The function f will be called an *isomorphism*. We will also use the notation G ≈H.

Isomorphism for groups is quite similar to bijection for sets. For sets, we have no other structure to keep track of, other than the elements. For groups, we want to also keep track of the way the elements interact with each other by addition.

Note that the meaning of 0 in the left and right hand sides of equation (1) is different. On the left, 0 is an element of G and on the right, it is an element of H. Similarly, the meaning of + is different on the two sides of equation (2).

You might want to think about why the examples we gave above are examples of isomorphic groups.

Here is a less obvious isomorphism. $\mathbf{Z}_{12} \approx \mathbf{Z}_3 \times \mathbf{Z}_4$

INSERT HERE PICTURE OF CLOCKS

Recall the meaning of the group on the right. It consists of pairs of "numbers"[46] say (2,2) and (1,3). We add them coordinatewise: (2,2)+(1,3) = (3,5) = (0,1). 3 = 0 in the first coordinate since that is the $\mathbf{Z}_3$ coordinate. 5 = 1 in the second coordinate, since that coordinate is a $\mathbf{Z}_4$.

The isomorphism is defined by [1] → (1,1). Everything is forced by that: we must have [n] → (n,n). We must see that this is an isomorphism. It will soon follow from general principles, but for now, we can just make a table.

0	1	2	3	4	5	6	7	8	9	10	11
(0,0)	(1,1)	(2,2)	(0,3)	(1,0)	(2,1)	(0,2)	(1,3)	(2,0)	(0,1)	(1,2)	(2,3)

This means that we can tell time with two watches exactly as well as with one, provided that the pair of watches is designed to have one turn around once each three hours and the other every four hours.

On the other hand $\mathbf{Z}_4$ is not isomorphic to $\mathbf{Z}_2 \times \mathbf{Z}_2$. Why not? $\mathbf{Z}_4$ has only one element of order 2, namely [2], while $\mathbf{Z}_2 \times \mathbf{Z}_2$ has 3 of them, namely (1,0),(0,1), and (1,1).

[46] I am deleting from the notation the [ ]'s we used above, since it only makes it clunky.

Exercise: Show that if f: G → H is an isomorphism, then ord(g) = ord (f(g)) to complete the above argument.

Problem: How many commutative groups can you construct with 8 elements (up to isomorphism)? What about with 16 elements?

The following theorem is extremely useful and also justifies the notation G/A for the quotient group.

Theorem: If A is a subgroup of a finite group G, then #(G/A) = #G/#A.

Proof: We have a function f: G → G/A, f(g) = [g]. It is clearly onto. To prove the theorem, by the multiplicative principle, it suffices to show that for all [g], {h | f(h) = [g]} = {h | [h] = [g]} } has the same number of elements. Because, for [e], this set is exactly A, and we would obtain:

$$\#G = \#(G/A) * \#A.$$

However there is an obvious function j:A → {h | f(h) = [g]}, given by j(a) = g+a. We must therefore show that this is one to one and onto. Suppose j(a) = j(b), then g+a = g+b, so a = b and j is 1-1. Now suppose that h ∈ {h | f(h) = [g]}, then g-h ∈ A. Consequently, so is h-g (= -(g-h)). j(h-g) = h-g+g = h, so j is onto. QED

Corollary 1 (**Lagrange's theorem,** Lagrange, 1736-1813). If G is a commutative group, then for any g, ord(g)|#(G).

Corollary 2: If G is a commutative group with #G = p a prime, then G ≅ $\mathbf{Z}_p$.
Proof: Let g∈G be a nontrivial element, then, by Lagrange's theorem, ord(g) = p, so every element of G must be of the form ng for some n =1,2,3,….p. Therefore the function f: $\mathbf{Z}_p$ → G given by f([a]) = ag is onto. It is therefore an isomorphism (since $\mathbf{Z}_p$ and G have the same number of elements.)

Problem: Show that any commutative group of order pq, where p and q are different primes is isomorphic to $\mathbf{Z}_{pq}$. (Hint: If ord(g) = p and ord(h) = q, what is ord(g+h)?)

Corollary 3: In $\mathbf{Z}_p$, ord([k]) = p for any k not divisible by p.

Exercise: Verify this.

Let us tease out some of the significance of this corollary (that we will later reprove by other methods). Since ord[k] = p, there is some n so that n[k] = [1]. In other words,

Corollary 4: If p is a prime number and p does not divide k, then there are numbers n and m so that

$$nk - mp = 1$$

Proof: That is exactly the meaning of the statement that n[k] = [1]. QED.

Corollary 5: If p is a prime number and p|qr, then p|q or p|r,

Proof: Suppose p does not divide q, and n and m are as in the corollary. Then

$$r = nqr - mpr$$

Since, by hypothesis p|qr it divides the first summand, and since it obviously divides mpr, it divides r, as we had claimed.

We now can prove the fundamental theorem of arithmetic, whose statement we now recall.

Fundamental Theorem of Arithmetic. There is one and only one decomposition of an integer as a product of primes, up to factors of ±1 and the order of the factors.

We had already shown the existence of a prime decomposition for any integer. We are left with the uniqueness issue, which we will do by induction. Suppose n is the first natural number with two prime decompositions:

$$n = p_1p_2\ldots p_k = q_1q_2\ldots q_l.$$

Without any harm, let's assume that all of the primes are positive. If p_1 were among the q's then we can divide the two compositions by this prime and a smaller number (n/p_1) with two different decompositions.

Since p_1 divides $q_1(q_2\ldots q_l)$, it divides either q_1 – in which case $p_1=q_1$ as q_1 is also a positive prime number--- or it divides $(q_2\ldots q_l)$. This is a product of a smaller number of primes, so, clearly, by induction on l, we see that p_1 must actually be among the q's, which proves the theorem. Q.E.D.

Appendix: Some "Applications"

The somewhat abstract material that we have studied so far, about the genral theory of finite commutative groups, has had a very good payoff in our understanding of the integers. At this point, we deserve to get some more interesting consequences.

Theorem (Fermat's, 1601-1665, "little" theorem): If p is a prime number, then for any a not divisible by p, $a^{p-1} = 1$ in $\mathbf{Z}_p$.

People often write = in $\mathbf{Z}_n$ as = mod n (read *modulo n*).

Proof: The idea is to break the habit of thinking that + means "plus". Recall, it is just a binary operation with certain properties.

Let us consider the nonzero elements of $\mathbf{Z}_p$ with the operation * of multiplication. Denote this by $\mathbf{Z}_p^*$. It is a group by corollary 4 above (1 is the identity and that corollary provides us with inverses).

Exercise: Show that if a = a' mod n and b = b' mod n, then ab = a'b' mod n (so that multiplication actually makes sense).

So we shall now write things multiplicatively. For any a, $a^{ord(a)} = 1$ (by the definition of order). By Lagrange's theorem ord(a)|#($\mathbf{Z}_p^*$) i.e. p-1 = kord(a) for some k. Hence,

$$a^{p-1} = a^{kord(a)} = (a^{ord(a)})^k = 1^k = 1 \text{ mod } p,$$

proving the theorem.

Example: Let's consider the number 5*7 = 35, that we secretly know to be composite. We can apply Fermat's theorem to see this *without actually finding any factors*.

Let's compute 2^{34}.

This is a moderately large exponent, too high for my calculator to handle, so it gives us an opportunity to take advantage of tricks of modular arithmetic. $2^6 = 64 = -6$ mod 35. $2^{12} = 36$ mod 35. $= 1$. So $2^{36} = 1$.

$2^{34} = 4^{-1}2^{36} = 4^{-1}$

which is not 1.

Remark: Computing powers in modular arithmetic is quite fast. This together with choosing random a's gives rise to a probabilistic test for primality. Nowadays many algorithms are "probabilistic" in nature and are faster than any known method that avoids randomness.

Exercise: What are the last two digits of 102^{128}?

A completely different sort of application is to the classification of *Pythagorean triples*. Here the question is to understand all of the solutions in integers of the equation

$$A^2 + B^2 = C^2 \qquad (*)$$

These are the possible lengths of the sides of right triangles, if we ask that the sides all have integer length. Some examples are (3,4,5), (5,12,13), and (8,15,17).

We can assume that A,B and C have no common factors, since given one solution, we can multiply all three by the same common factor and get another solution.

Exercise: Show that if 2 out A,B and C have a common factor, then all three of them have that factor.

Exercise: Consider {(A/C, B/C)} that are rescaled Pythagorean triples (also known as the pairs of rational numbers whose sums of squares are 1). Show that they form a commutative group under the operation (a,b)*(c,d) = (ac-bd, ad+bc). The unit is (1,0).

Theorem: We can generate all Pythagorean triples (up to order of A and B) via the formula

$$A = m^2 - n^2,\ B = 2mn,\ C = m^2 + n^2$$

Where m and n have no common factor and exactly one of them is even.

Exercise: For which m and n do we get the Pythagorean triples mentioned above.
Exercise: Show that every odd number is the shorter leg of some right triangle.

Proof of theorem: We can assume that B is the even number. We will write the equation (*) as $B^2 = C^2 - A^2 = (C+A)(C-A)$. Notice that C and A have no common factor.

For a product of two numbers to be a square, the fundamental theorem of arithmetic implies that they are each squares (or have a common factor).

Exercise: Prove this!

Notice that if C and A have no common factor, neither can C+A and C-A --- aside from 2 that, indeed, must be a common factor as A and C are odd. (If k|(C+A) and k|(C-A), then k|2C by adding and k|2A by subtracting.)

Thus

$C-A = f^2$

$C+A = g^2$

Where f and g have no common factor, except, perhaps 2. So $2C = f^2 + g^2$. and $2A = f^2 - g^2$ with B = fg. However, since B is even, we can assume that at least one of f and g is even. However, the equation $2A = f^2 - g^2$ then immediately implies that the other is even as well. We write m = f/2, n =g/2, and that completes the proof of the theorem.

QED.

This theorem is one of the first in the subject of *Diophantine equations* that looks for solutions to polynomial equations in many variables in integers. It has been at the core of number theory for millennia.

Our next applications are to the irrationality of certain numbers.

Theorem: If p is prime, then $\sqrt{p}$ is not a rational number.

Proof: For p =2 the following argument can be expressed in terms of adds and evens and is likely to be familiar to you.

Suppose $\sqrt{p} = a/b$ is a fraction, written in lowest terms (i.e. a and b have no common factors).

Then, squaring and clearing denominators we have

$$a^2 = pb^2.$$

Since p divides a^2, p|a, so a = pa'. Plugging into the equation we get

$$p^2a'^2 = pb^2.$$

i.e. $b^2 = pa'^2$. This then forces p|b, so p is a common factor of a and b giving a contradiction. QED

Note that this theorem can also be phrased as saying that the Diophantine equation $a^2 = pb^2$ only has the trivial solution of a=b=0.

Exercise: Prove that if n is not the square of an integer, then $\sqrt{n}$ is not a rational number.

The next application concerns the irrationality of logarithms. If you do not recall them now, you can skip this part and perhaps return to it after we review them in the last chapter. If x and y are (positive real) numbers then $\log_x y$ is that number which when x is raised to that exponent gives y. In a formula:

$$x^{\log_x y} = y.$$

So $\log_{10}100 = 2$, for example. (The trickery is necessary for situations where this quantity is not an integer, and we have to think about what the meaning of exponentiation some crazy number of times means.)

Theorem $\log_{10}2$ is irrational.

Proof: As usual, we will argue by contradiction. Suppose $\log_{10}2 = p/q$. Then

$$10^{p/q} = 2$$

raising both sides to the qth power, gives

$$10^p = 2^q.$$

Since 5 divides the left hand side and not the right we obtain a contradiction.

Exercise: Show that, more generally, $\log_m n$ is rational if and only if there is some integer d, so that both m and n are integer powers of d.

This has an interesting implication for the length of powers of 2. The length of powers of 10 is really easy: 10^n has n+1 digits. The number of digits that 2^n has is the

smallest integer larger than $n\log_{10}2$ ($=\log_{10}2^n$). There is no simple rule, like for each increase of 10 in n, the number of digits increases by 3. That would cause the logarithm to be 3/10. The true value is .3010299…. (This rule of thumb does work pretty well for $n<1000$ because .3 is a pretty good approximation. Of course 2^{999} is a big enough number that you might have given up calculating your exponential tables or had made a miscalculation by then.)

