Ryerson Polytechnic University

Foundations of Mathematical Thought
MTH 599 Course Notes

by P. Danziger

Contents

Introduction

This course will attempt to convey insight into the way mathematics is done, and the way it developed historically as well as considering some of the main philosophical questions surrounding mathematics. There are three major themes here, the first is the philosophy of mathematics. This includes questions about the nature of mathematics as well as questions arising from mathematics itself. The second theme is historical, how did mathematics develop. Finally there is the question of the application of mathematics to other disciplines. You should keep these ideas in mind as we proceed.

We will start by considering ancient civilisations in Mesopotamia and Egypt, then we will consider the Greeks and the development of deductive logic. We will look at one of the most influential Greek works, Euclid. Euclid's books were standard texts, in one form or another, for over two thousand years. Many of the theorems in geometry you may of learned in High School appear in Euclid (as well as many you didn't). We then jump nearly 2000 years and consider the infinite and the infinitesimal in mathematics. We then move on to the "discovery" of non-Euclidean geometry in the 19th century and the subsequent development of axiomatic theory. We will look at Russell's paradox and Gödel's theorem.

One point that should be made is that it is not really possible to talk *about* mathematics without *doing* some. So this course will include a fair amount of mathematics. I hasten to add though that the mathematics we will be doing bears little relation to the mathematics you may have done in high school.

Chapter 1

What is Mathematics?

1.1 A Definition of Mathematics?

Mathematics has consistently defied definition despite repeated attempts. It is even difficult to classify mathematics in terms of arts and sciences, many universities offer a Bachelor of Arts (B.A) in mathematics, many others a Bachelor of Science (B.Sc.), others (such as Waterloo) avoid the problem by inventing a new designation altogether such as Bachelor of Mathematics.

Most people associate mathematics with science; however while it is true that the sciences rely on mathematics, mathematics itself is not a science. In mathematics there are no experiments and thus it does not use the scientific method. Science deals with the real world and our interpretation of it, while mathematics is a purely abstract endeavour.

There is much speculation on why it is that a purely abstract endeavour such as mathematics should have so many applications in the real world. The usual answer is that mathematics is done by people, and people live in the real world; they form working 'real' models. Indeed until relatively recently most mathematics arose out of answers to physical problems. However, it is certainly possible to come up with completely abstract ideas of mathematics, which have no basis in reality. Since the 19th century this has been the norm for pure mathematics. The strange thing is that the most abstract of 19th century mathematics suddenly turns out to be useful in physical theories of the 20th century. The Physicist Eugene Wigner has called this "The unreasonable effectiveness of Mathematics in the Natural Sciences." It is interesting to speculate on why it is that the universe should obey the ordered laws of mathematics.

Many mathematicians have talked of the aesthetic value, or beauty of mathematics[1]. The delight at seeing a clever or well formed proof, the indescribable feeling of insight and accomplishment at discovering a mathematical truth. This is the side of mathematics that is rarely seen by the outside world, though many people glimpse it in a hard won solution to a high school problem, or a logic puzzle. It is this side of mathematics which explains the popularity of logical puzzles.

Mathematics is often characterised by extreme precision, not so much precision of quantitative measurement – the mathematician doesn't care if something is three inches or three miles – but rather precision of expression and thought. The following joke emphasises this point.

> One day a mathematician, physicist and an engineer go on holiday together to Texas. There they see a field full of black sheep.
>
> "Gee", says the engineer, "I guess all sheep in Texas are black".
>
> "Don't be ridiculous", says the physicist, "we can only conclude that all the sheep in this field are black".
>
> "No my friend, you are wrong" says the mathematician, "all we know is that all the sheep in this field are black *on one side.*"

Broadly speaking mathematics is about abstraction by logical reasoning: finding patterns and order in complex phenomena. Certainly logic and abstraction play a central role in mathematics, however as a definition this is woefully inadequate. It is really too broad since it includes *any* form of logical abstraction.

In addition most abstraction, in fact, comes from an intuitive understanding of a problem; the logical explanations come later. Indeed the role that intuition plays in *doing* mathematics is enormous. We should note here the difference between doing mathematics and being taught mathematics, receiving pronouncements of great truths leads to little insight. Whereas coming to these truths by careful thought and diligence usually means great insight to the problem, and its relevance to other problems, has been gained.

It is also worth noting that the finished mathematical theory often bears little relation to the process that went into developing it.

Mathematics could be seen as a general methodology of problem solving. It is definitely true that one of the great achievements of mathematics is to provide powerful mental tools for tackling problems. Here I do not mean the specific tools of mathematics, but rather the general frame of mind of logical reasoned thought. When recommending mathematics for

[1] See *The Art of Mathematics* by Jerry P. King

study it is not thought that, for example, planar geometry, is in itself a particularly useful subject to know about. But rather that the student should be introduced to the modes of thought and problem solving which arise. It is interesting to note in this context that many of the great philosophers of history were also mathematicians, and those that weren't were usually at least adept in the mathematics of the day. Mathematics provides a wonderful exercise in logical, precise reasoned thought.

Much of this side of mathematics has been completely lost by modern educational systems which merely seek to provide a prospective scientist or engineer with tools, such as calculus. It is abhorrent to many students that they should be asked to think and reason ("Do we have to do *proofs?* "), they expect to be provided with formulas which when applied in the correct manner yield the correct answer. Yet it is undoubtedly one of the great strengths of mathematics that it provides an education in the ideas and methods of problem solving.

Mathematics is generally interested in *form* rather than specifics. Mathematicians are generally uninterested in specific examples (unless they highlight a general point), but rather in the general phenomena. Thus, for example, in Euclidean geometry we may make statements about triangles; however we are completely uninterested in any particular triangle, but rather in the behaviour of triangles in general. Similarly in calculus we define the derivative of a function; an engineer or scientist may be very interested in what the derivative of a particular function is, but the mathematician only really cares about the *notion* of differentiation and what implications it may have to the theory of functions.

Mathematics is the development of an axiomatic system from which theorems are derived by logical rules. This definition again denies the intuitive component of mathematics, however it is generally accepted to be the closest definition we have.

1.2 Mathematical Systems

In fact mathematics can define itself (to some extent). We now look at a mathematical definition of mathematics.

Definition 1.2.1 *An* Axiom *is a rule or law which is assumed to be true.*

We generally chose the minimum number of axioms needed to generate the desired system.

Definition 1.2.2 *A* Rule of Inference *is a rule by which new truths (theorems) may be deduced from old.*

Rules of inference indicate how the objects under consideration relate to one another. The rules of inference are usually themselves stated as axioms, that is rules of inference are a type of axiom.

Definition 1.2.3 *A* Mathematical System *or* Formal System *consists of a set of axioms usually including a set of rules of inference. Any statement which can be derived from these axioms is called a* theorem *or* proposition.

We may use formal logic to aid us in showing how the axioms lead to new theorems of the system. Once a theorem has been 'proved' (shown to be a consequence of the axioms) it may be used to prove subsequent theorems. In this way we may build up complex systems from simple axioms.

Example 1.2.4

We consider strings of 0's and 1's, for example 0100010 or 1101001.

Our 'mathematical system' will allow some strings, but not others. If a string is allowed we call it valid. The valid strings are the theorems of this system.

Axioms:

- A1 The string 1 is valid.
- A2 If x is a valid string then so is $x1$ (x with a one appended).
- A3 The only valid strings are those found by a finite number of application of axioms A1 and A2.

This system consists of one basic axiom (A1) and one rule of inference (A2), the final basic axiom (A3) closes the system, i.e. tells us that A1 and A2 are the only way to generate strings. We can see that:
1 is valid
11 is valid
111 is valid
1111 is valid
etc.

It is intuitively obvious that the set of valid strings (theorems) are all those consisting entirely of 1's. However, this is an intuitive result; we must prove it using logic.

In mathematics we have four central objects: Definitions, Axioms, Propositions and Examples. Generally a mathematical work will start with precise definitions of the concepts involved, followed by the axioms. It will then continue with propositions which are proved using logical reasoning. Examples are often provided to illustrate various aspects of the propositions though they are not strictly necessary. In addition more definitions may occur as needed.

There are in fact two kinds of definitions in a mathematical theory. The first kind are fundamental definitions given at the beginning which set the stage as it where. It is these fundamental definitions which cause the most philosophical problems, we will consider these problems in detail later. The other kind of definitions are those which arise as a result of the theory. As we develop a theory certain concepts or collections of ideas become useful, it is often helpful to give these a name via a definition. While it is possible to define anything in a mathematical system a definition is only useful if the thing exists within the confines of the system. It is not really very useful to define irrational numbers when dealing with a system which contains only 0's and 1's.

Propositions occur in three flavours: Lemmas, Theorems and Corollaries. In general a lemma is a result which is needed in order to prove a theorem, but is not of general interest (in fact some of the most powerful results are often referred to as lemmas). A Theorem is the main result, that which generalises the concept to be investigated as much as possible. A Corollary is a result which is derived directly from a theorem or was proved incidentally in the course of proving the theorem. Often a result which is used in a practical application will be a corollary of a more general theorem.

To continue Example 1:

Definition 1.2.5

- *The* Null String *is the string with no entries (no 0's, no 1's).*
- *A string is called* unary *if it is not the null string and contains only 1's (no 0's).*
- *A string is called* non-unary *if it contains at least one 0.*

Theorem 1.2.6 *There exists a valid unary string.*

Proof:
By axiom 1, 1 is a valid string and it is unary. □ [2]

Lemma 1.2.7 $x0$ *(string x with 0 appended) is not a valid string for any string x.*

Note that we can't assume that every valid string ends in 1 (which intuitively it surely does) since we haven't proved this yet. In fact that is exactly what this lemma says.

Proof:
The only way to produce a new string is to append a 1 at the right (axioms 2 & 3).
If x is the null string $x1$ is the string '1', which is valid by axiom 1.
Thus the rightmost digit must always be a 1. □

[2] The symbol □ has largely replaced the classical Q.E.D. to denote the end of a proof

Lemma 1.2.8 *If x is not the null string and $x1$ (string x with a 1 appended) is valid, then so is x.*

Proof:
We assume that x is a string, which is not the null string.
By axiom 3 the only way that $x1$ could be formed is by application of axiom 2 on x, this is only possible if x was valid in the first place. □

Notice how we must consider the case of the null string separately, since it is not true that the null string is valid (see the next lemma). It is worth noting that it is extremely easy for errors to creep into proofs because there is some special case which must be considered separately.

Lemma 1.2.9 *The null string is not valid.*

Proof:
If the null string where valid, by axiom 3, it would have to either be the string 1 (axiom 1), or have been formed by appending a 1 to some previously valid string (axiom 2), but the null string contains no entries, and so cannot contain a 1. □

Theorem 1.2.10 *All valid strings in system 1 are unary.*

Another way of stating this is that there are no non-unary valid strings.
Proof:
The null string is not valid by Lemma 1.2.9, so any valid string must contain some entries.
Suppose that there where a valid non-unary string. Then there is a valid string containing a 0, x say.
By Lemma 1.2.9 x is not the null string (since x is valid).
Using Lemma 1.2.8 we may obtain valid strings by stripping of any rightmost 1's one at a time.
Thus we may produce a valid string in which a 0 appears as the rightmost digit.
But by Lemma 1.2.7 this is not a valid string and so neither is x. □

This is an example of proof by contradiction. We assume the opposite of what we wish to show, and then show that this leads to a contradiction, and so the original assumption must be false. In fact what we really want to show is the converse of the previous result.

Theorem 1.2.11 *All unary strings are valid*

However, in order to show this we are required to show something is true for an infinite set. To do this we need to use proof by induction. We include the proof for completeness.

Proof:
We proceed by induction on the number of digits in the string.

Base Case: By axiom 1, 1 is a valid string.
Inductive Step: We suppose that for all strings x with n or less digits, where $n \geq 1$, x is valid if and only if x is unary.
Let x be any valid string with n digits, the only strings with $n+1$ digits are of the form $x1$ by axiom 2, which is unary since x was. □

Corollary 1.2.12 *A string is valid if and only if it is unary.*

Proof:
This is an immediate consequence of Theorems 1.2.10 and 1.2.11. □

This seems to be a lot of work to prove something which seemed obvious in the first place, but now we are *sure* that our intuition was correct. It has been said that mathematics is the art of stating the obvious. Of course this is a relatively trivial system and the real power of mathematics comes to light only when we consider more complex systems.

Of course the purpose of this exercise is not to consider unary strings, but to introduce you to the structure of a mathematical work. Notice the structure of the propositions above:
Definition, Theorem, Lemma, Lemma, Theorem, Theorem, Corollary.
Consider how they are related, the lemmas are used to prove the theorems, which in turn are used to prove the corollary. Also note that the final result appears as a corollary rather than a theorem.

Finally note that the mathematical definition of mathematics has something of the same flavour, but only uses definitions and examples.

We will continue our discussion of formal systems and logic later. For now we ask, how did mathematics develop into such an intricate and subtle tool? What were its origins? To do this we start by considering the early history of mathematics.

Chapter 2

Early History

The very earliest use of mathematics was undoubtedly the use of number for counting. Every known language has some representation for number and counting is considered a "cultural norm", i.e. something which every culture has. Indeed it appears that the earliest forms of writing were in fact tallies marked out on wooden sticks. However, as Darwin noted, most of the higher mammals seem to be able to at least conceive of number and differentiate between small sets of different sizes.

Speculation about the earliest development of number is difficult as there is not much information. In addition it is of questionable value as far as mathematics is concerned. We continue on to the earliest civilisations which used writing and for which some historical record exists

2.1 Pre-Hellenic Civilisations

There were four civilisations which developed in fertile river valleys of the ancient world, between roughly 4,000 BC and 2,000 BC. These were the civilisation in the Yangtse valley in China, one in the Indus valley in India, the Mesopotamian (Babylonian) civilisation between the Tigris and Euphrates and the ancient Egyptians along the Nile.

The Mesopotamian and Egyptian civilizations are the ones which have had the most influence on modern mathematics. In addition very little has survived of the early Yangtse and Indus civilisations, though more has come to light recently. Thus we will not consider these civilisations further at this time.

The Egyptians used papyrus scrolls, copies of some scrolls of a mathematical nature have survived to the present day, notably the Rhind or Ahmes papyrus [1]. The Mesopotamians wrote on soft clay tablets which could then be baked for permanent records, the resulting clay tablets were very resistant to the ravages of time and a remarkable number have survived. Both Egyptian hieroglyphs and Mesopotamian cuneiform have been deciphered in the last century, mainly through the discovery of Greek translations written alongside the original script.

2.1.1 The Early Development of Mathematics

It is very difficult at this remove in time to guess at what prompted these early civilisations to develop any sort of mathematical ability. One possible reason is that both the Egyptians and the Mesopotamians where builders, building large and complex structures requires some degree of mathematical ability.

One of the natural features of a river valley is regular flooding, as a result these civilisations had to become adept at time measurement in order to predict the time of the next flood. This resulted in the development of some rudimentary form of astronomy, requiring a mathematical underpinning. The Nile valley is very regular in its flood patterns, flooding at roughly the same time every year. Thus the resulting system was fairly simple, it did however require the measurement of relations between astronomical objects.

The Tigris and Euphrates valley represents a much more complicated river system and flooding was not as regular and predictable. As a result the Mesopotamians developed a much more sophisticated system of time measurement. The remnants of this system are available to us today in the form of astrology. This tends to indicate that the time measurement systems had a mystical as well as practical component.

After each flood the lie of the land would change and a re-division of the land would be in order. It was general practice to ensure that the amount of land owned by any one individual remained more or less constant. This practice was very definitely in use in Egypt, and the resultant area calculations lead to a fairly sophisticated, if somewhat haphazard, formulation of geometry, particularly as regards area calculations. This explanation of the origins of geometry is due to the Greek historian Herotodus.

It seems that mathematics in these societies had a certain mystical flavour, Babylonian astrology is still with us today. This is probably because it was often priests who developed

[1] Henry Rhind discovered it in 1858 (AD), Ahmes copied it in c. 1650 BC from material dating from the middle kingdom c. 2000 to 1800 BC

mathematics. Aristotle (writing many centuries later) in fact ascribes the development of geometry in the Nile delta to the fact that the priests had the leisure time available to develop theoretical knowledge.

Recent work (by M. Ong and others) suggests that literate societies, i.e. those with writing, have a much greater ability in abstract reasoning than purely oral ones. Thus it may have been the invention of writing itself which allowed the development of abstract concepts, as well as allowing ideas to be recorded and built upon by subsequent generations.

2.1.2 Ancient Number Systems

The Egyptians

Our main source of knowledge on Egyptian arithmetical systems comes from the Ahmes papyrus mentioned above. This papyrus begins by stating that the work gives us "The correct method of reckoning, for grasping the meaning of things, and knowing everything that is." The papyrus then gives a table for doubling (or halving) odd fractions (more on this below). The body of the work then gives us 87 solved problems. The first 40 of these problems are essentially arithmetical in nature, the next 20 involve area and volume (geometric) calculations, and the remainder are applications to commerce.

The Egyptians used a base ten number system which was similar to Roman numerals. Indeed Roman numerals are a direct descendent, through the Greeks, of this system. A different glyph was used for 1, 10, 100, etc. up to 1,000,000. These glyphs would then be repeated as necessary to obtain a number. Thus, using roman numerals in place of glyphs, the number 1261 would be MCCXXXXXXII (M = 1,000, C = 100, X = 10, I = 1). This system is workable but becomes cumbersome for large calculations. Note that the use of IX for 9, XC for 90 etc. is a later invention. This sort of notation involves using the order in which a number is written down. For an ancient Egyptian MCCXXII was the same number as CMCIXIX.

Using this type of number system, doubling a number becomes a very easy operation. Simply repeat each numeral and then collect the terms. Thus to double MCCXXXXXXIII, we write MMCCCCXXXXXXXXXXXXIIIIII, collecting terms from the right: MMCCCCCXXIIII (= 2524).

The Egyptians used this ease of doubling to perform multiplicative arithmetic. In order to multiply two numbers together they would double one of the numbers enough times and then add. The following examples show how the method works.

Example 2.1.1 *Find* 123×12.

123	$\times 1$
246	$\times 2$
492	$\times 4$
984	$\times 8$

Now $12 = 8 + 4$, thus $123 \times 12 = 492 + 984 = 1476$.

Example 2.1.2 *Find* 146×15.

146	$\times 1$
292	$\times 2$
584	$\times 4$
1168	$\times 8$

Now $15 = 8 + 4 + 2 + 1$, thus $123 \times 12 = 1168 + 584 + 292 + 146 = 2190$.

This method inherently uses the *distributive law*: $a(b + c) = ab + ac$. However, there is no indication that the Egyptians had an explicit statement of this rule. Similarly, the method actually involves writing the smaller number in binary (sums of powers of 2), but again there is no indication that there was a formalisation of this process.

In order to do division, the same process was used in reverse.

Example 2.1.3 *Find* $1476 \div 123$.

123	$\times 1$
246	$\times 2$
492	$\times 4$
984	$\times 8$
1968	$\times 16$

Now $1968 > 1476$, so we know that the answer must be less than 16.
$1476 - 984 = 492$, and so the answer is $8 + (492 \div 123)$.
From the table $492 \div 123 = 4$, and so the answer is $8 + 4 = 12$.

This method essentially finds those numbers on the right which add up to the number to be divided, the answer is then the sum of the corresponding numbers on the left.

Example 2.1.4 *Find* $1606 \div 146$.

146	$\times 1$
292	$\times 2$
584	$\times 4$
1168	$\times 8$
2336	$\times 16$

Now $2336 > 1606 > 1168$, and so the answer must be between 8 and 16.
$1606 - 1168 = 438$, and so the answer is $8 + (438 \div 146)$.
$584 > 438 > 292$, so the next largest factor is 2.
$438 - 292 = 146$, and so the answer is $8 + 2 + (146 \div 146) = 8 + 4 + 1 = 11$.

This is all very well if the number divides perfectly, but what if there is a remainder? The modern approach would be either to write the fraction and simplify (giving an answer in fractional form) or to continue the division past the decimal point to get an answer in decimal form. However the Egyptians did not represent fractional numbers in the way we do. Indeed it is questionable whether they thought of fractional numbers in anything like the way we do.

In order to consider this case we must consider the way in which the Egyptians represented fractions. The only basic fractional units where those of the form $\frac{1}{n}$, for example $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, etc. They would place a dot on top of the number to indicate that it was fractional. Thus $\frac{1}{2}$ would be represented by $\dot{2}$, $\frac{1}{3}$ would be represented by $\dot{3}$, etc. In addition they had a special symbol for $\frac{2}{3}$, $\ddot{3}$.

Example 2.1.5 *Find* $7 \div 3$

3	$\times 1$
6	$\times 2$
12	$\times 4$
1	$\times \dot{3}$

Note the final line which tells us that $1 = 3 \times \frac{1}{3}$.
As before we find the numbers on the left which add up to the number to be divided, and add up the corresponding numbers on the right to get the answer.
So, since $7 = 6 + 1$, we have that $7 \div 3$ is $2 + \dot{3} = 2\ \dot{3}$.

But what happens when a divisor does not have a remainder of one? In order to deal with this general case it is necessary to know how to multiply and divide fractions by two. Using the dot notation division is easy if the denominator is even. $\frac{1}{2}\ \dot{n} = \dot{(n/2)}$, if n is even. So, for example $\frac{1}{2}\ \dot{12} = \dot{6}$. However, if n is odd $(n/2)$ is itself a fraction and the method fails.

In order to deal with fractions where the denominator is odd the Ahmes Papyrus begins with a long table of conversions for fractions of the form $\frac{2}{n}$, where n is odd. So, for example, the table tells us that a number such as $\frac{2}{5}$ should be represented as $\frac{1}{3} + \frac{1}{15}$, and $\frac{2}{13}$ should be represented as $\frac{1}{8} + \frac{1}{52} + \frac{1}{104}$.

Though experts have seen some patterns in the way parts of this table were put together there seems to be no discernible overall pattern. This system leads to obvious complications when dealing with arbitrary fractions. However, despite its weaknesses, this system was adopted and used successfully by the Greeks.

Armed with this table we can now do arbitrary divisions.
The general algorithm to find $n \div d$ is:

1. First find the non fractional part by the doubling method above.

2. Now create a line of the form $1 \;\; \dot{d}$.

3. Now double this line repeatedly (using the table if n is odd) to get the non fractional part.

It should be noted that while this appears to be the general algorithm used, there is no explanation of this algorithm anywhere in the text. Further ad hoc methods are used whenever the author feels like it.

Example 2.1.6 *Find* $13 \div 5$

5	$\times 1$
10	$\times 2$
15	$\times 4$
1	$\times \dot{5}$
2	$\times \dot{3}\ \dot{15}$ (from the table)

.

Now $13 = 10 + 2 + 1$, so $13 \div 5 = 2\ \dot{5}\ \dot{3}\ \dot{15}$.

Notice how addition of fractions is represented by leaving a space.

Example 2.1.7 *Find* $12 \div 13$

13	$\times 1$
1	$\times \dot{13}$
2	$\times \dot{8}\ \dot{52}\ \dot{104}$ (from the table) .
4	$\times \dot{4}\ \dot{26}\ \dot{52}$
8	$\times \dot{2}\ \dot{13}\ \dot{26}$

Now $12 = 8 + 4$, so $10 \div 13 = \dot{2}\ \ \dot{13}\ \ \dot{26}\ \ \dot{4}\ \ \dot{26}\ \ \dot{52} = \dot{2}\ \ \dot{4}\ \ \dot{13}\ \ \dot{13}\ \ \dot{52} =$ (using the table) $\dot{2}\ \dot{4}\ \dot{8}\ \dot{52}\ \dot{104}\ \dot{52} = \dot{2}\ \dot{4}\ \dot{8}\ \dot{26}\ \dot{104}$.

So the ancient Egyptians would have represented $12 \div 13 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{26} + \frac{1}{104}$.

The Mesopotamians

The Mesopotamians used a number system which is easily recognisable as equivalent to our decimal system, the main difference being that they used a mixture of base 60 and base 10 for their number system. Base 60 is called a hexagesimal system (as opposed to base 16 which is called hexadecimal and is widely used in computer science). By 2,000 BC the use of a special symbol (equivalent to zero) as a placeholder was in regular use.

It should be noted that this is a very sophisticated number system. It subsequently fell into disuse as the Mediterranean world adopted the Egyptian system. It was not until 2,800 years later, around 800 AD that the Hindu civilisation reinvented the use of zero as a placeholder.

In this system the number 1223 would be represented as 20,23 (20 60's and 23). They had an equivalent of the decimal point and could represent fractions in hexagesimal notation, thus $\frac{1}{2}$ would be .30 (30 is half of 60), $\frac{2}{5}$ would be .24 (12 is one fifth of 60, times 2). The use of base sixty makes enormous sense from a mathematical standpoint as 60 is divisible by 2, 3, 4, 5, 6, 10, 12, 15, 20 and 30. This means that the representation of the corresponding fractions $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$ etc. is very simple. On the other hand the Mesopotamians handicapped their system because they would not consider using repeating decimals. Thus they would be unable to represent a number like $\frac{1}{7}$. The vestiges of this system survive today in our use of a base 60 system for measuring time (hours, minutes, seconds) and angle (degrees, minutes of arc, seconds of arc).

The arithmetical and algebraic systems of the ancient Mesopotamians was arguably more sophisticated than those of the ancient Egyptians, though the Egyptians probably had the edge in geometry. The Mesopotamians were able to calculate products reciprocals, squares, cubes, square and cube roots. An existing tablet gives an approximation for $\sqrt{2}$ as $0.42, 30$, $\frac{42}{60} + \frac{30}{3600} \approx 0.7083$, a very good approximation of $\frac{1}{\sqrt{2}} \approx 0.7071$.

Surviving Mesopotamian texts give examples of relatively complex algebraic problems. For example a common exercise gives the sum and product of two numbers, the reader is then asked to find the two numbers. This is actually equivalent to the modern problem of solving a *quadratic* equation $x^2 + bx + c = 0$ in the case where the solution is rational. To see the equivalence let α and β be the solutions of this equation. Then $(x - \alpha)(x - \beta) = 0$. Multiplying out the bracket gives $x^2 + (\alpha + \beta)x + \alpha\beta = 0$, so $a = (\alpha + \beta)$ and $b = \alpha\beta$.

Summary

The ancient civilisations had a fair degree of sophistication when it came to practical specific problems. They were able to solve simple quadratic equations and calculate areas and volumes, they new about π, though their estimates were not enormously accurate. However, they don't appear to have had a systematic approach to solving problems. Most of the surviving texts are lists of problems which are then solved, there is no explicit generalisation (theorems) and no evidence of an explicit logical deductive system. Of course the solutions contain implicit assumptions and generalisations but these are not explicitly stated. It seems that it was required to solve each problem from "first principals" in an *ad hoc* manner.

2.2 The Early Greeks

The great mathematical achievement of the Greeks is summed up in the thirteen books (we would probably call them chapters today) of Euclid and in the logic of Aristotle, around 300 BC. Euclid includes at the beginning of book I his 5 basic postulates:

1. It is possible to draw a straight line from any point to any point.
2. It is possible to produce a finite straight line continuously in a straight line.
3. It is possible to describe a circle with any center and diameter.
4. All right angles are equal to one another.
5. If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles then the two straight lines, if produced indefinitely, meet on that side on which are the angles less than two right angles.

The first two postulates essentially say that we have a ruler with which we can draw a straight line between two points, or extend an existing line. The third proposition states that we have a (collapsing) compass with which we can draw circles. We thus speak of Euclid's constructions as 'ruler and compass' constructions. The third postulate is about the nature of geometric space, and the final one is Euclid's much famed fifth postulate, the parallel postulate.

Euclid also introduces the five Common Notions:

1. Things which are equal to the same thing are equal to one another.

2. If equals are added to equals then the wholes are equal.
3. If equals are subtracted from equals the remainders are equal.
4. Things which coincide with one another are equal to one another.
5. The whole is greater than the part.

From these axioms he proves various propositions, 47 in book I, by means of logic. It should be understood that most of Euclid is a collection of earlier work, and not due to Euclid himself. Though the presentation and many of the details are due to him. As Proclus (410 - 485 AD) says

> "... Euclid, who put together the Elements, collecting many of the theorems of Eudoxus, perfecting many others of Theaetetus, and bringing to irrefragable demonstration the things which had been only somewhat loosely proved by his predecessors"

It should be understood that the ancient Greeks did not have the algebraic machinery available to us today. Most of the early theorems were geometric in nature, though some of them mirror equivalent algebraic results.

The question then arises of how the ancient Greeks developed such a sophisticated system of reasoning about number and geometry. Many of the questions asked by the Greek philosophers of the 3rd and 4th centuries BC are still relevant today. Many of the others where not satisfactorily explained until recently.

The Greek civilisation arose on the shores of the Eastern Mediterranean sometime around 800 BC. They were a vibrant civilization of traders with a well developed civil and literary tradition. The first Olympic games were held in 776 BC and by that time Homer's Iliad was already widespread.

It is important to realise when considering the ancient Greeks that they did not just inhabit what is now modern Greece, but also much of the coastline of what is now Italy and Turkey around the Aegean and Ionian Seas. They were not a unified state, but rather a loose collection of city states with common cultural values.

The Greek city states, by and large, practised a democratic form of government, and are generally hailed as the inventors of democracy. This entailed the notion of *citizenship,* i.e. those who where eligible to vote. In return for the privilege of voting the citizen would agree to fight in the army if necessary and carry out civic duties, which the Greeks held to be very important. On the negative side the Greeks were very elitist and had a slave based economy, generally believing the other civilisations around them were barbarians and beneath contempt.

The Greeks were curious about the nature of the world and began to delve into its subtleties. In order to do this they developed a complex form of logical reasoning. The Greeks also placed great emphasis on mathematics in their world view, seeing it as an ideal abstraction to which the real world strives. In this they realised that mathematics was an abstract endeavour, conforming only to the rules of logic.

2.3 Thales of Miletus

The first Greek mathematician is generally taken to be Thales of Miletus (c. 600 BC). Very little is known about him, none of his works have survived and even references to him are third or fourth hand. However he is generally accredited as being the first true mathematician and the first of the "Seven Wise Men".

Thales travelled in Egypt and Mesopotamia, possibly as a merchant. There he learnt what mathematical lore he could and brought it back with him to Miletus. But that was not the end of it, he then improved on what he had learned and supposedly (according to Aristotle) provided proofs of four geometric theorems, given below.

It is not really clear how much of this material is really due to Thales, and how much he cribbed from the Egyptians, or indeed how much was in fact due to later mathematicians. It seems the ancient Greeks would often attribute things to Thales if all that they knew was that the result was old. It is also not known what might have constituted "proofs".

Thales is also said to have predicted an eclipse of the Sun, generally assumed to be the eclipse of 585 BC, to have measured the heights of the pyramids in Egypt and calculated distances of ships at sea. The most interesting point is that he was applying the same principal in different situations, generalising a result.

Aristotle relates an interesting story about Thales.

> One day Thales was asked by one of his merchant friends what point there was in studying philosophy (philosophy here refers to all of the natural sciences as well as mathematics and, of course, philosophy) and challenged Thales to show what profit there was in it. Thales nodded sagely and said "You will see, I will show you". Puzzled by this cryptic reply his friend went about his business. Thales

then used his deductive powers to predict a particularly good crop of olives, and cornered the market in olive presses, thus securing himself a fortune. The friend was needless to say impressed and immediately took up geometry!

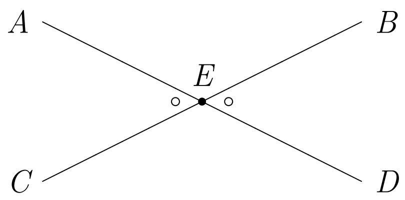
Thales is accredited with founding the Ionian school of mathematics and this center of learning and research flourished in rivalry with the Pythagoreans.

We give proofs of the four theorems usually attributed to Thales here, to illustrate the focus of Greek geometry.

2.3.1 Thales 1st Theorem

The pairs of opposite angles formed by two intersecting lines are equal (Proposition I.15 of Euclid[2]).

Figure 2.1: Opposite angles are equal (Euclid I.15)



If AD and BC are straight lines and E is their point of intersection then $\angle AEC = \angle BED$ and $\angle AEB = \angle CED$ (see figure 2.1).

A proof that $\angle AEC = \angle BED$ might go as follows:
Consider $\angle AEC + \angle AEB = 180^{\text{O}}$, since CEB is a straight line.
So $\angle AEC = 180^{\text{O}} - \angle AEB$
Now $\angle BED + \angle AEB = 180^{\text{O}}$, since AED is a straight line.
So $\angle BED = 180^{\text{O}} - \angle AEB$
Thus $\angle AEC = \angle BED$. $\square$

Our modern algebraic terminology obscures some of the reasoning which underlies this proof. The Greeks did not have such a system and would have to rely on a proof in words, possibly

[2] References to Euclid's propositions usually follow the format of book.proposition, where 'book' is the number of the book in which the proposition appears, in Roman numerals (I - XIII), and 'proposition' is the number of the proposition within that book

with reference to a picture. To see how difficult this can be try to give such a proof without using any symbols.

Underlying this proof is the idea that things which are equal to the same thing are equal to one another, this is the first *common notion* of Euclid's Elements. This proof also uses the second and third common notions implicitly in the algebra.

2.3.2 Thales 2nd Theorem

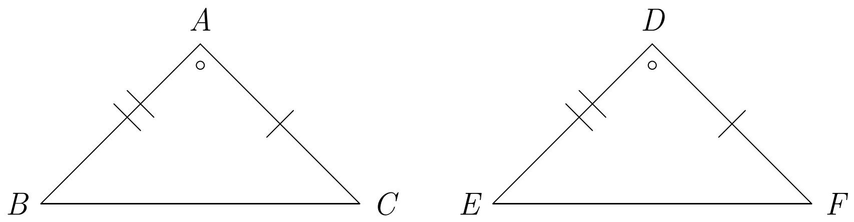
The base angles of an isosceles triangle are equal. (Proposition I.5 of Euclid)

Aristotle gives a proof of this theorem, which is significantly different to that found in Euclid. Aristotle's proof relies on circles and circular angles (angles between a straight line and a circle). While there is nothing wrong with this, it seems unnecessarily complicated. Indeed, the proof given in Euclid is considered to be one of his cleverest proofs. Interestingly in giving his proof Aristotle explicitly states one of the common notions (Common Notion 3).

Since Aristotle predates Euclid, Thales proof (if he did in fact give one) would have been closer to Aristotle's version. The proof given here is that found in Euclid.

Euclid's proof relies on his Proposition I.4, which is a slightly weaker version of the 'side angle side' theorem:
If two triangles have two sides equal to two sides respectively, and the angle contained by these two equal sides is equal, then the triangles are equal and the remaining side will be equal to the remaining side, and the remaining angles will also be equal. (See figure 2.2)

Figure 2.2: Euclid's Proposition I.4



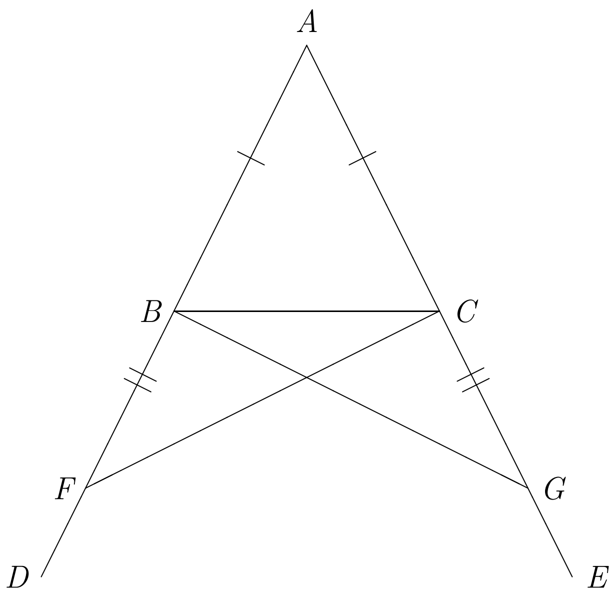
That is, if $AC = DF$, $AB = DE$ and $\angle BAC = \angle EDF$, then $\angle ABC = \angle DEF$, $\angle ACB = \angle DFE$ and $BC = EF$. Thus the triangles are 'equal' (we would say 'congruent' today).

In fact Euclid's proof of Proposition I.4 is flawed, one of the few cases where Euclid does not give an adequate proof. Euclid can hardly be blamed since it turns out that this, or some equivalent of it, must be assumed as a postulate. We will discuss this further when we consider Euclid in detail in Chapter 4. For now we will just assume this proposition.

We will also need Euclid's Proposition I.3:
Given two lines, AB and CD, with the length of AB greater than the length of CD, it is possible to cut off a segment of AB which is of equal length to CD, using only a ruler and compass.

We are now ready to prove Thales 2nd Theorem.

Figure 2.3: The base angles of an isosceles triangle are equal (Euclid I.5)



We start with the triangle $\triangle ABC$, the theorem we wish to prove states that if $AB = AC$ then $\angle ABC = \angle ACB$ (see figure 2.3).

We assume that $AB = AC$.
Extend the lines AB and AC to D and E respectively. (Postulate 2).
Let F be any point on BD.
Let G be the point on CE such that $BF = CG$. (I.3)
Draw the lines CF and BG. (Postulate 1)

Now consider the triangles $\triangle ABG$ and $\triangle ACF$.
These triangles both contain the angle $\angle FAG$ between the two lines AF and AG.

But $AB = AC$ (by assumption) and $BF = CG$ (by construction of G), also ACG and ABF are straight lines, so $AF = AG$.
So by I.4 $\triangle ABG$ is equal (congruent) to $\triangle ACF$.
In particular $\angle ABG = \angle ACF$ and $BG = CF$.

Now consider the triangles $\triangle BCF$ and $\triangle BCG$.
From the argument above $BG = CF$.
Also $BF = CG$ by the construction of G, and BC is common to both triangles.
Thus all the sides of these two triangles are equal and hence all the angles are equal.
In particular $\angle BCF = \angle CBG$.

So we have that $\angle ABG = \angle ACF$ and $\angle CBG = \angle BCF$.
Subtracting likes from likes gives $\angle ABG - \angle CBG = \angle ACF - \angle BCF$.
But $\angle ABG - \angle CBG = \angle ABC$ and $\angle ACF - \angle BCF = \angle ACB$.
Thus $\angle ABC = \angle ACB$. $\square$

2.3.3 Thales 3rd Theorem

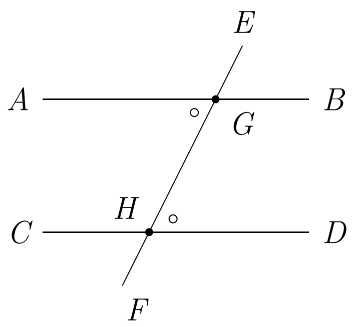
The sum of the angles in a triangle is 180^o (Proposition I.32 of Euclid)

The proof of this theorem is actually quite sophisticated. The reason for this is that this theorem is actually equivalent to Euclid's fifth 'parallel' postulate. That is, in order to prove this theorem we need the fifth 'parallel' postulate, and conversely if we assume this theorem we can prove the fifth postulate. It seems unlikely that Thales would have had such a sophisticated postulate. In fact the version of the postulate as stated in Euclid is the first known occurrence of it in this form.

It is possible that Thales actually assumed this result and used it to prove further things, but this seems unlikely. Perhaps a more likely possibility is that he proved this theorem by assuming another equivalent of the parallel postulate:
Lines which are parallel to the same line are parallel to one another. (Proposition I.30 of Euclid)

We will probably never know how Thales proved this theorem (if he in fact did), but he must have had some equivalent of the fifth postulate in order to do so. Whatever version of the fifth postulate was known to Thales, there are two things which he must have felt comfortable with in order to prove this theorem. Firstly, given a line, ℓ, and a point, A, it is possible to draw a line through the point A parallel to the line ℓ, using only a ruler and compass (Proposition I.31 of Euclid). The second is the 'opposite angles' theorem (Proposition I.29 of Euclid).

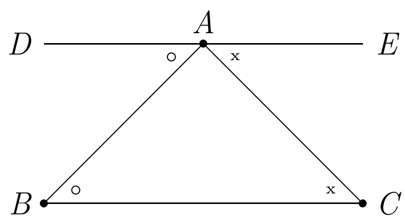
Figure 2.4: The opposite angles Theorem (Euclid I.29)



The *opposite angles* theorem states that if we have two parallel lines, AB and CD, and a third line, EF, which intersects them at G and H respectively (see figure 2.4), then $\angle AGH = \angle DHG$ and $\angle BGH = \angle CHG$.

If we assume these two results the proof that the angles of a triangle add up to 180^{O} can be given as follows (see figure 2.5).

