2011

The Mathematical Landscape

Antonio Collazo
Claremont McKenna College

Claremont McKenna College

The Mathematical Landscape

A guided tour

Submitted To

Professor Asuman G. Aksoy

and

Dean Gregory Hess

By

Antonio Collazo

For

Senior Thesis

Spring 2011

April 25, 2011

Abstract

The intent of this paper is to present the reader will enough information to spark a curiosity in to the subject. By no means, is the following a complete formulation of any of the topics covered. I want to give the reader a *tour* of the mathematical landscape, hoping maybe to see them again. There is plenty of further details to explore in each section, I have just touched the tip the iceberg. The work is basically in four sections: Numbers, Geometry, Functions, Sets and Logic, which are the basic building block of Math. The first sections are a exposition into the mathematical objects and their algebras. The last section dives into the foundation of math, sets and logic, and develops the "language" of Math. My hope is that after this, the read will have the necessarily (maybe not sufficient) information needed to talk the language of Math.

What is Mathematics?

Take a random sample from our population, and ask them "What is Mathematics?" Their answer, most likely, is that Mathematics *is* about numbers; some will add that Mathematics is the study of numbers *and* their arithmetic. This view of math has been held, incorrectly, by society for almost 2500 years; dating back to the Greeks. However, to fully appreciate the beauty and the power of Math, one must dismiss this view as a definition of math. A clever joke goes: "What is Mathematics? Well, mathematics is what a Mathematician does?" This is clearly does no better in a our search for a definition for math, it just moves the burden from defining Math to defining what a mathematician does; equally as formidable.

So then, what is a useful definition of math? I propose that Mathematics is the study of Structures and their Relationships. But, this definition again moves the burden from defining Math, to defining two things: Structures and Relationships. The rest of the section will discuss definitions for these two terms; from which the section's title can be answered. Also, I hope I can convince the reader that this definition does encompass the true essence of math, in a way that the "old definition" could not.

For instance, if Math is "numbers", then where does geometry settle in? The "old" definition plainly states that geometry is not Mathematics, for it's is not the study of numbers; yet, this is unsatisfying situation. If geometry isn't a Math, then what? However, by way of the proposed definition: Geometry is the study of the structure of Shapes: triangles, squares, polygons, etc, and their relationships: congruent, similarity, length, area, etc. More specifically, Geometry, as defined by the Erlanger Program, is the study of shapes, and the invariant properties under a group of transformations. Either way, there is a clear notion of the structure and of the relationships being studied, thus geometry *is* a math.

The true power of Math is that all one needs to do Math is some thinking, maybe a pencil and paper. Although most mathematicians, will argue no math worth doing is done without a pencil. To which I do not argue, insofar as convergent thinking practice is a necessary component to creative thinking. However, the ability to endure a race through the landscapes of your mind, is a powerful and necessary tool for math. This is the other component of creativity: divergent thinking. As an exercise of such thinking, lets construct an abstract notion of numbers.

Natural Numbers, $\mathbb{N}$

What is the common pattern in this collection of objects: three sheep, three cows, three dogs? Although there are several answers to the question, as in a collection of four-legged animals (mammals) or animals found on farm, all have in common the number three. The recognition of this pattern is the creation of an abstract concept of a number. Three is property that each element in the collection satisfies. However, the collection: three ties and three shirts, also has the common pattern of three; yet, none of the elements in the collection are the same in anyway. This abstract notion of a number devoid of anything else but the pattern mentioned above.

We find some other patterns that hold for this abstract notion of a number. For instance, if we take three cows and then get two more cows, we will have five cows. The same holds true for dogs or sheeps instead of cows, and so suppose that this pattern is a pattern of Numbers. When we take a number then add another number the result is also a number. This is called closure, written $x + y \in \mathbb{N}\ \forall x, y \in \mathbb{N}$.

The property of associativity can also be inferred by similar reasoning. Given three numbers, it does not matter if I add the first two and then add the third, or if I add the last two and then the first. Much more precisely stated, $(a + b) + c = a + (b + c)$. Effectively this property states that there is no ambiguity in writing $a + b + c$, since it does not matter which addition is done first. One also can test that $a + b = a + b$, which the Commutativity law.

Consider a bag of marbles and some containers, the above mentioned properties can easily be verified. Suppose, one marble represents the number one. The addition $3 + 4$ can be visualized by putting three marbles in a container, then putting four marbles in the same container. The resulting number of marbles in the container is the answer to the calculation. The commutativity law is easily verified, putting three then four is the same as putting four then three marbles into the container. Associativity is also easily checked, $(2 + 4) + 7 = 2 + (4 + 7)$ represents putting two then four marbles in one container, then putting seven marbles into that container. Which is the same as putting four then seven marbles into a container, then putting two more marbles in.

There is an equally as appealing representation of multiplication using the bags of marbles and containers. If there are three cups, each with four marbles, then how many marbles are there? Twelve, but how did I calculate that? I could pour all the marbles

in one cup, and then count them one-by-one; however we could *define* a new operation: multiplication to be answer to this question. $3 \times 4 = 12$. Using the same ideas as with addition, we want to consider if this new operation has the properties of closure, associativity, and commutativity. Each can be verified by similar means, for instance commutativity: there are the same amount of marbles if there are two containers each with five marbles, and if there are five containers each with two marbles.

The next logical question is how do addition and multiplication relate? Symbolically, what is $2 \times (4 + 6)$? Pictorially, how many marbles are there if I double the amount of marbles in a container with four marbles and a container with six marbles? Notice, this question is not incalculable, even without a formal law of how multiplication and addition interact. $4{+}6 = 10$ and then $2{*}10 = 20$. Yet, there is value in recognizing the pattern of the Distributive Law $a{\times}(b{+}c) = a{\times}b{+}a{\times}c$. Note that $2{\times}4{+}6 = 8{+}6 = 14 \neq 2{\times}(4{+}6) = 20$, the parenthesis are important in defining the order of operations.

Collecting terms, we have inferred that Numbers obey the above mentioned properties, at least in every case we've tried. There is no guarantee, however intuitive it seems, that under the evidence presented the numbers *must* obey these properties. Here is the first rather disappointing/unsatisfying step, although we will find this dissatisfaction can never truly be overcome. We *assume* that Numbers must obey these properties. Any system must have some terms and relationships that must be assumed rather deduced; if not, the resulting is circular logic. i.e.: All men are male. All males are men. (Although one could argue the need a certain maturity factored needed to socially be a "man"). A more sufficient treatment of why this is so will be handled later.

The main point is that any system must start with some rules that must be *assumed* to be true; and consequently, the system developed from these assumed rules is dependent on the validity of these rules to be valid. I.e.: if one of the assumed axioms is invalid, then so is the developed system *and* if all of the assumed axioms are valid then so is the developed system. In this light, the dissatisfaction in just plainly assuming these properties subsides, insofar as the validity of theses assumed rules is plainly obvious. In the case of the above mentioned properties, the assumed axioms are sufficiently self-evident that one can rest assured that any system developed from the properties will have some valid value.

To be more specific the "numbers" just described are the Natural Numbers, denoted $\mathbb{N}$. They are obey the following properties:

The Addition Properties, $\forall a, b, c \in \mathbb{N}$

- Closed: $a + b \in \mathbb{N}$
- Associative: $(a + b) + c = a + (b + c)$
- Commutativity: $a + b = b + a$

The Multiplicative Properties, $\forall a, b, c \in \mathbb{N}$

- Closed: $a \times b \in \mathbb{N}$
- Associative: $(a + b) + c = a + (b + c)$
- Commutativity: $a + b = b + a$

Distributive Property: $\forall a, b, c \in \mathbb{N}$

- $a \times (b + c) = a \times b + a \times c$

These are *not* sufficient conditions to characterize $\mathbb{N}$, but are necessary.

Zero

Unlike the origin of the concept of a number, the origin of the idea of Zero can be traced back to the Arabs. In fact, the Greek schools of Math in their geometric approach, did not "see" the concept of zero as useful. They felt that geometry was the language of nature, and that geometric properties such as length and area could therefore not obtain such flawed values such as zero or negative numbers They would argue: what is the use in considering a triangle that spans an area of zero? However, the Arabs had a very different approach to Mathematics. They developed what we now call Algebra, from the Arabic word *al-jabr*. The usefulness of zero has profound effects in the language of Algebra; particularly zero is the *additive identity*; meaning take any number and add zero, the result is the same number. Immediately we ask does multiplication have an identity? That is, does there exist a number, such that when multiplied by another number results in that other number? Well, one sounds like a candidate, and in fact it is the Multiplicative identity.

There is yet another subtle use of zero, again first employed by the Arabs; its use as a place holder. So far, when I have written numbers, I have typed them out (except in equations); I did so on purpose. The art of writing numbers, as we do today, is a direct affect of the Arabs contributions to math, and given the name Arabic Numerals.

So $\{10, 45, 1023, 20000000004\}$ are numbers written in Arabic Numeral form. We could write the same numbers in Roman Numeral form like so $\{X, VL, MXXIII, (IDK)\}$. However, Roman numerals are rather difficult to write and perform operations on them: $X + VII = XVII$ but $V + IV = IX$. The former seem relatively simple just concatenate the sums, however the latter example shows the pattern is not so simple.

Notice, I did not write the Roman Numeral expression for the last number in the list. Reason being there does not exist a Roman Numeral representation for this number. These representation work by declaring a new number every time we past a certain level. This means that we need to define a new symbol at every level. On the other hand, Arabic numerals eliminate the need for a defining a new symbol the bigger the number gets. All we do is move the place over, possibly using zero as place holder. For example, X is 10 in Arabic Numeral, while C is 100, and M is 1000. At every new level, in Roman numerals we need a new symbol, where as in Arabic numeral we just move over the one.

Now, with the addition of zero to the collection of Natural Number, we arrive at a new notion of numbers, called Whole Numbers. Although most authors, as will I, do not make a strict difference between these two collections (except in this paragraph). The whole numbers satisfy an additional property that the Natural Numbers do not, namely that of additive identity. Now, our notion of numbers encompass the following properties:

Additive Identity (0): $a + 0 = a$

Multiplicative Identity (1): $a \times 1 = a$

Integers, $\mathbb{Z}$

Consider the bag of marbles and some containers, we can ask the question: How many marbles must I add to have a total of five marbles, if I already have two marbles in the container? Three! Generalizing this question we can *define* a new operation: subtraction. As curious mathematicians, we ask whether subtraction obeys the laws of closure, associativity, and commutativity. Unfortunately, this operation is not as "nice" as addition nor multiplication.

Lets investigate this operation in the framework of our trusted marbles and containers. Subtraction can be visualized by taking away marbles from a container. So, the equation $4 - 3$ represents a container initially with four marbles from which I take three away,

then count how many I have left. Immediately, we run into problems if the number of marbles we are trying to take away is greater than the marbles present initially in the container. Consider the equation $3-4$. If we try to carry out this operation, using marbles and containers, the operation is not well-defined, I can not take out four marbles, from a container with three marbles. Notice, commutativity is not satisfied either, since $4-3 \in \mathbb{N}$ but $3-4 \notin \mathbb{N}$.

Do we accept subtraction as this "ugly" operation? The certain elegance Math achieves has no room for ugliness. However, there is another *view* of this operation, that will bring it into a more elegant light. The initial question was posed in terms of addition; a more prudent way to look at this operation may be as the inverse of the operation of addition. Consider the equation, $3-4$, the answer is -1; yet -1 is not a natural number: there is no such thing as -2 cows, so how can we find the common pattern between -2 cows and -2 sheep. Yet, we find use in subtraction, so how can we remedy this.

Lets consider a number line, marked off at unit intervals. In this representation, addition by 5 corresponds to moving to the right five units; while subtraction by 5 corresponds to moving to the left five units. Now consider the addition of a negative number, well this corresponds to moving to the left 5 units. This subtle difference is the unification of addition and subtraction: i.e. $a-b=a+(-b)$. In effect, we have now defined subtraction in terms of addition. Notice now, commutativity of subtraction is inherited from the commutativity of addition, i.e. $a+(-b)=(-b)+a$. If we extend the notion of a number, to include not only the Natural Numbers (with zero) and the negative numbers, we can verify on the number line that the properties of closure, associativity, commutativity, and distributivity all are satisfied. This is a much more appealing situation.

In effect, we have defined the additive inverse of all the Natural numbers. The inverse of an element is the number such that when added to that element the result is the identity, 0. Thus if $x+y=0$ then y is called the inverse of x, denoted $-x$. The new notion of numbers arrived at by adding the negative numbers—the additive inverses, to the natural numbers is called the Integers. This set is denoted $\mathbb{Z}$, because Zahlen is the German word for numbers. They obey the following properties:

The Addition Properties, $\forall a, b, c \in \mathbb{Z}$

- Closed: $a+b \in \mathbb{N}$
- Associative: $(a+b)+c=a+(b+c)$

- Identity (0): $a + 0 = a$
- Inverses: $\exists b \in \mathbb{Z}$ such that $a + b = 0$, let b be denoted by $-a$
- Commutativity: $a + b = b + a$

The Multiplicative Properties, $\forall a, b, c \in \mathbb{Z}$

- Closed: $a \times b \in \mathbb{N}$
- Associative: $(a + b) + c = a + (b + c)$
- Identity (1): $a \times 1 = a$
- Commutativity: $a + b = b + a$

Distributive Property: $\forall a, b, c \in \mathbb{N}$

- $a \times (b + c) = a \times b + a \times c$

Divisibility

Along the same lines, we can seek an inverse operation to multiplication; from elementary school, we know this operation is division. We can ask: how many containers would it take to spread out ten marbles evenly? By evenly, I mean that every container has the same amount of marbles. What about eleven marbles? After some trial and error, one finds that in response to the first question either one, two, five, or ten containers. The solution to the latter is one and eleven. Now, lets agree that any number can always be spread out evenly into either one container with all the marbles, or one marble in each container, and discard these from consideration. So, ten marbles can be spread evenly into two or five containers; eleven into none. This notion, although close to the pattern of division, is really a distinct pattern: divisibility. We make a definition to call any number that can not be spread out evenly into containers is a *prime*, like 11. If not, then call it *composite*, like 10.

The pattern of division arises when we ask the question: How many marbles are in each container, when ten marbles are spread out evenly into five containers? However, this operation is just as "ugly" as subtraction was. For instance, the question: how many marbles are in each container, when eleven marbles are spread out evenly into five containers? We could say there are $2\frac{1}{5}$ marbles in each container, but $2\frac{1}{5} \notin \mathbb{Z}$ and how do

we interpret $\frac{1}{5}$ of a marble. We could also say, there are two marbles in every container, *and* one marbles left over. This avoids discussion of the nonsensical $\frac{1}{5}$ of a marble, however the notion of spread out evenly is not satisfied fully, there is one marble remaining.

So what path do we take? Should we somehow extend $\mathbb{Z}$ to included these fractional units, or do we somehow change the notion of evenly, to incorporate this remainder? Both paths lead to very interesting and different notions of numbers. The former leads to the extension of $\mathbb{Z}$, called the Rational Numbers, denoted $\mathbb{Q}$, and the abstract notion of a field. The latter leads to the concepts of rings and primeness, basically this is the road to Number Theory. For now, let's concentrate on the Number Theory side of this issue, returning to the rationals in the next section.

Back to our marbles and containers, we can ask for given a number of marbles and containers, how many marbles are spread evenly in the containers, allowing for a remainder amount? An algorithmic procedure for answering this question is to grab all the marbles, and put one marble in every container, continue until you can not put one marble in *every* container. Through trials we can see that this procedure works for any pair of numbers representing the number of marbles and containers. This is know as the Euclidean Algorithm [4] We notice after more trials that given *any* two numbers, one representing the total number of marbles in consideration, while the other is how many containers, that there exist uniquely a pair of numbers, one which is the number of marbles in each container, the other is the remainder.

Algebraic Properties

The Integers, $\mathbb{Z}$, have acquired a lot of algebraic properties, which have useful abstractions and applications. This is a concrete example of the power of math to generalize and abstract notions. For example, the algebraic concept of a Group. Simply put, a *Group* is an ordered pair of a set and a binary relation on the set, $G = (X, +)$, that satisfies the following conditions:

Group Axioms

- Closure: $x + y \in X$, $\forall x, y \in X$
- Associativity: $(x + y) + z = x + (y + z)$, $\forall x, y, z \in X$
- Identity: $\exists e \in$, $e + x = x + e = x$, $\forall x \in X$

- Inverses: $\forall x \in X$, $\exists y \in X$, such that, $x + y = y + x = e$. In this case denoted, $y = x^{-1}$

Under these axioms, we see that $(\mathbb{Z}, +)$ is a group. Take $e = 0$ and the inverse to be $x^{-1} = -x$. Another *very* important group the symmetric group $(S_n, \circ)$. This describes the set of permutations of n objects, where $\circ$ operation means "followed by" and is called composition. For instance, $(S_3, \circ) = \{(1), (1\ 2), (1\ 3), (2\ 3), (1\ 2\ 3), (1\ 3\ 2)\}$. The way this group works is to think of 3 chairs and 3 people; how many ways are there for all 3 people to sit on all 3 chairs? Another way to think about this is, suppose the 3 people are each sitting in a chair, then how many distinct way of rearranging the seating assignments are there? The answer is 3!, read 3 factorial, which means $3! = 3 \cdot 2\dot{1}$, or in general $n! = n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 3 \cdot 2 \cdot 1$. This only says how many ways there are! We can also ask, what are the specific rearrangements? This is the function of the symmetric group.

Suppose the chairs are labeled $1, 2, 3$ respectively, and the people named A, B and C. The initial seating arrangement is $A \mapsto 1$, $B \mapsto 2$, and $C \mapsto 3$, which we more succinctly denote $\{A, B, C\}$. If the people in chair 2 and chair 3 switch, that is denoted $\{A, C, B\}$. The elements of the symmetric group *act* on this arrangement. So, $(12) \in S_3$ acts on $\{A, B, C\}$ by switching the people in chair 1 and chair 2; so $(12) \cdot \{A, B, C\} = \{B, A, C\}$. Also $(123) \cdot \{A, B, C\} = \{C, A, B\}$. Intuitively the element (123) of S_3 means send the person in seat 1 to seat 2, the person in seat 2 to seat 3, and the person in seat 3 to seat 1.

We can also perform multiple permutations, for instance $(1\ 2) \circ (1\ 2\ 3) \cdot \{A, B, C\} = (1\ 2) \cdot \{C, A, B\} = \{A, C, B\}$. But also: $(2\ 3) \cdot \{A, B, C\} = \{A, C, B\}$. This implies that $(1\ 2) \circ (1\ 2\ 3) = (2\ 3)$; this is the group structure of S_3. The axioms of a group are interpreted in this light: closure means that given any two permutations their composition is also a permutation. There exist an identity element: namely the "do nothing" permutation. There exists inverses means that for any permutation there is a permutation that "un-does" the action.

There is simple way to compose two permutations $(1\ 2\ 3) \circ (1\ 2)$. First, start with 1 and see where it goes. A word of caution composition is read from right-to-left, so $(1\ 2)$ happens first then $(1\ 2\ 3)$. Anyway, $(1\ 2)$ sends $1 \mapsto 2$, then $(1\ 2\ 3)$ sends $2 \mapsto 3$, thus $(1\ 2\ 3) \circ (1\ 2)$ sends $1 \mapsto 3$. So we start the calculation by writing $(1\ 3...)$, which just says 1 is mapped to 3. Then we need to see where 3 is sent by $(123) \circ (1\ 2)$. Well, (12) does not affect 3, or another way to think about it is that it sends $3 \mapsto 3$. Then

where does (1 2 3) send 3: remember this notation wraps around so $3 \mapsto 1$. This implies for our calculation that (131...) but this is really just (13) because of the wrap around notation. Alright, we are almost done with this calculation, we just need to see what happens to 2. Well (12) sends $2 \mapsto 1$, and (123) sends $1 \mapsto 2$. Thus the total effect of $(1\ 2) \circ (1\ 2\ 3)$ sends $2 \mapsto 2$, and thus does nothing to person sitting at chair 2. So, $(1\ 2) \circ (1\ 2\ 3) = (1\ 3)$. Composing permutations is as simple as just following where every element is mapped. We make the abbreviation to leave out of the notation any element that is left unchanged: i.e. $(1\ 3) = (1\ 3) \circ (2)$. Notice, that is this operation is *not* commutative: $(1\ 2) \circ (1\ 2\ 3) = (2\ 3) \neq (1\ 2) = (1\ 2\ 3) \circ (12)$. Here is the full composition table for S_3:

Composition table for S_3

$\circ$	(1)	(1 2)	(1 3)	(2 3)	(1 2 3)	(1 3 2)
(1)	(1)	(1 2)	(1 3)	(2 3)	(1 2 3)	(1 3 2)
(1 2)	(1 2)	(1)	(1 3 2)	(1 2 3)	(2 3)	(1 3)
(1 3)	(1 3)	(1 3 2)	(1)	(1 3 2)	(1 2)	(2 3)
(2 3)	(2 3)	(1 2 3)	(1 3 2)	(1)	(1 3)	(1 2)
(1 2 3)	(1 2 3)	(1 3)	(1 2 3)	(1 2)	(1 3 2)	(1)
(1 3 2)	(1 3 2)	(2 3)	(1 2)	(1 3)	(1)	(1 2 3)

We can see from the table that $(1\ 2)^2 = (1\ 2) \circ (1\ 2) = (1)$, and $(2\ 3)^2 = (1\ 3)^2 = 1$ thus they are each their own inverse! Also, $(1\ 2\ 3)^2 = (1\ 2\ 3) \circ (1\ 2\ 3)^2 = (1\ 2\ 3) \circ (1\ 3\ 2) = (1)$ Intuitively this seems very likely: if the persons in chairs 1 and 2 switch, and then switch again, nothing has changed! Also, (1 2 3) is like a cycling of of the people on the chairs, and (1 3 2) is the cycling in the other direction. The same procedure of composition works for all S_n for any n.

I said above that is a *very* important group: the reason is that *all* finite groups arise as a subgroup of the Symmetric group for some n. This reduces the study of finite group theory to just studying this group. A subgroup, $(H, \times)$ of a group $(G, \times)$ is such that $H \subset G$ in terms of sets, $(H, \times)$ is a group. We can think of this as restricting the set G to H, while preserving the group operation $\times$. This is a result is called Cayley's Theorem.[5]

The concept of a group is a very natural extension derived from this common pattern between these two sets. Notice the operation is different in each case, yet the exact same structure, namely that of a group, is afforded to both of these sets with their respective

operations. Yet, $(\mathbb{N}, +)$ is not a group. One reason is that it does not have an identity element, however this is merely a definition error; since one can always add Zero to the set of Natural Numbers. Furthermore, there does not exist inverse elements. Given a natural number, say 11, there does not exist an inverse element; namely, there does not exist a natural number such that when added to 11, the resulting sum is zero. Clearly, the number we seek is $-11 \notin \mathbb{N}$. This is just to show that the group structure depends both on the given set and the operation defined on it.

Technically, we should define the group structure as the quadruple $(X, +, e, x^{-1} = -x)$ Since the identity and inverse operations are as important to the group structure as was just mentioned about the Set and the Operation. However, in most cases context makes it clear which is the identity and inverse operation, thus in practice the group is normally just represented by the ordered pair $(X, +)$, or even just X when no ambiguity arises.

Now, we can also consider the algebraic properties of a *ring*. This is the generalized notion abstracted from the set of Integers. A ring is an ordered triple $R = (X, +, \times)$ such that the following axioms are satisfied:

Addition:

- Closure: $x + y \in X$, $\forall x, y \in X$
- Associativity: $(x + y) + z = x + (y + z)$, $\forall x, y, z \in X$
- Identity: $\exists e \in$, $e + x = x + e = x$, $\forall x \in X$
- Inverses: $\forall x \in X$, $\exists y \in X$, such that, $x + y = y + x = e$. In this case denoted, $y = x^{-1}$
- Commutativity: $x + y = y + x$, $\forall x, y \in X$

Multiplication:

- Closure: $x + y \in X$, $\forall x, y \in X$
- Associativity: $(x + y) + z = x + (y + z)$, $\forall x, y, z \in X$
- Identity: $\exists e \in$, $e + x = x + e = x$, $\forall x \in X$

Distributive:

- $x \times (y + z) = x \times y + x \times z$, $\forall x, y, z \in X$
- $(x + y) \times z = x \times z + y \times z$, $\forall x, y, z \in X$

Or in other words, $R = (X, +, \times)$ is a ring if and only if $(X, +)$ is an commutative group and $(X, \times)$ is a semi-group, and the multiplication is distributive over the addition. A commutative group, or abelian group, is a group with the added axiom of Commutativity. Note, $(\mathbb{Z}, +)$ is an abelian groups $(S_N, \circ)$ is not. The term semi-group means that $(X, \times)$ is an associative closed binary relation with an identity element; so, the first three axioms of a group. $(\mathbb{N}, +)$ was not a group because it lacked inverses, thus it can be properly called a semi-group. Another word of note, there does exist theories of ring where there does not have to be a Multiplicative identity element; if there is, we say it is a ring with identity.

$(\mathbb{Z}, +, \times)$ is a ring with identity; one can easily verify all the axioms above. Just like in the case of groups, there is a common pattern arising. Consider the set of polynomials with integer coefficients, denoted $\mathbb{Z}[x]$; I claim this is also a ring. Consider two elements of $\mathbb{Z}[x]$, namely $q(x) = 5 + 6x + x^2$ and $p(x) = 2 + x^2 + x^3$. Now, in order to verify the claim that $\mathbb{Z}[x]$ is ring, we must show first that $(\mathbb{Z}[x], +)$ is a group. So, I need to show closure: that is, given any two elements of the group, their sum is also in the group; i.e, the sum of two polynomials is again a polynomial. Clearly, $q(x) + p(x) = x^3 + 2x^2 + 6x + 7$. However, this does not prove that all elements of $\mathbb{Z}[x]$ satisfy the additive closure axiom, it only proves it for these two specific elements. In order to show this generally, I need a way to represent polynomials. For instance, any polynomial, by definition even, is $\sum_{i=0}^{n} a_i x^i$, where a_i is just the listing of the coefficients of the polynomial. Ergo, $q(x) = \sum_{i=0}^{n} a_i x^i$, where $(a_i) = (a_0 = 5, a_1 = 6, a_2 = 1)$ and $p(x) = \sum_{i=0}^{n} b_i x^i$, where $(b_i) = (2, 0, 1, 1)$, so $q(x) + p(x) = \sum_{i=0}^{n} c_i x^i$, where $(c_i) = (7, 6, 2, 1)$. If we allow the adding of extra zero at the end of the sequence (a_i), if necessary, then we can view the above as component-wise addition of these sequences. So, $p(x) + q(x) = (5, 6, 1, 0) + (2, 0, 1, 1) = (5 + 2, 6 + 0, 1 + 1, 0 + 1) = (7, 6, 2, 1)$. This pattern prevails for all polynomials, and so all the algebraic properties of $(\mathbb{Z}, +)$ are inherited component-wise to the group $(\mathbb{Z}[x], +)$, where $e = (0, 0, 0, 0, ...)$, and $x^{-1} = -x = (-x_0, -x_1, -x_2, ...)$.

Furthermore, we need to show that $(\mathbb{Z}[x], \times)$ is a semi-group, in order to finish classifying $(Z[x], +, \times)$ as a ring. Well, we know that multiplying two polynomials results in a

polynomial. Although here is a way to characterize the multiplication of polynomial using the above "representation" of polynomials, it does not result in the nice component-wise representation like addition. We can use the closure law of the multiplication and the distributive law of the Integers, like so.

$$\begin{aligned}
p(x) &= a_0 + a_1x + a_2x^2 + a_3x^3 + ... \\
q(x) &= b_0 + b_1x + b_2x^2 + b_3x^3 + ... \\
p(x)q(x) &= (a_0 + a_1x + a_2x^2 + a_3x^3 + ...)(b_0 + b_1x + b_2x^2 + b_3x^3 + ...) \\
&= (a_0(b_0 + b_1x + b_2x^2 + b_3x^3 + ...) + a_1x(b_0 + b_1x + b_2x^2 + b_3x^3 + ...) + \\
&+ a_2x^2(b_0 + b_1x + b_2x^2 + b_3x^3 + ...) + ...)
\end{aligned}$$

from which a second use of the distributive law, and a summing of like terms (or the component-wise addition), shows that the product of any polynomial is a polynomial. Thus we have the closure axiom. Associativity is more difficult to show, but only in computation; but it basically is inherited by the associativity and distributivity of the multiplication on the Integers. The identity element is the polynomial 1; and thus $(\mathbb{Z}[x], \times)$ is a semi-group. Therefore, $(\mathbb{Z}[x], +, \times)$ is a ring.

If you look back, the only properties used to show that $(\mathbb{Z}[x], +, \times)$ is a ring, are the properties that represent the ring structure of $(\mathbb{Z}, +, \times)$. This is general pattern of rings: R is a ring if and only if R[x] is a ring.[5] Although the above does not merit a proof of this statement, it does outline the general procedure.

There are more algebraic properties of $\mathbb{Z}$ that can be abstract and considered in general, analogous to the connection between the structure of $\mathbb{Z}$ and $\mathbb{Z}[x]$. For instance, the Fundamental Theorem of Arithmetic states that every integer can be written as a unique product of primes. This concept generalizes to the notion of a Unique Factorization Domain, and the dissolving of two distinct notion of "primeness": prime and irreducible.[5]

A Unique Factorization Domain (UFD) is a ring such that the Fundamental Theorem of Arithmetic is satisfied. The theorem involves two statements: 1) The existence of a decomposition of every element in the ring into a product of primes. 2)The uniqueness of this decomposition. The definition given before of a integer being prime is that it could only be divisible by itself and one, $2, 3, 5, 7, 11, ...$ It turns out that this definition has

pitfalls.

$$5 = (\sqrt{5})(\sqrt{5}) = (1+2i)(1-2i)$$

So, 5 is a prime in the ring $\mathbb{Z}$, but not in the rings $\mathbb{Z}[\sqrt{5}]$ or $\mathbb{Z}[i]$. $\mathbb{Z}[\sqrt{5}]$ is the ring of number of the form $a + b\sqrt{5}$, where $a, b \in \mathbb{Z}$. Similarly, $\mathbb{Z}[i]$ are the numbers of the form $a + bi$, where $a, b, \in \mathbb{Z}$. In this light: 5 is said to be irreducible, yet its primeness depends on the ring in question. In a UFD, these two notion coincide.

The Integers also satisfy some less specific axioms, from which more algebraic structures have been abstracted. If there are no zero-divisor, that is if $a \times b = 0$ then either $a = 0$ or $b = 0$; generalizes to the concept of Integral Domain. The fact that for any two integers there exist a greater common divisor, abstracts to the concept of a Principal Ideal Domain.

Recall in the Divisibility section the procedure described for division with remainder on the containers and marbles. Given 10 marbles and 4 cups, the division with remainder yields 2 marbles in each container and 2 left over. This procedure generalizes to the notion of a Euclidean Domain, named after the Euclidean Algorithm which is effectively the division with remainder procedure described before. The precise statement is $\forall x, y \in \mathbb{Z}, \exists! q, r \in \mathbb{Z}$ such that $x = qy + r$, where $0 \leq r < y$. $\exists!$ mean there exists uniquely. [5]

To gather the notions discussed, there is a nice chain of inclusions showing the hierarchy of rings:

$$\text{Rings} \supset \text{Rings with identity} \supset \text{Integral Domains} \supset \text{Principal Ideal Domains} \supset$$
$$\supset \text{Euclidean Domains} \supset \text{Unique Factorization Domains}$$

Each inclusion is proper, in that none of the categories are equal; in the extreme, not all Rings are Unique Factorization Domains. Each of these categories are a generalization of a specific property of the Integers, all leading up to the Fundamental Theorem of Arithmetic.

Modular Arithmetic

What time is it eight hours after ten in the morning? Six o'clock p.m.

The above makes perfect sense as written, but replacing the letters with numbers it becomes, 10:00 a.m. + 8:00 hours = 6:00 p.m. which is quite strange. But the Euclidean

Algorithm comes to the rescue, and in this light abstract to a the notion of modular arithmetic.

