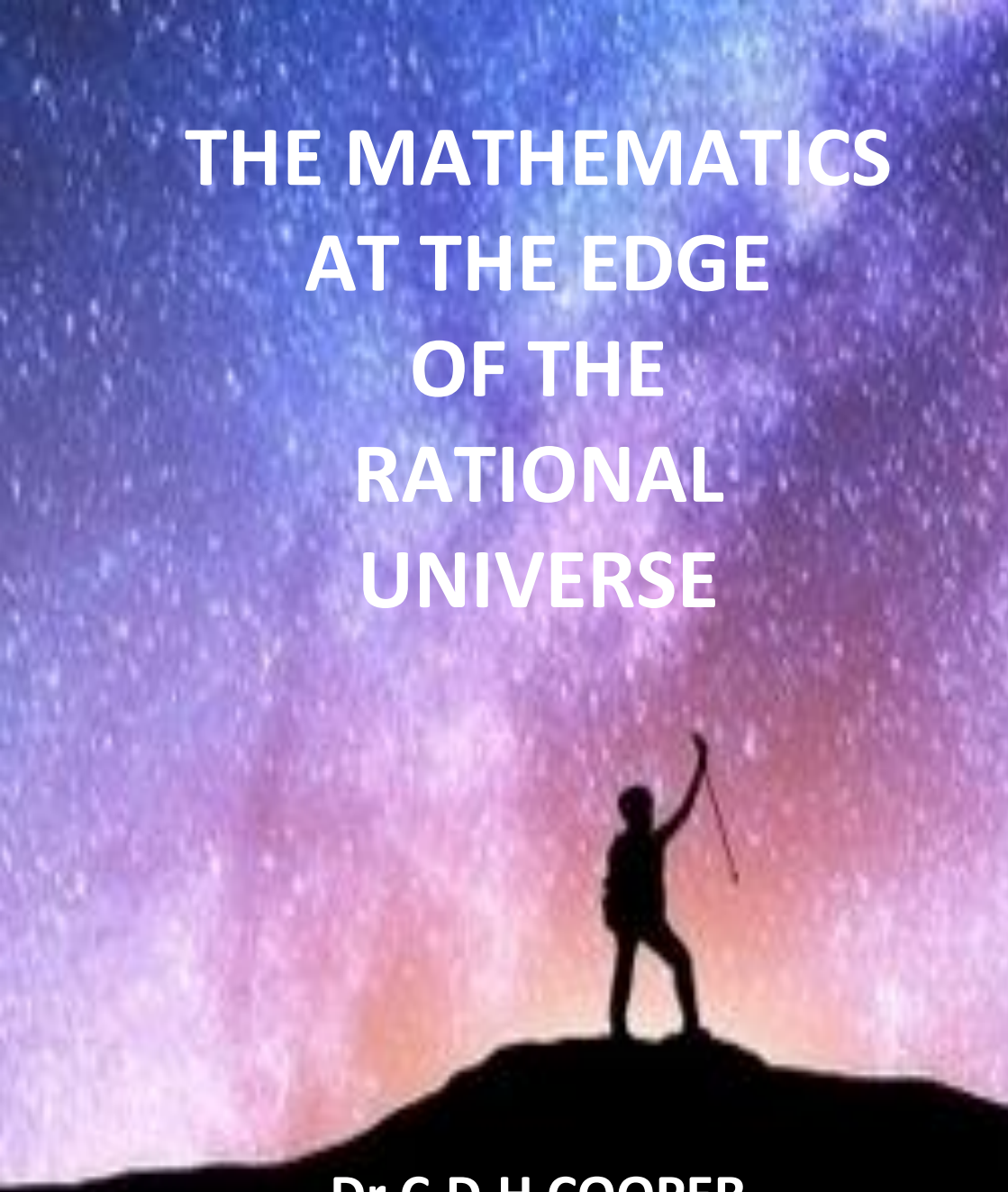
THE MATHEMATICS AT THE EDGE OF THE RATIONAL UNIVERSE



Dr C.D.H COOPER

THE MATHEMATICS AT THE EDGE OF THE RATIONAL UNIVERSE

7th edition Feb 2021

These notes were prepared for students at Macquarie University in Australia but are freely available to anyone. However if you make use of them and are not a Macquarie University student it would be nice if you could email me at christopherdonaldcooper@gmail.com to let me know where you are from. And, if you are from outside of Australia perhaps you could send me a postcard of where you are from to pin up on my wall (Christopher Cooper, 31 Epping Avenue, EASTWOOD, NSW 2122, Australia).

I dedicate this book to Rachel Chalmers, whose email many years after she attended one of my courses encouraged me to write up those notes for publication. Rachel, I can barely remember who you were, but if it hadn't been for your chance email this book would probably never have been finished.

PRELUDE

Mathematics is the art of story-telling. Nobody has ever seen a perfectly round circle or an infinitely long line of zero width. They're pure figments of the mathematical imagination. As for imaginary square roots of −1, ideal points where parallel lines meet, and 6-dimensional space …! What fantasies can be dreamt up by the fertile mind of a mathematician!

Stories, parables, fables, myths and legends can carry profound truths that have a powerful impact on the lives we lead. Mathematical stories are no exception. This gossamer web we mathematicians spin might be pure fancy. But it's the best tool we have to understand and predict the material universe. And it reaches far beyond.

In this book we'll go on a journey to the edge of the rational universe. Our motivation will be that of an explorer. We simply want to know what's out there. Whether any practical use can be made of what we find there is not our prime concern. This book is not written for the practitioner in logic or mathematics or computing science.

Having said that let me add that the inspiration for the book came from having to teach this material to embryonic mathematicians and computer scientists in several courses at Macquarie University. I began to realise that, stripped of some of the formal technicalities, much of the material I had taught to third

year students, to honours students and even to postgraduate students could be made accessible to a wider audience.

Material which had hitherto remained locked up in courses with such intimidating names as Advanced Algebra, Axiomatic Set Theory and Theory of Computation is too fascinating to leave there. All it needs is a little less emphasis on symbolic formality and a little more imaginative presentation.

That's not to say that having read this book you'll be on a par with the students who graduate from my courses. I like to think that what I've done is to build a road into a national park that has hitherto only been accessible on foot.

I taught this material many years ago as a continuing education course for mathematical laymen (and laywomen) at Macquarie University. I even wrote an earlier, and much less complete, version of this book to give out to the students. And there it lay.

More recently, many years later, I received an email out from one of those students. She had attended the course with her father and said how much she'd enjoyed it. In fact she said, "I was thinking last night, it's still the best maths class I ever took, and one of the most fun things I ever did with my Dad." I thought it went well, but surely that must be an exaggeration! However, it inspired me to dust off those old notes and fashion them into this book.

I'm certainly not the first to have attempted to bring deep ideas of logic and mathematics to a wider audience. Lewis Carroll was one of the first in *Alice*

Adventures in Wonderland – a book which delightfully introduces many ideas of logic. I have also been influenced by Abbott's *Flatland* and the writings of Martin Gardener and Douglas Hofstadter.

The final chapter goes beyond transcendental mathematics to consider the philosophical/theological question of the existence of something beyond the material world and proofs of the existence of God. This isn't a technical discussion of epistemology but rather a drawing together of some of the ideas in the earlier chapters.

After each chapter there's a little treat – a story, poem or joke, reflecting the ideas developed in that chapter. These may or may not aid the understanding of the chapter but at least they provide some breathing space before the next one and hopefully maintain the whimsical frame of mind in which this material can best be appreciated.

This book isn't for everybody. Is it for you? Here's a check list. If you can answer "yes" to some or all of them then go ahead and read this book.

(1) Are you intrigued by the logical reflexiveness of the sentence "this sentence is false?

(2) Have you read and enjoyed *Alice's Adventures in Wonderland*?

(3) Can you cope with the symbols in the following?

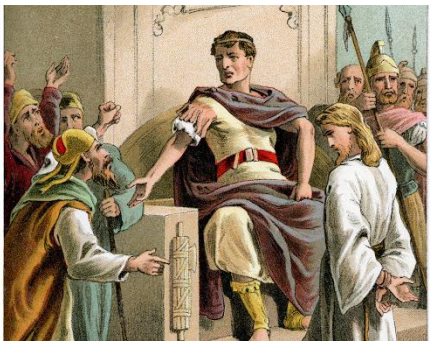
Let P denote a computer program and let D denote some data on which it acts. Suppose we denote the output by $P \rightarrow D$. So if P is a program for duplicating data then $P \rightarrow D = DD$. And if such a program is given its own description to duplicate, we have the equation $P \rightarrow P = PP$.

(4) Would it interest you if one could prove the existence of God?

CONTENTS

1. THE IMAGINARY

§1.1. Mathematics and Truth



"What is truth?" asked a famous Roman governor. Indeed, what *is* truth and how are we to know it? When we were young we soon learnt to tell the difference between truth and lies. Indeed we learnt to tell lies almost as soon as we could talk. "It wasn't me – Sarah did it!"

As we got older we learnt that things are not always what they seem. Optical illusions, and the sleight of hand of a magician, fascinated us.

As adults we've learnt that truth can be relative. Things are not always black and white. Even lies can be all shades of grey from despicable black to the purest of white.

Of all the subjects that we learnt at school, mathematics is the one where truth is most clearly defined. "What I like about mathematics," I'm often told, "is that things are either right or wrong – you really know where you stand."

Well, it's true isn't it? In our history essays it wasn't so important what conclusions we reached, we were told, but rather how well we supported them.

History isn't just about names and dates and 'facts' but more about explanations of why things happened the way they did. And *your* explanation may be quite different to *mine* yet be considered equally good. Even the facts of history undergo change as scholars revisit contemporary sources and discover that what we've been taught all these years was not actually the case.

Science is a very objective study, based as it is on observation and experiment. Yet how often has there been radical change there. The sun no longer travels round the earth as it did for centuries until Galileo. The atom is no longer an indivisible piece of matter. Light, which once travelled in a straight line, now curves in a gravitational field.

But the theorems of Euclid are still as valid as they were over two thousand years ago. With mathematics you know where you stand. Things are either true or false and when we prove that something is true that's the end of the matter. Or is it?

§1.2. Do Imaginary Numbers Exist?

People often ask "does God exist?" It's generally agreed that one cannot prove that there is a God. Some

people talk about having faith and say that they believe in God. Atheists believe that he doesn't. Agnostics say that you can't tell, and leave it at that.

What has this to do with mathematics? One of the themes of this book is to reflect on the similarities between mathematics and religion. In terms of truth you might think that they are poles apart. In mathematics everything that is believed to be true can be proved while religious truth is purely a matter of belief. But things are not quite as simple as they sound.

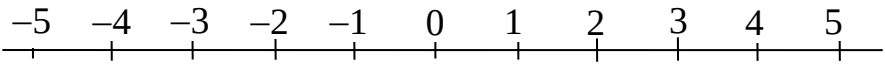
Now this is a book about mathematics not religion, so that if it should change your ideas about religious belief that's your own business. My main purpose here is to show that there are many mathematical truths that seem to be contrary to intuition and that even the concept of mathematical truth itself is not quite what you might have imagined it to be.