Figure 2.5: The sum of the angles in a triangle is 180^{O} (Euclid I.32)

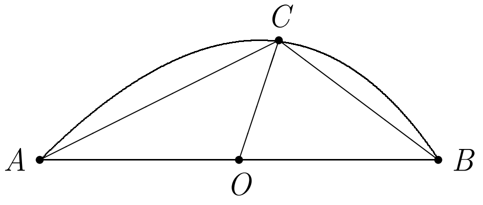


We start with the triangle $\triangle ABC$.
We use our first assumption (I.30) to draw a line, DE, parallel to BC, through A.
Now by the opposite angle theorem, $\angle ABC = \angle BAD$ and $\angle ACB = \angle CAE$.
But DAE is a straight line, so
$180^{\mathrm{O}} = \angle DAE = \angle CAE + \angle BAC + \angle BAD = \angle ABC + \angle BAC + \angle ABC$.
That is $\angle ABC + \angle BAC + \angle ABC = 180^{\mathrm{O}}$. $\square$

2.3.4 Thales 4th Theorem

An angle inscribed in a semicircle is a right angle.

Figure 2.6: An angle inscribed in a semicircle is a right angle



We start with the triangle $\triangle ABC$ inscribed in the circle ACB, O is the center of the circle, and the line AOB is a diameter (see figure 2.6). The theorem states that $\angle ACB$ is a right angle.

Draw the line OC (Postulate 1).
This cuts the original triangle $\triangle ABC$ into two new triangles $\triangle AOC$ and $\triangle BOC$.
Now since OA, OB and OC are radii of the circle they are all equal. Thus these two new triangles are both isosceles.
Thus $\angle OCA = \angle OAC = \angle BAC$, (I.4, Thales 2nd Theorem) call this angle α.
And $\angle OCB = \angle OBC = \angle ABC$, (I.4, Thales 2nd Theorem) call this angle β.
Further $\angle ACB = \angle OCA + \angle OCB = \alpha + \beta$.
Now the sums of the angles in $\triangle ABC$ is 180^o, so $180 = \angle BAC + \angle ABC + \angle ACB$.
Now by Thales 3rd Theorem, $180 = \alpha + \beta + (\alpha + \beta) = 2(\alpha + \beta) = 2\,\angle ACB$.
That is $180 = 2\angle ACB$.
Dividing by 2 gives $90 = \angle ACB$. $\Box$

It is interesting to note that the proof of this result requires both Thales 2nd and 3rd Theorems above.

2.4 Pythagoras and the Pythagoreans

2.4.1 Pythagoras' Theorem

The first thing which should be said about Pythagoras is that he certainly did not invent the theorem which bears his name, it was well known in Egypt and Mesopotamia, not to mention

China. However he is universally accredited with the first proof. Euclid gives Pythagoras' theorem as proposition I.47.

Pythagoras' Theorem (Euclid's Proposition I.47):
In a right angled triangle the square of the hypotenuse is equal to the sum of the squares on the other two sides. (The hypotenuse is the line across from the right angle.)

Here is a proof of Pythagoras' theorem, note that this is not the proof given by Euclid.

Figure 2.7: Pythagoras' Theorem (Euclid I.47)

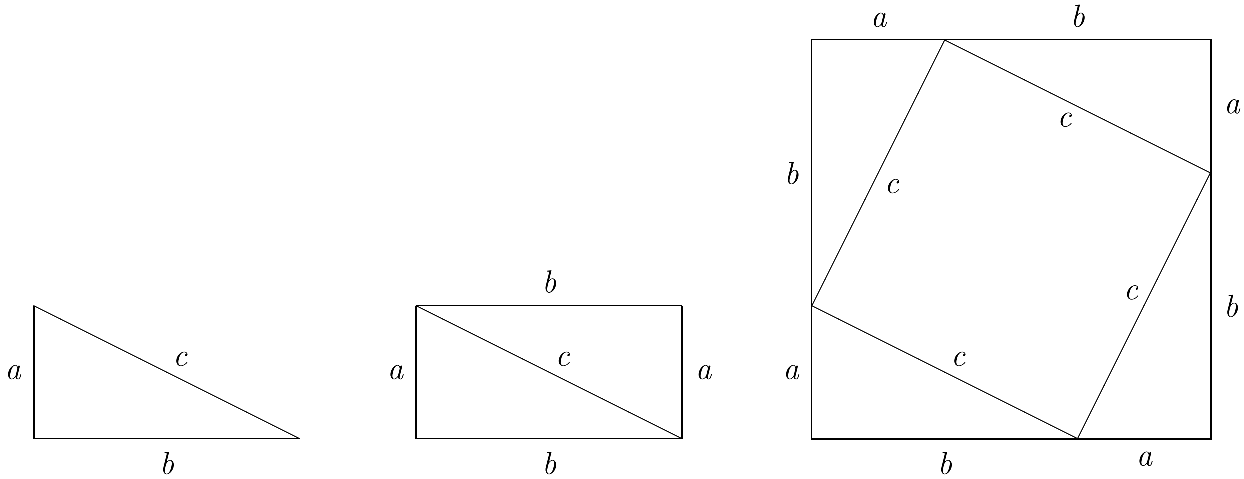
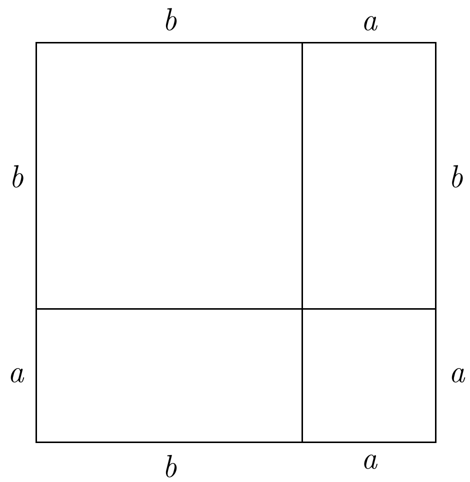


Figure 2.7 (a) Figure 2.7 (b) Figure 2.7 (c)

Consider the triangle with sides a, b and c (figure 2.7 (a)), with c being the hypotenuse.
Now consider figure 2.7 (b), we can see that the area of two a, b, c triangles is the area of the rectangle, that is ab.
Now consider figure 2.7 (c), the large outer square has area $(a+b)^2$.
This square is broken up into a square of area c^2 and four of the original a, b, c triangles.
We know that together these four triangles have area $2ab$.
Thus $(a+b)^2 = c^2 + 2ab$.
If we multiply out the bracket we get $(a+b)^2 = a^2 + b^2 + 2ab$.
Thus $a^2 + b^2 + 2ab = c^2 + 2ab$.
Cancelling the $2ab$'s gives $a^2 + b^2 = c^2$ the required result. $\square$

You may be sceptical of the step of multiplying out the bracket. It is worth remembering that the Greeks did not have our algebraic techniques, in fact they did not use variables as such, and would not have been able to perform this step. However we can prove geometrically that $(a+b)^2 = a^2 + b^2 + 2ab$ using figure 2.8.

Figure 2.8: $(a+b)^2 = a^2 + b^2 + 2ab$



Each side has length $a+b$, so the total area is $(a+b)^2$.
Counting the smaller parts individually we have one square with area a^2, one with area b^2, and two rectangles with area ab.
Since these must represent the same area we have that $(a+b)^2 = a^2 + b^2 + 2ab$. $\square$

This illustrates how algebraic results can be proved geometrically. Many of the Greek arguments where geometric in nature and they placed a great emphasis on geometry. For example rather than thinking of fractions they would consider ratios of lengths of line segments.

But are these really proofs? Do the drawings cover every situation? Can we trust the accuracy of the drawings? Certainly not, every drawing, however carefully rendered, contains some inaccuracies. It is certainly possible that a picture may be misleading, missing some special case. The reliance of the Greeks on geometric intuition is one of the things that was to cause problems for later generations. Indeed the Eleatic school exemplified by Zeno were already questioning the use of geometrical figures in mathematical arguments as early as 450 BC. We will consider the answers given to these questions by Plato and Aristotle later (see sections 3.4.1 and 4.3.2)

2.4.2 The Pythagorean School

Pythagoras was born on the Mediterranean island of Samos sometime around 572 BC This makes him more or less a contemporary of Zorastor, the Persian founder of the Zorastrians, Mahavisra, founder of the Jainists in India and Lau Tzu, founder of Taoism in China. Buddha and Confucius lived only a short time later (c. 500 BC).

In contrast to Thales, very much a practical man, Pythagoras was a mystic. He left his home of Samos purportedly to escape the oppressive rule of the Tyrant Polycrates[3], literally a tyrant. He travelled widely in Egypt, Mesopotamia and possibly even in India. When he returned in 530 BC he settled in the Greek town of Croton in what is now the south-eastern tip of Italy. There he set up a secret mystical society, known as the Pythagoreans. The Pythagoreans remained an influence in Greek intellectual life for over 100 years.

It is often difficult to disentangle what was discovered by who, where the early Pythagoreans are concerned. They had a tendency to ascribe every result to the master. It may be that it was in fact a disciple who first gave the proof of Pythagoras' theorem. In addition the Pythagoreans where a *secret* society. Though they did share much of their material relating to mathematics and Physics, they held two sets of lectures one for initiates and one for everybody else.

The tenets of the Pythagorean cult were based on "Orphic" principles, which was the religion in Pythagoras' homeland of Samos. This was the classic Greek religion with a pantheon including Apollo, Athena, Zeus et. al. The Pythagoreans held Apollo, the god of knowledge, to be supreme.

The major difference from the standard Greek religion was a predominance of the importance of philosophy and mathematics and, above all, number. Pythagoras taught that natural numbers were at the center of all things. Everything, he held, could be explained by whole numbers.

This belief may well stem from the discovery that if a string is held halfway down its length it produces an octave, $\frac{1}{3}$ of the way down a musical third, $\frac{1}{5}$ of the way down a musical fifth and so on. Thus it appeared that the pleasing musical harmonies were indeed related to ratios in whole numbers. Whether this discovery actually led to the central role of number in the Pythagorean school or was just used to support it is not known. However it is interesting to note that after this time music and mathematics were closely related in Greek culture. Of the seven liberal arts recommended for study by Plato the four "mathematical" subjects were arithmetic, geometry, music and astronomy, the other three were grammar, rhetoric

[3] See *Wisdom of the West* by B. Russell

and dialectic.

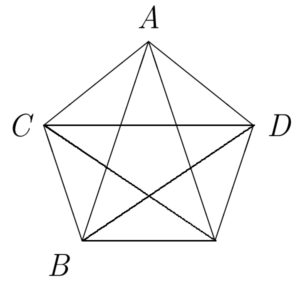
The Pythagoreans developed a sophisticated numerology in which odd numbers denoted male and even female:
1 is the generator of numbers and is the number of reason
2 is the number of opinion
3 is the number of harmony
4 is the number of justice and retribution (opinion squared)
5 is the number of marriage (union of the first male and the first female numbers)
6 is the number of creation
etc.

The holiest of all was the number 10 the number of the universe, because $1+2+3+4=10$. They provided long involved explanations to justify these presumptions. The point is that the explanations were logical, though the premises were suspect.

2.4.3 The Golden Ratio

The Pythagoreans also investigated geometry, their symbol was a pentagon with a five pointed star inscribed in it (see figure 2.9). They noticed that any two lines of the inscribed star intersected in the "golden ratio". In the figure the line AB cuts the line CD in the golden ratio.

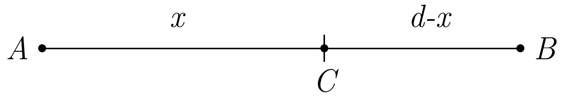
Figure 2.9: The Pythagorean Symbol



Given a line, if it is cut in such a way that the ratio of the smaller piece to the larger is the same as the ratio of the larger to the whole then it is said to be cut in the golden ratio, or golden mean.

In figure 2.10, C cuts the line AB in the golden ratio if $CB : AC = AC : AB$.
If the length of the line AB is d, and the length AC is x, then the length of CB is $d - x$.

Figure 2.10: The Golden Ratio



Turning the ratio into a fraction gives $\frac{d-x}{x} = \frac{x}{d}$.
Cross multiplying yields $d(d-x) = x^2$.
Which simplifies to $x^2 + dx - d^2 = 0$.
We can solve this equation by using the quadratic formula for a solution to $ax^2 + bx + c = 0$,

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

In this case $a = 1$, $b = d$ and $c = -d^2$.
This gives

$$x = \frac{-d \pm \sqrt{d^2 + 4d^2}}{2} = \frac{1 \mp \sqrt{5}}{2} \cdot (-d).$$

The number $g = \frac{1+\sqrt{5}}{2} \approx 1.618034\ldots$ is called the golden ratio, it has the property that $\frac{1}{g} = g - 1$.

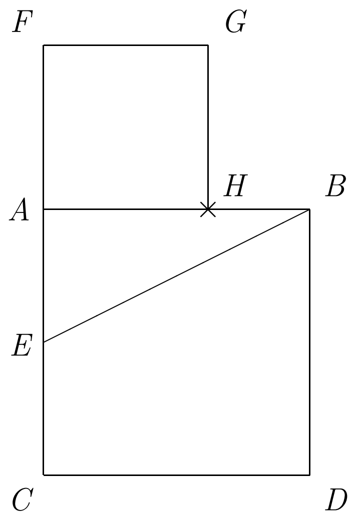
Euclid gives a method for constructing the golden ratio on a line segment AB (see figure 2.11). In order to perform this construction we must assume that given a line segment we can accurately bisect it and construct a square with the given line as one side, using only a ruler and compass. Euclid does indeed provide a method for doing these things (Propositions I.10 and I.46 respectively). In addition we must be able to cut a segment of a given length from a line, this is Euclid's Proposition I.3.

Given the line AB which we wish to divide it in the golden ratio.
First construct the square $ABCD$ (I.46).
Next bisect the line AC at E, i.e. E is halfway from C to A (I.10).
Now draw the line BE (Postulate 1).
Extend the line CA to F so that EF is the same length as EB (I.3).
Form the square $AFGH$ (I.46), H now intersects the line AB at the desired point.

Of course we have not proved that this construction actually gives the desired ratio. To complete the proof, as Euclid does, we would have to show that $HB : AH = AH : AB$.

We have noted that the pentacle has the desired property, why not just create a pentacle on the original line? The reason is that though we can construct a square with a given line as

Figure 2.11: Euclid's Construction of the Golden Ratio



a side we cannot create (by the rules of Euclid) a pentacle with a given line as an edge. In fact Euclid does give a method of creating a regular pentagon in IV.11, however he uses this construction to do so.

2.4.4 Pythagoreans and Number Theory

For obvious reasons the Pythagoreans were very interested in whole numbers, and the relations between them. They made the first inroads into the branch of mathematics which would today be called Number Theory. This material formed much of what became Euclid's books VII VIII and IX. How much of this was known to the Pythagoreans and how much was later material is not now clear.

The Pythagoreans were particularly interested in factors of numbers. As a result they had the notion of prime and composite numbers.

Definition 2.4.1 *(Prime Composite and Divisibility)*

- *A whole number, p, is* prime *if and only if $p > 1$ and for all whole numbers r and s if $n = rs$ then one of r or s is 1.*

- *A whole number, n, is* composite *if and only if there are whole numbers $r \neq 1$ and $s \neq 1$ such that $n = rs$.*

- *If n and d are whole numbers we say that n is* divisible *by d if and only if there is a whole number, k say, such that $n = kd$. We also say that d divides n, or that d is a factor of n. We write $d \mid n$*

So for example $2, 3, 5, 7, 11, 13, \ldots$ are prime, whereas $4, 6, 9, 10, 12, \ldots$ are composite. 2 and 5 are all the factors of 10, so we write $5 \mid 10$, but $3 \nmid 10$.

Prime numbers may be characterised as those numbers whose only factors are one and themselves.

The Pythagoreans were interested in relations between numbers and their factors. They called a number *perfect* if it was equal to the sum of its factors. 6 and 28 are examples of perfect numbers, these are the only examples less than 100.

One surprising thing about the sequence of prime numbers is that there is no known way to generate it short of checking each number in turn. That is we can generate the sequence of even numbers by $a_n = 2n$ for $n \geq 1$, but there is no known similar sequence for the prime numbers. It is this seemingly random distribution which holds part of the fascination of prime numbers. Many modern encryption algorithms work on the principle that it is hard to check whether a number is prime or composite, and if it is composite to find its prime factors.

It is thus reasonable to ask whether the sequence of prime numbers ever terminates. That is: is there a largest prime number? In order to answer this question we will need the following lemma.

Lemma 2.4.2 *For any whole number n and any prime number p, if $p \mid n$ then $p \nmid (n+1)$*

Proof: This is a proof by contradiction, so we assume the opposite and show that this leads to a contradiction.
Thus let n be some whole number, and p a prime such that $p \mid n$ and $p \mid (n+1)$.
Thus, by the definition of divisibility, there are whole numbers r and s such that $n = pr$ and $n + 1 = ps$.
Thus $1 = (n+1) - n = ps - pr = p(s-r)$
So $1 = p(s-r)$, that is $p \mid 1$. But the only divisor of 1 is 1.
Thus $p = 1$, but p is prime, so $p > 1$ which is a contradiction. $\square$

Theorem 2.4.3 *There are an infinite number of prime numbers.*

Proof:
This is a proof by contradiction, so we assume the opposite and show that this leads to a contradiction.
So, suppose not, that is suppose there are a finite number, N, of prime numbers. We will

enumerate them and denote them by $p_1, p_2, p_3, \ldots, p_N$, where p_N is the largest prime.
Now consider the number $M = (p_1 p_2 p_3 \ldots p_N) + 1$.
It is evident that $M > p_N$.
So M cannot be prime, since p_N is the largest prime.
Thus M must have prime factors, since it is composite.
But from the way it was constructed and the previous lemma it is evident that none of p_1 to p_N divide M.
This is a contradiction, and so the premise must be false. $\Box$

This theorem and its proof is one of the two classic theorems of antiquity, the other being the proof that the $\sqrt{2}$ is irrational (Theorem 2.4.13).

Another area of interest was notion of primality relations between numbers.

Definition 2.4.4 *(Relatively Prime, gcd)*

- *Two numbers n and m are* relatively prime *if and only if they share no common factors other than one.*

- *The* greatest common divisor *of two whole numbers is the largest whole number which divides them both.*

We usually write $\gcd(a, b)$ to denote that greatest common divisor of a and b. It is clear that a and b are relatively prime if and only if their gcd is 1.

So, for example gcd(3,6) = 3, $\gcd(10, 20) = 10$, gcd(5, 26) = 1, thus 5 and 26 have no common factors other than 1, and are thus relatively prime.

Note that if $d = \gcd(a, b)$ if and only if d satisfies two properties:

1. $d \mid a$ and $d \mid b$.

2. d is larger than any other number which divides both a and b.

Lemma 2.4.5 *If $d \mid a$ and $d \mid b$ then $d \mid a + b$ and $d \mid b - a$.*

Proof:
Let a and b be whole numbers and suppose that d is a whole number which divides both of them. That is $d \mid a$ and $d \mid b$.
Thus, by the definition of divides, there exist whole numbers r and s such that $a = dr$ and $b = ds$.
Now $a + b = dr + ds = d(r + s)$.
$r + s$ is a whole number since both r and s are and whole numbers are closed under addition.

Taking $n = r + s$, we get $a + b = dn$ for some number n. i.e. $d \mid a + b$.
The proof for the case $b - a$ is similar and left as an exercise. $\square$

Note that the converse of this theorem is not true. That is if $d \mid a+b$ it does not follow that $d \mid a$ and $d \mid b$. To see a counterexample take $a = 7$ and $b = 9$, then $a + b = 16$, so $4 \mid a + b$, but 4 does not divide either 7 or 9.

Theorem 2.4.6 *Given two whole numbers, a and b, with $b > a$ then* $\gcd(a, b) = \gcd(a, b-a)$.

Proof:
Let a and b be two whole numbers with $b > a$.
Let $d = \gcd(a, b)$.
We must show that $d \mid a$ and that $d \mid b - a$.
Since $d = \gcd(a, b)$, $d \mid a$ and $d \mid b$, and so by the previous lemma $d \mid a - b$.

It remains to show that there is no larger common divisor.
We have that $d \leq \gcd(a, b-a)$ since d divides both a and $b-a$, and gcd is the largest number which does this.
But any number which divides a and $b-a$ also divides their sum $a+b-a = b$, by the lemma.
Thus if there is a larger common divisor of a and $b - a$ it is also a larger common divisor of a and b, which contradicts the assumption that $d = \gcd(a, b)$. $\square$

This theorem is the basis of the well known *Euclidean Algorithm* for finding the greatest common divisor of two numbers. We need one more lemma before we can state the algorithm.

Lemma 2.4.7 *For any whole number n gcd$(0, n) = n$.*

Proof:
We must show that $n \mid 0$.
That is we must show that there exists a whole number k such that $0 = nk$.
Take $k = 0$ to get the result.

We must also show that $n \mid n$.
That is we must show that there exists a whole number k such that $n = nk$.
Take $k = 1$ to get the result.

Finally we must show that n is the largest whole number which divides both itself and 0.
But the largest whole number which can divide a number is itself. $\square$

Euclidean Algorithm
We are given two numbers a and b, with $b > a$ and asked to compute $\gcd(a, b)$.

$\gcd(a, b)$
    If $a = b = 0$, (no gcd) return ERROR.

If $a = b$ return a.
If $a = 0$ return b.
If $b = 0$ return a.
If $a > b$ return $\gcd(b, a - b)$.
Else return $\gcd(a, b - a)$.

Consider finding gcd(124, 220)

$a = 124$	$b = 220$	$220 - 124 = 96$	return gcd(124, 96)
$a = 124$	$b = 96$	$124 - 96 = 28$	return gcd(96, 28)
$a = 96$	$b = 28$	$96 - 28 = 68$	return gcd(68, 28)
$a = 68$	$b = 28$	$68 - 28 = 42$	return gcd(42, 28)
$a = 42$	$b = 28$	$42 - 28 = 14$	return gcd(28, 14)
$a = 28$	$b = 14$	$28 - 14 = 14$	return gcd(14, 14)
$a = 14$	$b = 14$	—	**return 14**

This algorithm works because of Lemma 2.4.7 and Theorem 2.4.6.

This is an example of a recursive algorithm, that is it calls itself repeatedly until the answer is found.

We now turn our attention to composite numbers. The Pythagoreans also were interested in different types of 'compositeness'. They had the notion of odd and even, those numbers which are divisible by two, and those which are not. The modern definition of these concepts differs very little from those known to the Pythagoreans.

Definition 2.4.8 *(Even, Odd)*

- *A whole number n is* even *if and only if it can be expressed in the form $n = 2m$ for some whole number m.*
- *A whole number n is* odd *if and only if it can be expressed in the form $n = 2m + 1$ for some whole number m.*

The Pythagoreans went further and defined what they called evenly-even and oddly-even numbers.

Definition 2.4.9 *(evenly-even, oddly-even)*

- *A whole number is* evenly-even *if and only if it is divisible by 4.*
- *A whole number is* oddly-even *if and only if it is even but not divisible by 4.*

Thus $4, 8, 12, 16, \ldots$ are evenly-even, whereas $2, 6, 10, 14, \ldots$ are oddly-even.

These are special cases of what we would call today *modular classes*.

Definition 2.4.10 *(mod)*

- *Given two whole numbers a and b, $a \bmod b$ is the remainder when a is divided into b*
- *Two numbers, a and b, are said to be in the same class modulo n if and only if $a \bmod n = b \bmod n$.*

For example $1, 3, 5, \ldots$ (the odd numbers) all are 1 modulo 2, i.e. they have remainder 1 when divided by two.

$2, 4, 6, \ldots$ (the even numbers) all are 0 modulo 2, i.e. they have remainder 0 when divided by two.

$4, 8, 12, \ldots$ (the evenly-even numbers) all are 0 modulo 4, i.e. they have remainder 0 when divided by 4.

$2, 6, 10, \ldots$ (the oddly-even numbers) all are 2 modulo 4, i.e. they have remainder 2 when divided by 4.

This suggests two other classes $1, 5, 9, 13, \ldots$ which are those numbers which are 1 modulo 4 and $3, 7, 11, 15, \ldots$ which are those numbers which are 3 modulo 4.

We can define modular classes for other numbers:
$1, 6, 11, 15 \ldots$, those numbers which are 1 modulo 5.
$3, 6, 9, 12 \ldots$, those numbers which are 3 modulo 6.
And so on.

2.4.5 Figurate Numbers

The Pythagoreans developed a geometric form of counting which involved placing stones in a geometric pattern. Thus they developed the idea of *triangular numbers*. Place stones regularly in a triangular formation of n tiers, each tier one greater than the last then the total number of stones is a triangular number (see figure 2.12). This is summed up in the formula:

$$N = 1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}.$$

Where N is the total number of stones

Figure 2.12: Pythagorean Triangular Numbers

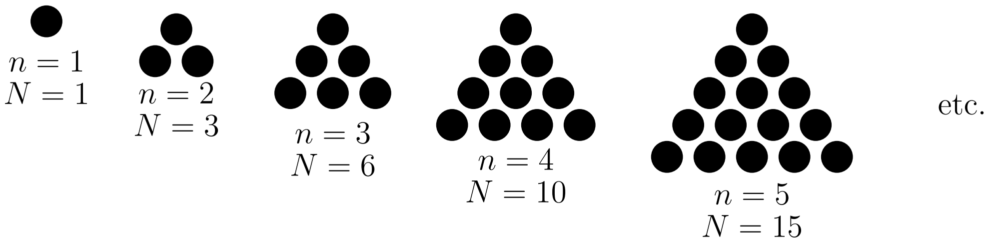


Figure 2.13: The Figurate Numbers

Name	Sequence	Sum N
Triangular Numbers	$1+2+3+\ldots+n$	$\frac{1}{2}n(n+1)$
Square Numbers	$1+3+5+\ldots+(2n-1)$	n^2
Oblong Numbers	$2+4+6+\ldots+2n$	$n(n+1)$
Pentagonal Numbers	$1+4+7+\ldots+(3n-2)$	$\frac{1}{2}n(3n-1)$
Hexagonal Numbers	$1+5+9+\ldots+(4n-3)$	$n(2n-1)$

They also developed number of similar systems for other figures, these are summarised in figure 2.13.

In each case the next number indicates how many stones must be added to maintain the figure but one unit larger. As an exercise see if you can construct the square numbers.

It is interesting to note that this method of geometrical counting shows how the Pythagoreans thought of things in units. Points are seen as having extent, fleshing out a figure as they are added.

This is the first known instance of the investigation of a series. Notice that the formulation is purely geometrical, the Greeks could not express the results in terms of the formulas given above, but they probably knew how to generate these numbers geometrically.

2.4.6 The Later Pythagoreans

The Pythagoreans were, to say the least, controversial. They took an active interest in local politics, as secret societies are wont to do. This was not appreciated by the locals, and in 510

BC Pythagoras was exiled from Croton. He then went to Metapontion where he remained until his death around 500 BC.

At about this time a rival sect, from Sybaris, surprised and murdered many of the Pythagorean leaders. The survivors dispersed throughout the Greek world, setting up centers of Pythagorean activity wherever they went. They carried their mathematical ideas with them and stimulated much interest in mathematics and philosophy, if only in order to refute their arguments.

One branch of the sect moved to Tarentum (now Taranto in southern Italy) where its main disciple, Philolaus, wrote a book on the tenets of Pythagoreanism. It is from this book that most of our knowledge of the Pythagoreans comes, through Aristotle. A student of Philolaus, Archytas (born c. 428 BC), as well as being a mathematician, rose to become a popular, kind and gentle leader of the city, he was consistently re-elected. It seems that the original political conservatism of the Pythagoreans had softened by this time. Archytas was a friend of Plato's and it was Archytas who introduced Plato to mathematics.

The fifth century BC is sometimes referred to as the Heroic age of mathematics. The rules of deductive logic where developed, most, if not all, of the material covered in the first two books of Euclid was proven. The philosopher became a part of Greek life, often they were involved in politics. Many of the great leaders of the Greek world had as tutors some of the greatest minds. The free thinking attitude of the philosophers often got them into trouble and quite a few of them spent some time in prison for heresy of one sort or another.

By the middle to end of the fifth century BC the Pythagorean sect itself had fallen into decline, though many of its ideas, stripped of their religious zeal, were still current. A number of Greek philosophers of the fifth century started out as Pythagoreans and then struck out on their own, when they found the dogma too restricting. One such was based in Elea in what is today Southern Italy, Parmanides of Elea (born c. 520 BC) was originally a Pythagorean, but he later rejected the doctrine and wrote a scathing refutation of Pythagoreanism. He went on to found his own school, the Eleasts. One of his students was Zeno (born c. 490), still famous today for his paradoxes, which we will consider later (see section 2.5).

2.4.7 Irrational Numbers

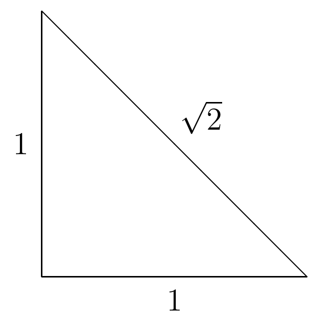
The final decline of the Pythagoreans was sealed by the discovery of incommensurate ratios, what we would call today irrational numbers. This discovery is usually ascribed to Hippasus around 420 BC. Originally a Pythagorean he later left the sect, reputedly after publishing his scandalous discovery against the wishes of the order, who hoped to keep the existence of irrational numbers a secret. When Hippasus was drowned in a shipwreck the Pythagoreans claimed divine retribution, some say that the Pythagoreans had him murdered.

A number is irrational if it is not expressible as the ratio of two whole numbers, that is a fraction. It is clear why this would have horrified the Pythagoreans. Here were numbers which denied the basic faith that whole numbers where supreme. If it is not possible to express irrational numbers in terms of whole numbers, how can whole numbers be the foundation of everything?

The problem was however even deeper than that. Many of the proofs of this period relied on lines being "commensurable". So great was the belief in whole number nobody had questioned this. The discovery of incommensurate (irrational) quantities threatened to invalidate many of the geometrical constructions of the day.

It is ironic that the simplest example of an irrational number comes from Pythagoras' theorem itself. Suppose we have a right angle triangle with the two shorter sides both being one unit in length. Then the hypotenuse must have length equal to a number whose square equals $1^2 + 1^2 = 2$, that is $\sqrt{2}$ (see figure 2.14).

Figure 2.14: The Irrational Triangle



We now provide a proof that $\sqrt{2}$ is irrational, that is cannot be represented in the form $\frac{a}{b}$ for two whole numbers a and b.

To see that $\sqrt{2}$ is irrational we must first realise that fractions represent an *overprescribed* system. That is the same number has many different representations as a fraction. So, for example, $\frac{1}{2}, \frac{2}{4}, \frac{3}{6}, \ldots$ all represent the same number.

Definition 2.4.11 *A fraction is in* lowest common form *if the denominator and the numerator have no common divisors (i.e. there is no number which divides both).*

Thus $\frac{1}{2}, \frac{2}{5}, \frac{3}{10}$, etc. are all in lowest common form, whereas $\frac{3}{6}, \frac{4}{10}, \frac{30}{100}$ etc. are not.

Note that a fraction $\frac{a}{b}$ will be in lowest common form if and only if $\gcd(a, b) = 1$. Given a fraction $\frac{a}{b}$, $\dfrac{(a/\gcd(a,b))}{(b/\gcd(a,b))}$ will be in lowest common form.

We will need the following lemma.

Lemma 2.4.12 *If n^2 is even, then so is n.*

Proof:
We consider the cases of n modulo 2.
If n is even then there exists a whole number m such that $n = 2m$, by the definition of even (Definition 2.4.8).
Thus $n^2 = (2m)^2 = 4m^2 = 2k$ for some whole number k ($k = 2m^2$).
Thus n^2 and n are both even and the lemma is true.

If n is odd then there exists a whole number m such that $n = 2m + 1$, by the definition of odd (Definition 2.4.8).
Thus $n^2 = (2m + 1)^2 = 4m^2 + 4m + 1 = 2k + 1$ for some whole number k ($k = 2m^2 + 2m$).
Thus n^2 and n are both odd and the lemma is true. $\square$

Theorem 2.4.13 *$\sqrt{2}$ is irrational. That is cannot be represented in the form $\frac{a}{b}$ for two whole numbers a and b.*

Proof:
This is a proof by contradiction, so we suppose that the result is false and derive a contradiction.
Suppose not, that is suppose that $\sqrt{2} = \frac{p}{q}$ (*) for some whole numbers p and q. Further we suppose that p and q have no common divisors, that is that $\frac{p}{q}$ is in lowest common form.
Now squaring (*) gives $(\sqrt{2}\,)^2 = 2 = \left(\frac{p}{q}\right)^2 = \frac{p^2}{q^2}$
So $2q^2 = p^2$ (**), thus p^2 is even.
But if p^2 is even then so must p be, by Lemma 2.4.12.
So $p = 2m$ for some number m, by the definition of even (Definition 2.4.8).
But this means that $p^2 = (2m)^2 = 4m^2$.
Thus $2q^2 = p^2 = 4m^2$, by (**).
Dividing through by 2 gives $q^2 = 2m^2$. So q^2 is even.
But if q^2 is even then so must q be, by Lemma 2.4.12.
But now both p and q are even, and hence divisible by 2, so they have a common factor and $\frac{p}{q}$ is not in lowest common form as assumed and we have a contradiction - there is no lowest common form for the fraction. $\square$

The idea that there is no common lowest form led the Pythagoreans to try to rescue the situation by the use of *continued fractions*. So for example

$$
\begin{aligned}
1+\frac{1}{1} &= 2\\
1+\cfrac{1}{1+\cfrac{1}{1}} &= \frac{3}{2}\\
1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1}}} &= \frac{5}{3}\\
1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1}}}} &= \frac{8}{5}\\
1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\dots}}}} &= \frac{1+\sqrt{5}}{2}
\end{aligned}
$$

However this just pushes the problem to that of the meaning of an infinite fractional sequence, intuitively similar to the idea of a limit in modern mathematics. Indeed the real point about irrational numbers is their relation to the infinite and infinitesimal. The Pythagoreans thought of number in very concrete terms, they conceived of each number as having extent, a breadth and width, as taking up space. This kind of thinking is exemplified in the stones method of counting cited above, in which the stones take up space to form a geometrical figure.

Irrational numbers on the other hand represent numbers with no extent, they are "infinitesimally" close together. Indeed the ancient Greek word *irrational* translates to *measureless*, not *bereft of reason* as we might assume. This change in meaning is directly attributable to the discomfort later philosophers felt when faced with irrational numbers.

The real confusion here is that between discrete quantities (whole numbers) and continuous quantities (such as lines). This dilemma was to haunt mathematics until the 19th century. It is very hard to avoid contradictions and self referential statements when talking about continuous quantities. At the core of the problem is that it is very hard to clearly describe exactly what we really mean by the word "infinitesimal". These problems were only resolved in the 19th century, we will discuss this further later.

The Pythagoreans, despite some questionable doctrines, were crucial in the introduction and popularisation of the deductive method and logical thought throughout the Greek world. They developed some far reaching ideas and represented a quest for knowledge about the nature of the world in which we live.

2.5 Zeno of Elea and his Paradoxes

Zeno of Elea (fl. 450 BC) was a student of Parmenedes who had rejected the Pythagorean doctrine. This rejection was carried to its logical extreme by Zeno around 450 BC. Zeno put forward four paradoxes, all related to the nature of space, time and motion.

Zeno seems to have been the first to use paradoxes to refute arguments. Though earlier philosophers had used *reducto ad absurdium* arguments, essentially proof by contradiction, Zeno used paradoxes. In a proof by contradiction we assume a premise and show that it leads to a false conclusion. Zeno would not only show that his opponents argument led to a false conclusion, but that its negation also leads to a false conclusion. This means that the set of presumptions are not just untrue, but impossible, they make no sense at all. Here we make the distinction between something which is merely false and something which is both true and false, paradoxical. We will discuss this distinction further when we consider logic.

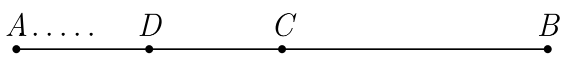
Zeno's aim was to show that the Pythagorean doctrine of number as supreme was untenable. In the Pythagorean view lines would consist of a discrete number of points, each having some extent. In the second two paradoxes, the Arrow and the Stade, Zeno shows that such a model denies the possibility of motion. Perhaps more worrying, in his first two paradoxes he argues that motion is also impossible in a continuous notion of geometry, that is one in which lines consist of a continuously infinite number of points, which is the modern view.

2.5.1 Zeno's Paradoxes

Zeno's first Paradox: The Dichotomy

Consider a racehorse moving in a straight line from the starting gate A, to the finishing line, B. Before the horse can reach B it must reach the halfway point between A and B, C say. But before it reaches C it must reach the halfway point between A and C, D say. But before it reaches D ... And so on (see figure 2.15).

Figure 2.15: Zeno's Dichotomy Paradox



The point here is to demonstrate that in order to travel from A to B we must traverse an infinite number of points in a finite time. Thus motion is impossible.

Zeno's Second Paradox: Achilles

Achilles, famed for his speed, is racing against the tortoise, who is not. To make the race fairer the tortoise has been given a head start. Suppose that the tortoise starts a given distance down the track. Now the race begins and at some point in time Achilles reaches the point where the tortoise started. But during that time the tortoise will have moved, however slowly, to a new point further along. So Achilles runs on to this new point only to find that the tortoise has moved just that little bit further down the track. And so it goes, Achilles can never catch the tortoise no matter how fast he runs.

Zeno's first two paradoxes argue that motion is impossible if one assumes infinite subdivisibility of space. The last two argue the opposite, that if space can be subdivided into finite units, then motion is again impossible.

Zeno's Third Paradox: The Arrow

This paradox can be hard to visualise, as we are not used to thinking of space as discrete.

Consider an (incompressible) arrow in flight, at any point in time it occupies a space "equal to itself", it is then at rest, how then is it to move forward? It would have to jump to the next point. Thus motion is impossible.

Zeno's Fourth Paradox: The Stade

This paradox relies on both time and space being discrete.

We consider three lines, ℓ_1, ℓ_2 and ℓ_3 say. By the assumption of discrete space each of these lines consist of some discrete number of points as indicated. The lines start as shown so that their points are lined up vertically. We suppose that ℓ_1 is moving to the right and that ℓ_3 is moving to the left, while ℓ_2 is stationary (see figure 2.16).

Since time is discrete, speed can be measured as the number of points we move past in a unit of time. We suppose that the lines ℓ_1 and ℓ_3 are moving as slowly as possible relative to the line ℓ_2. That is, they move past one point on ℓ_2 in each unit of time. So after one unit of time the picture becomes that given in figure 2.17.

Figure 2.16: Zeno's Stade Paradox – Initial State

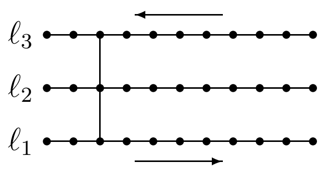
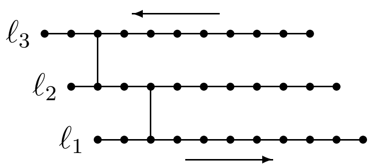


Figure 2.17: Zeno's Stade Paradox – After One Unit Time



But now the line ℓ_1 has moved two units along ℓ_3, so we can define a smaller unit of time, namely that which it takes ℓ_1 to cover one unit along ℓ_3, this will be half our original unit time.

Note that this also destroys the unital property of space, since with the new unit of time, moving as slowly as possible, ℓ_1 and ℓ_3 will both correspond to a place on ℓ_2 where there is no point. We have to double our number of points, but this puts us back where we started.

This paradox is particularly interesting in that it conveys the connection between time, space and motion. In particular, time is defined in terms of motion, this is the underlying principle in Einstein's theory of Relativity (1905).

2.5.2 Zeno's Paradoxes Explained

The third and fourth paradoxes, the Arrow and the Stade, are not too worrying as we can quite happily conclude that the premise about time and space being discrete is incorrect. However the first two paradoxes, the Dichotomy, and Achilles seem to be saying that motion is impossible in a continuously infinite space as well.

The point here of course is that the sequences are infinite, but diminishing. Thus in the Dichotomy the distance to be travelled decreases by half and, assuming a constant speed, can be covered in half the time. The time taken to cover the whole distance is $\dfrac{1}{2^0} = 1$, half the distance $\dfrac{1}{2^1} = \dfrac{1}{2}$, one quarter the distance $\dfrac{1}{2^2} = \dfrac{1}{4}$, etc. Again we are dancing around the idea of a limit, in modern notation we could say that $\lim_{n\to\infty} \dfrac{1}{2^n} = 0$.

In Achilles a similar argument prevails. The distance between Achilles and the Tortoise is constantly diminishing and so is the time taken to cover it. At first it seems that this wont help until we realise that the limits of both converge. So though it will take an infinite number of steps to reach the tortoise the amount of time taken will be finite.

Suppose that the tortoise starts off with a 10 km head start, and Achilles runs 10 times as fast as the tortoise. We take the unit of time to be the time it takes for Achilles to get to the tortoises original position. By this time the tortoise has travelled 1 more km. It takes Achilles $\dfrac{1}{10}$ units time to get to this new position. But now the tortoise has covered $\dfrac{1}{10}$ of a km, or 100m. It takes Achilles $\dfrac{1}{100}$ units time to cover this distance, and so on. In general it will take Achilles $\dfrac{1}{10^{n-2}}$ units time to cover the nth interval. If we sum up all these times we get

$$\sum_{n=1}^{\infty} \frac{1}{10^{n-1}} = \frac{19}{9}$$

(This sum is a geometric sequence and can be solved using the formula $\sum_1^n r^n = \dfrac{1-r^n}{1-r}$ and taking the limit as $n \to \infty$, using $\lim_{n\to\infty} r^n = 0$ if $0 < r < 1$, in this case $r = \frac{1}{10}$)

Notice how the solution again implicitly uses the idea of an infinite limit, this time we have an infinite series, also a limiting process. In fact what Zeno's paradoxes do is to highlight the similarity between infinity and infinitesimals.