In the case of hours, the number 12 has particular importance. So in search for the operation that is "common pattern" derived from the above question, it seems the number 12 has to play a role. Consider the operation called *addition modulo n*. First add the two numbers regularly, so $8 + 10 = 18$. Then write the unique decomposition of the result using the Euclidean Algorithm with 12 as the quotient, so $18 = 1(12) + 6$. The result of the operation is the unique r in the decomposition $nq + r$, so $10 +_{12} 8 = 6$ also written $10 + 8 \equiv 6 \pmod{12}$. Since the Euclidean Algorithm holds uniquely for any two integers, this operation is well-defined. Similarly, one can define a modular multiplication: by doing the ordinary multiplication of the numbers, then reducing modulo n.

Lets consider fixing a number n as the modulus. So then given any number, $x \in \mathbb{Z}$, can be written $x = qn + r$, where $0 \leq r < n$. r can only be a finite amount of numbers, namely $0, 1, 2, 3, ..., n - 1$. Also notice that if we fixed the modulus to 7, then $8 = 1(7)+1$, $15 = 2(7)+1$, $1 = 0(7)+1$, $-6 = -1(7)+1$, all the remainders are the same. This is all we really care about, and so we define a new equality to be such that two integers are "equal" if and only if their remainders modulo n are equal. Properly, this is an equivalence relation; more on this later, suffice it to say that the new equality defined works very similar to the equality on Integers.

Under this equivalence relation, called congruence modulo n, we see that $8 \equiv 15 \equiv 1 \equiv -6 \pmod{7}$, and there are infinitely many more. We make the choice to take a representative member of this infinite set of integers to be the smallest positive number in that set, and identify the entire equivalence class with the representative member. All this amounts to is that to study the Integers mod 7, is to study the set $\mathbb{Z}_7 = 0, 1, 2, 3, 4, 5, 6$ of representatives of the equivalence class under the equivalent relation of congruent. So, $0 \in \mathbb{Z}_7$ is not just $0 \in \mathbb{Z}$ but also represents $\{..., -14, -7, 7, 14, 21, ...\} \subseteq \mathbb{Z}$. In practice, we make no real distinction between the equivalence classes and the representative member; this depends on the controversial Axiom of Choice.

Lets explore the algebraic properties of this set, $\mathbb{Z}_n$, if any are available. The operations defined above for modular addition and modular multiplication turn this into a ring. For a concrete example lets go back to the clocks. In particular, consider the ring $(Z_{12}, \oplus, \otimes)$. Then $10 \oplus 8 = 6$, and thus this is the algebraic structure of our clock and timing systems. In the minute and second domain, time is measured mod 60, while in the weekdays mod

7.

Another familiar system that is modular arithmetic is rotation. If I'm facing north, then turn a 270° to the right, and then another 180° to the right, that is the same as just turning 90° to the right in the first place. This is addition mod 360; $180 + 270 = 450 = 1(360) + 90$.

The ring $\mathbb{Z}_n$ becomes a field only when n is prime. Whether n is prime or not, $(\mathbb{Z}_n, \oplus)$ is always a group. Thus $(\mathbb{Z}_n, \oplus)$ is only a group when n is prime, thus there exists multiplicative inverses. To be concrete, I present the addition and multiplication tables for $(\mathbb{Z}_5, \oplus, \otimes)$ and $(\mathbb{Z}_6, \oplus, \otimes)$.

Addition and Multiplication tables for $\mathbb{Z}_5$

$\oplus$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\otimes$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Addition and Multiplication tables for $\mathbb{Z}_6$

$\oplus$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\otimes$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Notice, the symmetry along the diagonal line: this implies that the operation is commutative. $\mathbb{Z}_5$ is a field, verify that every row in the addition and multiplication table has a 1: this implies the existence of inverses. $\mathbb{Z}_6$ is not a field, since not every row of multiplication table has a 1: for instance the 2 and 3 rows. Also, there are no zero divisors in $\mathbb{Z}_5$: if $a \otimes b = 0$ then either $a = 0$ or $b = 0$. Take the equations in the ring $\mathbb{Z}_6$: $2 \otimes 3 = 0$ or $3 \otimes 4 = 0$. These imply that there *are* zero divisors in the ring, namely $2, 3, 4$. These are exactly the elements that do not have a multiplicative inverse: i.e. there exist no element x of $\mathbb{Z}_6$ such that $2 \otimes x = 1$ or $3 \otimes x = 1$ or $4 \otimes x = 1$.

Rationals

Earlier, when we tried to answer the question: if i have x marbles and y containers how many marbles are in each container when evenly spread. We already explored one path to answering this question. By way of the Euclidean Algorithm were led to the roads of Rings, Modular Arithmetic and Number Theory.

The main objection was that "$\frac{1}{5}$" of a marble was nonsensical. There are some things where to speak of a fraction of would be permissible though: a pizza for instance. Algebraically, fractions represent the addition of multiplicative inverses. The $(\mathbb{Z}, \times)$ is not a group, since inverses are lacking: if $7 \times a = 1$ then $a = \frac{1}{7}$ but $\frac{1}{7} \notin \mathbb{Z}$. The rationals, denoted $\mathbb{Q}$ (for Quotients), is the set of numbers obtained when we add the axiom of the existence of a multiplicative inverse. This set can also described as the order pair (a, b), where $a, b \in \mathbb{Z}$, along with the equivalence relation $(a, b) = (c, d)$ if and only if $ad = bc$. If we look at the pair as $(a, b) = \frac{a}{b}$ then the equivalence relation means $\frac{a}{b} = \frac{c}{d}$ if and only if $ad = bc$, so $\frac{2}{4} = \frac{3}{6} = \frac{1}{2}$ since $2(6) = 3(4)$, $3(2) = 1(6)$, and, $1(4) = 2(2)$. Again, we make the choice to pick a representative member of these equivalence classes, namely the one with the smallest numerator and denominator. More precisely, we chose the member such that the numerator and the denominator are relatively prime, which means they have no common divisors. This ties into the greatest common divisor discussed above in the abstract notion of Principal Ideal Domains.

Addition is defined on the rationals as follows:

$$(a, b) + (c, d) = \frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} = (ad + bc, bd)$$

Multiplication is defined:

$$(a, b) \times (c, d) = \frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd} = (ac, bd)$$

The set of Rationals, $\mathbb{Q}$, along with the addition and multiplication operations defined above is an instance of the algebraic concept of a *field*. Both $(\mathbb{Q}, +)$ and $(\mathbb{Q}, \times)$ are abelian (commutative) groups, and in addition the multiplication is distributive over the addition; this is a field.

The Addition Properties, $\forall a, b, c \in \mathbb{Q}$

- Closed: $a + b \in \mathbb{Q}$

- Associative: $(a+b)+c = a+(b+c)$
- Identity (0): $a+0=a$
- Inverses: $\exists b \in \mathbb{Z}$ such that $a+b=0$, let b be denoted by $-a$
- Commutativity: $a+b=b+a$

The Multiplicative Properties, $\forall a,b,c \in \mathbb{Z}$

- Closed: $a \times b \in \mathbb{N}$
- Associative: $(a+b)+c = a+(b+c)$
- Identity (1): $a \times 1 = a$
- Inverses: $\exists b \in \mathbb{Z}$ such that $a \times b = 1$, let b be denoted by $\frac{1}{a}$
- Commutativity: $a+b=b+a$

Distributive Property: $\forall a,b,c \in \mathbb{N}$

- $a \times (b+c) = a \times b + a \times c$

A simple connection, albeit sightly abusive in notation, is to think of the algebraic structure hierarchy as:

Structure	Operations	Examples
Semi-group	one operation: $+$	$(\mathbb{N}, \times, 1)$, $(\mathbb{N}, +, 0)$
Group	two operations: $+$, -	$(\mathbb{Z}, +, 0)$, $(\mathbb{Q}_{>0}, \times)$, $(S_n, \circ)$
Ring	three operations: $+$, - , $\times$	$(\mathbb{Z}, +, 0, \times, 1)$, $(\mathbb{Z}[x], +, 0, \times, 1)$
Field	four operations: $+$, -, $\times$, $\div$	$(\mathbb{Q}, (+, 0, -), (\times, 1, \div))$

The above is abusive in notation since properly speaking we don't define $-$ as an operation, but define the inversion of the $+$ operation. However, it is useful picture to think about about algebraic structures.

The Rationals, as defined above, form a number system which is endowed with all the elementary algebraic operations. However, this is not the end of the line. There are still problems in this number field; one is the topological notion of completeness, the other is algebraic notion of solutions to polynomial equations. Each leads to a new set of numbers, the Real Numbers $\mathbb{R}$ and the Complex Numbers $\mathbb{C}$, respectively.

Real Numbers, $\mathbb{R}$

The construction of the Real numbers from the Rational numbers is a very interesting, and generalizable construction that shows the interplay between Cauchy sequences and completeness; which in turn abstracts to a general notion of a complete space. This will be more properly examined in the section on Sequences. For now, we will just consider adding to the set of Rationals those numbers which are Irrational.

An irrational number is one that is not rational, that is it can not be written as the fraction (ratio) of two integers. This is merely a definition, though, and there is still no assurance that any irrational numbers exist. An easy counter-example arises from the Pythagorean Theorem. Consider a right isosceles triangle with the legs equal to one, then by the theorem the hypotenuse must be $\sqrt{1^2+1^2}=\sqrt{2}$. So now we have this number, $\sqrt{2}$, and the question becomes is this number rational? The answer is no; the proof of which follows:
Claim: $\sqrt{2}\notin\mathbb{Q}$

Proof:

[11] Suppose, for a contradiction, that $\sqrt{2}\in\mathbb{Q}$.
So $\sqrt{2}=\frac{p}{q}$, for some $p,q\in\mathbb{Z}$.
Also, we assert that $\frac{p}{q}$ is in reduced terms, i.e. $gcd(p,q)=1$.
Then, $\sqrt{2}=\frac{p}{q}\Rightarrow 2=(\frac{p}{q})^2=\frac{p^2}{q^2}\Rightarrow 2q^2=p^2$.
This says that p^2 is divisible by 2, denoted $2|p^2$, i.e. p^2 is even.
If p^2 is even, this implies that p is even.
If p is even, then for some $r\in\mathbb{Z}$, $p=2r$.
So $2q^2=p^2=(2r)^2=4r^2\Rightarrow q^2=2r^2$.
Similarly, we see that q^2 is divisible by 2, so q^2 is even.
Then, q must be even!
Thus, we have reached our contradiction. Both p and q are even, but we assumed that they had no common factors.
Therefore, $\sqrt{2}\notin\mathbb{Q}$, and thus $\sqrt{2}$ is Irrational.

There is one claim in the proof that needs a little more discussion. The claim that if p^2 is even, then so is p. This was a crucial component to the above proof, and as such a

proof must be given. We will do this with the terminology developed in the last section on Integers, particularly modular arithmetic. Consider the modular group $\mathbb{Z}_2 = 0, 1$ along with the modular addition and multiplication defined above. For every $x \in \mathbb{Z}$ either $x \equiv 0 \pmod 2$ or $x \equiv 1 \pmod 2$. Intuitively, this group says whether a number is even or not; if $x \equiv 0 \pmod 2$ then x is even.

In this setting, it is easy to prove the above claim. Notice that $0^2 \equiv 0 \pmod 2$ and $1^2 \equiv 1 \pmod 2$; and the only way for a square to be even is if it was already even. This is sufficient to prove the claim, since the squaring operation "preserves" evenness. This may seem trivial, but this language helps prove other numbers irrational. For instance, $\sqrt{3}$ is irrational by a similar proof, and observing the following equations. Consider the group $\mathbb{Z}_3 = 0, 1, 2$. Intuitively, this group say whether a number is a multiple of three, or not. $0^2 \equiv 0$, $1^2 \equiv 1$, $2^2 = 4 \equiv 1$ all $\pmod 3$. These equations show that if p^2 divides 3 then so must p, because the only way to get a zero on the right side of the above equations is that there is a zero on the left! The rest of the proof follows almost exactly.

There is a funny anecdote in the history of Irrational Numbers. As it goes: one day at sea, one of Pythagoras's students constructed the number $\sqrt{2}$ using the Pythagorean Theorem. Upon reporting his result to Pythagoras, legend has it that the student was immediately thrown off the ship; for it was blasphemous that such a number existed. It went against all Greek tradition, that nature was purely "rational". As is seen through out this work, and through much of mathematical history, this dis-ease is salient, but was eventually transcended. The name "irrational" still bears the sigma associated with these numbers.

The discovery of these numbers leads to a new notion of numbers called the Real numbers, denoted $\mathbb{R}$. Simply, they are the union of the Rational numbers and the Irrational Numbers; but they have much more structure than that representation implies. They form a field, just like the rational numbers do; however there is a notion of Completeness that differentiates the two number systems. Also, it turns out that there are **a lot** of Real numbers, much more than any of the sets previous considered.

Intuitively, the notion of completeness means there are no "holes" in the space. Consider the set of all rational numbers whose square is less than 2. So, $P = \{x \in \mathbb{Q} : x^2 \leq 2\}$. What is numbers are in the set P? $0, 1 \in A$ since $0^2 = 0 < 2$ and $1^1 = 1 < 2$, but $-2, 2 \notin P$, since $(-2)^2 = 2^2 = 4 > 2$. So we see that the numbers in this set are at least in between 0 and 2.

What is the largest member of this set? Well, using some algebra to transform $x^2 < 2$ into $x < \sqrt{2}$, we can conclude that every member of the set is smaller than $\sqrt{2}$. Also, $\sqrt{2}$ is not in the set P, $\sqrt{2} \notin P$, since it is not rational. Yet, we found a bound for the largest member of P since it has to be smaller than $\sqrt{2}$. Great, then the largest member of the set P is the rational closest to $\sqrt{2}$. That sounds trivial enough, but $1.41^2 = 1.9881$, $1.414^2 = 1.999396$, and $1.4142^2 = 1.99996164$, and the situation gets worse. Imagine a sequence generated by adding the next digit of the following decimal number, which is the numerical approximation of $\sqrt{2}$

1.41421356237309504880168872420969807856967187537694807317667973799....

Then the sequence of squares of those numbers will approach 2, but never quite get there! For instance, $1.4142135623730952 = 1.999999999999999861968$, and $1.414213562373095048802 = 1.99999999999999999995$. Try out the rest of the sequence if you want, however a word of caution: don't trust the calculator! At some point in the sequence the calculator answer will equal 2. This is the fault of the calculator, depending on it precision level, not of the above claim that it never gets there. Try it out by hand, although it is lengthy, you will prove your calculator is wrong. In fact the calculator I am using says $1.41421356237309504880168872 = 2$, which is wrong. The calculator thinks that $1.99999999999999999999999999 = 2$, but that is *not* true.

The problem is the incompleteness of the Rational number system compared to the Real number system. Actually, your calculator uses an even worse system, with respects to completeness, than the Rational number system. It uses a restricted version of the Rational number system where the denominator of any fraction can not exceed a certain size, called the precision of the calculator.

Lets consider different system of numbers; call them A, B, C. A is the set of all decimal expansions that terminate, i.e. .5, .26374, .29238347, or generally $.a_1a_2a_3...a_n$. Here a_i represents a digit in the decimal expansion, so $0 \leq a_i < 10$ for all i. B is the set of all decimal expansions that have a certain pattern that repeats in the tail, i.e. $.33333... = .\overline{3}$, $.17463653653653653... = .1746\overline{365}$, or $.a_1a_2a_3...a_n\overline{b_1b_2b_3...b_n}$. The bar denotes the pattern repeats infinitely. Even though the decimal expansions in B are infinite, we can write them down in a finite fashion with the use of the bar. Note that every decimal expansion that terminates repeats zero in the tail, thus every number in A is in B; denoted $A \subseteq B$.

Finally, consider those decimal expansions such that no pattern ever exist in the tail, this is set C. These are the decimal expansions that proceed infinitely past the decimal point seemingly randomly; some reveal their secrets in a different light, but most transcend any and all appreciation for pattern.

The Rational number system can be viewed as a ordered of integers, as we saw before, where $(a, b) = \frac{a}{b}$. Also, we can view the Rational number system in terms of an integer and a decimal expansion. Lets denote the infinite decimal expansion $.a_1a_2a_3...a_{n-1}a_na_{n+1}...$ by a_i, then the Rational numbers can be represented by the ordered pair (a, b_i), where $a \in \mathbb{Z}$ and b_i is a decimal expansion of type B, $b_i \in B$. An Irrational number can be represented by the ordered pair (a, c_i), where $a \in \mathbb{Z}$ and $c_i \in C$. Then, the Real numbers can be represented by the ordered pair (a, c_i), where $a \in \mathbb{Z}$ and c_i are in either B or C, $d_i \in C \cup B$.

To wrap up the calculator discussion we need to consider one more number system. Let D_n be the number system with all decimal expansion containing less than n digits; to be concrete consider $D_1 = \{.0, .1, .2, .3, .4, .5, .6, .7, .8, .9\}$. Imagine the calculator with D_1 as its number system. Lets say we had to do some computation say $\frac{.1}{.3}$. We know the result is $\frac{.1}{.3} = \frac{1}{3} = .\bar{3} = .33333....$; yet, the finite precision calculator would say the result was .3; since the results lies in a "hole" of the number system, the calculator just thinks its .3. This number system is not a field, since multiplicative inverses do not exist for every element. In fact, its not even a ring, although a multiplication is defined its not a closed operation. Intuitively, D_n has way too many "holes" to be a field or ring.

Back to the discussion of the largest member of the set $P = \{x \in \mathbb{Q} : x^2 \leq 2\}$. We saw that the sequence of numbers, that arose from the addition of a extra digit of the numerical approximation to the $\sqrt{2}$, approached but never quite reached the number 2. As a consequence this set has no largest member! Given a rational number whose square is less than 2, there is always another rational number larger than the given number whose square is still less than 2. If the set was changed a bit to $Q = x \in \mathbb{R} : x^2 \leq 2$, where $\mathbb{R}$ denotes the set of Real Numbers. This set now has a largest member, namely $\sqrt{2}$. It is obviously a member of Q, since $\sqrt{2} \in \mathbb{R}$ and $\sqrt{2}^2 = 2 \leq 2$.

What is the difference in the rationals and the real numbers that this question is has such a different answer? The reason this happens is that the rationals are plagued by analogs of the holes in the finite precision calculators, they are incomplete. Since $\sqrt{2}$ is Irrational, it has a infinite decimal expansion in the set C, and thus never repeats a

pattern. In trying to find the largest member of P, we were using rational numbers, those which by definition have a pattern in the tail of their decimal expansion, to write a number which never settles into a pattern in the tail of their decimal expansion. Quite a difficult task. However, when trying to find the largest member of Q, we are allowed to use Real numbers thus we can express the irrational number $\sqrt{2}$. This property of completeness is an essential characteristic of $\mathbb{R}$ that plays an important role in the Calculus developed over the real numbers, this will be explored in a following section. For now, intuitively think of completeness as there being no holes in the space.

Complex Numbers, $\mathbb{C}$

In order to make this step in the progression of number system, we need to look at the polynomial ring over the real numbers. That is, we need to look at $\mathbb{R}[x]$ which is the set of all finite polynomials with coefficients from the field $\mathbb{R}$. We saw earlier, in the section on Integers, that R is a ring if and only if $R[x]$ is a ring. Since a field is also a ring, $\mathbb{R}$ is a ring, so $\mathbb{R}[x]$ is also a ring. Notice the power of that theorem; killed two birds with one stone, actually an infinite number of birds!

Take the polynomial $f(x) = x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$. What are the roots of this equation? The roots of an equation form a set given by $R = \{x \in \mathbb{R} : f(x) = 0\} = \{\sqrt{-2}, \sqrt{2}\}$. Let $p(x)$ be an polynomial. Its pointless, well fruitless, to look at the set $N_C = x \in \mathbb{R} : p(x) = C$ for some given $C = 0 \in \mathbb{R}$, for instance $N_2 = \{x \in \mathbb{R} : p(x) = 2\} = \{x \in \mathbb{R} : x^2 - 2 = 2\}$. The equation $x^2 - 2 = 2$ can be simplified to $x^2 = 0$, thus any set N_C for any polynomial is the set of roots of *some* other polynomial. Thus we only need to consider roots of polynomials.

Let's change the definition of the set of roots of a polynomial given above to a slightly more general definition. For any polynomial, $p(x)$, the set of roots $R = \{x \in \mathbb{D} : p(x) = 0\}$ where $\mathbb{D}$ denotes the Domain of the polynomial. The set of roots of $f(x) = x^2 - 2$ when the $\mathbb{D} = \mathbb{Z}$ is the empty set, i.e. There are no integers whose square is 2. Since $x^2 - 2 = 0 \Rightarrow x^2 = 2$, and $\{-\sqrt{2}, \sqrt{2}\} \notin \mathbb{Z}$, or for that matter in the rationals either. So, the domain of a polynomial is crucial in understanding its structure.

The question now becomes does there exist a field, with all its algebraic operations, such that every polynomial with coefficients in that field admits a set of roots that lies within the given field. This is akin to the closure property of the operations of addition

and multiplication, however this is the closure of the Polynomial Ring over the a field with respects to finding the roots of the polynomial. In the last chapter we saw the notion of completeness, or what can be called the topological closure of a set, with respects to taking limits of sequences (this is will discussed more thoroughly later). The point is we have seen three notions of a closure: the closure of an operation, the topological closure (completeness), and now the algebraic closure. The power of these concepts comes from the fact that given two elements in a closure of a set, then the applying the respective operation to those elements guarantees the result of the operation is a element of the set. For instance, $x + y, xy \in \mathbb{R}, \forall x, y \mathbb{R};\ \sup\{x \in \mathbb{Q} : x^2 < 2\} \subset \mathbb{R} \text{then} \sqrt{2} \in \mathbb{R};\ p(x) \in \mathbb{C}[x]$ then the roots of $p(x)$, $R = a \in \mathbb{R} : p(r) = 0$ must lie in $\mathbb{C}$. $\mathbb{C}$ denotes the field of Complex Numbers, which this section's work is to construct.

It turns out the field $\mathbb{R}$ is *not* algebraically closed. This means there is at least one polynomial in the polynomial ring $\mathbb{R}[x]$ which does not contain any real roots. Consider $q(x) = x^2 + 1$. Its roots are then $R = \{x \in \mathbb{R} : x^2 = -1\}$. We can immediately conclude that $q(x)$ has no real roots, since every real number squared is positive. In fact, $q(i) = i^2 - 1 = -1 - 1 = 0$, but $i \notin \mathbb{R}$. The previous statement should be looked at as a definition for a certain number i, which is not a real number, but satisfies the relation $i^2 = -1$.

The Complex Numbers are then the ordered pair $(a, b) = a + bi$; where $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$. Addition is defined component-wise, in that $(a, b) + (c, d) = (a + b, c + d) = (a + b) + (c + d)i$. Multiplication is defined by expanding the product $(a + bi)(c + di)$ using the distributive law. So,

$$(a + bi)(c + di) = a(c + di) + bi(c + di) = ac + adi + bci + bdi^2$$

Now replace i^2 with -1 by way of the definition of i. The product becomes $(a+bi)(c+di) = (ac - bd) + (ad + bc)i$. Under these two operations, $\mathbb{C}$ is a field. Addition is clearly a group operation inherited from the component-wise representation. Multiplication is a little harder to show that it is a group operation, but from the definition we immediately get the closure and from the associativity (commutativity) of the multiplication in the Polynomial Ring we inherit associativity (commutativity) of this multiplication operation. The multiplicative identity is $(1, 0)$, since $(1, 0) \times (c, d) = (1(c) - 0(b), 1(d) + 0(c)) = (c, d)$

All that is left to show is that there exists an multiplicative inverses for each element.

To show this we need to introduce a new operation called conjugation; it is denoted by a bar over the number, $\overline{(a,b)} = (a,-b)$ or $\overline{(a+bi)} = (a-bi)$. The product of any complex number with its conjugate is

$$(a+bi)(a-bi) = a^2 - abi + abi - b^2 i^2 = a^2 - (-1)b^2 = a^2 + b^2$$

Also, we need the analogous idea of the absolute value on Real numbers. Define a absolute value on $\mathbb{C}$ like so $|(a,b)| = |a+bi| = \sqrt{a^2+b^2}$. This is an extension of the absolute value defined on the real number if we notice that $|a| = \sqrt{a^2}$, so $|-2| = \sqrt{(-2)^2} = \sqrt{4} = 2$. Properly speaking, this is a "norm" on the number system, which intuitively is a distance defined on the space; further details on this will be given later. For now just take the above two operations as defined.

Back to the task of finding inverses. By definition, a inverse of $x = a+bi = (a,b) \in \mathbb{C}$ is an element, $y \in \mathbb{C}$ such that $x \times y = 1$ With the two operations defined above, this becomes trivial once we observe that $z(\overline{z}) = |z|^2$, since dividing by $|z|^2$ we obtain $z\frac{\overline{z}}{|z|^2} = 1$. Therefore, the inverse of z is $\frac{\overline{z}}{|z|^2}$. Since we chose the element z arbitrarily, then all elements of $\mathbb{C}$ have inverses. Well at least if $|z|^2 \neq 0$, since otherwise we would be dividing by zero. But, $|z|^2 = 0$ if and only if $z = 0$, so this is nothing new.

Thus $\mathbb{C}$ is a field. Also we have two other operations in $\mathbb{C}$, conjugation and absolute value. In reality the only "new" operation is conjugation, since there also exists an absolute value on $\mathbb{R}$. Besides its use in the definition of the multiplicative inverse, conjugation can also be used to determine whether a complex number is a real number. If $z = \overline{z}$ then $z \in \mathbb{R}$. Notice $1 \in \mathbb{R}$, and $1 = 1 + 0i \in \mathbb{C}$. The conjugate of 1 is $\overline{1} = 1 - 0i = 1$, also $i \notin \mathbb{R}$ since $\overline{i} = \overline{0+1i} = 0 - 1i = -i \neq i$.

Returning to the initial question of roots of polynomials with coefficients in Complex Numbers, I claim that $\mathbb{C}$ is an algebraically closed field. This is known as the Fundamental Theorem of Algebra. It states: Given a polynomial $p(x) \in \mathbb{C}[x]$ then if R is the set of roots of $p(x)$ and every element of R is in $\mathbb{C}$, or more succinctly $R \subseteq \mathbb{C}$. Actually, formally it states that every polynomial in complex coefficients has at least *one* root in the complex numbers. The second statement may seem weaker than the first, however the first statement is implied by the second statement. Notice that if a polynomial has a root, say z, then by definition $p(z) = 0$. One way to interpret that statement is to decompose the polynomial $p(x)$ into the product of two polynomials $p(x) = q(x)r(x)$ such that $q(z) = 0$

and $r(z) \neq 0$; observe that $p(z)$ still equals 0. How can we be sure that this decomposition exists? This properly in the domain of Unique Factorization Domains; recall, these where Rings where every element could be decomposed uniquely into a product of "primes". But for our discussion, recognizing the decomposition as $p(x) = q(x)r(x)$ is sufficient; the theory of Rings and UFD would help extend these notions to more complicated and abstract fields.

Utilizing the decomposition, we recognize that $r(x)$ is a polynomial and by Fundamental Theorem of Algebra it has a root, call it w. Then decompose $r(x)$ into $r(x) = s(x)t(x)$ where $s(w) = 0$ and $t(w) \notin 0$. Repeating this procedure until the polynomial can not decompose any more, i.e. it a prime polynomial. These turn out to the the linear polynomials such as $f(x) = x - c$, where $c \in \mathbb{C}$. By the end of the procedure, we have decomposed the original polynomial $p(x)$ into a product of "prime" polynomials, so that $p(x) = (x-c_1)(x-c_2)(x-c_3)\cdot ... \cdot (x-c_n)$, where each of the c_i's are in $\mathbb{C}$. So, the second version of the Fundamental Theorem of Algebra implies the first version.[13]

Actually, the Theorem asserts one more conclusion: there are exactly as many roots (distinct or not) as the degree of the polynomial. The degree of a polynomial is the number which is the highest power of x, for example if $p(x) = x^2 + x + 3$ then the degree of $p(x)$ is 2. More generally, any polynomial can be written $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + ... + a_nx^n$, then the degree is n. So to wrap things up, the Fundamental Theorem of Algebra states: Given $p(x) \in \mathbb{C}$ represented by $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + ... + a_nx^n$ then there exist a set of numbers $(c_i)_{i=0}^n$ such that $p(x) = (x - c_1)(x - c_2)(x - c_3)...(x - c_n)$. In that statement, there was no mention of roots of that polynomial, but notice that $p(c_1) = (c_1 - c_1)(x - c_2)(x - c_3)...(x - c_n) = 0(x - c_2)(x - c_3)...(x - c_n) = 0$ so the notion of roots is interplayed with the decomposition of the polynomial into a product instead of a sum. This is very analogous to the Fundamental Theorem of Arithmetic. Recall, it states that $\mathbb{Z}$ is a UFD: every element in $\mathbb{Z}$ can be written uniquely as a product of primes. The Fundamental Theorem of Algebra states that $\mathbb{C}[x]$ is UFD.

Recall, the difference between Rational number and Irrational numbers; which basically amounts to the Rational number possessing some type of finite character: either the decimal expansion terminated or there was a pattern that repeated in the tail. The Irrational number were those decimal expansion that where infinite in character: there was no repeating pattern in the tail. Finally, we put both types of numbers together to obtain the Real numbers. There is one more distinction to be made in the Irrational numbers,

that of an Algebraic number. $\sqrt{2}$ is an Algebraic number, while π is not an Algebraic number[5]; it is called a Transcendental number. What is the difference between these two numbers?

$\sqrt{2}$ is an Algebraic number, which I will start denoting $\mathbb{A}$. Intuitively, an algebraic number must posses some kind of finite character, just like the Rationals had a finite character. However, we know that in their decimal expansions, there is no rhyme, reason, or pattern to these numbers; thats why they are Irrational. But, as I hinted earlier, some of the Irrational number share their secrets but only under a different light, the algebraic numbers. There are also some Irrational numbers that transcend all appreciation for pattern: these are the Transcendental numbers. To find the secrets of the Algebraic numbers, we need to look at the Polynomials again. Properly, an Complex number is called Algebraic if there exist a *finite* polynomial in $\mathbb{Q}[x]$ (polynomials with coefficients in $\mathbb{Q}$), of which one of the roots is the given number. So, $\sqrt{2} \in \mathbb{A}$ since $p(x) = x^2 - 2 \in \mathbb{Q}[x]$, $p(\sqrt{2}) = 0$, and the degree of $p(x)$ is 2, which is finite. Here is the finite character of Algebraic numbers; they must have a finite representation as roots of a finite polynomial. Notice $i \in \mathbb{A}$ since $q(x) = x^2 + 1$, $q(i) = 0$ and the degree of $q(x)$ is 2.

Are there any numbers that are not Algebraic? I mentioned earlier that π is not Algebraic; therefore there does not exist any finite polynomial such that π is a root. Another way of saying this is, given any finite polynomial its value at π *must not* be zero. This is similar to the difference between Irrational and Rational; the former has an infinite character while the latter has a finite character. The proof that a number is not Algebraic is rather difficult. We would need to show that for *any* finite polynomial π is not a root; that an infinite number of cases, so more creativity and ingenuity are needed to prove this claim. A consequence of the transcendence of π is that if $p(\pi) = 0$ then the degree of $p(x)$ must be infinite! This in turn implies that $cos(x)$ and $sin(x)$ are infinite polynomials, since $cos(\pi) = 0$ and $cos(x - \pi) = sin(x)$, by the laws of Trigonometry, so $cos(\pi - \pi) = sin(0) = 0$. Thus π is a root of both of these functions, thus they must be

infinite polynomials!

$$\begin{aligned}\sin(x) &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \frac{x^{13}}{13!} + \dots \\ &= \sum_{i=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} \\ \cos(x) &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \frac{x^{10}}{10!} + \frac{x^{12}}{12!} + \dots \\ &= \sum_{i=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}\end{aligned}$$

The Algebraic numbers, $\mathbb{A}$, turn out to be exactly those numbers which are in the Algebraic Closure of $\mathbb{Q}$. We denote the algebraic closure (actually the topological closure too) by a bar over the given set: context should be clear as to which closure is involved, and context distinguishes between conjugation. So, $\overline{\mathbb{Q}} = \mathbb{A}$. The definition of an Algebraic closure ensures that any finite polynomial with coefficients in $\mathbb{Q}$ must posses a root that lies in the closure, i.e. for any polynomial in $p(x) \in \mathbb{Q}[x]$ and $p(z) = 0$ then that implies that $z \in \overline{\mathbb{Q}}$. This is exactly the definition of Algebraic numbers given above. As a set $\mathbb{A}$ is a subset of $\mathbb{C}$, also $\mathbb{Q}$ is a subset of $\mathbb{A}$; these statements should be interpreted like so: Every Algebraic number is a Complex Number, and every Rational number is an Algebraic number. But, $\mathbb{A}$ in not a subset of $\mathbb{R}$: so not every real number is an Algebraic number: $\pi \in \mathbb{R}$ but $\pi \notin \mathbb{A}$. Finally, the relationship between Irrational number and Algebraic numbers is more subtle. So see this consider the Complex numbers as the collection of both Transcendental numbers and the Algebraic number; just like the Reals where the union of the Irrational and Rational. Now consider the Algebraic numbers that are *not* rational. Then the Irrational numbers are exactly those Real Numbers that are Transcendental, or Algebraic but not rational.