Calculating the number of digits of 2^n for large n's boils down to computing the digits of $\log_{10}2$ – but with a big asymmetry. k –digits of the logarithm allows exponents all the way up to 10^k. The thoughtful reader might now wonder what tools can be brought to bear on the calculation of logarithms. While we will not be able to confront that in these notes (the only really good methods are based on calculus) we will later see some tools that we will apply to easier problems, that when sharpened can be used for this problem.

Appendix: Noncommutative Groups and Symmetry (*)

After all this time discussing commutative groups and eschewing the simpler adjective-free appellation group, I would be remiss if I didn't discuss the theory of groups that are not necessarily commutative. They have a much richer theory, but do taking getting used to. Some of their properties are generalizations of what we have seen in the commutative case, and some aspects are quite different.

Groups come into their own as a way of organizing symmetry, so perhaps we should start by considering a fairly symmetric object, the equilateral triangle.

PICTURE HERE

* This section is more demanding than the other sections of these notes, as it builds on a firm understanding of the previous sections --- the reader might have to reread it several times, with a paper and pen at her side, before being able to follow all of the arguments or do the exercises. The remainder of these notes do not use any of these ideas.

It has three vertices a,b,c and sides ABC opposite each. There are a number of symmetries: K and K^{-1} that are rotation clockwise and counterclockwise 120°.

These symmetries are all functions from the triangle to itself, and we can compose them: Remember that when we compose functions fg means first do g and then do f.

These rotations are inverses to each other. Moreover, $K^2 = K^{-1}$. In fact, it makes sense to count I, the identity, ($=K^3$) as a (kind of boring) symmetry as well. There are also 3 reflections across the lines from a to the midpoint of A, from b to the midpoint of B, and from c to the midpoint of C. I'll call the first one R_A and leave it to you to invent notation for the other 2.

We can make a multiplication table out of the six symmetries that we have found:

*	I	K	K^{-1}	R_A	R_B	R_C
I	I	K	K^{-1}	R_A	R_B	R_C
K	K	K^{-1}	I	R_C	R_A	R_B
K^{-1}	K^{-1}	I	K	R_B	R_C	R_A
R_A	R_A					
R_B	R_B					
R_C	R_C					

Exercise: Fill in the rest of the table.

If you think about it, actually the symmetries of the triangle are exactly the same as (or determine and are determined by) what they do to each of the vertices. This set of symmetries is the set of bijections {a,b,c} ↔ {a,b,c}. This is the set of permutations of 3 letters.

If you recall, we showed that functions always satisfy the associative law. The symmetries of the triangle, though does not satisfy the commutative law. (Compute KR_A and R_AK.) Therefore, it seems reasonable to encode the study of symmetry by the following axioms:

Definition: A group is a set G with a binary operation *:G x G → G, so that

(1) there is an identity element e, so that eg = ge = g for all g.

(2) for every g there is an element h, so that gh = hg = e

(3) for all g,h, and k, (gh)k = g(hk).

As before, each element has an order, so that $g^{ord(g)} = e$. If G is finite, this order is finite. It is also the case that every subset is contained in a smallest subgroup containing it.

Lagrange's theorem is also correct, but the proof is somewhat different than before. It is not the case that given any subgroup H of G we can define G/H as before and define a group structure on this. Multiplication is not always well defined.

Let us define G/H to be equivalence classes as before g~g' if there is an h in H, so that g = g'h. This is reflexive (h = e lies in the subgroup H), symmetric (g =g'h if and and only if gh^{-1}=g', and h^{-1}lies in H if h does), and transitive (g =g'h and g' = g"h' implies that g = g"h'h, and h'h is in H if h and h' are, as H is a subgroup).

So we have equivalence classes G/H. We cannot multiply them in a well defined fashion. Consider gg' and ghg'h'. The outside h' is fine: it doesn't change the equivalence class. However, there is no reason for [gh'g] to equal [gg'].

The argument we gave for Lagrange's theorem still applies, once one recognizes that we don't need a group structure on G/H: we only needed it to be a set and for the all of the equivalence classes to have the same size (they do: it is #H). Then the argument follows as before.

Exercise: Check this.

In order to define a group structure on G/H we need an extra condition. We embody this in a definition:

Definition: A subgroup H is a *normal subgroup* of G if for each h in H and each g, ghg^{-1} is an element of H.

Once we have this, then or problem regarding [gh'g'] goes away:

$$gh'g' = g(g'g'^{-1})h'g' = gg'(g'^{-1}h'g') = gg'h''$$

the last equality using the definition of normality. As a result [gh'g'] = [gg'], and the defintion of multiplication on G/H is well defined.

I will close this cursory introduction with a useful formula that often enables one to either do counting problems or prove results about groups. It uses the idea of a group action on a set. Suppose that G is a group and S is a set. An action of G on S consists of an assignment of a bijection b(g):S ↔ S for each g. The key points are that b(e) = id_S and b(gg') = b(g)b(g').

When there is only one group action involved, people will usually delete the b( )'s and denote the action by expressions like gs and so on.

Given a group action on S, and s ∈ S, then there are two important associated objects. The first is the *orbit* of s. It is {t ∈ S | t = gs for some g in G}; it is a subset of S. Notice that two orbits are disjoint or coincide: i.e. S is partitioned by the orbits.

Exercise: If H ⊂ G is a subgroup, it can act on G in two ways: either b(h)g = hg or b'(h)g = gh^{-1}. Check that these both define group actions. What are the orbits of these actions?

The second important object is the *stabilizer subgroup of s*, defined by G_S ={g ∈ G | s = gs}.

If O(s) denotes the orbit of s, then there is a bijection $G/G_S \rightarrow O(s)$. Just send g to gs. If we change g to gh, where h is in the stabilizer, then ghs = gs, as h(s) = s. Thus the function is well-defined.

Exercise: Show that this is a bijection. (Surjectivity is obvious; injectivity is only slightly less so.)

As a result #O(s) = #G/#G_S.

As a consequence of this, we then have

$$\#S = \sum \#G/\#G_S.$$

where the sum is taken over a set of representatives, one from each orbit.

Corollary: If P is a group of prime power order p^n and P acts on a set S, then

$$\#S = \#\{s \in S \mid gs = s \text{ for } \mathbf{all}\ g\} = S^G \mod p.$$

Proof: Each G_s has order p^k, so unless $G_s = G$, $\#G/\#G_s$ is divisible by p. These s's are exactly the elements of S^G. As a result, $\#S - \#S^G$ is divisble by p. QED

We call S^G the *fixed set* of the group action. The above corollary gives a useful condition for the existence of a fixed set even in situations where we know very little about the details of the action.

A beautiful consequence of this is a theorem of Cauchy (1789-1857).

Theorem: If H is a group, and p|#H then H has an element of order p.

Proof: We will let S be the subset of G^p consisting of p-tuples $(g_1,\ldots g_p)$, so that their product = e, i.e. $g_1,\ldots g_p = e$.

$\mathbf{Z}_p$ acts on this by having the generator send $(g_1,\ldots g_p)$,to $(g_p,g_1,\ldots g_{p-1})$. Note that $g_p g_1,\ldots g_{p-1} = g_p(g_1,\ldots g_{p-1}g_p)\, g_p^{-1} = g_p(e)\, g_p^{-1} = g_p g_p^{-1} = e$, so that we really do take elements of S back into S.

$\#S = (\#G)^{p-1}$ and is therefore divisble by p. The fixed set of the $\mathbf{Z}_p$ action exactly consists of the p-tuples of the form (g,….g) whose product is e, i.e. the elements of order 1 or p. Since (e,e,…e) lies in this fixed set, there must be at least p-1 other elements and, in particular, G contains an element of order p. QED.

Exercise: Suppose that G acts on S and s and t lie in the same orbit. What is the relationship between G_s and G_t?

Exercise: Let $G = S_n$ be the group of bijections of {1,2….n}. #G = n!. Now let G act on the $P(|n|)$ What are the orbits? What are the isotropy groups? What does the formula for $2^n = \#P(|n|)$ boil down to?

Question: How does the previous exercise give another viewpoint on the formula for the number of ways of choosing k elements from a set with n elements?

Exercise: Give a group of order 4 with no element of order 4.

Exercise: Suppose that #G is a prime power. Let S = G, and $b(h)g = hgh^{-1}$. Show that this is a group action and that S^G contains more than one element. Deduce that such a group contains an element g (other than the identity) so that gh = hg for all h in G. On the other hand, show that S_n does not contain such an element whenever n>2.

4. Rings and Fields

We are now ready to deal with algebraic systems that closely resemble the integers, **Z**. These systems are called *rings*.

Definition: A *ring* R, consists of a set with two binary operations, + and *, with special elements 0 and 1 that satisfy the following axioms:

(1) (R, 0, +) is an abelian group.
(2) * is a commutative binary operation on R, with 1 as an identity,
and that satisfies the associative law.
(3) The distributive law holds a*(b+c) = a*b + a*c.

We will often delete the * an signify a*b by ab.

Notice that we do not ask for multiplicative inverses for elements of R. The rings that abstract the peculiar features of **Q** are called *fields*.

Definition: A *field* F is a ring where every nonzero element has a multiplicative inverse (i.e. if r ≠0, there is an s, so that rs = 1.).

Let us begin with some elementary ring arithmetic.

Proposition 1: 0r = 0 for all r.
Proof: 0r = (1-1)r =1r-1r = r-r = 0.

Proposition 2: If 1 = 0, then R consists of just one element.
Proof: r = 1r = 0r = 0.

Definition: r is a *unit* or is *invertible*, if there is an s so that rs = 1.

Proposition 3: If r is a unit and ra = rb, then a = b.
Proof: a = sra (where s is the inverse to r) = sra –srb +srb = sr(a-b) +srb = a-b+b = b.

Proposition 4: If R is a ring, then its units form a commutative group under multiplication.

Definition: A *0-divisor* is a nonzero element r so that for some nonzero element s, rs = 0. 0-divisors are sort of the opposite of units.

Now for some examples:

Example 1. **Z** is a ring. It has no 0-divisors, but its only units are ±1.
Example 2. **Q** and **R** are rings where all the non-zero elements are units, i.e. they are fields.
Example 3. $\mathbf{Z}_n$ is a ring which has 0-divisors if and only if n is composite.

Proof: If n is prime, we have already seen that p|ab only if p|a or p|b which boils down to the absence of 0-divisors. If n = ab where a and b are nontrivial divisors (neither = ±n), then 0 = [a][b] shows that $\mathbf{Z}_n$ has 0-divisors.

Exercise: If n is prime, show that $\mathbf{Z}_p$ is a field. (Below we will show that any finite integral domain is a field.)

Example 4. **Z**[i] is the *ring of Gaussian integers*. Its elements are symbols of the form a + bi where a and b are ordinary integers. We add the elements as follows: a+bi + c + di = (a+c) + (b+d)i. The multiplication is forced by distributivity, and the rule $i^2 = -1$. Explicitly (a+bi)(c+di) = (ac-bd) + (ad+bc)i.

Exercise: Check that it is indeed a ring.

We claim that the units are ±1 and ±i and also that there are no 0-divisors. This can be done by verifying the following series of claims that I leave to you.
Claim 1: If we define $N(a+bi) = a^2 + b^2$, then $N(a+bi) = (a+bi)(a-bi)$.
Claim 2: For for any r and s in **Z**[i], N(rs) = N(r)N(s).
Claim 3: If $r \neq 0$, then $N(r) \neq 0$.
Claim 4: If r is a unit, then N(r) = 1.
Claim 5: If N(a+bi) = 1 then a = ±1 and b =0 or a = 0 and b = ±1.

Example 5. We can do the same construction with **Q**[i], defined exactly the same way as before, except that a and b are now allowed to be elements of **Q**. Now all nonzero elements are units, and **Q**[i] is a field. $(a+bi)^{-1} = (a-bi)/N(a+bi)$. (Note that the right hand side makes sense using claim 3 for this setting.)

Example 6. If we use **R** in place of **Q**, then we obtain the famous field of complex numbers, **C**.

Example 7. As yet another example, we can form **Z**[√2], **Q**[√2], and **R**[√2], These are **formal expressions** $a+b\sqrt{2}$ which are added and multiplied as before, except that the defining rule is that $(\sqrt{2})^2 = 2$.

We will now notice some differences and similarities.

Claim 1: If we define $N(a+b\sqrt{2}) = a^2 - 2b^2$ then $N(a+b\sqrt{2}) = (a+b\sqrt{2})(a-b\sqrt{2})$.
Claim 2: : For for any r and s in any of our rings, N(rs) = N(r)N(s).
Claim 3: If $r \neq 0$ in **Q**[√2], then $N(r) \neq 0$.

This follows from the irrationality of the √2. It is false in the ring **R**[√2], (Give an example!)