However, the ancient Greeks did not have the machinery of limits available to them. Any argument that relied on the infinite subdivision of intervals, or indeed any geometric quantity, had to be ready to explain away Zeno's paradoxes. Thus, for example, in statements of geometry the later Greeks avoided implying in any way that a line was made up of an infinite number of (indivisible) points.

It is possible that the arguments of Zeno dissuaded the Greeks from inventing Calculus, since to do Calculus one needs the concept of a limit, but without a precise definition you must confront Zeno. The definition of a limit is no easy feat, when modern mathematics invented calculus the idea of a limit was at first accepted more or less on faith. It took almost 200 years to come up with a precise definition.

Zeno was also instrumental in questioning the use of figures in geometrical arguments. How can a picture be accurately rendered? Can any information really be gleaned from such a drawing? Does it really have any generality? What if we have forgotten or misrepresented some possibility in our picture?

Plato provides something of an answer arguing that the figures are merely guides, an aid to intuition. When one draws a triangle it is understood to represent an ideal triangle, which can have no reality. This argument is fine, so long as there is not too much reliance on the pictures. Euclid provides a more pleasing response, by saying that we can draw a perfect line because the axioms of geometry, in effect, say that we can. We will consider these ideas in more detail when we cover these two men.

Chapter 3

The Athenians

3.1 Athens

The Greek world at the beginning of the 5th century BC was a loose collection of city states. Around 490 BC the Persians invaded the Greek mainland, they where held off by Athenian forces at Marathon. This invasion forced the Greeks into a tighter association called the Delian league, based on the island of Delos. By 479 BC the league had defeated the Persians. The brunt of the fighting had been born by Athens, the city was a ruin. The Athenians felt that as they had taken the lead in war, now they should take the lead in peace. They argued, successfully, that the treasury at Delos should be used to rebuild their city. It was at this time that the Parthenon was built in Athens.

Over time the Delian league eventually transformed itself into the Athenian empire and the treasury moved from Delos to Athens. It should, however, be understood that the basic city state was still the model and Athens was more first among equals. Other Greek states also flourished, notably Sparta, Corinth and Thebes. It has been suggested that it was the dichotomy between the unity of Greek culture and the strife between the various constituent states which led to such a vibrant culture in ancient Greece.

The Athenian empire flourished, largely under the rule of Pericles (475 ? - 429 BC). Though it was a democracy it was effectively run by this single enlightened man, who was constantly re-elected as *general* from 443 BC until his death from a plague in 429 BC which wiped out a quarter of the population of Athens. Pericles was well educated and came from an aristocratic family. His tutor was one of the first of the Athenian philosophers, Anaxagoras (? - 430 BC). Anaxagoras came from Clazomanae on the Ionian Sea. There is a story which relates how Pericles rescued his mentor from prison after he had run afoul of local religions by claiming that the Sun was not a God and that the Moon was but another Earth which reflected light from the sun, remarkably astute for his time.

Around 440 BC the Athenians restricted citizenship to only those born in Athens of Athenian parents. This caused a lot of bad feeling among the disenfranchised. In addition Sparta had become jealous of Athenian ascendancy. This lead to the Peloponesian wars (431 - 404 BC) which ended with the defeat of Athens.

There was a flourishing of art and literature as well as philosophy in and around Athens at this time. It is from this period, and slightly later, that we get Pheidias the sculptor who adorned the Acropolis, the playwrights Sophocles (496 - 406 BC), Euripides (480 - 406 BC) and Aristophanes (440 ? - 385 BC). Herodotus the historian (486 ? - 429 BC), from whom much of our knowledge of the ancient world comes, was also a native of Athens at this time.

3.2 Hippocrates of Chios and Quadrature

(see *Journey through Genius* by William Dunham)

Hippocrates of Chios was a mathematician who moved to Athens some time before 430 BC. He should not be confused with the physician Hippocrates of Cos, from whom we derive the Hippocratic oath taken by doctors. Hippocrates of Chios is said to have started life as a merchant, however he lacked any sort of business skills and he was financially ruined by pirates who took him for an easy mark. After this he moved to Athens to try his hand at teaching and took up mathematics.

Hippocrates is said to have written the first *Elements*, that is a collection of the mathematical (mainly geometrical) knowledge of the day. This work has not survived, but it would have been available to Euclid some 100 years later. Much of the content of books III & IV of Euclid are ascribed to Hippocrates.

The other claim to fame of Hippocrates is that his proof of quadrature of Lunes is the earliest surviving record of a proof. In fact what we have is Simplicius summary, written in 530 AD, of Eudumus' account of the proof, written in 335 BC. The original work is thought to have been done around 430 BC.

3.2.1 Quadrature

The idea of quadrature is to take some geometrical figure and produce a square with equal area. This must, however, be accomplished using only "ruler and compass". The ruler is a straight edge, it cannot measure distances, but can draw a straight line. The compass allows you to draw a circle centered at a given point. This type of construction is motivated by Euclid's first three postulates, and so these must have been known by this time. Almost all geometrical constructions of the ancient Greeks where by ruler and compass.

It is obvious that quadrature, if successful, would provide a useful method for calculating areas of figures, since the area of a square is trivial to determine. However it is clear that the interest in quadrature was more than just practical. It was a long standing problem of Greek mathematics to find the quadrature of various figures and it represented a real intellectual challenge.

The first step in Hippocrates' quadrature is to show that it is possible to find a square with the same area as a given rectangle. Then we show that this allows us to find the quadrature of a triangle. We then show how this leads to the quadrature of any convex polygon. A polygon is convex if a straight line drawn between any two vertices stays within the polygon.

In what follows we will assume that it is possible to bisect a given line into two equal parts using only a ruler and a compass (Euclid's proposition I.10) and to create a square on a given line (Euclid's proposition I.46).

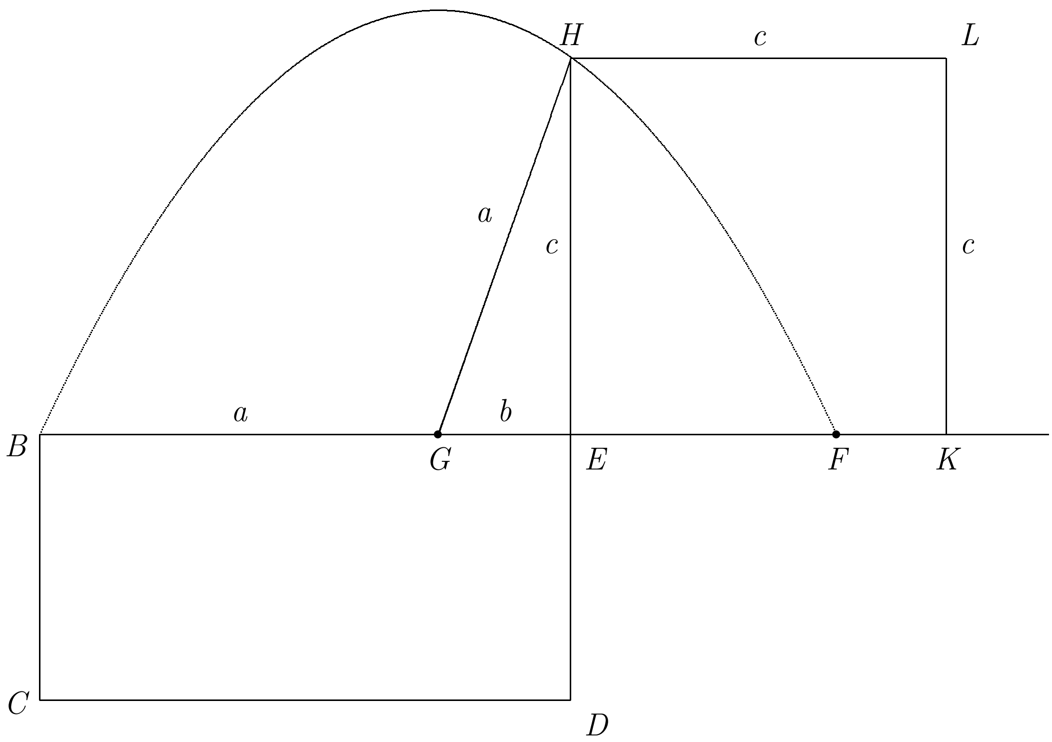
Step 1: The Quadrature of the Rectangle

This appears as Euclid's proposition II.14, the last proposition of the second book of Euclid. The proof given here and in Euclid is Hippocrates' original proof.

We start with a rectangle, shown in figure 3.1 as $BCDE$.
Extend the line BE to the right (Postulate 2).
Use the compass to mark the point F, so that $EF = ED$ (Postulate 3).
Bisect BF at G (I.10).
Now use the compass to put a circle with center G and radius $BG = GF$ (Postulate 3).
Extend the line DE to the point where it meets the circle, H say (Postulate 2).
Now create a square on the line EH, with the other two vertices being K and L (I.46).
$EHKL$ is the desired square.

We must now prove that this is indeed the required square. That is that the area of the rectangle is equal to that of the square.

Figure 3.1: Quadrature of the Rectangle



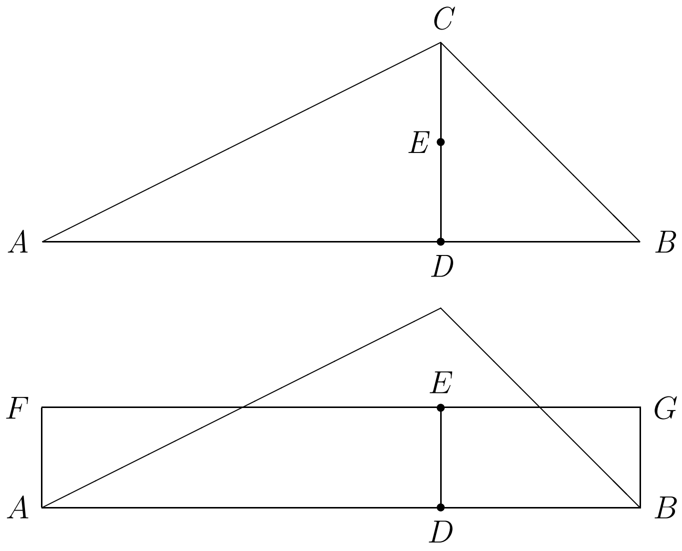
Let a be the radius of the circle. So $a = BG = GH = GF$.
Let c be the length of a side of the square. So $c = EH = EK = HL = KL$.
Let $b = GE$.
So by Pythagoras' Theorem we have that $a^2 = b^2 + c^2$. So $c^2 = a^2 - b^2$.
Also, since $GF = a$ and $GE = b$, we have that $EF = ED = a - b$.
Finally $BE = a + b$.
Now the area of the rectangle = base $\times$ height
$= BE \times ED$
$= (a + b)(a - b)$
$= a^2 - b^2$
$= c^2$
= the area of the square $EHLK$. $\square$

As an exercise try to come up with a geometrical proof that $(a + b)(a - b) = a^2 - b^2$.

Step 2: The Quadrature of the Triangle

We will assume that we know that the area of a triangle is $\frac{1}{2}$ base $\times$ height. We will also assume that given a straight line and a point not on the line it is possible to draw a straight line perpendicular to the original line through the point, using only a ruler and compass (Euclid's proposition I.12). We will also need that it is possible to draw a line parallel to a given line through a point (Euclid's proposition I.31).

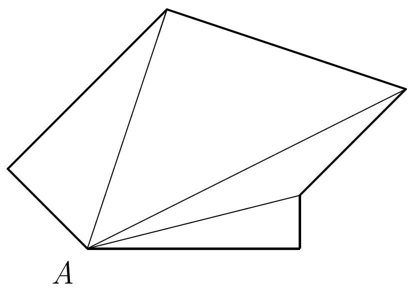
Figure 3.2: Quadrature of the Triangle



Consider the triangle ABC as shown in figure 3.2. We will consider AB to be the base.
Draw the line perpendicular to AB through the point C, call the point where this line intersects AB, D (I.12).
The line DC is the height of the triangle.
Bisect DC at E (I.10).
Now DE is $\frac{1}{2}$ the height.
Draw the line AF, parallel to DE at A (I.31).
Draw the line BG, parallel to DE at B (I.31).
The rectangle formed by $ABFG$ will have area $|AB| \times |DE| = \frac{1}{2}$ height $\times$ base.
By step 1 it is possible to find the quadrature of this rectangle. $\square$

Step 3: The Quadrature of the Polygon

Figure 3.3: Quadrature of a Convex Polygon



We start with any convex polygon, pick any vertex of the polygon, A say (see figure 3.3). Draw the lines from A to each of the other vertices of the polygon. Since the polygon is convex each of the regions thus defined is a triangle. By step 2 we can find the quadrature of each of these triangles. The area of the polygon is the sum of the areas of these triangles. $\square$

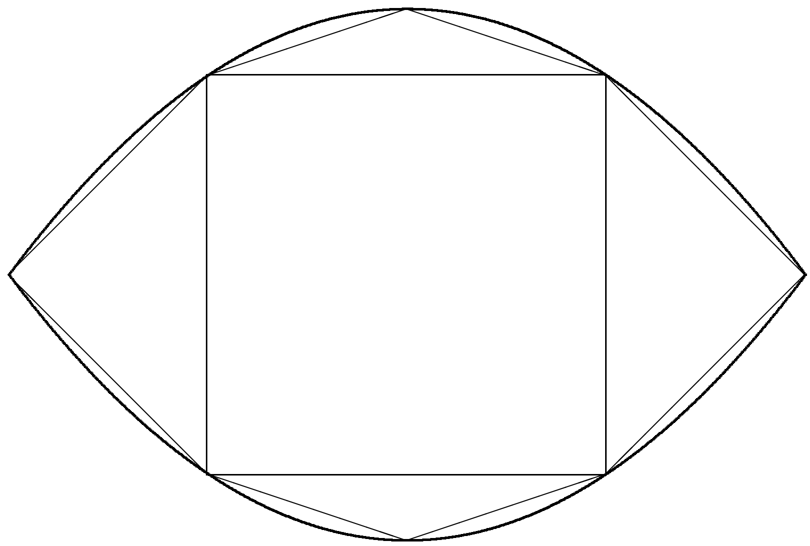
In fact Hippocrates went on to show that it is possible to find the quadrature of a Lune (a crescent moon shape).

3.2.2 The Method of Exhaustion

Given that we can find the quadrature of any convex polygon this suggests a method of finding the quadrature, and hence the area, of any (convex) curved figure. The idea is to approximate the curve by polygons, which we can find the area of by quadrature. We then let the size of each side get smaller and smaller, while increasing the number of vertices. Thus the area of the polygons approaches that of the curved figure. In the limit as the number of sides approaches infinity, and the length of each side approaches 0 we get the area of the curved figure.

In practice we start with a regular polygon inscribed in the curved figure, like the rectangle inscribed in figure 3.4. At each step double the number of vertices, and hence the number of sides of the inscribed polygon. In figure 3.4 this gives the inscribed octagonal polygon. At each stage the difference between the area of the curved figure and the inscribed polygon gets smaller. It is in fact possible to show that this difference in area decreases by more than half each time.

Figure 3.4: The Method of Exhaustion



The Greeks where well aware of this method of finding areas and called it the method of exhaustion, Eudoxus was to use it to great effect (see below). The method of exhaustion is actually equivalent to the modern technique of integration. However, its use once again raises the sceptre of Zeno and his paradoxes, since it involves the idea of infinitely subdividing the area between the figure and the polygon.

It is, in fact, necessary to show that the limit of the areas of the successive polygons tends to some definite number, which would be the area of the figure, thus we implicitly need the idea of a limit. The Greeks did this by a *reducto ad absurdium* type of argument, showing that the limit could not be less than the required area and that it could not be greater than the required area.

The trouble with this type of argument is that you must essentially know the area of the figure to begin with. Archimedes' (287 - 212 BC) book *The Method* , which was only recovered in 1905, gives hints on how to guess the areas of figures for use in this context. However the Greeks were unable to put the method of exhaustion on a rigorous foundation, at least until the time of Eudoxus (408 ? - 355 BC).

3.2.3 The Three Great Problems of Antiquity

Quadrature is related to one of three great problems which where current at this time and which became the three great unsolved mathematical problems of antiquity. All of these are ruler and compass constructions.

Problem 1: Squaring the Circle

Given a circle, find a square of equal area, using only ruler and compass. This is of course asking for the quadrature of the circle.

Problem 2: Doubling the Cube

Given a cube it is required to create a new cube with exactly twice the volume, using only a ruler and compass. There is a story of how this problem came about.

> During the great plague of Athens (429 BC) a deputation was sent to consult the oracle at Delphi as to what should be done to appease the gods and relieve their suffering.
>
> The Oracle replied that the (cubical) alter at the temple should be doubled. So the supplicants dutifully provided a new alter which was exactly twice the size in every dimension.
>
> However the plague continued unabated, because of course, the Delphic oracle had required an alter of twice the *volume*.

Problem 3: Trisecting an Angle

Given an angle it is required to divide it into 3 equal angles, using only a ruler and compass. Euclid gives a method of bisecting an angle into two equal parts with a ruler and compass (proposition I.9).

It was finally proved, some 2200 years later (around 1800), that none of these problems can be solved using ruler and compass alone.

3.3 The Athenian Philosophers

3.3.1 Socrates

The main Philosopher of note in Athens its heyday is Socrates (470 BC - 399 BC). Socrates is one of those pivotal people who define the step from one era to the next. Many of the so called pre-Socratic philosophers where in fact contemporaries of Socrates.

Socrates is usually portrayed as an ugly, absent-minded ragamuffin, though the Delphic oracle is said to have pronounced him the most wise of men. He was said to have enormous strength and courage, once risking his own life in battle to save a friend. Parmenedes and his student Zeno are said to have travelled to Athens at some point and to have debated with Socrates. However Socrates was not particularly mathematically inclined, indeed Socrates is said to have been somewhat opposed to the study of mathematics.

In 399 BC, at the age of 70, Socrates was convicted of non-conformism with the state religion and of corrupting the young with his teachings. He was sentenced to death, a sentence which he carried out himself by drinking Hemlock. Two of the 'corrupted youth' were Xenophon, who went on to become a renown general and historian, and Plato who recorded the ironic and sarcastic defence given by his teacher at his trial in Plato's *Apology.* In fact it is through Plato's dialogues, in which Socrates often appears as a character, that we know of Socrates and his teachings.

3.3.2 The Sophists

There is another group of philosophers who were active in Athens (and the rest of the Greek world) around the end of the fifth century BC, these where the Sophists. The Sophists held that truth was ephemeral and what really mattered was informed opinion. They generally acted as itinerant teachers and in this sense performed a useful function. However, they were the lawyers of the ancient world. They taught how to argue to win rather than how to argue to find the truth. Citizens in ancient Greece had to defend themselves in court and the sophist teaching of how to defend oneself in such circumstances was not far removed from a lawyers function. They developed a form of argument, called the *eristic,* which attempted to twist arguments and statements to a particular point of view and generate paradoxes in opponents statements.

Socrates and his school were vehemently opposed to the Sophists, since they believed in ultimate truth, and that it was the goal of philosophy, and hence philosophers, to find the truth. Another point of contention was that the Sophists charged for their knowledge, Socrates and his disciples held that knowledge was sacred and should be free to all.

Indeed the negative connotations of the modern words *sophist* (One who reasons with clever but fallacious arguments) and *sophism* (a false argument, especially one intended to deceive) come from this time, originally *sophist* meant *wise or learned person*.

An amusing story is told of Protagoras, the founder of the Sophists.

> Protagoras agreed with one of his poorer students that he could withhold payment until he had won his first court case. So convinced was Protagoras that his teaching was foolproof that he agreed that if the student didn't win his first case Protagoras would forego payment.
>
> However when the student had finished his tutelage he did not begin to practice.
> So Protagoras prosecuted in order to recover his fee.
> He argued that the student must pay:
> by the bargain if Protagoras lost and by the verdict if Protagoras won.
> The student gave as good as he got, arguing that the payment was forfeit:
> by the verdict if he won and by the bargain if he lost. [1]

The only mathematician of note to emerge from the Sophist tradition is Hippias of Ellis (fl. 440 BC ?). He is described by (a rather biased) Plato as "vain, biased and acquisitive". His main contribution to mathematics was that he constructed 'the trisectorix or quadratirix of Hippias', the first known example of a curve in geometry (as opposed to a straight line) other than the circle.

3.4 Plato

It is actually after its political heyday that Athens emerges as a center of philosophical and mathematical activity. The period from about 400 BC to 323 BC is when two men lived there who are still household names today, Plato (428 - 348 BC) and Aristotle (384 - 323 BC).

The first of these is Plato. Surprisingly nearly all of Plato's works have survived to the present day. His usual style of writing was the *dialogue* in which real characters come together and discuss their views in a the manner that they would have in a real circumstance, at least in

[1] Russell *Wisdom of the West*

Plato's view. Many of the dialogues actually portray real events, such as the *Pheada* which portrays Socrates' last moments with his devoted pupils.

Plato was born of an aristocratic family in Athens, he was slated to become a ruler of the city, but was so disturbed by the trial and death of his friend and mentor Socrates, that he opted for a life as a philosopher. After Socrates' death Plato travelled throughout the Greek world, possibly even in Egypt.

It is known that he was friends with Archytas, the mathematician ruler of Tarentum (now Taranto in Southern Italy), who was one of the last of the Pythagoreans. It was Archytas who introduced Plato to mathematics. In fact Plato does not appear to have been much of a mathematician himself, despite there being a number of mathematical entities named after him (the Platonic solids and Platonic numbers). However he had a high regard for mathematical endeavour and encouraged it in his students. In addition he had a strong understanding of number and geometry, much of the philosophy of Plato has a mathematical underpinning, with particular emphasis on geometry.

3.4.1 The Platonic Ideal

An essential idea which runs through much of Plato's works was that of the ideal 'Platonic' world. That is a world in which things take their true form, the real world in which we live is but a pale reflection of this ideal world. For example we still speak today of 'Platonic love', which is an ideal form of pure love, unsullied by the realities of sex. Plato gives the allegory of the cave as an example of the relation between the real world and the ideal world, and the philosophers connection between the two.

> Those without philosophy are like dwellers in a cave, behind them is a fire and in front of them a blank wall, on this wall they see their shadows reflected. As this is all that they see they come to believe that these shadows are in fact reality.
>
> One day a man finds his way out of the cave and sees the glory of the light of the sun illuminating the things of the real world. He goes back to the cave to tell his fellow cave dwellers of his discovery; that theirs is but a dim reflection of the real world. He tries to show them the way towards the light, to explain his momentous discovery.
>
> But he has become dazzled by the brightness of the sun, and now finds it hard to see the dim shadows of the cave dwellers. They thus think him addled, as he cannot see what they consider to be reality and do not want to follow him.

This idea of the importance of form is thought to originally have been due to the Pythagoreans, but Plato's version is far more sophisticated. This idea is obviously influenced by mathematics. If we consider a triangle, it is thought, in Plato's view, to represent an ideal triangle. The essential idea of form is to abstract the general nature of a thing and to talk only about this general nature. Hopefully we can then see what it is that is fundamental, what it is that makes a thing work in the way that it does.

Thus, it is an ideal triangle about which we prove theorems, not the imperfect triangle of the real world. It is the form of the triangle that is important, not the specifics of any particular triangle. If we want to consider a specific aspect or type of triangle we can specialise from the *genus* triangle to a more specific type of triangle. Thus for example if we consider Pythagoras' theorem about right angled triangles, it is the right angled aspect of the form that is important. This also points to another theme in Plato's work, the importance of precise definitions. We may define a genus or class, for example 'triangle', and then identify subclasses of this general form. Nowhere is precision of definition more evident than in mathematics.

It is in this sense that we are interested in the ideal *form* of things rather than in their imperfect representation in the real world. This idea permeates much of Plato's writing. It could be said that whereas the Pythagoreans placed number at the center of all things Plato placed form. It is interesting to note that in mathematics we talk today of *formal* theories, and *formal* systems, both derived from the fundamental Platonic idea of form being of ultimate importance. Ironically it is the evolution of these formal theories which now stands in direct contradiction of the Platonic view of an idealised system.

Aristotle attributes the statement "Numbers cannot be added" to Plato. At first glance this appears to be a completely ridiculous statement, but as is often the case with Plato there is much hidden beneath the surface. We should consider what we mean by (whole) number. Generally a number is an abstraction, a *form* not a thing. Thus when we speak of the number three, we mean the class of all things containing three objects[2]. Similarly the number two is the class of all things containing two objects. Thus, while we can add a collection of three objects to a collection of two objects to obtain a collection of five objects, we cannot add the number 2 to the number 3. This argument also emphasises the difference between types of object. Thus we cannot add quantities that are not of the same type, apples to oranges, an area to a length, a length to a volume etc.

[2] This definition leaves something to be desired since it is self referential, the word triplet is often used.

3.4.2 Plato's Academy

When Plato returned to Athens in 387 BC he set up a school. The land on which the school was founded was linked to the legendary hero Academus, thus the school was called 'The Academy'. This school lasted for over 900 years, longer than any present day institution, until it was closed by the Roman emperor Justinian in 529 AD. The Academy was based on the contemporary Pythagorean schools in Southern Italy, with which Plato had had some contact. The Academy was the model for the modern day University, it had a library, equipment for experimentation and carried on seminars and lectures.

It is not known whether the Academy charged tuition, it seems that those who attended must have provided for the upkeep of the place, but the existence of less well off students seems to indicate that there may have been some form of scholarships. Plato himself was wealthy and while he lived money was not an issue. With the existence of such a school the Sophists rapidly declined. It was much cheaper to send a child to the Academy, where they would be exposed to many of the greatest minds of the day, than to hire a sophist of dubious moral character.

There is a story told of Plato. One day a student at the Academy asked what gain there was to be had studying philosophy. Plato ordered a retainer to give the student a penny "for he feels he must profit from what he learns". The point here of course is that the profit is not material but spiritual.

The Academy regarded mathematics as a supreme art, and was a great center of mathematical activity. It is said that there was an inscription over the doorway which read "All those who dislike mathematics refrain from entering".

The mathematicians of the Academy included Theodorus of Cyrene (fl. 390 BC), Eudoxus of Cnidus (408 ? - 355 BC), Theaetetus (d. 369 BC), Menaechmus (fl. 350 BC) and his brother Dinostratus (fl. 350 BC), and Autolycus of Pitane (fl. 330 BC). One other student of note at the Academy was Aristotle (384 - 322 BC).

Theaetetus considered incommensurate ratios and quantities in detail. Much of the material in Book X of Euclid is ascribed to him. He is also credited with the theorem that the 'Platonic Solids' are the only regular polyhedra.

A polyhedra (three dimensional polygon) is regular if it is made up of only equal sized regular polygons. There are only five regular polyhedra, the Tetrahedron (made up of four triangles), the cube (made up of six squares), the octahedron (made up of eight triangles), the dodecahedron (made up of twelve pentagons) and the icosahedron (made up of twenty triangles). These five solids are known as the Platonic solids. They seem to have fascinated the Greeks who gave them a central role in their philosophy of the nature of things, at least as far back as the Pythagoreans. Why there are five and only five such regular solids does seem an intriguing question.

On Theaetetus' death in 369 BC from wounds received in battle Plato wrote a dialogue *Theaetetus* to commemorate his friend. This dialogue portrays a young Theaetetus visiting Socrates in his prison cell.

Menaechmus (fl. 350 BC) and his brother Dinostratus (fl. 350 BC) were actually students of Eudoxus. Menaechmus is accredited with the discovery of the curves known today as *conic sections*, the parabola, ellipse and hyperbola.

3.5 Eudoxus of Cnidus

Eudoxus was born on the island of Cnidus off the coast of modern day Turkey. He is said to have been poor in his youth. He travelled and studied around the Greek world, finally ending up at Plato's Academy. Eventually he moved on and set up his own school at Cyzicus in Asia Minor.

Eudoxus was undoubtedly the greatest mathematician of his era, however he was also a physician and legislator. Much of the material which appears in books V, VI and XII of Euclid is due either directly or indirectly to Eudoxus. It was Eudoxus who finally provided the answer to incommensurate ratios and went part of the way towards resolving Zeno's paradoxes.

3.5.1 Eudoxus' Definition of Ratio

At the heart of Eudoxus' work is a precise definition of ratio which did not rely on numbers. The definition is the same as that found in Euclid book V.

> Magnitudes are said to be in the same ratio $a : b = c : d$, if for any integers n and m if $na > mb$ then $nc > md$ and if $na = mb$ then $nc = md$ and if $na < mb$ then $nc < md$.

This has been rendered into modern notation, and n and m can be any integers. This definition is extremely sophisticated, in fact it resembles the 19th century definition of number.

It should be noted here that when Eudoxus says 'magnitudes' he is talking about not only whole numbers but irrational numbers, lengths, areas, volumes, etc. While this definition does not explain Zeno's paradox for numbers it does explain how to deal with infinitely regressing *ratios.*

What makes this definition so impressive is that it implicitly uses the idea of a universal quantifier. That is we are not saying that this is true for some particular n and m, or even for some range of n and m, but that it is true for every possible n and m. It is this idea that gives the definition its power.

At first glance this seems like too much to handle. How are we to prove that something is true for *every* number? In fact it turns out to be easier to do than one might expect, the use of universal statements is commonplace in modern mathematics. In fact the modern definition of a limit uses just such a universal declaration, coupled with an existential declaration. Eudoxus uses just such a coupling in his next step, the exhaustion property.

The other feature of this definition is what would be called today *division into cases*. That is we consider the all possible cases $>, =, <$ in turn. Eudoxus is certainly not the first to use this kind of argument, but that he manages to use it within a universal statement is impressive.

To see how much harder to understand these ideas are without the use of modern algebra we give a translation of Eudoxus' original definition.

> Magnitudes are said to be in the same ratio, the first to the second and the third to the fourth, when, if any equimultiples whatever be taken of the first and third, and any equimultiples whatever of the second and fourth, the former equimultiples alike exceed, are alike equal to, or are alike less than, the later equimultiples taken in the corresponding order.

This definition allowed Eudoxus to fix many of the Pythagorean theorems which had relied on the ratios of lines. It also allowed him to put the method of exhaustion of a firm footing by stating the *exhaustion property.*

> If from any magnitude there be subtracted a part not less than half, and if from the remaining one again subtracts not less than its half, and if this process of subtraction is carried on indefinitely there will remain a magnitude less than any preassigned magnitude of the same kind.

Again we are talking about any kind of magnitude. This appears as Euclid's proposition X.1. This is really saying that given any magnitude $\epsilon > 0$ we can (by the process outlined) find a magnitude smaller than ϵ. Since this is true for *any* ϵ, no matter how small, we can generate arbitrarily small quantities.

This kind of argument is at the heart of limits. In modern mathematical notation this would be written
$\forall\ \epsilon > 0, \exists\ \delta > 0$ such that $\delta < \epsilon$.
Where $\forall$ means *for all* and $\exists$ means *there exists*.
The reason that this property carries so much power is that it avoids the problem of talking about arbitrarily small magnitudes, instead it says: 'if you give me any quantity, no matter how small I can find a smaller one'.

Eudoxus used this property to prove the following theorem.

Theorem 3.5.1 *The ratio of the areas of two circles is the same as the ratio of the square of their diameters.*

Proof:
Suppose that we have two circles c and C with diameters d and D and areas a and A respectively. We wish to show that $\frac{d^2}{D^2} = \frac{a}{A}$.

We will assume that the ratio of the area of two polygons inscribed in two circles is equal to the ratio of their diameter.

This is a proof by contradiction, so we assume the opposite, that is $\frac{d^2}{D^2} \neq \frac{a}{A}$ and derive a contradiction.
Now if $\frac{d^2}{D^2} \neq \frac{a}{A}$ we must have either $\frac{d^2}{D^2} > \frac{a}{A}$ or $\frac{d^2}{D^2} < \frac{a}{A}$.
We will assume the second case, the other case is similar and is left as an exercise.
That is we assume that $\frac{d^2}{D^2} < \frac{a}{A}$.
Thus, there is a magnitude $b < a$ such that $\frac{d^2}{D^2} = \frac{b}{A}$.
Let $\epsilon = a - b$ be our preassigned magnitude (in this case an area), note that $\epsilon > 0$.
Now let our circles, c and C, each be inscribed with polygons having the same number of sides n.
Let these areas of these polygons be p_n and P_n respectively.
Let the remaining areas, that part that is outside of the polygons, but within the circles, be q_n and Q_n respectively.
If the number of sides of the polygons is doubled the remaining areas would be reduced by more than half (we will assume that this is true).
That is we are saying that $q_{2n} < \frac{1}{2}q_n$ and $Q_{2n} < \frac{1}{2}Q_n$.
Thus by the exhaustion property it is possible to increase the number of sides until $a - p_n < \epsilon$.
Now since $\epsilon = a - b$ we have that $p_n < b$.

By the assumption at the beginning the ratio of the areas of the polygons p_n and P_n is equal to the ratio of the square of the diameters.
That is, $\frac{p_n}{P_n} = \frac{d^2}{D^2} = \frac{b}{A}$.
But $p_n < b$, so in order to maintain equality we must have that $P_n > A$.
But the polygon with area P_n is inscribed *within* the circle C and so cannot have area greater than that of C. That is $P_n \leq A$.
This is a contradiction and so the premise, that $\frac{d^2}{D^2} < \frac{a}{A}$, is false. $\square$

3.6 Aristotle

The last in the line of Athenian philosophers is Aristotle (384 - 322 BC). Aristotle was born at Stagira in Thrace, in the kingdom of Macedonia. Macedonia was unusual among Greek states in that it had a King and royalty, indeed Aristotle's father was the court physician to the kings of Macedonia. It should be noted that the rest of the Greek world, based on city states, considered royalty a barbarian form of government.

At the age of 18 (366 BC) Aristotle was sent to study at Plato's academy, he remained there until Plato's death in 348 BC. During this time Macedonia expanded, conquering all of mainland Greece under king Phillip II. Athens became a vassal to the new power of the Greek world. After Plato's death Aristotle left Athens and went to in Mysia, wither an old school friend, Hermais a local ruler, had invited him. He married Hermais' niece and the couple moved to the island of Lesbos in 345 BC.

In 343 BC Aristotle was called to the Macedonian court to tutor the son of king Phillip II, Alexander (356 - 323 BC), later known as Alexander the Great. It is a major source of speculation as to what teachings one of the greatest philosophers might have imparted to the greatest conqueror who has ever lived. It is known that they did not see eye to eye on many issues.

In 340 BC Aristotle returned to his home town of Stagira, where he lived until 335 BC. At this time he returned to Athens where he set up a school of his own, called the Lyceum. He remained in Athens until 323 BC, leaving only in the last year of his life. It is from this second period in Athens that most of his surviving works date. Many of them are thought to be textbooks for use at the Lyceum.

After Alexander's death in 323 BC the empire that he had forged fragmented, Athens once again reasserted its independence, but it was no longer the power it had once been. During this period Aristotle, a Macedonian, feared that he would not be welcome in a newly independent Athens, so he fled to Chalcis where he died a year later.

3.6.1 Aristotelian Logic

Aristotle's main contribution to mathematics is that he codified and wrote down the rules of logic for posterity, thus we speak of Aristotelian logic. In the context of logical thought he discusses many of the basic definitions of geometry, we will discuss Aristotle's ideas in this regard when we consider Euclid.

Most of the logic recorded by Aristotle was undoubtedly known to Plato, but it is scattered through Plato's works. Aristotle synthesised and compiled this scattered knowledge into a complete whole. Nearly 1500 years later when Aristotle's works on logic re-entered Europe from the Arab world they would cause a storm.

Aristotle again emphasises the importance of form, this time in logic. One of the fundamental ideas of logic is that the statements themselves are not important, but rather the form in which they appear is. Logic may talk about any subject, the subject will change, but the *form* of the statements remains the same no matter what the subject. Aristotle gives the following example of the use of form in logic.

1. Socrates is a man.
2. All men are mortal.
3. Therefore, Socrates is mortal.

Statement 1 is of the form 'a is a b', statement 2 is of the form 'all b's are c's'. Aristotle holds that whenever we have two statements of this form, whatever a, b and c may represent, we can conclude that 'therefore a is a c', the third statement. In this case we have substituted the variables a for 'Socrates', b for 'man' and c for 'mortal'. Of course the use of variables is a modern idea, but the fundamental idea of the form being the important part of an argument was known to Aristotle.

Aristotle identified 4 essential types of logical statements:

1. All a's are b's.
2. No a is a b.
3. Some a's are b's.

4. Some a's are not b's

Aristotle's logic then codifies the relations between these statements. Thus in the example above we have two type 1 statements, which lead to a further type 1 statement of a particular form, related to the original two.

It is interesting to note that the form of these statements contains the idea of the modern notion of the universal quantifier 'For all' ($\forall$) and the existential quantifier 'There Exists' ($\exists$), as well as their negations.

3.7 Euclid and Alexandria

The death of Aristotle, or more precisely the death of Alexander in 323 BC, represents the final end of the Athenian age. Already during Aristotle's life Athens had been marginalised, a vassal to Macedonia. The conquests of Phillip and Alexander had irrevocably changed the shape of the ancient world, no longer was the city state predominant.

Alexander the Great conquered Egypt in 332 BC and Persia in 330 BC. By 325 BC he was undisputed lord of a region which included all of the Middle East, Pakistan and extending into India. However Alexander died unexpectedly at the age of 33, usually attributed to Malaria or possibly poison, leaving no clear successor. After his death his generals fought over his empire, eventually parcelling it out amongst themselves.

Egypt came under the rule of Ptolemy I who set up a school of learning in Alexandria, the city founded by Alexander when he conquered Egypt in 332 BC. The focus of the mathematical world moved from Athens to Alexandria in Egypt. Ptolemy summoned the most learned men he could find to staff his new school. One of these was Euclid who wrote the *Elements*, sometime around 300 BC. Very little is known about Euclid's life, he appears to have been a teacher and the elements is by way of an introductory (*element*ary) textbook. In the next chapter we consider Euclid's *Elements* in some detail.

Chapter 4

Euclid

(See *The Thirteen Books of the Elements* by Euclid, Vol. 1 (Books I and II))

Sometime around 300 BC Euclid wrote his thirteen books of the Elements. These books became the foundation of much mathematical endeavour over the next 2200 years. Many authors over the centuries have worked on problems associated with Euclid, but it was not until the latter half of the 19th century that Euclid was really superseded in its own field, that of planar geometry. This was already after the discovery of so called 'non-Euclidean geometries' by Gauss (1777 - 1855), Lobachevsky (1793 - 1856), Janos Bolyai (1802 - 1860) and Riemann (1826 - 1866).

By the latter half of the 19th century a number of authors sought to put Euclid on a firm footing with regards to problems that had been recognised by that time. Notable in this endeavour were Killing, Dedekind (1831 - 1916) and Hilbert (1862 - 1942). In 1900 Hilbert published a revamped version of Euclid which put planar geometry on a sound axiomatic footing. Another mathematician who wrote on Euclid about this time was Charles Dodgson (1832 - 1898), better known as Lewis Carroll, the author of *Alice in Wonderland.*

We start our discussion of Euclid with a brief general description of each of the thirteen books. Each book consists of a number of definitions of useful concepts relating to the topic at hand followed by a number of propositions, each of which is rigorously proved. Book I also includes the five basic postulates and the five common notions. We then give the definitions and postulates as well as selected proposition from Book I, without comment. The theorems in this section are generally as they would have appeared in the original, though some liberties have been taken in translation in order to give a more fluid account to a modern ear. Finally we continue on to a detailed analysis of Book I.

4.1 Summary of the 13 Books of Euclid

Book I (47 Propositions)

Euclid's first book covers the basic geometry of planar figures. It includes the five basic postulates and the five common notions. Often when referring to Euclid in general people actually mean the first book. Most of the first book is probably due to the Pythagoreans, or at least to this period. The final proposition (number 47) is Pythagoras' theorem.

Book II (14 Propositions)

This book covers the concepts of geometrical algebra, showing what we would consider algebraic formulas by geometrical methods. Again this material is probably due to the Pythagoreans or their contemporaries. It includes the quadrature of the rectangle as the last proposition (number 14).

Book III (37 Propositions)

This book covers the elementary geometry of circles, and associated chords, angles, tangents, etc. This material is believed to be due to Hippocrates or his contemporaries.

Book IV (16 Propositions)

This book covers regular polygons, from the triangle to the hexagon, either inscribed or circumscribed by circles. Again this material is assumed to come from Hippocrates.

Hippocrates original *Elements* would have covered roughly books I to IV of Euclid.

Book V (25 Propositions)

This book covers Eudoxus' theorem of ratio. Eudoxus' definition of ratio is included in the definition section of this book, and the propositions deal with ratios of magnitudes due to geometrical figures. This material is almost certainly due to Eudoxus.