Another interesting fact about the Complex Numbers is that we have lost an the ordering. In $\mathbb{R}$, there exist a total ordering relation, such that:
Total Ordering Relation $(\mathbb{R}, \leq)$, for all $x, y, z \in \mathbb{R}$
Reflexive: $x \leq x$ Transitive: $x \leq y$ and $y \leq z$ implies $x \leq z$ Anti-symmetric: $x \leq y$ and $y \leq x$ implies $x = y$ Linear (Total): for each pair x, y, either $x \leq y$ or $y \leq x$ Compatibility with Operations

- Addition: if $x \leq y$ then $x + c \leq y + c$

- Addition: if $x \leq y$ and $a \leq b$ then $x + a \leq y + b$
- Multiplication: if $x \leq y$ and $0 \leq z$ then $xc \leq yz$
- Multiplication: if $x \leq y$ and $z \leq 0$ then $yc \leq xc$

$\mathbb{R}$ under the normal ordering relation satisfies the above properties. We can define the "dual" order as such $x \geq y$ if and only if $y \leq x$. Also we can define the strict inequalities $x < y$ by ($x \leq y$ *and* $x \neq y$). So in effect, by defining the one ordering relation, we get all six relations $\leq, \geq, <, >, =, \neq$. Notice how equality is defined by the Anti-Symmetric Property; from which we can also define $\neq$. We will study the ordering in the next chapter, here I just want to show that there can be NO total order imposed on $\mathbb{C}$. To see this let's try to order i, in that either $i \leq 0$ or $0 \leq i$. Take the first case $i \leq 0$. Multiply by i on both sides, since $i \leq 0$, so by the second multiplication property $0 \leq i^2$. But $i^2 = -1$ so we have $0 \leq -1$: Absurd! Trying the other case: $0 \leq i$, again multiply by i on both sides and we obtain $0 \leq i^2 = -1$: Absurd! In either case we are led to a contradiction; thus there can be no Total ordering on $\mathbb{C}$ as there is on $\mathbb{R}$. This is curious phenomenon. Until now, at every step in the construction, in every extension we have acquired and inherited the properties of the previous set; here we have a lost the fundamental property of ordering. This loss is balanced out by the gain in the closure of the roots of polynomials, which was the main purpose in constructing the Complex numbers anyway.

A word on notation, the defining relation for i: $i^2 = -1$ implies notationally $i = \sqrt{-1}$. However, this notation leads to contradictions. Recall, the formula $\sqrt{a}\sqrt{b} = \sqrt{ab}$. This implies the following absurd calculation:

$$-1 = i^2 = \sqrt{-1}^2 = \sqrt{-1}\sqrt{-1} = \sqrt{(-1)^2} = \sqrt{1} = 1$$

Letting $\sqrt{-a} = i\sqrt{a}$ then restricting the square root operation to positive numbers, avoids contradictions in denoting the $\sqrt{-1}$ as i.

Vectorspaces

A vectorspace is another algebraic structure, where there is an addition defined and a so-called scalar multiplication. A element of a vectorspace space is called a vector; it looks like $v = (v_1, v_2, v_3, ..., v_n)$, where all the v_i's come from the field F. $(V, F, +, \cdot)$ is a

vectorspace if $(V,+)$ is a commutative (abelian) group. F is called the scalar field, we say that V is vectorspace over F. Context should present the scalar field clearly, so normally it is omitted from the quadruple and just represented $(V,+,\cdot)$. The scalar multiplication is defined as follows: Let $\alpha \in F$ then $\alpha \cdot (v_1, v_2, v_3, ..., v_n) = (\alpha \times v_1, \alpha \times v_2, ..., \alpha \times v_n)$. This operation is technically *not* a binary operation on the algebraic structure, since a binary operation acts on two elements of the algebraic structure; the operation acts on one element from the scalar field while the other element is from the vectorspace. This operation is properly called a group action; it *acts* on the additive group of the vectorspace as defined above. Notice that the scalar multiplication is a component-wise multiplication **in** the Scalar Field. Addition is also defined component-wise (as we have seen plenty before); the addition is also happening **in** the Scalar Field.

A vectorspace has a notion of dimension: given any two elements,x, y of V then their representation as a vector is $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$, then the dimension of the vectorspace is n. There are two cases: V is finite dimensional, or V is infinite-dimensional. The representation of a element of the vectorspace into the vector form is not unique. For instance, consider $(1,0)$ and $\frac{1}{2} \cdot (2,0)$ as two element of a 2-dimensional vectorspace; they are clearly equivalent. Formally the equivalence relation is defined component-wise, in that $v = (v_1, v_2, ..., v_n) = (w_1, w_2, ..., w_n) = w$ if and only if $v_1 = w_1$, $v_2 = w_2$, ... , $v_n = w_n$. So $(1,0) = \frac{1}{2} \cdot (2,0) = (\frac{1}{2} \times 2, \frac{1}{2} \times 0) = (1,0)$. However, any two vectors can be written uniquely if we specify a Basis. Observe, in a 2-dimensional vectorspace, any vector is represented (a,b) where $a, b \in F$. But I could also write, $a \cdot (1,0) + \cdot(0,1) = (a,0) + (0,b) = (a,b)$. Thus we see that given the basis $(1,0), (0,1)$ we can write uniquely any vector in the that vectorspace. The vector (a,b) is called the coordinate vector with respects to the Basis. The vector (x,y) has the coordinate vector $(x,y) = a(1,0) + b(0,1)$ so $(a,b) = (x,y)$. The fact that the coordinate vector is equal to the vector itself is a special property of this basis; which is why its called the Standard Basis. There exist an infinite number of other basis, like $\{(2,0),(0,2)\}$. In this basis the vector $(1,1)$ has coordinate vector $(\frac{1}{2}, \frac{1}{2})$, since $\frac{1}{2}(2,0) + \frac{1}{2}(0,2) = (1,1)$. Basis can look much more complicated: if $\{(1,2),(2,1)\}$ is a basis for a vectorspace, then the vector $(1,1)$ has coordinate vector $(\frac{1}{3}, \frac{1}{3})$ since $\frac{1}{3}(1,2) + \frac{1}{3}(2,1) = (\frac{1}{3}, \frac{2}{3}) + (\frac{2}{3}, \frac{1}{3}) = (1,1)$.

The connection between basis and dimension is that a vectorspace is n-dimensional if and only if there is a basis that has exactly n vectors. Technically the definition goes like this: A basis for a vectorspace is a linearly independent set of vectors such that they

span the vectorspace. Then the dimension is defined as the number of vectors needed for a basis; so, all basis have the same number of vectors. The condition of Linear Independence is an important premise, it guarantees the coordinate vectors are unique. For instance if the the basis is $\{(1,2),(2,4)\}$ then the vector $(3,6)$ has at least two coordinate vectors $(3,0)$, since $3(1,2)+0(2,4)=(3,6)$, and $(1,1)$, since $1(1,2)+1(2,4)=(1,2)+(2,4)=(1+2,2+4)=(3,6)$. So with this defect of the basis vectors we lose the uniqueness of the coordinate vectors; this set of vectors is called Linearly Dependent.

This set of vectors has another flaw: it does not span the space. Intuitively, the span of a set of vectors means that every vector in the vectorspace has a coordinate vector with respect to the given set. Properly, $span\{v,w,x,...,y\}$ (Note, $v,w,...$ are vectors, so each has a vector expansion) is the set of linear combinations of the set of vectors. So, $p \in span\{v,w,x,...,z\}$ if there exist $c_1,c_2,...,c_n \in F$ such that $p=c_1\cdot v+c_2\cdot w+c_3\cdot x+...+c_n\cdot z$. Notice the use of the scalar multiplication! Let's find the coordinate vector of $(-1,5)$ in the span of $\{(1,2),(2,4)\}$. Well it turns out that this is impossible, to see this start out with the span of $\{(1,2),(2,4)\}$. So, $a(1,2)+b(2,4)=(a,2a)+(2b,4b)=(a+2b,2a+4b)$. Any vector in the $span\{(1,2),(2,4)\}$ is represented like so. Then we must find an a and b in F such that $(-1,5)=(a+2b,2a+4b)$. This is equivalent to solving both pairs of equations simultaneously. So, $-1=a+2b$ and $5=2a+4b=2(a+2b)$. Dividing the second equation by 2, we obtain $\frac{5}{2}=a+2b$. Now, we see the impossibility of this situation since $-1\neq\frac{5}{2}$, clearly! The defect in this set of vectors is that they are do not span the set.

Basically, a set of vectors is a basis for the vectorspace if 1) there are enough vectors in the set to ensure *there exists* a coordinate vector for every vector in the vectorspace, and 2) the vectors in the set are such that there is a *unique* coordinate vector for every vector in the vectorspace. It turns out that every vectorspace space has at least one basis, even the infinite dimensional vectorspaces: for instance the standard basis $\{e_1,e_2,e_3,...\}$, where $e_i=(0,...,0,1,0,...)$. That is a vector with a 1 in the i-th coordinate.

Now, it may seem like vectorspaces are a random algebraic structure, but I claim we have already seen in the above alot of vectorspaces, we just did not "look" at them as vectorspaces. For instance, the polynomial rings are vectorspaces, actually they are infinite dimensional vectorspaces. The basis is the set $\{1,x,x^2,...,x^n,...\}$, so the

$$span\{1,x,x^2,...,x^n,...\}=a(1)+bx+cx^2+...+nx^n+...$$

Therefore, we can see that this vectorspace is the space of all the polynomials with coefficients coming from the scalar field of the vectorspace. The coordinate vector for the polynomial $p(x) = 1+3x+4x^3$ is $(1, 3, 0, 4)$, likewise for the polynomial $q(x) = 2+6x+5x^2+7x^3$, the coordinate vector is $(2, 6, 5, 7)$. Addition is defined component-wise, so $p(x) + q(x) = (1, 3, 0, 4) + (2, 6, 5, 7) = (1 + 2, 3 + 6, 0 + 5, 4 + 7) = (3, 9, 5, 11) = 3 + 9x + 5x^2 + 11x^3$. We allow the addition of extra zeros at the tail of a coordinate to perform the addition. Also, scalar multiplication corresponds to $3p(x) = 3 \cdot (1, 3, 0, 4) = (3, 9, 0, 12)$. For an example of a finite dimensional vectorspace, look at the $span\{1, x, x^2\}$. This is the space of all quadratic polynomials, that is the set of all polynomials whose degree is less than or equal to 2. We have also seen another infinite dimensional vectorspace; the space of all decimal expansions. There is a component-wise addition and a scalar multiplication defined, refer back to the section on Real numbers to verify this.

One of the main example of vectorspaces is $\mathbb{R}^n$. This is the vectorspace of dimension n, such that every component of the vector is coming from $\mathbb{R}$. For instance, $\mathbb{R}^3$ is a 3-dimensional space, which is very similar to the 3-D world we live in. Newtonian physics describes the mechanics of the world in terms of force vectors in $\mathbb{R}^3$. It is a familiar situation that if I push you with a certain force and you push me with the same force then we shouldn't move. Furthermore, if I push you with a stronger force that you push me then you should move backwards. This situation is easily described using vectors, like so: Let v represent the vector of force I exert on you. Let w be the vector of force you exert on me. The first statement then becomes $v + w = 0$, from which we conclude that $v = -w$; which is our intuitive understanding of what's physically going on. If two forces act on the same object, but in exactly opposite directions, then the resulting force is zero. The second statement says $v + w = z$, where $z \neq 0$ and z is in the same direction as v, which is represented by $c \cdot v = z$ where $c > 0 \in F$.

Consider the two dimensional vectorspace, $\mathbb{R}^2 = span\{(1, 0), (0, 1)\}$. This *is* the plane studied in high-school geometry. Notice every point on the plane can be represented by a pair of numbers (a, b), but is precisely what is meant by the $span\{(1, 0), (0, 1)\}$. The fact that every point on the plane can be represented by a pair of numbers is a notion first described by Rene Descartes. In his honor, this representation of the plane is referred to as the Cartesian plane, or Cartesian Coordinates. Descartes apparently came to this idea one day laying in bed watching a fly crawling on his ceiling. He realized that the fly's position could be described has the distance from the North wall *and* the distance from

the East wall, in his room.[9] Properly this notion is called Analytic Geometry; but its power comes from its unification of Algebra and Geometry. By describing the points in the plane as (a, b) where a is the distance to the North wall, and b is the distance to the East wall, we transform the plane into a vectorspace, and thus the transformation of a purely geometry idea to the purely algebraic structure.

The operations in vectorspace have very nice geometric interpretations, especially in R^2 since I can draw them on a paper. Start by picking a specific point: the origin. Every vector is represented by an arrow from an origin. The effect of scalar multiplication is to grow or shrink the vector; by multiplying by -1 the vector flips all the around to the other side of the origin. See figure 1.

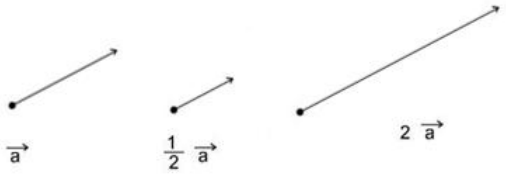

Figure 1:

The addition operation can be seen geometrically as follows. Draw two vectors on the paper, with an angle between them. Erect the rectangle formed. The addition of two vector is the diagonal of the rectangle. If there is no angle between the vectors, the addition reduces the same effect as scalar multiplication: growing and shrinking. Theres another's view of vector addition, which I like to refer to as triangle addition. $x, y, x+y$ as vectors form a triangle like so: x is the arrow from the origin, then place y at the endpoint of x, then $x + y$ is the triangle formed. Also, $x, y, x - y$ form a triangle. This time both x and y are starting at the origin, and $x - y$ is the other side of the triangle. See figure

When we consider a certain basis in a vectorspace, we impose a coordinate system on the vectorspace. Geometrically, this amounts to putting a grid on the paper so that we can assign every point a unique "address" (a, b). On the paper, one will have to first draw the axes: two specially denoted lines that intersect at the origin (this implies there is some angle between the vectors. Also, a unit interval on each axis must be established; place a special mark on each axis. This denotes what the distance "one" is, and also the positive direction on the axis. Lastly, for each axis make a mark at each unit length distance

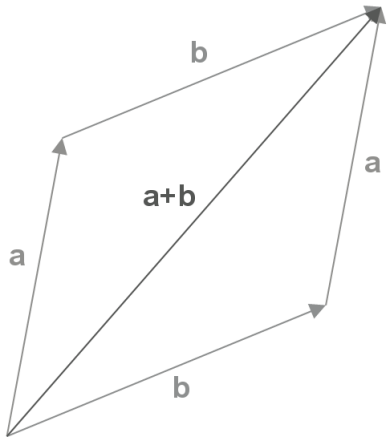

Figure 2:

and then draw the line parallel to the other axis. This creates a grid of quadrilaterals. Now any vector has a coordinate vector with respects to this basis (grid). The coordinate vector $(2, 3)$ with respect to a basis represents the vector starting at the origin and going 2 marks on the first axis, and then 3 marks on the second axis, then drawing the arrow from the origin to that point. Lets consider drawing a new grid, that goes through the same origin but uses different axes and unit intervals. The vector now has a new coordinate vector with respects to this new basis. See figure 3

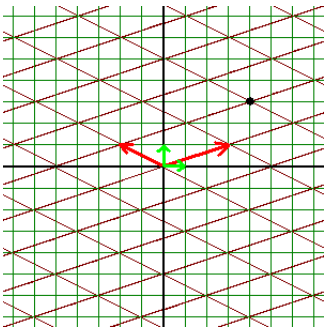

Figure 3:

The above example is important in that it highlights the difference between the vector and the coordinate vector with respects to a basis. The vector is just the arrow, its

is only when the grid is constructed that the coordinate vector arises. When we speak of a vector v in a vectorspace V, we mean just the arrow description. When we write $v = (v_1, v_2, v_3, ...v_n)$ that is referencing a certain basis and grid, and the equality says that v is represented this way in *that* basis. But $v = (w_1, w_2, w_3, ..., w_n)$ with respects to another basis. The question becomes is there a way of determining the w_i's from the v_i's?

Geometrically, the only pieces of information that is needed to create the the grid are the axes and their respective unit length. If we knew how the axes related and how there unit length related, this would be sufficient information to described the transformation from one basis to the other. Notice that for each axis and unit length, we could just draw the arrow starting at the origin ending at the given unit length. Then scalar multiplication of the vectors with the elements of the scalar field would generate the whole axis. The grid could then be constructed.

Have you ever seen a plane touch down on a landing strip with a lot of wind blowing? The plane has a vector associated with it, describing its velocity and its direction. The wind also has a vector associated with it describing its velocity and direction. When the plane tries to land it needs to take into account the wind vector; since its final direction and velocity is the sum of the plane vector and the winds vector. So in order to land straight on the runway, the plane must adjust its vector in the exact opposite way of the wind vector, so that the sum is where the pilot wants to land. This leaves the plane slanted as it comes in to land.

Algebras

The Complex numbers are also a 2-dimensional vectorspace over $\mathbb{R}$, with a basis $\{1, i\}$. That means that any complex number can be written in the $span\{1, i\} = a \times 1 + b \times i = a + bi$, for which the coordinate vector is (a, b). This coincides with the previous definition of $\mathbb{C}$. Recall, that addition was defined component-wise and if we defined a scalar multiplication as $\alpha \cdot (a, b) = (\alpha \times a, \alpha \times b)$, where $\alpha \in \mathbb{R}$. This shows that $\mathbb{C}$ is vectorspace over $\mathbb{R}$. Well, more precisely, once we show all this satisfies the axioms but those easily follow; in fact, we showed them in our discussion of $\mathbb{C}$.

In the algebraic structure of a vectorspace as defined above, there is no notion of multiplication of a vectors analogous to the multiplication in a field. However, there is a multiplication defined in $\mathbb{C}$. One can define a new algebraic structure called an *algebra*,

which is basically the union of a vectorspace and a field. Specifically, $(A, +, \cdot, \times)$ is an algebra if $(A, +, \cdot)$ is a vectorspace over a field F, and there is a new operation $\times$ such that $a \times b \in A$ when $a, b \in A$, and it distributes over the addition, that is $a \times (b+c) = a \times b + a \times c$ and $\alpha \cdot a \times \beta \cdot b = (\alpha\beta)(a \times b)$, where $\alpha, \beta \in F$. Note that $(\alpha\beta)$ is multiplication happening in the scalar field.

In order to turn the above definition of $\mathbb{C}$ as a vectorspace into an algebra, we need to define a multiplication operation. Well the one given earlier works: $(a, b) \times (c, d) = (ac - bd) + (ad + bc)i$. Also, we can define the multiplication operation *to be* an algebra, by assuming the distributive laws and the relation $i^2 = -1$. In that case we get,

$$(a + bi)(c + di) = a(c + di) + bi(c + di) = ac + adi + bci + bdi^2 = (ac - bd) + (ad + bc)i$$

This implies that $(\mathbb{C}, +, \cdot, \times)$ is an algebra. $(\mathbb{R}, +, \cdot, \times)$ is also an algebra, but this is kind of redundant. $(\mathbb{R}, +, \cdot)$ is a 1-dim vectorspace over itself. In effect, the $\cdot$ and $\times$ operation *are* the same.

Another interesting example which arises in Number Theory are the Number Fields. For instance, $\mathbb{Q}(\sqrt{2})$ is the 2-dimensional vectorspace over $\mathbb{Q}$ with basis $\{1, \sqrt{2}\}$. These are the number of the form $a + b\sqrt{2}$. Addition and scalar multiplication are defined as you would expect:

$$3(2 + 5\sqrt{2}) + (6 - 3\sqrt{2}) = (6 + 15\sqrt{2}) + (6 - 4\sqrt{2}) = (6 + 6) + (15 - 3)\sqrt{2} = 12 + 12\sqrt{2}$$

. Multiplication defined as in $\mathbb{C}$ instead with the relation $\sqrt{2}^2 = 2$. So,

$$(a+b\sqrt{2})(c+d\sqrt{2}) = a(c+d\sqrt{2})+b\sqrt{2}(c+d\sqrt{2}) = ac+ad\sqrt{2}+bc\sqrt{2}+bd\sqrt{2}^2 = (ac+2bd)+(ad+bc)\sqrt{2}$$

This makes $Q(\sqrt{2})$ an algebra, however it is more typically looked at as just a field since the multiplication implies the scalar multiplication. The same could be said about the Complex numbers above.

Consider the example $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. This a four-dimensional algebra with basis $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ and the relations: $\sqrt{2}^2 = 2$, $\sqrt{3}^2 = 3$, $\sqrt{2}\sqrt{3} = \sqrt{6}$, and $\sqrt{6}^2 = 6$. What is the product $(3 + 6\sqrt{3})(5 + \sqrt{2})$? We have glance over an important notion in these number fields: linearly independence of the basis. For instance, $Q(\sqrt{2}, \sqrt{3}, \sqrt{6})$ is the same as $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ because $\sqrt{6}$ is dependent on $\sqrt{2}$ and $\sqrt{3}$.

Solving Equations

We have seen a lot of different algebraic structures. Also, we have seen some structures, like $\mathbb{C}$, that could be view under different algebraic lights, a vectorspace over $\mathbb{R}$ or the algebraic completion of $\mathbb{R}$. This section present the connection between a algebraic structures and solving equations. In the above we gave the axioms that define an algebraic structure. A group was $(G, \cdot)$ such that the operation was close and associative, and there exists inverses and an identity. These four axioms are sufficient to guarantee a unique solution to a the equation $A \cdot x = B$, for all $A, B \in G$, and that $x \in G$. The following derivation shows why:

$$\begin{aligned} A \cdot x &= B \\ A^{-}1 \cdot (A \cdot x) &= A^{-}1 \cdot B \\ (A^{-}1 \cdot A) \cdot x &= A^{-}1 \cdot B \\ I \cdot x &= A^{-}1 \cdot B \\ x &= A^{-}1 \cdot B \end{aligned}$$

If you want a unique solution to the equation $Ax + B = C$, the elements A, B, C must be from a field.

$$\begin{aligned} Ax + B &= C \\ (Ax + B) - B &= C - B \\ Ax + (B - B) &= C - B \\ Ax + 0 &= C - B \\ Ax &= C - B \\ A^{-}1(Ax) &= A^{-}1(C - B) \\ (A^{-}1A)x &= A^{-}1(C - B) \\ Ix &= A^{-}1(C - B) \\ x &= A^{-}1(C - B) \end{aligned}$$

To solve equations of the form $ax^2 + bx + c = 0$, use the quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

The proof of the quadratic formula uses a method called completing the square. The idea is that if somehow we transform the equation to something of the form $(x + b)^2 = c$ then the answer is simply $x = \sqrt{c} - b$. Without loss of generality, assume that $A = 1$, since if not we could just divide through by $A^{-}1$ and obtain the equation $x^2 + \frac{b}{a}x + \frac{c}{a} = x^2 + \hat{b}x + \hat{c} = 0$. These are called the monic polynomials, leading factor equal to one. Notice $(x + b)^2 = x^2 + 2xb + b^2$. The proof is finished by:

$$\begin{aligned}
x^2 + bx + c &= 0 \\
x^2 + 2\left(\frac{b}{2}\right)x &= -c \\
x^2 + 2\left(\frac{b}{2}\right)x + \left(\frac{b}{2}\right)^2 &= \left(\frac{b}{2}\right)^2 - c \\
\left(x + \frac{b}{2}\right)^2 &= \left(\frac{b}{2}\right)^2 - c \\
\left(x + \frac{b}{2}\right)^2 &= \left(\frac{b^2}{4}\right) - c \\
\sqrt{\left(x + \frac{b}{2}\right)^2} &= \pm\sqrt{\frac{b^2}{4} - c} \\
\left(x + \frac{b}{2}\right) &= \pm\sqrt{\frac{b^2 - 4c}{4}} \\
x &= -\frac{b}{2} \pm \frac{\sqrt{b^2 - 4c}}{2} \\
x &= \frac{-b \pm \sqrt{b^2 - 4c}}{2}
\end{aligned}$$

In order to ensure that there exists a unique solution to the equation: $ax^2 + bx + c = 0$, the coefficients a, b, c must lie in the number field $\mathbb{Q}(\sqrt{b^2 - 4ac})$. If we where consider the

quartic equation:

$$(x^2 + ax + b)(x^2 + cx + d) = x^4 + (a + c)x^3 + (b + ac + d)x^2 + (ad + bc)x + bd$$

then, the number field needed is $Q(\sqrt{a^2 - 4b}, \sqrt{c^2 + d})$.

There does exist an analog to the quadratic formula for the cubics (degree 3) and quartic (degree 4), although the formulas are rather lengthy and complicated. However, once we get to the quintic (degree 5), Evariste Galois proved that there does not exist formula for solving the general quintic polynomial. [5] There are some quintic polynomials that can be solved explicitly: like $x^5 = 32$ then $x = \sqrt[5]{32} = 2$. This statement is for the *general* quintic: $ax^5 + bx^4 + cx^3 + dx^2 + ex + f = 0$. One can not write a the analogous "quadratic formula" for this polynomial. The proof of this statement is quite simple it follows from the fact that S_5 is not solvable. The language needed to describe what a is solvable group and how this connects to groups is the realm of Field Extensions and Galois Theory. The basic idea involves the number fields, and permutations of the roots of polynomials, that is why S_n is used.

The number fields are a vectorspace over $\mathbb{Q}$, they extended the field of $\mathbb{Q}$ to a bigger field $\mathbb{Q}(\sqrt{D})$ or $\mathbb{Q}(\sqrt[3]{D})$. Galois noticed a correspondence between these field extensions and subgroups of the symmetric group S_n. This is truly amazing connection between vectorspaces, field, and groups: three fundamental algebraic structures. This is Galois Theory.

Geometry

Geometry is the study of shapes, and the properties left invariant under a certain group of transformations; as defined by the Erlanger Program.[2] The Elements, by Euclid, showed that geometry could be defined as an axiomatic system, where there are certain undefined terms: points, lines, and a relation of incidence (i.e. a point is incidence on a line, or a line is incidence on a point) and assumed axioms. Any theorem, in this system, is valid if and only if there is a finite string of statements from which the theorem can be inferred using a suitable system of logic.

The axiomatic system Euclid gave for, what is now called Euclidean Geometry, is as follows:[6]

Common Notions

- Things which are equal to the same thing are also equal to one another
- If equals be added to equals, the wholes are equal.
- If equals be subtracted from equals, the remainders are equal.
- Things which coincide with one another are equal to one another.
- The whole is greater than the part.

Undefined Terms

- Point
- Line
- Relation of Incidence
  - Point incident to a line
  - Line incident to a point

Postulates

1 Given any two points, there exist a unique line incidence to both points.

2 Any line segment can be extended infinitely.

3 Given a point and a length, there exist a unique circle with center as the given point, and the radius the given length.

4 All right angles are congruent.

5 If a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced infinitely, meet on that side on which the angles are less than the two right angles.

He also included some definitions for things like Right Angles, Circle, Perpendicular and Parallel; however, these are *defined* terms, in that they depended on the undefined terms. For instance, parallel lines are those lines which are incident at no point. The definition for Right Angles is: if a line is incident on another line, and the adjacent angles are congruent, then the angles are called Right Angles, and the lines are said to be Perpendicular. Note that these definitions only make the language of the geometry much

simpler; they are technically not needed, for i can just state the definition of Right Angles, without any reference to the term itself only appealing to the undefined terms. Although, it is much more convenient to introduce these notions as part of the language used.

The Elements also contains 48 Theorems of plane geometry; the cornerstone of which is the Pythagorean Theorem, and its converse. As an aside, it is *not* known whether Pythagoras himself proved or even conceived the theorem which now bears his name; however, it was the school he founded who did prove this theorem. Along the same lines, Euclid himself did not prove nor conceive all the ideas and proofs in plane geometry, however he collected most Greek mathematics and assembled them into The Elements. This is just to give credit where it is due, and to consider that when we talk about Pythagoras' Theorem, it is his school and not him we are immortalizing.

As elegant as this theory is, it does suffer problems. The most controversial being his Fifth Postulate, which will be looked at closely in a following section. Other subtle problems started to arise during the Renaissance, where a new spark for Mathematics (along with all scientific queries) initiated a systematic review of Euclid's postulates and errors where discovered, ironically embedded into even the Axioms are errors. For instance, the Fourth Postulate: All right angles are congruent; has no meaning in the axiomatic system described above. Indeed, there is no meaning for the term congruent, furthermore as Hilbert later showed, the relationship of Congruency is an independent relation (can not be described in terms of the already listed undefined terms and relations). This problem is surmountable, in that when we can append to the undefined terms the relation of Congruency.

David Hilbert, along with others, tried to remedy these subtle "assumptions" Euclid made. The idea of an axiomatic system, which will be formally discussed later, is that no reference to intuition is needed to make deductions. All that is needed is the undefined terms, the axioms, and a suitable system of logic from which a deduction is clearly either valid or not. Euclid was on the right track, but he still he made hidden assumptions (hidden in that they were not explicitly stated). For instance, in the first theorem of the Elements (which is actually more a construction) Euclid makes an assumption about the continuity of circles; however, the concept of continuity is never discussed. Here is the theorem and its proof. Can you spot the hidden assumption?

Proposition 1[2]

On a given finite straight line, construct an equilateral triangle.

Proof:

Let AB be the given finite straight line.
By Axiom 3, there exist a unique circle with center A, radius AB.
Similarly, there exist a unique circle with center B, radius AB.
These circles intersect at a point C.
By Axiom 1, there exist a unique finite line through A and C, call it AC.
Similarly, there exist a unique finite line through B and C, call it BC.
The lengths AB and AC are equal, by the definition of the radius of a circle.
The lengths AB and BC are equal, by the definition of the radius of a circle.
The lengths BC and AC are equal, by the Common Notion of Transitivity.
Therefore ABC is an equilateral triangle.

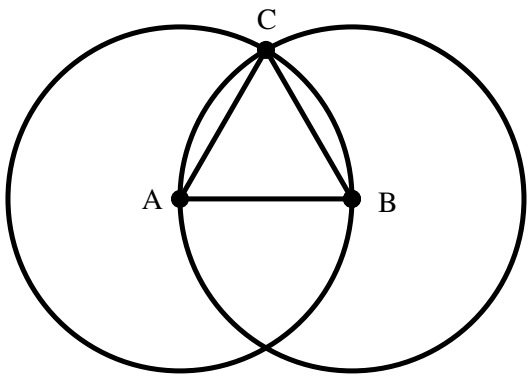

Figure 4:

There is only one line in the proof that does not reference a postulate or a common notion. "These circles intersect at a point C." Intuitively speaking, and especially looking at the picture, it is absolutely clear these circles must intersect! Try it yourself with a compass and a ruler, and for every line segment one draws, this construction will result in the assumed intersection. So, it is understandable that Euclid, nor anyone else in the almost 1500 years since it was written, did not see the flaw in assuming such an

intersection. However, in light of a complete formalism of geometry (or math in general) this was an unsound foundation from which geometry could be developed.

Hilbert remedied this and other problems, such as the relationship of betweenness (which Euclid also assumed), and formulated was is now considered a consistent and complete (in so far as it can be) formalism of the so called Euclidean Geometry. Although, Hilbert formulation is longer, has six undefined terms (point, line, plane, incident, congruence, betweenness) and 16 postulates; it serves as an adequate foundation for plane geometry. This resolution is a much more appealing situation, since there are no "hidden" assumptions, and thus no need to appeal to intuition.

A model of an axiomatic theory is a designation of the undefined terms, and if this interpretation satisfies the axioms of the theory, then any Theorem which is valid in the axiomatic theory, also applies to the specific model. So, the main problem with Euclid's foundation was not that it didn't describe plane geometry, it was that the hidden assumptions, made it difficult to apply this theory to any other type of mathematical structure. I do not think Euclid was concerned with this, however in this new age of Abstraction, this *is* the main goal of Mathematics: to study a specific concrete example, but then to generalize and abstract the key concepts to apply to a vast range of concrete examples: that is the power of Mathematics. More on this in the section on Axiomatic Theories.