The German mathematician Leopold Kronecker (1823 – 1891) once said "God created the natural numbers; all else is the work of man." He meant that the so-called "natural numbers", the ones that we count with, exist in an obvious way in the world around us. But fractions, negative numbers and decimals are artificial constructions.

Mathematicians began to become bothered with existence of certain numbers in the 17th century when

they made up some new numbers. Up to then it was believed that you can't have square roots of negative numbers.

At this stage numbers were synonymous with points on the number line. In the 'middle' of an infinitely long line you have the number zero. To the left are the negative numbers and to the right are the positive ones.



Whole numbers step out in both directions in uniform steps and fractions, and more generally decimals, fill in the space between them. To use the correct mathematical terms we have the **integers** represented by points made by equal sized steps and **rational numbers**, and more generally **real numbers**, fill up the number line.

There are certain facts that can be proved about these numbers. One of these is that if you multiply two negative numbers you get a positive one. And of course if you multiply two positive numbers you get a positive one.

It follows that negative numbers don't have square roots. If you multiply a number by itself the answer will always be positive (or zero if you square

zero). Here's something which was once considered to be a true fact – but not any more!

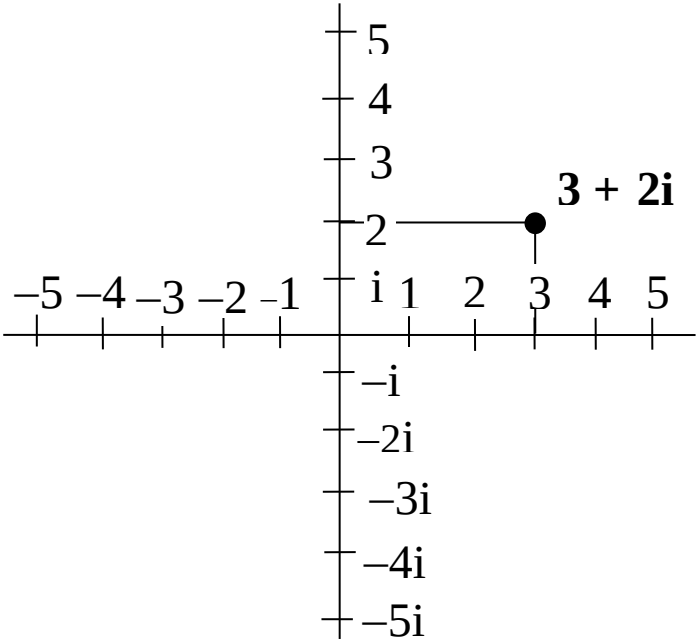
In the 17th century, in order to solve certain practical problems, mathematicians found it useful to *invent* square roots for negative numbers. These were called **imaginary numbers** because it was felt, useful as they are, that they didn't really exist. The square roots of –1 were called i and –i, with the symbol 'i' being used to remind us that they are only 'imaginary' – they don't really exist.

But how did they reconcile this with the fact that if this imaginary number i is either positive or negative its square must be positive. How could it possibly be equal to –1? Well, who's to say that this imaginary number i must be either positive or negative? Can't it be neither?

Today we live in a complex world where the simple axiom "everybody is either male or female" no longer holds. Why can't non-zero numbers be neither positive or negative? But surely the point that represents a non-zero number must be either to the left or the right of zero. On a line there are only two directions, left or right. That's true, but in a plane you can also go up and down. To accommodate these imaginary numbers we need to go into two dimensions.

If we're going to invent the imaginary number i we must allow it to combine arithmetically with the

ordinary real numbers, so that we must invent numbers such as 2i and 3 + 2i. More generally we'll have invented numbers of the form $a + b$i where a and b are ordinary real numbers. These are called **complex numbers**. The name 'complex' doesn't refer to the level of difficulty but simply to the fact that these numbers have a complex structure, being made up of two parts. This diagram shows how every complex number can be placed on what is called the **complex plane**.



It might seem a cheat when asked to add the numbers 3 and 2i and to be told that it is 3 + 2i. Is this the question or is it the answer? In fact it's both. It's not possible to simplify this answer. It is just like

asking someone to divide 2 by 3 and being told that the answer is $\frac{2}{3}$.

Doing arithmetic with these complex numbers is not difficult. To add 3 + 2i to 5 + 7i we get 8 + 9i. And $(2 + 3i)(5 + 7i) = 10 + 14i + 15i + 21i^2$. Remembering that $i^2 = -1$ we can simplify this to $10 + 29i - 21 = -11 + 29i$. Division is a little bit trickier.

These complex numbers proved to be extremely useful. Indeed modern electronics couldn't exist without them. But the philosophic question remains "do these imaginary numbers really *exist*?" Mathematicians used to struggle over such questions but the modern mantra is "if it doesn't exist then you just invent it".

So when faced with fact that parallel lines in a plane don't meet, mathematicians just invented new points where they *do* meet and so created a new geometry called the **Projective Plane**. There's a sense in which mathematics is just a game of make-believe! But it has proved to be such a useful make-believe that we don't fuss about whether these made up entities exist. It's a non-question.

Many people say that God is just something made up by people in order to explain the natural phenomena such as thunder (that was before science) or to make sense of their lives. Others say that God existed long before man was capable of thinking about

things. I leave it to you to make your mind up on that matter. I just want to point out that many mathematicians when they make up some mathematical entity that proves to be useful they get an uncanny feeling that they're not inventing but rather discovering something that was already there. In some sense complex numbers existed before the world began. If there's an advanced civilization on some far distant planet they will also have 'invented' complex numbers in one form or another.

§1.3. Mathematics Contains No Facts

Mathematics is the subject *par excellence* when it comes to logical foundations. Yet in another sense mathematics isn't about truth at all – certainly not in an absolute sense. When we prove that the angles of a triangle add up to 180° we think we've proved an absolute truth about the real world. Not so!

Mathematics is not about absolute truth (if there is such a thing) but rather about relative truth. Everything in mathematics is based on definitions and fundamental assumptions. Take the case of the angle sum theorem we've just mentioned. Quite apart from needing to define angles and triangles and addition we must accept the axioms on which the proof of the theorem is based.

Euclid began by setting down some basic axioms, or assumptions. Some of these were attempts

at definitions, others were considered as “self-evident truths”. Clearly, through any two distinct points there’s exactly one straight line. You don’t need to prove it – you can see that it’s obvious. If anyone is so obtuse as to say they don’t agree with it you simply have to ask them to put two points on a piece of paper and draw two *different* straight lines between them.

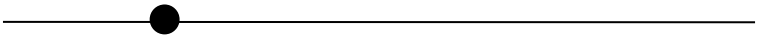
But our grounds for accepting this axiom are rather shaky. We’re arguing as a physicist might. We carry out many experiments with points and lines on a sheet of paper and are never able to construct two straight lines between the same two points. I’m not belittling the scientific method, but if we allow it to operate within geometry we may just as well go off and measure lots of triangles and conclude the truth of the angle sum theorem by experiment.

We might argue that light travels in straight lines so a ray of light that begins at point A and is seen at point B must have gone along a single path. Imagine if the light had to make up its mind as to which straight line to follow!

Unfortunately we now know that light doesn’t always travel exactly in straight lines. The more gravitation is around the more curved the path. And as

for light not being able to make up its mind as to which path to follow, even stranger things have been observed in the laboratory since the advent of quantum physics.

Another fundamental ‘truth’ is that if we have a straight line, and a point off that line, there’s exactly one line passing through the point parallel to the given line. It is on the basis of this that the Angle Sum Theorem is proved.



Now experimental evidence for this ‘fact’ is very strong. But remember that drawing lines on a piece of paper is neither particularly accurate nor particularly general. Perhaps two lines can be drawn, so close to each other that you’d only notice the difference if they were drawn with considerable precision and the sheet of paper was many light years across. Indeed there’s speculation that the geometry of space doesn’t quite follow Euclid’s axioms.

How do mathematicians cope with all this? Scrap thousands of years of Euclidean geometry? Not at all! “There’s nothing wrong with Euclidean geometry,” they say. “If the axioms hold then so do the theorems. It’s the job of the cosmologist to decide

whether the axioms are true for our universe, not the mathematician!"

What mathematicians did do, when it was discovered that this 'fact' didn't follow logically from the other axioms, was to develop non-Euclidean geometries where there can be more than one line through two distinct points, or none at all. So by the time physicists began to doubt whether our universe followed Euclidean geometry there was a mature study of non-Euclidean geometry for them to choose as an alternative.

§1.4. The Disembodied Angel

Mathematics isn't about absolute truth. Mathematicians create stories about imaginary systems. Each one is logically consistent but it's up to the physicist, or economist, to select one off the shelf to fit their observations.

If the universe were to disappear tonight, physics and chemistry would be no more. Biology and psychology would disappear, not to mention economics. Of all branches of learning only mathematics (and perhaps theology) would remain! It's a nice thought, though whether logic exists outside the hard-wiring of the human brain is yet another question. But certainly a mathematical truth shouldn't be dependent on physical observation. After all we must think of the disembodied angels!

Years ago one of my colleagues, Alan Macintosh, had to teach an advanced class in geometry. To emphasise the fact that geometry can be studied without recourse to spatial intuition he had a pair of walkie-talkies (these day's he'd use mobile phones). An accomplice was positioned in the next classroom with one handset and Professor Macintosh had the other. A student from the class was chosen and was given the job of explaining some geometrical concept to the 'disembodied angel' in the other room. The idea was that the 'angel' was infinitely intelligent but had no concept of space, living as she did in a purely spiritual realm. The results were quite amusing.

Mathematics has reached the level of maturity that it can now be taught to disembodied angels! That is, when studying it at an advanced level, students are required to empty themselves of all their intuition concerning number, space and even sets. The fundament objects of study are to be considered as undefined entities. We have to capture our intuition by writing down our assumptions as axioms. They might be self-evident to us but not to disembodied angels. Both they and we accept these axioms and proceed from there. At no time in the proofs of our theorems must we fall back on intuition. Everything must proceed using the tools of logic.

That's not to say that intuition has no place in modern mathematics. Mathematicians are not machines that manipulate symbols mindlessly to create

theorems. There's an old joke that mathematicians are machines for turning coffee into theorems but this perhaps reflects the fact that coffee may help to stimulate a mathematician's intuition. What happens is an interaction between intuition and logic. An intuitive insight causes a mathematician to 'see' that such and such must be true. He, or she (women are now quite active in the world of research mathematics), will then set out to prove the fact, using sound logic. Sometimes they will fail, but their efforts will help them to see that they were wrong. More often than not they will be able to prove that they were right. Either way the struggle towards a proof will strengthen their intuition.

For some laymen, the phrase 'mathematical research' is an oxymoron. I am often asked "Hasn't everything in mathematics been discovered a long time ago?" Well, mathematics may not be quite the oldest profession but it comes close. It's probably the oldest academic profession.

Mathematics has been building for thousands of years. And because it's highly structured you can only understand the more recent bits once you understand the earlier bits. So all of the mathematics you learnt at school, even at the most advanced level, would be a couple of hundred years old. If you continued on to university mathematics you might be brought up to the end of the nineteenth century, with a couple of exceptions, though you'd still learn only a tiny fraction of what was known to that point.

But mathematics has been producing new theorems, even whole new branches, at an ever increasing rate. Some years ago, before reviews of mathematical papers went online, *Mathematics Reviews* was putting out monthly volumes, each the size of a small telephone book, that contained short summaries of the more important mathematics papers that had been published that month in research journals around the world.