Claim 4: $1+\sqrt{2}$ is a unit in **Z**[√2],. It's inverse is $-1+\sqrt{2}$. As a consequence there are infinitely many units in this ring, e.g. $(1+\sqrt{2})^n$.

These are actually all different.

Exercise: Show that if $= a_n + b_n\sqrt{2} = (1+\sqrt{2})^n$, then for n>0, a_n is an increasing function of n (i.e. if m>n, then $a_m > a_n$). On the other hand , $a_n = a_{-n}$ and $b_n = -b_{-n}$.

We can give a different argument for this using an important idea: the homomorphism. We can write a function f: $\mathbf{Z}[\sqrt{2}] \to \mathbf{R}$ by defining $f(a + b\sqrt{2}) = a + b\sqrt{2}$, Before you think I have gone crazy, let us note that the symbol $\sqrt{2}$ means two different things on the left and on the right. On the left, it is a place keeper. It is an element of this abstract ring we have constructed, so that a certain multiplication table holds. On the right it is the real number 1.4142135…. whose square is the real number 2. So the function f is not crazy.

f has a number of good properties. The key for us now is that it is well defined (duh!) and that f(rs) = f(r)f(s) (which I leave to you). As a result, $f(a_n + b_n\sqrt{2}) = (1 + \sqrt{2})^n$ which are all different numbers (in the real numbers if $x^n = x^m$ for some $m \neq n$, then either $x = 0, \pm 1$ and $1 + \sqrt{2}$ is none of those numbers).

Exercise: Deduce from our work to this point that the Diophantine equation $x^2 - 2y^2 = 1$ has infinitely many solutions.

Claim 5: If $r \neq 0$ in $\mathbf{Q}[\sqrt{2}]$, then r is a unit, so $\mathbf{Q}[\sqrt{2}]$ is a field.

Exercise: What is its inverse?

Example 8. $\mathbf{Z}_p[i]$ where we do as in examples 4,5, and 6 except using elements of $\mathbf{Z}_p$ for the coefficients.

Here the basic questions work out differently for different primes. There are two cases: depending on an analysis of whether the function N is so that N(r) = 0 is possible for $r \neq 0$. If there is such an r, then the ring has 0-divisors. If not, then the analysis above gives that every non-zero element is a unit.

For p = 5, (2+i) is a 0-divisor, but for p = 3 or p =7 all the non-zero elements are units. (We will later see that the answer only depends on p mod 4.)

Example 9. Here is an example that is completely different in spirit than the previous ones:

Let R be any ring, and X be a set, then we can consider F(X: R) the set of functions[47] from X to R. We add and multiply functions in the obvious way: (fg)(x) = f(x)g(x). The constant function 1 is the multiplicative identity, and the constant function 0 is the additive constant. It is easy to check that this is a ring.

Notice that if X has more than 1 element, then this ring has 0-divisors. (Why?)

Definition: An integral domain is a Ring with no 0-divisors.

So $\mathbf{Z_n}$ is an integral domain if and only if n is prime.

Theorem: If R is a finite integral domain, then it is a field.

Proof: Consider *s :R → R for any nonzero s, as we had seen that cancellation is possible for any non-zero-divisor.

Since R is finite, the 1-1 function *s: R → R is onto. So there is an r, so that sr = 1, and that shows that s is a unit. QED

5. Ideals and Homomorphisms

A key to analyzing rings is their *ideals*. We will use them for the construction of new rings and also to help analyze homomorphisms.

Definition: An *ideal* A in a ring R, is a subset that

(1) is closed under addition. (i.e, *A* is an additive subgroup of R).

(2) if r is any element of R and a∈A, then ar ∈A.

Examples:

1 In any ring, R, {0} and R itself are both ideals. We will call {0} the trivial ideal.

[47] The example becomes more interesting when we put some kind of adjective on the types of functions considered. But, we will keep things simple at this point.

2. In **Z** the multiples of any number k, form an ideal, k**Z**. In fact, for any r∈R, we can consider rR, the multiples in R of r= {s∈R | s =rt for some t ∈R}. These ideals are called "principal ideals".

3. In **Q**, and **R** the only ideals are {0} and the whole ring. This is because of the following:

Observation: If an ideal A contains a unit, then A = R
Proof: Let t be the inverse of the unit u, then 1 = tu, is in A (by condition (2)), and therefore r1 = r is also an element of A.

Those rings are among those that we saw wherein all elements other than 0 are units. Such rings are called fields, and will be studied in the next section.

4. In the "function ring" situation we can now define some other ideals that are not necessarily principal.

For instance, we can look at the set of f that vanish (i.e. evaluate to 0) at a given element x of X.

Exercise: Show that this is actually a principal ideal (although we did not define it that way.) On the other hand, if X is infinite, and we instead look at the {f | f(x) = 0 for all but a finite subset of X}, then this is an ideal, and it is not principal for R a nontrivial ring.

The reason that ideals are called "ideal" is because, for some purposes we can think of them as "ideal elements" of the ring. Sometimes, it is convenient to think of elements as being essentially the same as the principal ideal they define, and then we "enlarge our universe" by also considering the non-principal ideals, if there are any.

Theorem: Principal ideals in an integral domain, rR and sR, are equal if and only if there is a unit u, so that r = su.

Notice the hypothesis of the theorem is symmetric so the conclusion of the theorem should be, as well. Observe that it is! We say r and s are *associates* if they differ multiplicatively by a unit.

Proof. Since r∈sR, r =ts. And, s = vr, so it follows that r =vtr.

Corollary of the proof: If $rR \subset sR$ then $s|r$.

Proof: Exercise.
Remember the slogan: To contain is to divide.

Exercise: Show that the intersection of two ideals is an ideal.
Exercise: Show that if A and B are ideals, the $\{a+b \mid a \in A \text{ and } b \in B\}$ forms an ideal, that we will call A+B.
Exercise/Warning: Show that $2\mathbf{Z}+3\mathbf{Z} = \mathbf{Z}$. So don't get too carried away by the notation or by thinking of ideals as "generalized elements".
Exercise: If R is any ring, then $N = \{r \mid r^k = 0 \text{ for some } k\}$ is an ideal. (Hint: Check first that if $r^2 = 0$ and $s^2 = 0$ then $(r+s)^3 = 0$.) This is called the ideal of *nilpotent elements.*
Exercise: Show that if r is a nilpotent element, and u is a unit, e.g. if u = 1, then u+r is a unit. (Hint: Do the case of u =1 first. Also try to find inverses in the special case of $r^2 = 0$ and maybe $r^3 = 0$ before trying to tackle the general case.)

We can use any ideal in a ring to define a new ring.

Definition: If R is a ring and A is an ideal, then R/A is defined as follows:
(1) As a group, it is the group R/A.
(2) To define $[r]*[r']$ we just take $[rr']$.
Of course, we need to see that this definition makes sense, and that it defines a ring. I will do the first, and I leave the second part to you.

Proof: Suppose $[s] = [r]$ and $[s'] = [r']$. That means that we can find $a,a' \in A$, so that $s = r+a$ and $s' = r'+a'$.

Let us compute $ss' - rr'$.

$$ss' - rr' = (r+a)(r'+a') - rr' = rr' + ar' + r'a + aa' - rr'$$
$$= ar' + a'r + aa' \in A$$

the last sum is a sum of three elements of A (why?) and hence lies in A and $[ss'] = [rr']$.

QED.

Exercise: Check that $\mathbf{Z}/n\mathbf{Z} \cong \mathbf{Z}_n$.

The function R → R/A that sends r to [r] has important properties:

Definition: A *homomorphism* between rings is a function f:R → S so that

(1) f(0) = 0 (although 0 means different things on the two sides of this equation)
(2) f(1) = 1.
(3) f(a+a') = f(a) + f(a')
(4) f(rs) = f(r)f(s)

The map **Z[**√2] → **R** we used before is also a homomorphism.

Exercise: Check that it is automatic that f(-a) = -f(a).
Exercise: Check that the function R → R/A that sends r to [r] is indeed a homomorphism.
Exercise: Check that if f: R → S is a homomorphism and g: S → T is a homomorphism, then so it gf: R → T.

Definition/Theorem: If f: R → S is a homomorphism between rings, then ker(f) = {r ∈ R | f(r) = 0} (called the *kernel* of f) is an ideal in R. Moreover there is a homorphism f':R/ker(f) → S defined by f'([r]) = f(r). Moreover, if A is any ideal in R **contained in** ker(f), then we can define a homomorphism h: R/A → S defined by the same equation h([r]) = f(r).

Exercise: Prove the theorem.

As a first application, let's try to analyze **R[**√2], a ring that we saw has 0-divisors. There are two homomorphisms **R[**√2] → **R** that we shall consider: the first sends the symbol √2 to the real number √2. The second sends it to - √2. In other words, we send a+b√2 to (a+b√2, a-b√2).

The point is that the defining property of the ring has $(\sqrt{2})^2 = 2$, something true of two different real numbers. We can use each for defining a homomorphism. (If we had an additional structure, like an order, then perhaps we would not be able to define both homomorphisms preserving that additional structure.)

The kernel of this pair of homorphisms, thought of as a single homomorphism to the product of rings **R**x**R** is the intersection of the kernels of the 2 homomorphisms. The

kernel of the first homomorphism is { $a+b\sqrt{2} \mid a = -b/\sqrt{2}$}. The kernel of the second homomorphsm is { $a+b\sqrt{2} \mid a = b/\sqrt{2}$}. The only (a,b) satisfying both requirements is (0,0)

Observation/Exercise: A homomorphism h is 1-1 if and only if ker(h) = 0.

Thus, we have a 1-1 homomorphism of $\mathbf{R}[\sqrt{2}] \rightarrow \mathbf{R}x\mathbf{R}$. It is not hard to see that it is actually onto, so it is an isomorphism. This isomorphism can then be used to analyze all sorts of questions about the ring $\mathbf{R}[\sqrt{2}]$ that might not have been obvious before.

Exercise: What are all of the 0-divisors of $\mathbf{R}[\sqrt{2}]$? What are the solutions to the equation $x^2 = x$? to $x^2 = 1$? Which $a+b\sqrt{2}$ are squares?

As another application, we see that if m|n then there is a homomorphism $\mathbf{Z}_n \rightarrow \mathbf{Z}_m$. Let us repeat this for all the prime powers dividing an integer n. This gives us a homomorphism

$$j\colon \mathbf{Z}_n \rightarrow \mathbf{Z}_{p_1^{a_1}} \times \mathbf{Z}_{p_2^{a_2}} \times \ldots\ldots \times \mathbf{Z}_{p_k^{a_k}}$$

just by combining these individual homorphisms. Notice that the # of elements on both sides is the same =n. As a result this map is an isomorphism (i.e. is 1-1 and onto) if and only if it is 1-1. In light of the observation above, we would like to check that ker j = 0.

Let [t] be an element of ker j. That means that t, thought of as an integer is divisible by each $p_i^{a_i}$. But, by the fundamental theorem of algebra, this means that t is divisible by their product --- i.e. by n. Said alternatively, [t] = 0 in $\mathbf{Z}_n$, proving that indeed ker j = 0, and hence that j is an isomorphism.

We have proved:

Theorem (The Chinese Remainder Theorem). The ring $\mathbf{Z}_n$ can be described as a product of rings $\mathbf{Z}_{p_1^{a_1}} \times \mathbf{Z}_{p_2^{a_2}} \times \ldots\ldots \times \mathbf{Z}_{p_k^{a_k}}$ according to the prime power divisors.

Exercise: Show that if k and l have no common factor then there is an isomorphism $\mathbf{Z}_{kl} \cong \mathbf{Z}_k \times \mathbf{Z}_l$. On the other hand, if d is a common factor of k and l, then there is a

homomorphism $\mathbf{Z}_k \times \mathbf{Z}_l \rightarrow \mathbf{Z}_d$ so that the image of $\mathbf{Z}_{kl}$ in $\mathbf{Z}_k \times \mathbf{Z}_l$ lies in the kernel of the homomorphism to $\mathbf{Z}_d$.

You can think of the result of this exercise as saying the following. Suppose that I have 3 clocks. One is a 4 hour clock, the second a 3 hour clock, and the last a 1 hour clock. Since 4 and 3 are relatively prime, the result of considering the first two clocks is exactly the same as using the last. Moreover, any pair of positions on the first two clocks occurs for some hour on the third clock. I can specify 2 mod 4 and 1 mod 3 and find an appropriate integer that accomplishes both.