Parallel Postulate

The length and the complicated language of the Euclid's Fifth Postulate, the so-called Parallel Postulate, led most to assumed that this postulate was properly a Theorem; that is, it could be deduced from the other axioms. This notion was more intuitive than anything else. No one had any particular reason to suspect this, other than it was an "ugly" assumption.

"The mathematician's patterns, like the painter's or poet's, must be beautiful. The ideas, like the colours or the words, must fit together in a harmonious way. Beauty is the first test: There is no permanent place in the world for *ugly* mathematics." – G. H. Hardy

Still, everyone search for the dependence of the Parallel Postulate. In this search, which was not fruitful, a list of equivalent axioms was discovered. The notion of two axioms being equivalent is that if an axiom where replaced by an equivalent axiom, then the theory generated by the new set of axioms is the "same" as before. It is in this light that the foregoing axioms are equivalent.[2] Playfair's Axiom: Given a line and a point not on the line, there exist exactly one parallel line to the given line. The sum of the angles of a triangle is exactly $\pi(= 180^\circ)$. The area of a circle is exactly πr^2. There exists a pair of similar triangles. (Similar means that all respective angles are congruent, while their respective lengths may differ.) There exists a pair of straight lines everywhere equidistant from each other. (i.e. There exist a pair of parallel lines) Given any three non-collinear points (there does not a exist one line through all three points), there exist a unique circle through all three points. If three angles of a quadrilateral are right angles, then so is the the fourth angle.

The issue people had with the Parallel Postulate, as stated by Euclid, was still not resolved by these equivalent axioms. Although, one may argue the "beauty" of Playfair's axiom was enough to shadow the "ugliness" of the Parallel Postulate. But at this point in the historical development, so many attempts had tried and with no success that the temptation to keep trying to prove its dependence did not subside.

Girolamo Saccheri (1733) tried to establish the Parallel Postulate, by negating it, and trying to arriving at a contradiction. Notice, this is backwards from what people were trying to prove, in that Saccheri was trying to establish its independence, while most of the work up to that point was in trying to prove its dependence.[3] If successful, this would unequivocally establish the independence of the Parallel Postulate forever more. This is the right idea, however, he was still diseased by the Greek/Euclidean view of the world's geometry. The Greeks where very geometric in their math, in part since algebra was not on the footing it is today, but also they held the view that there is *one* geometry that of the real world, and that the geometry described by Euclid is that geometry; i.e. Euclidean Geometry *is* the geometry of the real world. Saccheri was still infected by this disease as he proceeded in his attempt to prove the independence of the Parallel Postulate. His main flaw was that he was working under the assumption that there is *one and only one* geometry. Although, he did make a mistake in his proof, of which I will outline in the following, he set the stage for curing of the Euclidean disease and for society to accept that there is other kinds of geometry. In particular, Einstein's General Relativity Theory

showed that space-time itself is a curved space, not the flat space of Euclidean Geometry.

There are two different ways to negate the Parallel Postulate, amounting to three different case. For more aesthetic purposes than anything else, let's negate Playfair's axiom; which is equivalent to Euclid's 5th Axiom, anyway. It is easy to show that the negations of the Fifth Axiom are equivalent to the following negations.

Three case of Playfair's Axioms

1 Given a line and a point not on the line, there exist **no** parallel line to the given line.

2 Given a line and a point not on the line, there exist **exactly one** parallel line to the given line.

3 Given a line and a point not on the line, there exist **more than one** parallel line to the given line.

Saccheri was convinced of existence of the "Absolute Geometry", and unfortunately since these axioms, each in turn, with the rest of Euclid's axioms (or more properly the rest of Hilbert's axiom) each generate a new Geometry; to which surely would bear his name as the discover. But, the diseased mind sees only what it was want, if only he had more faith in his logic rather than faith in society. He showed, albeit incorrectly, that the first possibility led to a contradiction. So all he needed to find was a contradiction arising from the third possibility, however he could not find one. I want to reiterate his disease at this point, since he was trying to prove there was only one geometry, the lack of a contradiction in the third possibility probably led him to doubt in own mind. If he would have taken a step back, he could have realized that this third possibility led to an entirely new, yet consistent geometry; which would now probably be called Saccheri Geometry. We have already seen this disease, in practically every step in the construction of the complex numbers, there was a social inertia that needed to be overcome. The concept of zero, and negative numbers, the very name of imaginary numbers alludes to the dis-ease in which these numbers where accepted. But like in the case of imaginary numbers, non-euclidean geometry (that which is based on a negation of Playfair's axiom) is a much more unifying theory. Indeed, Einstein's theories imply this universe is far from the euclidean view of universe and its' geometry.

Hyperbolic Geometry

About 100 years later, Gauss was the first known reference to "curious geometry, quite different from ours, but thoroughly consistent".[3, 165] He was working under the assumption that the sum of the angles of a triangle is less than $\pi(= 180^\circ)$; which is an equivalent axiom 2 above. However, he did not publish his findings, for fear it would "harm his considerable reputation if he were to go on record as saying that Euclidean geometry is not the only one possible".[3, 165] János Bolyai, a Hungarian artillery officer, published results of this new geometry as an appendix to his father's book; only to get a letter from Gauss explaining that he had "developed [the subject] to his satisfaction" already. Three years prior Nikolay Lobechevsky had published similar results under the title *Imaginary Geometry*. Notice how even though more and more results are being produced, every one is still tainted by the Euclidean dogma of a universal geometry: the name *Imaginary Geometry*, the reluctance of Gauss to publish anything, the placement of the Bolyai's publishing as an appendix.

This new geometry was shown to be consistent, in that a Theorem could not be proved both true *and* false. Properly speaking, hyperbolic geometry is the axiomatic theory generated from Euclid's first four axioms, the hidden axioms filled in by Hilbert, and the Hyperbolic Axiom: Given a line and a point not on the line, there exist **more than one** parallel line to the given line. Euclid's first 28 theorems only used the first four axioms, hence all of those propositions hold in this geometry too. Also, the list of equivalent axioms of the Parallel Postulate hold in this geometry; well, their negations!

Given a line and point not on the line, there are more than one lines that are parallel to the given line. The sum of the angles of a triangle is less than $\pi(= 180^\circ)$. The area of a circle is less than πr^2. If three angles of a quadrilateral are right angles, then the fourth angle is less than right. There does not exists a pair of (properly) similar triangles. (If all respective angles are congruent, then triangles are congruent) There does not exists a pair of straight lines everywhere equidistant from each other.

The negation of above axioms are properly in the realm of logic. Notice that the "pure" negation to Playfair's Axiom: Given a line and a point not on the line, there does not exist one line that is parallel to the given line. However, there are two distinct cases: either no line, or more than one lines are parallel to the given line. Along the same lines, we negate the above accordingly, as in axioms 2 and 3 where "exactly" was replaced by

"less than". The power of logic starts to show face here, a more sufficient treatment is done in a latter section on Logic. Notice how since the equivalences are valid in the logic of euclidean geometry and hyperbolic geometry is essentially the same geometry less with the hyperbolic axiom. So, the burden of proof of the above statements lies in the logic of axiomatic theories, from which a consistent and rigorous negation schema is inherited. This is also to show how even though the pursuit of the "proof" of the parallel postulate wasn't as fruitful as desired (especially by the Greeks) but it led to this equivalent list, which leads (logically) directly into this new geometries discovered.

Proclus, a Greek mathematician, is the earliest source of criticism about Euclid's Parallel Postulate.[3] He revised the definition of parallel from: Parallel lines, are straight lines which being produced infinitely in the same plane, do not meet one another in either direction. to Parallel lines are equidistant everywhere. Proclus aim was to find a pair of lines that are parallel in Euclid's sense yet not equidistant everywhere. His goal was never accomplished, and yet his distinction is significant in Hyperbolic geometry, as is shown by (Second to last) negation axiom above. There does exist parallel lines in hyperbolic geometry, however there does not exists lines that are everywhere equidistant: a strange notion indeed. However, this is where one's intuition must be suppressed and trust the logic, at least until their intuition becomes acquainted with this new geometry.

The concept of Parallelism takes one a new form in this geometry (Euclid's sense of parallel). For instance, there are two distinct notions of parallelism: sensed parallel and ultra parallel. The new concept is properly sensed parallelism, in that ultra parallelism is very much that same as the euclidean notion of parallel. In that if one line is ultra-parallel to another, and a third line is ultra-parallel to the latter, then the former line is ultra-parallel to the third. Simply stated as ultra-parallelism is a transitive relation.

Sensed-parallels are those lines which "meet" at infinity; there are two distinct cases (at least in 2-D hyperbolic geometry) right-sensed parallel and left-sensed parallel. A line is left-sensed parallel to a given line, if and only if their only point of intersection is the so-called point at infinity, but in the left direction; respectively for right-sensed parallels. Basically, the left-sensed parallel of a given line through a given point is the "first" line parallel (in that they do not intersect) in the left direction. In that if you rotate counter-clockwise it is the first line through the given point that is does not intersect the given line. Also, these are prime examples of lines that are parallel, yet are not equidistant. If AB represents the distance from a line to it's left-sensed parallel through a given point, then

the length AB approaches zero as the lines are produced infinitely in the left direction. These parallels are uniquely determined by a line, a point, and a direction. Sensed-parallelism is not a transitive relation, which is why above I stated that ultra-parallelism is more analogous to the notion of parallelism in Euclidean Geometry.

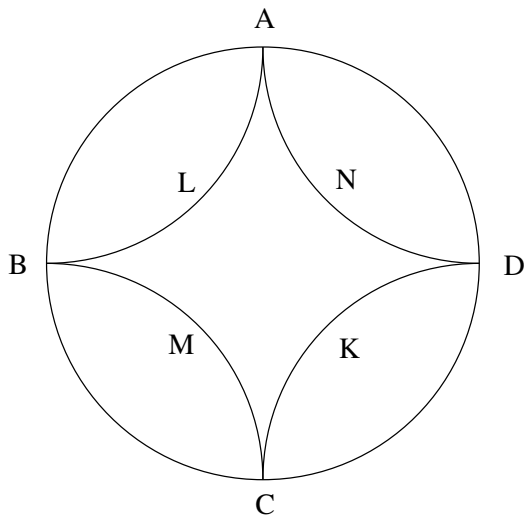

Figure 5:

In the figure 5, line L and K are ultra-parallel, while N is the left-sensed parallel to L and M is the right-sensed parallel to L. What are the lines right-sensed, left-sensed and ultra-parallel to M? Notice in the "quadrilateral" that is formed, the angles sum to more than $360°(= 2\pi)$. The important point about this model is that it "preserves" angles. However, it does not preserve distance, in that, for any two point of the model (interior to the fundamental circle) the their distance in the model does not represent the distance between the points in the hyperbolic geometry.

The above is common place in the abstraction of mathematical structures. We notice that both the concepts of equidistant/parallel in euclidean geometry break apart to form two distinct notions in hyperbolic geometry; similarly, with parallelism. We have already seen this in our brief discussion of prime/irreducible. This is just to point out, that sometimes the deeper we dive into a subject, the concepts inherited from the parent theory may dissolved out into distinct concepts.

How do we visualize the hyperbolic geometry? We need to find a model for it. A

model is simply a designation of the undefined terms, and if this interpretation satisfies all the axioms, then it is a model for the axiomatic system. There are several models for hyperbolic geometry: Poincaré disc model, his upper-half plane model, and Klein disc model. Let's explore Poincaré disc model. Accordingly, we need to assign some concrete meaning to our undefined terms: point, line, incidence We consider a circle, called the fundamental circle, in the Euclidean plane. Then a *point* is a euclidean point that is interior to the circle. A *line* is that part of a circle lying interior to the fundamental circle and is orthogonal (perpendicular) to the fundamental circle or a diameter of the fundamental circle The incidence relation is inherited from the euclidean plane relation of incidence. As we will see later, for a Axiomatic system to be consistent all that is needed is to present a model for it. However, since our model is embedded into the Euclidean plane, which is desirable in that we inherit all its' properties, the consistency of Hyperbolic geometry is dependent on the consistency of Euclidean plane geometry. In fact, the converse is also true; so Hyperbolic Geometry is as consistent as Euclidean geometry. This is as satisfying as the situation can get, as we will show later.

In the figure 6, there are a lot of interesting figures. First, these are all *straight* lines in the hyperbolic geometry. What would the arc of a circle that does not intersect the fundamental circle perpendicularly? Find some triangles and sum the angles; its less than $180°$.

Elliptical Geometry

Saccheri's attempt to prove the independence of the parallel postulate led him a to indirect proof, where by negating the postulate and attempting to arrive at a contradiction he would unequivocally prove its independence. The case where the negation leads to the statement: Given a line and a point not on it, there does not exist any parallel lines to the given line. This will be called the Elliptical Axiom. Saccheri reached a contradiction here, however the contradiction was based on the infinitude of lines: i.e. Euclid's Second Axiom. Bernard Riemann was the first to make a distinction between the notions of infinitude of a line and its unboundedness. For instance, consider the surface of the earth and the equator. This is line which has a specific length, namely its' circumference, and hence is not an infinite line. Yet, it is unbounded in the sense that if I were to walk on the equator, I could keep walking forever and never reach an end.

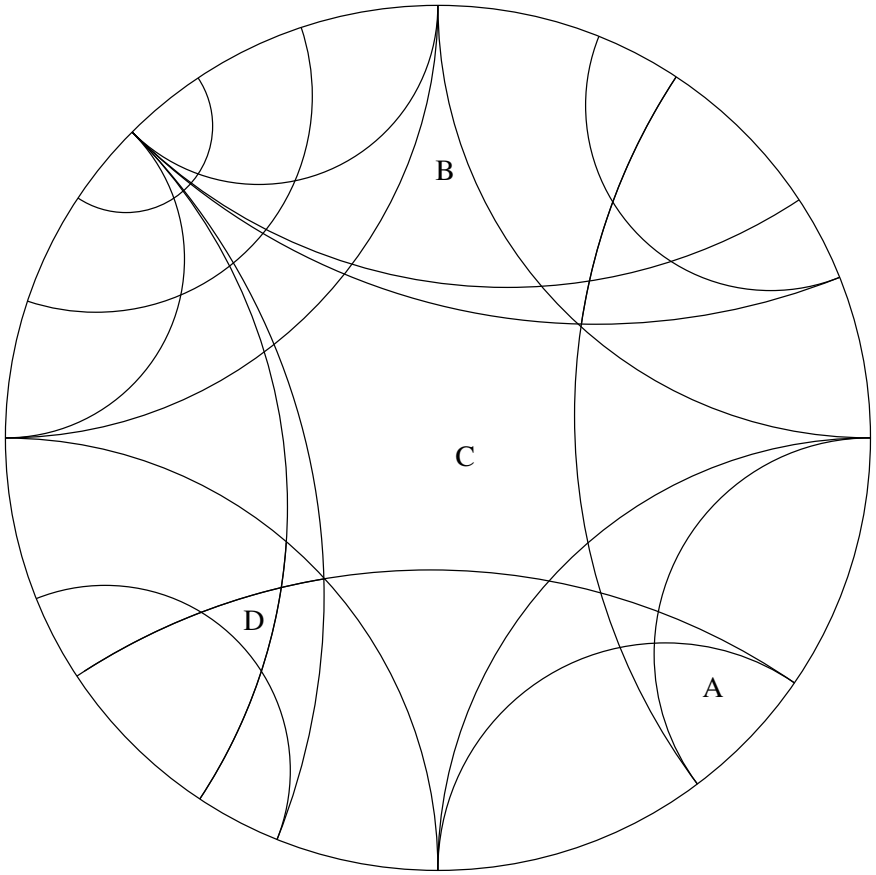

Figure 6:

With this distinction in mind, Euclid's Second Postulate can be restated to say: A finite line can be produced continuously in a line; the line obtained is unbounded but *not* necessarily of infinite extent. However, the resulting axiomatic system, Euclid's first four postulates and the Elliptical Axiom, does not result in a consistent system. For instance, Proposition 27 proves the existence of parallel (non-intersecting) lines. This proposition was proved using only the first four axioms, thus should be logically valid in Elliptic Geometry as well, however this leads to the inconsistency since the proposition proves the existence of parallel lines, yet the Elliptical Axiom assert its negation. A system is inconsistent if a statement can be proved true and false. Examining the proof of proposition 27, one needs to invalidate this proof in order to make Elliptical Geometry

consistent. The proof is presented in the following:

Proposition 27[2]

If a straight line falling on two straight lines make the alternate angles equal to one another, the straight lines will be parallel to one another.

Proof by Contradiction:

Let EF be the transversal of lines AB and CD, such that the alternate angles, AEF and EFD, are congruent. Assume that AB is not parallel to CD, i.e. there exist a point, G, that is incident to both AB and CD. Consider the triangle formed by GEF, then the exterior angle AEF is equal to the interior and opposite angle EFG. By Proposition 16: In any triangle, if one of the sides be produced, the exterior angles is greater than either of the interior and opposite angles. We have reached a contradiction! Thus AB is parallel to CD.

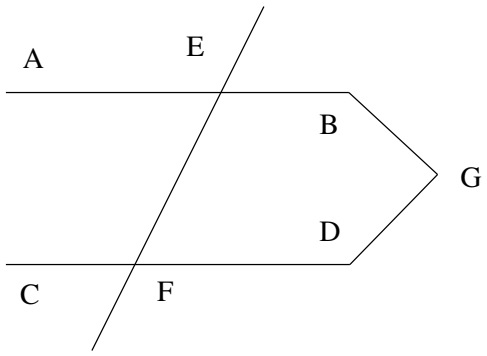

Figure 7: Proposition 27

The main tool in this proof was proposition 16, which we will now examine its proof.

Proposition 16[2]

In any triangle, if one of the sides be produced, the exterior angles is greater than either of the interior and opposite angles.

Proof:

Let ABC be a triangle, and let the side BC be produced to D.
Let AC be bisected at E and let BE be produced in a straight line to F: By Prop. 10
Let EF be made equal to BE: By Prop. 3
Draw the line from F to C: By Axiom 1
Let AC be drawn though to G: By Axiom 2

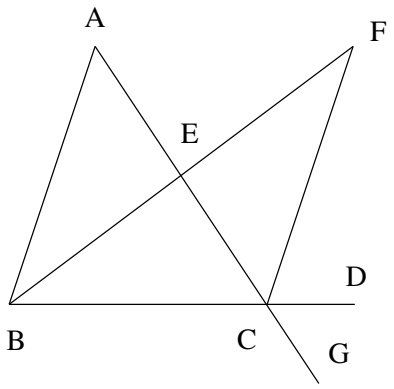

Figure 8: Proposition 16

Since $AE \cong EC$, $BE \cong EF$, the two sides, AE, EB are equal respectively to CE, EF
The angle AEB is equal to FEC, for they are vertical angles: By Prop 15.
Therefore, the triangles ABE and CFE are congruent. By SAS.
So, all respective lengths and angles of the triangles are congruent.
So in particular, BAE is equal to ECF.
But, the angle ECD is greater than the angle ECF.
Therefore, ECD is greater than the angle BAE.
Similarly proved for the angle BCG, if BC is bisected and the above proof repeated.

The main error in this proof is made when assuming that the line BF can be drawn. In effect, Euclid constructed a line segment, BF, double the length of the original line segment, BE. The question is then, how can one be sure such a line segment exist, i.e. how does one know that F doesn't wrap around and lie *on* the line segment BE. This is the distinction already pointed out by Riemann: the difference between unboundedness and infinite in extent. So, to fix the consistency of Elliptical Geometry, we must make the

revision in the second axiom of Euclid as stated above.

However, there is since inconsistency in this geometry. Consider the following proof of the existence of parallel lines, different from Prop. 27 above.

Proof A

Let A and B be two points on a line l.
Let m and n be lines perpendicular to l at A and B, respectively. By Prop. 11
Assume m and n are not parallel, let C be their intersection.
Construct the segment AC', where C' lies on the opposite side of l that C does, and such that length AC is equal to length AC'.
Draw the line $C'B$.
Triangles ABC and ABC' are congruent, by SAS.
Thus, angle ABC' is a right angle.
By Prop 14: C', B, C are collinear (lie on the same line).
But, now we have constructed two distinct lines both passing through the points C and C'.
We have reached our contradiction! Thus m and n are parallel.

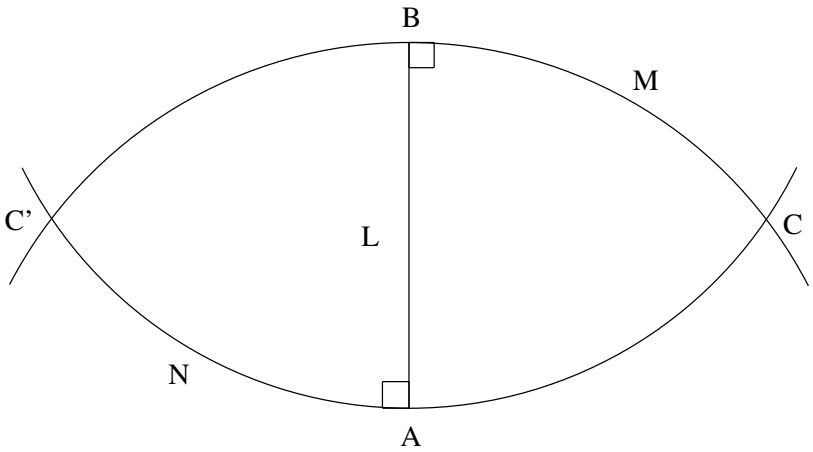

Figure 9: Proof A

We can see the contradiction was arrived at because of Proposition 14 and Axiom 1.

So, one of the inconsistencies arises from the first axiom: Given two points there is a unique line through both. The other inconsistency arises in the proof of Proposition 14. The main issue is the assumption that a line separates the plane. This is equivalent to what is now called Pash's Axiom: A line containing a vertex of a triangle and a point interior to the triangle will intersect the opposite side of the triangle. Either way, we are left with two statements, either of which if negated will result in a consistent geometry. If the first is negated and the second keep, then the geometry generated is called Double-Elliptical Geometry. Indeed, this is the geometry of the surface of the Earth. On the other hand, if the second is negated and the first kept, then the geometry generated is called Single-Elliptical Geometry.

This geometry, although probably more useful for real world application, like flying airplanes, lacks the compatibility afforded to hyperbolic geometry by Euclidean geometry. In hyperbolic geometry, we only replaced the parallel postulate with the hyperbolic axiom, and as such there was already all of Euclid's 28 first propositions at our disposal. Yet, in elliptical geometry there was some more modifications need to result in a consistent geometry, and as such, most of Euclid's work does not apply to Elliptical geometry, unless one examines each proof to assure that none of the modified axioms were used (at least incorrectly in their new senses).

Consider a plane flying from Miami to Madrid. The geometry of the surface of the earth is the model for Double-Elliptical Geometry, and thus the "straight lines" are the circle that radius equal to the radius of the earth. What does this imply about the flight of the plane? It must travel along the "straight" line path, thus will travel on the arc of the circle joining Miami and Madrid. This is easier to think of on a **flat** map: draw the straight line from Miami to Madrid. This path is **not** the path the plane would take. The plane must curve its trajectory toward the north pole because this is the "straight" line in the Double-elliptical geometry. To avoid confusion between "straight line" and straight line, the word *geodesic* is used for the former. In reality, there is another force that will curve the flight of the plane even more: the Coriolis effect.

The sum of the angles in a triangle in Elliptical Geometry is more than 180°. Take the surface of the earth as the model. Consider the triangle formed by the equator, the prime meridian, and any other meridian. Each meridian is perpendicular to the equator, thus there is 180° with just two angles. The other angle has to be greater than zero (if not the figure is not a triangle) but less than 180°. Thus the sum of the angles of this

triangle is in strictly greater than 180° and strictly less than 360°.

Mappings and Functions

A mapping is a correspondence, relationship, between two collection of objects. Consider the Name mapping; this assigns to each person in the world their name. There are plenty examples of these "social" mappings: like height, weight, age, race, etc. When considering mappings, there is a Domain and Range associated with each mapping. Notationally this is denoted by $M : \mathbb{D} \to \mathbb{R}$ (Here $\mathbb{R}$ does not denote the Real numbers, but the Range. Past this section $\mathbb{R}$ will once again denote the Real numbers.) So, Name has as its domain the set of all people in the world, and as its range the set of all names in the world. We could also talk about the "inverse" name mapping, Name^{-1}: {set of all names} $\to$ {set of all people}.

There are two other notions also associated with a mapping: injectivity and surjectivity. The former, also know as one-to-one, states that the mapping sends different elements to different elements. Precisely, it states that if $x \neq y \in \mathbb{D}$ then $M(x) \neq M(y) \in \mathbb{R}$. The notion of surjectivity, also called onto, describes that the mapping hits *every* element in the range. Formally, it states that $M(\mathbb{D}) = \mathbb{R}$, where $M(\mathbb{D})$ is the set of all point in the range that are mapped from some element in the domain. The equation $M(\mathbb{D}) = \mathbb{R}$ says that all the points in the range have some corresponding element in the domain.

Pictorially, we can represent collections of objects by a set of points (infinite or finite). A mapping is then an assignment of arrows starting from every point in the domain, ending at some point in the range. A mapping is injective if every point in the range has *one and only one* arrow pointing to it. A mapping is surjective if every point in the range has some arrow point to it. Note, that the mapping is the collection of all arrows.

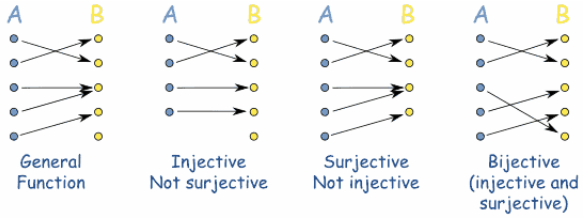

Figure 10:

Looking back at the Name mapping, it is clearly surjective since everyone person in the world has a name. Is it injective? No, since there are people with the same name, therefore in the range there are some points that have more than one arrow pointing to it. Is the mapping $Name^{-}1$ surjective? Injective? This mapping goes from the set of names to the set of persons. Thus this mapping is surjective: every person has a name. The mapping is Injective: every person has a one name, thus there is only one arrow pointing.

There is a subtle problem with the $Name^{-}1$ mapping, is not a Function. This addition condition asserts that there is exactly one arrow coming from every element in the domain. The $Name^{-}1$ maps a name to every person that has that name. So if n people have the name Antonio, then there are n arrows coming from Antonio.

A mapping then has five notions associated with it: Domain, Range, injective, surjective, and function; The first two notions describe the starting and ending places for the arrows; while the last three describe how the arrows behave. Strictly speaking, a mapping, its domain and range are atomic (indivisible), that is when we talk about a mapping it necessarily has a domain and range. The latter properties are each independently true or false.

There is one more property we could talk about in terms of mappings: whether every point in the domain has a arrow coming from it. This is usually assumed! Since if the Domain includes points that are not mapped to any point in the range, just restrict the domain by excluding those points. The same could be said about the the Range. A mapping can be surjective if we restrict the range by omitting those points that have no arrow coming to them. The restricting of the Domain causes no problems in practice, however the surjectivity of a mapping is a useful notion, just thus we do not restrict the range. Whether a mapping is a function describes two notions, 1) every point in the domain has at least one arrow, and 2) every point in the domain has at most one arrow. The former notion is that of restricting the Domain, that latter notion is the distinguishing property of a function. This is why Name^{-1} is not a function: there is more that one arrow coming from at least some, if not all, of name.

We can also talk about the inverse mapping. Pictorially, the inverse mapping is just reversing the direction of the arrows. Can we say anything about the inverse mapping, knowing the properties of the original mapping? Well, the domain of a mapping is the range of the inverse mapping, similarly the range of the mapping is the domain of the inverse mapping. However, it is convenient to restrict the domain of the mapping, and

make it a function. If so, then surjectivity of the mapping guarantees the inverse mapping has at least one arrow coming from every point in its domain (the range of the original mapping). If the mapping is injective, then in the inverse mapping if it has an arrow coming from a point, then there is only one arrow. If the original mapping is a function, then its inverse is a function if and only if the function is surjective and injective. These are easily verified using a pencil and some paper and drawing some mappings.

A function is said to be bijective if it is surjective and injective. In the following, we will use bijective functions to reach some rather surprising conclusions about the number of elements in the set described in the Chapter on Numbers. $\mathbb{N}$ and $\mathbb{Z}$ each have an infinite number of elements, but is there a way to show that there is more elements in $\mathbb{Z}$ than $\mathbb{N}$? This seems reasonable since it seems there are double the amount of $\mathbb{Z}$ than $\mathbb{N}$. In fact, $\mathbb{Z} = -\mathbb{N} \cup 0 \cup \mathbb{N}$, so this seems ever more likely. If I had a some sheep on my farm, and I buy more sheep then clearly I have more sheep than before. By the same logic, there are more Integers than Natural numbers.

It turns out, I think rather surprisingly given the above argument, that there are exactly the same number of elements is $\mathbb{Z}$ as $\mathbb{N}$. To see this we should talk about some counting principles. How can I determine if I have more pennies than nickels? I should simply count how many pennies I have, then count how many nickels I have, then compare the two numbers and see which is bigger. Clearly the bigger number corresponds to the coin of which I have more of. Alternately, I could throw all my pennies onto a table alongside a pile of all my nickels. Then, I could grab one penny and one nickel and put those in container. Continue this pairing until one of the sets of coin is empty. There are two cases that arise: 1) Both set of coins are now empty, or 2) one of the set is still non-empty. In the first case, there would be exactly the same number of pennies and nickels. In the latter case, there are more elements in the remaining set thus it is bigger than the other. To be concrete, imagine I continued the procedure of pairing nickels and pennies and I ran out of pennies. This mean that there are more nickels than pennies.

Lets call the first procedure: Counting, while the second procedure: Mapping. Notice that in describing the Mapping procedure, we paired every penny with exactly one nickel. This is a function, actually is a bijection! I hope the naming convention is now clearer. Which procedure is better? Well, if the sets considered are finite then each procedure will produce the exact same result. The problem becomes what if the sets are infinite? In this case, the counting procedure says both of the sets have infinite elements, thus both have

the same number of elements. However, the situation is not so easy. It turns out that the Counting procedure can not apply to the infinite sets, and we must resort to the Mapping procedure.

As an example consider Hilbert's Hotel. This is an imaginary hotel that has an infinite number of rooms, and assume that the hotel has no-vacancies: every room is occupied. Hilbert is an savvy business man, and realizes that even though the hotel is fully occupied, he could still accommodate new customers. When a new customer enters the lobby, Hilbert tells everyone in the hotel to move down a room, i.e. if you are in room 6, move your stuff to room 7. So, room 1 is now empty and can accommodate the new customer. With every new customer that enters the lobby, Hilbert can re-perform the move and accommodate the new customer. Although this may be a little painful for the old customers, they do not really care as long as they are not just walking forever; besides, the novelty of staying in an infinite hotel should be enough to hold back their frustrations.

This is a weird hotel, but the weirdness has not even begun. Imagine a tour bus of infinite capacity; they plan on staying at the Infinite Hotel. When the driver calls Hilbert to make reservations for the bus, Hilbert is reluctant to turn down their business, yet he is worried he can not fit everyone. The procedure Hilbert was using earlier would not work, since none of the hotel guest would ever be in a room. Every time they move down a room, to accommodate another person, they would have to keep moving to accommodate the next person coming; never settling down in a room. The customers do not mind moving down a room, but they would mind walking forever and next settling in a room. Hilbert knew that this approach wouldn't work; he would lose all his customers, but he did not want to lose the business of the Infinite bus. What is he to do?

Finally, Hilbert comes up with a plan to accommodate everyone, and yet avoid the problem of the customers just walking forever down to their room. He tells everyone in the hotel already to look at their room number, multiply it by 2, and move to that room. Do you see what happened? Every Odd numbered room is now empty, all the old customers are in the even rooms. Hilbert the directs all the riders of the bus to the even rooms. Hilbert was very fond of his idea: the old customers only had to make one change of room, and he accommodated everyone in the bus.