§1.5. Propositional Logic

So mathematics is founded on logic and uses its tools to create proofs. What are the tools of logic? To begin with we consider things called ‘statements’ or ‘propositions’. We could regard a statement as an undefined entity but it helps to think of it as a sentence for which it is meaningful to say that it is true or that it is false. Statements have things called ‘truth values’ – TRUE and FALSE.

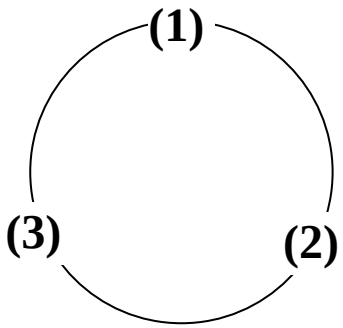
Not every sentence is TRUE or FALSE. “Come here!” is a command, not a statement. But even things that look like statements might just be pretending. Consider the sentence: “THIS STATEMENT IS FALSE”.

If it’s TRUE then it’s FALSE and if it’s FALSE it must be TRUE. So it can be neither TRUE nor FALSE. This doesn’t invalidate logic – it simply means

that “THIS STATEMENT IS FALSE” isn’t really a statement.

The reason why “THIS STATEMENT IS FALSE” can’t be considered a statement is because it is *self-referential* – that is it refers to itself. A similar situation exists if we have a collection of statements that refer to other statements, but where the references go around in a circle such as:

(1) Statement (2) is FALSE
(2) Statement (3) is FALSE
(3) Statement (1) is FALSE



If any of these is TRUE, the next, around the circle is FALSE.
If any of these is FALSE, the next, around the circle is TRUE.

So the TRUE and FALSE tags must alternate, which around a circle with an odd number of points is impossible.

You might think that the recipe to avoid such paradoxes is to insist that no statement is allowed to refer to itself either directly or indirectly. But there are instances where there’s no hint of self-referentiality

and where we still get such a paradox. We'll see such a system in a later chapter. This shows that it's very difficult to define what is meant by a statement.

We represent statements by lower case letters *p*, *q*, … It's just like algebra except that the symbols stand for statements instead of numbers. The only property of a statement that logic can deal with is its truth value. Whether the statement is long-winded or amusing, or contains a certain four letter word, lies outside the realm of logic. So you can think of the variables *p*, *q*, *r*, … as undefined entities having one of two possible values T or F (shorthand for TRUE and FALSE).

Logic thrives on constructing complex statements from simple ones, and then asking whether the complex statement is true or false. It does this using 'logical operators'.

The basic one is '**not**'. When we say '*not p*' we mean the statement '*p* is FALSE'. So if *p* is TRUE then *not p* is FALSE but if *p* is FALSE then *not p* is TRUE.

For example, if *p* = "2 + 2 = 4" then

not p = "2 + 2 ≠ 4".

Here *p* is TRUE while *not p* is FALSE.

If *q* = "3 > 9" then *not q* = "3 ≤ 9".

Here *q* is FALSE and *not q* is TRUE.

If g = "God exists" then not g = "God does not exist". In this case you must decide for yourself whether p or not p is TRUE. We can't have both being TRUE and our logic insists that at least one of them is TRUE. (There are other logics that logicians study where statements may be neither, but mathematicians are usually intuitive about our logic.)

Two statements p and q can be combined in several ways: "**and**", "**or**" or "**implies**".

The complex statement p and q means what it says – we assert that both of them are TRUE. We can express this to a disembodied angel by means of a table that sets out the truth value of p and q under all four combinations of the truth values of p and q separately. Such an angel doesn't need to have any concept of what 'TRUE' means.

p and q		
↓p q→	T	F
T	T	F
F	F	F

Sometimes we say p but not q. We might say "mathematics is interesting but economics is not". Here 'but' just means 'and', at least at the basic level of logic. There may be overtones of surprise or contrast but such subtlety is beyond basic logic.

So if p is "mathematics is interesting" and q = "economics is interesting" then what we're saying can be encapsulated in symbols as p and not q.

Or we might say p or q. Here our intuitive grasp of the word 'or' can more or less define what we mean. But there's some ambiguity. There is the *exclusive* 'or' and the *inclusive* 'or'.

At a party, if we're offered a glass of wine, and are asked whether we want red or white, our host would be quite taken aback if we said "both". Here the word 'or' is used in a polite sense, that is, it means the *exclusive* 'or'. But mathematicians are impolite. We reserve the right to say "both" – perhaps not in a social situation but in our mathematics. When we say "$x = 0$ or $y = 0$" we include three possibilities:

(1) x is zero but y is not,
(2) y is zero but x is not,
(3) they are both zero.

Of course there are situations when we need to be exclusive, but then we'd have to spell it out: p or q and not(p and q).

For the benefit of the disembodied angel we should simply set out our meaning in a table.

p or q		
↓p q→	T	F
T	T	T
F	T	F

The third logical operator is rather more confusing: "if *p* then *q*". We call this 'implication' but we don't mean to imply any causal connection between the two – simply a connection between their truth values.

Let's see how far our intuition might go to defining implication. In the case where *p* is TRUE and *q* is TRUE then of course we want "if *p* then *q*" to be TRUE. True statements imply true statements.

And in the case where *p* is TRUE and *q* is FALSE we want "if *p* then *q*" to be FALSE. True statements don't imply false ones.

What do FALSE statements imply? We may be tempted to say "nothing". In other words we may think we want "if *p* then *q*" to be FALSE whenever *p* is FALSE. But do we? Look at the table that would result from that decision.

↓p q→	T	F
T	T	F
F	F	F

Our disembodied angel would say, "This is the same table that you gave me for 'and'. Do you mean that 'if *p* then *q*' is just a complicated way of saying '*p* and *q*'?"

Rather than try to tease anything more out of our intuition we'll simply present the correct table and be done with it. As Humpty Dumpty said in *Alice's Adventures in Wonderland*, "When I use a word it means just what I choose it to mean." Just accept that in mathematics "if ... then" means this.

if p then q		
↓p q→	T	F
T	T	F
F	T	T

But surely it's wrong, surely, to have a FALSE statement implying anything! There's a technical explanation that doesn't become apparent until we meet quantifiers. At this stage just pretend that you're a disembodied angel and simply accept the table.

§1.6. Quantifier Logic

Here we move up to the next level of logic. It's going to involve some symbols so perhaps you're ready to skip to the end of the chapter. By all means, if you've got symbol phobia, then do just that. But let me encourage you to persist. Just remember that symbols are just short words and have to be read more slowly than most text.

In a short while we'll encounter the sentence

$$x^2 - y^2 = (x + y)(x - y).$$

Perhaps you might call it an "equation" but equations are really sentences. They have verbs and nouns. The verb in an equation is the symbol "=" that is shorthand for "equals". The nouns in the above sentence are x and y.

Shortly we'll we introducing the strange symbol '$\forall$'. Roughly speaking it's shorthand for 'all'. We could write $\forall$cats, meaning 'all cats'. So a complete sentence could be: $\forall$cats are cute.

More precisely, $\forall$ is shorthand for 'for all'. What follows is a variable, so '$\forall x$' stands for 'for all x'. Now on its own this doesn't make sense. It must be followed by a statement about x. So $\forall x$[x is cute] asserts that all x's are cute. But not everything is cute. We need to limit the x's to come from some 'universe of quantification'. We don't normally incorporate this into our notation but we have to have some sort of context that implies what this is. For example if we were discussing arithmetic we would be assuming that x is a number.

Perhaps you don't like symbols and would prefer to have everything spelt out in words. For example we *could* avoid using symbols and write the statement:

$$\forall x \forall y[(x^2 - y^2 = (x + y)(x - y))$$

as ‘if you take any two numbers and subtract the second number squared from the first number squared you’ll always get the same answer as if you had added the sum and difference of the original two numbers and then multiplied the sum by the difference.” Do you really think that this makes it any easier to understand? Symbols are used in mathematics not to scare away the uninitiated but to make life easier.

The quantifier $\forall$ is called the ‘universal quantifier’. The other quantifier is the ‘existential quantifier’. It is written $\exists$ and is shorthand for ‘for some’. So $\exists x$ means ‘for some x’.

If the universe of quantification consists of all human beings we might write $\exists x$[x can run a mile in under 4 minutes] to mean that some human can run a four-minute mile]. In fact there are quite a few possible x’s – the x does not have to be unique.

In todays world this is a TRUE statement, but in the early part of the twentieth century it was FALSE. Of course the truth of a statement shouldn’t be time-dependent. We should be more precise in defining the universe of quantification. If it is the set of all humans alive in 1920 the statement is FALSE, but if it is the set of all humans alive in 2020 it is TRUE.

We could also consider the statement:
$\exists x$[x can run a mile in under 3 minutes].

We could also consider the statement:

$\exists x$[x can run a mile in under 3 minutes].

If the universe of quantification is the set of all humans who were alive in 2020 this would be FALSE. If it was the set of all humans who have ever lived or who will ever live in the future we can't decide whether this is TRUE or FALSE.

If the universe was the set of all mammals alive today then it is most certainly TRUE as there are animals in the cat family wo can run a mile in well under two minutes.

We might write $\exists x[x + 2 = 0]$, meaning, if our universe of quantification is the universe of all numbers, that there is some number which when added to 2 gives zero. There is only one such number, namely -2.

You might be scared of these two strange symbols that are used to represent quantifiers. But quantifiers themselves are things you use in everyday speech. It's just that you probably don't know the technical jargon for them, or the symbols that represent them.

Without "quantifiers" there would be no mathematics. Come to think of it, without quantifiers

our everyday conversation would be at the level of a caveman's grunt.

"Children of today don't know what hardship is". Here we're not referring to a particular child but to children in general. "Someone's taken my icecream out of the fridge!" There you have the two types of quantifier. You use them all the time!

In mathematics we mostly make general statements involving variables. It wouldn't be edifying to come across a theorem that said "3456 + 9876 = 9876 + 6543". It's not a theorem we'd use very often! On the other hand there is a theorem that says:

$$x + y = y + x.$$

Now notice that here we have two variables x and y. What's implied by this is that we can substitute any number for x and any number for y and we get a true result.

$$5^2 - 4^2 = 25 - 16 = 9 = 9 \times 1 = (5 + 4)(5 - 4).$$
$$9^2 - 7^2 = 81 - 49 = 32 = 16 \times 2 = (9 + 7)(9 - 7).$$
$$10^2 - 1^2 = 100 - 1 = 99 = 11 \times 9 = (10 + 1)(10 - 1).$$
$$3^2 - 5^2 = 9 - 25 = -16 = 8 \times (-2) = (3 + 5)(3 - 5).$$

Such a theorem is much more useful than a specific one without any variables. It represents infinitely many individual true statements all at once.

To make it clear that we mean for *x* and *y* to represent any number we can use the **universal quantifier** and write:

$$\forall x \forall y[x + y = y + x].$$

But mathematicians get lazy and often leave out universal quantifiers. Any free variable in a statement is assumed to be bound by the universal quantifier.