Book VI (33 Propositions)

This book uses the results derived in book V to treat similar figures. Again this material is believed to be due to Eudoxus.

Book VII (19 Propositions)

This book is *arithmetical*, it deals with natural numbers and their ratios (fractions). Numbers are represented as line segments and the proofs are geometrical. This book really just reiterates the material in book V in the special case of line segments. It seems odd that Euclid considered it necessary to include this book, rather than simply using the results as implied by book V. Many theories have been suggested as to why Euclid did this. Possibly the intent was to emphasise Eudoxus' view of the difference between number and magnitude. Or perhaps Euclid wished to stress the difference between continuous (lines) and discrete (numbers) quantities.

A partial explanation is provided by noting that if we look at the definitions in book I, nowhere is *distance* mentioned, all magnitudes are only defined by their relation to each other (greater, lesser or equal). In book VII Euclid essentially allows the existence of 'A unit magnitude' and thus can talk about whole and fractional numbers as geometric quantities in ratio to this unit.

Book VIII (27 Propositions)

This book deals with algebraic relationships between ratios of numbers. Thus the final proposition, number 27 states that for any whole numbers a, b and c $(ma \cdot mb \cdot mc) : (na \cdot nb \cdot nc) = m^3 : n^3$.

Book IX (27 Propositions)

This book again deals with numbers. It is an odd mixture of theorems which would have been known to the Pythagoreans and later more sophisticated results. It includes an elementary treatment of odd and even numbers. But also includes (Proposition 20) the famous theorem on the infinitude of primes (see section 2.4.4).

Book X (115 Propositions)

This is the longest of Euclid's books, it is concerned with the classification of certain incommensurate (irrational) quantities. It is presumed to be the work of Theaetetus and Theodorus, both students of Plato.

Book XI (39 Propositions)

This book introduces the subject of solid (3 dimensional) geometry. It includes definitions of solid angles, the angle between planes and so on. It also introduces the five platonic solids as well as other solid figures, such as the prism, pyramid, sphere and cone.

Book XII (18 Propositions)

This book relates to the *measurement* of solid figures, in particular spheres, cones, pyramids and cylinders all have their volumes and surface areas calculated. The method used is exhaustion and all of this book is attributed to Eudoxus.

Book XIII (19 Propositions)

This book deals with the five regular (platonic) solids. The final proposition is the theorem that these five are the only regular polyhedra. Much has been made of the fact that Euclid ended his account with this theorem, involving the semi-mystical Platonic solids. Indeed some have tried to claim that the whole purpose of Euclid's thirteen books is just to prove this theorem. This is clearly not the case, whole books could have been left out if this was all that Euclid wanted to show. Most of this material is ascribed to Theaetetus.

4.2 Euclid – Book I

4.2.1 Definitions

1. A *point* is that which has no part. (A *point* is that which indivisible into parts).
2. A *line* is a breathless length.

3. The extremities of a line are points.
4. A *straight line* is a line which lies evenly with the points on itself.
5. A *surface* is that which has length and breadth only.
6. The extremities of a surface are lines.
7. A *plane surface* is that which lies evenly with the straight lines on itself.
8. A *plane angle* is the inclination to one another of two lines in a plane which meet one another and do not lie in a straight line.
9. When the lines containing the angle are straight the angle is said to be *rectilineal*.
10. When a straight line set up on a straight line makes the adjacent angles equal to one another, each of the equal angles is a *right angle*, and the straight line standing on the other is called *perpendicular* to that on which it stands.
11. An *obtuse* angle is an angle which is greater than a right angle.
12. An *acute angle* is an angle which is less than a right angle.
13. A *boundary* is that which is an extremity of anything.
14. A *figure* is that which is contained by any boundary or boundaries.
15. A *circle* is a plane figure contained by one line such that all the straight lines falling upon it from one point among those lying within the figure are equal to one another.
16. This point is called the *center* of the circle.
17. A *diameter* of the circle is any straight line drawn through the center and terminated in both directions by the circumference of the circle, and such a straight line also bisects the circle.
18. A *semicircle* is the figure contained by the diameter and the circumference cut off by it. The *center* of a semicircle is the same as that of the circle.
19. *Rectilineal figures* are those which are contained by straight lines. *Trilateral* figures being those contained by three. *Quadrilateral* those contained by four. An *multilateral* those contained by more than four lines.
20. Of trilateral figures an *equilateral triangle* is one which has all three of its sides equal. An *isosceles triangle* is one which has exactly two of its sides equal. A *scalene triangle* is one which has all three sides unequal.

21. Further, of trilateral figures, a *right angled triangle* is that which has a right angle. An *obtuse angled triangle* is that which has an obtuse angle. An *acute angled triangle* is one which has all three angles acute.

22. Of quadrilateral figures, a *square* is one which is both equilateral and right angled. An *oblong* is one which is right angled but not equilateral. A *rhombus* is one which is equilateral but not right angled. A *rhomboid* is one which has opposite sides and angles equal but is not a neither equilateral nor right angled. Let any quadrilaterals other than these be called *trapezia*.

23. *Parallel* straight lines are straight lines which, being in the same plane and being produced indefinitely in both directions, do not meet one another in either direction.

4.2.2 Postulates

Let the following be postulated:

1. It is possible to draw a straight line from any point to any point.
2. It is possible to produce a finite straight line continuously in a straight line.
3. It is possible to describe a circle with any center and diameter.
4. All right angles are equal to one another.
5. That if a straight line falling on two straight lines make the interior angles on the same side less than two right angles then the two straight lines, if produced indefinitely, meet on that side on which are the angles less than two right angles.

4.2.3 Common Notions

1. Things which are equal to the same thing are equal to one another.
2. If equals be added to equals then the wholes are equal.
3. If equals be subtracted from equals the remainders are equal.
4. Things which coincide with one another are equal to one another.
5. The whole is greater than the part.

4.3 Analysis of Euclid's Book I

In what follows all propositions definitions, etc. referred to are assumed to be from book I, unless explicitly stated otherwise. This is not an exhaustive analysis, we will only consider salient portions.

It should be stated at the outset that if we seem to be constantly criticising Euclid it is only to move forward, to see the things he may have missed. Euclid has stood over 2000 years of probing by many of the greatest minds that ever lived remarkably well. Indeed Euclid represents a seminal achievement, none of his 467 propositions is incorrect and his work has stood as *the* standard text on planar geometry for over two thousand years. Though we may find fault with a couple of his proofs and definitions the vast majority are sound.

4.3.1 The Definitions

It is clear that many of the problems associated with fundamental definitions were known even before Euclid's day, both Aristotle and Plato concern themselves with discussions of the fundamental definitions of planar geometry. In addition much was written after Euclid's time, one name in particular that crops up is that of Proclus (410 - 485 AD) who wrote *Commentary on Book I of Euclid.* The importance of Proclus' work is that he had access to texts written in antiquity, now lost to us, to which he makes reference.

The problem that Euclid runs into is that of defining fundamental concepts. How are we to start? If there is no ground to begin with, what are we to base our definitions on? The definitions may sound reasonable once we have an idea of what it is we are trying to define, but we shouldn't have to rely on previous knowledge or experience in a mathematical definition.

Euclid's first definition is probably the most problematical. It is attempting to define something by what it is not, never a good idea. Love could be said to have no part, but it is not a point. Aristotle accuses earlier scholars (apparently Plato) of defining a point as the extremities of a line, which has not yet been defined. Aristotle himself defines a point as "something which has position, but no magnitude, is indivisible". But this itself leads to a problem, for if a point has no magnitude then no collection, however large, of points could make up anything of magnitude. But lines may be cut anywhere to 'create' a point, how can this be? Of course Aristotle did not want to allow for infinite numbers of points to avoid Zeno. Worse yet, this definition assumes knowledge of another fundamental concept, that of *magnitude.* Even the word for point changed several times during the early Greek period.

Proclus states that the problem may be solved if we restrict ourselves to *the subject of geometry.* He also states that it is all right to use negations when defining first principles. Even Aristotle has to admit that certain terms, such as blindness, can only be defined in terms of what they lack.

This argument over the very first definition illustrates the problem of defining basic concepts. It is a reasonable definition only if we already have an idea of what the concept is. This is well illustrated by the second definition. If we know what a line is, defining it as a breadthless length seems reasonable. But give this definition to someone who has never seen or experienced the idea of a line and they will look at you blankly.

This problem also arises in other disciplines, in physics no attempt is made to define space or time, these are things which *are.* But physics talks about the real world, and so may at least claim the actuality of the objects referred to. It is not really reasonable to assert that points exist as anything but a mathematical abstraction, and if that is what they are we should be able to define them mathematically.

The question is really even deeper than this. Do we assert that mathematics, in particular planar geometry, has some sort of existence outside of our conception of it, in which case we are merely seeking to *describe* it. Or does it only exist in our conception, in which case we must *define* it. While it is possible to imagine that planar geometry has some sort of existence in reality, it becomes harder and harder to justify such an existence when considering more and more abstract constructions, such as for example a 7 dimensional hyperbolic geometry.

The modern way out of this dilemma is to consider fundamental terms, such as point and line as *undefined.* They become defined only by the way that they appear in the axioms, and nothing more. In this view it is a useful fiction to believe that points or lines represent planar geometry, they merely interact in the way stated in the axioms. We will consider this viewpoint later when we consider *formal systems.*

In fact it can be seen that there are undefined terms appearing in Euclid. For example in definitions 4 and 6 he talks of ‘lying evenly’, it is evident what is intended from the context, since we already know what a straight line is.

Notes on the Definitions

Definition 2: *A* line *is a breathless length.*

It should be noted that lines in Euclid's definition are not necessarily straight. It is also worth noting that Euclid defines the class *lines* but only defines one subclass, straight lines.

Definition 3: *The extremities of a line are points.*

The only points on a line are its endpoints, no mention is made of interior points. This is important to realise when considering definition 4. This is undoubtedly an effort to steer clear of Zeno, Euclid did not want to imply that lines where made up of an infinite number of indivisible points.

It should be noted that not all lines have endpoints, for example infinite lines or lines which enclose a circle. Thus Euclid does not assert that all lines have endpoints, only that if they do then these are points. In the case of a circle, or any line which encloses a region, it is sometimes useful to designate some point(s) on it as the endpoint(s).

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

'Lies evenly' has already been discussed. Plato defines a straight line as "straight is whatever has its middle in front of (i.e. so placed to obstruct the view of) both its ends"[1]. This definition is perhaps more evocative, but relies on sight or view. It seems that Euclid is attempting to say the same thing without using this physical concept from the real world.

Some more recent definitions are given in terms of distance, "A straight line is that line between two points which is shorter than any other line between those two points." It should be noted though that Euclid never in fact uses the concept of distance as such, only equality lesser or greater *magnitudes*, probably a wise move.

Definition 6: *The extremities of a surface are lines.*

Again surfaces are not necessarily planar.

Definition 8: *A* plane angle *is the inclination to one another of two lines in a plane which meet one another and do not lie in a straight line.*

This uses another undefined concept *meet*, though this causes less difficulty in understanding try to define 'meet' using only those concepts introduced thus far. It should be noted that here angle is defined for curved as well as straight lines. In this context the use of "do not lie in a straight line" seems odd.

There is some question as to what *category* angles lie in. The usual interpretation is that they are magnitudes. But consider the angle that a circle makes with its tangent. This angle "is less than any other angle" (but not zero), seeming nonsense. If on the other hand angle is a *quality* how can we talk about greater or less in relation to angles, rather we should say 'more angular' or 'less angular'. It is worth noting that Aristotle talks of *similar* angles,

[1]Plato *Parmenedes*

rather than equal angles.

Definitions 11 and 12: *An* obtuse *angle is an angle which is greater than a right angle. An* acute angle *is an angle which is less than a right angle.*

These require us to understand greater and less, particularly in the context of angles.

Definition 15: *A* circle *is a plane figure contained by one line such that all the straight lines falling upon it from one point among those lying within the figure are equal to one another.*

We should remember that the line containing a circle is not straight. The "one point" is of course the center of the circle. So this definition says that all straight lines from one boundary point to another boundary point, which go through the center, have the same length.

Definition 17: *A* diameter *of the circle is any straight line drawn through the center and terminated in both directions by the circumference of the circle, and such a straight line also bisects the circle.*

This uses the undefined term circumference, but this is surely an oversight, as we may readily define the circumference of a circle to be its boundary. In fact some later copies of Euclid include this in Definition 15.

This definition actually makes a claim about diameters of circles "they bisect the circle". Strictly speaking this should be proved as a proposition. The proof of this is usually attributed to Thales and so had been around for a long time. It may be that Euclid considered this to be self evident from his definition, which may well have been more sophisticated than that used by Thales.

Proclus gives a proof which argues that if the two bisected pieces were not equal then, if placed one over the other, one would have to be smaller. Thus the lines from the center would have to be longer on one or the other of the pieces. This proof uses the notion of *supersition by transport*, we move one piece and place it on top of the other, and then argue that they are equal. But this assumes a 'linear transport', that is that figures do not deform as we move them. Euclid runs into this problem in the proof of Proposition 4 (see below).

Definition 20: *Of trilateral figures an* equilateral triangle *is one which has all three of its sides equal. An* isosceles triangle *is one which has exactly two of its sides equal. A* scalene triangle *is one which has all three sides unequal.*

This requires us to know what equality is in the context of straight lines.

We note that the definition of parallel lines is left till the last definition.

While we have discussed the problems associated with fundamental definitions, most of the later definitions are actually compound definitions arising from the fundamental notions. In this case the form of the definitions is not so problematical as they can be built up from fundamental concepts, which we can now be presumed to know. However, it should be noted that the existence and any property ascribed to a defined figure must be demonstrated. Euclid is generally very careful to do this, showing or postulating the existence of each defined object before using it. Thus, for example, Proposition 1 may be seen as a proof that equilateral triangles (Definition 20) actually exist.

4.3.2 The Postulates

The first thing that should be noted is that both the *common notions* and the *postulates* would probably, in a modern account, all be stated as *axioms.*

Postulate 1: *It is possible to draw a straight line from any point to any point.*

This postulate is basically saying that we have a straight edge (ruler). Thus given any two points we may join them with our straight edge.

This definition implicitly postulates the existence of lines and their endpoints. It also implies that the line thus drawn is unique. That is "Any two straight lines drawn between two given points are one and the same", Euclid implicitly assumes this in Proposition 4. This also implies that two straight lines cannot enclose a space. It is clear from Proclus' commentary that Euclid did indeed mean this to be implied.

In addition it answers Zeno's concern about imperfect figures, we can draw perfect straight lines because we postulate that we can. Whether we can *in reality* do such a thing is immaterial, in our hypothetical construction (planar geometry) we have postulated that it is possible. To quote Aristotle

> The geometer's demonstration is not concerned with the particular imperfect straight line which he has drawn, but with the ideal straight line of which it is an imperfect representation.

Postulate 2: *It is possible to produce a finite straight line continuously in a straight line.*

A straight edge may also be used to continue an existing straight line, this possibility is covered here.

Again the assumption is implicit that the extension is unique, in either direction. Thus different straight lines cannot have common segments. This is explicitly stated as a consequence of this axiom in Book XI, however it is implicitly used immediately, in Proposition 1.

Postulate 3: *It is possible to describe a circle with any center and diameter.*

This postulate says that we have a compass with which to draw circles.

Some translations give the last word as "distance", but this seems inconsistent with Euclid's avoidance of the notion of distance, except in terms of lengths of given lines, in this case the diameter. In fact radius would be a better word as this is what is actually meant.

It should be noted that the only thing we can do with our compass is draw circles. In particular we are not allowed to measure off a distance, move our compass and apply this distance elsewhere. This is what is known as a collapsing compass, as soon as it is removed from the page it collapses. In fact Propositions 2 and 3 are a proof that we can in fact 'measure off' distances by lines in this way. However it is not until Proposition 31 that we are given a method of drawing a line parallel to a given line.

The prevalence of 'ruler and compass' constructions by the 5th century BC indicates that these postulates must have been well known in some form by that time. They are probably due to the Pythagoreans or Thales.

Postulate 4: *All right angles are equal to one another.*

At first sight this seems 'obvious', but in mathematics one would be well advised to be extremely wary of the word obvious. Euclid defines right angles by the relation of straight lines, it is not necessary, from the definition, that all angles created in this way must be equal, hence we must postulate that this is indeed so.

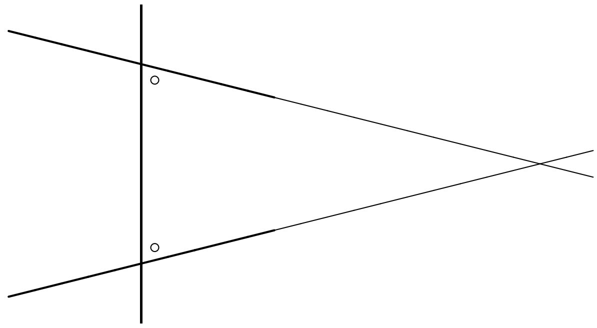
In fact this also implicitly stipulates that angles (at least right angles) are to be regarded as magnitudes, since right angles can be *equal* to each other. This also gives us a standard from which other angles can be measured as in Definitions 11 and 12.

This axiom has another important consequence, it implies the homogeneity of (geometrical) space. That is, if we move figures around right angles are preserved. Euclid seems to think that this implies that figures and lengths are also preserved, in fact this does not necessarily follow.

One other thing to note about this 'postulate' is that it is in some sense really an axiom. That is in the first three postulates we postulate that we can *do* something, whereas in this postulate we are making a statement about the *nature* of geometrical space.

Postulate 5: *That if a straight line falling on two straight lines make the interior angles on the same side less than two right angles then the two straight lines, if produced indefinitely, meet on that side on which are the angles less than two right angles.*

Figure 4.1: Euclid's Parallel Postulate



This postulate is illustrated in figure 4.1. if the indicated angles are both less than right angles then the lines will meet if continued. Note that this postulate does not say that the angles formed on one side by a line crossing two non parallel lines *must* be less than two right angles, only that if they are, then the lines will meet.

This is Euclid's much famed fifth postulate, also known as the parallel postulate. It seems unnecessarily complicated when compared to the other four. The idea of postulates is that they be as few as possible and as simple as possible, describing only the necessary basic principles. The first three postulates describe what is needed for our ruler and compass, the fourth states a seemingly basic obvious property, but the fifth states a fairly complex property of non parallel lines.

It is obvious that Euclid himself shared this concern, he delays the use of this postulate as long as possible, it is not used until Proposition 29. As far as we know this postulate is due to Euclid himself, no earlier references exist to it, though earlier geometers had used equivalent forms.

Over the centuries enormous effort has been put into showing that this axiom can be derived from the others or replaced by a simpler one, all to no avail. Some of the great names of mathematics who have tried their hand at this problem are Ptolemy (c. 200), Proclus (410 - 485), Nasiraddin at-Tusi (1201 - 1274), John Wallis (1616 - 1703), Gerolamo Saccheri (1667 - 1733), Johan Heinrich Lambert (1728 - 1777), Laplace (1749 - 1825), Legendre (1752 - 1833), Gauss (1777 - 1855), Lobachevsky (1793 - 1856), Janos Bolyai (1802 - 1860) and Riemann (1826 - 1866). In fact the last four names on this list are the mathematicians who finally cracked the secret of the fifth postulate and hence invented non Euclidean geometry.

So much wasted effort had been expended on the investigation of this postulate by the 19th century that when Janos Bolyai's father Wolfgang, himself a mathematician who had worked on the parallel postulate and a friend of Gauss, heard that his son was considering working in this area he begged him to reconsider:

> You must not attempt this approach to the parallels. I know this way to its very end. I have traversed the bottomless night, which extinguished all light and joy of my life. I entreat you, leave this science of parallels alone. You should detest it just as much as lewd intercourse; it can deprive you of all of your leisure, your health, your rest, and the whole happiness of your life. This abysmal darkness might perhaps devour a thousand towering Newtons.

Over the years many equivalent statements of this postulate have been found, probably the most well known of these is Playfair's Postulate:

> Given a straight line and a point not on it, then there is exactly one straight line through the point parallel to the given line.

In fact this form was known to Proclus. Another equivalent statement given by Proclus is:

> Straight lines parallel to the same straight line are parallel to one another. (Proposition 30)

Yet another equivalent form is:

> The angles of a triangle add up to two right angles (Proposition 32).

4.3.3 The Common Notions

The common notions are interesting in that they relate to things which may be held to be true outside of geometry. The form of their statement keeps them as general as possible. We should note though that Euclid is absolutely correct in realising that they must be stated explicitly. In modern terms they represent essentially algebraic laws.

The most contentious of the common notions is number 4 (*things which coincide with one another are equal to one another*). Unlike the others it seems to be talking specifically about geometric objects, and so some have argued that it really should be included with the postulates. What is more contentious is that it seems to be an *axiom of congruence*, that is an axiom which allows the use of supersition by transport to show that things coincide and are hence equal.

We must salute the incredible achievement of Euclid and the ancient Greek world in realising that these are the only axioms (postulates and common notions) needed (well... almost). Any more would be superfluous, any less and we would not be able to prove all of the theorems. The only thing that is missing is a principle of continuity (see below).

4.3.4 Propositions 1 - 4

Proposition 1: *It is possible to construct an equilateral triangle with a given finite straight line as a side.*

We have already mentioned that Euclid is implicitly using postulate 2 to ensure that the lines do not have a common segment. If this were possible the lines AC and BC might meet before C, and extend with a common segment to C.

We must also assume a similar property for circles. That is that two circles, if they meet, have no common part. It is actually proved in Book III, Proposition 10 that two circles cannot meet in more than two points, exactly what we need. But we are not allowed to assume a proposition which has not yet been proved. One way out of this dilemma is to realise that we only need *a* point of intersection, to find our equilateral triangle. The only problem then would be that the picture might be misleading (which it is not by III.10).

There is in fact a deeper problem which suffuses much of Euclid. Nowhere does he state that if two lines cross then they must meet. This is essentially a principle of continuity, i.e. that there are no holes in lines, however small, through which other lines may pass. The problem is that Euclid does not want to define lines in terms of interior points (indivisibles or atoms), otherwise he runs into the problem of Zeno's paradoxes. But all known definitions of continuity (particularly Dedekind's) use the idea that a line is made up of points. Killing in the 19th century proposed adding the following postulate, which covers most cases:

> Given a figure which divides into two parts (the inside and the outside) and a line, one of whose endpoints is inside the figure and the other of whose endpoints is outside the figure, then the line must meet the boundary of the figure at some point.

Dedekind (1831 - 1916) gave an axiom of continuity, which relies on lines being made up of points.

> If all points of a straight line fall into two classes such that every point of the first class lies to the left of every point in the second class, there exists one and only one point which produces this division of the straight line into two parts.

We could take issue with the use of “to the left of”, but this axiom may be stated in precise mathematical language. The statement “one and only one” implies that the point of the cut exists and is unique. Such a point is sometimes referred to as a Dedekind cut. It should be noted that this axiom applies only if we consider lines as made up of points, and thus really only makes sense in terms of the wider revamping of planar geometry which was taking place at the time. Dedekind’s axiom can be used to show Killing’s postulate and in addition this axiom can be used to show that similar theorems hold for arbitrary lines and even angles.

Proposition 2: *It is possible to place a straight line at a given point which is equal to a given straight line.*

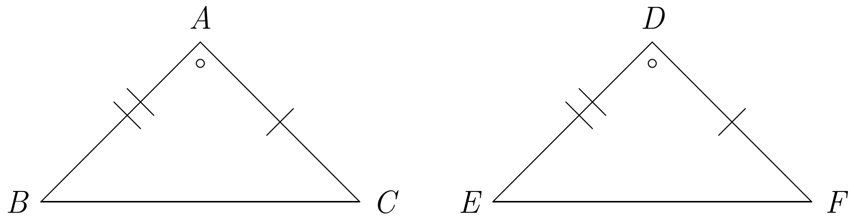
In fact the construction given does not cover the case when the given point A lies on the given line BC or an extension of it. It is quite possible that Euclid considered this a trivial case and left it as ‘an exercise for the reader’. As an exercise try to prove this proposition in the case when A lies on BC, or an extension of it.

Proposition 3: *Given two unequal straight lines, it is possible to cut off, from the greater a straight line equal to the less.*

Propositions 3 gives us a non collapsible compass. We can measure off any line against any other. Note that Proposition 3 uses Proposition 2 in its proof.

Proposition 4: *Given two triangles which have two sides of the first equal to two sides of the second, and the angle contained by this pair of sides equal, then the triangles will be equal, thus the third side of the triangles will also be equal and the remaining angles will be equal.*

Figure 4.2: Euclid’s Proposition I.4



This proposition has probably received more attention than any other single proposition in

Euclid. This proposition is essentially a special case of the well known 'Side Angle Side' theorem.

A serious problem is encountered in the proof. The proof is one of supersition by transport. Euclid essentially moves one triangle on top of the other (this is what is meant by *applied*) and compares the result. Only two propositions ago Euclid went to great lengths to show that we could move lines around, now he is cavalierly doing the same thing for a whole triangle and assuming everything stays the same. Specifically he is assuming that the act of transport does not change the enclosed angles. Nowhere in the postulates or common notions do we state that such a thing is true. If such a method of proof were generally allowed the proof of Proposition 2 would be trivial, simply place the line at the required point. It seems likely that Euclid was at least to some extent aware of the problems associated with this method of proof as he avoids it whenever possible.

In fact there is a completely insupportable statement in the proof. Around line 20 Euclid states

> If the point A is placed on the point D and the straight line AB on DE, then the point B will coincide with E, because AB is equal to DE.
> The straight line AC will also coincide with DF because the angle $\angle BAC$ is equal to the angle $\angle EDF$.

No explanation or reference is given for this startling presumption, because Euclid has none. What Euclid is assuming is the converse of common notion 4, "If two things are equal then they will coincide". But this does not follow from any of the stated notions or postulates.

The situation can be rescued to a certain extent by appealing to Proposition 2 to place the line AB at D, let us call this new line DB'. The equality of AB and DE does not imply that AB lies along DE, to see this imagine that the triangles in the picture are rotated with respect to each other. However we can draw a circle with center D radius equal to AB and conclude that E lies on this circle. An exactly similar procedure can be conducted on the line AC to produce a new line DC'. The problem now is that we have no way of knowing anything about the angle $B'DC'$. Certainly there is no reason to conclude that it is equal to BAC from the axioms and propositions proved thus far. The reason for this is that the act of *transport* may have distorted the angles.

In order to get a feeling for the problem, imagine that the space in which we are carrying out this procedure is not flat but bumpy. Now if we slide the triangle ABC around on this surface the angle BAC will change as A goes over the bumps. In fact this proposition is exactly asserting that geometrical space is *not* bumpy, but flat.

Stated in these terms it seems a pretty fundamental property of geometrical space and might well be a candidate for being a postulate. In fact when Hilbert did his new formulation of planar geometry in 1900 this is exactly what he was forced to do, including Proposition 4, in a slightly weaker form, as an axiom.

4.3.5 The Other Propositions

Proposition 5 is the theorem which states that the base angles of an isosceles triangle are equal, attributed to Thales.

Proposition 15 is Thales' theorem about opposite angles.

Proposition 26 is the well known 'angle side angle' theorem on triangles, this theorem is also often attributed to Thales.

Proposition 32 is the theorem which states that the angles of a triangle add up to two right angles (180^o), also attributed to Thales. This proposition is actually equivalent to the parallel postulate, so Thales would have to have been aware of the parallel postulate, or some equivalent in order to give a truly correct proof, this seems unlikely.

Proposition 29 is the opposite angles theorem for parallel lines. The proof given previously of Proposition 32 assumes this result, it seems likely that this is the way that Thales approached the problem. This is the first proposition in Euclid which requires the parallel postulate. Note that Proposition 28 is actually the converse of Proposition 29.

Proposition 47 is Pythagoras' theorem. Note that the proof given in Euclid is not that given on page 31.

In previous constructions we have assumed, at various stages, the following.
Proposition 2, that we may move a line.
Proposition 3, that we can cut a line of a given length from a longer line.
Proposition 10, that we may bisect a line.
Proposition 12, that we can draw a line perpendicular from a given line through a given point not on the line.
Proposition 31, that we can draw a line parallel to a given line through a given point.
Proposition 46, that we may construct a square on a given line.

The other propositions given result from the desire to state Proposition 46 with a complete proof. Looking through the proof of Proposition 46 we see that the square is constructed by drawing a line perpendicular to it from A, this is Proposition 11. Then we draw a line parallel to AB, this is Proposition 31. Finally to show that the result is in fact a square we must appeal to propositions 29 and 34. Thus propositions 11, 29, 31 and 34 must be given.

Figure 4.3: The Dependencies of Propositions in the Proof of Proposition I.46

Proposition	requires	is used to prove[1]
46	11, 29, 31, 34	-
34	4, 26, 29	46
31	23, 27	46
29	13, 15	24, 46
27	16, 23	31
26	4, 16	34
23	8, 22	27, 31
22[2]	-	23, 27, 31
16	3, 4, 10, 15	26, 27
15	13	16, 29
13	11	15, 29
11	1, 3, 8	13
10	1, 4, 9	16
9	3, 8	10
8	7	9, 10
7	5	8
5	1, 3, 4	7
4[3]	-	5, 10, 16, 26, 34
3	2	5, 9, 11, 16
2	-	3, 9, 11, 16
1	-	10, 11

[1] Only those propositions already in the list are given.
[2] Proposition 22 makes reference to Proposition 20 in its statement.
[3] Proposition 4 should be an axiom, note how many of the other propositions rely on it.
We do not include the dependencies on the postulates or common notions

A perusal of the proof of Proposition 34 shows that it in turn relies on propositions 4, 26 and 29, so these must also be given. If we follow the chains down the propositions we get a tree of dependencies, described in the table in figure 4.3.

Thus we must give all the propositions in this list in order to give a full proof of Proposition

46. In order to prove Proposition 47 as Euclid does propositions 14, and 41 must be added. Proposition 41 in turn relies on propositions 34 and 37.

Thomas Hobbes (1588 - 1679) the well known political philosopher was introduced late in life, aged 40, to Euclid. When he was visiting a friend he was asked to wait, a copy of Euclid lay open on the table, John Aubrey recounts:

> Euclid's *Elements* lay open at Proposition I.47. He read the proposition. "By G−−" said he, (he would now and then swear emphatically by way of emphasis) "*this is impossible* !" So he reads the demonstration of it, which referred him back to such a proposition; which proposition he read. That referred him back to another, which he also read, and so on, that at last he was demonstratively convinced of the truth. This made him in love with Geometry.

Chapter 5

Euclid to the Renaissance

5.1 The Later Greek Period

It would be misleading to imply that Greek mathematical development ended with Euclid. The 3rd century BC saw further mathematical growth and development, though much of the substance is beyond the level of this course. The great library at Alexandria in Egypt continued to be a center of activity for several hundred years.

Two names that stand out from the 3rd century BC are Archimedes and Apollonius, both of whom are said to have studied at Alexandria, but later moved elsewhere.

5.1.1 Apollonius of Perga

Apollonius (262 - ? BC) was born in Perga in Asia Minor, he was probably educated in Alexandria and later taught in Pergamum. Very little is actually known of his life and none of his works have survived in the original.

Apollonius is probably best known for his eight volume work *Conics* containing 467 propositions. This was to the geometry of conic sections (ellipses, circles, hyperbolae and parabolae) what Euclid was to basic geometry. *Conics* represented a summary of the known work in the area, it completed the work of Eudoxus' student Menaechmus.

The first four books have been recovered from Greek manuscripts dating from the 12th or 13th centuries. The next three books have survived in Arabic translations by Thabit ibn Qurra. But the eighth book has been lost. Edmund Halley (1710) provided Latin translations of the first seven books and attempted to piece together the contents of the eighth book from statements given by Pappus (284 - 345 AD).

5.1.2 Archimedes of Syracuse

Archimedes (287 - 212 BC) was born in Syracuse in Sicily. He also studied in Alexandria and then returned to Syracuse until his death at the hands of an unknown Roman soldier.

Archimedes was incredibly prolific with wide ranging interests in mathematics, physics and engineering. He set the foundations of modern engineering introducing concepts such as center of mass, buoyancy and leverage. It is to Archimedes that the immortal line "Give me a place to stand and I shall move the Earth" is ascribed. He invented the compound pulley and the Archimedian screw (for raising water).

He is known to have published at least 10 books, however unlike Euclid and Apollonius these were all original work. These books are *On the Equilibrium of Plane Figures* (2 books), *Quadrature of the Parabola* , *On the Sphere and Cylinder* , *On Spirals* , *On Conoids and Spheroids* , *On Floating Bodies* (2 books), *The Measurement of the Circle* , *The Sandreckoner* and *The Method.* There are several more books which have been lost, including *On levers* and *On centers of Gravity.*

Archimedes is often accredited with inventing the method of exhaustion, which he used in his quadrature of the parabola. This is not strictly speaking correct, the method had been known for some time, at least since the time of Eudoxus.

The method of exhaustion required that the answer essentially be known beforehand. Archimedes book *The Method* contains tricks for guessing the areas and volumes of figures so that a proof by exhaustion can be given, thus the 'method' is that of exhaustion. This book was in fact thought to have been lost until recently. In 1906 J. L. Heiberg was investigating a set of Greek manuscripts dating from about 900 AD. Among these was a parchment which had been wiped clean for reuse (parchment was expensive). The original text could just be made out, it turned out to be a copy of Archimedes' *The Method.*

There are two stories told of Archimedes. The first relates to his discovery of the laws of buoyancy "a body displaces an amount of water equal to its volume". It is said that he made this discovery while in the bath (presumably the public "Roman" style baths), he was so excited that he leapt up and ran naked down the street shouting "Eureka" (I found it).

The second story relates to the fall of Syracuse to the Romans in 212 BC and the method of his death at the age of 75. This is related by the Roman historian Plutarch. The Romans under the direction of the general Marcellus approached Syracuse in 215 BC, Archimedes was ready for them, inventing a series of deadly weapons to protect his city. One of them was said to have been the catapult, which the Romans adopted as their own. To quote Plutarch (circa 100 AD):

> For the truth was that all the rest of the Syracusians merely provided the manpower to operate Archimedes' inventions, and it was his mind which directed and controlled every manoeuvre. All other weapons were discarded, and it was upon him alone that the city relied for both attack and defence. At last the Romans were reduced to such a state of alarm that if they saw so much as a length of rope or a piece of timber appear over the top of the wall, it was enough to make them cry out 'Look Archimedes is aiming one of his machines at us!' and they would turn their backs and run. When Marcellus saw this he abandoned all attempts to capture the city by assault and settled down to reduce it by blockade.

Eventually, after a siege lasting $2\frac{1}{2}$ years, Marcellus' strategy paid off, and the city fell. Reportedly Marcellus had become a great fan of Archimedes and ordered that his life be spared, unfortunately his order was not obeyed. Plutarch gives three different possible accounts of Archimedes' death:

> Nothing afflicted Marcellus so much as the death of Archimedes, who was then, as fate would have it, intent on working out some problem by a diagram, and having fixed mind and eyes alike upon the subject of his speculation, he never noticed the incursion of the Romans, nor that the city was taken. In this transport of study and contemplation, a soldier, unexpectedly coming up to him, commanded him to follow to Marcellus; which he declined to do before he had worked out his problem to a demonstration. The soldier, enraged, drew his sword and ran him through.
>
> Others write that a Roman soldier, running upon him with a drawn sword, offered to kill him; and that Archimedes, looking back, earnestly besought him to hold his hand a little while that he might not leave what he was then at work upon inconclusive and imperfect; but the soldier, nothing moved by his entreaty, instantly killed him.
>
> Others again relate that, as Archimedes was carrying to Marcellus mathematical instruments, dials, spheres and angles, by which the magnitude of the Sun might be measured to the sight, some soldiers seeing him, and thinking that he carried gold in a vessel, slew him.

> Certain it is that his death was very afflicting to Marcellus; and that Marcellus ever after regarded him that killed [Archimedes] as a murderer; and that he sought for [Archimedes'] kindred and honoured them with signal favours.

One common feature of the first two stories is the way that Archimedes is so abstracted in thought that death means little when compared to solving the mysteries of the Universe.

One of the features of Archimedes' work was that he never liked to present anything that was not a finished whole. This attitude is evident in many of the great minds of mathematics and science. It explains why often results of great import where not widely disseminated by their inventors until long after their original inception, since they wished to have every detail perfect. This was a less than perfect strategy when such an endeavour took more than a lifetime.

5.1.3 The Alexandrians

The Library and Museum at Alexandria remained a center of learning and mathematical endeavour for nearly 700 years. Some of the names associated with the great library at Alexandria in its early period are Eratosthenes of Cyrene (276 - 194 BC), curator of the library who is famed for having measured radius of Earth with great accuracy, he also worked on the foundations of what we would today call trigonometry. Another who worked in this field was Aristarchus of Samos (310 - 230 BC). Hiparchus of Nicaea (180 - 125 BC) compiled first trig table and is said to be the first to use 360 degrees in a circle. Whether this was his invention or adopted from the Babylonian astronomers of the day is uncertain.

In 31 BC Egypt fell to the Roman leader Octavius Caesar (later the emperor Augustus) after a long struggle immortalised in Shakespeare's 'Anthony and Cleopatra'. After this Alexandria became part of the Roman Empire, however, the center of learning in Alexandria continued under its new masters.

One of the figures from this period is Menelaus of Alexandria (100 AD) who worked on spherical geometry. Sometime around 200 AD Claudius Ptolemy wrote a compendium of mathematical and astronomical knowledge to that date, this book is known today by its Arabic tittle *Almagest* (The Greatest). About the same time was Heron of Alexandria who was interested in mensuration, and is probably best known today for "Heron's Formula" for the area of a triangle, given the lengths of its sides, a, b and c, Area $= \sqrt{s(s-a)(s-b)(s-c)}$, where $s = \frac{1}{2}(a+b+c)$.

In the later period we have Diophantus (c. 300 BC), who is sometimes called the father of algebra. He invented notational systems which are much closer to our modern algebraic notations. However his material was never really widely distributed and was not available to later scholars, who had to reinvent algebra. Around the same time lived Pappus of Alexandria (284 - 345 AD) who wrote *Synagoge* (Mathematical Collection), eight of the original ten books have survived. Pappus is also remembered in the modern "Pappus' theorem" on centroids of rotated bodies. Pappus is considered the last of the "Greek" mathematicians. By the time of Hypatia (415 AD) the hegemony of Alexandrian learning was over.

The last product of Alexandria was Proclus (410 - 485 AD). Though born in Alexandria, significantly, he moved to Athens. He was more of a historian than a mathematician, and he wrote a detailed *Commentary on Book I of Euclid* . He would have had access to the *History of Geometry* by Eudumus, as well as Pappus' *Commentary on the Elements* both now lost. It is from Proclus' work that much of our knowledge of the ancient Greek mathematical world stems.

Finally in 651 AD the early Muslim Arabs under the Caliph Omar spread westwards through Alexandria. The Great Library was destroyed in the fighting and untold numbers of works from antiquity which were stored there were destroyed and lost for all time. This represents the final chapter of the Alexandrian period, which had lasted for almost a thousand years from 300 BC, though the city had ceased to be a center of learning from about 400 AD.

5.1.4 Mathematics in the Roman World

It should be noted that during the Roman period, while there where great advances in optics, astronomy, geography, medicine and mechanics, there were few significant advances in mathematics. There were advances in trigonometry, but this was viewed in purely practical terms as a device of measurement.

The Roman world in general was noted for its feats of engineering, but almost no significant mathematics is attributable to this period. What mathematics there was in the Roman world seems to have been almost exclusively carried out by people who were culturally Greek living in Egypt under Roman rule. In the 500 years from Hiparchus (150 BC) to Pappus (350 AD) there were far fewer advances in mathematics than in the 200 years from Pythagoras to Euclid. When Greek mathematics was reintroduced into Europe in the 15th and 16th centuries, the invention of Calculus followed less than a hundred years later.

This raises the question of what element was missing from the Roman world which the earlier Greeks seem to have had aplenty. The Roman empire was a flourishing empire with wealth and trade. There certainly was leisure time for mathematical study, which would seem to argue against Aristotle's idea that it was too much spare time in the priestly classes which led to the initial development of mathematics in Egypt.

Some argue (see M. Kline *The loss of Certainty*) that mathematics must be close to applications to move forward. But the Romans *used* mathematics extensively to build things and for navigation. They seemed uninterested in mathematics for mathematics sake the way the Greeks were. Indeed it seems that too much effort and concern was being given to the purely mechanical tasks of measurement and not enough to theoretical endeavour.

Another possibility is that the Roman world was a unified state, with very centralised structures. It thus lacked the dynamism produced by individual, yet culturally similar Greek city states. This dynamism was also apparent in the disparate yet culturally similar nature of European states in the Renaissance.

It has also been suggested that the Greek systems of notation (algebra) had reached its limit, that this system just couldn't bear the needed calculations. But this seems a weak explanation for over 500 years of relative inactivity. Surely if they had been concentrating on it someone would have invented better notations. In fact Diophantus of Alexandria around Pappus' time (350 AD) did invent a form of algebraic notation, but it came too late, the mathematical culture in Alexandria was almost gone and it didn't exist significantly elsewhere.