The next day, learning from the Bus driver of this amazing Infinite Hotel, two more Infinite buses arrive at the hotel. Hilbert knows exactly what to do to accommodate all the new customers. He announces to the hotel guest that they must move to the room

which is three times their current room number. This moves every current hotel guest to all room number which are multiples of three, leaving all room number that have a remainder of 1 or 2 when dividing by 3; equivalently the remaining rooms number are either $n \equiv 1 \pmod 3$ or $n \equiv 2 \pmod 3$. Hilbert then assigns every one in Bus 1 to those rooms which admit a remainder of 1 $\pmod 3$, similarly assigns everyone in Bus 2 to those rooms which admit a remainder of 2 $\pmod 3$. Using this exact procedure he can accommodate any number of buses, if n infinite buses comes, he tells all the hotel guests to go to the new room, determined by n times their current room. Then assigns each bus the number is the residue classes $\pmod n$.

What happens if an infinite number of Infinite buses come to the hotel, can Hilbert accommodate all these new customers? Well lets try the procedure described above. So Hilbert tells all the current guest to move to the the new room, given by $\infty \times r$, where r is their current room number and ∞ denotes infinity. Immediately, we see this is an unfortunate situation, what does $\infty \times r$ mean? The customers all get frustration by this announcement; since either the equation $\infty \times r$ makes no sense to them and so they do not know what room to move to, or we could say $\infty \times r = \infty$ but this weird too, since every guest would move to the same room, the ∞ room. Yet that is also an unsatisfying situation, since there is no room in the Infinite hotel that corresponds to the room number ∞. Either way we are led to the conclusion that Hilbert *can not* accommodate in his Infinite hotel an infinite number of Infinite buses.

Okay, now that we have an intuitive understanding: let's make precise the notions discussed in Hilbert's Infinite Hotel. Since we are dealing with infinities we have to use the Mapping Procedure. Recall, the case of one Infinite bus arriving at the hotel. Hilbert told all the current guest to move to a new room determined by $R(x) = 2x$, where x is their current room. So if I were in room number 48, I would move to room $R(48) = 2(48) = 96$. The domain of this function is the Natural Number, while the range of the function $R(x)$ is all the even numbers. So, $R(x) : \mathbb{N} \to \mathbb{E}$, where $\mathbb{E}$ is the set of even numbers. This function $R(x)$ is bijective. Surjectivity is from the fact that, by definition an Even number is divisible by 2. Clearly, $2x$ is divisible by two for any number x, so $2x$ is always even: therefore $R(x)$ is surjective. Injectivity can be seen intuitively by the fact that no pair of customers that start out in different room end up in the same room. Formally, if $x \neq y$ then $2x \neq 2y$, so $R(x) \neq R(y)$. This shows that $R(x)$ is injective, and thus bijective. By the Mapping Procedure, this shows that there are the Even numbers and the Natural

Numbers have the same Cardinality: that is they have the same number of elements.

Similarly, we can show that there is a bijection of the people on the Infinite bus to the non-occupied Odd Numbered rooms. Number every person in the bus with a unique natural number, then their room number is given by $P(x) = 2x+1$, where x is the natural number they were assigned on the bus. So, $P(x) : \mathbb{N} \to \mathbb{O}$, where $\mathbb{O}$ is the set of Odd numbers. $2x+1$ is always an odd number, to see this divide by 2: $\frac{2x+1}{2} = \frac{2x}{2} + \frac{1}{2} = x + \frac{1}{2}$. If a Integer is even, then division by two would result in an Integer. However, the calculation shows that the division of $2x+1$ by 2 leads to a Rational number, thus $2x+1$ is always Odd. This shows that $P(x)$ is surjective. Injectivity follows from: if $x \neq y$ then $2(x)+1 \neq 2(y)+1$ so $P(x) \neq P(y)$. Thus $P(x)$ is a bijection. Therefore, by the Mapping Procedure: $\mathbb{N}$ and $\mathbb{O}$ have the same cardinality.

Notationally, we denote cardinality by these vertical bars, like absolute value (context should avoid ambiguity): the cardinality of $\mathbb{N}$ is $|\mathbb{N}|$. We have seen that $|\mathbb{N}| = |\mathbb{E}|$ and $|\mathbb{N}| = |\mathbb{O}|$. This also implies $|\mathbb{O}| = |\mathbb{E}|$. The last equality is probably the most intuitively clear, however whether they coincide with intuition or not, the logic of the discussion is infallible. Another notational note, $\mathbb{E}$ can be more succinctly represented by $2\mathbb{N}$. The latter can be interpreted as multiplying every Natural Number by 2; this is clearly the set of Even number. Also $\mathbb{O}$ is better written $2\mathbb{N}+1$. Extending this notation, the multiples of three is denoted by $3\mathbb{N}$, while the numbers whose remainder mod 3 is 1 is denoted $3\mathbb{N}+1$. Similarly $3\mathbb{N}+2$ is the set of number whose remainder mod 3 is 2. Using the same logic as above, just more steps, we can show $|3\mathbb{N}| = |3\mathbb{N}+1| = |3\mathbb{N}+2| = |\mathbb{N}|$. Actually, for any $n \in \mathbb{N}$: $|n\mathbb{N}| = |n\mathbb{N}+1| = |n\mathbb{N}+2| = ... = |n\mathbb{N}+(n-1)| = |\mathbb{N}|$.

This is truly a strange fact, since intuitively it would seem as we divide the Natural numbers into more and more parts, we should not have the same amount of numbers we started with. Imagine a pizza that every slice you cut was the exact same size as the pizza itself. If you sliced your slice you would still have the same amount of pizza left as the original pizza. This implies that every bite you take is the full amount of the pizza, and that no matter how many bites you take you will never finish. You better be hungry!

Now we can begin to classify the cardinalities of the Number systems: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$. Since all the sets are infinite, we need to use the Mapping Procedure.
Claim: $|\mathbb{N}| = |\mathbb{Z}|$. To see this imagine putting every positive number in a even room, and

every negative number in a odd room. Specifically: $P(x) = \begin{cases} 2x & \text{if } x \geq 0 \\ 2x+1 & \text{if } x < 0 \end{cases}$ $P(x)$ is a bijection; which we have actually proved already for each "part" of $P(x)$ when we showed $|\mathbb{N}| = |\mathbb{E}| = |\mathbb{O}|$. So this proves the claim $|\mathbb{N}| = |\mathbb{Z}|$.

I also claim that $|\mathbb{N}| = |\mathbb{Q}|$. I think, this is actually one of the most absurd (unintuitive) facts about numbers. If we think about it, there are so many (an infinite amount actually) rational numbers in between 0 and 1, like $0, 1, \frac{1}{2}, \frac{2}{3}, \frac{1}{5}, \frac{1}{100}, \frac{1}{1000000}, ...$, and yet only two integers. This hold of any interval we consider, take all the rational between -4 and 4. There are still an infinite number of rationals in that interval, yet the only integers are $|-4, -3, -2, -1, 0, 1, 2, 3, 4| = 9$. But, however intuitive or not the Mapping Procedure should definitively either prove or disprove the claim. So we are searching for a Mapping between $\mathbb{Q}$ and $\mathbb{N}$ that is bijective.

A bijective mapping can be described by a line, in that if we arrange $\mathbb{Q}$ in some way such that I could draw a line touching every element of $\mathbb{Q}$ that is a bijection from $\mathbb{N}$ to $\mathbb{Q}$. The mapping is defined by: the first point of $\mathbb{Q}$ touched by the line corresponds to the 1 of the natural numbers, the second point hit by the line corresponds to 2, so on and so forth. So lets arrange $\mathbb{Q}$ in the following manner. The heads of each column and row are each numbered with the natural numbers, then the grid is filled in by writing the rational number where the column head is the numerator and the row head is the denominator.

0	1	2	3	4	...
1	$\frac{1}{1}$	$\frac{2}{1}$	$\frac{3}{1}$	$\frac{4}{1}$	...
2	$\frac{1}{2}$	$\frac{2}{2}$	$\frac{3}{2}$	...	
3	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{3}{3}$	...	
4	$\frac{1}{4}$	$\frac{2}{4}$		$\ddots$	
	$\vdots$	$\vdots$			

The above represents all positive rational numbers, although it does repeat some: the diagonal entries are all one and there are plenty others. But, we can just throw out the repetitions from consideration. If we can find a line that would touch every point on that diagram, then I would have the bijection and thus proving the claim. As a first attempt, we can draw the line that starts at the top left corner, and draw it straight out to the left, or straight down. This however would be not a bijection since it only covers at most one row, or column. If we zig-zag (technical term) around like the figure 11 shows, we do obtain out bijection, and thus proving the claim: $|\mathbb{N}| = |\mathbb{Q}|$. Technically, we only showed that the positive rational number are equal in cardinality to the natural number. Once we have that though, we can map all the positive rational numbers to the even numbers and the negative rational numbers to the odd numbers and thus obtain the total statement.

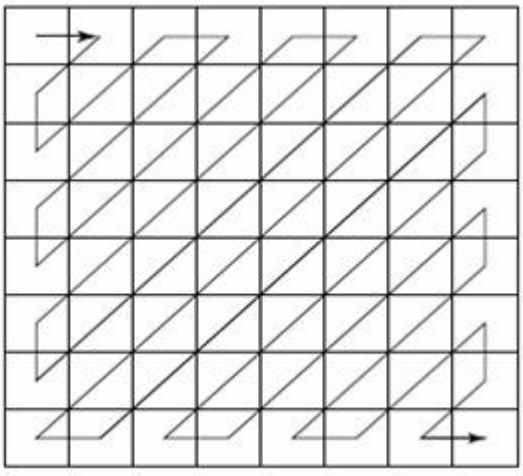

Figure 11: The zig-zag!

Consider the relationship between $|\mathbb{R}|$ and $|\mathbb{N}|$, are they equal or different? By that phrasing you might suspect that in fact they are not equal. To shows this we use a proof by contradiction and what is called Cantor's Diagonalization Argument.[11] Before we prove that, lets shows that $|\mathbb{R}| = |[-\frac{\pi}{2}, \frac{\pi}{2}]|$, where $[-\frac{\pi}{2}, \frac{\pi}{2}]$ is the set of real numbers between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ inclusive. So we need to find a bijection. Consider $P(x) = \tan^{-1}(x)$ and $P(x) : \mathbb{R} \to [-1, 1]$. The domain of $P(x)$ is $\mathbb{R}$ which on the graph is the x-axis. The range of $P(x)$ is the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. We can deduce the injectivity and surjectivity of $P : \mathbb{R} \to [-\frac{\pi}{2}, \frac{\pi}{2}]$ from its graph.

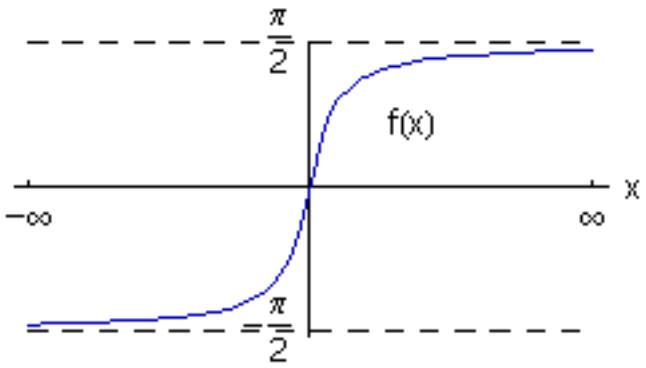

Figure 12:

Take a point on the graph, draw a vertical line at that point. The vertical line will intersect the x-axis at some point, x. This line could also intersect another point of the graph, for instance in a circle. But it is not necessarily true that the vertical line intersects the graph more than at the point. Suppose that for *every point* of the graph, the vertical line at that point intersect only the x-axis, then the function is injective. The same could be said about the horizontal line and the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ on the y-axis, except the function is then surjective.

So, $|\mathbb{R}| = |[-\frac{\pi}{2}, \frac{\pi}{2}]|$; which is also unintuitive, although perhaps not that surprising anymore. In fact, $|\mathbb{R}| = |[0,1]| = |[a,b]|$ for any $a, b \in \mathbb{R}$. The proofs are similar to the above, with different functions. For instance, to prove $|[0,1]| = |[a,b]|$ Use the function $f(x) = (b-a)x + a$, so that $f(0) = a$ and $f(1) = (b-a) + a = b$. The graph of $f(x)$ is a line with slope $(b-a)$ and y-intercept at a. The bijective of f follows immediately.

Claim: $|[0,1]| \neq |\mathbb{N}|$, actually $|[0,1]| > |\mathbb{N}|$. This would imply that $|\mathbb{R}| > |\mathbb{N}|$.

Proof:[11]

To reach a contradiction, assume that $|[0,1]| = |\mathbb{N}|$.
By definition, this implies there exist a bijection $P(x) : \mathbb{N} \to [0,1]$
The set $[0,1]$ is the set of all infinite decimal expansions, whether terminating in a pattern or not.
Since $P(x)$ is a bijection, we can list all the decimal numbers in a list, were the first item in the list is the decimal given by $P(1)$, the second item is given by $P(2)$, so on.

Pictorially, we have this diagram. Note in the expression $.a_1a_2a_3...$, the a_i's are digits so elements of $\{0,1,2,3,4,5,6,7,8,9\}$.

$$\begin{array}{ccc} 1 & \mapsto & .a_1a_2a_3a_4\ldots \\ 2 & \mapsto & .b_1b_2b_3b_4\ldots \\ 3 & \mapsto & .c_1c_2c_3c_4\ldots \\ 4 & \mapsto & .d_1d_2d_3d_4\ldots \\ \vdots & & \vdots \end{array}$$

So now we have everything setup. Here comes the Diagonalization Argument. Let $\hat{a_i}$ be *any* digit that is not a_i; so if $a_1 = 1$ then $\hat{a_1}$ is any digit but one, so $\hat{a_1} \in 0,2,3,4,5,6,7,8,9$. Construct the number, $\alpha = .\hat{a_1}\hat{b_2}\hat{c_3}\hat{d_4}\ldots$. Do you see the where the name comes from? Now, two decimals are equal if everyone of the digits of the decimal expansion are equal. Obviously, $\alpha \in [0,1]$. α is a decimal. But, α does not equal **any** of the decimal numbers in that infinite list, since it differs in at least one digit from every decimal in the list, by construction. We have reached the contradiction! We assumed that every single decimal was listed somewhere in that list, yet we constructed a number which is clearly a decimal but also is clearly not in that list! This explicitly contradicts the assumption that there existed a bijection $P(x)$. This implies there are more numbers in $[0,1]$ than $\mathbb{N}$. Thus proving $|[0,1]| > |\mathbb{N}|$.

A set of cardinality equal to $\mathbb{N}$ is called *countable*, while a set of cardinality larger than $\mathbb{N}$ like $\mathbb{R}$ is called *uncountable*.

Circles and Lines

There is another interesting question which can be answered in this light. What has more points the circle or the line? The circle has exactly one more point than the line! To see this: pick a point (the origin), draw a horizontal straight line (the axis). Then, draw a circle that intersects the axis exactly at the origin. There is a unique point that lies diametrically opposite the origin; that is, on the opposite side of the diameter containing the origin. This point will be call the pole. Now, draw the unique line through the Pole and a point on the axis; it will intersect the circle at exactly one point. This is a mapping from the line to the circle. For every point on the line, there corresponds a *unique* point

on the circle. This is an injective function. However, it is not surjective: there is no point on the line that maps to the pole on the circle. Therefore there is exactly one more point on the circle than the line, namely the pole. This point is called the "point at infinity".

Consider the inverse mapping from the circle to the line, such that the unique line through the pole and a point on the circle maps to the intersection of the axis and the line. Trying to see where the pole gets mapped to leads to drawing a line through the pole and itself. There are an infinite number of lines that go through one point. In an effort to maintain injectivity, we can eliminate any line that intersect the circle. That leaves exactly one line: the one parallel to the axis intersecting the circle at the pole. But, that means the line and the axis do not intersect, and so the definition of the map is undefined. Therefore, this inverse mapping from the circle to the line can not be injective.

This is just another "proof" for the above statement, but highlights a new test for cardinality: $|A| > |B|$ if there does not exist an injective mapping from $A \to B$. If we think about finite sets, this makes perfect sense. Consider a house 10 rooms; another with 4 rooms. Let M be a mapping from the 10-house to the 4-house. Try to make M injective; that is, try to draw an arrow from *every* room in the 10-house to some room in the 4-house, but there can only be one arrow pointing to each room in the 4-house. Impossible!

We could extend the procedure for the above mapping (called stereographic projection), so that is goes from the plane to the surface of the sphere. This is how stereographic maps are drawn. The pole usually (in this hemisphere) is the North Pole: this is why Greenland looks twice as big as Africa. A little distance close to the pole maps to a large distance on the map. This also implies that there is exactly one more point on the surface of the sphere than on the plane, which is also true for higher dimensional analogs of the plane and sphere.

A interesting unification of the concepts of a line and a circle uses the stereographic projection from the plane to the surface of the sphere. A circle drawn on the plane will map to a circle on the sphere, this circle will not intersect the north pole. A line drawn on the plane will map to a circle that goes through the north pole on the sphere. So, a line can be considered a circle of infinite radius. What would other curves map to under this projection?

Can you see why the (infinite) cylinder has exactly one more line than the plane? Hint: imagine stacking the picture for the mapping between the line and the circle in such a way

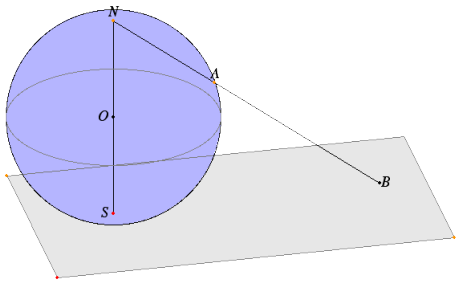

Figure 13:

that the circles line up to make a cylinder and the lines join to make a plane. In that same light, verify the result that a sphere has one more point that the plane by rotating the picture for the make mapping between the line and the circle about the diameter through the pole and the origin; the lines will sweep out a plane and the circle will sweep out the surface of the sphere.

Claim: $|\mathbb{C}| = |\mathbb{R}|$. The proof can be very similar to that of showing $|\mathbb{N}| = |\mathbb{Q}|$, by forming that table and zig- zagging around. We saw that $\mathbb{C}$ is a 2-dimensional vectorspace over $\mathbb{R}$, which means $\mathbb{C}$ can be visualized as a plane. The fact that $\mathbb{C}$ and $\mathbb{R}$ have the same cardinality means that there exists a bijection from $\mathbb{R} \to \mathbb{C}$. This also implies that there are the same number of points on the line and the plane. Also, that I can draw a continuous line on the plane and cover the entire plane.

Sequences

Consider a 100 foot race between a tortoise and Achilles. Suppose Achilles starts 10 feet behind the tortoise, and that Achilles is 10 times faster than the tortoise. By the time, Achilles has run the 10 foot difference in starting position, the tortoise has moved a foot. When Achilles has run that extra foot, the tortoise is still $\frac{1}{10}$ a foot ahead. When Achilles covers the extra $\frac{1}{10}$th of a foot, the tortoise has advanced another $\frac{1}{100}$ of a foot. This keeps going and so Achilles never passes the tortoise. There is clearly something wrong in this argument, but what?

This paradox is known as Zeno's Paradox of Motion.[3, 9] Another one of his paradoxes goes like this: in order to get somewhere I must get half-way first. Then in order to get

half-way, I need to get half-way there so a quarter of the way, and so on. The paradox is then how do you travel an infinite number of points in a finite amount time. This is like walking half-way to the door, then from there walk half-way again. Continue to do so and you will never reach the door, only approach it.

To resolve these issues, we need to consider a mathematical object called a sequence. Formally, a sequence is a function from the $\mathbb{N} \to \mathbb{R}$. Although, the range of a sequence can be any complete field or vectorspace like $\mathbb{C}, \mathbb{R}^n$, or $\mathbb{C}^n$. Notionally, $(a_i) = (a_0, a_1, a_2, a_3, ...)$ is a sequence. We can picture a sequence by placing a point in $\mathbb{R}$ (or whatever range) for every index of the sequence. Connecting to the tortoise and Achilles, form the sequence of points of how far the tortoise is ahead $(10, 1, \frac{1}{10}, \frac{1}{100}, \frac{1}{1000}, ...)$. In the second example the sequence is $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, ...)$. We could write these more succinctly by exploiting the indexing like so: $(a_i) = (10\,\frac{1}{10^i})$, $(b_i) = (\frac{1}{2^i})$, respectively.

These sequence seem like they are approaching zero; but not all sequences have to approach zero. For example, $(1, 1, 1, 1, ...)$ or $(1, 2, 3, 4, 5, ...)$. The latter example shows that a sequence may not even approach something. This idea is expressed in the language of limits, convergence and divergence. Intuitively, the limit of a sequence is the number which it approaches. If it does in fact approach a number, then it is said to converge; otherwise, it diverges. The sequences $(a_i), (b_i)$ defined above are convergent; the limit is zero. $(1, 1, 1, ...)$ converges to 1, while $(1, 2, 3, 4, ...)$ diverges. Notationally, we express the ideas of limit and convergence like: $\lim_{x\to\infty} a_n = a$ or simply, $(a_n) \to a$. In the end though, if we can learn to understand sequences that approach zero that is enough. Since, if $(x_n) \to x$ then $(x_n - x) \to 0$.

The precise notion of convergence and limits was developed by Cauchy and Weierstrass.[7] The definition goes like this: (a_n) converges to a if

$$(\forall \epsilon > 0)(\exists N \in \mathbb{N}) \text{ such that } d(a_n, a) < \varepsilon \ (\forall n > N)$$

The function $d(a_n, a)$ represents the distance from a_n to a. Lets break the definition up into two parts to analyze what is going on. First pick an $\epsilon > 0$, say 1, then the second part of the definition becomes $(\exists N \in \mathbb{N})$ such that $d(a_n, a) < 1 \ \forall n > N$. This says that after a certain index, every point of the sequence is closer to the limit point, a, than a distance one; equivalently, after a certain index every point of the sequence lies within the circle of radius 1 centered at the limit point.

Lets examine this pictorially: consider a table, a bunch of containers, and a bag of marbles. A sequence is then represented by infinite succession of marbles each thrown onto the table landing at a particular point; indexed by the order in which the marbles are thrown. In practice, we can not throw an infinite number of marbles in the finite time of our lives; however, lets imagine we can.

There are three distinct cases: unbounded, bounded, and convergent sequences. The first means that the after a certain index, the marbles are not even landing on the table. Actually, this would mean that there is no index such that after that index all the points of the sequence stay within a circle of some radius. Thus in the physical world, an unbounded sequence would be those that after a certain index would land outside even the universe!! In practice, this would be impossible; but that does not preclude the fact that unbounded sequences exits in $\mathbb{R}$. The sequence $(1, 2, 3, 4, 5, ...)$ is an unbounded sequence, thus it does not converge to a point. One can think of this as the sequence converging to ∞, but $\infty \notin \mathbb{R}$, so we say the limit does not exist.

Bounded sequences are those that after a certain index land entirely on the table, or some other finite area. By the above discussion of unbounded sequences, it follows that any sequence in the real world is bounded; by the universe itself: any marble I throw will land somewhere in the universe! Does it necessarily follow that a bounded sequence is a convergent sequence? No, but it does follow that a bounded sequence must admit a convergent subsequence. This result is know as Bolzano-Weierstrass Theorem.[11] To see the Bolzano-Weierstrass Theorem, consider the bounded sequence $(a_n) = (-1, 1, -1, 1, -1, ..., (-1)^n, ...)$. It is bounded, since the circle centered at the origin with radius 2 surely contains the points $-1, 1$. It does not converge to a limit point, since it just bounces back and forth from -1 to 1. However, we can extract two subsequences each which converge to -1 and 1, respectively. The proper definition of a subsequence is to remove points from the original sequence, leaving the order of the remaining elements the same. So then $(b_n) = (1, 1, 1, ...)$ and $(c_n) = (-1, -1, -1, ...)$ are two subsequence of (a_n), although these are not the only ones. We denote that like $(b_n) = (a_{2n})$, likewise $(c_n) = (a_{2n+1})$; note (a_{n_k}) means that it is a subsequence of (a_n) and the last equality is the definition of the subsequence. $(b_n) \to 1$, and $(c_n) \to -1$.

Intuitively, a sequence bounded, by say the table, has to stay on the table for an infinite number of points. There are only so many places the sequence can go. Eventually it must land on a point twice, and then three times, and ad infinitum. This point is

then the limit of some subsequence of the original sequence. This shows we can think of bounded sequences as those sequences that "converge" to multiple points. Consider the sequence of points of your life indexed every second. This sequence will be all the place on earth that you have gone. This sequence is clearly bounded: by the surface of the Earth (or again, the universe); thus there must be some convergent subsequences. The limits of these sub-sequences are exactly those places you frequent most: your house, school, work, etc. If you imagine plotting the sequence of your life on the surface of the Earth, then you will find that some areas have a lot more points than others: those are the limit points of the subsequence. Again, in practice this wont really work, since sequences are infinite in natural, while life is necessarily finite; but I hope the analogy helps.

Lets go back to convergent sequences. In order to discuss this we need to put a grid on the table. This amounts to defining a metric on the table space; a metric space is a space such that the distance between every pair of points is defined and given by the function $d(x, y)$. One can use the same procedure described in the section on Bases and Vectorspaces. So we create a grid of rectangles on the table. We could also use circles, but that makes the picture a harder. The definition of the metric is what determines whether circles or rectangles or some other shape is used in constructing the grid. However, the metric of rectangles and the metric of circles turn out to be the same, with respects to convergence anyway.

Recall the definition of convergence, which we split into parts. Particularly, the part: $(\exists N \in \mathbb{N})$ such that $d(a_n, a) < 1 \ \forall n > N$. Suppose the table grid is constructed with squares of side length one. This part of the definition of convergence means that after a certain index, all the marbles I throw must land in the SAME square. Specifically, I can throw the first $N - 1$ marbles anywhere I want, but after the Nth marble *all of them must* fall into the same square. This partial definition of convergence is not enough to distinguish between bounded sequences and convergence. Take the sequence $(a_n) = (-1, 1, -1, ...)$ which we know is bounded yet does not converge. Consider the table with a grid of squares of side length 4. Then the sequence (a_n) are thrown into the square at the origin, for all indexes.

The extra piece needed in the full definition of convergence is $(\forall \epsilon > 0)$. Suppose that $a < \varepsilon$ for all $\varepsilon > 0$, what does that imply about a? Well, let's pick an $\varepsilon > 0$, say 1, then $a < 1$. But this happens for all ε that are greater than zero; so pick $\varepsilon = \frac{1}{2}$, then $a < \frac{1}{2}$. Let $(\varepsilon_n) = (\frac{1}{n})_{i=0}^{\infty}$, then each element of this sequence is positive and thus by the

assumptions $a < \varepsilon_n$ for all n. So $a < \frac{1}{1000}$, and $a < \frac{1}{100000}$, etc. This implies that $a = 0$. This is the key step in the proper definition of convergence.

The table is setup with a square grid of length ε. The question now becomes is there an index, after which all the marbles fall into the same square. This is same as we consider above with the square length of 1, now instead of one we use ε. The question of convergence is then, does there exist such an index, for all $\varepsilon > 0$. Pictorially, this is the same as repeating the above procedure except now with a smaller ε trying to find that special index; then continue shrinking the ε so that it approaches zero. With this definition in mind, we can see that $(-1, 1, -1, 1, ...)$ does not converge to limit. When the table grid is setup with the square length 4, there certainly exist an index after which all points of the sequence fall into the same square. Yet, when the table grid is setup with square length of 1, then there exists no such index: for half of the marbles are falling in the square around 1, and the other half are falling in the square around -1.

Consider the subsequence $(b_n) = (1, 1, 1, 1, ...)$, it converges to 1. Lets show this using the definition. First lets assign a metric on the space, that is $d(x, y) = \max\{x, y\}$ and $d(x, y) = 0$ whenever $x = y$. This is the metric that setups up a grid of squares. To test for convergence to the limit point 1, we need to look at $d(b_n, 1)$. This becomes $d(1, 1) = 0$ for all $n \in \mathbb{N}$. So this shows that $(b_n) \to 1$: pictorially, this can be seen by imagine a sequence of shrinking squares around 1, then for all indexes and for all squares *all* the points of the sequence fall in the given squares.

A more interesting convergent sequences is given by $(a_n) = (\frac{1}{n})_{n=0}^{\infty}$. Lets show it converges to zero. Examine the $d(a_n, 0) = \max a_n, 0 = a_n$. The last equality follows since $\frac{1}{n} > 0$, $\forall n \in \mathbb{N}$. So the definition says: $(\forall \varepsilon > 0)(\exists N \in \mathbb{N})$ such that $d(a_n, 0) < \varepsilon$, $\forall n > N$. So given an ε we need to explicitly find N. So $d(a_n, 0) = a_n$, thus we need $a_n < \varepsilon$. But $a_n = \frac{1}{n}$ thus $\frac{1}{n} < \varepsilon$ implies $\frac{1}{\varepsilon} < n$. So, if $N = \frac{1}{\varepsilon}$ then $\varepsilon = \frac{1}{N}$. Also, if $n > N$ then $\frac{1}{n} < \frac{1}{N}$. These are all the pieces we need to show convergence, now to put it all together: Given an $\varepsilon > 0$, and $N = \frac{1}{\varepsilon}$, this implies that $d(a_n, 0) = d(\frac{1}{n}, 0) = \frac{1}{n}$. By the conditions observed before $\frac{1}{n} < \frac{1}{N} < \varepsilon$, $\forall n > N$. That shows $(a_n) \to 0$, or $\lim_{n\to\infty} \frac{1}{n} = 0$. Intuitively, convergence means that given an $\varepsilon > 0$ I can explicitly find that N that makes the definition always true. In effect, the convergence of $(\frac{1}{n})$ is given by the equation $N = \frac{1}{\varepsilon}$. For the sequence $(\frac{1}{n^2})$, can you verify the convergence is given by the equation $N = \frac{1}{\sqrt{\varepsilon}}$? What is the limit?

In order to connection sequences to Zeno's Paradox, we need to look at series. It

turns out that the theory of series and the theory of sequences are exactly the same, we just have to look at series in the right light. A series is an infinite sum denotes $\sum_{n=0}^{\infty} a_n = a_0 + a_1 + a_2 + ...$ A finite sum, denoted $\sum_{n=0}^{M} a_n = a_0 + a_1 + a_2 + ... + a_m$, always converges to a number. That is easily seen, just add up all the numbers. We have seen that the Real numbers are closed with respect to addition, thus the sum of any two real number is a real number and therefore, finite sum must converge! The infinite sums, as you may have suspected, may converge or diverge. Take the series $\sum_{n=0}^{\infty} 1 = 1+1+1+....$ It clearly diverges. What about the series $\sum_{n=0}^{\infty}(-1)^n = 1+-1+1+-1+...$? It seems like it converges to zero, if we place some parenthesis like so $(1+-1)+(1+-1)+(1+-1)+... = 0+0+0+... = 0$. That seems alright, except if we place parenthesis around in a different way: $(1+-1+1)+(1+-1+1)+... = 1+1+1+...$ which clearly diverges, but also is a different answer than before. Notice that I implicitly used commutativity by rearranging the terms.

How can we efficiently describe the convergence of series? By forming a sequence of Partial sums, and then using the definition of convergence for sequences. Consider the convergence of the series $\sum_{i=0}^{\infty} a_i$. Let S_n be the sequence of Partial sums formed like this: $(S_n) = (\sum_{i=0}^{n} a_n)$. Then the series converges if and only if the sequence of Partial sums converges. As an example, consider the series $\sum_{i=0}^{\infty} 1$. The sequence of partial sums is $(\sum_{i=0}^{n} 1) = (\sum_{i=0}^{1} 1, \sum_{i=0}^{2} 1, \sum_{i=0}^{3} 1, ...) = (1, 1+1, 1+1+1, 1+1+1+1, ...) = (1, 2, 3, 4, ...)$. The sequence of partial sums does not converge, thus the series does not converge: it diverges. Another sequence that diverges, called the harmonic series if $\sum_{i=0}^{\infty} \frac{1}{n} = 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+...$ Can you show it diverges?