Things get interesting when we get mixtures of quantifiers. Suppose *x*L*y* represents some statement involving two variables. Is there any difference between $\forall x \exists y[x\mathrm{L}y]$ and $\exists y \exists x[x\mathrm{L}y]$? Can quantifiers be swapped around just like numbers in a multiplication problem?

Well, suppose that L represents "loves" so that *x*L*y* means '*x* loves *y*'. If *x* = your mother and *y* = you, then hopefully *x*L*y* is TRUE.

Now we need to specify our universe of quantification. Let's make it as general as possible as consisting of all beings, alive or dead, who are capable of loving. Consider what statements you can get by putting quantifiers in front of *x*L*y*.

To say $\forall x \forall y[x\mathrm{L}y]$ would be to make the rather optimistic claim that everybody loves everybody. At the other extreme is the cynical claim that

$\exists x \exists y[x L y]$.Somewhere, somebody loves somebody. The world is not totally devoid of love.

What about mixtures of universal and existential quantifiers?

$\forall y \exists x[x L y]$ says that for all y there is somebody *x* who loves them. Nobody is unloved. I'm sure you would agree that is, rather weak, statement is TRUE.

But swap the quantifiers around and we get:

$$\exists x \forall y[x L y]$$

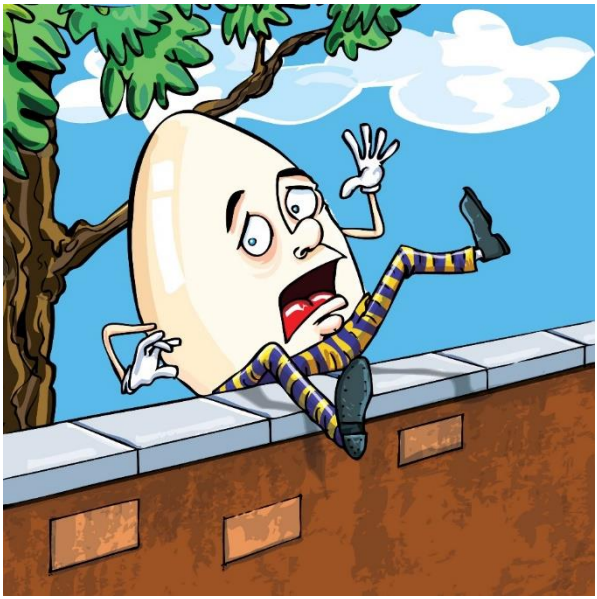
This is almost the theological statement that God exists, for it asserts that there exists a being who loves everybody. You may believe that this is FALSE, but even if, like me, you believe that it is TRUE you must admit that it's a much stronger statement than $\forall y \exists x[x L y]$.

It's amazing! You made it to the end of the chapter! It wasn't quite so readable as *Alice's Adventures in Wonderland* was it? The difference is that Lewis Carroll was content with throwing fragments of logic around in his delightful story. My aim is more ambitious. I want to take you on a *real* mathematical journey. The stories and poems between the chapters are just resting places.

And I hope you weren't put off by all the symbolic expressions. The secret is not to read mathematics as you'd read a novel. When you come to a symbolic expression you need to slow down and examine it symbol by symbol. It's a well-known fact that when reading English prose your eye can easily ignore a spelling error. You read whole words and if the word is mispelt you may not even notice. (I bet you didn't even notice that "misspelt" was missing an "s".)

Anyway, your brain now needs a rest. That's the reason for the following Humpty Dumpty poem.

A POEM: HUMPTY DUMPTY



There existed an egg who sat on a wall,
And the wall being short implies this story is tall.
Now if that fat egg had had a great fall
Or slipped off the top, but not jumped, then not all
The king's horses and all the king's men,
If they worked through the day and the evening, then
They could not succeed if and only if when
They attempted to put Humpty together again.

2. THE IMPOSSIBLE

§2.1. Nothing is Impossible!?

It's impossible! It can never be done! Dangerous words! How often has the short-sightedness of man placed limits on what can be achieved?

Man will never fly in a heavier-than-air machine and certainly will never stand on the moon. Total 'impossibilities' yet we've seen them come about. Computers will never be able to play a game of chess to grand-master standard. Yet it has happened.

But of course some things are eternally impossible. As children we grappled with the idea of the impossible.

"Bet there's nothing God can't do."

"Bet there is."

"What, then? Bet you three marbles you can't think of something God can't do."

"He can't make 2 plus 2 make 5."

"Yeah, but that's not possible. I mean God can do anything that's *logically* possible."

"So he could lift up the world?"

"Sure, he could even lift up the sun with his little finger!"

“Well then, is he able to find something so big he can’t lift it?”

“But that’s impossible.”

“No it’s not. I can easily find something so big that I can’t lift it, so why can’t God?”

Now of course there’s nothing impossible about something being logically impossible. We can all make up problems that have no solutions. And if a problem is impossible it’s important to know that, otherwise we can waste a lot of time.

This book is about a lot of impossible things. The more important of them have helped to delineate the boundaries of rational thought. Because they involve things close to the limits of human reasoning we may from time to time look over the fence into philosophy, but our feet will stay firmly on the side of logical thought as we dabble in the mathematics at the edge of the rational universe.

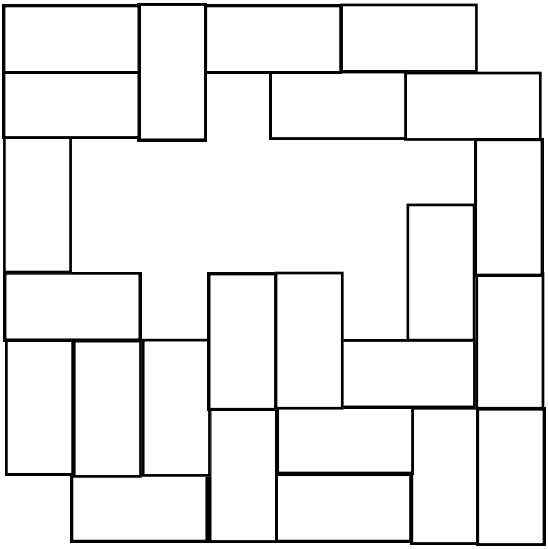
§2.2. The Domino Puzzle

How are we able to prove that a problem has no solution? The most obvious way is to check through, and eliminate, every possibility. But how can we prove that something is impossible if there is an enormous number of possibilities? Get a computer to

do the checking? But what if there are *infinitely* many possibilities?

The following puzzle involves a huge, but finite, number of possibilities, though it could easily be adapted to one with an infinite number.

Take 31 dominos and place them on a chessboard (each one covering two adjacent squares) so that the two squares that remain uncovered are at diagonally opposite corners. At first sight the following appears to be well on the way to a solution but closer inspection reveals that it can't be completed.



That doesn't prove it *is* impossible of course. It may simply mean we started wrongly. But what an

enormous number of possible ways of starting we'd need to check!

We might try unsuccessfully for quite some time and declare in disgust that "it's impossible" but the nagging thought would remain, "maybe just one more try will do it".

Yet it is indeed impossible. You can take that as a challenge if you wish, but you're really wasting your time. It's known to be impossible, not because many have tried and failed and not because a computer has worked through every conceivable possibility. It has been *proved* to be impossible. And the proof involves a clever but exceedingly simple idea.

Use a chessboard pattern of black and white squares. Each domino must, of course, cover one black square and one white one. The 31 dominos therefore cover the same number of black and white squares and so the two remaining squares must be of *opposite* colours. But diagonally opposite corners of a chessboard, as every chess player knows, have the *same* colour. A contradiction is reached if a solution were to exist. So of course no solution could possibly exist.

§2.3. Proof By Contradiction

Not every proof of impossibility is as short and transparent as this one. But they all rely on the simple idea that any assumption that leads to a contradiction must be false.

We begin by assuming that whatever we're trying to prove impossible is in fact *possible*. We then attempt to use logic to reach a contradiction, that is, something which is both true and false. If we succeed in producing this nonsense we know that our assumption of possibility must be false and we'll have proven impossibility.

Some people get worried about the validity of this type of reasoning. "You can't make assumptions in proofs." It's true that if you're allowed to assume that what you're trying to prove true *is* true, then naturally you'll succeed all the time, no matter what you're trying to prove.

Assume that the moon is made of green cheese. Therefore if you land on the moon and dig up a sample, it will be green in colour, and will have a strong cheesy flavour. Therefore the sample will be green cheese. And if you repeat this experiment at numerous other locations on the moon you must get the same result. Hence the moon *is* made of green cheese! But of course that proves nothing. We might get away with such a fallacy if our chain of arguments is so long that our

listener forgets what we'd assumed in the first place. But fallacious reasoning it is, nevertheless.

So of course it *is* fallacious to assume what you're trying to prove. But that's not what we're doing in a proof by contradiction. In such a proof we're assuming that what we're trying to prove is *false*, or that the so-called impossible is in fact possible. And that's a totally different thing.

Proof by contradiction is not some esoteric rule thought up by logicians or mathematicians. It's just ordinary common sense that we use all the time. "You couldn't have put the milk away because it's still on the bench." Analysing the logic behind this assertion we find that it's a proof by contradiction.

Theorem: You didn't put the milk away.
Proof: Suppose that you *did* put the milk away.
Then the milk is in the refrigerator.
[Here there's the unspoken assumption that no-one else has been around to take it out again.]
But the milk is still on the bench and so is not in the fridge.
[That we are talking about the same bottle is another unspoken assumption.]
Contradiction! Therefore you did not put the milk away.

You, the accused, might still dispute this argument. Nothing in life is quite as clear-cut as in mathematics. But the only way you could validly attack it would be to draw out and dispute one or other of these unspoken assumptions. The underlying logic of the argument itself is perfectly sound.

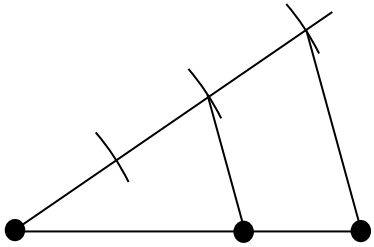
Impossibilities are everywhere, not just at the edges of rational thought. Before we journey out to the uttermost parts of the rational universe we'll look at a number of other perfectly ordinary impossibilities. Some are quite famous in the history of mathematics. Others are mere curiosities. Our purpose in examining them is to help us feel quite at home with proof by contradiction because that's the tool we'll need on our journey.

§2.4. The Square Root of 2 is Irrational

An irrational number isn't one which is crazy. It simply means one which cannot be expressed exactly as a *ratio* of two whole numbers, like 2/3 or 22/7. Ratios have a geometric significance. The Greeks were able to divide any given line in any ratio of two whole numbers.

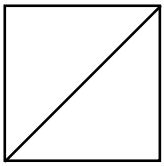
For example to find the point which is two thirds of the way along a given line segment, you construct a second line from one end of the first and mark off three equal lengths (with a compass, of course — using rulers to measure was considered unacceptable).

The third point is joined to the other end of the original line segment and other lines are drawn parallel to it as in the following diagram. If you remember from school how to draw parallel lines using ruler and compass, well and good. If you've forgotten, it doesn't matter. There *is* a way.



This can be easily adapted to construct a line segment that is any rational multiple of the one given.