However, if I consider, say, the first and the third. These are not independent. Using the common divisor 2, I note that the parity of the hours that those two clocks show must be the same. If I'd use the common factor 4, then I'd notice that the hour on the 12 hour clock determines the hour on the 4 hour clock.

Exercise: How many positions are possible with a pair of clocks: one with period 9 and the other with period 6.

6. The Euclidean Algorithm

In this section, we will give another proof of one of the main steps in the fundamental theorem of arithmetic, following Euclid. This method yields quite a bit more information.

We will start by working in the ring **Z**. The goal is to understand the set of "common divisors" of two numbers m and n. That is {t | t|m and t|n}. We will see that there is a largest number that has this property and that any number that divides both m and n divides this number.

It will also show that all ideals in **Z** are principal.

Let's start with say m = 130 and n =45. If d|m and d|n then d|(m-kn) for any k. Noting that 130 = 2*45 +40, the common divisor must divide 40. Since d|40 and d|45, d|(45-40)=5, d|5. As 5|45 and 5|130, it is the greatest common divisor.

Here's the algorithm for finding the gcd of integers.

1. Input m and n
2. Sort them so that m≥n.
3. If n = 0, output “the greatest common divisor is m”
4. replace (m,n) by (n, the remainder of m divided by n)
5. goto 3

The process ends, because the numbers keep getting smaller. Indeed, after one step the larger number is at most as large as the smaller was. After a second step, the smaller of the two is at most ½ of what it once was. Consequently, the number of steps it takes to terminate is at most a small multiple (like 3*) the number of digits in the smaller number.

The same algorithm computes the smallest ideal containing m and n. (Notice that the argument we gave before for subgroups applies equally well to ideals. In any case, it is easy enough to see that the ideal generated by m and n is the set of elements of the form ms+nt, or mR+nR, in the notation of the previous section.)

Instead, one outputs on line 3 “the ideal is the principal ideal m**Z**”.

The reason this works is that the ideal generated by m and n is the same as the ideal generated by n and the remainder r of m when divided by n. Because if m = kn+r, then

$$am + bn = a(kn+r) + bn = (ak+b)n + ar$$

so every element in the ideal generated by m and n is in the ideal generated by n and r, and in the reverse

$$an + br = an + b(m-kn) = bm + (a-k)n$$

shows that any element of the ideal generated by n and r lies in the one generated by m and n.

Let us summarize our conclusion in the following important theorem:

Theorem: In **Z** the elements of the form am+bn equals the set of all multiples of an integer d. We have d|m and d|n and any common divisor of m and n divides d.

Note that this includes the fact that if m and n have no common factor then there are an a and b so that am+bn =1. When m is a prime, we had proved this by group theoretic methods --- and it was one of the main steps in our proof of the fundamental theorem of arithmetic.

Corollary: In $\mathbf{Z}_n$ the units are {m| m and n have no common factor}.
Proof. To say [m] is a unit, means that for some a, [am] = [1]. This means that am = 1 +bn, so it is equivalent to am-bn =1. By the theorem, that is equivalent to saying that m and n have no common divisor (other than ±1).

Corollary: (Euler's (1707-1783) **theorem)** If a and m have no common factors, then $a^{\phi(m)} = 1 \bmod m$, where $\phi(m)$ is the number of integers <m that have no common factor with m.

Proof: $\phi(m)$ is the order of the group of units of $\mathbf{Z}_m$ and a is an element of that group. QED.

Exercise: Show that in **Z**, all ideals are principal (not just the ones generated by two elements.)

Let us now apply exactly the same reasoning to a different class of rings.

Definition. Let R be a ring. R[x] is the *polynomial ring* on R, Its elements are things of the for $r_0 + r_1x + r_2x^2 + r_3x^3 + \ldots r_nx^n$ where the r's lie in R (they are called the coefficients of the polynomial). The rules for adding and multiplying polynomials are those of elementary algebra.

$$(r_0 + r_1x + r_2x^2 + r_3x^3 + \ldots r_nx^n) + (s_0 + s_1x + s_2x^2 + s_3x^3 + \ldots s_mx^m) =$$

$$(r_0 + s_0) + (r_1+s_1)x + (r_2+ s_2)x^2 + \ldots..$$

and $(r_0 + r_1x + r_2x^2 + r_3x^3 + \ldots r_nx^n) * (s_0 + s_1x + s_2x^2 + s_3x^3 + \ldots s_mx^m) =$

$$(r_0s_0) + (r_0s_1 + r_1s_0)x + (r_0s_2 + r_1s_1 + r_2s_0)x^2 + \ldots + (r_ns_m)x^{n+m}.$$

Exercise: Show that in $\mathbf{Z}[x]$, show that the greatest common divisor or 2 and x is 1. The ideal generated by 2 and x is not principal.

The interesting observation that I would now like to make is that if F is a field, then F[x] has a version of the Euclidean algorithm, and that therefore all ideals are principle, and furthermore, there is a fundamental theorem of arithmetic for arithmetic in this system.

Let us discuss the main points one by one. I will leave verifications to you to think through.

a) Polynomial have a degree. We can use this to make a notion of division with remainder. If P(x) and Q(x) are polynomials of degree p and q respectively, we can write $P(x) = S(x)Q(x) + R(x)$ where $\deg(R) < q$.

You can do this by the method of long division. If P starts with $a_p x^p$ and Q starts with $b_q x^q$ then S(x) starts with $a_p/b_q x^{p-q}$ and then one does the usual algorithm. (Notice that it is the initial coefficient that requires that we be in a field.) The Remainder = 0 if and only if P(x) is divisible by Q(x).

b) We can now form the Euclidean algorithm, interpreting remainder as in the previous point.

c) The process always lowers the degree of the polynomials involved and therefore must terminate in a finite number of steps.

d) As before, it produces the greatest common divisor, and a single polynomial G(x) whose multiples are exactly the polynomials of the form $A(x)P(x) + B(x)Q(x)$.

Now let us consider the issue of factorization. A polynomial P(x) is *prime* if it is not a product of lower degree polynomials (aside from constants). Every polynomial is a product of Prime polynomials by induction on the degree. As before, the issue is uniqueness.

As before, it suffices to show that if P(x)|Q(x)R(x) then P(x)|Q(x) or P(x)|R(x). We will approach this by considering the analogue of **Z**/p**Z**.

Claim: If P(x) is a prime polynomial, then the quotient ring F[x]/P(x)F[x] is a field. Proof: Let [Q(x)] be an element of F[x]/P(x)F[x]. If [Q] ≠ 0, then P(x) does not divide Q(x). This means, given that P(x) has no nonconstant divisors, that there are polynomials A(x) and B(x) so that A(x)Q(x) + B(x)P(x) = 1. Consequently, [A(x)] is an inverse to [Q(x)]. QED

Since this quotient is a field, it has no 0-divisors, which is exactly what we wanted to prove: that if P(x)|Q(x)R(x) then P(x)|Q(x) (i.e. [Q(x)] = 0) or P(x)|R(x) (i.e [R(X)] = 0.)

With this property of primes, the proof of the fundamental theorem of arithmetic is exactly as before. We can phrase it as follows:

Fundamental Theorem of Arithmetic in F[x]: If F is a field, then every polynomial is a product of a constant and monic prime polynomials. This factorization is unique up to the order of the prime factors.

A *monic polynomial* is one where the coefficient of the highest order term is 1. We can always accomplish this by multiplying by an element of F.

A final comment: The claim above now gives us a way for producing many new fields. For instance, the only finite fields we had seen before are of the form $\mathbf{Z_p}$. We now can produce finite fields with p^2 elements.

Exercise: Show that if P(x) is an irreducible quadratic polynomial, then F[x]/P(x)F[x] is a field with $(\#F)^2$ elements.

Exercise: The number of monic degree one polynomials is #F. Therefore the number of non-rime quadratic polynomials is (#F)(#F+1)/2. (Hint consider the product and study the role of order.). On the other hand, the number of monic quadratic polynomials is $(\#F)^2$. Deduce that there is always some irreducible quadratic polynomial (if #F is finite!).

Remark: An elaboration of the above argument actually shows that for each k there is a prime polynomial of degree k over F[48]. As a result for every prime power, there is a field with that many elements. These are sometimes called Galois fields after the famous and tragic French mathematician Evariste Galois (1811-1832).

Exercise: Show that if F is a finite field then #F is always a prime power.
Hint: Consider ord(1) in the additive group. Argue that it is a prime p. Therefore pf = 0 for all f ∈ F. Using Cauchy's theorem from the appendix on group theory, argue that #F is a power of p.

Remark: Galois fields have entered applied mathematics in the theory of error-correcting codes. These are ways of sending information over channels in such a way that if some errors are introduced during transmission, the receiver can efficiently figure out what the most likely signal had been.

We leave the following to the reader, its proof being essentially the same of another theorem with the same name.

Chinese Remainder Theorem: If P(x) = Q(x)R(x) is a product of polynomials that have no common factor, then F[x]/P(x) ≅ F[x]/Q(x)×F[x]/R(x).

Appendix: Formal Power Series.

There is an interesting variant on the polynomial ring R[x] called the formal power series ring R[[x[[. It is polynomials of infinite degree – we don't insist that they ever end. (They are called "formal power series" because we don't impose any conditions on them so that they actually make sense as functions. We will discuss this point in the last chapter.)

Let us consider this ring for R = **Q.**

---

[48] If p is prime, the number of irreducible degree p polynomials is $((\#F)^p-(\#F))/p$. Fermat's little theorem tells us that this is an integer! You could do worse than trying to verify this for p = 3. Do you see why this counting formula implies the existence of Galois fields of all prime power order?

Theorem. **Q**[[x]] is an integral domain. Moreover, a power series $f(x) \in \mathbf{Q}[[x]]$ is a unit if and only if its constant term is nonzero. The ideals are all principal. They are generated by x^k for some natural number k.

The first statement follows from the second. The second can be accomplished by long division starting at the constant term, and just continuing forever.

These series are completely formal. It is not the case that one can "plug in" any value of x at all except x = 0. You might find it interesting to think about expressions like $\Sigma n! x^n$.

We will now see that this ring can be very useful for doing counting problems, or organizing solutions to counting problems, especially when those problems have an inductive or recursive aspects.

Example 1: $1/(1-x) = 1 + x + x^2 + x^3 + x^4 + \ldots$.

as you can see by multiplying both sides by (1-x). The left hand side is 1, the right is

$(1-x) + x(1-x) + x^2(1-x) + \ldots$

$= \quad 1-x + x - x^2 + x^2 - x^3 + - \ldots$

$= \quad 1$

Now we can plug 2x in for x. This gives the formula

$1/(1-2x) = 1 + 2x + 4x^2 + 8x^3 + 16x^4 + \ldots = \Sigma 2^n x^n$

Another "plug in" trick gives the invertibility of the power series starting with a non-zero constant, and maybe makes the reason for its truth more transparent. Without loss of generality assume the constant is 1. Then f = 1 +xg. We can then write

$1/f = 1 - xg + x^2 g^2 - x^3 g^3 + x^4 g^4 - + \ldots$

The terms are formal power series, but we can add up an infinite number of them since any particular coefficient only gets changed finitely many times. (After all the nth term is a multiple of x^n and therefore does not influence any of the coefficients of lower degree.)

Another useful trick that can help in manipulating these series is that if

$F(x) = \Sigma a_n x^n$, then $F(x)/(1-x)$ is $\Sigma b_n x^n$, where $\Sigma a_k = b_n$. This should be clear just by multiplying out.

For example if we want to add up the first n powers of 3 we might note that these sums are the coefficients of

$$1/(1-3x)(1-x) = 2/3\ [3/(1-3x)-1/(1-x)]$$

Examining the coefficients of the right hand side we get that the sum = $2/3(3^{n+1}-1)$, as you might have already known.

Example 2: (Fibonacci revisited): Let $F(x) = \Sigma F_n x^n$.

Notice that $xF(x) = \Sigma F_n x^{n+1} = \Sigma F_{n-1} x^n$. This means that $F(x) + xF(x) = (F(x) - 1)/x$. Rearranging, we get the equation $(x^2+x-1)F(x) = -1$, so that $F(x) = 1/(1-x-x^2)$. The fact that the golden mean $((1+\sqrt{5})/2)$ is a root of the denominator can't be an accident so we should study it.

To get a better formula for the Fibonacci numbers out of this expression, we make use of an idea from the section we just completed. If P and Q are polynomials without common factor, then we write $1/PQ = B/P + A/Q$ where A and B are polynomials so that $AP+BQ = 1$.