In Zeno's paradox of the tortoise and Achilles, we can obtain the series of the distance Achilles travels. First form the sequence of point $(a_i) = (10, 1, \frac{1}{10}, \frac{1}{100}, ..., 10\frac{1}{10^i}, ...)$. Then the distance Achilles travels is $\sum_{i=0}^{\infty} a_i = 10 + 1 + \frac{1}{10} + \frac{1}{100} + ...$ To test convergence

of this series, form the sequence of partial sums (S_n). To make that easier, notice that $\sum_{i=0}^{\infty} a_n = \sum_{i=0}^{\infty} 10\frac{1}{10^i}$ just by plugging in the "general term" of the sequence. Also, we can factor the 10 out, since it is a part of every summand. Then, $\sum_{i=0}^{\infty} 10\frac{1}{10^i} = 10\sum_{i=0}^{\infty} \frac{1}{10^i}$. So we only have to check the convergence of $\sum_{i=0}^{\infty} \frac{1}{10^i}$, since once we find that limit point just multiply by 10 to get the original answer.

There is another sequence associated with the series $\sum_{i=0}^{\infty} \frac{1}{10^i}$, namely $(a_i) = (\frac{1}{10^i})$. Expanding this sequence $(1, \frac{1}{10}, (\frac{1}{10})(\frac{1}{10}) = \frac{1}{100}, \frac{1}{1000}, ..., \frac{1}{10^i}, ...)$, we see to obtain the next element in the sequence multiply the current term by $\frac{1}{10}$. This is called a geometric sequence. As an aside, an arithmetic sequence is one where the next term is the current term plus a constant; these are only interesting as finite sequences, the infinite series always diverge. The series $\sum_{i=0}^{\infty} a_i$ is called a geometric series, if the sequence (a_i) is geometric: thus the series $\sum_{i=0}^{\infty} \frac{1}{10^i}$ is geometric. Any geometric sequence is of the form $(a_n) = (\alpha r^n)$, by definition. So all geometric series have the form the series $\sum_{i=0}^{\infty} \alpha r^i$. The α is not really important, since it can be factored out. These series have a particularly nice form; the question of convergence is completely classified by the equation $|r| < 1$. If $|r| < 1$ then the series $\sum_{i=0}^{\infty} \alpha r^i = \alpha\frac{1}{1-r}$; otherwise, it diverges.

Thus, $\sum_{i=0}^{\infty} \frac{1}{10^i} = \frac{1}{1-\frac{1}{10}} = \frac{1}{\frac{9}{10}} = \frac{10}{9}$. Likewise, the series considered in the other paradox: walking repeatedly half-way to the door, is $\sum_{i=0}^{\infty} \frac{1}{2^i} = \frac{1}{1-\frac{1}{2}} = \frac{1}{\frac{1}{2}} = 2$. This answer needs to be interpreted correctly in order to see the answer to the paradox. Look at the first term in the sequence of the terms in the series, it starts at 1, then $\frac{1}{2}$, etc. In the paradox, our first step was half-way to the door, the next was a quarter of the way. We need to just multiply the series by $\frac{1}{2}$ so that the first term in the sequence is $\frac{1}{2}$. Therefore, the actually series considered by this paradox is given by $\sum_{i=0}^{\infty} \frac{1}{2}\frac{1}{2^i} = (\frac{1}{2})2 = 1$.

This means that in the limit as the number of steps approaches ∞ we finally reach the door.

How does this language of sequences and series resolve the paradox presented? Reconsider, the tortoise and Achilles. The main issue was how can Achilles catch the tortoise if in order to catch him he must make up the difference, but in that time the tortoise would have moved another distance, so Achilles needs to make up that distance, ad infinitum. The above discussion showed that Achilles would get there in an "infinite" number of step; but then how can an infinite task be accomplished in a finite amount of time? The answer also comes from geometric series. We have been consider the distance needed to be covered, but when can also consider the duration of time! The relation of speed relates time and distance: $r = \frac{d}{t}$, which implies $t = \frac{d}{r}$. This means that if the sequence of distance is (d_n), then the sequence of durations is $(t_n) = (\frac{d_n}{r}) = \frac{1}{r}(d_n)$. Thus, if (d_n) is geometric so is (t_n), furthermore if $\sum_{i=0}^{\infty} d_n = D$ then $\sum_{i=0}^{\infty} t_n = \sum_{i=0}^{\infty} \frac{1}{r}(d_n) = (\frac{1}{r})D$. In terms of the paradox, this implies that the time needed to run over the "infinite" number of points is finite; so Achilles will catch up to the tortoise in a finite amount of time (depending on his speed), and eventually pass him. This makes our intuition happy.

Cauchy Sequences

In the definition for convergence, we select a particular limit point and then verified with the definition that it indeed converges. In would be nice if there was some criterion for determining whether a sequence converges or not without referencing a limit point. This is the characterization of a Cauchy sequence. If we are in a complete metric space like $\mathbb{R}^n$, the a sequence (a_n) is called Cauchy if $(\forall \varepsilon > 0)(\exists N \in \mathbb{N})$ such that $d(a_n, a_m) < \varepsilon$, for all $n, m > N$. In a Cauchy sequence, after a certain index the terms between arbitrarily close.

This is a particularly useful criterion, since we can check if a sequence is Cauchy or not just by examining the terms of the sequence; there is no reference to the limit point (which is not necessarily not a term of the sequence). It is easy to show that every convergent sequence is a Cauchy sequence. The converse: Every Cauchy sequence is a convergent sequence is only true in Complete Metric spaces. Actually, a space is complete if and only if every Cauchy sequence converges.

Consider the set of bounded sequences in $\mathbb{Q}$. Recall, the example of finding the max-

imum of the set $x \in \mathbb{Q} : x^2 \leq 2$. There is no maximum: we found a sequence of rational that approached $\sqrt{2}$: $(1, 1.4, 1.41, 1.414, ...)$. With the language of convergence, we say this sequence converges to $\sqrt{2}$. With the additional fact that $\sqrt{2} \notin \mathbb{Q}$, we can see that this sequence does not converge to an element of $\mathbb{Q}$; that is, $\mathbb{Q}$ is incomplete.

In fact, any real number is the limit of some sequence of rational numbers. To see this recall the decimal expansions. Say $a = .a_1a_2a_3...$ is an infinite decimal expansion (terminating in a pattern or not), then the sequence:

$$(a_n) = (\, .a_1,\, .a_1a_2,\, .a_1a_2a_3,\, .a_1a_2a_3a_4, ...)$$

gives a sequence of rationals converging to the real number a. To finish the claim the remember $|\mathbb{R}| = |[0, 1]|$. In fact, every one of the sequences considered like that will be a Cauchy sequence. (Why?)

What this discussion outlines is the construction of $\mathbb{R}$ out of $\mathbb{Q}$. Consider the set of all Cauchy sequences in $\mathbb{Q}$. Some of those will converge to elements of $\mathbb{Q}$, but some will not converge to an element of $\mathbb{Q}$. Create a new set R that has all the rational numbers. Append to this set, a new number for each distinct Cauchy sequence that does not converge to a point in $\mathbb{Q}$. Since every Cauchy sequence converges, by definition, R is complete. The set R is the set of Real numbers, that is $R = \mathbb{R}$.

Calculus

Sequences were the study of function from $\mathbb{N}$ to $\mathbb{R}$. We now turn to studies of functions from $\mathbb{R}$ to $\mathbb{R}$. In all generality the Domain and Range of these function can be any complete space into another, but $\mathbb{R}$ will suffice here. In this section we aim to develop a brief understanding of Calculus. In order to do this, we need to define continuity. A calculus is studied over continuous functions. An intuitive picture of Continuity is a curve with no breaks or holes. For example, if a draw a curve on a draw *without* lifting my pencil, then the curve is continuous.

A simple characterization for continuity using sequences is: $f(x)$ is continuous if whenever $(x_n) \to x$ then $f(x_n) \to f(x)$. Thus continuous functions map converging sequences to converging sequences; we say, a continuous function preserves the limit operation. Notationally this means we can interchange the limit operation and the function: if $\lim_{x\to\infty} x_n = x$

then $\lim_{x\to\infty} f(x_n) = f(\lim_{x\to\infty} x_n) = f(x)$. Formally the definition of continuity is: $f(x)$ is continuous if $(\forall \varepsilon > 0)(\exists \delta > 0)$ such that $d(x_n, x) < \delta$ whenever $d(f(x_n), f(x)) < \varepsilon$. These definitions are slightly different but coincide in metric spaces.

Sir Isaac Newton is probably most know for his work in Physics, primarily his equation on Gravity. He was the first to realize that the same force that makes an apple fall to the earth is responsible for the movements of the planets. Also, he had a big impact on the Mathematical world, by starting the development of calculus.[7] Although his version was by no means foundationally sound, he managed to perform calculations. The particular reasons his theory was not foundationally sound will come up later in this section.

In his search for the mechanics of physics, Newton realized that the trajectory of an object could be described by a continuous function. For instance, the path of baseball hit off the end bat could be described by an upside down parabola, which is a quadratic polynomial, which is continuous. So for certain parameters, depending on the hitter and pitcher, the baseball would travel the path described by $p(x) = -ax^2 + bx + c$. The parameter a is related to the force of gravity pushing down on the ball, b is related to the initial velocity of the baseball when it hits the bat, and c is related to the height off the ground where the bat and ball make contact.

It would be nice if there was an operation perform on the function $p(x)$ that would give us the speed of the ball. Speed is the ratio of the distance traveled to the time elapsed: $r = \frac{d}{t}$. For a particular time interval, say one second, we could calculate the average speed the ball travels in the time interval. This is not exactly what we set out to find. We want the speed of the object at a particular time, not its average speed over some time interval. We could take the limit as the time interval goes to zero. This limiting process turns the average speed to the instantaneous speed, and now we have a function such that it tells us the speed of the baseball at any particular time.

This operation is called Differentiation. If a function f describes the trajectory of an object with respect to time, then its derivate, which Newton denoted by f' (said f prime), is a function describing the velocity of the object with respect to time. Since f' is a function, we can take its derivative,f''. This function describes the change in velocity with respects to time: aka the acceleration of the object. Newton used these notions to describe the motion of objects.

At practically the same time, Gottfried Leibniz was developing the same idea under a completely different light.[9, 3] Leibniz starts out with a function and ask is there some

way to calculate the slope of the tangent line of function at a particular point. To find the slope of a line use the definition $m = \frac{\Delta y}{\Delta x}$, where $\Delta y = y_1 - y_2$, and $\Delta x = x_1 - x_2$. At a particular point, to find the slope of the tangent line create a sequence of points on the curve that approach the point. Also, create the sequence of slopes determined by the lines going through the particular point and the corresponding point in the sequence just created. Then the limit of the sequence of slope is the slope of the tangent line.

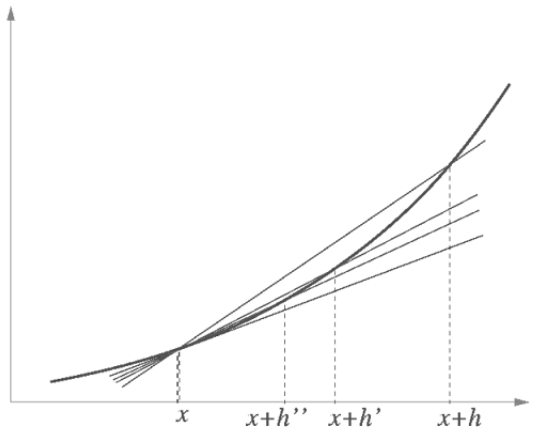

Figure 14: The sequence of secant lines: derivative

Both of these notions of the derivative are the same: loosely speaking, Newton considered the derivative of a function with respect to time, while Leibniz considers the derivative with respect to space. In their description, I use the terms limit and sequence. Neither of these men had this language (nor their techniques) at their disposal. Not until Cauchy and Weierstrass, formally define the algebra of sequence and limits, does this definition reach the rigor and stability it has now. But, both still had an incredible impact on Math. Newton's main impact was to think of the derivative as an operation on a function that returns another function. Leibniz's view was much more general (not just applying to trajectories of objects) but he viewed differentiation as a method for calculating the slope, not as an operation on functions. Both aspects are needed for a sound foundation of calculus.

Both men had the right ideas, by thinking of the limiting process as a dynamic process.

However, math is inherently static, going back to Zeno's Paradox. A solid foundation of this limiting process needed to be described in order for the mechanics of calculus to be fully understood. The definition of Sequences correctly moves the dynamic limiting process into a static process that can be examined in the Math framework. Still, they both knew it worked, but did not know why.[7] That is not a satisfactory position.

The integral is the inverse operation of the derivative. This is known as the Fundamental Theorem of Calculus.[15] In terms of Newton, the integral of the velocity function will a give a function that describes its trajectory; also the integral of the acceleration function is the velocity function. Notice how much this simplifies Newtonian physics, since to have any one of the position, velocity, or acceleration function gives us the rest. In terms of Leibniz, the integral of a function calculates the area under the curve.

To calculate the integral, we need to first approximate the area with the Riemann Sum. Consider a function, and its graph. Then, partition the x-axis into subintervals, and at each endpoint of a subinterval, draw the vertical line to the curve. Finally, connect all the points on the curve: either by a horizontal line creating a bunch of rectangle, or by connecting consecutive terms with a line creating a bunch of trapezoids. Either way does not effect the limiting process. Calculate the area of this given partition; by summing up the areas of each of the rectangles (or trapezoids). Finally, take the limit as the size of the partition on the x-axis approaches zero. This is the area under the curve. Technically, a function is Riemann Integrable if the Upper and Lower Riemann Sums coincide. The Upper Riemann Sum is defined as the rectangles setup up like above on a partition of the x-axis such that the area of the sum of the Rectangles is greater than the area under the curve. To ensure this, take a partition of the x-axis and draw the vertical lines to the curve. Then at each sub-interval, connect the vertical lines (possibly extending one of them) so that the area of the rectangle is greater than the area of the curve at that particular sub-interval. The Lower Riemann sum is defined similarly replacing every instance of "greater than" with "less than". Finally, a function is Riemann Integrable if and only if the limit of the Lower Sum equals the limits of the Upper Sum.[11]

The problem with this formulation of the Integral (and hence the derivative) is that is it not a continuous operator of the space of functions. Consider a sequence of Riemann Integrable function (f_n) such that $(f_n) \to f$. Then is not necessarily true that $\lim_{n\to\infty} \int f_n = \int (\lim_{n\to\infty} f_n) = \int f$. This is the sequence characterization of continuity: the interchange

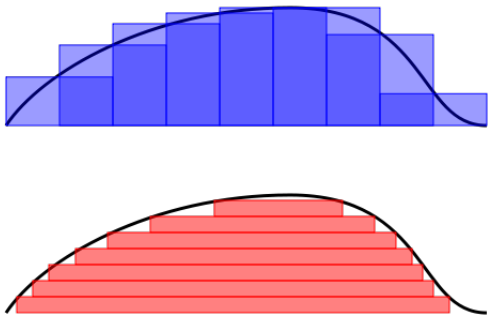

Figure 15: The top figure is the Riemann Integral. The bottom figure is the Lesbegue Integral

of lim and the operator $\int$. The Riemann integral as defined above does not satisfy this property. This shows that the space of Riemann Integrable function is not a complete space.[15]

To fix this, we introduce the Lebesgue Integral. This integral is in fact a continuous operation, and furthermore the space of Lebesgue Integrable function is a complete space.[15] We have seen the the property of completeness is an important property, especially when dealing with abstract notions and spaces. The fact that every Cauchy sequences must converge in a Complete space makes dealing with the Lebesgue Integral much nicer than the Riemann Integral; since to show that the Lebesgue Integral exists amounts to showing that the sequence it determines is a Cauchy sequence. This is much easier than showing the sequence converges: since there is no explicit mention of the limit point (which in this case is a function).

The definition of the Lebesgue Integral depends intimately on the notion of a measure. The measure of a set E describes the volume (length if in 1-D, area if in 2-D, and the higher dimensional analogs). It does so by first considering a cover of the set by other "special" sets. The special sets are those whose volumes can be determined easily, like cubes or spheres. Let's uses cubes. There is a theorem about the structure of sets in $\mathbb{R}^n$: Any open set in $\mathbb{R}^n$ can be written as a countable union of almost disjoint closed cubes. A set of cubes is almost disjoint if the cubes only intersect at the boundary. With this theorem, we can decompose any set in $\mathbb{R}^n$ into a disjoint union of almost disjoint cubes.[15]

The measure of the set is then the sum of the volumes of each individual cube; which are calculable. The proper definition uses the inf operator, and finds the smallest cover of the set in question, and then takes the sum of those cubes. This is just an introduction into Measure Theory; properly, one defines an exterior measure (analogous to the Upper Riemann sum) and an interior measure (lower Riemann sums) and then if they coincide in the limit, the set is measurable, and the measure equals the exterior measure (or the interior: they are the same).

Once the properties of the measure have been established the definition of the Integral follows really easily. The characteristic function of a set E is denoted χ_E. This is a function such that: $\chi_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}$

The integral of a the Characteristic Function of E *is* defined to be the measure of the set E. Thus, $\int \chi_E = m(E)$, where $m(E)$ is the measure of the set E. One important property of the Integral is that it is Linear. This means that $\int f + g = \int f + \int g$, this looks like the distributive property of multiplication over addition, in fact multiplication is a linear operation. Also, $\int \alpha f = \alpha \int f$. If we accept these properties of the integral, we can extend the definition above to include finite sums of the characteristic functions. For instance, $\int \sum_{i=0}^{n} \alpha_n \chi_(E_n) = \sum_{i=0}^{n} \alpha_n \int \chi_(E_n) = \sum_{i=0}^{n} \alpha_n m(E_n)$

Now, we have an Integral that is defined on any finite linear combination of characteristic functions. There is structure theorem developed in Measure Theory about non-negative measurable function: Any non-negative measurable function is the limit of a sequence of finite linear combinations of characteristic functions.[15] Given this structure theorem, we can develop an integral for all non-negative measurable functions. The theorem implies that if f is a non-negative function, then there exists a sequence of finite linear combinations of characteristic functions whose limit is f, i.e. $\exists(f_n) \to f$ such that $f_n = \sum_{i=0}^{n} \alpha_n \int \chi_(E_n)$. We observed that the problem with the Riemann integral was it lacked the property of continuity. Here we see why we need this property. $\int f = \int \lim_{n\to\infty} f_n = \lim_{n\to\infty} \int f_n$. Notice $\int f_n$ was defined above: f_n is a finite linear combination of characteristic functions.

So, we have defined an Integral that works on all non-negative functions. To obtain the general integral for any function (non-negative or not) we use one more fact about functions: any function is the difference of two non-negative functions.[15] To see this consider the function f. There is a "sub-function", f^+ such that it is those part of f that

are positive. Similarly, we define f^- to be those parts of f that are negative, except f^- is the negative of those negative parts, thus positive. This make both f^+ and f^- non-negative functions. Also $f = f^+ - f^-$. The minus is important since it sends the positive parts of f^- back to the negative side. The general Lebesgue Integral is then defined as $\int f = \int f^+ - f^- = \int f^+ - \int f^-$, where both of the latter integrals are of non-negative functions which we defined above.

The Lebesgue integral is completely determined by the measure function. There are certain properties a function must satisfy to be consider a measure. Once this is done it determines an Integral with respect to that measure, as is outline above. This is the abstraction of calculus. A measure space is a abstract space on which there is a measure function, which in turn determines an integral and differentiation operation on the space of function of the measure space.

Topology

A very specific example of a topology we have seen is a metric space. The open sets are the $\{B_r(x)\}$ for all $r > 0 \in R$ and $x \in X$, where $B_r(x) = \{y \in X : d(y, x) < r\}$ A metric space is vectorspace, along with a function that is called the metric, $d(x, y)$. It satisfies the following properties:

Metric Space $(X, d(x, y))$, for all $x, y \in X$ $d(x, y) \geq 0$ $d(x, y) = 0 \Leftrightarrow x = y$ $d(x, y) = d(y, x)$ $d(x, y) \leq d(x, z) + d(z, y)$ for all $z \in X$

A metric is the distance between those points. The first two properties state that the distance between two different point is always positive, and the distance is zero only when the points are equal. The third says the distance from here to there is the same as the distance from there to here. To that I say: if you walk from here to there, then you arrive here not there: so how can you move ever get there when you are always here? The last property is the defining property of a metric. This is known as the triangle inequality: imagine a triangle with points x, y, z. The property states that the sum of two side of the triangle is greater that the third side.

The definition above for convergence is based on a metric; and thus defines convergence on Metric spaces. However, not every space is a Metric Space. However, there are spaces where a notion of length is defined: called the Norm. These space are collectively called Normed Linear spaces. The norm of a vector in the space is the distance from the origin.

That is $\|x\| = d(x, 0)$. That is if the metric induces a norm. There are certain metrics that do not induce norm, most do though. If a metric does satisfy two additional conditions: $d(\lambda x, \lambda y) = \lambda d(x, y)$ and $d(x+t, y+t) = d(x, y)$ then it does induces a norm, given by the above equation.[14] Going the other way, a norm always induces a metric: $d(x, y) = \|x-y\|$. Normed linear spaces are an important area of study in Math. With the additional hypothesis that the space is complete, these spaces are called Banach spaces. In conclusion, not all metric spaces are normed spaces, but all normed spaces are metrics spaces though.

Norm $(X, \|x\|)$, for all $x \in X$ $\|x\| \geq 0$ $\|x\| \Leftrightarrow x = 0$ $\|\alpha x\| = \alpha\|x\|$ $\|x+y\| \leq \|x\|+\|y\|$ for all $y \in X$

There is one more space considered in this light: An inner product space. This is a vectorspace with a inner product denoted (x, y). (Not an order pair.) It satisfies the following properties:

Inner Product Space $(X, (x, y))$, for all $x, y \in X$ $(x, y) \in \mathbb{R}$ or $(x, y) \in \mathbb{C}$ $(x, x) \geq 0$ and $(x, x) = 0 \Leftrightarrow x = 0$ $(x, y) = (y, x)$ or if in $\mathbb{C}$ $(x, y) = \overline{(y, x)}$ $(x + y, z) = (x, z) + (y, z)$ for all $z \in X$

If (x, y) is an inner product, then the norm of a vector x is $\|x\| = \sqrt{(x, x)}$, and the norm induces a metric. Thus every inner product space is a normed space and thus a metric space; however, not every normed space is an inner product space. In order for a norm to induce an inner product the norm must satisfies the Parallelogram identity: $\|x + y\| + \|x - y\| = 2(\|x\| + \|y\|$.[14] If the norm satisfies that identity then the inner product (x, y) is given by the Polarization Identity: in a real vectorspace $(x, y) = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2)$. An inner product space that is complete is called a Hilbert space.

An inner product space defines the angles between its elements. $(x, y) = |x|\,|x|\cos(\alpha)$, where α is the angle between them. These spaces have strong connections to the physical world we live in. $\mathbb{R}^3$ is the mathematical idealization of the 3-dimensional world, $\mathbb{R}^4$ the 4-D world of space-time; both are concrete examples of an Inner Product Space. The

inner product, norm, and metric in this space are

$$\begin{aligned}(x,y) &= ((x_1,x_2,...,x_n),(y_1,y_2,...y_n)) \\ &= x_1y_1 + x_2y_2 + ... + x_ny_n \\ |x| &= \sqrt{(x,x)} = x_1^2 + x_2^2 + ... + x_n^2 \\ d(x,y) &= \sqrt{|x-y|} = \sqrt{|(x_1-y_1,x_2-y_2,...,x_n-y_n)} \\ &= \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + ... + (x_n-y_n)^2}\end{aligned}$$

The last equation is the usual distance formula of Euclidean geometry. The inner product in the space is called the dot product. This inner product along with an integral (and derivative) is the mathematical structure used in introductory physics and mechanics. Newtons second law is that: $\sum F_i = m \cdot a$, where F_i are the vector of forces acting on the object, m is the scalar mass of the object, a is the vector of acceleration. The sum of the forces on an object equals the mass times the acceleration. All the F_i live in the vectorspace $\mathbb{R}^3$ and so the sum of them well-defined. The m is in the scalar field of the vectorspace, and thus the scalar multiplication $m \cdot a$.

Recall, Newton's use of the derivative was to transform the position function to the velocity function, then to the acceleration function. So, $a = x''$, where x is the position function. Hence, $\sum F_i = m \cdot x''$. This is a differential equation and, although greatly simplified, represents the motion of objects in the space we live in, at least in the translational motion (no rotation). Extension of this equation include rotational and vibrational motion; also, adding the term bx' introduces "friction" to the model. Rotation motion needs another equation $\sum \tau_i = I\alpha$: the sum of the torques equals the moment of inertia times the angular acceleration, exactly analogous to the translational equation. Quantum Mechanics describes the physics of the very small through the use of an infinite dimensional complete inner product space like $\mathbb{C}^\infty$.[14]

Anyway, Topology is the general study of spaces, where the notions of convergence are defined. A topology over a set, X is a distinguished subset, T, of the set of all subsets of X, such that any any set in T is called an *open* set. A topological space is the umbrella class of spaces: every one of the space above is a topological space. There does not have to be a notion of distance, length, or angles in a general topological space, the only thing that is defined is convergence. There is an abstract notion of a sequence called a *net*. The

T of open sets in the topology are used to define a neighborhood N of a point $x \in X$: if there exists an open set, $O \in T$ such that $x \in U \subset U$. A net $x_{\alpha\alpha\in I} \to x$ if for any neighborhood N of x, there is a $\beta \in I$ so that $x_\alpha \in N$ if $\alpha \succ B$.[14] This is similar in form to the definition of convergence in a metric space: and in fact, in a metric space they are the same. This definition holds for any topological space. Continuity is characterized in these spaces employing just topological notion by: the inverse image of an open set is an open set.

Sets

We have been using the notion of a set through this paper, yet I have not presented a definition of a set. Let a set be a collection of objects. For instance the set of Even numbers is the set of number divisible by two. The set of people on Earth is the collection of animals on earth that are human. To make the definition more practical, define a set to be the collection of objects each of which satisfies a certain property. Notationally, $A = \{x : P(x)\}$, this is read the set is A is the collection of objects x such that each x satisfies property P. (The colon means such that, some authors use a | instead of :) Notice $P(x)$ does not represent a function, but rather it represents that x has the property P. The set of people is $\{x \in Animals : Human(x)\}$. Human(x) means that x is a human. The even numbers are $\{x \in \mathbb{N} : x \equiv 0 \pmod 2\}$

This intuitive definition of a set was first presented by Cantor. However, this axiomatization does not lead to a consistent theory. The following contradiction is know as Russell's Paradox.[11] Suppose there was barber, B who only shaves everyone in his town who does not shave himself. Consider the set of residents who shaves himself, R. Does the barber shave himself: is $B \in R$ true? If the barber does not shave himself: $B \notin R$, then since the barber must shave anyone who does not shave himself, he must shave himself: $B \in R$. If the barber shaves himself: $B \in R$, and since the barber only shaves people who don't shave themselves, he must not shave himself: $b \notin R$. One of these must be true: $B \in R$ or $B \notin R$, but in either case we reached a contradiction.

Bertrand Russell did not state his paradox with the use of this barber. He instead considered the set of all sets that do not contain themselves, $R = \{X : X \nsubseteq X\}$. The paradox is $R \in R$ or $R \notin R$. The former means that $R \nsubseteq R$ which means that $R \notin R$. The latter case $R \notin R$ implies $R \subseteq R$ which means $R \in R$.

This issue rocked the foundation of Math to its core. Sets are the basic building block of all of math, and Russell exposed the cracks in the foundation. If not fixed, this inconsistency pertubates through all of math, rendering every and all theorems incorrect. Zermelo–Fraenkel Set theory aims to correct this inconsistency, but does so at the expense of a more complicated definition of a set.

Given any two sets A, B, there are certain operations we can perform on them. Let U be the "universe" in which A and B live. $A \cup B$ is the set of elements x, such that $x \in A$ or $x \in B$. In English we use or in the exclusive way: I want that or this (but not both); this is inclusive or: I want that or this or both. The operation $\cup$ is called union. Similarly, we define intersection, $A \cap B = \{x \in U : x \in A \text{ and } x \in B\}$. Also, there is a complement operation. A^C (read A compliment) is the set of objects in U that are not in A, so $A^C = \{x \in U : x \notin A\}$. There is a difference operator on sets: $A \setminus B$ is the set of all elements of A not in B. If we you at the picture of two overlapping circle as model for the sets A, B, then we see that every section of the picture can be uniquely written as one of $\{A, B, A \cup B, A \cap B, A \setminus B, B \setminus A, A \triangle B\}$ and each of their compliments. Can you find all 12 distinct areas created?

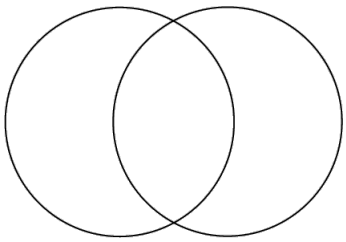

Figure 16:

Actually only two operations need to be defined: complements and intersection. The union and difference operators follow:

$$A \cup B = (A \cap B)^C$$
$$A \setminus B = A \cap B^C$$

These operation define a Boolean Algebra. The follow properties finish the characterization of the above operations

Associativity

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

Commutativity

- $A \cap B = B \cap A$
- $A \cup B = B \cup A$

Distributivity

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

Identity

- $A \cup \varnothing = A$
- $A \cap U = A$

Pseudo-Inverses

- $A \cup A^C = U$
- $A \cap A^C = \varnothing$

Idempotent

- $A \cup A = A$
- $A \cap A = A$

There are a couple more useful properties; these can be verified from the above:

- If for all A, $A \cup B = A$ then $B = \varnothing$
- If for all A, $A \cap B = A$ then $B = U$

Unique Compliments

- If $A \cup B = U$ and $A \cap B = \varnothing$, then $B = A^C$

- $(A^C)^C = A$
- $(\varnothing^C) = U$ and $U^C = \varnothing$

De Morgan's Laws

- $(A \cap B)^C = A \cup B$
- $(A \cup B)^C = A \cap B$

Absorption Laws

- $A \cup (A \cap B) = A$
- $A \cap (A \cup B) = A$
- $A \cup U = U$
- $A \cap \varnothing = \varnothing$

We can see that the algebra of sets is very similar to the ring structure of $\mathbb{Z}$. However, this is not a ring at all. Notice that there is no inverse. Identities exists for each operations: $\varnothing$ is the identity of $\cup$, and U the identity of $\cap$. The inverses for the $\cup$ operation must return a $\varnothing$, but how can we take union of two sets such that the resulting set is empty? The inverse of the $\cap$ operation must return a U, yet how can we take the intersection of two sets and get the Entire sets. The inverse of both operations are therefore not defined: that is why I called them pseudo-inverses.

These in turn define the inclusion relation, $\subseteq$. $A \subseteq B$ if $A \cap B = A$ or $A \cup B = A$, also $B^C \subseteq A^C$. The inclusion relation defines a partial ordering on the set of subsets. A partial ordering is the same as a total ordering (presented above) but without the condition of the total (linear) ordering; that is, not every pair of elements is comparable. In a total ordering it is necessary that for any given pair of elements, a, b, then exactly one of the following is true: $a \geq b$ or $a \leq b$ (or the Trichotomy Law). In a partial ordering, we relax this condition to say at exactly one of the following is true : $a \geq b$, $a \leq b$, $a \parallel b$ (read a is incomparable to b).

To illustrate this, we need to consider one more operation on sets: the power set. This is an operation that returns a set that contains all the subsets of the given set. A subset of a set is any set of elements from the given set. For example, $A = \{1, 2, 3\}$ then the

power set of A, denoted $2^A = \{\{\varnothing\}, \{1\}, \{2\}, \{3\}, \{1,2\}, \{2,3\}, \{1,3\}, \{1,2,3\}$. Note, the whole set $\{A\}$ and the empty set $\{\varnothing\}$ are subsets. One important point here is that $\{A\}$ is not the same as A. The former is a set with the single element A; the later is the set A which contains 3 elements. Is $\{1,2\} \in \{\{1,2,3\}, \{1,3\}, 1, 2\}$?

The notation of Power sets 2^A is suggestive; in the sense that if $|A| = n$ then $|2^A| = 2^n$. Another way to think about this is given a set A with n elements, then for each element it is either in a certain subset or not. If there is only one element in the set $A = \{1\}$, then for each subset of 2^A, the element 1 is either in a subset or not, this leaves only two subsets: $2^A = \{\{\varnothing\}, \{1\}\}$.