5.2 The Post Roman Period

5.2.1 The Hindus

After the fall of Rome (450 BC) to barbarian invaders, mathematical culture moved elsewhere. Though the Eastern Roman Empire survived, in one form or another for another 1000 years. The center of mathematical activity moved at this time to the Hindu culture in the Indus valley. The Hindu's had had some contact with the Greek world, at least from the time of the conquests of Alexander.

It should be noted though that the chronology of events in India at this time are somewhat confused. Few texts survive, and those that do have little or no chronological context. Much of the work in India was in algebra (solution of equations) and trigonometry.

The form of Indian mathematics really represents a step back from the achievements of the Greeks. Much of the early material was written in verse, which did not allow for detailed proofs. The form of the material is more reminiscent of Egyptian and Babylonian texts, lacking the logical deductive flavour of Greek mathematics.

While this period can hardly be called prolific, there was one enormously significant achievement between about 600 AD and 900 AD. This was the introduction of the decimal number system (in use by 595), and the use of zero as a place holder (in use by 876). However they don't appear to have used the system for representing decimal fractions, thus losing much of the power of such a system.

The other major contributions of the Hindu culture to mathematics were in the field of trigonometry (introduction of the sine function) and solutions to algebraic equations, though they did not use algebraic notation.

5.2.2 The Arabic World

Mohammed (570 - 632) founded the Muslim religion about 620, he had set up an Islamic state by 630, which had expansionist views. The Islamic world expanded into the Middle East, North Africa and Spain in a paroxysm of religious fervour during the period from about 630 AD to 750 AD. They conquered Alexandria in 651 AD. At first the new conquerors showed little interest in mathematical, or any knowledge not religiously sanctioned. There is a story (almost certainly untrue) which tells how the Caliph Omar, when asked what to do with the books in the great library at Alexandria, said to burn them since "Either they agree with the Koran, and are hence superfluous, or they don't, in which case they are worse."

By 750 the Muslim world had settled down and the conquerors absorbed some of the learning and culture in the region. Caliph al Mamun (809 - 833) established a "House of Learning" in Baghdad, similar to the library in Alexandria. It is thanks to this institution that much of the work of the Greek period was preserved. Baghdad was ideally situated in between the Greek and Indian worlds, and on top of the remains of the ancient Babylonians. Its scholars drew from all of these sources and synthesised much of what they learned.

A member of the faculty at the House of Learning was Mohammed ibn-Musa al Khwarizmi (780 ? - 850). Al Khwarizmi wrote at least twelve astronomical and mathematical texts, summarising much of what was known at the time. Significantly he drew on Indian as well as Greek sources. In one of his books *De Numero Indorum* ("On the Indian system of number") he gave a full account of the Indian decimal place holding system. When al Khwarizmi's works became known in Europe they attributed the number system to the Arabic author, thus giving rise to the false idea that our current system of numeration derives from the Arabs, rather than the Indians. A Latinisation of his name was used to indicate the new system, which became known as "algorismi", from which we derive the word "algorithm".

The title of another of al Khwarizmi's works was *Hisab Al-jabr Wa'l Muqabalah* ("The Science of Equations") 'Al-jabr' means transposing a quantity from one side of an equation to another, and 'muqabalah' means cancellation of like terms on two sides of an equation. From the title *Al-jabr* we derive the word "algebra". The treatment is of the solution of certain types of quadratic equations, probably much derived from the Indian world. The notation is still completely rhetorical, that is not algebraic, and represents a step back from that given by Diophantus 500 years earlier. However this book became widely known and was the foundation from which modern algebra was built. *Al-jabr* includes rules on manipulation of algebraic equations, and identifies six fundamental types of quadratic equation. The book also includes much of the material of Book II of Euclid, but considered from an algebraic, rather than geometric standpoint.

A couple of generations later we have Thabit ibn Qurra (836 - 901). He copied and translated many Greek works into Arabic. It is from Thabit that the only existing copies of the work of Appollonius come.

Another Arabic mathematician was Omar Kayyam (1050 - 1123). Omar was the court astronomer at the court of the Caliph of Merv. He considered solutions to certain cubic and quartic equations, but he believed, incorrectly, that the general solution to cubics was impossible, and solutions would have to involve the use of conic sections. Omar Kayyam also worked in geometry, and was familiar with Euclid. He attempted, unsuccessfully, to prove the parallel postulate.

By the 12th century incursions by the Europeans and Mongols caused political and social upheavals in the Muslim world and the flame of learning dimmed.

The Arabic period saw some development in the area of algebraic solutions of equations. However the Arabic mathematicians never really developed an algebraic notation. They used the rhetorical form of the ancient Greeks, a step back from the Roman Greeks such as Diophantus.

Significantly this period saw the inclusion of the material covered by the Hindu civilisation in India in the general mathematical literature. It seems likely that the interest in solving algebraic equations came from this source.

5.2.3 Fibonacci

We Now consider a man who really lives on the cusp of the European and Arabic worlds, Leonardo of Pisa (1175 - 1250), better known as Fibonacci. Fibonacci was the son of a merchant of Pisa, who served as customs officer in North Africa. Thus as a boy Fibonacci travelled widely and learnt Arabic. He was taught by a Muslim teacher and learnt Arabic mathematics.

When he finally returned to Italy he wrote a book detailing a number of Arabic mathematical systems, including the Hindu system of numeration (complete with zero) in Latin. This was the first significant mathematical work in Europe for almost a thousand years.

Fibonacci also wrote a number of other mathematical works, but they were generally too sophisticated for his European contemporaries. Thus his work went largely unnoticed in Europe. Interestingly Fibonacci used letters to indicate variables, a significant algebraic advance.

The Fibonacci Numbers

Included in this work was a description of the sequence which bears his name, the Fibonacci numbers.

> How many pairs of rabbits will be produced each month, beginning with a single pair, if every month each 'productive' pair bears a new pair which becomes productive from the second month on.

We assume that deaths do not occur.

We denote the number of pairs alive in the nth month by u_n. So $u_1 = 1$.
In the next month this pair produces a new pair, so $u_2 = 1 + 1 = 2$, the original pair and the new pair.
In the nth month the number of pairs of rabbits alive will be a number of new rabbits, produced by the productive pairs, which are those which are two or more months old, u_{n-2}, as well as the number already alive in the previous month u_{n-1}.
Thus $u_n = u_{n-1} + u_{n-2}$.
The sequence thus produced is $1, 1, 2, 3, 5, 8, 13, 21, \ldots$. This is an example of a recursively defined sequence, in which the next term is defined by preceding terms. This is in contrast to an explicit formula, in which each term is defined in terms of n. An example of such an explicitly defined sequence would be $v_n = 2n + 1$, which generates the sequence of odd numbers $1, 3, 5, 7, 9, 11, \ldots$ It is clear that the Fibonacci numbers do not follow any obvious explicit formula (try to find one).

Interestingly an explicit formula for the nth Fibonacci number u_n was found in the 19th century this gives:

$$u_n = \frac{1}{\sqrt{5}}\left(g^n - \left(\frac{1}{g}\right)^n\right)$$

where $g = \frac{1}{2}(1+\sqrt{5})$, the golden ratio!

If we consider the ratio of successive terms, $\frac{u_n}{u_{n-1}}$, we get

$$\frac{u_n}{u_{n-1}} = \frac{g^n - \left(\frac{1}{g}\right)^n}{g^{n-1} - \left(\frac{1}{g}\right)^{n-1}}.$$

Now $\frac{1}{g} < 1$, so as $n \to \infty$ the second term $\left(\frac{1}{g}\right)^n$ tends to 0 in both the numerator and the denominator. Thus the limit of the terms is g, or in mathematical notation $\lim_{n\to\infty} \frac{u_n}{u_{n-1}} = g$. This convergence is in fact quite fast, $g = 1.61803\ldots$, $\frac{u_9}{u_8} = \frac{21}{13} = 1.615\ldots$, already correct to three figures.

5.3 The Renaissance

After the fall of the Roman empire in Europe a period of chaos ensued. Indeed Boethius (480-524) wrote many translations into Latin of Greek works, but these were 'dumbed down' versions with the harder parts, and hence the logical rigour, removed. Nevertheless a significant amount of classical material was preserved by remote monasteries in the far flung corners of the empire. These institutions kept the flickering flame of civilisation alive through the dark ages.

The empire of Charlemagne represented a brief hiatus in this chaos. Charlemagne encouraged learning in his subjects in the face of great opposition. However his edict that Cathedrals and Monasteries should have schools had a long lasting effect.

The 12th and 13th centuries saw the crusades and some of the Arabic knowledge slowly filtered into Europe.

The 14th century in Europe was an awful time, with wave after wave of outbreaks of the Black Death. This terrible scourge wiped out between one third and one half of the population of Europe. Most people had enough to worry about just surviving, mathematics was not an option.

By the middle of the 15th century Europe had staggered to its feet and a number of events triggered an explosion of learning and exploration there, first in Italy, then slowly spreading northwards. This is what is known today as the Renaissance. The first of these events was the invention of movable type (1447), making books widely available. The next was the fall of Constantinople (1453) and the collapse of the Byzantian Empire which led to an influx of well read refugees, particularly in Italy. The final event was the conquest of Muslim Spain (1450 - 1492). This resulted in much closer contact with the Muslim world and the discovery of libraries containing classical Greek and Roman works as well as contemporary Arabic works, which were translated.

The most significant development in mathematics in this period was the development of modern algebraic notation. The focus was, as inherited from the Arabs, the solution of algebraic equations. Significant names from this period are Scipione del Ferro (a.k.a Tartaglia) (1465 - 1526), Geronimo Cardano (1501 - 1576), Luigi Ferrari (1522 - 1560) and Francois Viete (1540 - 1603).

This is the age of Copernicus (1473 - 1543), Keppler (1571 - 1630) and Galileo (1564 - 1642). This period saw a war between the Christian Church, who had been the uncontested power and authority in the Medieval period, and the new logical *scientific* theories. Many of these people had to be very careful of what they said in view of Church sanction.

Much of the mathematical activity from the fall of Rome till about 1600 was concerned with finding solutions to algebraic equations. In particular the following standard equations, a, b, c, d and e are given constants.

- The quadratic equation $ax^2 + bx + c$.
- The cubic equation $ax^3 + bx^2 + cx + d$.
- The quartic equation $ax^4 + bx^3 + cx^2 + dx + e$.

Many authors gave solutions to particular cases of these equations. Though the general form as written of these equations, as written above, was not recognised until fairly late in the game. Part of the problem was recognising that a particular problem was indeed simply asking for a solution to one of these equations. Consider the Mesopotamian problem of finding a pair of numbers, given their sum and product, it is not immediately obvious that this is in fact asking to the solution to a particular quadratic equation.

One of the main achievements of this period was to develop algebraic notation, using variables in the way we now take for granted. The use of variables clarifies the problems enormously.

By 1600 the algebraic notations were recognisable as modern, though the full modern algebraic notation is more properly due to Descartes (1569 - 1650), Leibniz (1646 - 1716) and Euler (1707 - 1783).

The modern age is generally taken to begin about 1600. A full account of this period would require a book all on its own, we will skim some of the high points.

5.4 Early Rationalism

The first great mathematicians of the modern age where Rene Descartes (1569 - 1650) and Pierre de Fermat (1601 - 1665).

5.4.1 Descartes

Descartes is probably better known today for his philosophy, particularly his statement 'Cognito ergo Sum', 'I think therefore I am'. However, often what Descartes was thinking about was mathematics.

Descartes is usually accredited with the invention of co-ordinate geometry, thus we talk of the standard xy co-ordinate system as the *Cartesian* plane. In fact, in its modern form, it is probably more accurately due to Fermat. Interestingly both of these men came up with the idea independently when considering a problem posed by Apollonius over one thousand years earlier.

5.4.2 Fermat

Pierre de Fermat (1601 - 1665) was less of a philosopher than Descartes, but undoubtedly more of a mathematician. The son of a leather merchant he trained as a lawyer and became the Kings Councillor at Toulouse. Fermat was quiet and retiring, spending many hours shut away in his study. He published very little during his lifetime, but kept up a correspondence with other mathematicians of the day, on his death his journals and diaries where published.

In the margin of one of his books, a copy of Diophantus translated into Latin, was a famous note. The problem of Diophantus (Problem 8 of Book II) was to 'divide a square number into the sum of two squares', i.e. to find integer solutions to the equation $x^2 + y^2 = z^2$. Indeed this is the problem of finding integer 'Pythagorean' triples, $3, 4, 5$ is a solution. Fermat had noted in the margin:

> To the contrary it is impossible to separate a cube into two cubes [i.e. find integer solutions of $x^3 + y^3 = a^3$], a fourth power into two fourth powers [i.e. find integer solutions of $x^4 + y^4 = a^4$], or, generally any power above the second degree into powers of the same degree [i.e. find integer solutions of $x^n + y^n = a^n$ where $n > 2$]: I have discovered a truly marvellous demonstration which this margin is too narrow to contain.

Alas Fermat seems not to have written down his proof, for truly marvellous it must have been. This problem was only finally solved in 1995, despite enormous effort over the years. This proposition became known as Fermat's last theorem, there has been much debate as to whether Fermat actually had a proof or whether his proof was in fact flawed.

The next generation of mathematicians saw the development of the general method known today as Calculus. In fact much of the material was known to Fermat, but he failed to bring it together into a whole. We will consider this in the next chapter.

Chapter 6

The Infinitesimal in Mathematics

6.1 The Development of Calculus

The two independent developers of Calculus where Sir Isaac Newton (1642 - 1727) in England and Gotfried Wilhelm Leibniz (1646 - 1716) in Germany. Newton had first developed the theory of *fluxions* as he called it by 1666, however he did not publish. In 1684 Leibniz published his first paper on the subject, though he appears to have been in possession of the basic ingredients by 1676. A probable cause for the delay was the realisation of the problems associated with limits.

When Leibniz first published all went well, Newton merely noted that he also had developed a similar theory, which he then made public. However as the years passed other people intervened. Leibniz was accused, by British mathematicians, of stealing the result from Newton. Leibniz was scandalised and denied all accusations. As time progressed the debate became more heated with the English supporting Newton and the continental Europeans supporting Leibniz. Eventually the two main protagonists Newton and Leibniz where themselves drawn into the argument permanently souring relations between them.

The general feeling today is that these two men did develop Calculus independently, while it is possible that Leibniz saw some of Newtons material it is unlikely to have been much use to him.

6.1.1 Newton

Newton was undoubtedly the greatest Physicist who ever lived, as well as being one of the greatest mathematicians. Einstein's statement 'We stand on the shoulders of giants' referred to Newton (Newton himself made this statement in reference to Archimedes). The achievements and life of Newton could easily fill a book on their own, and have. Newton investigated the theory of light, explained gravity and hence the motion of the planets. He is also famed for inventing 'Newtonian Mechanics' and explicating his famous three laws of motion. Possibly even more important he formulated a rigorous method of mathematical modelling, in which a model is proposed and then refined.

At the end of his book *Optica,* on the study of light, Newton makes 31 'Queries', or conjectures which he thinks are worthy of further study. These queries are fascinating, in that they show that Newtons thought was years ahead of its time. The very first one

> Do not Bodies act on light at a distance, and by their action bend its rays?

This predicts one of the central tenets of Einstein's theory of relativity over 200 years before it was posed. Another query states:

> Are not gross Bodies and Light convertible into one another, and may not Bodies receive much of their activity from the particles of light which enter into their Composition? ...
> The changing of Bodies into Light, and Light into Bodies, is very conformable to the Course of Nature, which seems delighted by such Transformations.

This predicts Einstein's formula $E = mc^2$, energy and mass are interchangeable forms of the same thing.

Though Newton was an accomplished scientist and mathematician, it must be noted that he was a man of his time, and as such spent much time engaged in alchemical research, searching for the elixir of life and the philosophers stone. Newton was the first great scientist, he was also the last great alchemist.

6.1.2 Leibniz

Leibniz was something of a philosopher as well as a mathematician, he could well be compared with Descartes. However many of his philosophical beliefs are considered outdated, Bertrand

Russell referred to them as 'fantastic'. Leibniz invented a mechanical calculating machine which would multiply as well as add, the mechanics of which where still being used as late as 1940. He also did work in discrete mathematics and the foundations of logic.

Leibniz believed that language was paramount, he thought that if we could speak in a perfect language it would be impossible to make untrue or incorrect statements. As a result of his focus on language his notations were very good, he was the first to use the notation $f(x)$. The notation used today in Calculus $\frac{\mathrm{d}f}{\mathrm{d}x}$ and $\int f(x)\mathrm{d}x$ is Leibniz' notation. He paid a great deal of attention to language in his formulation of Calculus, as a result his notation was far superior to Newton's, and has been adopted.

6.1.3 English Mathematics

Around this time there was a flourishing of mathematical activity, many people who have given their names to theorems in Calculus lived in this period. As a result of the dispute between Newton and Leibniz the mathematical world fractured into British mathematicians, and the rest of Europe. There was still considerable contact, but approaches varied. In England Colin McClaurin (1698 - 1746) and Brook Taylor (1685 - 1731) showed that any 'differentiable' function could be approximated by polynomials, this result is known today as 'Taylor's Theorem', with 'McClaurin Series' being a special case.

Abraham De Moivre (1667 - 1754) was originally a French Huguenot, but he left for England fearing religious persecution. He is responsible for De Moivre's Theorem, a version of which may be succinctly stated as

$$e^{i\pi} + 1 = 0$$

What is interesting about this theorem (aside from its very real mathematical applications) is that it takes the three most esoteric, but naturally occurring numbers known to mathematics, e the base of the natural logarithm ($e \approx 2.71828\ldots$), π the circumference of a unit circle ($\pi \approx 3.14159\ldots$), and the square root of minus one $i = \sqrt{-1}$, and puts them together with the simplest numbers, 0 and 1, in a natural way.

6.1.4 The Bernoullis

In Europe there was also a flourishing of activity, the most significant players where the Swiss Bernoulli brothers, Jean Bernoulli (1667 - 1748) and Jacques Bernoulli (1654 - 1705). Jean and Jacques were the staunchest supporters of Leibniz. The Bernoulli family kept producing mathematicians, Jean's three sons, three grandsons, a great grandson and a great

great grandson (Jean Gustave Bernoulli (1811 - 1863)) where all mathematicians, though none them of the eminence of Jean and Jacques.

Both brothers did extensive work in the early development of Calculus. In addition they created an extensive culture of mathematics, posing interesting problems for solution (Newton would enter these competitions anonymously).

Jean Bernoulli is accredited with writing the first Calculus textbook in 1692. He was living in Paris at this time, earning money by teaching. One of his students was a French marquis, the marquis de L'Hopital (1661 - 1704). L'Hopital, a wealthy man, saw an opportunity for immortality through Jean and took it. He offered to pay Jean a regular salary if he could have all of Jean's mathematical results to do with as he pleased. As a result of this deal Jean Bernoulli's theorem about limits is known to this day as L'Hopitals rule.

6.2 Limits and the Infinitesimal

We have already seen how the idea of infinitesimals played an important part in Greek mathematics, through the paradoxes of Zeno and the method of exhaustion. These ideas where finally brought together in the field of Calculus.

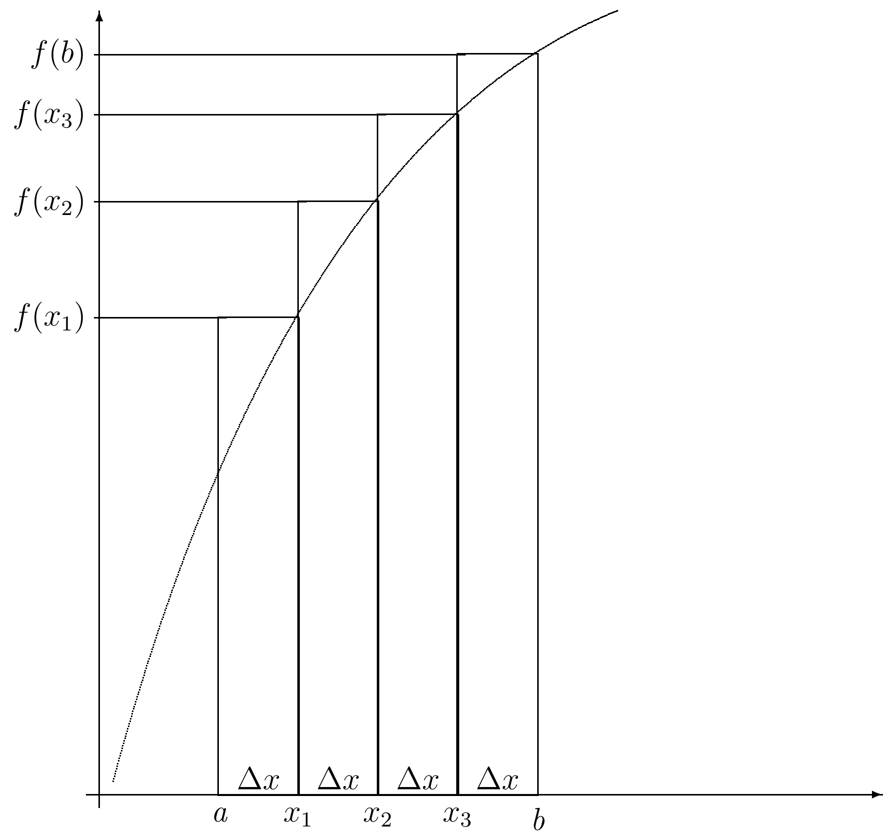
6.2.1 Integration

In Calculus there are two major threads, differentiation and integration. Integration seeks to find the area bounded by a curve between two points a and b. Thus integration is a method of quadrature, and was originally thought of as such. The idea is similar to the method of exhaustion. In order to find the area we divide the region under the curve into n equal intervals along the x axis, creating n thin strips, each strip has width Δx. The area of each of these strips is approximately $f(x_i)\Delta x$ where x_i is the rightmost point in the nth strip. [1] The sum of the areas of all these strips will be approximately the area under the curve. As the number of strips goes to infinity and the width of each strip, Δx, goes to 0 the approximation becomes more accurate. *In the limit,* as n tends to infinity and Δx tends to zero, the approximation becomes exact, thus the integral is defined as:

$$\int f(x)\mathrm{d}x = \lim_{n\to\infty} \sum_{1}^{n} f(x_i^*)\Delta x$$

[1]In fact x_i can be any point in the ith strip

Figure 6.1: Integration



In figure 6.1 we have split the region from a to b into four pieces, generating the points $a(= x_0), x_1, x_2, x_3, b(= x_4)$. Each of the resulting rectangles has width Δx, and the ith rectangle has height $f(x_{i+1})$. Thus the area of the rectangles is $f(x_{i+1})\Delta x$, and the area under the curve is (very) approximately the sum of the areas of the rectangles. To get a better approximation we halve the width Δx and double the number of regions. In the limit as Δx tends to 0, the approximation becomes exact.

From this formulation of integration the relation to the Greeks method of exhaustion is clear. It should be noted however that this formulation of integration is essentially due to Riemann, and is a 19th century account.

6.2.2 Differentiation

Differentiation seeks to find the slope of the tangent line to a curve. Any slope may be given as 'rise over run'. We wish to find the slope of the tangent to the curve given by $f(x)$ at the point a (see figure 6.2).

Consider a point nearby a, at $a+\delta x$, where δx is a small amount. The value of the function at $a+\delta x$ is $f(a+\delta x)$. Now the line joining $(a, f(a))$ to $(a+\delta x, f(a+\delta x))$ is a fair approximation of the tangent. This line has run (x length) $a + \delta x - a = \delta x$ and rises $f(a + \delta x) - f(a)$. *In the limit* as δx tends to zero $a + \delta x$ tends to a and this line tends to the tangent line. Thus we define the derivative $\dfrac{\mathrm{d}f}{\mathrm{d}x}$ by

$$\frac{\mathrm{d}f(a)}{\mathrm{d}x} = \lim_{\delta x \to 0} \frac{f(a + \delta x) - f(a)}{\delta x}.$$

The derivative is also denoted by a prime on the function, thus $f'(a) = \dfrac{\mathrm{d}f(a)}{\mathrm{d}x}$.

The two ideas of integration and differentiation are related by *The Fundamental Theorem of Calculus*

> Differentiation and integration are inverse operations to each other, that is
>
> $$\frac{\mathrm{d}}{\mathrm{d}x}\left(\int f(x)\mathrm{d}x\right) = f(x).$$

6.2.3 Limits

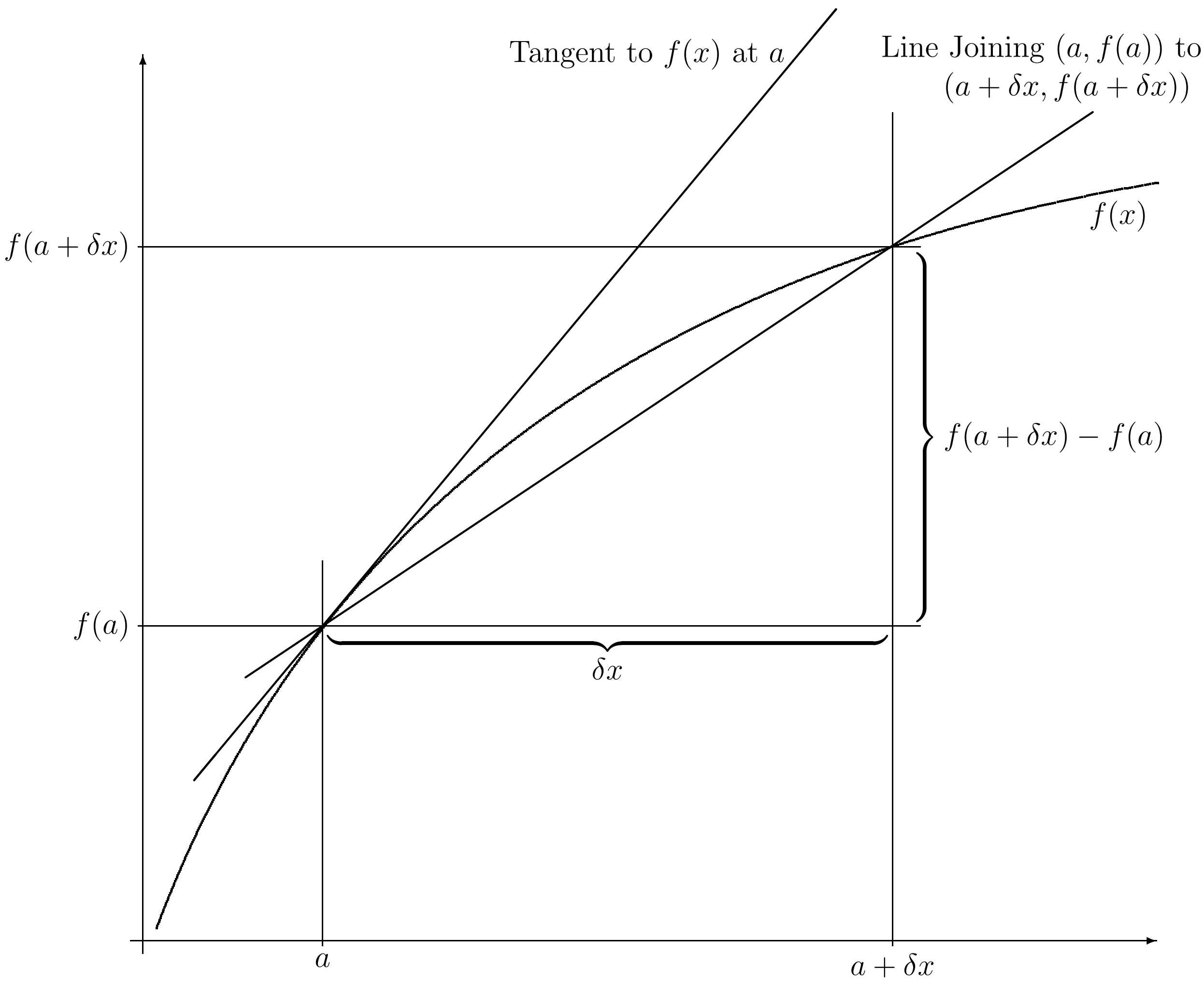
Ever since the time of Zeno man has been dancing around the idea of a limit. Finally with the development of calculus this idea became formalised. We can see that both integration and differentiation use the idea of a limit in a fundamental way.

Unfortunately nobody was able to come up with a precise definition of limit. It seems hard to avoid talking of 'infinitesimally small quantities' in such a definition. What do we mean by 'infinitesimally small'? Newton tried the following using 'evanescent ratios':

> [By the limit of a ratio] is to be understood the ratio of the quantities, not before they vanish, nor after, but that with which they vanish.

This makes no sense, Newton is trying to say that a limit is both zero and non zero at the same time. Newton realised the problem and reformulated his approach using what he called *the prime and ultimate ratios,* but this was not much better.

Figure 6.2: Differentiation



Leibniz also tried to define a limit, with about as much success.

> ... when we speak of...infinitely small quantities (i.e. the very least of those within our knowledge), it is understood that we mean quantities that are...indefinitely small...

Which ends up defining 'infinitely small' quantities as 'indefinitely small' ones, not much help.

In order to find the derivative of $f(x) = x^2$ at $x = a$ Newton and his contemporaries might have tried the following argument.

We start with the function $f(x) = x^2$ at some point a, then $f(a) = a^2$.
We now consider a small increment δx in x, so a goes to $a + \delta x$ and $f(a + \delta x) = (a + \delta x)^2$.
Now $f(a + \delta x) - f(a) = (a + \delta x)^2 - a^2$.
Expanding the bracket gives $f(a + \delta x) - f(a) = a^2 + 2a\delta x + (\delta x)^2 - a^2 = 2a\delta x + (\delta x)^2$.
Now we divide by δx to get $2a + (\delta x)$.
Finally we let δx become zero to get that $f'(a) = 2a$. $\square$

The problem is that in the penultimate step we divide by something which, in the ultimate step, we set equal to zero. Thus we are dividing by zero. When we divide we get $\dfrac{2a\delta x}{\delta x} + \dfrac{(\delta x)^2}{\delta x}$. Taking $\delta x = 0$ gives $\frac{0}{0} + \frac{0}{0}$. But what are we to make of the quantity $\frac{0}{0}$? It is in fact an undefined quantity.

To understand the problem consider the function $f(x) = \frac{x}{x}$. What is the value of this function at zero? $f(0) = \frac{0}{0}$. Now consider $\dfrac{1}{0}$ as representing infinity [2] (∞). Now we could write $f(0)$ as $0 \times \dfrac{1}{0} = 0 \times \infty$ [3]. But we know two contradictory rules:

1. Any finite number times zero is zero
2. Any non zero quantity times infinity is infinite[4].

---

[2] If we consider $\frac{1}{x}$, the closer x gets to zero the larger $\frac{1}{x}$ becomes, thus we use the symbol ∞ to represent this unbounded behaviour. Note though that depending whether $x > 0$ or $x < 0$ we will get ∞ or $-\infty$, thus an "infinitesimal" change in x from "just bigger than 0" to "just less than 0" takes $\frac{1}{x}$ from ∞ to $-\infty$.

[3] This is an absolutely abhorrent thing to do, as any mathematician will tell you. ∞ is not a number, it is a symbol which represents the notion of unboundedness, thus it is not (strictly speaking) possible to multiply or divide by it. However for the purposes of this argument I have taken some liberties. What this really represents is the *form* of the function near zero.

[4] Well, it seems reasonable doesn't it? What we are really saying is that if x is unbounded and $a \neq 0$ is a constant then ax will also be unbounded.

What then are we to make of $0 \times \infty$? It has no meaning, it is undefined.

However, if $x \neq 0$ then $f(x) = \frac{x}{x} = 1$, by cancelling the x's. Thus at any point near 0, but not equal to it, $f(x) = 1$. It seems reasonable then that we should allow $\lim_{x \to 0} f(x) = 1$, since all the values of f 'infinitesimally' close to 0 are 1.

To see that $0 \times \infty$ is really not well defined, consider the function $g(x) = \frac{ax}{x}$, where a is a constant. $g(0)$ again has the form $0 \times \infty$. Now if $x \neq 0$ we have that $g(x) = a$, so we may conclude that $\lim_{x \to 0} g(x) = a$. But a is any constant, so $0 \times \infty$ can take any value, depending on the function from which it came.

The early view of limit was that we either understand a limit as being infinitesimally close to the value, but not equal to it (Leibniz view). But then how do we define 'infinitesimally'? Alternatively we may view δx as an 'evanescent increment', one which is both zero and non zero (Newton's view).

The problems with limits caused enormous worries for the early practitioners of the calculus. How could calculus be justified when it ultimately rested on such a shaky foundation? This was compounded by the fact that much new science was being based on calculus. If the calculus was flawed, what of the scientific theories that rested on it?

In 1734 Bishop George Berkeley (1685 - 1753) published *The Analyst.* The central idea of this book was an attack on mathematics and hence science. Bishop Berkeley's argument essentially was that science and calculus ultimately rests on an act of faith, belief in limits. Thus, why not equally make that act of faith a belief in God? In Bishop Berkeley's words:

> ... a Discourse Addressed to an Infidel Mathematician. Wherein It Is Examined Whether the Object, Principles, and Inferences of the Modern Analysis are More Distinctly conceived, or More Evidently Deduced, than Religious Mysteries and Points of Faith.

The 'infidel mathematician' was probably Sir Edmund Halley (1656 - 1742), famed for predicting the orbit of 'Halley's' comet.

Berkeley's book was not just a vitriolic attack on mathematical ideas, it was well argued by someone who was well versed in current notions of the theory. By the methods of the time a derivative or integral was calculated by incrementing the variables, and then later on assuming that these increments are zero (after having divided by them). Berkeley lambasted the notion of limit as it then existed.

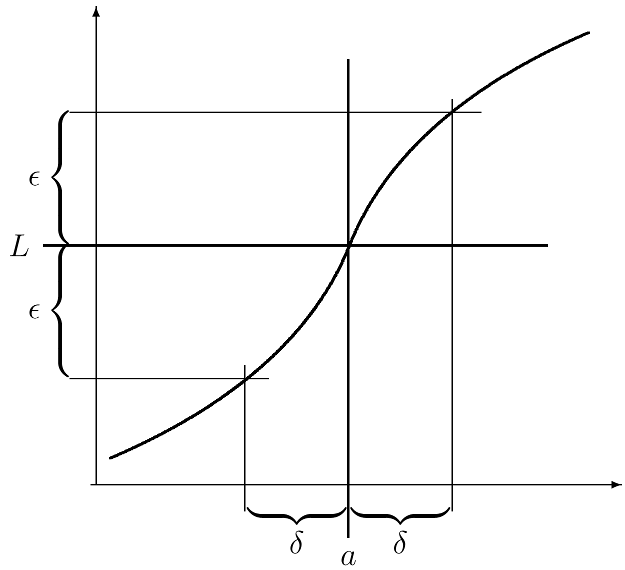
> And what are these fluxions? The velocities of evanescent increments. And what are these same evanescent increments? They are neither finite quantities, nor

quantities infinitely small, nor yet nothing. May we not call them ghosts of departed quantities?

...he who can digest a second or third fluxion, a second or third differential, need not, methinks, be squeamish about any point in divinity.

In its favour calculus had never been shown to give an incorrect answer, this led the mathematicians of the day to believe that there must be some solid foundation, if only they could find it!

Figure 6.3: Definition of a Limit



The first real breakthrough came when d'Alambert considered limits in 1768. He first of all rejected both Newton's and Leibniz' view.

... a quantity is something or nothing; if it is something, it has not yet vanished; if it is nothing it has literally vanished. The supposition that there is an intermediate state between these two is a chimera.

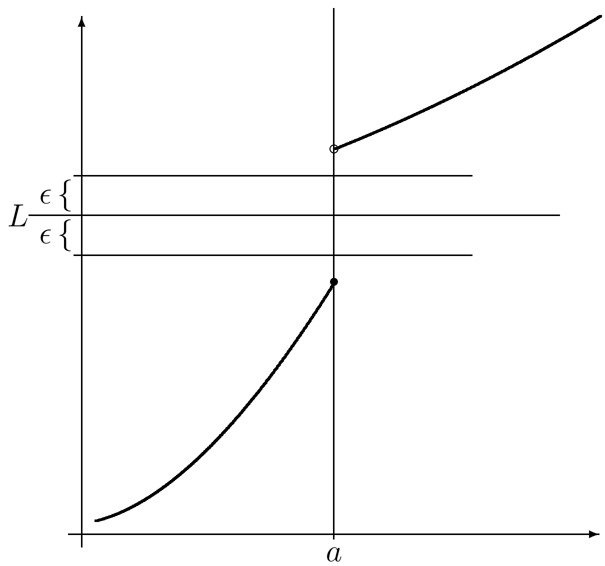
He then went on to define the limit of one variable in terms of another:

One magnitude is said to be the limit of another magnitude when the second may approach the first [to] within any given magnitude, however small, ... so that the difference of such a quantity to its limit is absolutely unassignable.

The fundamental difference here is that we are viewing a limit as being defined by a process, that is a variable. Limits only make sense, in this view, as limits of functions of one variable with respect to another. In this view there is no necessity to talk about 'infinitesimally small' magnitudes, all we are saying is that for any given magnitude we can find a a smaller one, one closer to the limit, ala Eudoxus. The main problem with this definition is that it is still somewhat vague. Over the next few years other mathematicians worked on making this definition more precise, notably Gauss and Dedekind.

This definition is reminiscent of Eudoxus' definition of ratio and the principle of exhaustion. What d'Alambert is saying is that we can say that $\lim_{x \to a} f(x) = L$, if whenever we are given some fixed (positive) magnitude, $\epsilon > 0$ say, then there is some other magnitude, δ say, such that whenever x is within δ of a, $f(x)$ will be within ϵ of L. The real power of course comes from the fact that this is true for any ϵ, no matter how small. In mathematical notation $\forall \epsilon > 0, \exists \delta > 0$ such that $|x - a| < \delta \Rightarrow |f(x) - L| < \epsilon$. In figure 6.3 the given value of ϵ gives rise to a value of δ.

Figure 6.4: Jump Discontinuity

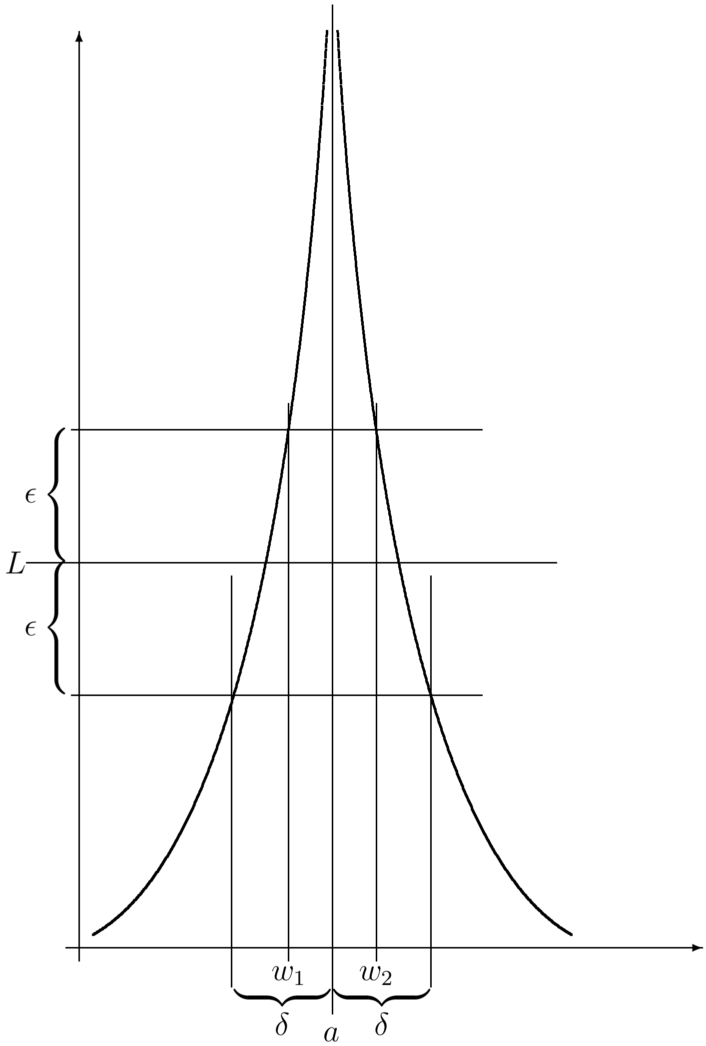


To see how this definition works it is instructive to consider those cases when the limit does not exist, when the definition breaks down. One way that this can happen is if the function has a *Jump discontinuity* at a, that is the function jumps from one value to another. In figure 6.4 $f(x)$ 'jumps' at $x = a$. In this case, wherever we take L, if ϵ covers any part of the discontinuity, no matter how close we are to a, it is impossible to get $f(x)$ to be within ϵ of L.

Another type of discontinuity is the essential singularity, that is somewhere where the function becomes progressively bigger, becoming unbounded at some point. An example of such an essential singularity is the function $\frac{1}{x-a}$ at $x = a$.