Now (x^2+x-1) is irreducible in **Q**[x] so we cannot immediately exploit this trick, but if we go further to **R**[x] it does break up into a product of monic polynomials:

$$(x^2+x-1) = (x - (1+\sqrt{5})/2))(x-(1-\sqrt{5})/2)))$$

(There is no mystery here. A monic quadratic that has 2 roots can always be written as (x-r)(x-s) where r and s are the roots. This point will be discussed further in the coming section.)

It is easy enough to just guess the A's and B's in this case. It gives the formula

$$F_n = ½\ [(1+\sqrt{5})/2)^n + (1-\sqrt{5})/2)^n]$$

This formula is easy enough to verify by induction: formal power series provide a way of guessing it.

Remark: A remarkable consequence of this formula is that $(1+\sqrt{5})/2)^n$ is remarkably close to an integer, namely twice a Fibonacci number, (over and undershooting it). The number of digits that are 0 or 9 after the decimal place is at least n/5 since the other term is $(1-\sqrt{5})/2)^n$ and $(1-\sqrt{5})/2)^5$ = -.090…, so every 5 additional multiplications gives another power of 10.

For more prosaic numbers like 1.5 we have very little understanding about the distribution of fractional parts.

Another consequence of this formula is the ratio of successive Fibonacci numbers give a very good approximation to the irrational number $(1+\sqrt{5})/2)$. We will discuss this aspect more in the next chapter.

Example 3. Here is another counting problem that we can easily turn into a recursion problem and then using formal power series, solve. We shall count rooted binary trees.

INSERT PICTURE HERE.

When we remove the root of a rooted tree with k vertices, we get 2 rooted trees each of which has some number of vertices, say a and b, where a+b = v-1.

If we call T_n = # trees with n vertices. T_0=0 and T_1=1. $T_n = \Sigma T_a T_b$, where(a+b) = n-1. Thus $T(x) = \Sigma T_n x^n$ satisfies the equation, $T(x) = 1 + xT(x)^2$. And therefore[49] T(x) = $(-1 +\sqrt{(1+4x)})/2x$.

Now the question becomes how to tease information out of the explicit formula $T(x) = (-1 +\sqrt{(1+4x)})/2x$. Calculus is a suitable tool for this problem – formal methods can be used to relate the T_ns to exponentials and binomial coefficients.

[49] Using the quadratic formula

This ironic situation is not that uncommon. It is often possible to find an exact expression that doesn't – at least by itself – help us answer crude questions, like around how many digits can I expect T_{100} to have.

Example 4. We will prove an identity of Euler (by his method). He showed that the number of partitions of a number into odd parts is the same as the number of partitions into unequal parts. A partition of n is the description of n as a sum of integers. We will care about decompositions where the pieces are odd or decompositions where no two are the same.

For instance: 2 = 1+1 (a partition into odd parts) and 2 =2 (partition into unequal parts!) 3 = 1+1+1 or = 3 (two partitions into odd parts) 3 = 2+1 and = 3 (two partitions into unequal parts).

We shall write expressions for $\Sigma a_n x^n$ where we set a_n = # of the relevant kinds of partitions.

The function that corresponds to "all pieces are different" = $(1+x)(1+x^2)(1+x^3)(1+x^4)\ldots$ After all, think about multiplying this all out. Each x^i comes up once – and whenever we take a product of these, they contribute an x to their sum to the total. A power of x occurs exactly according to the number of decompositions into unequal pieces.

Now the function that corresponds to decompositions with odd pieces (but allowing the odd numbers to be counted twice or more) is $(1-x)^{-1}(1-x^3)^{-1}(1-x^5)^{-1}(1-x^7)^{-1}$. (Each of these factors can be thought of as $1 + x^r + x^{r+r} + x^{r+r+r} + \ldots$ for an odd number r, and when we multiply them together, we again get a contribution to x^k exactly for each partition into odd pieces.)

Thus, to prove Euler's identity, we should prove that
$(1+x)(1+x^2)(1+x^3)(1+x^4)\ldots = (1-x)^{-1}(1-x^3)^{-1}(1-x^5)^{-1}(1-x^7)^{-1}\ldots$.

Let's multiply the two sides together:
$(1+x)(1+x^2)(1+x^3)(1+x^4)\ldots (1-x)(1-x^3)(1-x^5)(1-x^7)\ldots$.
$= (1+x^2)(1+x^4)\ldots (1-x^2)(1-x^6)(1-x^{10})\ldots$. (here the second product now goes over 2 mod 4 exponents)

$= (1+x^4)\ldots (1-x^4)(1-x^{12})(1-x^{20})\ldots$ (here the second product now goes over 4 mod 8 exponents)

$=\ldots. = 1$

Each time we multiply out and simplify we get higher powers of x left, so the first large number of coefficients are all 0 (after the first). QED

7. Polynomials and their roots.

We shall now study a bit more systematically polynomials, and their roots for various rings. We will be most interested in $\mathbf{Z}_n$, **Z**, **Q**, **R**.

Our main goals are to understand how many roots a polynomial of a given degree can have and also to see that there are nontrivial things that can be said about how specific polynomials behave in different $\mathbf{Z}_n$.

There is much to be learnt even from trying to solve for x in the simple equations:

$$x^2 = 0$$
$$x^2 = x$$
$$x^2 = 1$$
$$x^2 = -1$$

and we will start with these.

The first three equations are straightforward enough in **Z**, **Q**, and **R**. For the first, there is only the solution x = 0 as they are integral domains; the second factors as x(x-1) = 0, so it has 2 solutions, x = 0 or x = 1 and the last factors as (x+1)(x-1) = 0 so x = ±1.

However, for $\mathbf{Z}_n$ the answer depends a great deal on n.

Theorem: The equation $x^2 = 0$ has one solution in $\mathbf{Z}_n$ if and only if n is "squarefree" i.e. not divisible by any square (other than 1).

Proof: First, if n is squarefree, then n is a product of distinct primes. Using the Chinese Remainder theorem, we are thus trying to solve $x^2 = 0$ in each $\mathbf{Z}_p$ separately. But in each of these there is only one solution.

Any n can be written as m^2n' where n' is squarefree. Notice that $[mn'] \neq 0$ (if $m \neq 1$), but it satisfies the equation.

Exercise: How many solutions are there? In particular for n a square how many are there?

The second equation $x^2 = x$ has a different behavior:

Theorem: The equation $x^2 = x$ has two solutions in $\mathbf{Z}_n$ if and only if n is a prime power. If n is divisible by exactly k distinct primes, then it has 2^k solutions.

Proof: The analysis is similar, using the Chinese remainder theorem. The idea is to show that if n is a prime power then there are exactly 2 solutions namely 0 and 1. Then for a product of these, you can pick at each factor separately whether you want an element x to be 0 or 1 at that factor. Any such x with satisfy $x^2 = x$ (since it is true at each factor) and there are 2^k possible choices.

So let us analyze the equation x(x-1) in $\mathbf{Z}/p^k\mathbf{Z}$. Note that in $\mathbf{Z}$, t and t+1 are always relatively prime, so only one of x or x-1 can be divisible by p. The element not divisible by p is a unit, so we can cancel it. The element divisible by p is therefore divisible by p^k, i.e. x = 0 or x = 1. QED

The upshot of this discussion is that for $\mathbf{Z}_n$ equations can have many more solutions than you might expect – because we can combine: mix and match solutions at the various primes at will. For the prime powers, things seem a bit less wild, but there still can be many solutions, like in the $x^2 = 0$ example. In any case, we will concentrate our attention henceforth on the situation of integral domains, or better fields.

Problem: How many solutions are there to $x^2 = 1$ in $\mathbf{Z}_n$.

Hint: The hard case is prime powers. The case of p = 2 works out differently than odd p. After some trial and error, you should be able to verify your final conclusion by induction on the exponent[50]. Even without dealing with the prime 2, you will have a formula for odd n.

Theorem: The number of roots of a degree d polynomial in any field is at most d.

So, in a field, linear equations can have at most one root, (and they always do), quadratics can have at most two, cubics at most three, and so on.

Notice that the conclusion applies to **Z** as well – because any root in **Z** is also a root in **Q**, so there can't be more than d integer roots! Indeed, any integral domain "embeds" in a field, so the same argument can be made to work in general. It is however, not hard to modify the argument we give for the theorem to prove it in this extra generality.

We will prove this using a lemma.

Lemma: Let P(x) be a polynomial in a field F. Suppose f is an element of the field, then P(x) = (x-f)Q(x) + P(f) for some polynomial Q(x).

Note that by P(f) we mean the result of evaluating the polynomial P at x = f. (There is a homomorphism for each f, F[x] → F, that is the identity on coefficients and sends x to f.).

This is the result of the division algorithm so P(x) = (x-f)Q(x) + R. The remainder had degree 0, so is a constant. We can evaluate this constant by setting x = f.

Corollary: If P has d roots and is monic, then P(x) = $(x-r_1)\ldots (x-r_d)$ where the r's are the roots.

Proof: If P(f) = 0 we get P(x) = (x-f)Q(x). Now we can induct on the degree.

Proof of theorem: If f is any element of the field that is not one of the r's, the right hand side is visibly nonzero. So there can be no more than d roots. QED

---

[50] When you have finally succeeded, and I am sure you will, you can look back in satisfaction at your proof of Hensel's (1861-1941) lemma – although the official statement of that lemma is more general.

This theorem that we proved, interesting enough in its own right, especially in contrast to what occurs for the cyclic rings, is also the key step in the proof of the following beautiful and useful result about Galois fields.

Theorem: If F is a finite field, then the non-zero elements form a cyclic group under multiplication.

For instance $\mathbf{Z}_{37}$ is a field, since 37 is a prime. There is therefore some g, so that g, g^2,…. $g^{36} = 1$ are all distinct and fill up the 36 nonzero classes in $\mathbf{Z}_{37}$.

The proof follows from the previous theorem and the next, which is a result in pure group theory.

Theorem: If G is a finite commutative group, and for each n there are at most n elements of order dividing n, then, G is cyclic.

The non-zero elements have at most n solutions to any equation of the form, $x^n - 1 = 0$ by the previous theorem, so the non-zero elements are a cyclic group.

We will prove this by induction. The result is clear for n =1. For each d<n, and d|n, we can assume by induction that {g | dg = 0} is cyclic. We want to see that there is some element of order n. The G will be the cyclic group generated by that element.

Every element has order that divides n. How many are there of order = d. These are exactly the generators of $\mathbf{Z}_d$, (if the cyclic group {g | dg = 0} is as large as it could be) of which there are exactly $\phi(d)$. As a result, the number of elements of order < n in our group G is at most $\Sigma\phi(d)$, where the sum is over proper divisors of n. When we check the following equality, we will be done because there will be of necessity some element of order n.

Proposition: $\Sigma\phi(d) = n$ when the sum on the left is take over **all** divisors of n.

This can be seen by the same reasoning applies to the cyclic group of order n. The left hand side is the sum of the number of elements of order d for each d|n. Since every element has such an order, and there are n elements in that cyclic group we are done.

Finally we can figure out for which primes -1 has a square root in $\mathbf{Z}_p$. Certainly -1 is an element of order 2 in the multiplicative group of nonzero elements of $\mathbf{Z}_p$.

If p = -1 mod 4, then the multiplicative group is cyclic of order 2*odd number. The element of order 2 cannot be twice anything else.

If p = 1 mod 4, then the multiplicative group is (using the Chinese remainder theorem) the product of a cyclic group of order 2^k*odd order group. The element of order 2 lies in the first factor, and of course has 2 square roots, namely the two elements of order 4.

Problem: Show that for p = 1 mod 8, 2 has a square root in $\mathbf{Z}_p$, Hint: Let i denote the square root of -1. Observe that $(1+i)^2 = 2i$, so it suffices to show that i has a square root.

Exercise: For which p is there a nontrivial solution to the equation $x^2+x+1 = 0$?
Exercise: For which p is there a solution to $x^5 = 1$ other than x = 1?

10. Real Numbers

This chapter is devoted to getting a better understanding of our workaday number system the real numbers.

We tend to think of the real numbers as things that look like 1.234567891011121314…. and "keep on going". We learnt so long ago the rules of manipulating them that we take for granted certain features of this system. For instance, it is not obvious at all how to multiply two real numbers.

Consider 1.11111111111111111111111111111*1.11111111111111111111…

If you multiply them out the usual way, then you get something that looks like

1111111111111111111111111111111111111111….
11111111111111111111111111111111111111111….
111111111111111111111111111111111111111….
1111111111111111111111111111111111111111….
….