But the Greeks soon learnt of a theorem that's associated with the name Pythagoras. The square on the hypotenuse is equal to the sum of the squares on the other two sides. Construct a square (which the Greeks could do easily, using their rulers and compasses) and by this famous theorem the length of the diagonal of the square must be $\sqrt{2}$ times the length of the side. (Diagonal squared = $1^2 + 1^2 = 2$.)



Since the only numbers that existed at that time, or rather the only numbers that had been invented, were rational numbers, it was obvious that $\sqrt{2}$ had to be the ratio of two integers. It simply remained to find the two integers.

The rational number 10/7, when squared, gives 2.04..., which is close to, but not exactly equal to, 2. A better approximation is given by 1393/985. Its square is 1.9999989... very much closer. Try 8119/5741.

Then came the embarrassing truth. They discovered an argument that demonstrated the impossibility of finding such integers. No rational number gives exactly 2 when squared. So here was a line which had no length! That can't be! To get around this difficulty, new numbers had to be invented.

Theorem: $\sqrt{2}$ is irrational
Proof: Suppose that, on the contrary, it is rational. (Here's a classic Proof By Contradiction.)
Let m and n be the two whole numbers whose ratio, when squared, gives exactly 2.
Then $\left(\frac{m}{n}\right)^2 = 2$.
But this means that $\frac{m^2}{n^2} = 2$ and so $m^2 = 2n^2$, that is, m^2 must be exactly twice as big as n^2.

Now consider the number of factors of 2 which divide these numbers. However many factors of 2 there

are that divide n, clearly exactly double that number divide n^2. In fact the number of factors of 2 in any perfect square, n^2 or m^2, must be even. But that means an odd number of factors of 2 divides $2n^2$ and an even number dividing m^2. This can't happen if they're equal!

This contradiction rests firmly on a single assumption – that $\sqrt{2}$ is rational. This assumption cannot stand. The square root of 2 must be irrational.

Now a professional mathematician might argue that this proof relies on the Unique Factorisation Theorem for whole numbers: There is essentially only one way of factorising a whole number and the number of factors of 2 will always be the same. He will argue that there is a more basic proof that avoids the need to assume the Unique Factorisation Theorem.

However I have found that most people find it harder to follow than the above proof. Remember that I'm not trying to develop number theory but rather to explain the concept of Proof By Contradiction.

§2.5. It is Impossible to Trisect a Given Angle by Ruler and Compass

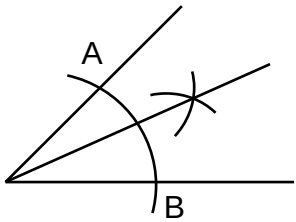
One of the famous classical impossibilities concerns ruler and compass constructions. This type of geometric construction was a highly developed art form in the time of the ancient Greeks because for them, arithmetic was built on the foundation of

geometry. Ruler and compass construction was as important a tool then as the calculator is today.

The ruler wasn't used to measure lengths. In fact any straight edge would do. What the Greeks had against measurement was that it wasn't exact. No matter how fine the divisions, a length may fall between two of them and the human eye is called upon to estimate.

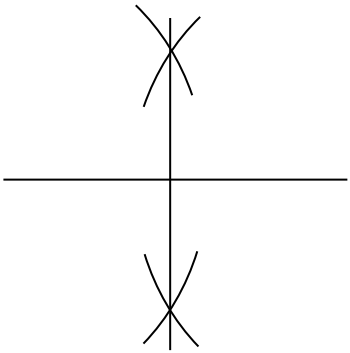
The Greeks were intoxicated by perfection. Any method had to be theoretically exact. And this they could achieve with straight-edge and compass – at least for many problems.

They could, for example, bisect any given angle. Most school pupils learn how to do this. With the compass point on the vertex of the angle, draw an arc cutting the two arms of the arc at points A, B. Now with any convenient radius (but the same for each) draw intersecting arcs, one with A as centre and one with centre at B. Joining the intersection of these arcs to the vertex of the angle exactly bisects the angle.



The method is mathematically exact. Using theorems of congruent triangles one can prove that the two angles created at the vertex are equal, each exactly half the original. Of course to do it in practice, no matter how carefully you carry out the construction, all sorts of little errors creep in. But the method is *mathematically* exact.

Lines of any given length can be bisected by a similar construction. Here the centres of the arcs are the endpoints of the line.



What was really tantalising was that although lines can be trisected (3 equal pieces) there appeared to be no method for trisecting *angles*. This really disturbed them because it was obvious to them that it could be done. Why should there be any difference? Aren't lengths and angles just different geometric manifestations of the same numbers?

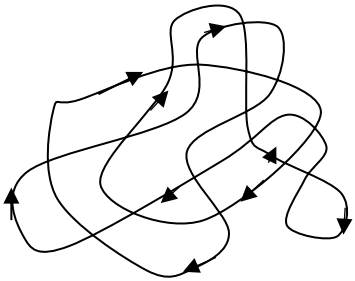
Much effort went into looking for such a construction without success. All was wasted effort. Many, many centuries passed before a proof that such a construction is impossible was discovered. It's too technical to present here, but it's worth pointing out that, like many proofs of impossibility, the breakthrough came by cleverly converting the problem to one involving whole numbers.

In the trisection case, there's a number which can be associated with any ruler and compass construction called the "degree of the corresponding field extension". Never mind what that means. Suffice to point out that it starts at 1 and with each stage in a ruler and compass construction it either remains the same or it doubles. So only exact powers of 2 are possible: 1, 2, 4, 8, 16, ...

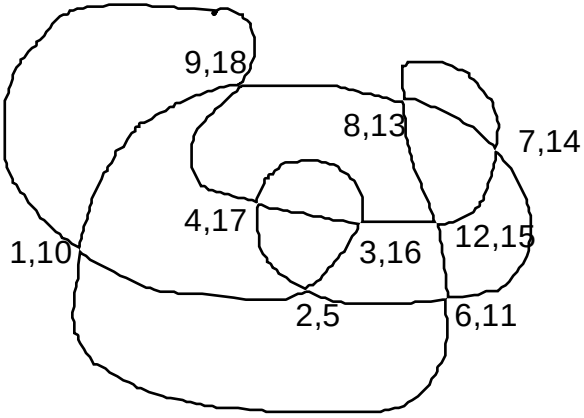
But it can be shown that a method which trisects a 60 degree angle, must be capable of producing a field extension whose degree is exactly 3. Clearly 3 is not a power of 2 and so we get a contradiction if we assume that angle trisection is always possible.

§2.6. Scribbles

Draw a **scribble**. By this I mean a continuous line which crosses itself many times and ends up where it starts. Oh, and you are not allowed to pass through a previous crossing.



You'll have a number of crossings where two parts of the scribble cross over. Start at any crossing you like and number them in order: 1,2,3, ... When you visit a crossing you must give it a second number. Continue until all crossings have been given two numbers.



Now a **double crossing** is one where one of its two numbers is double the other and a **triple crossing** is one where one of its numbers is three times the other,. There's no difficulty in producing double

crossings. This scribble has two of them: 7, 14 and 9, 18. But there are no triple crossings.

Can we create a scribble which includes at least one triple crossing? It might have to be an exceedingly complicated scribble with millions of crossings, one of which might be a crossing labelled as (123123, 369369).

The problem can't be solved. There is no solution. You might like to try to find one just to get the 'feel' of it, but don't try too hard because the puzzle is really quite impossible. But how can we be sure of this? After all there's no limit to the complexity of the scribble so it's just not possible to check all cases.

§2.7. Why No Triple Crossings?

Suppose it *can* be done. (Notice that we've started in the usual way for a Proof by Contradiction.) Then we'd have a (k, $3k$) crossing somewhere. We visit when the count is k and revisit when the count has reached $3k$.

Question: How many times will we pass a crossing between these two visits?

Answer: An odd number of times.

This might seem wrong because the difference between k and $3k$ is $2k$ which is even. But think again.

How many numbers are there between 4 and 12? Do you think there are 8, because 12 – 4 = 8? No, there are only 7 numbers between 4 and 12. They are 5, 6, 7, 8, 9, 10 and 11. When we subtract 4 from 12 and get 8 we're counting the one-unit sections between 4 and 8. But the number of dividing points is never equal to the number of sections. It's always one more if we're counting both ends and or one less if we're not (as in this case).

	4	5	6	7	8	9	10	11
sections	1	2	3	4	5	6	7	
intervening numbers	1	2	3	4	5	6		

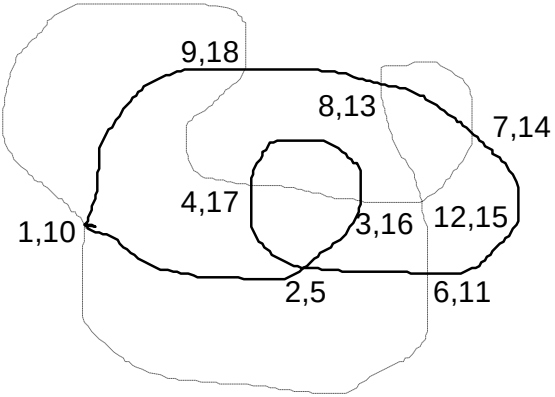
Always, between the two visits to a (k, $3k$) triple crossing, there are $2k$ sections and so $2k - 1$ visits to other crossings. So the answer to the question is "an odd number of times".

Second Answer: An even number of times.

Before you start pointing out that this contradicts what we concluded earlier, consider the supporting argument.

If you start at any crossing in a scribble and move around till you revisit that crossing you'll have traced out a smaller scribble. Those parts you haven't traced will also be a smaller scribble. What you'll have

done is to decompose the original scribble into two simpler ones, linked at the crossing you started with.



You can think of one of these smaller scribbles as being the boundary of a region and the other scribble as being a closed path ('closed' here just means that it ends where it starts) which cuts across the first scribble in a number of places. Now because of the principle that "what goes in must come out" (this must hold because the scribbles don't have any free ends), the two smaller scribbles must cut each other in an *even* number of places.

So when you go from a (k, $3k$) triple crossing at visit k, until you revisit it at visit $3k$, you'll have passed through an *even* number of crossings.

But wait a minute, we've overlooked places where the scribble that's traced out on this journey may cross itself. As well as the even number of places where the two scribbles cross we must add the places where

the first scribble crosses itself. And couldn't that be an odd number?

Yes, of course. In the above example the solid line and the dotted line are the two smaller scribbles that together make up the whole scribble. Now you can see that the solid line cuts itself just once. So doesn't that destroy the evenness of the number of intervening crossings?

Not at all. Remember, we're not counting crossings but *visits*. So, when the scribble we're following crosses itself, that counts as *two* visits. Including these self-crossings merely adds an even number to an already even number.

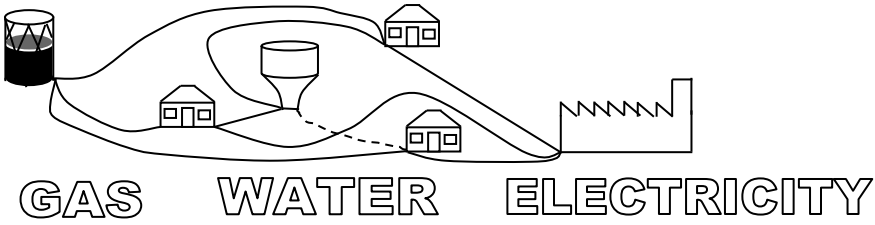
So for these reasons the number of visits to crossings between the first and second visits to our mythical triple crossing is even.