In figure 6.5 the ϵ values appear to generate δ values, but for any point between w_1 and w_2 the curve does not lie within ϵ of L. Since the curve is unbounded this will happen whatever L value we take.

Figure 6.5: Essential Singularity



Chapter 7

The Modern Age

7.1 The Age of Rationalism

The 18th century is often known as the age of rationalism. Newton's theories of motion and gravity had conquered fundamental questions of how the universe worked. Calculus by this time had been developed into a powerful tool, and mathematical modelling became the rule. During this period 'rational thought' (i.e. mathematical and logical analysis) was brought to bear on almost every sphere of human activity.

7.1.1 Euler

The greatest mathematician of the 18th century was undoubtedly Leonhard Euler (1707 - 1783). Euler was born in Basle Switzerland, the son of a Calvanist Pastor and amateur mathematician. Euler entered Basle university and had earned his masters degree by the age of 17. At Basle he met Jean Bernoulli's two sons Nicolaus Bernoulli (1695 - 1726) and Daniel Bernoulli (1700 - 1782), both of whom took positions at the University of St. Petersburg in Russia, these two men secured a position at St. Petersburg for Euler. Euler stayed at St. Petersburg from 1727 to 1741, however the political political situation in Russia at this time was very unstable, government spies were everywhere, purges where the rule. When later asked about his reticence to speak by the Queen of Prussia, Euler is reported to have replied 'Madam, I come from a country where if you speak, you are hanged'.

In 1741 Euler accepted a new post at the Berlin academy, court mathematician to Frederick the Great of Prussia, Euler stayed in this position for 25 years. By 1766 the political situation in Russia had settled down and Catherine the Great of Russia was in power. At this time Euler accepted an new appointment in St. Petersburg where he was hailed as a returning hero. Now 59 Euler started to lose his sight, but amazingly he managed to keep up his mathematical work by dictating to his sons. He was apparently able to do extensive and complex calculations in his head, remembering every step so that he could recite them for his sons to record. Euler died in 1783 at the age of 76, active to the last.

Euler's work was prodigious, he published more than 500 books and papers during his lifetime, with another 400 appearing posthumously, his collected works span 74 volumes. Euler worked in nearly every field of mathematics, and theoretical physics enhancing and unifying what he found. Euler was also active in the popularisation of mathematics and science, he designed textbooks for Russian elementary schools, his lessons to Frederick the Great's niece Princess Anhalt-Dessau where written up in *Letters to a German Princess,* which was translated into seven languages and became a best seller. In addition Euler represents the last step in the development of totally modern notations. He is accredited with the first use of the Greek letter π to represent π, e to represent the base of the natural logarithm and i for $\sqrt{-1}$.

Euler represents the last mathematician who was able to encompass the whole subject. By the end of his life mathematics had become a series of many subdisciplines.

There is a story told of Euler during his second stay in Russia.

> Euler proclaimed that he had a proof of the existence of god. In order to contest and demonstrate his proof a famous philosopher of the day, Diderot (a non mathematician), was brought in to debate the supposed proof. Euler arrived and proclaimed 'my proof that God exists is:'
>
> $$x = a + \frac{b^n}{n}.$$
>
> Poor Diderot was dumbstruck, he had no recourse to a mathematical argument and was forced to retire in shame and ridicule.

This story illustrates another point, the distinction between mathematically literate and the non mathematically literate. Anyone with any mathematical knowledge would immediately have questioned Descartes on the meaning of the symbols, and would soon have revealed that the statement was so much gobbledegook.

This illustrates the way in which knowledge of mathematics, or the lack of it causes a stratification in society. Those who know mathematics are able to use it to support arguments, whereas those who do not are mystified by such arguments, whether or not they have merit. There is a temptation to label all such arguments as spurious (as they sometimes are). On the other hand the mathematical cognoscenti often feel a sense of superiority, dismissing the muddle headed arguments of their non mathematical colleagues. This leads to two societies, two worlds separated by a gulf of mathematical knowledge.

7.1.2 The French School

In the 18th and early 19th centuries there grew up an influential school in France. These men of science, many of whom were mathematicians were active in the happenings of the day. Many where caught up in the French revolution (1793), and many held governmental office.

The foremost mathematicians of this group were Jean d'Alambert (1717 - 1783), Joseph Louis Lagrange (1736 - 1813), Pierre-Simon de Laplace (1749 - 1827), Adrien-Marie Legendre (1752 - 1830), Fourier (1768 - 1830), Simeon-Denis Poisson (1781 - 1840) and, possibly the greatest of them all Augustin-Louis Cauchy (1789 - 1857). All of these men were active in physics as well as mathematics, the form of mechanics used in Quantum mechanics is known as 'Lagrangian Mechanics'. 'Laplace's equation' is used to calculate heat transfers (and much else).

By 1820 there had been established in Paris a special school, the Ecole Polytechnique, which specialised in mathematical education. This school had some of the best mathematical minds on its staff and a regular journal. In addition the French Academy of Sciences particularly encouraged mathematics.

Another mathematician who comes at the end of this period is Evariste Galois (1811 - 1832). Galois is one of the most tragic and romantic figures of mathematics. The son of the mayor of a small town outside Paris, he undoubtedly had the potential to become one of the greatest mathematicians of all time. Galois proved that no polynomial of degree greater than five could be solved analytically. In the course of doing this he single-handedly laid the foundations of a whole new area of mathematics called Group Theory, he is remembered in the title of the subject Galois Theory. He did all of this before the age of 20.

Galois' early schooling was at the infamous lycee of Louis Grande in Paris. The conditions in this school were so bad that the students staged a rebellion, demanding better conditions. The rebellion was put down and Galois returned to the torture of lycee Louis Grande. At the age of 16 he recognised his mathematical genius and applied to the prestigious Ecole Polytechnique. However so sure was he of acceptance that he did not bother to study for the entrance exams, his lack of formal qualifications and a less than scintillating academic record in areas outside mathematics also counted against him, he was turned down.

In the Ecole Normal he disregarded his regular schooling and investigated the further reaches of higher mathematics. He was fascinated by Legendre's work on solutions to polynomial equations. In 1828 Galois came under the wing of Louis Paul Emile Richard, who was noted for fostering young mathematical talent, at last it seemed that things were going Galois' way. Galois secured a promise from Cauchy to review his first paper for submission to a *Memoirs* of the Academy of Sciences. However Cauchy, an old man by this time, forgot about the promise and the paper lay forgotten in his office.

Disillusioned by this unintentional rejection Galois tried again, submitting a paper summarising his results to the Academy of Science, in the hopes of winning the much coveted Grand Prize in Mathematics. Once again Galois' bad luck won out, Fourier received his paper for review, but died before he could read it. A final effort to get the paper accepted by the Academy was returned with a request by Poisson for further proofs. This was too much for Galois and he abandoned mathematics, joining the National Guard. He became involved in the aborted rebellion of 1830.

Tragically Galois died at the age of 21 in a duel with a cavalry officer over a woman. Galois was no marksman and expected to die. It is said that he spent the night before the duel furiously writing down his theories to send to his friend Chevalier for publication. His theories finally came to light in 1846 when Joseph Liouville (1809 - 1882) published this letter together with the two forgotten publications, and provided proofs and explanations of the new concepts.

7.1.3 Gauss

In Germany Carl Frederick Gauss (1777 - 1855) was known as the Prince of Mathematicians. Gauss was born to humble parents in Brunswick in Germany. His father was a labourer, with not much understanding of his son's special gift, he died when Guass was 31. However Gauss' mother gave the Gauss all the encouragement she could. After his father's death Gauss' mother came to live with him, which she did until her death in 1838 at the age of 97.

Gauss was a child prodigy, there is a story which he used to recount of his schoolmaster, one Bütner, a bit of a slavemaster. Bütner would assign long sums to do in order to keep his pupils quiet. When Gauss was 7, Bütner asked the pupils to sum all the numbers from one to one hundred. No sooner had the question left his mouth than Gauss' slate arrived at his desk, with the correct answer (the only correct one as it turned out). The young Gauss had found the Pythagoreans formula for triangular numbers $1+2+3+\ldots+n = \frac{1}{2}n(n+1)$, where in this case $n = 100$.

Luckily Bütner recognised Gauss' genius for mathematics and encouraged him by giving him special lessons. Gauss soon outstripped his teacher, but as luck would have it the school appointed a new 17 year old assistant, J.M Bartels (1769 - 1836). Bartels had a passion for mathematics, and went on to become a professor of mathematics himself. Bartels taught the young Gauss, by the age of 12 Gauss was able to criticise Euclid, by age 13 he was already groping toward the ideas of non Euclidean geometry. Bartels also had some influential contacts, bringing Gauss to the attention of the Duke of Brunswick, Duke Ferdinand. The Duke paid for Gauss to be sent to the local Gymnasium (high school) at 15. Here Gauss learnt the new Calculus.

In 1795 Gauss entered the university of Göttingen, with the generous Duke again footing the bill. Like Leibniz Gauss was a linguist of some note, when he first entered university there was some question as to whether he would study Classics or mathematics, luckily for mathematics he chose the latter. Gauss graduated from Göttingen in 1798, already he was publishing widely. After an aborted attempt to support himself by taking on students, the Duke once again came to the rescue, paying Gauss a yearly stipend.

In 1806 Duke Ferdinand, now 70, was killed while leading the Prussian forces against Napoleon. Gauss, whose fame was by now spreading, obtained a position at Göttingen, as director of the university observatory, and later as a professor of astronomy.

Gauss is one of the great men of mathematics, he did work in a wide range of fields, as well as working in astronomy, surveying techniques, and electromagnetism. The unit of magnetic force, 'the Guass' is named after him. As mentioned Gauss was known as the Prince of Mathematicians, and was considered the greatest mathematician of his day. When the occupying Napoleonic French demanded 2000 francs from Gauss in taxes, a sum which was totally out of his reach, he refused all offers of financial help. The sum was paid by Laplace in Paris without Gauss' knowledge or consent. Laplace said "[It is] an honour to assist the greatest mathematician in the world". Later when Gauss had the money he repaid Laplace in full, with interest.

7.2 Non Euclidean Geometry

There are three names usually associated with the initial development of non Euclidean geometries, these are Gauss, the Hungarian Janos Bolyai (1802 - 1860) and the Russian Ivanovich Lobachevsky (1793 - 1856).

It seems that the first person to investigate non Euclidean geometries was Gauss. However Gauss feared that his ideas would not be accepted and did not want to suffer the ridicule of those who did not understand his work. In fact during his life Gauss never published anything on non Euclidean geometry, it was only after his death in 1855 when his private notes where published that it was realised how far he had gotten with non Euclidean geometry before abandoning it as too controversial. However privately Gauss encouraged both of the others whose names are connected with non Euclidean geometry.

Bolyai's father, Farkas, had worked on the parallel postulate and was a good friend of Gauss. Young Janos decided to take up the problem of the 'parallel postulate', against the advice of his father. Janos soon realised that it was possible to develop a completely consistent geometry in which one allows infinitely many lines through a point parallel to a given line. The elder Bolyai, much impressed with his sons work, published it as an appendix to a textbook he was writing. This book was copyrighted 1829, but did not actually appear until 1832. Farkas also showed the work to Gauss, Gauss commented that he had had very similar notions himself. Gauss privately encouraged the young Bolyai, but would not support him in print.

Rather than suffer ridicule the young Bolyai's work suffered a much worse fate, it was largely ignored. Bolyai became discouraged and did not pursue his new idea further, thus it is Lobachevsky who gets most of the credit for the invention of non Euclidean geometry.

Ivanovich Lobachevsky (1793 - 1856) was the son of a minor government official who died when Lobachevsky was seven. Despite hardship and poverty Lobachevsky managed to attend Kazan University, Kazan is about 700 km east of Moscow. One of his professors there was none other than Gauss' boyhood friend and teacher Bartels. In 1814, at the age of 21, Lobachevsky started teaching at Kazan, where he spent the rest of his life.

By 1829 Lobachevsky had developed a notion of non Euclidean geometry similar to that of Bolyai, however Lobachevsky published his work immediately. Lobachevsky originally called his new geometry 'Imaginary Geometry', and from 1835 to 1855 he developed the theory into a whole, publishing 4 books on the subject. It is not clear how much influence Bartels had on Lobachevsky's early work, Bartels would have known of Gauss' work on the subject. Gauss learned of Lobachevsky's work and once again encouraged him privately, sponsoring him for entrance into the prestigious Göttingen academy of sciences, but he would not commit himself in public. As Gauss had predicted the subject caused enormous controversy, becoming known and accepted only very slowly. However when it did become accepted it led to a very different view of mathematics, and eventually the development of *formal systems* which we will consider later.

7.2.1 Riemann

The man who finally took non Euclidean geometry to its logical conclusion, and was responsible for publicising it widely was Georg Friedrch Bernhard Riemann (1826 -1866). Riemann was the son of a village pastor, again despite an impoverished childhood he attended Berlin university, going on to Göttingen to do his doctorate. He joined the faculty at Göttingen in 1854.

As was customary he gave an inaugural lecture, his topic was geometry. This lecture shook the foundations of the mathematical world. Riemann's notions of geometry were very different from that of either Euclid or indeed Lobachevsky. Riemann did not work from the parallel postulate, but rather gave an entirely new formulation of geometry, interpreted in the broadest of senses.

Rather than taking lines and circles as the fundamental objects of geometry, Riemann took points and functions to be the fundamental concepts. Riemann's idea of geometry considered points to be n-tuples of real numbers $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Just as (x, y) defines a point in the plane, and (x, y, z) defines a point in 3 space, an n-tuple describes a point in 'n dimensional' space, thus a geometry could have any number of dimensions. Any geometrical object, such as a line or surface can be described as a collection of points. Which points are on a surface is given by the solution to an equation involving a given function, $f(\mathbf{x}) = c$, where c is a constant. Lines are described by 'one parameter functions' $f(t)$, as t varies this describes the line.

Riemann identified a very special function, the distance function or metric. Thus in Riemann's view different geometric spaces may vary by having different notions of distance. In Euclidean geometry distance is defined by the Pythagorean formula $s^2 = x^2 + y^2$, this appears as a proposition, but in Riemann's formulation this is an axiom. Further, Riemann realised that other 'distance functions' where permissible. For example $s^2 = x^2 - y^2$, which allows for negative distances, such a distance function gives a 2 dimensional hyperbolic geometry (the geometry of the surface of a hyperbolae), one of the spaces investigated by Lobachevsky. Riemann then *defines* a straight line as the shortest distance between two points. In order to distinguish these 'straight' lines from the common notion of straight they are called *geodesics.* These geodesics may be far from straight in the usual sense of the word, especially if the metric is complicated.

In subsequent work it was discovered that many varied distance functions are possible, giving rise to many weird and wonderful geometries. Riemannian Geometry provides the cornerstone to Einstein's General Theory of Relativity (1915), current versions of this theory hold that (on a global scale) the distance function of the Universe is something like:

$$s^2 = -t^2 + R(t)(x^2 + y^2 + z^2), \text{ where } R(t) \text{ is a (growing) function of time.}$$

This geometry represents an expanding 4 dimensional hyperbolae. A 'Euclidean' 3 dimensional geometry in the x, y and z co-ordinates, the function $R(t)$ says that Universe is expanding in these three dimensions. The t (time) is negative, which indicates that the Universe is hyperbolic in this dimension. On a local scale the distance function would be far more complex, representing 'bumps' which correspond to stars and planets.

Riemann also realised that geometries could be defined solely by functions which act only on the geometry concerned, this allows the geometry of surfaces to be described without any reference to the space in which they are embedded. For example, using Riemannian geometry it is possible to completely describe the geometry of the surface of a sphere, without any reference to the surrounding space. This is particularly important when considering theories of the Universe, as when describing the Universe there can be no reference to space outside the Universe.

7.3 Cantor and the Infinite

7.3.1 Cardinal vs. Ordinal Numbers

(For an interesting fictional account of infinity see *White light* by Ruddy Rucker)

If we look closely at our notions of number we will see that in fact we have two different ways of conceiving of numbers. The first is as a method of counting (cardinality), the second is as an order relation (ordinality).

When we count a set of n objects we assign to each object an integer in sequence, starting from one. So each object is assigned a unique number from 1 to n. Such an assignment is called a one to one correspondence (or in mathematical parlance a *bijection*).

Definition 7.3.1 *A* one to one correspondence*, or* bijection *is a function, f, which maps a set S to a set T (we write $f : S \to T$) in such a way that*

1. *Each element of S is mapped to a unique element of T.*
2. *Every element of T is the image of some element of S.*

In a similar way we may use such correspondences to measure distance, by putting space in one to one correspondence with the real numbers. Each distance corresponds to exactly one unique real number. Of course we are ignoring any imprecisions of the physical act of measurement, in our ideal mathematical world this is the way things would be. This kind of view of number is known as cardinality.

The other notion of number is ordinality, in this notion we are only concerned with statements like 'x is bigger than y'. This is actually the notion of number we have used in the discussion of limits in section 6.2.3. Nowhere do we consider actual magnitudes, only relations of magnitudes.

While this distinction may seem pedantic it has two important ramifications. The first is in the field of measurement, If we are not able to give quantitative measure to an object of study we can still get some information from an ordinal measure, this is often the case in the social sciences. Thus we may say 'John is more emotional than Mary', though we may not quantify by how much, since we have no scale of measurement.

The second ramification is in the notion of infinity. With an ordinal representation of infinity we may only talk about infinity representing an unbounded quantity, as in the discussion of limits in section 6.2.3. The ordinal interpretation of infinity has a safety valve for dealing with the infinite. The infinities encountered when considering limits where 'potential infinities', meaning that they represent potentially unbounded quantities. A function never actually equals infinity, it merely gets larger and larger. At the point where it would become infinite it actually becomes undefined. Thus calculus avoids dealing directly with infinity.

On the other hand with a cardinal viewpoint we may consider a bijection between two infinite sets, the existence of such a bijection would imply that they have the same size. Thus we may consider sets with infinite cardinality, for example the set of natural numbers, or the set of real numbers between 0 and 1. This gives a much finer representation of the notion of infinity than an ordinal representation. The notion of an 'actual infinity' has always been contentious.

7.3.2 Cantor

The first person to really consider the notion of infinity in this way was Georg Cantor (1845 - 1918). Cantor was born in St. Petersburg, Russia, but moved to Berlin when he was 11. Cantor studied mathematics at Zurich, and ended up teaching at Halle university in Germany. Cantor never gained the recognition he felt he deserved, and during his life his ideas were ridiculed by the mathematical community. This effected him deeply and he spent much of his life in and out of mental institutions, suffering repeated breakdowns. However by the time of his death in 1918 his ideas were becoming widely accepted and his genius started to be recognised.

Cantor's basic idea was to use cardinality as a definition of infinity. Thus we may talk about the number of natural numbers, or the number of rational numbers and so on. If two infinite sets can be put into one to one correspondence with each other, then they represent the same infinity, just as any two sets which can be put into one to one correspondence with each other represent the same number.

We start with the number of natural (counting) numbers $\{1, 2, 3, \ldots\}$, this infinity is called countable infinity, a set of this size is called *countable.*

As an illustration of Cantorean arguments, consider the set of even numbers, $\{2, 4, 6, 8, \ldots\}$. Each even number may be represented as two times a natural number, thus m is even if and only if $m = 2n$ for some natural number n. But this represents a one to one correspondence between the set of natural numbers and the set of even numbers. For every even number, m, there is a unique natural number, $\frac{m}{2}$, and for each natural number, n, there is a unique even number, $2n$. The correspondence is represented by the function $f(n) = 2n$, it is because this function is *bijective* that the correspondence works.

Natural Numbers	1	2	3	4	5	...
Even Numbers	2	4	6	8	10	...

Thus the conclusion is that the number of natural numbers equals the number of even numbers, very strange. This leads to strange phenomena, such as Hilbert's hotel.

> Hilbert's hotel is a hotel with an infinite number of rooms, numbered 1,2,3 etc. One day the hotel is full, with an infinite number of guests. That day an infinite number of buses arrive carrying an infinite number of new guests. "No problem", says the manager, "we can accommodate you all". How does the manager accommodate the new guests?
>
> For each of the current guests the manager moves the person in room i to room $2i$, thus freeing up an infinite number of rooms for the new guests.

What about the number of integers? Consider the following correspondence, n is a natural number.

$$f(n) = \begin{cases} -\frac{n-1}{2} & \text{if } n \text{ is odd} \\ \frac{n}{2} & \text{if } n \text{ is even} \end{cases}$$

The output of this function goes as follows:

$$\begin{array}{cccccccc} n & 1 & 2 & 3 & 4 & 5 & \ldots \\ f(n) & 0 & 1 & -1 & 2 & -2 & \ldots \end{array}$$

It is evident that each integer corresponds to a unique natural number and visa versa. Thus the number of integers is the same as the number of natural numbers, it is countable.

What about the rational numbers, **Q**? In order to count the rational numbers we create an infinite table on which to count, the denominator increases as we move to the right, and the numerator increases as we move down (see figure 7.1). This enumerates all possible values of numerator and denominator. We now count diagonally starting at the top left, every time we reach the top we move right, and every time we reach the left hand side we move down.

$$\begin{array}{cccccc} \frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \cdots \\ \frac{2}{1} & \frac{2}{2} & \frac{2}{3} & \frac{2}{4} & \frac{2}{5} & \cdots \\ \frac{3}{1} & \frac{3}{2} & \frac{3}{3} & \frac{3}{4} & \frac{3}{5} & \cdots \\ \frac{4}{1} & \frac{4}{2} & \frac{4}{3} & \frac{4}{4} & \frac{4}{5} & \cdots \\ \frac{5}{1} & \frac{5}{2} & \frac{5}{3} & \frac{5}{4} & \frac{5}{5} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Figure 7.1: Cantor's Counting of the Rationals

This assigns a unique natural number to each rational number.

$$\begin{array}{cccccccc} n & 1 & 2 & 3 & 4 & 5 & 6 & \ldots \\ f(n) & \frac{1}{1} & \frac{1}{2} & \frac{2}{1} & \frac{3}{1} & \frac{2}{2} & \frac{1}{3} & \ldots \end{array}$$

Thus the rational numbers are countable, i.e. they have the same cardinality as the natural numbers.

We have seen that the natural numbers, the even numbers, the integers and the rational numbers are all countable, i.e. are the same size of infinity.

What about the real numbers **R**? We can simplify things by only considering those real numbers between 0 and 1, there are certainly an infinite number of them. We claim that this infinity is different than the previous ones, that is the set of real numbers between 0 and 1 is not countable. This means that there is *no* one to one correspondence between this set and the natural numbers. We are thus faced with the problem of showing that something does not exist, always much harder than showing that something does.

The answer is Cantor's much famed diagonalisation proof. This proof was highly contentious when Cantor first proposed it, but now it has become an accepted method of proof. It has been adapted to many circumstances, including the proof that the problem of deciding whether a machine (computer) will eventually halt on any input is insoluble (the halting problem).

Cantor's diagonalisation proof is essentially a proof by contradiction, thus the starting point is to assume that such a one to one correspondence exists. That is, to each real number between 0 and 1 we can assign a unique natural number. If this were so we could take an infinitely large sheet of paper and write out the natural numbers, n corresponding to the real numbers x. An example of how such a table might look is given in figure 7.2.

Figure 7.2: Cantor's Diagonalisation Proof

n	x
1	$0.\underline{1}986759143598725309861532\ldots$
2	$0.6\underline{5}69872345023458796234509\ldots$
3	$0.29\underline{3}8745723450972345234534\ldots$
4	$0.985\underline{4}918273450912346598764\ldots$
5	$0.1987\underline{5}23444098234734598723\ldots$
6	$0.23418\underline{9}7123487912349876123\ldots$
$\vdots$	$\vdots \qquad \ddots$

Now we will create a real number between 0 and 1 which cannot be on the table. On the ith row we identify the ith digit after the decimal point. In this example this gives the digits $1, 5, 3, 4, 5, 9, \ldots$.
We can consider this as a number between 0 and 1, namely $0.153459\ldots$.
Now, for each identified digit we change it to something else, for example we might add 1, changing 9 to 0.
This gives a new number, in this case $0.264560\ldots$.
This number, which lies between 0 and 1, cannot be on the list.

The reason is that it differs from the first number on the list in the first position after the decimal point, from the second number on the list in the second position after the decimal point, from the third number on the list in the third position after the decimal point, and so on.
In general it differs from the ith number on the list in the ith position after the decimal point, since we started with that digit and then changed it.
Thus we have created a number between 0 and 1 which does not have a natural number assigned to it, if it did it would be on the list.

This shows that the number of real numbers between 0 and 1 is not countable, this is a different infinity! This infinity is called the *continuum.* While it was surprising that seemingly different sized sets have the same infinite number of elements, once we have accepted this fact, it is even more surprising that there should be more than one type of infinity. It should be noted that it can be shown that the set of real numbers has the same cardinality as the set of real numbers from 0 to 1, thus the set of real numbers also has the size of the continuum.

This naturally raises the question of whether there are more infinities, indeed there are, a countably infinite number of them! For example, the number of functions from $\mathbf{R}$ to $\mathbf{R}$ represents a new level of infinity.

It is evident from the proof above that the size of the continuum is in some sense larger than countable, this raises the question of whether there are any infinities 'in between' these two? Nobody has been able to find one, but nobody has been able to show that one does not exist. In fact in 1963 Paul Cohen (1934 -) showed that this question is *undecidable*, and so when dealing with transfinite numbers it is necessary to take as an axiom that there is no infinity between the countable and the continuum, this is known as the *continuum hypothesis.* Of course we can equally well make the opposite assumption that there is such an infinity, each gives rise to a different mathematical system.

Chapter 8

Formal Logic

(See Discrete Mathematics, Susanna Epp)

Aristotle is credited with codifying and recording the rules of logic, though most of the logic recorded by Aristotle was undoubtedly known to Plato, it is scattered through Plato's works. Aristotle synthesised and compiled this scattered knowledge into a complete whole. Thus we speak of Aristotelian logic. His books where rediscovered by European renaissance mathematicians and, along with Euclid, formed a large part of the foundation of subsequent scientific and mathematical investigation.

The rules of logic he described have undergone a number of revisions over the intervening years, particularly by George Boole (1847) who gave the subject a much more algebraic flavour; hence today we talk of Boolean logic. Since then, the definitions have undergone further refinement, some important names in this context would be DeMorgan (1806 - 1871), Georg Cantor (1845 - 1918) and Bertrand Russell (1872 - 1970).

Historically logic is seen as having three basic stages of development.

1. The Greeks
2. Scholastic logic
3. Mathematical logic

The first of these is the logic of Aristotle, in which logical formulas consist of words, subject to the limitations of syntactical and grammatical rules. The second stage was abstracted from normal language, but still had to obey a limited set of rules related to their use in language. The third stage represents a further level of abstraction in which an artificial (mathematical) language was developed to obey certain well determined rules; "the rules of logic". This was the great achievement of George Boole (1847).

The main distinction between the first two stages and the third is that the first two are derived from normal language, whereas in the third there is a purely formal manipulation of logical constructs which then may be translated into normal language. Thus in the first two language underpins the logic, whereas in the third the logic is paramount. We may be able to translate logical statements into English, but the grammar may be awkward. On the other hand we can represent English sentences as logical statements, but we must beware of the fact that often statements in English are ambiguous in a logical sense.

In this account we give the basic 'rules of logic'. It should be noted however, that the greatest difficulty usually encountered by beginners in constructing logical arguments is the translation from formal logical syntax to language, particularly in the case of quantified statements.

8.1 Statements

Definition 8.1.1 *A* statement *is a sentence which is either true or false, but not both.*

Note that a statement must be either true or false, it cannot be neither, and it cannot be both. For example the following are statements.

1. This room is green.
2. My dog has flees.
3. The Goldbach conjecture is true.

The first example is a statement of fact. Provided we have a common understanding of what is green and what is not, then either the room is green or it is not. Note though that we don't care which, only that it must be one or the other.

The second example is also a statement of fact, either my dog has fleas or it does not. However this is slightly complicated by the fact that I don't have a dog. It doesn't seem reasonable that the mere non existence of the subject (my dog) should suddenly turn this sentence from a statement into something which is not a statement. Now, there is actually an ambiguity in the language, we could be implicitly implying that I have a dog, so the statement would more accurately be “I have a dog which has fleas”, which is false since I don’t have a dog. On the other hand we could be saying “All the dogs which I own have fleas”, which is vacuously true, since I don’t own any dogs. In either case it is a statement since it is definitely true or false. The usual interpretation in logic would be the latter.

We might try to argue that “if I had a dog then it would have flees”, or “if I had a dog then it might or might not have flees”, but these are actually compound statements consisting of an if...then... conjunction of statements, which we will deal with later.

The third example talks about a *conjecture*, that is something which has been asserted but is not known to be either true or false. It is not strictly speaking necessary to know what the Goldbach conjecture is, only whether it must be either true or false. In fact this conjecture was stated by Christian Goldbach in 1742 and proposes that every even integer greater than two is the sum of exactly two primes. The important point is that we do not know the answer, no one has been able to find an even integer which is not the sum of two primes. On the other hand no one has been able to provide a proof that this *must* be so. However, as far as logic is concerned this is irrelevant. The Goldbach conjecture must be either true or false, it cannot be both or neither. Just because we don’t know which doesn’t change the sentence being a statement.

It is useful to look at some sentences which are not statements.

1. This sentence is a lie.
2. This is a false statement.
3. All bus drivers write memos.
4. $y = x + 3$

The first example represents a paradox. If it is true, then by its statement it must be false. On the other hand if it is false it must be true. So we conclude that paradoxical statements are not allowed.

We carry this further in the second example. In this case the sentence asserts that it is a statement, if it is a statement it is a paradox and so not allowed, if not then it is not a statement, so it is false and hence a statement, oh dear! The problem is easily resolved by realising that we are attempting to assign a definite truth value to something which we have already said is not a statement and hence has ambiguous truth value.

The essential property which both of these statements have is that of being *self referential,* they talk about themselves. We can see that these types of statements are on the borders of logic. Self referential statements are instrumental in showing Gödel's incompleteness theorem[1].

The third example is a bit more subtle. The problem here is how we define "All bus drivers". Do we mean those people who are at this moment driving buses? Or perhaps all people who have at any time driven a bus? Or perhaps all members of the bus drivers union? We can see from this that logic is related to the theory of sets. If we had an accepted definition for the set of all bus drivers then we could state categorically whether the sentence was true or false, and hence it would be a statement.

The fourth example suffers from a similar problem, we do not know what x and y are, they could be planets for all we know. Even if we know that they represent integers or numbers we still don't have enough information, since we need to know how they relate. Thus

> For every integer x there is an integer y such that $y = x + 3$.

is a statement, which happens to be true. On the other hand

> For all integers x and y, $y = x + 3$.

is also a statement, which happens to be false. We will see later on how to deal with statements of this type.

In general we use the letters $p, q, r, \ldots$ to represent statements. Note that any given statement has associated with it a truth value T for true or F for false. Though we may not *know* which is associated with a given statement it must be one or the other. With the advent of computers the convention has arisen that 0 means false, and 1 (or any non zero value) means true.

8.2 Logical Operations

8.2.1 NOT, AND, OR

There are three basic logical operations which we can apply to statements, they are *not, and, or.* In order to give these operations precise definitions it is useful to introduce the idea of

[1]See *Gödel, Escher, Bach* by D. Hofstadter.

a truth table. In a truth table we tabulate all possible truth values of the input statements. We can show what the output of a compound statement would be for each set of truth values for the inputs. Thus for example, if the only input is one statement, p, and the output is not p we get the table in figure 8.1.

Figure 8.1: Truth Table for Not p

p	not p
T	F
F	T

If there are two or more inputs, we tabulate every possible combination of values (see figure 8.2).

Figure 8.2: Truth Table with Two Inputs

p	q	The output goes here.
T	T	X
T	F	X
F	T	X
F	F	X

We may use truth tables to provide definitions of logical connective statements. We may also use them to prove or disprove logical identities.

Definition 8.2.1

- **NOT**
  The negation of a statement p, denoted not p *or* $\neg p$ *has the opposite truth value as* p. *The associated truth table is:*

p	$\neg p$
T	F
F	T

- **AND**
  Given two statements p *and* q *we define* p and q, *also written* $p \wedge q$, *to be true only when both* p *and* q *are, and false otherwise. Thus the truth table is:*

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

- **OR**
  Given two statements p *and* q *we define* p or q, *also written* $p \vee q$, *to be true when either one of* p *or* q *is true, or both are true, and false otherwise. Thus the truth table is:*

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

Any *not* statement is carried out before any other, unless that part is bracketed. We may come up with complex juxtapositions of statements using *and, or* and *not*. Such statements are called *compound statements*. Thus we may have compound statements such as $p \vee (q \wedge r)$ and $(\neg p \wedge (q \vee r)) \wedge r)$.

Two compound statements are *logically equivalent* if they have the same truth value for any input. Thus if the columns of the truth tables for the two statements are the same the statements are equivalent. We denote equivalence with the symbol $\equiv$

We can use truth tables to prove the validity of statements. For example we can show that the negation of a negation of a statement is the original statement, that is $\neg(\neg p) \equiv p$.

p	$\neg p$	$\neg(\neg p)$
T	F	T
F	T	F

$\square$

We can use this method to prove DeMorgan's laws about the negation of and's and or's.

Theorem 8.2.2 (De Morgan's Laws)

For any statements p and q

1. $\neg(p \wedge q) \equiv \neg p \vee \neg q$
2. $\neg(p \vee q) \equiv \neg p \wedge \neg q$

Proof:
We give the proof of the first case, the second is left as an exercise. We prove this by showing that the corresponding columns in the truth table are the same. Note that we include some intermediary columns, the columns to compare are denoted by the double lines.

p	q	$p \wedge q$	$\neg(p \wedge q)$	$\neg p$	$\neg q$	$\neg p \vee \neg q$
T	T	T	F	F	F	F
T	F	F	T	F	T	T
F	T	F	T	T	F	T
F	F	F	T	T	T	T

$\square$

Note that all this has been done without any reference to language, or indeed the meaning of the statements p and q. There is in fact a linguistic point to be made here. English does not allow the use of brackets, and so to translate $\neg(p \wedge q)$ as 'not p and q' may lead to confusion, as we could mean '$\neg p \wedge q$'. In fact in English we implicitly use DeMorgan's law by saying 'neither p nor q'.

A compound statement is a *tautology* if all values in the corresponding column of the truth table are T. A compound statement is a *contradiction* if all values in the corresponding column of the truth table are F.

Thus the compound statement $p \vee \neg p$ is a tautology, always true.

p	$\neg p$	$p \vee \neg p$
T	F	T
F	T	T

The compound statement $p \wedge \neg p$ is a contradiction, always false.

p	$\neg p$	$p \wedge \neg p$
T	F	F
F	T	F

Note that a contradiction is a statement which is always false, not a paradox, which is not a statement at all.

8.2.2 Implication

There is another fundamental type of connectives between statements, that of *implication* or more properly *conditional statements.* In English these are statements of the form 'If p then q' or 'p implies q'.

Definition 8.2.3 *The compound statement $p \Rightarrow q$ ('If p then q') is defined by the following truth table:*

p	q	$p \Rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

In an implicative statement, $p \Rightarrow q$, we call p the *premise* and q the *conclusion.*

The first two rows make perfect sense from our linguistic understanding of 'If p then q', but the second two rows are more problematical. What are we to do if p is false? First note that we must do something, otherwise $p \Rightarrow q$ would not be a well defined *statement,* since it would not be defined as either true or false on all the possible inputs. We make the convention that $p \Rightarrow q$ is always true if p is false, this is often referred to as being *vacuously true.*

Consider the statement 'If I have a dog then it will have flees', by the rules above this is true since I don't have a dog. However the statement 'If I have a dog then it will not have flees' is also logically true, for the same reason. At first sight this seems odd, though we note that, since I don't have a dog the distinction is largely academic.

The major reason for defining things this way is the following observation. made by Bertrand Russell (1872 - 1970).

> From a false premise it is possible to prove *any* conclusion.

For example, I will now show that 'If $0 = 1$ then I am the Pope'.
We assume $0 = 1$ and show that 'I am the Pope' follows.
$0 = 1$, by adding 1 to both sides we conclude that $1 = 2$.
The Pope and I are two.
But $2 = 1$, hence the Pope and I are one and the same! $\square$

The word *any* is very important here. It means literally anything, including, of course, things which are true. It is a common mistake in proofs to assume something along the way which is not true, then proving the result is always possible. This is referred to as 'arguing from false premises'.

A caveat is in order here about what we mean by true or false. Generally we mean true or false *within the system.* Thus we could define a number system in which $0 = 1$, by adding this to the axioms of the integers. We could then develop whole theories about this system, but it is evident that this new system does not model 'numbers' as we generally understand the term. In fact it turns out that if you include this axiom in the axiomatic definition of the integers, then all numbers are equal, not a terribly interesting system, *but perfectly valid as a mathematical construct.*

When we wish to prove an implicative statement of the form $p \Rightarrow q$ we assume that p is true and show that q follows under this assumption. Since, with our definition, if p is false $p \Rightarrow q$ is true irrespective of the truth value of q, we only have to consider the case when p is true.

It should be noted that the truth value of an implicative statement is related to the relevance of the implication, rather than the input statements. Part of the problem is that in language we do not generally use implicative statements in which the premise is false, or in which the premise and and conclusion are unrelated. We usually assume that an implicative statement implies a connection, this is not so in logic. In logic we can make no such presumption, who would enforce 'relatedness'? How would we define it? Thus we allow statements like 'If this room is green then I have a dog', which is false if the room is green, but true if it is not, since I don't have a dog.

Given an implicative statement, $p \Rightarrow q$, we can define the following statements:

- The *converse* is $q \Rightarrow p$.
- The *contrapositive* is $\neg q \Rightarrow \neg p$

$p \Rightarrow q$ is logically equivalent to its contrapositive. That is $p \Rightarrow q$ is true if and only if $\neg q \Rightarrow \neg p$ is (see figure 8.3). Note that an implicative statement is *not* generally equivalent to its converse.

Figure 8.3: Proof that an Implicative Statement is Equivalent to its Contrapositive

p	q	$p \Rightarrow q$	$\neg q$	$\neg p$	$\neg q \Rightarrow \neg p$
T	T	T	F	F	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

$\square$

A common method of proof is to in fact prove the contrapositive of an implicative statement. Thus, for example, if we wish to prove that

For all p prime, if p divides n^2 then p divides n.

it is easier to prove the contrapositive:

For all p prime, if p does not divide n then p does not divide n^2.

There is one other type of statement to do with implication, the biconditional. This is usually represented in language by 'if and only if'. It means that p implies q *and* q implies p, we write $p \Leftrightarrow q$. Thus $p \Leftrightarrow q \equiv (p \Rightarrow q) \wedge (q \Rightarrow p)$.

p	q	$p \Leftrightarrow q$
T	T	T
T	F	F
F	T	F
F	F	T

The biconditional is actually the same as equivalence ($\equiv$) and is often used in place of it. Thus we would say 'p if and only if q', instead of p is equivalent to q.

When we wish to prove biconditional statements we *must* prove each direction separately. Thus we first prove $p \Rightarrow q$ (if p is true then so is q) and then independently we prove $q \Rightarrow p$ (if q is true then so is p).

8.2.3 Necessary and Sufficient

Definition 8.2.4 *Given two statements p and q*

- *'p is a necessary condition for q' means $\sim p \Rightarrow \sim q$ or equivalently $p \Rightarrow q$.*
- *'p is a sufficient condition for q' means $\sim p \Rightarrow \sim q$ or equivalently $p \Rightarrow q$.*

Notes

1. If p is a necessary condition for q this means that if q is true then so is p. However if p is true q may not be – there may be other things that must be true in order for q to happen.
2. If p is a sufficient condition for q then if p is true then q **must** happen as a consequence (all the conditions for q are fulfilled). However q may happen in some other way – so q True but p False is a possibility

8.3 Valid Forms

It would be possible, though tedious, to prove all of our logical statements by the use of truth tables. In order to avoid this we use *valid forms.* Valid Forms are sequences of statements which have already been shown to be true, leading to a new conclusion. To start with consider the statement

If statement 1 is true and statement 2 is true then statement 3 is true.

Suppose we have shown this to be a tautology, i.e. always true. Now in actual proofs we have axioms and previously proved statements which we can take as true. So the next step is to change the 'if...then...' form into a definite assertion. Thus

Statement 1 is true.
Statement 2 is true.
Therefore statement 3 is true.