With the use of the Power set, we can find a new understand for the relationship between $\mathbb{Q}$ and $\mathbb{R}$ and incidentally find an infinite number of sets, each which has cardinality larger the previous : i.e. there are an infinite number of distinct types of infinity. The theorem goes like this: $|2^A| > |A|$. That is, there is **strictly** more elements in the power set than the in given set. To prove this, recall the Mapping procedure used to discuss Cardinality of sets. In there, I mentioned that $|A| > |B|$ if there does not exist an injective map from A to B. Similarly if we can show that there does not exists a surjective map from B to A, then $|A| > |B|$. Consider the mapping $\mathcal{P}\colon A \to 2^A$. We need to setup up a correspondence: start by mapping every element in A to the subset formed by just that single element, i.e. if $x \in A$ then $\{x\} \in 2^A$ and $\mathcal{P}(x) = \{x\}$ Now, there are still plenty of sets in 2^A that have not been hit by the mapping, but we have run out of elements in A. I can not have more than one arrow coming from any element in A, so I'm done! This mapping can never be surjective. Thus $|2^A| > |A|$.

Okay, but how does this relate $\mathbb{Q}$ to $\mathbb{R}$? It turns out that $2^{\mathbb{Q}} = \mathbb{R}$. That is, the set of real numbers is the set of all subsets of $\mathbb{Q}$. This is quicker proof that $|\mathbb{Q}| < |\mathbb{R}|$. Also it generates a sequence of Sets, consider: $(\mathbb{Q}, 2^{\mathbb{Q}} = \mathbb{R}, 2^{\mathbb{R}}, 2^{(2^{\mathbb{R}})}, ...)$. The cardinality of each set is larger the previous, and each is a distinct "type" of infinity. What is the set $2^{\mathbb{R}}$?

Back to the inclusion relation and partial orderings. The power set of any set A is a partially ordered set with respect to $\subseteq$. Consider the set $\{1,2\}$, then the power set is $\{\{\varnothing\}, \{1\}, \{2\}, \{1,2\}\}$. Then, $\{\varnothing\} \subseteq \{1\} \subseteq \{1,2\}$ and $\{\varnothing\} \subseteq \{2\} \subseteq \{1,2\}$. But $\{1\} \nsubseteq \{2\}$ and $\{2\} \nsubseteq \{1\}$, so $\{1\} \parallel \{2\}$. For any set A, the power set will be partially ordered, actually any Partial ordering is of this form.[16] In a partially ordered set, $(P, \subseteq)$ there exists chains: which are subsets of P such that the partial ordering becomes a total ordering. So, $\{\varnothing\} \subseteq \{1\} \subseteq \{1,2\}$ is a chain. We see a in chain every pair of elements is

comparable. An upper bound for a subset of a partially ordered set is an element of the set such that every member of the given set is "smaller" than the upper bound.

There is an important Theorem concerning partial orderings called Zorn's Lemma.[11] Zorn's Lemma states: that in a partially ordered set, if every chain has an upper bound then the partially ordered set contains a Maximal element. In the example, both chains have $\{1, 2\}$ as an upper bound, thus there exists a maximal element. Zorn's Lemma is not a "constructive" lemma, in that it only guarantees the existence of such a maximal element, but does not give a procedure for finding it. In the example the maximal element is $\{1, 2\}$.

Zorn's Lemma ties to the controversial Axiom of Choice. This is one of the axioms in Zermelo–Fraenkel set theory. It states, that given a partition of the set, we can create a set with a representative member from each partition. To fully understand this, we need to consider the definition of an Equivalence Relation.

ØEquivalence Relation $\forall a, b, c$

- Reflexive: $a = a$
- Symmetric: If $a = b$ then $b = a$
- Transitive: If $a = b$, and $b = c$, then $a = c$.

Any relation that satisfies the above properties is an equivalence relation. We have seen some other equivalent relations through out the above, like the equivalence relation on fractions $(a, b) = \frac{a}{b} = \frac{c}{d} = (c, d)$ if and only if $ad = bc$. Also we have seen the component-wise equality of vectors: $(v_1, v_2, v_3, ...) = (w_1, w_2, w_3, ...)$ if and only if $v_1 = w_1$, $v_2 = w_2$, $v_3 = w_3$, ... and so on. The equivalence relation of modular arithmetic: congruence, is also an equivalence relation. The above properties as easily shown to be satisfied under all of these relations. The power of an equivalence relation is that it necessarily partitions the set into Equivalence Class. The converse is also true, given any partition of the set there exist a corresponding equivalence relation.[5] This is the abstract notion afforded from the regular equality.

The equivalence classes setup up by the equivalence relation are all those elements of the set which are "equal" to each other. For instance, the equivalence relation on fraction partitions the Rationals into equivalence classes: like $\{\frac{1}{2}, \frac{2}{4}, \frac{3}{6}, \frac{4}{8}, ...\}$. In a abstract landscape, this where Zorn's Lemma and the Axiom of Choice come in. The Axiom of

Choice says we can create a set with representative member from each equivalence class. This is the proper way to think of $\mathbb{Q}$, as the set alluded to in the Axiom of Choice. Zorn's lemma is more powerful in that it asserts (if the premises are satisfied) that there exists a maximal element. We could apply Zorn's lemma to each equivalence class to and pick out the Maximal element (note, we could also pick the minimal element if we considered the dual ordering). Specifically, the axiom of choice is all that is needed to create $\mathbb{Q}$ under this equivalence relation. Zorn's lemma is normally used in other more abstract contexts.

The Axiom of Choice is a very controversial axiom, some mathematicians do not even work under this assumption. An intuitive reason why this is so controversial: consider the equivalence relation of love. This partitions the set of people into two equivalence class: those who you do love, and those who you do not love. Axiom of Choice says we can pick a representative member of each equivalence class, but can you? This is clearly a more subjective argument, but *can* you pick a representative "person" that you love, or even one you do not love. If you are married, your spouse would most likely be your "representative" member. Otherwise, at least in my view, there is not a specific representative of these equivalence classes. This question really amounts to do you believe in soul mates? And more specifically *unique* soul mates? Some would answer affirmatively, others would not. This is basically the controversy in the Axiom of Choice; although most mathematicians *do* assume such an axiom.

The equivalence relation of sets is $A = B$ if $x \in A$ implies $x \in B$ and if $y \in B$ implies $y \in A$. This can also be seen through this light: $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$. Notice this is the Anti-Symmetric property of an ordering (partial or total). In practice, to show two sets are equal, one must show both containment directions. Now, that we have an equivalence relation, we can set up an Algebra to solve equations. This path becomes simpler if we introduce one more operation on sets: symmetric difference. Denoted $A \triangle B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$. The power of the operation is that $A = B$ if and only if $A \triangle B = \varnothing$. Consider the equation $X \cup A = B$, where $X \in U$ but is a "variable" set, and $A, B \in U$ are fixed, constant. Since these sets are equal, the symmetric difference must be $\varnothing$. So, using the first form of the symmetric difference operation: $((X \cup A) \setminus B) \cup (B \setminus (X \cup A)) = \varnothing$. Then using the relationship of

$P \setminus Q = P \cap Q^C$, we get:

$$
\begin{aligned}
B &= X \cup A \\
\varnothing &= [(X \cup A) \cap B^C] \cup [B \cap (X \cup A)^C] \\
\varnothing &= [(X \cup A) \cap B^C] \cup [B \cap X^C \cap A^C] \\
\varnothing &= (X \cap B^C) \cup (A \cap B^C) \cup [B \cap X^C \cap A^C] \\
\varnothing &= (X \cap B^C) \cup (A \cap B^C) \cup [B \cap X^C \cap A^C] \\
\varnothing &= (X \cap B^C) \cup [(A \cap B^C) \cap (X \cup X^C)] \cup [B \cap X^C \cap A^C] \\
\varnothing &= (B^C \cap X) \cup (A \cap B^C \cap X) \cup (A \cap B^C \cap X^C) \cup [B \cap X^C \cap A^C] \\
\varnothing &= \{[B^C \cup (A \cap B^C)] \cap X\} \cup \{[(A \cap B^C) \cup (B \cap A^C)] \cap X^C\} \\
\varnothing &= (B^C \cap X) \cup [(A \triangle B) \cap X^C]
\end{aligned}
$$

To continue the calculation, observe that $X \cup Y = \varnothing$ if and only if $X = \varnothing$ and $Y = \varnothing$. So,

$$(B^C \cap X) = \varnothing \text{ and } (A \triangle B) \cap X^C = \varnothing$$

The first equation can be written $X \setminus B = \varnothing$. So, when we take all the elements of B out of X, there is nothing left: this implies $X \subseteq B$. The second equation can be written, $(A \triangle B) \setminus X = \varnothing$. By exactly the same logic: $A \triangle B \subseteq X$.

Thus we see that X is a solution if and only if $A \triangle B \subseteq X \subseteq B$. But this can be further simplified since, $A \triangle B \subseteq B$ implies $A \subseteq B$. To see this, recall $A \subseteq B$ implies $A \cap B = A$ and $A \cup B = B$. So,

$$
\begin{aligned}
A \triangle B &= (A \cup B) \setminus (A \cap B) \\
&= B \setminus A \\
&= B \cap A^C \\
A \triangle B &= (A \setminus B) \cup (B \setminus A) \\
&\text{So,} \\
B \setminus A &= (A \setminus B) \cup (B \setminus A) \\
A \setminus B &= \varnothing
\end{aligned}
$$

So, when we take out every element of B from A we are left with nothing. This implies every element of A is an element of B, or more succinctly: $A \subseteq B$. At each step above try to recognize which of the Axioms for Boolean Algebras was used. As an aside: the power set of any set is group under the $\triangle$ operation and $\varnothing$ is the identity.

Cartesian Products

A very powerful use of sets is their unification of the concepts of functions, relations and operations. By considering a product on sets called the Cartesian product, functions, relations, and operations are reduced to subsets of this Cartesian product set. In addition, the use of the notation of exponentiation to denote the vectorspace $\mathbb{R}^N$ becomes apparent. Consider $\mathbb{R}^2$ as $\mathbb{R} \times \mathbb{R}$, where $\times$ is this the Cartesian product, read $\mathbb{R}$ cross $\mathbb{R}$. In general, $A \times B = \{(a, b) : a \in A, b \in B\}$ that is the set of ordered pairs where the first coordinate is an element of A, and the second coordinate is an element of B. So, $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. What is $\mathbb{Z}^2$ or $\mathbb{Q}^2$? The subset of all integer (rational) points of $\mathbb{R}^2$

A function, $f : A \to B$ is the set of ordered pairs $(a, f(a))$ for all $a \in A$ and $b = f(a) \in B$. This can be seen as a subset of $A \times B$, i.e. $f \subseteq A \times B$. Recall that a function is dependent on the Domain and Range; that is a function is a subset of $\mathbb{D} \times \mathbb{R}$, where $\mathbb{D}$ is the domain and $\mathbb{R}$ is the range. An operation on the set A, more properly a binary operation on A is a subset of $A \times A$. Addition and multiplication are binary operations. The addition (multiplication) on Integers, $\mathbb{Z}$, is a subset of $\mathbb{Z}^2$, while the addition (multiplication) on Rationals $\mathbb{Q}$ is a subset of $\mathbb{Q}^2$. This means that we can view operations as functions from a set into the same set. Addition is a function $+(a, b) = a + b$, also $\times(a, b) = a \times b$ (This product is regular multiplication not the Cartesian Product). In this view we can ask if addition (multiplication) are continuous operations: it turns out they are! Consider a two sequences in $\mathbb{R}$, $(a_k), (b_k)$ such that each converges to a point a, b, respectively. Notice by the completeness of $\mathbb{R}$, a and b must be a real numbers: $a, b \in \mathbb{R}$. The continuity of addition follows from: $+((a_k), (b_k)) = (a_k) + (b_k) = (a_k + b_k)$. Hence, $\lim_{n\to\infty}(a_k + b_k) = a + b$, thus the function $+$ is a continuous function. Similarly, it follows that multiplication is a continuous function (operation). The implications of these statements is the algebra of sequences is well defined, in that we can add and multiply convergent sequences: the limit of the sum (product) of some sequences is the sum of their limits. In general, let $(a_k)_{i=0}^N$ be a finite sequence of sequences, then $\lim_{n\to\infty} \sum_{i=0}^N (a_k)_n = \sum_{i=0}^N \lim_{n\to\infty} (a_k)_n$.

Relations, like the partial (total) orderings and equivalence relations are also subsets of a Cartesian product set. The ordering $\geq$ on $\mathbb{Z}$ is a subset of $\mathbb{Z}^2$ such that $(a, b) \in \geq$ if and only if $a \geq b$. Likewise the equality relation on $\mathbb{R}$ is a subset of $\mathbb{R}^2$ such that $(a, b) \in =$ if and only if $a = b$. We can conclude that all of the above structures are really just different "views" of sets, and their subsets. For instance, a group, $(G, +)$ is a the set G, along with a particular subset of $G \times G$ called $+$.

Ordered pairs, and their higher dimensional analogs can also be viewed in the set theoretic light. A set is not ordered, in that $\{a, b\} = \{b, a\}$, so we can not just write $\{a, b\}$ for an ordered pair. But, (a, b) is the set $\{\{a\}, \{a, b\}\}$. In this view, the first element (coordinate) of the ordered pair is clear, and thus we have persevered the ordering of the ordered pair. The ordered triple (a, b, c) is the set $\{\{a\}, \{a, b\}, \{a, b, c\}\}$. Now, we really have come full circle: every structure and concept discussed in the above has a foundation in the language of sets. This shows how foundational set theory is; math is a building constructed on set theory.

Peano's System

All the number system we constructed in the first section: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ where constructed out of the Natural numbers. We would like a set theoretic definition of the Natural numbers, and therefore extend the foundational nature of set theory. This construction is due to Giuseppe Peano.[16] He describes the natural numbers as the $(0, s(x))$, where 0 is the number zero, and $s(x)$ is called the successor function. The idea is successively using the successor function will generate all the rest of the numbers: i.e. $s(0) = 1$ and $s(1) = 2$ so on. In view of the above discussion, we can view this function as a particular set. Take $\{\varnothing\}$ to the set that describes the number 0. The number 1 is then the set $\{\varnothing, \{\varnothing\}\}$, and 2 is $\{\varnothing, \{\varnothing, \{\varnothing\}\}\}$. To see the pattern more efficiently: let notice that $1 = \{\varnothing, \{\varnothing\}\}$ so that $2 = \{\varnothing, \{1\}\}$, and $3 = \{\varnothing, \{2\}\}$. Or in this light: $1 = s(0)$, $2 = s(1) = s(s(0))$, $3 = s(2) = s(s(s(0)))$, and $s(n) = s^n(0)$, where s^n means n successive applications of the successor function. So the successor function appends a $\varnothing$ into the set: $s(n) = \{\varnothing, \{n\}\}$. Also think of the successor function as just adding one to the given number: $s(n) = n + 1$

Peano's proper description was in terms of the following axioms, and assuming that $=$ is an equivalence relation.

- 0 is a Natural number.
- If n is a Natural number, then so is $s(n)$.
- $0 \neq s(n)$ for any Natural number n, that is 0 is not the successor of any Natural number.
- If $s(n) = s(m)$ then $n = m$: that is s is injective.
- ***If $0 \in A$ and if $n \in A$ implies $s(n) \in A$, then $A = \mathbb{N}$.***

The first four axioms setup the natural numbers as we would expect. The last axiom is important enough to be named: the Principle of Mathematical Induction. This axiom gives a criterion to determine whether a set is equal to the Natural numbers. This axiom also presents a new method of proof: Proof by Induction. The proof procedure is a direct application of this axiom. Consider a certain property, $P(x)$; recall this notation is not a function, but rather shows that x satisfies the property P. First, prove the "base case": that is, show $P(0)$ is true. Then, make the "induction hypothesis": assume the property is true of n: $P(n)$ is true. Lastly, under this assumption prove the property holds for $n+1$: $P(n+1)$ is true. If there exist a proof for the last step, then by the axiom the set of number for which the property P is true is all the Natural numbers.

Intuitively, there is a domino effect going on. The last implication is: if $P(n)$ is true, then so is $P(n+1)$. If we manage to proof this statement, and we show that $P(0)$ is true (the base case) then all the dominoes fall. If $P(0)$ is true, then so is $P(0+1) = P(1)$, but then if $P(1)$ is true, so is $P(2)$, and $P(3), P(4), P(5)$ and so on. This is very powerful method of proof, with the swoop of two proofs: the base case, and the domino implication the inductive proof concludes an infinite number of results. Also, if instead of showing that $P(0)$ is true, but use that $P(1)$ or even $P(n)$ as the base case, then the domino implication shows that the property holds for all numbers greater than $1, n$, respectively.

For example, a prove by induction that

$$\sum_{i=0}^{n} i = 0 + 1 + 2 + 3 + ... + n = \frac{n(n+1)}{2}$$

The property $P(n)$ represents that $\sum_{i=0}^{n} i = \frac{n(n+1)}{2}$ is true.

Proof by Induction

Base Case: Show that $P(0)$ is true.

$$\sum_{i=0}^{0} i = 0 = \frac{0(0+1)}{2}$$

Induction Hypothesis: Suppose $P(n)$ is true, that is: $\sum_{i=0}^{n} i = \frac{n(n+1)}{2}$
Now, show that $P(n+1)$ is true.

$$\begin{aligned}
\sum_{i=0}^{n+1} i &= \sum_{i=0}^{n} i + (n+1) \\
&= \frac{n(n+1)}{2} + (n+1) = \frac{n(n+1)+2(n+1)}{2} \\
&= \frac{(n^2+n+2n+2)}{2} = \frac{n^2+3n+2}{2} \\
&= \frac{(n+1)(n+2)}{2} \\
\sum_{i=0}^{n+1} i &= \frac{(n+1)((n+1)+1)}{2}
\end{aligned}$$

The form of the last equation is suggestive that $P(n+1)$ is true, although technically all that is needed is the second to last equation. Notice, in the first equality how we make use of the Induction hypothesis.

Statement Calculus

We have seen that sets are the foundations for a mathematical language: the alphabet of math. We use this language to prove *theorems*. But, how are we sure that the given proof is correct? There should be a procedure to identify whether a given proof is logically valid. This is the burden of logic.

What can you conclude from the following pair of deductions?

If it is raining, I will stay inside.
I stayed inside.
Therefore: It is raining.

If it is raining, I will stay inside.
It is raining.
Therefore: I would stay inside!

Which one is logically valid? Or, are both valid? The first one may be false: I could be inside, while it is not raining. It is not *necessarily* true, and certainly not a valid argument. The latter argument is known as Modus Ponens; which is a logically valid inference. One might object noting that in certain situations one would leave the dryness of their home indifferent to the rain. Why does that not contradict the validity of the second argument? In this case, the first premise is no longer true, and thus the argument is not valid anyway!

Logic is the mathematical structure that tries present the theory of Reasoning. How and why arguments are correct and wrong. In order to discuss logic more succinctly, we adopt a notation such that a letter, say P, represents a *prime* sentence, like "It is raining." or "I will stay inside". Also, we use the sentential connectives: not, and, or, if ... then ... , and ... if and only if ... (where ... are replaced by sentences). Symbolically, we use: $\sim, \wedge, \vee, \rightarrow, \leftrightarrow$. A composite sentences is a Q such that it can be written using only prime sentences and the sentential connectives. For instance the composite sentence: It is raining and cold outside. Let R be the prime sentence: It is raining. And, let C be: It is cold outside. The composite sentence is then: $R \wedge C$. Notice, we make English grammatical corrections as necessary, in this case by omitting the common subject: It is raining and it is cold outside.

Suppose there was a set of prime sentences: $\mathcal{P} = \{P_1, P_2, ..., P_n\}$. Consider a new set $\mathcal{F}$ generated from the prime sentences in $\mathcal{P}$ and the sentential connectives. $\mathcal{F}$ is the set of all formulas like: $P_1 \wedge P_2$, $\sim P_3$, $P_n \vee (\sim P_2)$. Parenthesis may be introduce to unambiguous define the order of operations, and thus are also an element of the generating set. That is,

$\mathcal{F} = \langle \mathcal{P}, \{\sim, \wedge, \vee, \rightarrow, \leftrightarrow, (,)\}\rangle$, where the $\langle A\rangle$ is to mean the set generated from A. Note, an order of precedence of the connectives can be introduced to simply notation.

In connection to the English language, let $\mathcal{A}$ be set of letters in the alphabet, $\mathcal{P}$ be the set of all punctuations like: . , ! ? : ; " " , etc. Let $\mathcal{L} = \mathcal{A} \cup \mathcal{P}$. Those are the prime sentences of the English language. Consider the connective of concatenation: this just merges two letters to form words. We also need spaces to form sentences, and we could continue this pattern by adding paragraphs, chapters, books, etc. Let $\mathcal{C}$ be the set of all those connectives. Then, anything written previously, currently, or in the future is an element of the language generated by $\langle \mathcal{L}, \mathcal{C}\rangle$. This is meant more for an intuitive understanding of the concept of *being generated by*. We will do a less superficial connection to the English language in the Predicate Calculus.

An sight detour presents an interesting light on infinity. Suppose a monkey struck a typewriter randomly for an *infinite* amount of time. The resulting papers would also contain anything ever written.

Thus far, no means of determining the validity of the statements in the language has been discussed, only the way the sentences can be put together. To achieve this, we assign a truth value to each prime sentence: that is : P is either T or F; true and false respectively. Using these truth values, we can assign a value to each connective applied to prime sentences.

Initial values	Negation
P	$\sim P$
T	F
F	T

Initial values		Conjunction	Disjunction	Conditional	Bi-conditional
P	Q	$P \wedge Q$	$P \vee Q$	$P \rightarrow Q$	$P \leftrightarrow Q$
T	T	T	T	T	T
F	T	F	T	F	F
T	F	F	T	T	F
F	F	F	F	T	T

The headings of each connective correspond to their proper name. Negation, or *not*, is exactly as you would except: let R be the statement "It is raining outside". Then

$\sim R$ is "It is not raining outside". Note this is just manipulation in the language of the statements. If we assign a truth value of T to R, then $\sim R$ must have a truth value of F according to the table. There is subtle point here, in that it is only *when* we assign a truth value, that the table helps in simplifying the connectives.

The *and* and *or* connectives: conjunction and disjunction repetitively, work very similar to the English language's *and* and *or*. Recall the *or* is inclusive! A conjunction is only true when both of the sentences are true, false otherwise. A disjunction is only false we both of the sentences are false, otherwise true. For example, let C be "It is cold outside". Consider the composite sentences $A = R \wedge C$, and $B = R \vee C$. A is only true when both R and C are true, and B is false only when both R and C are false.[16, 8]

A word on the conditional truth table. This connective is the language equivalent of "If .... then .... ": $P \to Q$. It represents that Q is deducible from P If Q were true, then the true value of P should not matter, and the value of $P \to Q$ should be true; since we can deduce the truth of Q. Line 1 and 3 of the conditional truth table reflects these lines of reasoning. In the second line, the value of Q is false, and thus the conditional is false. Anytime P is true but Q is false, the sentence: if P then Q is false. The last line is true only to avoid certain pathologies. Consider the sentence $(P \wedge Q) \to P$. This says: If P and Q then P. Clearly, if I have four apples and five oranges, then I have four apples. However, the pathology arises if both P and Q are false, since in that case $P \wedge Q$ is false. But that reads, If I do not have four apples nor five oranges, then i do not have four apples. This is certainly a true statement, thus the last line of the conditional truth table. The Bi-conditional is very much like an equality: it is an equivalence relation. If one side is true, the so is the other; and if one is false then so is the other.

With these truth tables, it is now possible to discuss the validity of statement. Again, this can only happen once we have assigned truth value to each of the prime sentences in the statement being evaluated, since the tables will reduce any statements to either T or F. There are times when we would like to examine all possible values of a given statements. This can be done in a mechanical way: Setup an the initial value table for all possible combinations of truth values for each of the prime sentences in the given statement. Then, using the truth tables reduce the statement to a true value, depending on a certain line in the initial value table. For example, lets examine the validity of the statement: $(P \to Q) \longleftrightarrow (\sim P \vee Q)$

Initial values					
P	Q	$P \to Q$	$\sim P$	$\sim P \vee Q$	$(P \to Q) \longleftrightarrow (\sim P \vee Q)$
T	T	T	F	T	T
F	T	T	T	T	T
T	F	F	F	F	T
F	F	T	T	T	T

First of all, notice how the initial values are setup: they include every possible permutation of the truth value assignments to each prime sentence. The columns represent the steps take to reach the last column, which is the truth value assignments of the given statement. Each column is the reduction of the given statement using the truth table of the connectives. The last column happens to be all true: this is called a tautology. This implies that the given statement is always true, indifferent to the initial values. This also shows a new way of writing $P \to Q$, in terms of negation and disjunction.

Another illuminating example: Find the truth table for $[P \wedge (P \to Q)] \to Q$.

Initial values				
P	Q	$P \to Q$	$P \wedge (P \to Q) = A$	$A \to Q$
T	T	T	T	T
F	T	T	F	T
T	F	F	F	T
F	F	T	F	T

This is again a tautology! Actually this is the same as the argument given above called Modus Ponens. Intuitively, this says that if we have condition P and P implies Q, then we must have Q. Look back at the "raining" example given at the introduction to this section. This is the connection between a deduction and the statements we have been looking at. Consider the argument: If it is raining and below freezing, then it will snow. It is not snowing. Is there any logically valid conclusion to draw? Let R be "It's raining", C: "It's below freezing", and S: "It is snowing". Lets introduce some language for deductions: the premises are the statements given, in this case there are two: $(R \wedge C) \to S$ and $\sim S$. The conclusion or deduction is any logically valid inference from the premises. In this case, we can conclude that $\sim (R \wedge C)$. To verify this, we need to construct a truth table: But, how do we represent the premises and conclusion as one statement so that we *can* use the

truth table? Consider the statement:

$$P = \Big([(R \wedge C) \to S] \wedge (\sim S) \Big) \to [\sim (R \wedge C)]$$

Now, we can apply the truth table method to determine if this is a valid statement, that is we want to show this is a tautology.

Initial values							
R	C	S	$R \wedge C$	$(R \wedge C) \to S = A$	$A \wedge (\sim S)$	$\sim (R \wedge C)$	P
T	T	T	T	T	F	F	T
F	T	T	T	T	F	F	T
T	F	T	T	T	F	F	T
F	F	T	F	F	F	F	T
T	T	F	T	T	T	T	T
F	T	F	T	T	T	T	T
T	F	F	T	T	T	T	T
F	F	F	F	T	T	T	T

This arguments shows some very important properties. For instance, I claim this is effectively an Modus Ponens arguments. This is clear once the tautology $(P \to Q) \leftrightarrow (\sim Q \to \sim P)$ is verified. This is called the contrapositive. In English it is: "If you don't think about it, it makes sense" is the same as "If it doesn't make sense, think about it". There actually another hidden assumption here $\sim (\sim P) = P$, but this is verifiable in the framework of the truth table.

Armed with the contrapositive: the deduction goes like this: $P \to S$ is equivalent to $\sim S \to \sim P$. If $\sim S$ is true, then by Modus Ponens $\sim P$ is true. This is inference made in the above argument: however P was a more complicated statement, a composite sentence. This presents another useful point: In any tautology, we can replace each letter P_n in the statement with another more complicated statement and still maintain the tautology. In this case, we replaced P by $(R \wedge C)$ and concluded $\sim (R \wedge C)$, which is the replacement of $(\sim P)$.

Another tautology is $\sim (P \wedge Q) = (\sim P) \vee (\sim Q)$. Verify this with a truth table. This should remind you of the DeMorgan's Laws: $(P \cap Q)^C = P^C \cup Q^C$. We will explore the analogies between set theory and logic in a bit. For now, lets explore this tautology in

terms of the English language. Recall, the argument given above: "If it is raining and below freezing, then it will snow. It is not snowing." We conclude that: "It is not (raining and below freezing)." Substituting the above tautology we get: "It is not raining or it is not freezing". It is clear the English version of this deduction captures the intuitive mean we would expect.

As I alluded to, there is a strong connection between the Boolean Algebra of sets and the Statement Calculus just developed. Actually,the namesake of Boolean Algebra: George Boole was a logician. The analogs of $\cap$ and $\cup$ are $\wedge$ and $\vee$ respectively. Negation corresponds to complementation . The conditional in logic is the containment relation on Set, while the bi-conditional is equality on Sets. These operations behave the same way the operations on sets work. We have seen the logic version of DeMorgan's laws, directly above. This is not meant to say that logic and set theory are the same: they are not! But, there are strong parallels in each subject.

As we did with set theory: reducing the number operations to just complementation and union (or intersection), logic can be reduced to two operations: negation and the conditional. We can use disjunction (or conjunction) too, instead of the conditional. The rest of the connectives arise like so:

$$A \vee B = \sim A \to B$$
$$A \wedge B = \sim (A \to \sim B)$$
$$(A \leftrightarrow B) = (A \to B) \wedge (B \to A)$$

The sentential connectives: $\wedge$ and $\vee$ obey the laws of associativity, commutativity, are distributive over each other, and satisfy DeMorgan's Laws and the absorption laws $A \wedge A = A \vee A = A$. Also , the conditional connective is a partial ordering, since the following are tautologies: $A \to A$, $[(A \to B) \wedge (B \to C)] \to (A \to C)$, and from the definition of $\leftrightarrow$ we get the anti-symmetric property of a partial ordering. The bi-conditional then inherits its equivalence relation properties from the partial ordering $\to$.

There is another formulation of Boolean logic and the Statement Calculus, that connects logic to algebra. Consider the following correspondence:[16]

Logic Statement	Arithmetic Representation
$\sim P$	$1+P$
$P \wedge Q$	PQ
$P \vee Q$	$P+Q+PQ$
$P \rightarrow Q$	$1+P(1+Q)$
$P \leftrightarrow Q$	$1+P+Q$

A prime sentence can take on a value of 1 or 0 depending on its truth value, where $\mathsf{T}=1$, and $\mathsf{F}=0$ We also make certain restriction on the algebra of these numbers: $1+1=0$: this is basically addition modulo 2. So far in order to verify tautologies we have used the cumbersome method of truth tables. Now, we can use this representation, carry out the algebraic simplification and if the resulting number is 1 then the statement is true, if the result is 0 then the statement is false; also, if the resulting algebraic statement is equal.

Let use this representation to show that $P \vee \sim P$ is a tautology. Lets first make sure that it is in fact a tautology using the truth table method.

Initial values		
P	$\sim P$	$P \vee \sim P$
T	F	T
F	T	T

Thus it is a tautology. In English this mean: "It is either raining or it is not raining". This reminds me of a joke: What is the probability of seeing a dinosaur tomorrow? Well, 50-50: I will either see one, or I won't.

Okay, back on track, lets convert the logic statement $P \vee \sim P$ to the arithmetic representation: $P+(1+P)+P(1+P)$. Let x be an arbitrary truth value, so either 0 or 1. Then, $0+0=0=1+1$, so $x+x=0$, and $x(1-x)=0$, since $1(1-1)=1(0)=0$ and $0(1-0)=0$. So the arithmetic representation reduces to : $1+(P+P)+P(1+P)=1+0+0=1$. Therefore, this statement is a tautology.

As another example, lets shows that the definition of $\leftrightarrow$ in terms of $\rightarrow$ is a tautology:

$(A \leftrightarrow B) \leftrightarrow (A \to B) \wedge (B \to A)$. This becomes in the arithmetic representation:

$$\begin{aligned}(A \leftrightarrow B) &\leftrightarrow (A \to B) \wedge (B \to A)\\ 1 + A + B &= [(1 + A(1 + B)][1 + B(1 + A)]\\ &= 1 + B(1 + A) + A(1 + B) + A(1 + A)B(1 + B)\\ &= 1 + B + AB + A + AB\\ &= 1 + A + B\end{aligned}$$

With this method any statement in the statement calculus is reduced to algebraic manipulations. The algebra is a little tricky at first, but armed with three equations: $x + x = 0$, $x(1 - x) = 0$, and $x^2 = x$ and the axioms for a field everything works out.

Predicate Calculus

The statement calculus does well with most language concepts, but it does have "holes". There is no way to express the statement: "Some of the people invited to the party will not come". The sentences we deal with earlier had much simpler form, easily identified by a letter. In the Predicate Calculus, we dive into the structure of a sentence appealing to subject-predicate nature of a sentence.[16, 8]

Let $P(x)$ be a sentence, where P is the predicate and x the subject. Consider the sentence: I will come to the party. The predicate is: will come to the party, which is represented by $Party(x)$. Then, the sentence is $Party(I)$. We can use the statement calculus on sentences that are completely defined: like $Party(I)$. However, the point in dissecting the sentence was to reach a broader logical system. To state the original sentence, we need to introduce the *existential quantifier*, $\exists$, read there exists. $(\exists x \in Invited) \sim Party(x)$. Another way to write this is: $(\exists x)(Invite(x) \to \sim Party(x)$, read: There exists an x such that If x was invited to the party, then x will not go to the party.