The fact that we previously convinced ourselves that this number is odd, and the fact that no number can be simultaneously odd and even, completes the argument. If a triple crossing were to exist then we'd have a contradiction, and once even a single contradiction is allowed to creep in, others follow: odd = even , true = false, black = white and the whole edifice of knowledge crumbles to dust.

You may be getting the impression that the difference between odd and even is at the heart of every proof of impossibility. This is certainly true in many cases. However we'll now see a few examples where the impossibility uses other methods.

§2.8. The Utilities Puzzle

Imagine that you have three houses, each of which has to be connected to the three utilities of gas, water and electricity. Now the catch is this. Pipes and wires are not allowed to cross over one another. Why this should be is never properly explained in the puzzle. Perhaps the world is a sort of two-dimensional 'Flatland'.



It's very easy to get a solution that almost works, where only one pipe or wire remains to be installed. But no amount of ingenuity can come up with one that works completely. Try as you might, no matter how ingenious and how contorted you make the routes, nothing seems to work.

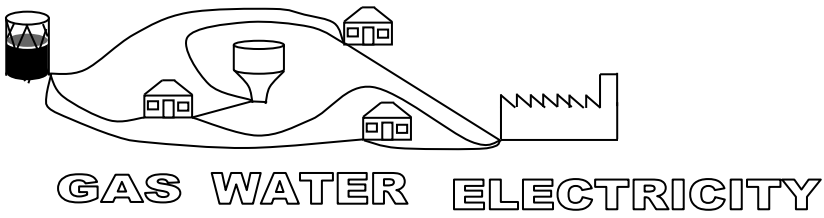
Now you might be a pragmatist and conclude, after a bit of fruitless experimentation, that it's

impossible. Or you might accept the authority of the experts and say that if nobody has managed to find a solution after all these years then it must indeed be impossible.

But if you're happy to leave it there then you don't possess mathematical curiosity. "Perhaps one day, a solution might be found." Unless the impossibility has been ruled out by a water-tight logical argument the problem would continue to tantalise mathematicians. But just such an argument *has* been found and a proof of impossibility can produce as much excitement in a mathematician as a solution would have.

The key to proving the impossibility of solving the Utilities Puzzle lies in counting. We suppose that a solution exists and count the number of points (well that's easy – there are 6 points, 3 houses plus 3 utilities), the number of connecting lines (that's easy too – there's a pipe or wire from each of the 3 houses to each of the 3 utilities, that's 9 altogether) and the number of regions enclosed by the lines.

For example if we take the above attempt at a solution and remove the incomplete pipe from the bottom house to the waterworks the number of regions, including the outside, is four.



How many regions will there be in a solution to the puzzle? We might say that there would be 5, one more than in the above because a ninth line would split one region into two. But remember that there's no way of successfully putting in a ninth line to the above picture. If there *is* a solution we'd have to start from scratch.

So how on earth can we count the number of regions until we've drawn the picture? And if the problem is impossible we can never draw the picture. Ah, but there's another, sneakier, way to do this.

You see, there's a connection between these three numbers which holds for any map on a plane surface. It is called Euler's Formula:

$$\mathbf{V + F - E = 2}$$

Here V is the number of 'vertices' (that just means points), F is the number of 'faces' (that just means regions) and E is the number of 'edges' (or connecting lines).

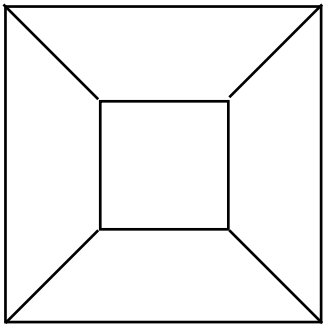
Usually this formula is quoted for solid figures, like cubes and pyramids, which are bound by a number of flat faces, and where each face is bounded by a

number of straight edges. The technical term for these solids is ‘polyhedra’. They’re the three-dimensional analogues of polygons.

A cube has 8 vertices, 6 faces and 12 edges and $8 + 6 - 12 = 2$, so it works for a cube. In fact it works for any polyhedron. But what’s this got to do with our two-dimensional problem? It’s amazing the way a mathematician is able to change the subject. Talk to him (or her) about one thing and the next thing you know he or she is talking about something completely different and apparently unrelated.

But the true mathematical approach is to draw together the original problem and the apparent ‘red herring’ and to show that they’re very much related after all.

Take a polyhedron, for example a cube, and remove one of the faces. Now stretch the rest out so that it lies flat. You might need to use your imagination for this because a cardboard cube is not sufficiently elastic. The edges of the faces may no longer be straight. That doesn’t matter. The important thing is which point is joined to what. Nothing in that department has changed; only the layout which now lies in a plane.



What we've produced is a 'map' with vertices, edges and faces (except that the faces would now be better described as regions). The numbers of vertices, faces and edges has not changed throughout this imaginative flattening. What about the face we removed to open it all up? Well that's just become the outside region of the map.

It's because $V + F - E = 2$ works for maps that it also works for polyhedra. But why does it work for maps? Well can we put that one on hold for a while. The best way to convince you is by a method called 'Mathematical Induction' and that's something we'll talk about later. Just be a good mathematical reader and accept it as fact. (Of course a really good mathematical reader will say "well, just for now, but eventually I want to know why".) But to give you enough faith to keep you going, draw a few maps and check it out. A few confirming examples is no proof, but they're comforting nevertheless! Well back to our supposed solution to the Utilities Puzzle.

We know that V = 6 and E = 9. We counted them. We have to have that number of vertices and that number of edges in any solution to the problem. Conceivably the number of faces, or regions could vary. But no. Euler's formula says that

$$F = E + 2 – V = 9 + 2 – 6 = 5.$$

So, indirectly, we can infer that any solution to the puzzle must have exactly 5 regions. Where does this get us?

The question to ask at this point is “what is the average number of edges per face?” Why this question? What led to asking that? That’s where mathematicians get really sneaky. Often it’s just a matter of asking the right question and it all falls out. So how does a mathematician develop the art of asking just the one question that will unravel a problem?

The answer is two-fold. Firstly, a mathematician, thinking about a certain problem, develops his or her intuition so that the ‘right’ question just pops out. It’s a common experience in the trade that after getting nowhere with a problem a mathematician puts it away and “sleeps on it”. Then suddenly the answer, or at least the right question which leads to the answer, comes as if from nowhere. He might be on a bus, she might be out walking. The problem is miles away. Then like a bolt from the blue, it comes.

The other explanation for why mathematicians seem to have this uncanny ability to hit on exactly the right question first time, is that they generally don't. You see, in practice a mathematician might ask dozens or hundreds of questions about the problem in hand. Scores of screwed up sheets of paper might litter the floor until finally "eureka" – the right one comes.

Now you don't think a mathematician is going to unravel all those crumpled-up pieces of paper and write up the whole investigation, false starts and all – of course not. You'd never want to read them and nor would any other mathematician. Only the right question, the right way of looking at the problem, gets into print. It appears to the reader that Euclid, Euler or Einstein just sat down one day and wrote a theorem as effortlessly as, we're told, Mozart wrote his music — flawless in first draft. All the pain and tears and sweat and sleepless nights and countless cups of coffee and conversations are hidden. All that appears is the finished product.

By the way, while we are talking about Einstein I should point out that, unlike what many people believe, he wasn't a great mathematician. He was a great theoretical physicist – probably the greatest that ever lived. And he was he had a good knowledge of mathematics, otherwise he couldn't have applied it so well. But frequently he had to ask his colleagues about some difficult mathematical technique. And he never

discovered any new mathematics. But a great theoretical physicist, yes.

Mathematicians are not the only ones who remove their scaffolding before displaying their finished edifice. But probably the process by which they achieve their results is less well understood than most.

What has all this to do with the problem in hand? Not a great deal. This was a real digression. If you remember, we had a map with 6 vertices, 9 edges and 5 regions. Well, we *supposed* we had such a map, because a solution to the Utilities Puzzle requires such a map to exist. And we were about to ask the RIGHT question. And this is ...

What is the average number of edges per face?

Easy! With 9 edges and 5 faces or regions the average number of edges per face is 9/5 = 1.8. Whoops! That's a bit on the low side! Think again!

Silly us. We forgot that each edge (boundary) separates *two* regions. Imagine that each edge is neatly sliced lengthwise into two half-edges. Now each half-edge is attached to only one region. Start again.

We have 9 edges, that’s 18 half-edges, to be shared among 5 regions. That’s 18/5 = 3.6 edges per face on

average. That's better. But maybe still a wee bit too small.

Each face has to have at least four edges. Why not 2? Well, that would mean two edges connecting the same two vertices and the puzzle specifies only one. Why not 3? Well a closed path has to alternate between utility and house. Three edges just wouldn't work.

So if 4 is the smallest number of edges surrounding any one face then the average must be at least 4? An average below 4 is just not possible. In fact it's *impossible*. Yet that impossible state of affairs is forced upon us if we assume that a solution exists. Therefore no solution can exist.

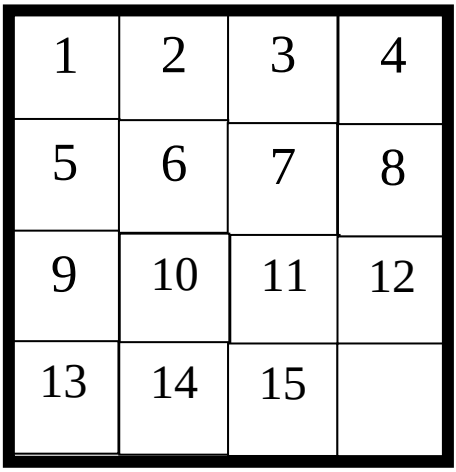
You see the infinitely many possibilities can be captured by a little piece of elementary arithmetic that in the end depended on the undeniable fact that 3.6 is less than 4. And all because we asked the right question!

§2.9. Is it possible to get "GODEL" from "LODGE"?

It would be understandable if you felt that you'd have enough impossibilities for now. By all means skip to the next chapter, taking a detour via the radio play *There Is No Time*. But if you're a glutton for algebraic punishment then read on.