Since mathematicians are lazy, and want to write as little as possible the final step is to remove the 'is true' and to assume it. We also use the symbol $\therefore$ for therefore. Thus a *valid form* appears like:

Figure 8.4: Valid Forms

Modus ponens $\begin{array}{ll} & p \Rightarrow q \\ & p \\ \hline \therefore & q \end{array}$	Modus tollens $\begin{array}{ll} & p \Rightarrow q \\ & \neg q \\ \hline \therefore & \neg p \end{array}$
Disjunctive addition a) $\begin{array}{ll} & p \\ \hline \therefore & p \vee q \end{array}$ b) $\begin{array}{ll} & q \\ \hline \therefore & p \vee q \end{array}$	Disjunctive syllogism a) $\begin{array}{ll} & p \vee q \\ & \neg q \\ \hline \therefore & p \end{array}$ b) $\begin{array}{ll} & p \vee q \\ & \neg p \\ \hline \therefore & q \end{array}$
Conjunctive addition $\begin{array}{ll} & p \\ & q \\ \hline \therefore & p \wedge q \end{array}$	Conjunctive simplification a) $\begin{array}{ll} & p \wedge q \\ \hline \therefore & p \end{array}$ b) $\begin{array}{ll} & p \wedge q \\ \hline \therefore & q \end{array}$
Hypothetical syllogism $\begin{array}{ll} & p \Rightarrow q \\ & q \Rightarrow r \\ \hline \therefore & p \Rightarrow r \end{array}$	Dilemma: Division into cases $\begin{array}{ll} & p \vee q \\ & p \Rightarrow r \\ & q \Rightarrow r \\ \hline \therefore & p \Rightarrow r \end{array}$

In addition we have The Rule of Contradiction, which states that if $\neg p$ implies a contradiction (a false statement) then p is true. Thus, if c denotes a contradiction:

Rule of contradiction

$$\begin{array}{ll} & \neg p \Rightarrow c \\ \hline \therefore & p \end{array}$$

$$\begin{array}{l} \text{Statement 1} \\ \text{Statement 2} \\ \hline \therefore \text{ Statement 3} \end{array}$$

All the valid forms are stated like this. Thus, for example *modus ponens* has the form:

$$\begin{array}{l} p \Rightarrow q \\ p \\ \hline \therefore q \end{array}$$

It is saying that we know that p is true and that p implies q, thus q must be true.

We list the most common valid forms in figure 8.4, along with their names, in this context 'disjunction' is *or* and 'conjunction' is *and.*

8.4 Set Theory

We have seen that logic is related to the theory of sets. We include a short account of some simple set theory.

8.4.1 Basic Definitions

Definition 8.4.1 *(Cantor) A* set *is a collection into a whole of definite and separate objects of our intuition or thought.*

(Russell) Such that given an object x and a set S it is possible to state unequivocally that either x belongs to the set S, or x does not belong to S.

Generally we use upper case letters to denote sets A, B, C, etc. Given a set A and an object x, if x is in A we write $x \in A$ and say x is in A or x is a member of A. If x is not in A we write $x \notin A$ and say x is not in A or x is not a member of A. Thus the symbol $\in$ may be read as 'in'.

Notice the similarity between this definition and the definition of *statement* above. Both use the idea of definite statements, definitely true or false, definitely in or out.

We have already seen how 'the set of bus drivers' can be ambiguous and thus, by this definition of set, it would not be a set since we can find objects (people) which may or may not be in the set. Something like 'the set of all real numbers between 0 and 1' is however a well defined set (provided we know whether 0 and 1 themselves are to be included) since we can state unequivocally, for any real number, whether it is between 0 and 1 or not.

8.4.2 Notation

1. If x is a member of a set S, we write $x \in S$.
   If x is not a member of a set S, we write $x \notin S$.

2. The definition of sets appears between curly braces $\{\ldots\}$, which is read 'the set' or 'the set of'.
   Example 8.4.2 $S = \{a, b, c\}$ *means that* S *consists exactly of the elements* a, b *and* c.
   Note that order is unimportant, thus $\{a, b, c\}$ and $\{b, a, c\}$ represent the same set. Also since a set is defined only by its members we ignore repitition. Thus the set $\{a, b, c\}$ is the same as $\{a, a, b, b, b, c, c\}$.

3. We use ellipses $(\ldots)$ to indicate that a pattern continues.
   Example 8.4.3 $\{0, 1, 2, 3, \ldots\} = \mathbb{N}$ - *The Natural numbers.*

4. We also use the symbol $|$ within the braces for 'such that'. Thus $T = \{x \in S \mid P(x)\}$ indicates the set of all those elements of S which satisfy the property P.

   T 'equals' 'the set of' x 'in' S 'such that' $P(x)$.

   Note that $|$ only has this meaning within braces, it usually is taken to mean something else outside.

   Example 8.4.4

   (a) If A is 'the set of all real numbers between 0 and 1' we would write

   $$A = \{x \in \mathbb{R} \,|\, x > 0 \text{ and } x < 1\}$$

   ($\mathbb{R}$ stands for the set of real numbers.) Which would be read 'A is the set of x in $\mathbf{R}$ such that x is greater than 0 and x is less than 1'. Note that 0 and 1 are not actually members of this set. In fact we often abbreviate $x > 0$ and $x < 1$ to $0 < x < 1$.

   (b) 'The set of all rational numbers between 0 and 1' would be

   $$B = \{x \in \mathbb{Q} \,|\, x > 0 \text{ and } x < 1\}$$

   ($\mathbb{Q}$ stands for the set of rational numbers.)

   (c) $T = \{x \in \mathbb{Z} \,|\, x \text{ is divisible by } 2\}$
   The set of even integers.

8.4.3 The Universal Set

Often a set will be defined in terms of a larger set, we generally wish to restrict the universe of possible objects to something less than everything. In the first example above we restrict ourselves to the set of people, in the second example we restrict the universe of possible objects to be the real numbers. This larger set is called the *universal set.* In general it is denoted by U, unless it is explicitly given to be something else. Thus if we are talking about 'the set of all real numbers between 0 and 1', the universal set is the real numbers. Whereas if we are talking about 'the set of all rational numbers between 0 and 1,' the universal set is the rational numbers.

8.4.4 Some Useful Sets

The Empty Set (5.3)

Definition 8.4.5 *The empty set is the set with no elements, denoted by ϕ.*

Number Sets

- $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ - The natural numbers.
- $\mathbb{N}^+ = \{x \in \mathbb{N} \mid x > 0\} = \{1, 2, 3, \ldots\}$ - The positive natural numbers.
- $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ - The integers.
- $\mathbb{Z}^+ = \{x \in \mathbb{Z} \mid x > 0\} = \{1, 2, 3, \ldots\}$ - The positive integers. $(= \mathbb{N}^+)$
- $\mathbb{Z}^- = \{x \in \mathbb{Z} \mid x < 0\} = \{\ldots, -3, -2, -1\}$ - The negative integers.
- $\mathbb{P} = \{p \in \mathbb{Z} \mid p \text{ is prime } \}$ - The prime numbers.
- $\mathbb{Q} = \{\frac{x}{y} \mid x \in \mathbb{Z} \wedge y \in \mathbb{N}^+\}$ - The rationals.
- $\mathbb{Q}^+ = \{\frac{x}{y} \mid x \in \mathbb{Z}^+ \wedge y \in \mathbb{N}^+\}$ - The positive rationals.
- $\mathbb{Q}^- = \{\frac{x}{y} \mid x \in \mathbb{Z}^- \wedge y \in \mathbb{N}^+\}$ - The negative rationals.
- $\mathbb{R} = (-\infty, \infty)$ - The Real numbers.
- $\mathbb{R}^+ = (0, \infty)$ - The positive Real numbers.
- $\mathbb{R}^- = (-\infty, 0)$ - The negative Real numbers.

- $\mathbb{I} = \mathbb{R} - \mathbb{Q}$ (all real numbers which are not rational) - The irrational numbers.
- $\mathbb{C} = \{x + yi \mid x, y \in \mathbb{R}\}$ - The Complex numbers.

Note: There are many real numbers which are not rational, e.g. π, $\sqrt{2}$.

8.4.5 Set Comparisons

Definition 8.4.6 *Given two sets A and B:*

1. *A is called a <u>subset</u> of B, written $A \subseteq B$, if and only if every element of A is also in B.*
   We also say B contains A, or A is contained in B.

   $$\textit{Formally: } A \subseteq B \Leftrightarrow \textit{ for every } x \in U, x \in A \Rightarrow x \in B$$

2. *$A = B$ if and only if A and B have exactly the same elements.*
3. *A is a <u>proper</u>, or <u>strict</u>, subset of B, written $A \subset B$ if and only if A is a subset of B, but there is at least one element of B which is not in A.*

Notes

- We say that two sets are *equal,* written $A = B$, if $A \subseteq B$ and $B \subseteq A$. Note that to prove equality in sets we must prove $A \subseteq B$ and then show $B \subseteq A$ independently. i.e.

  $A = B \Leftrightarrow (A \subseteq B) \wedge (B \subseteq A)$.
- By analogy with $\geq$ and $\leq$ we may define $A \supseteq B$ to be equivalent to $B \subseteq A$. Again by analogy with $>$ and $<$ we define $A \subset B$ to mean that $A \subseteq B$ and $A \neq B$. Similarly $A \supset B$ means that $A \supseteq B$ and $A \neq B$ i.e.

  $A \subset B \Leftrightarrow (A \subseteq B) \wedge (B \neq A)$ etc.

Example 8.4.7 *Let* $A = \{0, 1, 2, 3\}$, $B = \{1, 2\}$, $C = \{x \in \mathbb{N} \mid x < 4\}$.

Then the following statements are true:

$C \subseteq A$, $A \subseteq C$, $C = A$, $B \subseteq A$, $B \subset A$.

We can also say: $A \not\subseteq B$, $C \not\subset A$.

Note Sets may have other sets as elements, this is not the same as being a subset.

Example 8.4.8 $S = \{0, 1, \{2, 3\}\}$

$\{2,3\} \in S$, $\{2,3\} \not\subseteq S$, $\{0,1\} \subseteq S$, $\{0,1\} \notin S$, $0 \in S$.

How many elements are in the set ϕ?
How many elements are in the set $\{\phi\}$?

(Answer: 0 and 1 respectively)

Theorem 8.4.9

1. *ϕ is a subset of every other set, including itself.*

2. *Every set, including U, is a subset of the universal set U.*

Proof:

1. Must show that for any set S, $x \in \phi \Rightarrow x \in S$.
   But since there is no x such that $x \in \phi$ this is vacuously true.

2. For any set S, if $x \in S \Rightarrow x \in U$, since everything is in U by definition.

Theorem 8.4.10 *The empty set is unique. (i.e. there is only one empty set)*

Proof: (By contradiction.)
Suppose not, that is suppose that there are two distinct empty sets, ϕ_1 and ϕ_2.
Now, $\phi_1 \subseteq \phi_2$ (every set is a subset of the empty set).
And $\phi_2 \subseteq \phi_1$ (every set is a subset of the empty set).
Thus $\phi_1 = \phi_2$ (Definition of = for sets). $\square$

8.4.6 Operations on Sets

There are 3 basic operations on sets: complement, intersection and union.

Definition 8.4.11 *Let A and B be any subsets of a universal set U.*

1. *The* <u>*union*</u> *of A and B, denoted $A \cup B$ is the set of all elements of U which are in either A or B.*
$$i.e.\ A \cup B = \{x \in U \mid x \in A \vee x \in B\}.$$

2. *The* <u>*intersection*</u> *of A and B, denoted $A \cap B$ is the set of all elements of U which are in both A and B.*
$$i.e.\ A \cap B = \{x \in U \mid x \in A \wedge x \in B\}.$$

3. *The* <u>*difference*</u> *of B minus A, denoted $B - A$ is the set of all elements of U which are in B but not in A.*
$$i.e.\ B - A = \{x \in U \mid x \in B \wedge x \notin A\}.$$

4. *The* <u>*complement*</u> *of A, denoted A^c is the set of all elements of U which are not in A.*
$$i.e.\ A^c = \{x \in U \mid x \notin A\}.$$

8.4.7 Set Identities

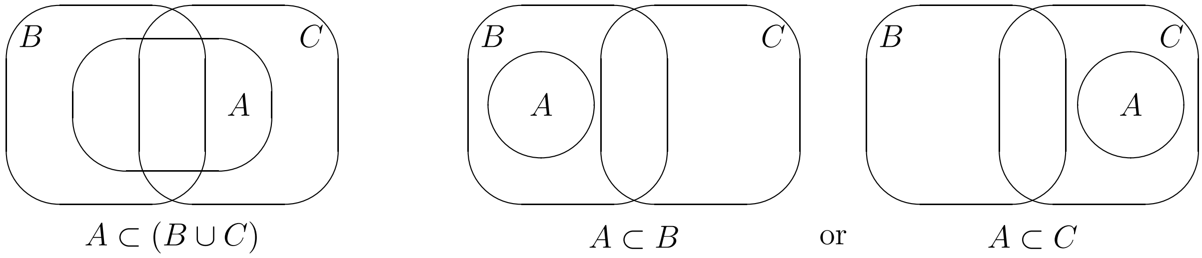
The way that the definitions of compliment, intersection and union use the notion of *not, and* and *or* and containment uses implication suggests a correspondence between logic and sets, with negation replaced by compliment, *and* replaced by intersection, *or* replaced by union and implication by containment. Thus, for example DeMorgan's laws $\neg(p \wedge q) \equiv \neg p \vee \neg q$ and $\neg(p \vee q) \equiv \neg p \wedge \neg q$ become $(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$ respectively.

However this correspondence between set theory and logic is not exact. The problem comes when we try to find the equivalent of *true* and *false.* The obvious equivalence would be that true corresponds to *is a member of,* and false corresponds to *is not a member of.* The problem is that whereas true and false are definite for each statement, member or not is multifaceted.

For example, in logic we may have a statement of the form $p \Rightarrow (q \vee r)$ (p implies either q or r) which is the same as saying that either p implies q or p implies r (($p \Rightarrow q$) or ($p \Rightarrow r$)). (As an exercise prove that these are equivalent using truth tables.) The original statement would translate as $A \subset (B \cup C)$, that is A is contained in the union of B and C. Thus (pushing the analogy), we would have to conclude that either $(A \subset B)$ or $(A \subset C)$, which is not necessarily the case. A may have members from both B and C, which together give containment, see figure 8.5.

While this correspondence is not exact it is striking how much of logic and set theory seem to be aspects of the same process.

Figure 8.5: $A \subset (B \cup C)$ verses $(A \subset B)$ or $(A \subset C)$



Here are some set identities, along with their analogues in logic. A, B, C are sets, ϕ is the empty set and U is the universal set.

p, q, r are statements t is the tautology and c the contradiction.

Law	Set Identity (5.2.2)	Logical Equivalence (1.1.1)
Commutativity	$A \cap B = B \cap A$ $A \cup B = B \cup A$	$p \wedge q \equiv q \wedge p$ $p \vee q \equiv q \vee p$
Associativity	$A \cap (B \cap C) = (A \cap B) \cap C$ $A \cup (B \cup C) = (A \cup B) \cup C$	$p \wedge (q \wedge r) \equiv (p \wedge q) \wedge r$ $p \vee (q \vee r) \equiv (p \vee q) \vee r$
Distributivity	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$ $p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$
Existence of Identity	$A \cap U = A$ $A \cup \phi = A$	$p \wedge t = p$ $p \vee c = p$
Complement/ Negation	$A \cup A^c = U$ $A \cap A^c = \phi$	$p \vee \sim p = t$ $p \wedge \sim p = c$
Double Complement	$(A^c)^c = A$	$\sim(\sim p) = p$
Idempotent	$A \cap A = A$ $A \cup A = A$	$p \wedge p = p$ $p \vee p = p$
DeMorgans	$(A \cap B)^c = A^c \cup B^c$ $(A \cup B)^c = A^c \cap B^c$	$\sim(p \wedge q) = \sim p \vee \sim q$ $\sim(p \vee q) = \sim p \wedge \sim q$
Universal bound	$A \cup U = U$ $A \cap \phi = \phi$	$p \vee t = t$ $p \wedge c = c$
Absorbtion	$A \cup (A \cap B) = A$ $A \cap (A \cup B) = A$	$p \vee (p \wedge q) = p$ $p \wedge (p \vee q) = p$
Identity Relations	$U^c = \phi$ $\phi^c = U$	$\sim t = c$ $\sim c = t$

Given a set with two binary operations and one unary operation defined on it, if it satisfies the first five of these properties (Commutativity, Associativity, Distributivity, Existence of Identities), then it will satisfy the rest. Such a system is called a *Boolean Algebra*.

In this case we have set of statements with $\wedge, \vee$ and $\sim$.

The set of all sets with $\cap, \cup$ and ^{c}.

Thus to develop the theory of Boolean Algebra, we assume that we have a set with two binary operations and one unary operation defined on it and take Commutativity, Associativity, Distributivity, Existence of Identities, of these operations as axioms. We assume no other properties that the actual operations may have. A particular instance (statements, or set theory) which realises an axiomatic system is called a model. We will discuss this idea further in the next Chapter.

8.5 Quantifiers

The last element in formal logic is the idea of quantifiers. We have seen that there is a problem with statements like $x + 3 = 4$, we don't know what x is and we don't know if we mean that this is true for every x, or some x. There are essentially two types of quantifier 'for all' $\forall$ and 'there exists' $\exists$.

The first type of quantifier is the *universal quantifier* 'for all', $forall$, it is used to indicate that a proposition $P(x)$ is true for all elements in some set. It occurs as $\forall\, x \in S, P(x)$, $\forall$ is read 'for all', S is a set and $P(x)$ is some proposition involving x as a variable, such as $x + 3 = 4$. Thus this is read as 'For all x in S, $P(x)$'.

The second type of quantifier is the *existential quantifier* 'there exists', $\exists$, it is used to indicate that a proposition $P(x)$ is true for at least one element of some set. It occurs as $\exists\, x \in S$ such that $P(x)$, $\exists$ is read 'there exists', again S is a set and $P(x)$ is some proposition involving x as a variable, such as $x + 3 = 4$. Thus this is read as 'There exists x in S such that $P(x)$'.

Definition 8.5.1

1. *A* predicate *is a sentence which contains a finite number of variables and becomes a statment if particular values are substituted for the variables.*
2. *The* domain *of a predicate variable is the set of possible values which that variable can take.*

Generally the truth value of a quantified statement depends on the domain of the variables.

Thus quantified statements have one of the following general forms:

1. $\forall x \in S,\ P(x)$.
2. $\exists\, x \in S$ such that $P(x)$.

Here S is a set (the domain of x), and $P(x)$ is a property which x may or may not have. $P(x)$ evaluates to either true or false for each choice of x from S.

The first form is called a *universal statement*, it is true if all elements of S have the property P and false otherwise.

The second form is called an *existential statement*, it is true if there exists at least one $x \in S$ with the property P and false otherwise.

Universal statements usually involve words like: 'every' 'all', 'any' or 'each'.
Existential statements usually involve words like: 'some' 'at least one', 'we can find' or 'there exists'.
The 'such that' in universal statements is sometimes abreviated to 's.t.'
Note $P(x)$ may be thought of as a function from S to $\{$T, F$\}$.

Example 8.5.2

1. "All buttercups are yellow" (True)
   Let $B = \{x \mid x$ is a Buttercups $\}$. $P(x) = x$ is yellow.
   $\forall x \in B, x$ is yellow.
2. "All cars are yellow" (False)
   Let $C = \{x \mid x$ is a car $\}$. $P(x) = x$ is yellow.
   $\forall x \in C, x$ is yellow.

Note that the only difference between 1 and 2 is the domain of x, yet their truth values are different.

3. "There is a yellow car" (True)
   $\exists\, x \in C$ such that x is yellow.
4. "All integers are positive" (False)
   $\forall x \in \mathbb{Z}, x > 0$.
5. "All natural numbers, except 0, are positive" (True)
   $\forall x \in \mathbf{N}, x = 0 \vee x > 0$.

6. There is a natural number x, which satisfies $x + 3 = 2$ (False)
   $\exists\, x \in \mathbb{N}$ such that $x + 3 = 2$.
7. There is an integer number x, which satisfies $x + 3 = 2$ (True)
   $\exists\, x \in \mathbb{Z}$ such that $x + 3 = 2$.
8. Every integer satisfies $x + 3 = 2$. (False)
   $\forall x \in \mathbb{Z}, x + 3 = 2$.

8.5.1 Domain Change

Given a statement of the form $\forall x \in D, P(x)$ we may rewrite it as $\forall x, x \in D \Rightarrow P(x)$.
Or if $D \subseteq S$: $\quad \forall x \in S, x \in D \Rightarrow P(x)$.
This is called if - then form, or a domain change.

Example 8.5.3

1. Write $\forall x \in \mathbb{Z}^+, 2x \in \mathbb{Z}^+$ in if - then form.
   'For every positive integer, twice it is also a positive integer'.

   $\forall x \in \mathbb{Z}, x \in \mathbb{Z}^+ \Rightarrow 2x \in \mathbb{Z}^+$.
   'For every integer, if it is positive then so is twice it.'
2. 'All valid arguments with true premises have true conclusions'
   Let $A = \{p \mid p \text{ is an argument form }\}$ and $V = \{p \mid p \text{ is a valid argument form }\}$.

   $\forall x \in A, x \in V \Rightarrow$ the conclusion of x is true.

We may do a similar thing for existential statements using *and*.
$\exists\, x \in D$ such that $P(x)$ can be rewritten as $\exists\, x, x \in D \Rightarrow P(x)$.
Or if $D \subseteq S$: $\quad \exists\, x \in S, x \in D \Rightarrow P(x)$.

Example 8.5.4

1. $\exists\, x \in \mathbb{Z}^+$ such that x is even.
   'There is a positive even integer'

   $\exists\, x \in \mathbb{Z},$ s.t. $x \in \mathbb{Z}^+ \wedge x$ is even.
2. 'Some flowers are yellow'. Let $F = \{x \mid x \text{ is a flower }\}$.
   $\exists\, x \in F$, such that x is yellow.

   $\exists\, x, x \in F, \wedge x$ is yellow.

8.5.2 Negations of Quantified Statements

Consider the negation of 'All flowers are Yellow' ($\forall x \in F, Y(x)$). This will be untrue if there is at least one non yellow flower i.e. ($\exists\, x \in F$ s.t. $\sim Y(x)$).

Theorem 8.5.5 *The negation of a universal statement of the form $\forall x \in S, P(x)$ is $\exists\, x \in S, \sim P(x)$.*

$$\textit{i.e. } \sim(\forall x \in S, P(x)) \equiv \exists\, x \in S, \sim P(x).$$

Consider the statement 'Some flowers are yellow', ($\exists\, x \in F$ s.t. $Y(x)$) this will be false if all flowers are not yellow, i.e. ($\forall x \in F, \sim Y(x)$).

Theorem 8.5.6 *The negation of an existential statement of the form $\exists\, x \in S, P(x)$ is $\forall x \in S, \sim P(x)$.*

$$\textit{i.e. } \sim(\exists\, x \in S, P(x)) \equiv \forall x \in S, \sim P(x).$$

Example 8.5.7 *Write negations of the following statements.*

1. 'All integers are even' (False)
   Logical form: $\forall x \in \mathbb{Z}, x$ is even.

   Negation: $\exists\, x \in \mathbb{Z}$ s.t. x is not even.
   'There is atleast one non even integer'.

2. 'There is an integer which divides 7' (True)
   Logical form: $\exists x \in \mathbb{Z}$ s.t. x divides 7.

   Negation: $\forall x \in \mathbb{Z}, x$ does not divide 7.
   'There is no integer which divides 7'.

3. 'If x is an integer then $x > 0$.' (False)
   Logical form: $\forall x, x \in \mathbb{Z} \Rightarrow x > 0$.

   Negation: $\exists x$ s.t. $\sim(x \in \mathbb{Z} \Rightarrow x > 0)$. $\equiv \exists x$ s.t. $x \in \mathbb{Z} \wedge x \geq 0$
   'There is a number which is both an integer and non negative'.

8.5.3 Vacuosly True Universal Statements

Consider the statement 'All my dogs have fleas' mentioned at the beginning of the chapter. This translates to $\forall x \in D, F(x)$, where $D = \{$ dogs owned by me $\}$ and $F(x) = x$ has fleas. But as mentioned I have no dogs, so $D = \phi$, the empty set. Thus we have $\forall x \in \phi, F(x)$.

Is this statement true or false? It should be one or the other since it doesn't seem reasonable that the mere non existence of the subject (my dog) should suddenly turn this sentence from a statement into something which is not a statement.

In general universal statements about the empty set are always true. We say that it is vacuously true.

In particular the following statements are true:

1. 'All my dogs have fleas' $(\forall x \in D, F(x))$
2. 'None of my dogs have fleas' $(\forall x \in D, \sim F(x))$

On the other hand Existential statements about the empty set are generally false, since they state that something exists in the empty set. thus

'I have a dog with fleas'. $(\exists x \in D, \text{ s.t. } F(x))$

is clearly false since it claims that, among other things, I have a dog.

This is also consistent with the negation rule. For example, the negation of the (False) existential statement $\exists x \in D$, s.t. $F(x)$ above, is the (True) statement 2 $\forall x \in D, \sim F(x)$, 'None of my dogs have fleas' above.

8.5.4 Multiply Quantified Statements

The real power of quantifiers is found when we use more than one in a statement. We may have statements with more than one quantified variable. If there is more than one variable with the same quantifier they are put togther if possible.

Example 8.5.8

1. 'The sum of any two natural numbers is greater than either of them'
   $\forall x, y \in \mathbf{N}, x + y > x \wedge x + y > y$.

The $\forall$ quantifies both x and y.

2. 'There are two integers whose sum is twice the first'.
   $\exists\, x, y \in \mathbb{Z}$ s.t. $x + y = 2x$.

The $\exists$ quantifies both x and y.

We can have more than two variables:

3. 'There are three integers whose sum is 25'.
   $\exists\, x, y, z \in \mathbb{Z}$ s.t. $x + y + z = 25$.

We can use the same quantifier for different sets:

4. 'There is an integer and a rational number whose sum is 2.5'.
   $\exists\, x \in \mathbb{Z}, y \in \mathbf{Q},$ s.t. $x + y = 2.5$.

More interesting statements can be formed by mixing quantifiers. $\forall x \in A, \exists y \in B$ s.t. $P(x, y)$ means that for every x in A we can find a y in B such that the property $P(x, y)$ is true. Whereas the statement $\exists\, x \in A$ s.t. $\forall y \in B, P(x, y)$ means that there is a fixed element x in A such that $P(x, y)$ is true for every y in B.

Note Order *is* important.

Example 8.5.9

1. There is an integer whose sum with any other integer is equal to the second.
   $\exists x \in \mathbb{Z}$ such that $\forall y \in \mathbb{Z}, x + y = y$.
   This is true, the number (x) with this property is 0.

2. For every integer, there is another integer whose sum with the first is 0.
   $\forall x \in \mathbb{Z}, \exists\, y \in \mathbb{Z}$ such that $x + y = 0$.
   This is true: $y = -x$, but note that the value of y depends on x. i.e. each x gets a different y.

3. There is an integer whose sum with any other real number is equal to twice itself.
   $\exists x \in \mathbb{Z}$ such that $\forall y \in \mathbb{R}, x + y = 2x$.
   This is false there is no such number.

4. For every integer there is a real number such that their sum is twice the integer.
   $\forall x \in \mathbb{Z}, \exists y \in \mathbb{R}$ such that $x + y = 2x$.
   This is true, take $y = x$.

5. All fractions may be expressed as the ratio of two integers.
   $\forall x \in \mathbf{Q}, \exists\, p, q \in \mathbb{Z}$ such that $x = \frac{p}{q}$.
   This is true - see the definition of $\mathbb{Q}$.

6. There is no greatest integer.
   $\forall x \in \mathbb{Z}, \exists\, y \in \mathbb{Z}$ such that $y > x$.
   Literally this translates as 'for every integer there is an integer which is greater than it'.
7. Between any two rational numbers there is an irrational number.
   $\forall x, y \in \mathbb{Q}, \exists\, z \in \mathbb{I}$ such that $x < z < y$ or $y < z < x$.
   This is true.

The negation rule still applies, so:

$\sim(\forall x \in A, \exists\, y \in B \text{ s.t. } P(x, y)) \equiv \exists\, x \in A \text{ s.t. } \forall y \in B, \sim P(x, y)$.

$\sim(\exists\, x \in A \text{ s.t. } \forall y \in B, P(x, y)) \equiv \forall x \in A, \exists\, y \in B \text{ s.t. } \sim P(x, y)$.

8.5.5 Scope

The scope of a quantified variable is as follows:

1. If the quantifier appears at the beginning of the statement the scope is the whole statement.

2. If the quantifier appears after a separator ($\wedge$, $\vee$, $\Rightarrow$) then the scope is until the next separator, or the end of the statement if there is none.

3. We may use brackets to change the scope.

Example 8.5.10 *Put the following in symbolic form, given:*
$B = \{x \mid x \textit{ is a bee}\}$ $\quad$ $F = \{x \mid x \textit{ is a flower}\}$ $\quad$ $L(x, y) = x \textit{ loves } y$
$U = $ *the universal set.*

1. 'All bees love all flowers'

   $\forall x \in B, y \in F, L(x, y)$.

2. 'Some bees love all flowers.'

   $\exists\, x \in B$ s.t. $\forall y \in F, L(x, y)$.

3. 'All bees love some flower'
   $\forall x \in B, \exists\, y \in F$ s.t. $L(x, y)$.
4. 'Only bees love flowers.'
   $\forall x \in U, \exists y \in F$ s.t. $L(x, y) \Rightarrow x \in B$.
5. 'No bee loves only flowers.'
   $\forall x \in B, \exists\, x \in U$ s.t. $L(x, y) \wedge x \notin B$.
6. 'Some bees love only flowers'
   $\exists\, x \in B$ s.t. $\forall y \in U, L(x, y) \Rightarrow y \in B$.
7. 'Every bee hates all flowers'
   $\forall x \in B, y \in F, \sim L(x, y)$.
8. 'No bee hates all flowers'
   $\sim(\exists\, x \in B$ s.t. $\forall y \in F, \sim L(x, y)) \equiv \forall x \in B, \exists\, y \in F$ s.t. $L(x, y)$.

8.5.6 Variants of Universal Conditional Statements

Given a statement involving a quantifier and an implication we may consider the contrapositive, converse and inverse of the statement.

Definition 8.5.11 *Given a statement of the form $\forall x \in S, P(x) \Rightarrow Q(x)$:*

1. *The* contrapositive *is the statement*
$$\forall x \in S, \sim Q(x) \Rightarrow \sim P(x).$$
2. *The* converse *is the statement*
$$\forall x \in S, Q(x) \Rightarrow P(x).$$
3. *The* inverse *is the statement*
$$\forall x \in S, \sim P(x) \Rightarrow \sim Q(x).$$

As before a statement is logically equivalent to its contrapositive, but not to its converse or inverse.

Example 8.5.12 *Write the contrapositive, converse and inverse of the statement 'The square of every real number greater than 1 is greater than itself.'*

The symbolic form of the statment is: $\forall x \in \mathbb{R}, x > 1 \Rightarrow x^2 > x$.

Contrapositive: $\forall x \in \mathbb{R}, x^2 \leq x \Rightarrow x \leq 1$.
'If the square of a real number is less than itself then that number is less than 1'

Converse: $\forall x \in \mathbb{R}, x^2 > x \Rightarrow x > 1$.
'If the square of a real number is greater than itself then that number is greater than 1.'

Inverse: $\forall x \in \mathbb{R}, x \leq 1 \Rightarrow x^2 \leq x$.
'Any real number less than 1 has square less than or equal to itself.'

Note that in this case the converse and inverse are false. (Take x negative.)

8.5.7 Necessary and Sufficient

We can also define the concepts of necessary and sufficient.

Definition 8.5.13

1. *$\forall x, P(x)$ is a <u>sufficient condition</u> for $Q(x)$, means $\forall x, P(x) \Rightarrow Q(x)$*
2. *$\forall x, P(x)$ is a <u>necessary condition</u> for $Q(x)$, means $\forall x, \sim P(x) \Rightarrow \sim Q(x)$*
3. *$\forall x, P(x)$ <u>only if</u> for $Q(x)$, means $\forall x, \sim Q(x) \Rightarrow \sim P(x)$*

Example 8.5.14

1. 'In order to pass it is necessary that the student complete the final exam.'
   Let $S = \{x \mid x \text{ is a student }\}$, $C(x) = x$ completes the final exam. $P(x) = x$ passes.
   $\forall x \in S, \sim G(x) \Rightarrow \sim P(x)$.
   'If any student does not complete the final, then they wont pass.'
   (But might complete final and still fail)

2. 'For an integer to be less than 6 it is sufficient that it be less than 3'

   $\forall x \in \mathbb{Z}, x < 3 \Rightarrow x < 6.$

   (Not necessary, for example take $x = 5$)

3. 'A necessary and sufficient condition for a positive integer to be even is that there is some positive integer which is half the first'

   $\forall x\ \mathbb{Z}^+, x$ is even $\Leftrightarrow \exists k \in \mathbb{Z}^+$ s.t. $x = 2k$

8.6 Methods of Proof

Note that these methods are only general guidelines, every proof has its own form. The guts of the proof still needs to be filled in, these guidelines merely provide a possible staring point.

Some Useful rules:

1. **Always** clearly state any assumptions you are making.
2. **Always** go back to the definition of objects under consideration.
3. **Never** use a variable which has not first been introduced.

Introduce variables with statements like: "Let $x \in S$ ..."

This identifies x as an element of the set S. x is assumed to be a particular, but arbitrary element of S.
We can assume that x has any of the properties associated with S, *but no other properties.*

Example 8.6.1

Let $y \in \{x \in \mathbb{Z} \mid x \text{ is even }\}$.

We can now assume that y is even, but nothing else.
With this restriction we can now use y as a particular element of the set.
Thus we may make statements like "$2y$ is even".
This is understood to hold for every y in the domain of y (even numbers).

However, we may not assume that y is divisible by 4, since this is not a property shared by every element of the set.

8.6.1 Direct Methods

Given a statement of the form $\forall x \in D, P(x)$:

We can show that such a universal statement is *false* by finding a counterexample. i.e. an element of the domain which does not satisfy the property.

Note that the negation of the statement above has the form $\exists x \in D$ s.t. $Q(x)$, so to prove a statement of this form we may find an element of the domain with the required property. This is called proof by construction.

If the domain, D, is finite, we can show that this statement is *true* by going through every value in the domain. This is called the method of exhaustion. This has obvious problems if the domain is large, or worse yet not finite.

Example 8.6.2 *Let* $S = \{2, 4, 6\}$ *and* $D = \{2, 4, 5\}$.

1. $\forall x \in S, x$ is even.

   2 is even, 4 is even, 6 is even, so true by exhaustion.

2. $\forall x \in D, x$ is even.

   A counterexample is $x = 5$, so false.

3. $\exists\, x \in S$ s.t. x is even.

   Take $x = 4$, so true.

4. $\exists\, x \in S$ s.t. x is not even.

   2 is even, 4 is even, 6 is even, so false by exhaustion.

Most statements in mathematics are of the form $p \Rightarrow q$, that is quantified implicative statements, for example 'For all integers n, if n^2 is even then n is even'.

When proving an implicative statement we assume that the premise p is true and show that the conclusion q must follow. For an implicative statement we don't care about the truth value of q when p is false.

To prove **biconditional** statements of the form '$p \Leftrightarrow q$'.
First prove $p \Rightarrow q$.
Then prove $p \Leftarrow q$.
Never try to prove a biconditional in one go.

We can mix and match these methods as needed. For example, a proof a statement of the form below might go as follows.

To Prove $\forall\, x \in S, \exists\, y \in T$ such that $P(x, y) \Rightarrow Q(x, y)$.

Let $x \in S$ [x is an arbitrary but specific element of S]
Consider $a \in T$ [a is a specific element of S, which you give]
Assume that $P(x, a)$ is true (Assume the predicate)

$$\vdots \quad \begin{bmatrix} \text{This is the guts of the proof, we may assume that } x \text{ has} \\ \text{any of the properties of } S\text{, and that } P(x,a) \text{ is true} \end{bmatrix}$$

Therefore $Q(x, a)$.

8.6.2 Indirect Methods

Contrapositive

When trying to prove a statement of the form $p \Rightarrow q$ it is often easier to prove the contrapositive, i.e. $\neg q \Rightarrow \neg p$. Since this is logically equivalent to the original statement a proof of the contrapositive proves the original statement.

For example in order to prove

For all integers n, if n^2 is even then n is even.

It is actually easier to prove the contrapositive:

For all integers n, if n is odd then n^2 is odd.

(Note that odd is the same as 'not even')

Proof by Contradiction

Proof by contradiction is probably the most common, as well as the most contentious method of proof. It is very useful because it is not restricted to implicative statements. This method of proof is based on the Rule of Contradiction.

In this method of proof we assume that the statement, p say, is false. We then consider a new statement, q say, which is known to be true, either because it is an axiom or because it has been shown in a previous theorem. We now show that the negation of p ($\neg p$) implies that q is false i.e. $\neg p \Rightarrow \neg q$. Note that if $\neg p$ is false we can always do this since we are arguing from false premises, and hence can prove anything. Now we have the pair of statements:

$$\neg p \Rightarrow \neg q$$
$$q$$

The contrapositive of the first statement is $q \Rightarrow p$, so by modus ponens we conclude p.

The objections to proof by contradiction come from the fact that we generally must manipulate objects which we know do not exist (since p is false). For example in the proof that $\sqrt{2}$ is irrational we started off by assuming that it wasn't. That is that there exist integers p and q, such that $\sqrt{2} = \frac{p}{q}$. But we know (since we in fact know that $\sqrt{2}$ is irrational) that p and q do not exist. This then makes our next couple of steps somewhat suspect, we square both sides to get $2 = \left(\frac{p}{q}\right)^2$, and multiply by q to get $2q = p$. But how can we do these things if p and q don't actually exist. Certainly we can imagine them, but what real meaning does it have to manipulate an imaginary quantity.

The usual answer to this is to argue that this is the way p and q would behave if they did exist, after all the whole point is to show that they must have some fantastic property which precludes their existence. I personally feel very comfortable with proof by contradiction, as do most mathematicians. But there is always a niggling doubt about this method of proof.

8.6.3 Induction

Suppose we wish to prove a statement of the form $\forall n \in \mathbb{Z}, n \geq a \Rightarrow p_n$, where a is some fixed number. (Usually 0 or 1, so this becomes $\forall n \in \mathbb{N}, p_n$ or $\forall n \in \mathbb{N}^+, p_n$.) That is the property p is indexed in some way by n. Note that p_n may involve quantifiers and hence other variables. We can prove a statement like this by using the method of induction.

Consider the following two statements

1. p_a is true.

2. For every natural number k, with $k \geq a$, if p_k is true then p_{k+1} is true.

If these two statements are both true then $\forall n \in \mathbb{N}, n \geq a \Rightarrow p_n$ will be true, by the domino effect.

The first part is celled the **base case**.
The second part is called the **inductive step**

Thus every proof by induction looks like:

Proof: By Induction on n
Base Case

> Let $n = a$.
> Prove that p_a is true.

Inductive Step

> Let $n = k$, with $k \geq a$.
> Assume p_k (this is called the inductive hypothesis)
>
> $\vdots$
>
> Hence p_{k+1}. $\square$

In the inductive step it is helpful to keep in mind what p_{k+1} is.

Example 8.6.3

Given $b_0 = 1,\ b_n = \dfrac{b_{n-1}}{1 + b_{n-1}}$ for $n > 1$, find a formula for b_n and prove it.

$$b_0 = 1, \qquad b_1 = \tfrac{1}{1+1} = \frac{1}{2}, \qquad b_2 = \frac{\frac{1}{2}}{1+\frac{1}{2}} = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}, \qquad b_3 = \frac{\frac{1}{3}}{1+\frac{1}{3}} = \frac{\frac{1}{3}}{\frac{4}{3}} = \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4},$$

Theorem 8.6.4 *Given $b_0 = 1,\ b_n = \dfrac{b_{n-1}}{1 + b_{n-1}}$ for $n > 1$, show that $b_n = \dfrac{1}{n+1}$.*

Proof: By Induction on n

Base Case:

> Let $n = 0$, $b_0 = 1 = \dfrac{1}{0+1}$.
> So true for $n = 0$.

Inductive Step:

Let $n = k \geq 0$.
$b_k = \dfrac{1}{k+1}$
Now $b_{k+1} = \dfrac{b_k}{1+b_k} = \dfrac{\frac{1}{k+1}}{1+\frac{1}{k+1}} = \dfrac{\frac{1}{k+1}}{\frac{k+2}{k+1}} = \dfrac{1}{k+1} \cdot \dfrac{k+1}{k+2} = \dfrac{1}{k+2}$.
So true for $k+1$. $\square$

8.6.4 Assumptions

When doing proofs we may assume any axioms and any previously proved theorems. In mathematical texts it will often be assumed that the reader knows some basic theorems, and they will also be assumed.

An example of this might be that you may assume the basic rules of algebraic manipulation (**Algebra**), where appropriate. But be careful about where these rules are appropriate.

Another type of assumption is **closure** under basic operations as appropriate. Closure means that given an operation defined on a set S the result of an operation is guaranteed to remain in the set. Thus adding two natural numbers is guaranteed to yield a natural number, so $\mathbb{N}$ is closed under '$+$'. On the other hand subtracting one natural number from another may not give a natural number, so $\mathbb{N}$ is not closed under '$-$'.

- $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ and $\mathbb{R}$ are all closed under addition.
- $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ and $\mathbb{R}$ are all closed under multiplication.
- $\mathbb{Z}, \mathbb{Q}$ and $\mathbb{R}$ are closed under subtraction.
  $\mathbb{N}$ is **not** closed under subtraction.
- $\mathbb{Q}$ and $\mathbb{R}$ are closed under division.
  $\mathbb{N}$ and $\mathbb{Z}$ are **not** closed under division.