Consider the argument: All dogs have four legs. I have two legs. Therefore I am not a dog. Let $Four(x), Two(x), Dog(x)$ represent the respective predicates. The third statements is: $\sim Dog(I)$. The second could be written $Two(I)$, but for these purposes lets write $\sim Four(I)$: I do not have four legs. I only chose this form to make the argument more clear. The first is a little more tricky. A partial representation: $Dog(x) \to Four(x)$. but, this just says if x is a dog, then x has four-legs. In order to specify the notion that **all**

dogs have four legs, we need the *universal quantifier*, $\forall$, read for all. The first statement is then $(\forall x)(Dog(x) \rightarrow Four(x))$. The full argument is then:

$$[(\forall x)(Dog(x) \rightarrow Four(x)) \wedge \sim Four(I)] \rightarrow \sim Dog(I)$$

To show the validity of the argument, we would like to reduce the formula to a sentence in the statement calculus then we can use the machinery developed. The second and third sentences are already statements. From the first formula, we can get the statement $Dog(I) \rightarrow Four(I)$, since the formula is true for all x, in particular I. Thus, we have reduced this argument of the Predicate calculus to the Statement Calculus, where we can show $[(D \rightarrow F) \wedge \sim F] \rightarrow \sim D$ is a tautology.

It turns out that the existential quantifier and the universal quantifier are related by negation. In particular: $\sim ((\exists x)A(x)) = (\forall x)(\sim A(x))$, or by negating this statement: $(\exists x)A(x)) = \sim (\forall x)(\sim A(x))$. The first form is suggestive of how to transfer the negation sign through the quantifiers; the second is can be viewed as the definition for the universal quantifier. Hence, the Predicate Calculus is just the statement calculus plus the existential quantifier.

An argument in the Predicate calculus is a finite set of premises: $A_1, A_2, ..., A_N$. Any formula B is consequence, or a logically valid conclusion of the set of premises if there exists a finite sequence $E_1, E_2, ..., E_M$ of formulas, such that $E_M = B$ and $E_i = A_j$, for some pair (i, j), and the remaining indexes in the sequence (E_i) are logically justified from a use of the Modus Ponens inference (and its contrapositive). In the statement calculus, we reduced the validity of a statement to algebraic manipulations. Consequence and validity are only slightly different: validity states whether a formula is valid: tautology. If B is consequence of a set of premises: $A_1, A_2, ..., A_N$: we write $A_1, A_2, ..., A_N \vDash B$. This is the form of an argument: start out with a set of premises, and conclude B. An argument can be turned into a formula, thus reducing consequence to validity. $\vDash (A_1 \wedge A_2 \wedge ... \wedge A_N) \rightarrow B$. The $\vDash$ denotes "tautology" in this case, in the argument form it denotes "is a logically conclusion". In effect, this sleight of hand equates the notion of consequence in an argument and the notion of validity in a statement.

In the statement calculus, we had the five sentential connectives, (the parenthesis), and the prime sentences. Also, we had the logic inference of Modus Ponens: $A \wedge (A \rightarrow B) \vDash B$, or $\vDash [A \wedge (A \rightarrow B)] \rightarrow B$. In the Predicate calculus, we have the statement calculus

plus the existential quantifier, and a couple new inference rules. The rules of: Universal Specification and Universal Generalization. Universal specification is: if $(\forall x)A(x)$ and t is a term that can be replaced for x, then $A(t)$. More specifically: $(\forall x \in \mathbb{D})A(x) \wedge (t \in \mathbb{D}) \vDash A(t)$. Another way of writing this using only connectives: $(\forall x)\mathbb{D}(x) \wedge A(x) \wedge \mathbb{D}(t) \vDash A(t)$, where $\mathbb{D}(x)$ is the predicate saying: "x is a member of $\mathbb{D}$". This just introduces the idea that we can a priori restrict a certain variable, say t, to lie in a specified domain. Then, the statement becomes simply: $(\forall x)A(x) \vDash A(t)$: much nicer!

The second rule of inference goes like so: if $\mathbb{D}(x) \to A(x) \vDash (\forall x \in \mathbb{D})A(x)$. Intuitively, what this says is if you take an *arbitrary* element of $\mathbb{D}$ and show that this element satisfies A, that is $A(x)$. Then, we can logically conclude that for all $x \in mathbbD$ satisfy A. Notice, the only condition imposed on x was that $x \in \mathbb{D}$, and then we showed that $A(x)$. This must imply that every element of $\mathbb{D}$ $A(x)$ must be true.

Collectively these two rules allow one to bounce back and forth from the predicate calculus to the statement calculus. The first rule will take us from the predicate calculus to the statement calculus, while the second allows us to go back to the predicate calculus. Both of these rules deal specifically with the universal quantifier, but how do we deal with the existential quantifier. $A(x_0) \vDash (\exists x)A(x)$ and $(\exists x)A(x) \vDash A(x_o)$. These are more properly definitions of the existential quantifier than logic inferences. The first states that if x_O has the property A, then clearly there exists an x such that $A(x)$. The second states that if there exist an x such that $A(x)$ then we can *choose* a x_0 such that $A(x_0)$. The second statement is very similar to the Universal specification, except that in that case *any* element we choose will satisfy the property; as opposed to in this case there exists *at least one* element that satisfies A, but no guarantees are made for the rest of the elements.

Let's consider an extension of the argument first presented in this section: All dogs have four legs. All humans have two legs. Therefore, All humans are not dogs. For simplicity: assume that the variable x is restricted to the domain of animals. The argument written symbolically is: $(\forall x)(D(x) \to F(x)), (\forall x)(H(x) \to\sim F(x)) \vDash (\forall x)(H(x) \to\sim D(x)$. The comma just represents a new premise. This is not a new operation just a simplification in notation. This can be properly written with a $\wedge$ between all the premises.

To show the validity of this argument, we would like to reduce it to an argument in the statement calculus, using the Universal specification, then, bring it back to the Predicate calculus by using the Universal generalization rule. The deduction goes as follows:

	Statements	Reasons
1	$(\forall x)(D(x) \to F(x))$	Premise A
2	$(\forall x)(H(x) \to\sim F(x))$	Premise B
3	$D(x_0) \to F(x_0)$	Universal Specification from Premise A
4	$H(x_1) \to\sim F(x_1)$	Universal Specification from Premise B
5	$x_0 = x_1$	Since we can *choose* them to be the same!
6	$H(x_0) \to F(x_0)$	Just substituting equals for equals
7	$\sim F(x_0) \to\sim D(x_0)$	The contrapositive of $D(x_0) \to F(x_0)$
8	$H(x_0) \to\sim D(x_0)$	Combining lines 4 and 7 through transitivity of $\to$
$\therefore$	$(\forall x)(H(x) \to\sim D(x))$	Universal Generalization from line 8

The symbol $\therefore$ means therefore. Do you see how we reduced the argument into the statement calculus, and then in the last line brought it back into the predicate calculus. This shows that $(\forall x)(H(x) \to\sim D(x))$ is a logical consequence of the premises. Or, in terms of validity: $\vDash (\forall x)(D(x) \to F(x)) \wedge (\forall x)(H(x) \to\sim F(x)) \to (\forall x)(H(x) \to\sim D(x))$.

Now, we have developed a logic system that can express most, if not all, concepts and notions of an argument enough so to be able to determine its validity, or equivalently the consequences from a set of premises. However, there is a subtle distinction in the logic system developed if we let the quantifiers modify not only the subjects the of sentences but also the predicates. For instance, $(\forall P)(D(x) \to P(x))$. This is a *second-order* logic, where the former would be called a *first-order* logic. There's a theorem of Analysis called the well-ordered property, this can not be expressed in terms of first-order logic. The property states: For any non-empty subset of X (some space), there exist a minimal element in the set. The latter part of the statement can be written: $(\exists x)P(x) \to (\exists x)(P(x) \wedge (\forall y)(P(y) \to (x \leq y)$. The predicate P is interpreted to mean that $x \in P$ for some subset $P \subseteq X$. Let's break up this statement and translate into English what it means. The first part: $(\exists x)P(x)$, just assert that P is non-empty; there exists at least one $x \in P$. The second part (after the first $\to$) can be broken into two parts on either side of the $\wedge$. The quantifier $(\exists x)$ applies to the whole part, and the x used in the second part, is *not* the same as the x used in the statement of non-emptiness. We could replace the first occurrence (or second) with a different variable name, z, and still have the same meaning.

Dissecting the second part, the statement $(\exists x)P(x)$ asserts that the element x is in P. The last part express that x is the minimal element. $(\exists x)(\forall y)(P(y) \to (x \leq y))$. Translating this direct says: There exists an x (which is in P) such that if $y \in P$ (thats

the $P(y)$) then $x \leq y$ for all y. So, we see the statement expresses the idea that for a certain subset P, there exists a minimal element. This statement is of first-order since the quantifiers only modified the subjects of the predicates: x and y. The full well-ordered property states that for any subset P of X, there exists a minimal element. We can write a $(\forall P)$ in front of the whole statement and thus expresses the well-ordered property. Notice the quantifiers now modify both the subjects and the predicates, hence this is a second-order logic statement.

$$(\forall P)[(\exists x)P(x) \to (\exists x)(P(x) \wedge (\forall y)(P(y) \to (x \leq y)]$$

There are n-order logics too, which just generalize the notion of what the quantifiers are allowed to modify. With these ideas, there are not many, if any, statements which can not be written in the predicate calculus.

As an aside, recall Russell's paradox. Written in the language of logic, consider the predicate $R(x)$ which is true when x is a set and $x \notin x$. Let $S(x)$ be the predicate: "is a set", while $M(x)$ be the predicate $x \notin x$. Then Russell's paradox is written: $R(x) = S(x) \wedge M(x)$. Consider the validity of $R(R)$ which is $S(R) \wedge M(R)$. Now, $R(R)$ implies that $R \in R$, but $M(R)$ implies that $R \notin R$, thus reaching the contradiction. Also, if $\sim R(R)$ then $R \notin R$ but also $\sim M(R)$ implies that $R \in R$, again reaching a contradiction. The problem is in the second-order nature of this statement: the x in $R(x)$ is replaced by R a predicate to form $R(R)$. The resolution to this paradoxes comes from denying the statement $S(R)$, thats is that R is *not* a set. This sleight of hand completely eliminates the paradox: since $R(R) = S(R) \wedge M(R)$ but $S(R)$ is false and thus the conjunction is also false, and so $R \notin R$ since R is not a set. The set R is in some sense too big to be a set. Properly it is called a *class*, or at times the jargon *family*, or *collection* are used.

Proofs

Armed with the language of logic, we can now explore Proofs and Theorems. A lemma, a proposition, a corollary are all synonymous with a Theorem, except that a lemma (proposition) are considered auxiliary proofs working towards a Theorem. A corollary is a direct consequence of a Theorem. In practice, though they are all effectively the same. A conjecture is a statement that has not been proved, a hypothesis.

In order to prove a Theorem, a logically valid proof must be presented. A proof is a *finite* sequence of logically valid deductions from the set of premises such that the last line in the proof is the statement we were trying to prove. In other words, the theorem must be a consequence of the set of premises. The validity of a proof is the burden of logic. Most Theorems have the form $(P_1 \wedge P_2 \wedge, ..., \wedge P_n) \rightarrow B$. Therefore, once the theorem has been proved, then by Modus Ponens if we prove the premises are satisfied, thats is $P_1 \wedge P_2 \wedge, ..., \wedge P_n$ is true, then we can logically conclude that B is true. For instance, Pythagoras' theorem states: if $\triangle ABC$ is a triangle, and $\angle ACB$ is a right angle, then $a^2 + b^2 = c^2$, where a corresponds to the side opposite the point A in the triangle (and the same of b and c). Thus, if I can show that $\triangle ABC$ is in fact a triangle with a right angle, which we name $\angle ACB$, then the equation $a^2 + b^2 = c^2$ is true.

We have seen three different types of proofs: direct proof, indirect proof, inductive proof. An inductive proof is completely determined by the principle of mathematical induction. This was discussed in the section on the Peano's Axioms for the natural numbers. Therefore the logic of this type of proof is valid only in the case where axiom of Mathematical induction is satisfied. A direct proof is exactly the same as the procedure to show the validity of a statement in logic. It is a finite sequence of statements such that each follows logically from the set of premises and the last line in the proof is statement we were trying to prove.

An indirect proof, or a proof by contradiction, in my opinion is the most interesting type of proof. The procedure is to assume the negation of the statement, and reach a contradiction. This implies that the negation of the statement can not possible be true, and since one of A or $\sim A$ is true: A must be true. This type of proof is so interesting is that one is trying to find a flaw. In the other types, we try to show the validity of the statement by taking logically valid steps reaching the desired statement. In general, there maybe many "paths" to prove a theorem directly, but the end point of the path is always the same: the statement of the theorem. In a indirect prove, there are still many "paths" to a proof, but in general, the end point is *not* the statement of the theorem, but some statement that contradicts an assumption or a theorem.

Consider the statement: There exists an infinite number of primes. How can we prove this? The following proof is due to Euclid and is presented in the Elements.

Proof by Contradiction[4]

Assume: There exists a finite number of primes, say $p_1, p_2, p_3, ..., p_n$. We are trying to find a flaw in making this assumption. That is, we are searching for a number q that *is* prime, and *not* in the set $\mathcal{P} = \{p_1, p_2, ..., p_n\}$. If we can do this, then we have shown that the assumption that the set of primes is finite is false (leads to a contradiction), and so it must be that there are an infinite number of primes. Let $q = p_1 \times p_2 \times ... \times p_n + 1$. Clearly, $q \notin \mathcal{P}$. Now, we need to show that q is prime. An equivalent definition of a prime is that it is not divisible by any prime number. So, if we can show that q is not divisible by any p_i then q is prime. For a number to be divisible by another, their division must result in a Integer. Consider, $\frac{q}{p_i} = \frac{p_1 \times p_2 \times ... \times p_n + 1}{p_i} = \frac{p_1 \times p_2 \times ... \times p_n}{p_i} + \frac{1}{p_i} = p_1 \times ... \times p_{i-1} \times p_{i+1} \times ... \times p_n + \frac{1}{p_i} = P_i + \frac{1}{p_i}$. Let P_i be the product $p_1 \times p_2 \times ... \times p_n$ without the i-th term. This is clearly not an integer, since $\frac{1}{p_i} \notin \mathbb{Z}$. Also, since we used the general term p_i by Universal Generalization, this implies that $\frac{q}{p_i} = P_i + \frac{1}{p_i}$ $(\forall\, 0 \leq i \leq n)$. So we have found a q which is not in $\mathcal{P}$, and that q is prime. Thus, we assumed that there were a finite number of primes, and listed them *all*. Then, we constructed a new prime! This contradictions our assumption that we listed them all. Therefore, there are an infinite number of primes.

Theorems are simultaneously the tools of math and its ultimate goal. We use theorems as tools to build other theorems so we can use those as tools, and so on. An helpful analogy: think of Mathematics as world, like the Earth. We would like to explore this world enough to be able to draw a map of it. The axioms of logic are analogous to the mechanics of physics, in that the only "steps" we are allowed to take in the Math world are the logically valid one. On earth, every step we take is a physically valid step. At first, we only have our legs to journey around: in the math world, we only have the axioms. Then, we can gather enough material and ingenuity to build a car. Analogously, we have explored the axioms and constructed proofs of certain statement, which we call theorems. Now, if the car has enough gas, we can further explore the world. If the premises are satisfied, the theorem's conclusion is true. We could walk around the entire planet, given enough time, without having to resort to building the car; however, it would be nicer to have a car to explore. The proofs of the theorems only depend on the logic axioms—the steps of the math world. A theorem is the mathematical car: it takes you from A to B in one step.

There are limits to this analogy. If the car has gas, it may not necessarily run: no keys! However, this can not happen in the mathematical world. If the premises of the

theorem are satisfied, then the theorem will always "run". This is power of deductive reasoning. Science uses the scientific method as their proof method: experimentation, observation, and data analysis. The data analysis is to compare the observations to the results predicted by a certain theory to test its accuracy. The main fabric of the scientific method is that scientist *can* create an experiment and see what happens. Mathematicians are not afforded this luxury. We can see or touch abstract structures only in our mind's eye or on our mind's skin.

That is why such effort has gone into axiomatizing math. It is truly remarkable to be able to think about something in my mind, and that through a proper axiomatization that something is transfered into someones mind. Artist, writers, dancers spend their whole life practicing and honing in their skill: transferring an idea (emotion) from their mind to another's through a certain medium. Art is judged in a subjectively: which in turn implies that different people may judge (perceive) different ideas or emotions from a piece of art. The goal of Mathematics is quite the opposite: a proof must be judged by different people in exactly the same way.

Axiomatic Theories

In order to ensure the idea transfered by a proof will be judged in the same way, we make use the notion of an Axiomatic Theory: Undefined terms, defined terms, axioms, a system of logic, theorems. The undefined terms are needed to avoid circular definitions, while the defined terms are used to simply the language. The system of logic is the predicate calculus (or some higher-order logic system). The axioms are set of formulas that are assumed to be valid. The theorems are any statement that is logically valid. The axiomatic theory generated is the set of all theorems under the assumption of the axioms.

Before, we dive into the axiomatic theories, a word on Languages and grammar. Which of the following are sentences in the English language?

A Scientist in Africa have discovered a new species of monkey: the calik.

B Dog bark tree pink in the banana.

C The will rise in the east and set in the west.

Obviously, B is not a sentence, but neither is A: calik is a word I made up. Yet, at first glance it does seem like a sentence, even though you did not know what calik meant. It is the structure, how the words are grouped, that determine whether or not it is an English sentence. This is the concept that is abstracted to the notion of a Language, where the grammar determines how th sentences are built. As a simple example of the English language, consider the grammar:[3]

$$\begin{aligned} \text{DNP VP} &\to \text{S} \\ \text{V DNP} &\to \text{VP} \\ \text{DET NP} &\to \text{DNP} \\ \text{A NP} &\to \text{DNP} \\ \text{N} &\to \text{NP} \end{aligned}$$

The first one reads: A definite noun phrase (DNP) followed by a verb phrase (VP) is a sentence (S). The rest read similarly with the association: V is a verb, VP is a verb phrase, DET is a determiner (like the), N is a noun, NP a noun phrase, and A is an adjective. We can now decompose the sentence: The plane is a marvelous invention.

The	plane	is	a	marvelous	invention.
DET	N	V	DET	A	N
DET	NP	V	DET	A	NP
	DNP	V	DET	NP	
	DNP	V	DNP		
	DNP	VP			
		S			

The first line is an assignment, in that we assume that "the" is a DET and "plane" is a N, and so on. Every line following is a application of the rules of the grammar. For instance, from the first line to the second we used $N \to NP$. Since the last line is S, the expression is a sentence of the language generated by the grammar. This is very simple grammar and by no means is a complete representation of the grammar of the English language. For instance, there is no mention of prepositions, pronouns, or adverbs which are an essential part of the language. That said, we would just need to keep adding more relations in the grammar to accommodate these new types. We could introduce the axiom

$PDNP \to PP$ and add to the language the notions of a preposition and a prepositional phrase, also we need to add axioms as to how PP's work with the other parts of the grammar, like $VPP \to VP$. Also add the notion of a pronoun PN: $PN \to NP$. Now, the following expression is a sentence of the language: I walked to the park. This particular type of study into the grammatical structure of language was founded by an American linguist Noam Chomsky.

In another view, let A^n be the set of all finite sequence of element in the alphabet, A. In relation to the English language, "asdpiugqwas" is a finite sequence of elements of the alphabet, however it is not a word. Let A be the English alphabet, and $\mathcal{F}$ be the set of words, then "asdpiugqwas", and "mathematics" $\in A^n$ but only "mathematics" $\in \mathcal{F}$. However, a better way to think of the relation between A^n and $\mathcal{F}$ is as the set of words and the set of sentences, respectively. $\mathcal{F}$ is called the Language generated from A and a certain grammar.

An axiomatic theory is effectively the same thing as a the language generated from a grammar. A axiomatic theory is $\mathcal{T} = (A, \mathcal{F}, \mathcal{A}, \mathcal{R})$ where A is the alphabet, $\mathcal{F}$ is the language generated by the grammar $\mathcal{R}$ and $\mathcal{A}$ is the set of axioms. $\mathcal{A}$ is a subset of $\mathcal{F}$ such that we assume the axioms are valid. Let Chomsky-style grammar be $\mathcal{R}$. Then $\mathcal{A}$ is the assignments assumed in the first line of the deduction used to show that: The plane is a marvelous invention, is a sentence. The axiomatic theory generated is the language of simple sentences.

To be concrete, I present the axiomatic theory of the predicate calculus:[16] $A \to (B \to A)$ $(C \to (A \to B)) \to ((C \to A) \to (C \to))$ $(\sim B \to \sim A) \to (A \to B)$ $(\forall x)(A \to B(x)) \to (A \to (\forall x)B(x))$ $(\forall x)A(x) \to A(t)$ for some acceptable t Recall, that the predicate calculus just needed the symbols $\to, \sim, \forall$, the rest of the connectives can be written in terms of $\to, \sim$, and the existential quantifier in terms of $\sim, \forall$. Any axiomatic theory that desires to obtain any logically valid results *must* assume the above axioms, that is any axiomatic theory contains the predicate calculus. In particular we have the rules of inference Modus Ponens and Universal Generalization. This ensures that the proof and theorems are logically sound. However, an axiomatic theory can assume other additional axioms: if so, they are called proper axioms. The axiomatic theory of groups is defined by the axioms of the predicate calculus, a predicate $P(x, y)$ which means $x = y$, and an operation $f(x, y) = x + y$, such that the following axioms are satisfied: (The colon is separating the same presentation in terms of the P and f; on the other side,

in terms of $=$, $+$) $(\forall x,y,z)f(x,f(y,z)) = f(f(x,y),z)) : (x+(y+z) = (x+y)+z)$ $(\forall x,y)(\exists z)P(x,f(y,z)) : (x = y+z)$ $(\forall x,y)(\exists z)P(x,f(z,y)) : (x = z+y)$ $(\forall x)P(x,x) : (x = x)$ $(\forall x,y)(P(x,y) \to P(y,x)) : x = y \to y = x$ $(\forall x,y,z)(P(x,y) \wedge P(y,z)) \to P(x,y) : x = y \wedge y = z \to x = z$ $(\forall x,y,z)(y = z \to (x+y = x+z \wedge y+x = z+x)$ The last four axioms are not necessarily particular to a group. Axioms four, five and six just say that $P(x,y)$ is an equivalent relation, and the last line says that adding equal things to equal thing is again equal: this is the compatibility of $+$ and $=$. This is a slightly different presentation than the one given before using identities and inverses, however both axioms generate the theory of groups. The definition given earlier for groups was that the operation was closed, associative and there exist an identity and a inverse for every element. The closure is guaranteed by the definition of the function $f(x,y)$, and thus does not have to been re-written. Associativity is the first axioms presented here. The next two axioms serve the same roles as the two axioms of identity and inverses.

For instance, the existence of identity in this presentation is given by: $(\forall x,y)(\exists z)(x = y+z)$, well by universal specification, let $x = y$, so there exists a z such that $x = x+z$. Similarly, using the other axiom, $(\exists z)(x = z+x)$. So, $x+z = x = z+x$ which implies that z is the identity element, call it e. For inverses, pick x and e and then the axiom implies $(\exists z)(e = x+z)$; using the other axiom we get: $(\exists z)(e = z+x)$. Both equations together imply that z is the inverse of X.

We have seen plenty of example of axiomatic theories through this work. The axioms for all algebraic structures are grammars (along the addition axioms for the predicate calculus) for the axiomatic theory they generate. Euclid's push for axiomatizing geometry is the first know attempt at using an axiomatic theory. Hilbert presented the first "full" axiomatization of geometry. Projective geometry, the umbrella geometry, is an axiomatic system. Zermelo-Fraenkel developed the axiomatic theory of sets. Peano's axioms for the Natural Numbers turn them into an axiomatic theory. Finally, the predicate calculus is an axiomatic theory, in fact in a sense is the smallest axiomatic theory.

This is how a mathematician transfers the ideas in his head to another's mind, without the risk of a subjective judgment (as opposed to art). He writes down the axioms he is assuming, and then present a claim in the language generated. If the axiomatic theories maintains certain properties, then I can determine unambiguously whether his claim is a theorem or not: that is to say, is his claim valid or invalid. This would rarely happen with an art piece: if I claim that a sculpture is beautiful, it is not necessary that anyone who

looks at the sculpture would be as enthralled as I was.

So what are the properties that an axiomatic theory must posses to ensure a reader of a proof can unambiguously determine the validity of a claim. There are two notions: Consistency and Completeness. A *consistent* axiomatic system is when either A is true or $\sim A$ is true, but not both.[16] It should be clear that an inconsistent axiomatic theory is worthless. We introduce the concept of a model to test consistency of a theory. A model for an axiomatic theory is an assignment of each of the undefined terms such that the axioms are satisfied. If there exists a model for an axiomatic system, then it is consistent. As an example, consider the following three axioms. The axiomatic theory generated is called Four-Point Geometry.[2] Axioms of Four-point geometry

Undefined terms: Point, Line, Incidence

1 There exists exactly four points.

2 Two distinct points are on exactly one line.

3 Each line is on exactly two points.

In order to show the consistency of this axiomatic system, we need to find a model. First, we need to assign a meaning to the undefined terms. We identify the undefined term point with a point on the plane, a line with the a line on plane, and the relation incidence to be the usual euclidean relation of incidence. Then, by the first axiom, draw four points (say, in a square) and between every pair of points draw a line according to axiom 2. Now, verify that axiom 3 is satisfied: that is, check that every line contains two points. That picture is enough to show the consistency of Four-point Geometry. There is subtle distinction here between absolute consistency and relative consistency. In drawing our model in the plane, we have shown that Four-point geometry is as consistent as plane geometry. If an inconsistency arises in plane geometry, then that inconsistency propagates to the Four-point geometry, however it is believe by most mathematicians that plane geometry is consistent. The difficulty in showing the consistency of plane geometry is that we have to draw a model that is not in the plane for that would be circular. We will see in the next section, that relative consistency is all we can hope.

In the figures, there are two geometrically different models for Four-point geometry. Each satisfies the axioms, so either one is enough to show the consistency of Four-point

geometry. In this case they are effectively the same model, we say they are isomorphic. The are some axiomatic theories that admit different models. For instance, in the axiomatic theory of groups $(\mathbb{R}, +)$ is a model, so is $(\{0, 1\}, \times)$ yet these are not the same, one has an infinite number of elements, the latter has exactly 2.

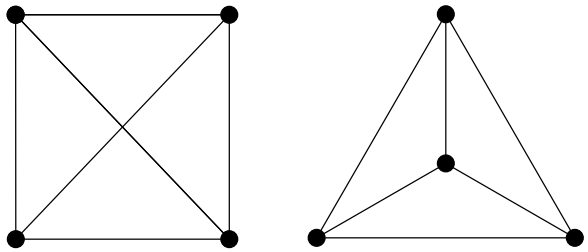

Figure 17:

There is another notion associated with an axiomatic system: *independence*. An axiom is called dependence if it can be deduced from the other axioms, that is, it is a theorem. An independent axiom is not dependent. An independent set of axioms is a more elegant way of describing the axiomatic theory, however there is no need to have independent axioms. The theory generated is the same whether the axioms are independent or dependent, but there will be more to prove as theorems in a independent system of axioms. Recall, the controversy of the Parallel Postulate. The popular belief was that it was a dependent axiom, yet we saw that it is a independent axiom. To show independence, we need to present a model of the axiomatic system replacing the axiom in question with its negations.[2] If there exists a model for this modified axiomatic system then the axiom is independent. The independence of the Parallel Postulate is established through the models of Elliptical Geometry and Hyperbolic Geometry. This notion is rather unimportant, it only affects the elegance of the presentation. As such, an axiomatic theory is usually developed with a dependent set of axioms, and then eliminate those are not needed (are dependent).

An axiomatic system is *complete* if every statement containing undefined and defined terms can be proved valid or invalid.[10, 2] In other words it is not possible to add a new independent axiom to the system while maintaining consistency. Intuitively, a complete system has enough axioms to prove any statement that is true is logically valid. A direct proof of completeness is usually impossible to find. But, for a complete axiomatic system

any two models are isomorphic. Which by the logically equivalent contrapositive: If two models for an axiomatic system are *not* isomorphic then the system is *not* complete. This shows that the axiomatic theory of groups is not complete, since we found two non-isomorphic models, while four-point geometry is complete.

There exists certain axiomatic theories that not complete, that means there are statements which we *can not* prove true nor false. These are the so-called *undecidable* statements. Gödel's Incompleteness Theorem states that in any rich enough axiomatic theory there *are* undecidable statements.[10] By rich enough, I mean that the theory contains the predicate calculus and the arithmetic of the natural numbers, or equivalently Peano's axioms. This is a profound result: no matter how careful and thorough we are in defining an Axiomatic system, there will be statements that are undecidable. Furthermore, the addition of an axiom to prove the given undecidable statement, does not solve the problem: there will be another undecidable statement.[12]

An example of a statement that is undecidable within the axiomatic system is the consistency of theory. Another example: Goldbach's conjecture: every even number > 2 is the sum of primes. Computers have tested this for a billion even numbers, and will continue to test this claim. Yet, no one can disprove the statement, and no one has proved it either. Whether it actually is an undecidable statement is also undecidable. If the statement is not undecidable then it is either true or false, but we do not know which, thus the statement is undecidable.

It's a Wrap

There is a lot of information in the above, but I hope that at least there is now an intuitive understanding of the What Mathematics is? and What Mathematics does? We saw that all sets apart from being the basic object in math, unified the notions of: equivalence relations, ordering relations, operations, even the integral. Logic is the language we use to write math effectively. The axiomatic theory is the cornerstone: any theory in math, any problem, any question, can be asked in a appropriate axiomatic theory. That is the the true function of Math: to ask questions and eventually answering them. Both aspects are crucial to practing Math. Most importantly, learn to ask the right question. It's meaningless to ask what the angle is between vectors in topology. But the skill and practice obtained in actually answering the questions is a life-long practical skill that

transfers to most jobs.

Math can show us how our reasoning works, and why the forces of the physical world drive our mechanics. It can talk about objects and structures that can not be visualized in the mind nor drawn in this world. As good as it is it does have it flaws, as Gödel's theorem implies. But, I rather take it in the light that there is always something more to discover, to learn. There is always another level of abstraction and new way to connect and relate things. In the end, I think Math, just like art, writing, dancing, music, are all languages that express idea and emotions: mathematicians have just figured out how to take the subjectivity out of it.

Bibliography

[1] S.J. Axler. *Linear algebra done right*. Springer Verlag, 1997.

[2] J.N. Cederberg. *A course in modern geometries*. Springer Verlag, 2001.

[3] K.J. Devlin. *The language of mathematics: making the invisible visible*. Holt Paperbacks, 2000.

[4] U. Dudley. *Elementary Number Theory*. Dover, 1978.

[5] D.S. Dummit and R.M. Foote. *Abstract algebra*. Wiley, 1999.

[6] E.B. Golos. *Foundations of Euclidean and non-Euclidean geometry*. Holt, Rinehart and Winston, 1968.

[7] J.V. Grabiner. *A Historian Looks Back: The Calculus as Algebra and Selected Writings*. Maa, 2010.

[8] D. Kalish, R. Montague, G. Mar, and R.J. Fogelin. *Logic: techniques of formal reasoning*. Harcourt Brace Jovanovich, 1980.

[9] M Kline. *Mathematics: A Cultural Approach*. Addison-Wesley, 1962.

[10] A.B. Manaster. *Completeness, Compactness, and Undecidability: An Introduction to Mathematical Logic*. Prentice-Hall, 1975.

[11] J.E. Marsden and M.J. Hoffman. *Elementary classical analysis*. WH Freeman, 1993.

[12] E. Nagel and J.R. Newman. *Godel's proof*. Routledge, 1971.

[13] H.A. Priestley. *Introduction to complex analysis*. Oxford University Press, USA, 2003.

[14] M. Reed and B. Simon. *Methods of modern mathematical physics*. Academic press, 1980.

[15] E.M. Stein and R. Shakarchi. *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton Univ Pr, 2005.

[16] R.R. Stoll. *Sets, logic, and axiomatic theories*. Freeman, 1961.