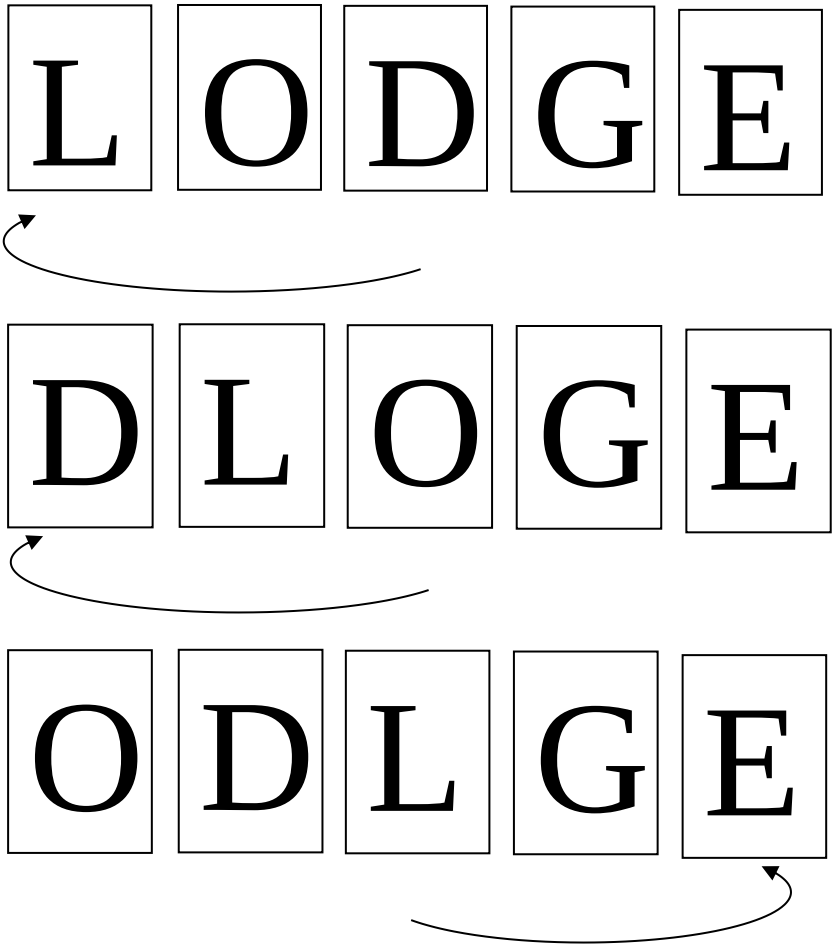
The word GODEL (the logician whose work shook the foundations of mathematics to the core in the 1930's was actually Gödel, but we'll drop the umlaut over the "o") and the word LODGE use the same letters, so a simple rearrangement will do the job of getting GODEL from LODGE.

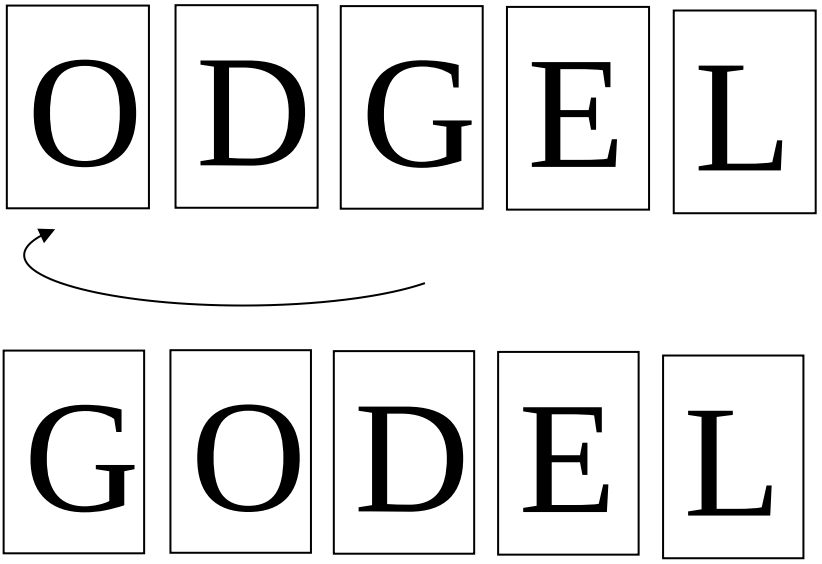
But suppose that the five letters are written on five cards and arranged in a row to spell LODGE, and suppose that a rule is imposed on how the cards are to be rearranged. Suppose that we're only allowed to move the middle card to either end, moving them up to close the gap. If that is all we're allowed to do, can we still get GODEL from LODGE?



This is one of many puzzles that involve permutations, or rearrangements. The Rubik's Cube™ is perhaps the most famous, and probably the most complicated. There is another puzzle which was in vogue many years ago, called the "Fifteen Puzzle". It consisted of fifteen small square tiles that could slide around in a 4 by 4 square frame.

With a permutation puzzle there are a number of pieces that can be moved and one or more possible moves that are permitted. In most cases the restrictions as to what rearrangements are allowed are automatically imposed by the engineering of the puzzle. In our LODGE-TO-GODEL puzzle, however, we've artificially imposed a restriction. Well, this puzzle is a pretty easy one to solve:





Not very interesting. (If you want a harder challenge, try converting LEDOG to GODEL by the same rules.)

But suppose we change the rules. Under these new rules there are three basic moves:

move	**description**	**effect on LODGE**
L	take the left card and transfer it to the right	ODGEL
R	take the right card and transfer it to the left	ELODG
V	reverse the order of the cards	EGDOL

Using these new rules, can we get from OGLED to GODEL? With a bit of experimenting we see that we can:

OGLED $\rightarrow$ ODELG (move L) $\rightarrow$ DELGO (move L) $\rightarrow$ OGLED (move V).

Now, can we get from LODGE to GODEL? The answer is "NO". It is IMPOSSIBLE. But how can we say this? There are infinitely many possible sequences of L, R and V. Have we tried them all? Of course not, yet we can *prove* that it is impossible!

To see this, first observe that we can do without R, because R is the same as doing L four times, which we write as L^4. Anything that can be achieved using L, R and V can be achieved using just L and V alone.

Next, it is obvious that doing L five times brings us back to where we were. We write the operation of doing nothing by the symbol I and so we write $L^5 = I$. Similarly $V^2 = I$. Reversing the order twice in a row gets us back to where we started.

We can write a sequence of moves as a product of powers of L and V, but we only need powers of L up to 4, and we don't need any powers of V, other than V^1 which, of course, is just V.

Suppose you came up with a recipe for transforming LODGE into DOLEG such as:

$$\mathbf{L^8V^3L^2VL^4V^5L^7}.$$

You might be tempted to collect all the L's together, and all the V's and write this is $L^{17}V^9$ and then simplify this to just L^2V using the fact that $L^5 = V^2 = I$. So the power of L can be reduced by removing blocks of 5, and the power of V can be reduced by removing pairs.

However you can't bring the L's together and the V's, like you would in ordinary algebra. These moves don't *commute*, that is, VL ≠ LV. If you start with LODGE, the move VL would turn it into GDOLE, while LV makes it LEGDO.

We can use the fact that $L^5 = V^2 = I$ to simplify our supposed solution to $\mathbf{L^3VL^2VL^4VL^2}$. Could we simplify it further?

Notice that $VL = L^4V$. Check it out. For example, VL turns LODGE into GDOLE. L^4 turns LODGE into ELODG and V then changes this to GDOLE, which is the same as VL. This means that we can move a V across L's but as $VL = L^4V$ so each L on the right of a V becomes L^4. But, since $L^5 = I$, $L^4 = L^{-1}$, so our rule for moving a V across an L is:

$$\mathbf{VL = L^{-1}V}.$$

This is called the **dihedral law** and is used in certain cases where two operations don't obey the **commutative law**.

This can be generalised to $VL^n = L^{-n}V$ by moving the L's one at a time. But instead of remembering the formula just remember the mantra:

TO MOVE A 'V' PAST A POWER OF 'L' SIMPLY INVERT THE POWER OF L.

So our supposed solution can be simplified as follows:

$L^3VL^2VL^4VL^2 \rightarrow L^3\,L^{-2}V\,VL^4VL^2 \rightarrow LV^2L^4VL^2 \rightarrow L^5VL^2 \rightarrow L^5\,L^{-2}V \rightarrow L^3V.$

In this way, any solution can be written as L^mV^n, and since we can take $m = 0, 1, 2, 3$ or 4 and $n = 0$ or 1, we only get 10 possibilities. So we only need to test those 10 possibilities.

I	L	L^2	L^3	L^4
LODGE	ODGEL	DGELO	GELOD	ELODG

V	LV	L^2V	L^3V	L^4V
EGDOL	LEGDO	OLEGD	DOLEG	GDOLE

Since GODEL is not one of these it cannot be achieved.

The area of mathematics that studies such non-commutative systems is called Group Theory. Groups were invented by a French mathematician Évariste

Galois at the age of 19. He used them as a way of solving a problem about polynomial equations.

So, in the end, the proof of impossibility came down to checking a finite number of possibilities. The breakthrough came when we realised that the infinitely many possibilities are equivalent to just 10. This is a common situation in mathematics. Something, with infinitely many possibilities, is proved to be impossible by reducing these infinitely many possibilities to a finite number which are then checked.

A classic example of this is the celebrated Four Colour Theorem. It began as a question in 1852 when Francis Guthrie, who was drawing and colouring a map of the counties of England, wondered whether four colours are enough. Guthrie had studied under the mathematician Augustus de Morgan at University College in London and his brother, Frederick was then studying mathematics under de Morgan. So Francis passed on the question to de Morgan through his brother.

De Morgan wrote: "A student of mine asked me to day to give him a reason for a fact which I did not know was a fact – and do not yet. He says that if a figure be any how divided and the compartments differently coloured so that figures with any portion of common boundary line are differently coloured – four colours may be wanted but not more – the following is

his case in which four colours are wanted. Query: cannot a necessity for five or more be invented.

Over the next few decades it became the Four Colour Conjecture. Many ‘proofs’ were published and several of them stood for a number of years before they were shown to be wrong. It took over 100 years before, in 1976, it was finally proved by Appel and Haken.

But the proof caused a lot of controversy in that it was the first theorem in history that was proved by a computer program. Of course no computer program could consider the infinitely many possible maps. What Appel and Haken did was to use standard mathematical reasoning to reduce this to 1,834 maps. If all these could be 4-coloured then every map could be 4-coloured. Here is the principal of reducing a proof of impossibility to checking a finite number of cases. However this time, the number of cases was rather large and required a computer to check them all. A computer program, laboriously, considered each of these maps and, indeed, showed that every one of them was 4-colourable.

Appel and Haken were at the University of Illinois when they published their proof and the local postal authorities were so proud of this discovery that

for many years they franked letters that passed through their hands with the words FOUR COLORS SUFFICE. Indeed they were still using this slogan in 1994 as this picture shows.

Motivational speakers often use the slogan:

NOTHING IS IMPOSSIBLE!

I hope that, as a result of reading this chapter, you will realise that it is not really true. You slogan should read:

SOME THINGS ARE IMPOSSIBLE BUT THEY ARE LESS COMMON THAN YOU MIGHT THINK!

INTERLUDE: RADIO SCRIPT

"It is Impossible – There is no Time"

Narrator: Mathematics and sport have this in common. They're both a young man's occupation. An historian reaches his peak in his sixties, an engineer at forty. A mathematician is said to be already on the decline at the age of thirty. Évariste Galois made his important discoveries in the theory of algebraic equations at the age of nineteen. At twenty he was dead.

Female Voice: Poor boy. What did he die of?

Narrator: He was killed in a duel.

Female Voice: Sounds like a character out of one of the Alexander Dumas novels.

Narrator: Almost. Alexander Dumas knew him and referred to him in one of his memoirs.

Female Voice: So Galois had to defend his mathematics with his rapier?

Narrator: Well no. For a start it was a duel fought with pistols, not swords. And secondly it was over a woman.

Female Voice: Just like a Frenchman!

Narrator: Perhaps. But in fairness I should point out that it was more than likely that she had been planted by his political adversaries to provide the excuse for a duel. It was really all to do with politics. You see, Galois had been very active in Republican politics and several times landed himself in trouble with the police. In fact much of his mathematics was done during spells in gaol. *(With feeling)* He wasn't afraid of death and he'd gladly have died for the Republican cause. But such glory was not to be.