So we start feeling good 1.23456…but when we get to the column with 10 items we see, oops there's a "carrying" to do so it's 1.23456791… but forever after there will be carrying from later on. How do we know that there isn't a tidal wave of carrying (after all, at the 100^{th} place we will carry forward two decimal places to the right and at 1000^{th} 3 decimal places)?

Many of us have learnt that fractions have "periodic decimal expansions". Why should the product of two decimals with periodic expansions be periodic? The process that produces the answer is anything but periodic.

The truth is that there really is no way to discuss the theory of real numbers without some careful thinking about approximations. Decimal expansions are really an infinite sequence of better and better approximation to the real numbers. We shall explain this here.

We also shall talk about various kinds of real numbers, rational and irrational, algebraic and transcendental. Finally, we will return to equations, and explain how it is

possible to get good approximations to roots of polynomials even when we don't have any formula for what they look like.

1. Rational Numbers.

The main goal of this section is to understand the connection between rational numbers and periodic infinite decimals. Here is the main theorem:

Theorem: Any rational number m/n can be expressed as a decimal, so that the part on the right part of the decimal point has a finite number of places that are sporadic, followed by a part that is periodic – e.g. that repeats every P places. P can be no larger than n-1 and can only be n-1 if n is a prime. The sporadic part is nontrivial if and only if n is divisible by 2 or 5 (when the fraction m/n is put in lowest terms). The decimal terminates if and only if the denominator has no prime factor other than 2 or 5.

The first few 1/n's are:
1/1= 1
½ = .5
1/3 = .3333333333333….
1/4 = .25
1/5 = .2
1/6 = .166666666666…
1/7 = .142861428614….
1/8= .125
1/9=.11111111111111…
1/10=.1
1/11=.090909090909…
1/12= .0666666666666…
1/13=,076923076923…

They satisfy the theorem of course.

.111111 = 1/9 of course then tells us that .444444… = 4/9, say. So the fractions with pure period = 1 are the multiples of 1/9.

Which suggests looking at the multiples of .01010101010101….. These are fractions whose decimals are pure with period = 2. They are the multiples of 1/99.

So the question of which period 1/n (and any a/n) will have boils down to "which P is it that $n|(10^P - 1)$. In other words, what is the order of 10 in the multiplicative group of $\mathbf{Z_n}$.

This already tells us that pure periodicity is impossible when the denominator has a factor of 2 or 5.

On the other hand, if they arise as 2^k5^l , then we can multiply by $10^{max(k,l)}$ to clear the denominator, and then move the final answer to the left by max(k,l) places: this explains the sporadic part.

Since we are now dealing with something that has no factor of 2 or 5, the gcd(n, 10) =1 so, indeed, the relevant order is finite. It must be <n (e.g. by Lagrange's theorem.) Indeed it must be smaller than $\phi(n)$ – which =n-1 if and only if n is prime. (Prove this!)

Problem: What are the possible periods of the sum of two fractions whose decimal periods are P and P' repsepctively?
Harder Problem: What are the possible periods of the product of two fractions whose decimal periods are P and P' respectively?
Exercise: Compute the complete decimal expansion of $(1.11111111111111\ldots..)^2$.

2. How to tell if an Algebraic Number is rational.

An algebraic number in **R** is a number that is the root of some polynomial in **Q**[x] (i.e. a polynomial with rational coefficients). We have seen that $\sqrt{p}$ for p a prime is not rational. What about, say, $\sqrt{2} + \sqrt{3}$?

First we should prove the following:

Theorem: The algebraic numbers form a field.

But, since a completely elementary proof will be much more complicated than the standard, "conceptual" but more advanced proofs, we will not prove it – but instead go through a construction in one case.

Consider $x = \sqrt{2} + \sqrt{3}$. $y = x-\sqrt{2}$ satisfies $y^2 = 3$. Expanding out, we get $x^2 - 2\sqrt{2}x +2 = 3$ or better $x^2 - 2\sqrt{2}x -1= 0$. This looks pretty bad: it's an equation true for x, no doubt, but its coefficients are not rational.

Aha, we note though that we can take this polynomial and multiply it by another polynomial and perhaps benefit. We try $(x^2 - 2\sqrt{2}x -1)(x^2 + 2\sqrt{2}x -1)$ perhaps out of experience from our Fibonacci days.

We get that the product polynomial is $(x^2 -1)^2 - 8x^2$. It clearly has rational coefficients, so we win. Expanded out, it is $x^4 -10x^2 +1$.

This polynomial has four roots, actually, and they are $\pm\sqrt{2}\pm\sqrt{3}$ in all combinations. (Why?) We shall simultaneously see that none of these are rational. Our method reduces the check of whether a polynomial with rational coefficients has a rational root to a finite number of candidates.

Theorem: Suppose $p(x) = a_n x^n + \ldots + a_0$ is a polynomial with integer coefficients. Then if p/q is a rational root of p(x) (written in lowest form) then $p|a_0$ and $q|a_n$.

Exercise: Why is it enough to consider polynomials in **Z**[x] if our interest is in **Q**[x]?

Thus, the only possible rational solutions to is $x^4 -10x^2 +1$ are ±1, and these are not roots, so the polynomial has no positive roots.

The proof of the theorem is direct: Let's plug in x = p/q and multiply through by q^n. This gives:

$$a_n p^n + a_{n-1}p^{n-1}q + a_{n-2}p^{n-2}q^2 \quad \ldots + \quad = 0$$

Noting that all but the last term is divisible by p, $a_0 q^n$ must be. Since q has no factor in common with p, $p|a_0$. The same argument applied to q, shows the other division.

Exercise: If u is algebraic, show that u^2 is algebraic as well.

3. Transcendental Numbers

Having introduced the word algebraic, it must be that not all real numbers are algebraic. That is our first theorem[51]. A transcendental number is a number that is not algebraic.

Theorem: There are transcendental numbers.

Proof: (Cantor) How many elements are there in **Q**[x]? Certainly it is countable, as in each degree d the polynomials are surely in a bijection with $\mathbf{Q}^{d+1}$.

Now, each polynomial has only finitely many roots (as we saw in the last chapter). So the set of algebraic numbers is a countable union of finite sets and is countable, as well. As **R** is uncountable, "most" real numbers are not countable.

π is a transcendental number, but the proof is rather difficult. Even the irrationality of π is beyond the scope of this book. However, we will now give an example of transcendental number (and therefore another proof of the theorem).

The basic idea of the following example/theorem is the one that underlies almost all transcendence proofs: a number, which is irrational, and is too well approximated by rational numbers, must be transcendental. Of course, then you must build clever approximations to your number.

Theorem (Liouville (1809-1882)) If α is irrational and the root of a degree n polynomial in **Q**[x]. then there is a contant C, so that for any rational number $|\alpha - p/q| > C/q^n$.

Proof: The idea is this. Let f(x) be a polynomial with integer coefficients. Let's assume that we are restricting our attention to some small interval around α. Then one can show that for any x in this interval, there is a constant K so that $|f(x) - f(y)| \leq K|x-y|$.

Here's a sketch proof: Think of y as a constant. f(x)-f(y) = (x-y)Q(x,y) for a polynomial Q with integer coefficients, of two variables --- by our work on roots of polynomials. In any bounded region in the x,y plane, Q cannot get too large: take K to be a bound on how large it gets.

Now, here is Liouville's brilliant observation. If α-p/q is small, but f(x) has no rational root, then $f(p/q) \neq 0$ implies that it can't be too small $|f(p/q)| \geq 1/q^n$. Think first

[51] It was the last theorem of the course on which this book is based.

about the case where q = 1. f(integer) = an integer, so if nonzero, it is at least 1. Same about the general case. Clear denominators, repeat the argument, and then divide by q^n.

Now $1/q^n \le |f(p/q)| = |f(p/q) - f(\alpha)| \le K|p/q - \alpha|$
gives the proof. QED

All we have to do is give a number that can be so well approximated that for every C and every n we violate Liouville's inequality. The basic idea is to take a number whose decimal expansion has 1's very separated out. If we truncate at the nth 1, the denominator is around 10^{an} and the error is around 10^{an+1} (It certainly is less than twice that amount). So we essentially want the ratio to get arbitrarily large, and we win.

A concrete example is $\sum 10^{-n!}$.

Exercise: Verify that this number is transcendental using Liouville's theorem.

Liouville's theorem is the first result in the field of Diophantine approximation: of the problems of approximating real numbers by rational numbers. A very interesting device that, in some sense, gives the best approximations to a real number is the continued fraction.

Let x be any positive real number. We let [x] denote the greatest integer less than x. By {x} we denote x-[x], the fractional part of x.

Now we can consider the following process.

Let $0<x<1$. Consider forming 1/x and the fraction 1/[1/x] +{1/x}. The [1/x] is an integer that is at least 1, and {1/x} is again between 0 and 1. So we can start over. For instance if x = ½ then we get x = ½ boring. If x = 2/3, we get x = 1/(1+1/2). In general, we get a "continued fraction".

We can write

where the a's are relevant integer parts. If we stop at some stage, then the error is rather small.

The standard notation for this awkward expression is $[a_0 ; a_1, a_2 ,\ldots]$.

Exercise: Show that a number is rational if and only if the continued fraction terminates. (What is the relation between this and the Euclidean algorithm.)

An interesting special case is then the process repeats.after a while. For example, if $x = ½(\sqrt{5}-1)$. Then x = .618033…. $1/x = ½(\sqrt{5}+1) = 1 + x$. So $x = \{1/x\}$. This means the process cycles immediately!

We get a continued fraction for $½(\sqrt{5}-1) = [1,1,1,….]$

When we truncate, the fractions are ratios of Fibonacci numbers. For $x = \sqrt{2}-1$ we compute that $1/(\sqrt{2}-1) = \sqrt{2}+1$ So $\{1/x\} = 2 + x$.

Thus $\sqrt{2} = [1; 2,2,2,….]$

Again, we have periodicity hitting immediately. There is again a simple recurrence relation for the numerators and denominators:

$a_n = 2a_{n-1} + a_{n-2}$.

Where $a_0 = 1$ and $a_1 = 1$. The formula we get is $a_n = ½ ((\sqrt{2} +1)^n + (-\sqrt{2}+1)^n)$.

However for $x = \sqrt{3}-1$ we get $1/x = (\sqrt{3}+1)/2 = 1 + x/2$. $\{1/x\} = x/2$. Let $y = x/2$. A similar calculation shows that $1/y = 2+x$. We thus get $\sqrt{3}-1 = [1,2,1,2…]$.

In all of these cases, although the decimal does not repeat, the continued fraction gives us a simple way to record excellent rational approximations to the number we are interested in.

Exercise: Show that a continued fraction that repeats satisfies a quadratic equation over **Q**.

Problem: Prove a converse statement..

Exercise: Show that if the a_n (that arise in a general continued fraction) are rapidly growing, then the continued fraction represents a transcendental number.

4. Real numbers, what are they really?

Let us return to a better understanding of the real numbers themselves, and, for instance, the issue of how to multiply them.

What does it mean to be given a real number? You somehow must have some way to get at its digits.

So let us think of real numbers as being produced by sequences of rational numbers (e.g. decimals) that ultimately make a decision about what the nth digit shall be.

This is actually pretty good, except for that nuisance about .999999999… = 1.0.which we will come back to.

It's clear, though, that we will view two sequences as giving the same real number if for each decimal place, they ultimately agree about what it should be.

If two numbers agree through the first n digits (after the decimal point) then they can differ by at most 10^{-n}. If we had 1.0 and .999… that same principle would be true.

So the formal definition most mathematicians use is this.

A real number is the *limit* of a *convergent sequence* of rational numbers. The limit is the decimal that these rational numbers are ultimately settling down on. Convergence, means that if we specify a tolerance, say of .0001, then there is a stage, beond which no two members of our sequence differ from each other by more than .0001.

A decimal expansion (even with the .9999 troubles) is a sequence with this property. So are the continued fractions that we considered before. In fact, in a few cases (essentially of quadratic irrationals), we saw that the fractions converged exponentially quickly. This means that after k steps you get ck decimal places for some c>0.

Example 1: Products. Now let us turn to multiplication.

Indeed it is the case that to compute the n-th decimal place requires the knowledge of what occurs later in the n+kth decimal place of the factors. For instance when you multiply by 10, the new nth decimal place is the old $n+1^{st}$. What we have to do is show that the k involved can always be controlled --- at least in any one multiplication problem.

So let us consider explicitly the problem of multiplying two real numbers X and Y. Suppose we truncate each after the k-th digit after the decimal place. Then we get new real numbers X' and Y' and here is the key:

$|X-X'| < 10^{-k}$

and $|Y-Y'| < 10^{-k}$.