Another assumptions is that the rules (axioms) of logic hold. Most of the proof methods rely on some standard valid form from logic.

8.6.5 Common Mistakes

When doing proofs, keep in mind the goal toward which you are working. Try to develop an understanding of the problem. Study particular examples and come up with reasons why things are, or are not true.

Always go back to the definitions of the objects under consideration. This often provides a starting point, or a clarification of the statement. Note that definitions are always in logical, symbolic form.

Above all proofs should be **clear** and **precise**

Do not ramble on, hoping that no one will notice that you don't know what you are talking about. Such arguments are called *hand waving,* due to the large amount of gesticulation that usually accompanies them.

Here are some common mistakes made by beginners.

1. **Proof by Example**
   A universal statement cannot be proved by an example.

   Though you may wish to think about an example to gain insight into a problem an example never constitutes a proof of a universal statement.

2. **Incorrect Variable Use**
   Don't use the same variable name for two different things.

   > m and n are even, so $m = 2k$, and $n = 2k$ for some integer k.

   m and n are not necessarily equal, so this may cause confusion later on.

3. **Jumping to a Conclusion**
   Beware of making unwarranted assumptions (even if they happen to be true).

4. **Begging the Question**
   Don't assume the result, whether implicitly or explicitly.

5. **Misuse of if**
   Don't use *if* when you mean *since* or *because.*

6. **Arguing from false premises**
   Don't assume something which is not true, then proving the result is always possible. ($p \Rightarrow q$ is always true when p is false.)

Chapter 9

Formal Systems

(See *The Nature and Power of Mathematics* by Donald M. Davis)

9.1 Axiomatic Systems

We have seen how mathematics rests on an axiomatic framework. In this chapter we investigate this framework more closely.

We have seen how Euclid had trouble defining fundamental concepts, this is a general problem with fundamental entities. To avoid this problem in a *Formal System* we start with a set of undefined terms, rather than trying to give definitions to fundamental concepts we initially leave them undefined. We then have a number of axioms which use the undefined terms. The undefined terms become defined by the way that they are used in the axioms.

There are essentially two kinds of terms, objects and relational terms. The first define the objects under study, *point* and *line* are examples from geometry. The other type of terms are relational, explaining how the objects interrelate, an example from geometry would be *is incident with,* or *meets.*

Example 9.1.1

We start with the undefined terms *point,* *line* and *meet.*
Axioms:

1. Every pair of points uniquely determines a line.

In this case we have only one axiom, which is Euclid's first axiom, however we have no guarantee that we are in fact talking about geometry, nothing states that lines or points behave in the way we might expect. We might be talking about some wildly different system which happens to obey this axiom. An example of a system which satisfies this axiom is if we take our 'points' to be the numbers 1 to 7 and the 'lines' to be the sets

$$\begin{array}{llll} \{1,2,4\} & \{2,3,5\} & \{3,4,6\} & \{4,5,7\} \\ \{1,5,6\} & \{2,6,7\} & \{1,3,7\} & \end{array}$$

every pair of numbers between 1 and 7 determines one and only one of these sets.

It may be misleading to call objects by common names, since this may lead to confusion. We may inadvertently ascribe properties to objects which the axioms do not imply, simply because we are thinking of them as points or lines or whatever. A general rule is that we should be able to replace any of the undefined terms by any words, including nonsense words, and still be able to prove the same theorems.

Example 9.1.2

Undefined terms are *lorn, pom* and *mort.*

Axioms:

1. There is at least one lorn.
2. For any two distinct poms, there is one and only one lorn that morts both of them.
3. Each lorn morts at least two poms.
4. For any lorn, there is at least one pom which it does not mort.

From these axioms it is clear that lorns and poms are objects, while mort is a relation, lorns mort poms. We can now prove things about lorns and poms.

Theorem 9.1.3 *There are at least three poms, which are not morted by the same lorn.*

Proof:
By Axiom 1 there is at least one lorn.
By Axiom 3 there are at least two poms that it morts.
By Axiom 4 there is at least one pom that it doesn't mort.
Thus there are at least three poms not morted by the same lorn. □

Theorem 9.1.4 *If there are exactly three poms, then there are exactly three lorns.*

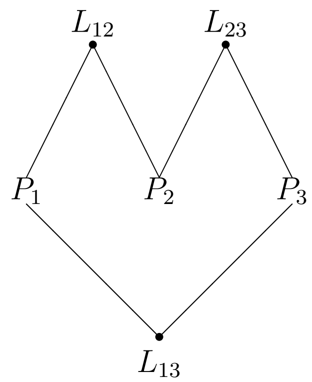
Proof:
We prove this by first showing that if there are three poms, then there must be at least three lorns. We then show that there can be no more than three lorns.

Let P_1, P_2 and P_3 be the three poms.
By Axiom 2 there is a unique lorn that morts both P_1 and P_2. Let us call this lorn L_{12}.
Axiom 4 tells us that L_{12} cannot mort P_3, since there must be at least one pom which L_{12} does not mort.
But by Axiom 2 there must be a lorn which morts both P_1 and P_3, call this lorn L_{13}.
By Axiom 4 L_{13} does not mort P_2.
But there must be a lorn that morts P_2 and P_3, call this lorn L_{23}.
Thus we have shown that there must be at least three lorns.

We now show that there can be at most three lorns.
Suppose that there was such a lorn, call it L.
Then by Axiom 3 it must mort at least two poms.
But every pair of poms already has a lorn morting them, thus the new lorn would be in violation of Axiom 2. $\square$

We can see that as we build up theorems and begin to work with 'lorns' and 'poms' and their 'morting', we begin to build up an idea of how they interrelate.

In the case where there are exactly three poms, the relation of morting between lorns and poms can be represented by the following picture. lorns and poms are represented by points, and the relation morting is represented by lines joining lorns to the poms which they mort.



Such a picture is a *model* for the axiomatic system, there may be many models for a given axiomatic system. A model is any system that obeys the axioms.

By the argument above we can see that any model with three poms, and hence three lorns, will essentially boil down to the model above, we say that the models are *isomorphic.* However in general there may be wildly differing models which obey the same set of axioms.

If we consider what happens when we have exactly four poms we can see that there are models which are *non-isomorphic.* There are valid models with three, four and five lorns. (Exercise: see if you can find them.) Clearly these models represent very different systems since they have different numbers of lorns.

In general two models are isomorphic if there is a permutation of the objects which preserves type (lorns to lorns and poms to poms in this case) which leads to the same relational structure (morting).

9.2 Another System

We now consider another example. In this example we use the undefined terms to define new terms.

Example 9.2.1

The undefined terms are *nick*, *stane* and *bate*.
In addition we make the following definitions:

- Two stanes are *conicked* if they are bated by the same nick.
- Two stanes are *non-nicked* if there is no nick which bates both of them.

Axioms:

1. For every distinct pair of nicks A and B there exists a unique stane which is bated by A and B.
2. For every stane there exist at least two distinct nicks which bate it.
3. There exist three distinct nicks which are not bated by the same stane.

It is evident that *nick* and *stane* are objects, while *bate* is a relation between nicks and stanes. We will adopt the convention that nicks are represented by upper case letters A, B, etc. And that stanes are represented by lower case letters x, y, z. We now prove some theorems:

Theorem 9.2.2 *If x and y are distinct stanes which are not non-nicked, then there is a unique nick which bates both of them.*

Proof:
Since x and y are not non-nicked there must be at least one nick which bates them both, call this nick A.
Suppose that there is another nick, B say, which bates both x and y.
But this contradicts Axiom 1, which says that each pair of nicks bates a *unique* stane. $\square$

Theorem 9.2.3 *For every stane there is at least one nick which does not bate it.*

Proof:
Suppose that there is a stane which bates every nick, this would contradict Axiom 3. $\square$

Theorem 9.2.4 *There is at least one stane.*

Proof:
Axiom 3 tells us that there exist at least three nicks.
Axiom 2 tells us that any two of these must bate a stane, so there must be a stane for them to bate. $\square$

If you are not totally confused by this nick, you might want to consider a career as a pure mathematician! If you are not confused it is probably because you have managed to build up an effective model for this system.

The problem is that while using nonsense words protects us from introducing unwarranted assumptions, it provides no intuitive model from which to work. This brings us back to the Platonic argument about idealised worlds only on a whole new level.

The Platonic view is that mathematical objects exist independently, in some idealised world. We may then manipulate them, using our intuition to guide us. The axiomatic or formal viewpoint holds that mathematical objects exist only as defined by their axioms, and to approach them in any other way invites disaster.

While the axiomatic view is clearly *mathematically* correct, we can rely only on the axioms and their consequences. We, as humans, require some intuitive framework from which to work. Thus, in practice, we start with a model or framework from which to work and build up theorems based on this model, we then go back and build up an axiomatic system to show that the model is mathematically well founded. We may go through this process a number of times before a reliable theory is developed.

It is not even clear when two axiomatic systems are in fact talking about the same set of models. In this case we say that the systems are *equivalent*. If we replace the words nick,

stane and bate by pom, lorn and mort respectively we can see that
Theorem 9.2.4 is Axiom 1 of the mort system.
Axiom 1 of the bate system is Axiom 2 of the mort system.
Axiom 2 is Axiom 3 of the mort system, and
Theorem 9.2.3 is Axiom 4 of the mort system,
Thus the mort system can be *proved* from the bate system.

Conversely, Axiom 2 of the mort system is Axiom 1 of the bate system,
Axiom 3 of the mort system is Axiom 2 of the bate system,
Theorem 9.1.3 is Axiom 3 of the mort system.

Thus we have shown that it is possible to prove the axioms of the mort system from the bate system, and to prove the axioms of the stane system from the mort system. We therefore say that these two systems are *equivalent*. Though this is not at all obvious at the outset. In this case any theorem in one system has an equivalent in the other. This means that any models given for the mort system are also models of the bate system and vice versa.

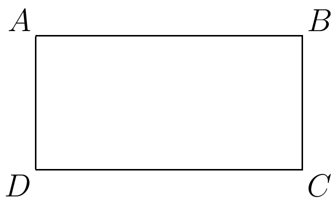
9.3 Models

We could look for more models of this axiomatic system. One possible model is the following:
'nick' means a letter A, B, C etc.
'stane' means a set of two letters $\{A, B\}, \{A, C\}, \{B, C\}$ etc.
'bate' means "is a member of".

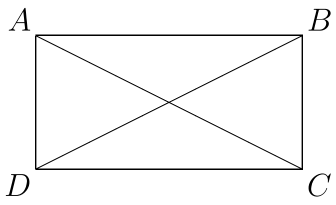
This is a perfectly valid model which clearly satisfies Axioms 1 and 2 and, provided there are at least three nicks, satisfies Axiom 3.

Another model would be to consider the system as a triangle, with nicks being the vertices and stanes the edges. Bate then means 'lies on'. This model is in fact isomorphic to the previous one when there are exactly three nicks.

What about trying to extend this model to the case of four points? The obvious generalisation is a rectangle. However a rectangle does not satisfy Axiom 1, both A and C, and B and D have no common edge.



We could add these edges to get the following model.



This is not the only model with exactly four nicks. Consider the stanes:

$$P_1 = \{A, B, C\} \quad P_2 = \{C, D\} \quad P_3 = \{A, D\}$$

In this case we allow stanes which are not bated by the same number of nicks.

If we want to look for other models where stanes are always bated by the same number of nicks we might try the following:
'nick' means a letter A, B, C etc.
'stane' means a set of three letters $\{A, B, C\}$ etc.
'bate' means "is a member of".

In the case of three letters we get the single stane $\{A, B, C\}$, but this does not satisfy Axiom 3.

What about four nicks? In this case we may arbitrarily take our first stane as $\{A, B, C\}$ with D being the fourth point, which does not bate this stane.
But now when we try to place a second stane we run into a problem, it must contain D and two other elements from the first stane, but then these two elements do not bate a *unique* stane, they are in both stanes.

A similar argument shows that such a system with 5 or 6 nicks is also impossible (Exercise: try and show this). It seems as if this model does not work. However when we come to seven nicks we get the following model:

$$\begin{array}{cccc} \{1,2,4\} & \{2,3,5\} & \{3,4,6\} & \{4,5,7\} \\ \{1,5,6\} & \{2,6,7\} & \{1,3,7\} & \end{array}$$

Here we have used the numbers 1 to 7 to denote the nicks, rather than letters. This is exactly the same system as introduced at the very beginning of this chapter. We will refer to this system as $S(7)$.

Continuing this scheme shows eight nicks to be impossible. Nine nicks gives us the following 12 stanes:

$$\begin{array}{ccc} \{1,2,3\} & \{4,5,6\} & \{7,8,9\} \\ \{1,4,7\} & \{2,5,8\} & \{3,6,9\} \\ \{1,5,9\} & \{2,6,7\} & \{3,4,8\} \\ \{1,6,8\} & \{2,4,9\} & \{3,5,7\} \end{array}$$

We will refer to this system as $S(9)$.

In general it can be shown that such a scheme exists if and only if the number of nicks is 1 or 3 mod 6, that is the remainder of the number of nicks when divided by 6 is either 1 or 3. These are known as finite geometries.

There is another interpretation of these axioms, we let nicks be points and stains be lines in the Euclidean plane. Bate is then taken to mean 'lies on'. In this case the number of nicks and stanes is infinite. The axioms are the 'incidence axioms' introduced by Hilbert in his axiomatisation of Euclidean geometry. These axioms essentially define what we mean by 'lies on' or 'is incident with'. Thus this axiomatic system represents the first step toward axiomatic geometry. However we can see that there are many models which are consistent with this set of axioms. By restricting ourselves to planar geometry we may miss some of the subtleties or applications of this axiomatic system.

9.3.1 Incompleteness

What about Euclid's parallel postulate? We can state it in the form of Playfair:

Playfair's Postulate: For every line, l, and every point a not on the line there is a unique line through a which does not meet l.

Stated in the language of nicks and stanes this becomes (note that *conick* corresponds to *meet*):

Postulate X: For every stane, x, and every nick, A, not bating that stane there is a unique stane which is bated by A and does not conick x.

We consider this statement in connection with the finite geometries $S(7)$ and $S(9)$ above. In this context the postulate becomes:

Postulate X': For every stane, x, and every nick, A, not in the stane there is a unique stane which contains A and does not contain any element of x.

Consider the system $S(7)$, this system does not satisfy this postulate. For example if we consider the stane $\{1, 2, 4\}$ and the nick 5, 5 appears in $\{2, 3, 5\}, \{4, 5, 7\}$ and $\{1, 5, 6\}$, each of which contain a point of $\{1, 2, 4\}$.

Now consider $S(9)$, we find that it does indeed satisfy X'. For example, if we consider the stane $\{1, 2, 3\}$ and the nick 4, we can see that there is one and only one stane which does not contain 1, 2 or 3 and which does contain 4, namely $\{4, 5, 6\}$. To prove that $S(9)$ does satisfy X' we have to check that it holds for every stane, and every nick not on that stane, this is straightforward but tedious. In fact we can see that each row, as given, defines a complete set of 'parallel' stanes.

But what are we to make of this? We have two models, both of which satisfy the axioms, but which do not agree on postulate X. We must conclude that postulate X is *undecidable* within this axiomatic system. That is these axioms alone do not provide us with enough information to prove or disprove postulate X. An axiomatic system which contains undecidable statements is called *incomplete.*

In order to provide a complete system we must either take postulate X as a new axiom, or take its converse as a new axiom. If we do the latter there are, in fact, two ways of contradicting postulate X. This gives us three possibilities for a new axiom:

Axiom X_0: For every stane, x, and every nick not bating the stane x there are no stanes which are bated by that nick and do not conick x.

Axiom X_1: For every stane, x, and every nick not bating the stane x there is a unique stane which is bated by that nick and does not conick x.

Axiom X_∞: For every stane, x, and every nick not bating the stane x there is more than one stane which is bated by that nick and does not conick x.

Axiom X_0 is the postulate that leads to *spherical geometry,* it is also obeyed by $S(7)$.
Axiom X_1 is Euclid's parallel postulate of *planar geometry,* and is also obeyed by $S(9)$.

Axiom X_∞ is the postulate that leads to *hyperbolic geometry,* it is also obeyed by $S(13)$:

$$\begin{array}{ccccc}
\{1,2,5\} & \{2,3,6\} & \{3,4,7\} & \{4,5,8\} & \{5,6,9\} \\
\{6,7,10\} & \{7,8,11\} & \{8,9,12\} & \{9,10,13\} & \{1,10,11\} \\
\{2,11,12\} & \{3,12,13\} & \{1,4,13\} & \{1,3,8\} & \{2,4,9\} \\
\{3,5,10\} & \{4,6,11\} & \{5,7,12\} & \{6,8,13\} & \{1,7,9\} \\
\{2,8,10\} & \{3,9,11\} & \{4,10,12\} & \{5,11,13\} & \{1,6,12\} \\
\{2,7,13\} & & & &
\end{array}$$

For example taking the stane $\{1,2,5\}$ and the nick 3, $\{3,4,7\},\{3,12,13\}$ and $\{3,9,11\}$ contain 3 and do not contain any point of the given stane. In fact, in this case, for every stane and every nick not in that stane, there are exactly three stanes which contain the nick and do not intersect the original stane.

We might be tempted to invent a new form of the axiom to cover this case:

Axiom X_3: For every stane, x, and every nick not bating the stane there are exactly three stanes which are bated by that nick and do not conick x.

In General we could have:

Axiom X_i: For every stane, x, and every nick not bating the stane there are exactly i stanes which are bated by that nick and do not conick x.

It should be noted that for each of the resulting axiom systems created by adding one of either X_0, X_1 etc. there are many possible models.

In the model of standard Euclidean planar geometry we take the lines to be lying on a flat plane.
In the model of spherical geometry we take lines to be the great circles of a sphere, each 'line' then meets every other in exactly two points.
In the model of hyperbolic geometry we take lines to be the hyperbolas lying on a hyperbolae. In this case there are an infinite number of 'lines' through a point which do not intersect a given line.

It is interesting to note that current theories (relativity) actually state that the universe is a four dimensional hyperbolic geometry, with the fourth dimension being time. For more on spherical and hyperbolic geometries see Davis' book.

When Gauss and Lobachevsky first deduced these so called non Euclidean geometries around 1850 it caused an such a stir, not because the geometries were in themselves so interesting, but because this revealed a whole new way of looking at mathematics.

The axiomatic or formal method released mathematics from the constraints of what we can visualise, anything was fair game, any set of axioms could be taken and the consequences deduced to form a whole new area of endeavour. Conversely the theorems derived from a set of axioms were no longer constrained by the model used but had general significance to any model which obeyed the given axioms.

9.3.2 Consistency

But can we take *any* set of axioms? Are there constraints on this new found power? What if we were to add both the axioms X_0 *and* X_1 to our system? In this case we have two contradictory axioms, clearly this is not a good idea. But the problem is more subtle, in this case it is obvious that the axioms contradict each other, what if it were not so obvious? What if after some arduous chain of reasoning from the axioms we were able to prove two mutually contradictory statements?

If it is possible to prove two mutually contradictory statements from a set of axioms, then this set of axioms is called *inconsistent*, otherwise a set of axioms is said to be *consistent.*

But how do we know if a set of axioms is inconsistent beforehand? It may be that the axioms appear perfectly reasonable but in fact, far down the logical path, contain an inconsistency. A general rule of thumb is the following:

> A set of axioms is consistent if an only if it has a model which obeys those axioms.

In order to actually show that a set of axioms is inconsistent we must actually prove two contradictory statements.

There is one other thing to keep in mind when dealing with formal axiomatic systems. It is possible that we have in fact taken extra unnecessary axioms. This is true if it is possible to derive one of the axioms from the others. If this is the case we may drop the derived axiom as we can prove it as a theorem. If it is not possible to prove any of the axioms from the others we say that the system is *independent.* It is always desirable to have as few axioms as possible, this is the same as saying that axioms should be independent.

In this Chapter we have been introduced to the idea of a *formal system.* We have seen how a given system may have many *models*, which may be *isomorphic* or *non-isomorphic* to each other. A set of axioms may be *incomplete* if there are *undecidable* (i.e. unprovable) statements. The axioms should be *consistent* and *independent.* It is possible that seemingly different systems of axioms are *equivalent*, representing the same set of models.

Chapter 10

The Axiomatisation of Mathematics

The introduction of formal systems began with the discovery by Gauss and Lobachevsky of non Euclidean geometries in the 1850's. George Boole (1847) had already provided an essentially axiomatic formulation of logic, which was honed in subsequent years. Much effort went into trying to discover the fundamental axioms associated with various areas of mathematics over the next few years. By 1880 the stage was set and the rush was on to axiomatise all areas of mathematics. This was done over the next few years, though not without controversy.

10.1 The Axiomatic Foundation of Numbers

10.1.1 The Natural Numbers

In 1894 Peano (1858 - 1932) provided an axiomatic foundation for the natural numbers.

Undefined terms: *zero, number* and *successor.*
Axioms:

P1 Zero is a number.

P2 If a is a number then the successor of a is a number.

P3 Zero is not the successor of any other number.

P4 Two numbers with the same successors are themselves the same.

P5 If a set S of numbers contains zero and also the successor of every number in S, then every number is in S.

We then give names to the various successors, for example we call the successor of zero 'one', the successor of one is called 'two' and so on. The final axiom essentially says that there are no numbers except as defined by the successor relationship.

We may go on to define addition as repeated successions, and then multiplication in terms of addition. Thus we would define $a+b$ to be the number obtained by applying the successor operation b times to a. Then $a \times b$ would be the result of adding a to itself b times. One potential problem with this approach is that we use the notion of number (as a counting mechanism) to define addition of numbers. However if we realise that numbers as a counting mechanism are a model of these axioms we can use this model in further definitions.

10.1.2 Addition – The Integers

It is however usual to take a slightly different approach to the definition of the algebraic operations of addition and multiplication, defining them in terms of their formal properties. In this case addition and multiplication become specific models of a more general axiomatic scheme. We must of course prove that the formal definition below fits with the definition of addition as repeated successions.

Undefined terms: *number* , '+' and '='
Axioms:

A0 *Closure*
For all numbers a and b, $a+b$ is also a number.

A1 *Commutative Law*
For all numbers a and b, $a+b=b+a$

A2 *Associative Law*
For all numbers a, b and c, $(a+b)+c=a+(b+c)$.

A3 *Existence of Identity*
There exist a number, denoted 0, such that for every number a, $a+0=0+a=a$.

A4 *Existence of Inverse*
For every number a, there exists a number, denoted $-a$, called the *additive inverse of a,* such that $a+(-a)=(-a)+a=0$.

Note how the number 0 is defined in Axiom A3 and then used in Axiom A4.

These axioms are so general that any set, S, on which a binary operation [1] is defined, which satisfies axioms A0, A2, A3 and A4, is called a *Group.* If this operation also satisfies Axiom A1 it is called an *Abelian Group*[2] or *Commutative Group.* There are many examples of things which are groups, which bear no obvious relation to addition. One example is rotations of the plane, another is the symmetries of a square.

In order to connect these axioms to those of Peano we must identify number with Peano's number and Peano's 'Zero' with the '0' of these axioms. In fact the natural numbers do not satisfy Axiom A4, for this reason we use the general term 'number' rather than 'natural number' in the axioms. It is necessary to extend the natural numbers to the integers by defining for each natural number, a, another number $-a$ with the property that $a+(-a) = 0$. Thus the integers come as a result of defining the inverse of each element with respect to addition.

10.1.3 Multiplication – The Rationals

To axiomatise multiplication we note that multiplication is also a commutative group, that is it satisfies exactly the same axioms as above, with $\times$ in place of $+$.

Undefined term: $\times$
Axioms:

M0 *Closure*
For all numbers a and b, $\times b$ is also a number.

M1 *Commutative Law*
For all numbers a and b, $a \times b = b \times a$

M2 *Associative Law*
For all numbers a, b and c, $(a \times b) \times c = a \times (b \times c)$.

M3 *Existence of Identity*
There exist a number, denoted 1, such that for every number a, $1 \times a = a \times 1 = a$.

M4 *Existence of Inverse*
For every non zero number a, there exists a number, denoted $\frac{1}{a}$, called the *multiplicative inverse of a,* such that $a \times \frac{1}{a} = \frac{1}{a} \times a = 1$.

[1]A binary operation is one which takes two inputs from the set and produces a single member of the set.
[2]Named after the Norwegian mathematician Niels Abel (1802 - 1832)

Of course the integers do not satisfy Axiom M4, so as above we must extend our set to include all the numbers $\frac{1}{a}$. In addition we must include all possible multiples of these numbers, otherwise $b \times \frac{1}{a}$, (where b is not a multiple of a) would not be a valid number in our set and $\times$ would not always produce a valid number (M0). We denote this number as $\frac{b}{a}$. In this way we have defined the rational numbers.

10.1.4 Fields

Finally we need axioms to relate multiplication and addition:

Distributive Laws
For all numbers a, b and c,

D1 $a \times (b + c) = a \times b + a \times c$.

D2 $(b + c) \times a = b \times a + c \times a$.

Given a set S with two binary operations defined, both of which are Abelian groups, and which satisfy the distributive axioms D1 and D2 above it is called a *Field.*

10.1.5 The Order Axioms

We have not yet considered the order of numbers, we need a few axioms to define the order relations among numbers. These are known as the order axioms.

Undefined term: *positive*
Axioms:

O1 For any numbers a and b, if a and b are positive, then so are $a + b$ and $a \times b$.

O2 For every non zero number, a, either a is positive or $-a$ is positive, but not both.

O3 0 is not positive.

Of course theses axioms are actually satisfied by the natural numbers, the integers, the rationals and the real numbers. Any field which satisfies the order axioms is called an *ordered field.* We may now define the terms $<$, $>$ and *negative* from these order axioms.

Definition:

- $a < b$ means that $b + (-a)$ is positive.

- If $a < 0$ we say that a is *negative.*
- $a > b$ means that $b + (-a)$ is negative.
- $a \geq b$ means that either $a > b$, or $a = b$.

10.1.6 The Real Numbers

But what of the real numbers? Up to this point we have defined only the rational numbers. In order to generate the real numbers we need one more axiom which relates to the infinitesimal or continuous nature of the reals. In order to state this axiom we need another definition:

> For any set of (rational) numbers S, x is said to be a *least upper bound of S* if for every element a of S, $x \geq a$ and x is less than every other number which satisfies this property.

It should be noted that while all the elements of S may be rational the least upper bound may not be. We have already seen examples of sequences of rational numbers which tend (monotonically) to a given non rational number. Thus, though all the elements of the sequence are rational, the least upper bound is not. These sequences of rational numbers can be used to define the irrational number to which they tend. In order to formalise this notion we introduce one final axiom.

Axiom:

LUB *Least Upper Bound Axiom*
Any nonempty set of numbers that is bounded above has a least upper bound.

This axiom was first introduced by J.W.R. Dedekind (1831 - 1916) in 1872, though he may have had the idea as early as 1858. Dedekind came to this idea by considering the correspondence of the real numbers to geometric lines, and then considering the notion of continuity. He came up with the following idea.

> For every division of the rational numbers into two classes such that all members of the first class are less than all the members of the second class there is one and only one (real) number which produces this cut.

These cuts are known as Dedekind cuts. Thus $\sqrt{2}$ produces a unique cut of the rationals, which is not producable by any other number. All the rationals on one side have square less than two, on the other they have square greater than two. To get the above axiom consider the class which is less than the cut, this class has least upper bound equal to the cut.

As Russell pointed out some years later $\sqrt{2}$ may actually be defined as the segment or class of positive rationals whose squares are less than 2. Since there is only one number that produces this cut of the line ($\sqrt{2}$) this definition is unique. By using this kind of definition of number, as Dedekind cuts, we generate the real numbers.

Since Calculus can be described by operations on real numbers this formal definition of the reals also allows a formal description of calculus.

10.2 The Axiomatisation of Geometry

Meanwhile, in 1899 Hilbert (1862 - 1943) produced a completely axiomatised version of Euclidean (planar) geometry. Hilbert's axioms appear in five groups, at each stage new undefined terms are introduced. Each group defines a new idea, thus the first three 'incidence' axioms define the notion of incidence of points on lines. The reason for doing this is that the groups of axioms rely only on the previous groups, and we may adopt only some of the groups to get differing notions of geometry. Thus, for example taking the first three 'incidence' axioms alone gives the 'incidence geometries' discussed in the previous chapter.

Group I – Axioms of Incidence

Undefined terms: *point, line* and *incident with.*

H1 Any two distinct points A and B are incident with one and only one line, called the line AB.

H2 Every line is incident with at least two points.

H3 There exist at least three distinct points not all of which are incident with the same line.

Group II – Axioms of Order

Undefined term: *between*

H4 If a point B is between two other points A and C, then A, B and C are three distinct points incident with the same line, and B is between C and A.

H5 If A and C are two distinct points incident on a line then there is at least one point B between A and C.

H6 Of any three points incident on the same line, there is one and only one which is between the other two.

At this point it is possible to define a number of useful concepts.

Definitions:

- The *line segment* AB is the line consisting of all the points between A and B.
- The *ray* from A through B consists of all those points C on the line AB such that A is not between B and C.
- Given three points A, B and C, all on a line ℓ, B and C are said to be *on the same side* of A on the line ℓ if they both lie on the ray from A through B.
- Given two lines segments, or rays h and k with the same endpoint A they form an *angle,* $\angle(h, k)$. If h is AB and k is AC, we write $\angle BAC$ or $\angle CAB$.
- Given a line ℓ and two points A and B not on ℓ we say that A and B are *on different sides of* ℓ if the line segment AB intersects ℓ. Otherwise we say that A and B are *on the same side of* ℓ*.*
- Given two rays m and n, which form an angle, $\angle(m, n)$, the *interior points of the angle* are all those points which are both on the same side of m as n, and on the same side of n as m.
- Given three points A, B and C, not on the same line, the *triangle* ABC is described by the line segments AB, AC and BC.

There is one more axiom in this group.

H7 Let A, B and C be three points not all incident with the same line, and let ℓ be a line not incident with any of A, B or C. Then, if the line ℓ passes through a point of the segment AB, it will also pass through a point of the segment BC or a point of the segment AC.

Group III – Axioms of Congruence

Undefined term: *congruent to* $(\equiv)$

H8 If A and B are two points on a line ℓ and if C is a point on the same or another line m, then on a given side of C on m there is one and only one point D such that the segment AB is congruent to CD.

H9 *Transitivity of Congruence*
Given three line segments AB, CD and EF, if $AB \equiv CD$ and $CD \equiv EF$, then $AB \equiv EF$.

H10 Let AB and AC be two segments of the straight line ℓ, which have only the point B in common. Let DE and DF be segments of the same or another line m, which have only the point D in common. Then, if $AB \equiv DE$ and $AC \equiv DF$ then $BC \equiv EF$.

H11 Let an angle $\angle(h, k)$, a ray m of a line ℓ and a prescribed side of the line ℓ be given. Then there is exactly one ray n such that $\angle(h, k) \equiv \angle(m, n)$ and at the same time all interior points of the angle are on the prescribed side of ℓ.

H12 If, given two triangles, ABC and DEF, we have that $AB \equiv DE$, $AC \equiv DF$, and $\angle BAC \equiv \angle EDF$, then $\angle ABC \equiv DEF$.[3]

Group IV – Euclid's Parallel Postulate

H13 Given a line ℓ and a point A not on ℓ there is one and only one line incident with A, which does not have any point in common with ℓ.

Group V – Archimedean Axiom

H14 Given two points, A and B, on a line, ℓ, let $A_1, A_2, A_3, \ldots$ be points on ℓ such that the line segments $AA_1 \equiv A_1A_2 \equiv A_2A_3 \equiv \ldots$. Further suppose that A_1 lies between A and A_2, A_2 lies between A_1 and A_3, etc. Then in the sequence $A_1, A_2, A_3, \ldots$ there is always a point A_n such that B lies between A and A_n.

Axiom H14 is often attributed to Archimedes, hence the name, but it is essentially equivalent to Eudoxus' principle of exhaustion. Archimedes is quite clear in his statement that "...earlier geometers have used this lemma ...". Archimedes statement is in terms of area, rather than lengths as is the case here.

[3]This is essentially Euclid's Proposition 4

> That the excess by which the greater of two unequal areas exceeds the less can, by being added to itself, be made to exceed any given finite area.

These axioms clearly define what is meant by congruence of angles and lines. In addition the notion of 'between' is seen as fundamental. In fact it is used to define many common concepts in geometry. Note how angle is defined in terms of intersecting rays. This provides no notion of the measurement of angles, however the notion of two angles being congruent is defined by Axiom H11. Axiom H12 is a statement of the relevant part of Euclid's Proposition 4, which caused so many worries.

10.3 Topology

One of the great successes of the axiomatisation of mathematics was the development of *topology*. This was an area of mathematics which actually did not exist before, and only came into existence as a result of a long search for fundamental axioms. Some of the names associated with the early development of topology are Felix Hausdorff (1868 - 1942) and Maurice Frechet (1878 - 1973).

Through the development of axioms it was possible to consider and define the most basic of mathematical relations. This endeavour became known as topology. Topology is essentially a notion of set theory in which the nature of the elements is irrelevant, only the relations among elements are considered important. In 1914 Felix Hausdorff introduced the axioms of topology.

Undefined terms: *point, neighbourhood* and *contain.*

Notation:
The neighbourhood $U(x)$ is assumed to contain the point x.

Definition:
Given two neighbourhoods U and V, we say that U is a *subset* of V if and only if every point contained in U is also contained in V.

Axioms:

T1 To each point, x, there corresponds at least one neighbourhood $U(x)$ which contains x, and each neighbourhood contains at least one point.

T2 If $U(x)$ and $V(x)$ are two neighbourhoods containing the same point x, there must exist a neighbourhood containing x which is a subset of both. (i.e. every point of which is contained in both $U(x)$ and $V(x)$.)

T3 If the point y is contained in $U(x)$, there must exist a neighbourhood of y, $V(y)$ say, that is a subset of $U(x)$. (i.e. $U(x)$ contains every point which is contained in $V(y)$.)

T4 For any two different points x and y there are neighbourhoods $U(x)$ and $V(y)$, with no points contained in both.

A useful model to keep in mind when thinking about topology is the plane. The points are points in the plane, the neighbourhoods are regions of the plane with their boundaries removed.

The power of topology is that it is exceedingly general in nature, many things and relations can be considered from a purely topological standpoint. From these axioms it is possible to define the boundary and the interior of a neighbourhood. It is also possible to define a notion of continuity. In fact these topological axioms lead to what is known today as measure theory. One model of measure theory is integration, and hence differentiation. Thus topology allows a notion of calculus *that makes no reference to the real numbers!* This general notion of calculus can be applied to many strange objects.

10.4 Russell's Paradox and Gödel's Theorem

The power of these systems was phenomenal, it allowed whole new areas to be developed, as well as a completely new understanding of old areas. The idea that the same axiomatic systems could be applied to different models encouraged an extremely abstract approach to mathematical objects. The fundamental essence of what made a system work the way it does could be distilled. A good example of this was group theory, the original ideas where developed by Galois, but it was developed almost immediately as a purely axiomatic theory.

With this rush to axiomatise everything it seemed reasonable that it should be possible to write a book, which started with some basic axioms, and then derived all of mathematics from these axioms. At the beginning of this century Alfred Whitehead (1861 - 1947) and Bertrand Russell (1872 - 1970) attempted to write just such a book, *Principia Mathematica,* which actually appeared in three volumes from 1910 to 1913. The scope of this work was incredible, the proof that $1 + 1 = 2$ does not appear until page 362.

However Russell and Whitehead ran into a problem, they kept running into inconsistencies. Every time they tried to fix an inconsistency it would pop up again elsewhere. Though they did eventually publish, the work was flawed in that it was incomplete, there were mathematical theorems it could not address from the fundamental axioms.

The basic inconsistency that they found is known as Russell's paradox. Russell provided the following simple puzzle, known as the barbers paradox, to exemplify the problem:

> In a certain town there is a male barber who shaves all those men, and only those men, who do not shave themselves.
> The question is, who shaves the barber?

If the barber does not shave himself, then he should, since he shaves all those men who do not shave themselves. On the other hand the barber *only* shaves such men, and hence cannot shave himself.

The fundamental problem appeared in the axiomatisation of set theory. To understand the mathematical interpretation of Russell's paradox it is first necessary to understand that a set may itself have sets as members. Thus the set $A = \{0, \{1\}\}$ has the number 0 as an element and the set $\{1\}$ as an element.

To see that sets can have themselves as members consider the *universal* set of everything. We saw how universal sets are often used to define other sets, it seems reasonable that we should be allowed to make a universal set, from which other sets are drawn, which contains absolutely everything in the universe, real or imagined. But one thing in the universe (since we just defined it) is this universal set. Thus the universal set must contain itself as a member. Another example of a set with this property is the set of all sets. This is a set and so must contain itself as a member.

A more general way to describe sets which have themselves as members is to consider a sequence of sets B_i, indexed by the natural numbers i. Each of which with the property that they have the previous one as a member. That is $B_{i-1} \in B_i$, for each $i > 1$. Now consider the set $B = \lim_{i\to\infty} B_i$. It seems evident that this set has the property that it has itself as a member.

While the property of self containment is unusual it is not, in and of itself, paradoxical, we obtain a paradox when we consider 'the set of all sets which are not members of themselves'.

$$R = \{ \text{ Sets } A \,|\, A \notin A\}$$

The question is, is R a member of itself?

If R contains itself as a member, then by the definition of R, R should not be a member of R, since it contains itself, and R contains only those sets which do not contain themselves as members. On the other hand if R is not a member of R then it should be, since R is defined to contain all sets which are not members of themselves. Thus we have a paradox, R cannot not be a member of itself, but it must be. This is known as Russell's paradox.

If we go back to the definition of set given earlier, we can actually find a way out of this dilemma.

> A *set* is a collection of objects, such that it is possible to state unequivocally whether any given object is in the set or not.

Since we have an object, R, for which it is not possible to state unequivocally $R \in R$ or $R \notin R$, we must conclude that, by this definition, R is not a set. This is exactly analogous to the way in which paradoxical statements are not allowed. It should be noted that this definition of set is due to Russell and Whitehead, and was given precisely to get themselves out of the paradox.

The trouble with this approach is that it restricts the range of objects about which we can reasonably talk. The resulting system is incomplete. That is there are relevant statements which we cannot prove or disprove from the axioms.

After this mathematicians attempted to fix the axiomatic theories, to find some consistent set of axioms which allowed all of mathematics to be derived.

The final word in this story came from the Austrian mathematician Kurt Gödel (1906 - 1978). In 1931 Gödel came up with his famous theorem, which may be succinctly stated as follows.

> Every formal axiomatic system, with a finite number of axioms, is either incomplete or inconsistent.

That is to say that any formal system either contains a paradox (i.e. inconsistency), or it is possible to formulate relevant statements which are not provable within the system. Of course we can then take these statements to be new axioms, but then the new system must contain further new incompletenesses, and so on, ad infinitum.

What is truly incredible about this is that Gödel managed to use a formal mathematical approach to show that mathematics can never be both complete and consistent. [4]

Of course in mathematics we eshrew inconsistency and always go for incompleteness. Thus there are statements about mathematical objects which are undecidable within the formal system from which they came. The trouble is that, usually, there is no way to know whether we can't prove a given statement because we can't think of a proof, or because it is literally unprovable.

[4] See *Gödel, Escher, Bach* by D. Hofstadter for a more complete discussion of Gödel's theorem.

Bibliography

Hisory of Mathematics:

1. Carl B. Boyer & Uta C. Merzbach *A History of Mathematics*, John Wiley and Sons, ISBN 0-471-54397-7.
   A fairly detailed History of Mathematics, from antiquity to the modern day. Of the detailed histories this is the most recomended.

2. Donald M. Davis *The Nature and Power of Mathematics.* Princeton ISBN 0-691-08783-0 & 0-691-02562-2
   A laymans introduction. Covers Euclid in some detail. This book covers many of the topics which we will be looking at.

3. William Dunham *Journey Through Genius*, Penguin
   Gives accounts of a selected group of proofs. Also discusses the men who made them.

4. Euclid *The Thirteen books of the Elements*, Dover.
   The Original. Contains detailed bibliographic notes and commentary. However it is a fairly difficult read.

5. Stuart Hollingdale *Makers of Mathematics*, Penguin Good summary of the history of Mathematics.

6. Douglas Hofstadter *Gödel, Escher, Bach, an Eternal Golden Braid.* Random House.
   Not really a History at all. However this is an excellent book which introduces Gödels theorem to the layman.

7. M. Kline *Mathematics in Western Culture*
   A general history of mathematics.

8. Jerry P. King *The Art of Mathematics*
   A General discussion of mathematics as art.

9. M. Kline *A History of Mathemtics*
   A detailed history of mathematics, in three volumes.

10. Bertrand Russell *Wisdom of the West*, Fawcett
    A detailed history of Greek philosophy and conceptions of thought.

11. Bertrand Russell *History of Western Philosophy*
    A detailed history of western philosophy, contains most of the material in *Wisdom of the West.*