How far apart is our approximate multiplication from our genuine one?

$$|XY - X'Y'| = |XY - X'Y + X'Y - X'Y'|$$
$$\leq |Y||X-X'| + |X'||Y-Y'|$$

(here we use the inequality that the absolute value of a sum is at most the sums of the absolute values, and the fact that the absolute value of a product is the product of the absolute values)

$$< 2\max(|X'|,|Y|)10^{-k}.$$

In other words, if X and Y have no more than 5 digits to the right of the decimal place, we will never have to care about what happens after the $k+6^{th}$ digit to find the kth digit when forming a product. In other words, the technique we have always been using really gives a well defined product of real numbers (what a relief!).

Exercise: Prove that the real numbers, with this product, is a field (assuming that you already know that **Q** is a field.)

The essentially infinite nature of our description of real numbers gives us a great deal of flexibility in how to work with them. The rest of this section will just explain some functions that can be defined on **R** or on subsets of it.

Example 2: Roots How do we find mth roots?

One way is to just look at numbers whose m-th powers are below u, and then go up a little, and then when the m-th power is too big, one goes down a bit. This is a trial and error method. And, it does actually work. We will discuss the details of this type of argument in the next section in more generality.

In practice, there is a better method that converges more rapidly (i.e. approximates the answer much more quickly).

Suppose we have a number u and we have guessed an approximation p to the mth root. A reasonable thing to do then is to try the average $1/2(p + u/p^{m-1})$ If u were p^m, then this average would = p. If p were too large, then u/p^{m-1} would be too small, and vice versa – so taking an average might be getting us closer.

This is an iteration scheme. We start with one approximation, apply the scheme, and hope to get a better one.

If you try this on a computer, you discover that for m=2, that is for square roots, this works very well, for m = 3 it works pretty well. For m>3, it does not work at all.

A little tweaking of the idea makes it work. What works well is taking a weighted average: send p to (m-1)/m*p + (1/m)*u/p^{m-1} – for larger m it is important to weight you current estimate more than the "force of change" – and the (m-1)/m and 1/m weights turn out to be optimal.

The reason for this behavior is not hard to discern. We will have approach the issue obliquely.

Brief on Iteration and dynamics.

Let us think abstractly about some situation about a function f: X → X that we want to iterate, and try to understand its behavior.

Perhaps the simplest case is something like f(x) = ½x on the real line. No matter what number we start with, its size gets smaller and smaller, and as the number of iterates goes to infinity, we converge to 0. Notice that f(0) = 0, that is 0 is a *fixed point*, which should be the case if the iterates converge to anything at all.

The next case to consider might by f(x) = 2x. In that case 0 is still a fixed point, but iteration doesn't get us to it. Indeed it is a *repelling fixed point*. If you are close to the fixed point and you apply the function f you get further away.

For the ½x case, the fixed point is *attracting* (or *attractive*).

For iteration schemes to work, it is important that they attempt to converge to an attractive fixed point.

Let's consider X = [0,1], the interval between 0 and 1. Let f(x) = x^2. This function also has two fixed points 0 and 1. Their behavior is completely different: 0 is attractive and 1 is repellant. Let h be a very small number (1-h) = 1-2h + h^2. The image under f is around twice as far away than the original approximation (when h is small, h^2 is tiny, and we can usually ignore it in rough calculations).

At 0, f is *superattractive*: f(.1) = .01, 10 times closer, f(.01) = .0001, 100 times closer, and so on. For superattractive critical points, an iteration is likely to converge incredibly quickly. The number of decimal places of accuracy will grow exponentially.

For this function f, if you pick your start position to be the number 1, then all the iterates = 1, and you are lucky: you are at a fixed point. But, if you were not so lucky, then when you iterate you always converge (very rapidly) to 0.

Let's consider now f(x) = 1/(1+x). This is defined for $x \neq -1$. To avoid any problems, let's restrict attention to x>0, so that f(x) > 0 as well, and we can therefore iterate without fear of not being defined.

We know that f has 2 fixed points on the real line: $1/2(1\pm\sqrt{5})$, but only one of these is positive. We would imagine that when you iterate f, you should converge to the positive fixed point – at least if we verify that the fixed point is attractive (but it is not super-attractive).

Exercise: Show that if x,y>.1, then $|f(x) - f(y)| \leq .9|x-y|$

Now setting y to be the fixed point (it is larger than .2), applying the exercie, we see that after k iterations, our point will be within $(.9)^k$ times as close as we started (if we started above .1. (This is irrelevant, because if x>0, one application of f will increase its value to be above .1.)

Back to root finding. The first thing we tried was $f(x) = \frac{1}{2}(x + u/x^{m-1})$ Indeed, the mth root of u is a fixed point.

We shall see (at least heuristically) that it is superattractive for m = 2, attractive for m=3 and repellant for m>3. But, the fix of changing the weights in the average, will make it superattractive for all m.

Let's see what happens if we take the fixed point p (= $u^{1/m}$) and perturb it by a small h.

$f(p+h)-f(p) = h/2 + \frac{1}{2}\, u/(p+h)^{m-1} - \frac{1}{2}\, p$

$= h/2 - (m-1)h/2 +$ (things involving h^2 and higher)

The calculation rests on two points: the binomial formula $(p+h)^k = p^k + kp^{k-1}h$ + stuff involving h^2 and higher, and the formula that $1/(a+h) - 1/a = h/a(a+h) = h/a^2$ + terms involving h^2 and higher.

So, we see that for m=2 it is superattractive (for h small). For m=3, it moves nearby approximations around half the way closer. For m>3, it is repellant, and you can't get to the root with this iteration procedure.

However when we change the weightings, we get (m-1)/m*h – (m-1)/m*h + (higher order terms) = h^2 and higher terms. So points do move incomparably closer with this method. We gain superattractivity!

Remark: For 1-dimensional X like the line, dynamics is much simpler because there is only the possibility of attraction or repulsion near a fixed point: there is only one direction, so to speak. On the plane, we can have behavior like $(x,y) \rightarrow (2x, 1/3y)$ that attracts in one direction and repels in another. This allows for a great deal more richness and actually is typical of what occurs in physical dynamical systems. Liouville showed that there is a notion of "volume" that is preserved in such systems --- and that is incompatible with pure attraction or repulsion.

On the other hand, engineers and roboticists routinely try to use friction or other forces to make the dynamics of their machines behave more like an attractive fixed point, so that even if there are errors in the implementation of a plan, the robot or device ends up doing what it is supposed to.

Remark: Even in the one dimensional case, the iterates of a point under a fairly tame looking function can look very different than just converging to a fixed point. Consider $f(x) = 4x(1-x)$ mapping $\mathbf{R} \rightarrow \mathbf{R}$. If $x<0$ or $x>1$ then x moves off to infinity. If x is in the interval [0,1] its whole future is there. However, it can end up at a fixed point (there are 2), it can be (or end up) at periodic point, that is x goes to something else which goes to something else and so on k-times, till it ends up in its starting position. And there are orbits that go all around the interval staying p% of the time in the first half and 100-p% in the other half: you specify p!

By the way replacing the 4 by 4.1 changes the situation enormously. Now there are infinitely many disjoint intervals that escape to infinity (whose endpoints do not).

These phenomena sometimes are described by the word "chaos".

Example 3: We have dealt with no ado[52] with the exponential map $x \to x^n$, when n is an integer. When x>0, we can extend this to any rational number, once we know how to take m-th roots. $x^{n/m}$ = the mth root of x^n.

After having done this, one should check that the usual formulae about exponentials still hold. That is, you want to know that for all rational numbers r and r' we have $x^{r+r'} = x^r x^{r'}$ and $x^{rr'} = (x^r)^{r'}$.

And finally, we take a limit over rational approximations to define x^α for an irrational exponent. And, when that's done, one checks that the basic formulae all still hold.

Of course, one actually has quite a bit of work to actually execute all of these steps. But, if you are going to talk about exponentials and the like, that is the work that you must do.

Finally, once you have a theory of exponential maps, then it makes sense to talk about logarithms. $\log_x y$ is "defined" by the equation $x^{\log_x y} = y$. It makes sense when both x and y are positive.

I wrote "defined" in quotes, because, while it's true that there is only one real number satisfying that equation for any given positive x and y --- one has to prove that this number actually exists. This is done by the method of continuity that we will explain in the next section.

Once logarithms are defined, it is not hard to guess the basic formulae – they follow directly from the formulae for exponentials.

$$\log_x(yz) = \log_x y + \log_x z$$

$$\log_x y^z = z\log_x y$$

$$\log_x y = 1/\log_y x \qquad \text{if } x, y > 0.$$

Example 3. Power series.

This is a very powerful tool that covers a great many functions. In fact, it can be used as alternative to the exponential and the logarithmic functions of the previous paragraphs – and it comes with a built in method of calculation.

[52] What is ado anyway?\\

Remember our formal power series **R**[[x]]. Let us examine some elements of this ring with more care. Our goal will be to see that for some elements, it is possible to plug in small or even large values of x and evaluate the series to get a real number.

If we find a series with the right properties, then we can define all sorts of functions. Here are some examples:

Example 3.1: $1/(1-x) = 1+x+x^2+\ldots$.

We can plug in any number with $|x| < 1$ and the terms get smaller exponentially. Indeed it is easy to see that it only takes a linear in k number of terms to get the first k decimal places. For x =.1, for instance this is exactly the decimal expansion for 1/9 that we are seeing.

When x = 1 we get an undefined left hand side and the right hand side is infinite – a pretty comprehensible conclusion

When x = -1, the left hand side = ½ and the right hand side doesn't coverge at all: it oscillates between the values 0 and 1.

The equality between the two sides of the equation must be taken with a grain of salt. The left hand side has a broader range of applicability – but it equals the right hand side when that makes sense.

Example 3.2: $e^x = \sum(x^n/n!)$.

This is a strange function, at least at its face. One can check that for any x the right hand side converges, so we really do define a function.

The interesting thing occurs when we compute $e^x e^y$.

$$\begin{aligned} e^x e^y &= \sum(x^n/n!)\sum(y^n/n!). \\ &= \sum(\sum(x^n/n!)(y^m/m!)). \text{ (sum over k, and over m+n=k)} \\ &= \sum(x+y)^k/k! \text{ (by the binomial theorem)} \\ &= e^{x+y} \end{aligned}$$

So it looks like $e^x e^y$ = really is some number e taken to the xth power! (Moreover, we can find what e should be by setting x=1.) It satisfies the main rule for exponentials.

This guess is correct – and can be used as the basis of a general theory of exponentials – even for bases other than e. However, the details for such a development are also somewhat complicated.

Example 3.3: $(1+x)^{1/2}$. We already have perfectly excellent methods for taking square roots, but it seems worth pointing out that one can also use the method of formal power series for this purpose as well. (It can be applied to k-th roots also, with some more work).

Consider the problem in **Q**[[x]] of solving $u^2 = 1+x$. One can do so, term by term, and you get

$1 + x/2 +(1/2)(-1/2)/2!x^2 + (1/2)(-1/2)(-3/2)/3!x^3 + (1/2)(-1/2)(-3/2)(-5/2)/4!x^4 +\ldots$

(Of course, you would only get the coefficients – not the expression for the coefficients. Induction would prove the insightful guess I made). This converges if $|x| < 1$. Of course, this expression can be combined with our analysis of binary trees to give a formula for the number of those trees (something our other methods of computing roots won't do).

Perhaps these examples – and their incomplete nature – gives you some motivation to learn calculus. It moves some of these steps from the realm of craft to science and also from heuristic to rigor.

5. Odd degree polynomials have real roots.

The goal of this section is to explain why odd degree polynomials always have roots. The reason is ultimately geometric, and rests on several interesting ideas.

11. The first of these ideas is no doubt familiar. Polynomials can be thought of as functions, and functions (and other relations) can be graphed.

The first idea is then Cartesian coordinates – which we can think of as a method of going back and forth between algebra and geometry.

The second idea is that if we have a path that goes from below the x-axis, e.g. from the third quadrant to the first quadrant, which is above the x axis, then it must cross the x-axis.

This idea is quite subtle, and depends on having a proper definition of the word "path". We will make it rigorous in a moment via an important result "the intermediate value theorem".

The final ingredient is that we can understand the behavior as $x \to \pm\infty$ very easily for polynomials, and it only depends on its highest degree part. As a result we can see that indeed the graph of an odd degree polynomial does go from the third to first quadrants.

Definition: A function f: $\mathbf{R} \to \mathbf{R}$ is *continuous*, if for each point x, the value of f(x) is well approximated by the value of f at points nearby. In other words for each k, to determine the kth digit of f(x), there is an l so that any x' whose first l digits agree with those of x will have the same k-the digit.

As usual, we have ignored the issue of infinite 9's. One can interpret "having the same kth digit" as meaning having values that differ by at most 10^{-k}.

Theorem (**Intermediate Value Theorem**). If f is a continuous function, a < b, and f(a) < t < f(b), then there is a c, so that f(c) = t. In other words, continuous functions assume all the intermediate values, between any values they assume.

Proof: Replacing f by the function f-t we can assume f(a) < 0 and f(b) > 0 and our goal is to find a c with f(c) = 0.

Suppose there is no such c. $Y = \{x \mid f(x) < 0 \text{ and } x < b\}$ has a least upper bound. That is, there is a real number M that is larger than any (other) element of this set – and there is no smaller number with that property.

This is easy enough to see. c is an upper bound. Now look to see if there is any number with a smaller first digit. If there is use that number and start over. If not, work on the second digit. Keep on going. The kth digit of the the number M is determined in the kth step of this procedure.

We claim that f(M) < 0. It cannot be = 0, because we are assuming there is no such number. If it were positive, then the very first digit of f could not be determined by any of the digits, as there are elements of N whose first k digits agree with that of M, for any k. (If not, we could find an M'<M that is an upper bound.)

But, now we are ready for our contradiction. Think about numbers >M that agree through the first k digits. These all have positive f (else they would lie in N, and be above M). But then they differ by more that |f(M)| despite being as close as we want to M. This gives us a contradiction.

The truth, of course, is that for M defined as the least upper bound of N, f(M) = 0. M is our c! QED

Theorem: Suppose $p(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0$ is a polynomial, and the degree n is odd, then there is an M>0, so that if $x < -M$, $f(x) < 0$ and if $x > M$, $f(x) > 0$.

With this theorem and the intermediate value theorem, the fact that odd degree polynomials always have real roots is clear.

Proof: Let $M = \text{Max}(1, \sum |a_i|)$. If $x > M$, then

$$|x^n + a_{n-1}x^{n-1} + \ldots + a_0| \geq |x^n| - |a_{n-1}x^{n-1} + \ldots + a_0|$$

$$> M^n - M^{n-1}\sum |a_i|$$

$$> 0$$

by the definition of M. The argument for $x < -M$ is similar, completing our proof.

Exercise: Use the intermediate value theorem to show the existence of mth roots of positive real numbers for all m, and of all real numbers for odd m.

Appendix: An arithmetic proof of the impossibility of angle trisection.

In this appendix we will combine the Cartesian idea with some field theory to show that the regular 9-gon cannot be constructed by straightedge and compass. Since an equilateral triangle can be constructed, we will have also shown the impossibility of angle trisection.

Unfortunately, we have not developed the necessary technology to give a complete proof of this. The main missing ingredient is the theory of dimensions of

vector spaces, or the fallout of this in field theory, the notion of the index of a field extension.

To get around this, we will argue using finite fields. For them, the "index" of the field extension boils down to counting the number of elements. It is a shame to not develop the important idea of dimension, but there's always the possibility of another volume.

The overall strategy of the proof I am giving is essentially the classical one, so when you learn the rudiments of linear algebra,it won't be hard for you to modify the argument appropriately, and the key trick of reducing problems from the real world to the "abstract one" of finite fields is one of the best tricks of 20^{th} century number theory: it is a pleasure to show it to you.

Let us start with the rules of the game.

One starts with a line, with two points marked on it. We call the distance between these two points 1, and imagine this line to be the x-axis. The Greeks gave us two tools, the straightedge and the compass. The straightedge can be used to draw lines between any pair of "constructible points". The compass enables you to make circles centered at any constructible point, of any radius that is a constructible distance (a distance between constructible points). When you've constructed lines or circles, you get to intersect them with other lines and circles that you already made, and these intersection points are also constructible.

Theorem: The constructible points are the pairs (x,y) in the plane, where x and y are both constructible numbers. The constructible numbers are those that can be obtained from the integers by a sequence of $\sqrt{}$ operations.

So, for instance $\sqrt{(1+\sqrt{(19)} + 3/5\sqrt{7})}$ is constructible. Things like $\sqrt[3]{2}$ are not --- although we don't quite have the technology for proving that. The idea is that if the degree of the minimal polynomial that you satisfy is not a power of 2, then there's no way to get to that number by a bunch of square roots.

The degree of the equation that the length of a regular 9-gon satisfies is 6, so it is not constructible. But we will have to see this indirectly.

First let's prove the theorem.

Lemma: If (x,y) is a constructible point, so is (x,0) and (y,0) and conversely.
Proof: To get (x,0) and (0,y) one just constructs the perpindicular to the x and y axes respectively through (x,y). The picture is this:

INSERT PICTURE HERE

To get (0,y), you draw the circle from (0,0) through (0,y) and consider its intersection with the x-axis. The other direction is similar. QED

Lemma: The intersections between two constructible lines and circles is expressible in terms of square roots of the constructible numbers used in constructing these.

The square roots come from the equation for a circle of radius r around the point (a,b): $(x-a)^2 + (y-b)^2 = r^2$. The lines are $y = mx + c$ where m is the slope, and is in the field generated by the constructible numbers and c is an x intercept, a patently constructible notion on its own.

It takes a bit of work to show that every number of this form can be built by straightedge and compass, but it's not too hard, and not relevant to our impossibility proof.

Now let us think about the regular 9-gon. Constructing it, is the same as constructing the $2\pi/9$ angle whose bottom leg is the x-axis. The point that this line intersects the unit circle in is a 9^{th} root of unity (i.e $u^9 = 1$) if we think of the plane as the complex plane.

(There's nothing special about 9 here. It would be $2\pi/n$ for the regular n-gon. The proof is either an exercise in trigonometry, or in the geometric interpretation of multiplication of complex numbers: in polar coordinates, the product of 2 complex numbers adds their angles and multiplies their lengths.)

This complex number satisfies the equation $(x^9-1)/(x^3-1) = 0$. Doing the division gives the 6^{th} degree equation $x^6 + x^3 + 1 = 0$.

Let us imagine a putative construction of the solution u by straightedge and compass. It will start from 1, do some additions and divisions and so on, then a square root or so and so on till we are done.

If we were to try to mimic this construction in $\mathbf{Z}_p$ we could be stymied by some of the divisions: the denominator might be divisible by p. When we take a square root, the necessary element might not be in the field we are in (at this point in the construction). We might have to go to the field $F[v]/(v^2-u)F[v]$ and v will be that square root.

Why is that a field? Because v^2-u is a prime polynomial (had it not been, it would have a linear factor in F, and hence a root there, and we woldn't have done this extension step). And then we continue our mimicking of the geometric construction.

Corollary: If an algebraic number satisfying a polynomial p(x) in $\mathbf{Q}[x]$ is constructible, then for all large enough primes, p(x) has a root in some finite field that is a sequence of quadratic extensions of $\mathbf{Z}_p$.

Notice that $F[v]/(v^2-u)F[v]$ has $(\#F)^2$ elements (each element can be describe uniquely as $a+bv$, where $a,b \in F$). So the corollary asserts that there is a root of p(x) in a field with p^{2k} elements for some p.

We are now ready to finish off our proof. If $p(x) = x^6 + x^3 + 1$, any root of this will be of order 9 in the multiplicative group, as after multiplying by $x^3 - 1$ you get $x^9 - 1$. (Moreover, the order cannot be less than 9, since $x^6 + x^3 + 1$ and $x^3 - 1$ have no common factor.) Therefore $9|p^{2k}-1$for some k. If $p \neq \pm 1 \bmod 9$, then the 2^k powers of p are 2,4,7,4,7,…. Mod 9, never 1. (Check this!) So we have to find infinitely many such primes.

We will use a version of Euclid's proof of the existence of infinitely many primes. Suppose there were only finitely many primes that not $\neq \pm 1 \bmod 9$. Let P be their product. The quantity P^6+1 will be 2 mod 9, so cannot be a product of only primes that are ±1 mod 9, so there must be some prime that isn't ±1 mod 9 not on the list. This completes the proof of the theorem.

Remark: It is a deep theorem of Dirichlet that every congruence class a mod n, such that a is relatively prime to n, contains infinitely many primes. So there are infinitely

many primes that are 2 mod 9, for example. Dirichlet's theorem in general, however, is a much harder fact than the result about trisection! Happily, we were able to see directly just enough to get the proof to work.

Exercise: Suppose that $E \subset F$ is an inclusion of finite fields, then show that #F is a power of #E. We call this exponent the index of E in F.

Hint: Use the fact that #nonzero elements in a finite field is p^n-1, and that p^n-1 must divide p^m-1 if E has p^n and F has p^m elements (why?).

Exercise: Show that index is multiplicative: If $E \subset F \subset G$ are inclusions of finite fields, then $\text{ind}(E \subset G) = \text{ind}(E \subset F)\text{ind}(F \subset G)$.

Exercise: Deduce that if $E \subset F$ is an inclusion of finite fields, then if p(x) is a prime polynomial in E[x] satisfied by an element f of F, then $\deg(f) | \text{ind}(E \subset F)$.

This sequence of exercises, together, gives a more conceptual view of our proof. If the degree of an algebraic number is the degree of any prime polynomial it satisfies, then that degree must be a power of 2 if the number is to be constructible.

12. Further Directions
13.
14. This book is only a very beginning: after all, how far can a book that includes the theorems of the cavemen be expected to go?

If it must be one book, let me say “Courant and Robbins” What is Mathematics is great, if at times, austere. But, there not be just one. There are so many different directions one can go next. In all cases, it was hard to make the choice to stop here and not climb the next mountain, see the vista just around the corner. Let me point out some things one can do next. (The mathematics student will want to do all of these and more!)

First of all, we have just touched the surface of set theory. To go further, one should learn about the axiom of choice and some of its implications. These include the well-ordering principle and the method of transfinite induction. More advanced topics interact with logic, and that is another subject entirely.

Moreover, we have only introduced one area of logic: the theory of computable functions. The bible in this field is the book by Soare. It can be read without much specialized mathematical training. However, to seriously study logic, one must read a detailed proof of Godel’s theorem (popular accounts include Nagel and Newman, Goldstein, and Hofstadter) and learn some of its implications – such as the copactness theorem, and model theory.

Traditionally, the most conventional topics to continue on to are more algebra and number theory, calculus, and some geometry.

A good book on algebra is Herstein’s “Topics in Algebra”. I also like the book by Stewart and Tall. We stopped short of a number of important directions. We did not discuss any linear algebra. This subject is of unimaginable significance throughout science and mathematics. It can be viewed as a subject within algebra, as a numerical-algorithmic viewpoint, and as a topic in dynamical systems. All of these viewpoints help, and are worth learning. It’s first notion is that of dimension, and that is what is necessary to complete our proof of the impossibility of angle trisection by straightedge and compass.

All this would serve as an introduction to Galois theory, which is the theory of algebraic extensions of fields, but it does not yet get to its unique feature: the exploitation of symmetries that one field has with respect to the other (like complex conjugation for the complexes that preserves the real numbers). This theory explains why it is possible to generalize the formula that solves quadratic equations to cubic and quartics, but why it is impossible for degree five and higher.

In number theory, we have been particularly haphazard. Elementary number theory has many expositions, and the subject itself has many different strands. Even today, I love to recommend Hardy and Wright to beginners. However, I would suggest to the serious student that they also look at Ireland and Rosen for a more systematic view of the algebraic side. Alas, most of the rest requires calculus and its extensions to the complex plane, complex analysis.

So, it is necessary to learn calculus[53]. For this, there are hundreds of books and popularizations. I have no advice at the most elementary level. I love personally, Hardy's "Pure Mathematics", Rudin's "Principles of Mathematical Analysis", and C.H.Edwards's "Advanced Calculus of Several Variables".

I will stop here with my mention of standard mathematical directlons. I assume that readers will email me their favorites, and I hope to make more of such information available on the book's web page.

A lovely book on social choice written at the undergraduate mathematics level is "Chaotic Elections" by Donald Saari.

Again there are many introductory and popular books on game theory – in the sense that we did not discuss: games where players do not know each others moves and so on. These are the games at the foundations of economics and modern evolutionary biology.

[53] Of course, there are many other reasons to study Calculus as well. But people tell me that there are some reasons not to – so it doesn't hurt to tell you that number theory is another.