Galois: I beg patriots and my friends to forgive me that in dying I do not die for my country. I die the victim of an infamous coquette. My life is quenched in a miserable piece of slander.

Oh, why do I have to die for such an unimportant cause; to die for something so contemptible? Farewell! It was my wish to give my life for the public good. Forgiveness to those who kill me. They are of good faith.

Narrator: These were the words he wrote to his friends on the night before the duel. He seemed quite sure that

this night would be his last. He sat up all night writing some personal letters and then going over his mathematical papers. Scrawled across one of them he wrote the pathetic words ....

Galois: *(in despair)* I have no time.

Narrator: So many of his ideas had yet to be written down and there was just not enough time. He wrote ...

Galois: I hope some people will find it to their advantage to decipher all this mess.

Narrator: The duel took place on Wednesday 30th May 1832 just outside Paris. Galois was wounded and left lying by the roadside. Even his seconds deserted him. He was eventually found by a Good Samaritan and taken to hospital. It was in vain, for the next day he died.

His mathematical discoveries however were to lie on the roadside for a further eleven years. Not by the side of the road out of Paris but by the side of the highway of mathematical research. The Good Samaritan who rescued them was Joseph Liouville who in 1843 drew Galois' work to the attention of the French Academy.

Liouville: I hope to interest the Academy in announcing that among the papers of Évariste Galois I have found a solution, as precise as it is profound, of

this beautiful problem: whether or not a given polynomial equation is soluble by radicals.

****

Narrator: What was this theory that was "beautiful" and "as precise as it is profound"? It was in fact the culmination of over two thousand years of mathematical enquiry into the theory of polynomial equations.

Most people have heard of quadratic equations. Most people vaguely remember what they are. Roughly speaking they're equations involving x^2. Maybe you also remember that there's such a thing as a quadratic equation formula. Now I'm not expecting you to remember it — simply to know that it exists. It's a formula into which you put the numbers from the equation, do some arithmetic, and out pop the answers.

The arithmetic isn't hard, but at one stage it involves finding a square root. "Radical" means the same as "root", and solving a polynomial equation by radicals simply means finding a formula for such an equation into which you plug the numbers from the equation and do some arithmetic, including finding square roots, cube roots or whatever roots may be necessary.

Now the Babylonians could do it for quadratic equations. In the sixteenth century the Italians worked out how to solve the cubic (involving powers of x up to x^3) and the quartic (powers up to x^4). These formulae

are much more complicated than the one for the quadratic but they have a similar structure.

The next step should have been the quintic (powers up to x^5). But no such formula was forthcoming for the next three centuries. Finally in 1824 a 22 year-old Norwegian mathematician, Abel, called off the search — he proved that no such formula can possibly exist.

Abel's methods, however, were not very enlightening. They worked but they didn't make one feel that one knew why they worked. The methods of Galois a few years later were much more general and much more enlightening. Moreover he took the problem a stage further.

So, Abel has shown that there is no general formula for *all* polynomials involving x^5. But there *are* formulae that work for *some* of them. Which ones? Galois worked out exactly which ones are soluble by radicals and which ones are not.

Now make sure you understand what is being claimed. Not that some polynomial equations have no solutions. Solutions can be proved to exist, even if we can't find them. Not even that we can't find the solutions for practical purposes. There are methods, implemented by computers, which can find any solution to any degree of accuracy. It's a question of which polynomials can be solved *exactly*, by means of a *formula* involving radicals or roots.

Galois showed that corresponding to every polynomial equation is something called a group. And one can tell from the structure of this group whether or not the polynomial is soluble by radicals.

Now I think that rather than give you a formal, precise and technical definition of a group it would be better if I gave a broad and vague description, and then a specific example.

A group is a certain type of mathematical system where the things in it can be combined like multiplication. But the things needn't be numbers and the method of combination needn't be ordinary multiplication.

Pretty vague isn't it? Well I didn't want to get too technical. Now here's an example. It's called the "dihedral group of order 8". It crops up in many different guises. I could describe it to you the way Galois would have, in terms of substitutions of solutions of a certain polynomial equation, or, as it is presented in a modern course on Galois Theory, as automorphism groups of field extensions. But I won't. That's too hard.

Instead, let me dress it up as a children's party game. I've called it "duels" in memory of Galois. It's rather a fun sort of game that can be counted on to keep

a bunch of bored children amused - for a few minutes anyway. Who said mathematics can't be useful!

“Duels” is a game basically like “O'Grady Says” where players are “out” if they make a mistake in obeying the leader's instructions.

The instructions are RIGHT, LEFT and LOAD. The instructions RIGHT and LEFT require you to turn through 90 degrees, left or right and to LOAD, you hold your hand up with two fingers outstretched as if holding a pistol. But here’s the catch.

Whenever the gun is loaded you must do the **opposite** to what you are told.

If your gun is loaded and you’re told to load, you must unload, that is, fire. And if told to turn right with a loaded gun you must turn left and vice versa. But only when the gun is loaded do you do the opposite. At other times you must obey the instructions exactly.

It’s quite hilarious to watch when a number of people are playing and you really need to keep your wits about you to play it well.

Would you like to try it out right now? If you don’t feel like standing up and obeying the instructions overtly you can remember which wall you’re supposed to be facing, and discreetly raise your right hand whenever the gun is loaded.

Choose a particular starting direction as your “home” direction. Gun unloaded. Ready?

RIGHT
LOAD
RIGHT

Did you remember to obey this second right turn by turning left?

LOAD

You should once again be in your home position with your hands by your side having just fired the pistol.

Now there are eight positions you can be in during this game — four directions, each with a loaded or unloaded pistol. And there are basically eight different sets of instructions for getting you there.

We say that two sets of instructions are equal if they result in the final positions. So for example, LOAD LOAD LOAD would be the same as LOAD. (Never mind that in the first case you've fired a shot.) And three right turns would equal one left turn.

So you see, we've a mathematical system here consisting of eight things. The things aren't numbers — they're sets of instructions. And we can combine them like multiplication by doing one set of instructions after the other.

We get equations like:

RIGHT times RIGHT times RIGHT equals LEFT
and
RIGHT times LEFT equals LEFT times RIGHT

Now here's the interesting thing about this group which makes it quite different from groups of numbers. Are you in your starting position? Gun unloaded?

RIGHT
LOAD

I want you to remember which way you're facing. You just performed RIGHT times LOAD. Now go back to your home position, gun unloaded, and this time do

LOAD
RIGHT

that is, do the same two operations in reverse order. Notice that you've ended up in the opposite direction to before.

RIGHT times LOAD is not equal to LOAD times RIGHT

We have what is called a **non-commutative** group.

The difference between commutative and non-commutative groups is very important in Galois Theory. Commutative groups are those where x times y is always equal to y times x. The dihedral group of order 8 is *non-commutative*.

Suppose you think of solution by radicals as a sort of "abstract stomach" and commutative groups as

particles which can be absorbed by the stomach lining. Any group which can be broken up into commutative bits would therefore be digestible. The dihedral group, for example, can be broken into two commutative bits in a way that I won't attempt to describe.

Galois showed that these digestible groups (or "soluble groups" as he called them) – these groups which can be broken down into commutative bits – are precisely the groups that correspond to polynomial equations that are soluble by radicals.

Some polynomial equations ($3x^5 - 5x^3 + 1 = 0$ for example) correspond to groups which are not soluble, or to use our analogy, they are indigestible. They involve a non-commutative chunk which cannot be broken down further. These polynomial equations are therefore not soluble by radicals. Not even by Galois, who was something of a radical in the political sense.

It was quite an achievement for a young man who only scribbled his mathematics in his spare time and threw the major part of his energies in fighting for the freedom of his country. On the eve of his duel he wrote to two of his friends ...

(The sounds of "Le Marseillaise" are heard in the background.)

Galois: I have been provoked by two patriots and it is impossible for me to refuse.

Your task is simple. I want to let it be known that I am fighting against my will after having exhausted all means of reconciliation. Please remember me, since fate did not allow me a life that would make my name worthy to be remembered by my country.

I die your friend,

É. Galois

(The music of "Le Marseillaise" swells and reaches its dramatic conclusion.)

3. THE INFINITE

§3.1. Is Infinite Knowledge Possible?

We're told that God is omniscient. He knows all. We don't. We have finite brains so we can't know infinitely many different things. Or can we?

The finite brain has a large, but finite, memory capacity. So common sense tells us that there's only a finite amount of information it can contain. Yet in a certain sense we can know infinitely many facts. (Of course this falls far short of omniscience!)

Suppose I am in love with a woman, living in another city. I write and tell her that I love her. At the time of writing she doesn't know that. But the next day, when she receives the love letter, she knows. She knows that I love her but I don't know that she knows. Not until she writes back. When I get her reply, I know that she knows I love her, but she doesn't yet know this. She must wait for my reply to find out.

There are infinitely many facts of the type: " know that she knows that I know that she knows ...." After a few letters it may not be of very great interest to me to distinguish between these successive layers. But logically they're separate facts because they become true on successive days.

Now although there are infinitely many distinct facts here, at any time only a finite number will be known to me. Every second day when I open the next letter from my beloved I have a new fact to add to my collection, but only ever a finite number at any stage.

But suppose that instead I was in her presence, looking lovingly into her eyes. At the moment I issue the words, "I love you", all of the "I know that she knows ." facts are instantly known to me. It's not that the information travels back and forth rapidly at the speed of light. I don't need to wait for her eyes to light up. I can infer that because she is actually in my presence, he has heard what I said. And equally well she can deduce that because I am standing there, I will know that she knows. Infinitely many distinct facts are knowable in an instant!

I know that 2 is bigger than 1. That's one fact. And 3 is bigger than 2 – a second fact. And 4 is bigger than 3, and 5 exceeds 4, and so on. There are infinitely many facts like these, all of which I know. And the reason why I know them is that I have a rule: "one more than any number exceeds that number", or "$n + 1 > n$".

The only way a finite mind can know an infinite number of facts is to know a rule that will generate them. We will only ever use that rule a finite number of times, but the set of potential instances is infinite. In this sense, a finite mind can have infinite capability.

§3.2. What Does ‘Infinite’ Mean?

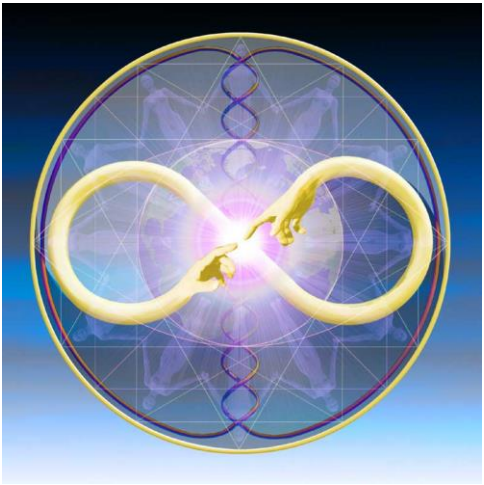
It’s an amazing thing that the finite human mind ever conceived of the idea of the infinite. As amazing as if a race of natives living on a small island which had always been completely cut off from the rest of the world, had a concept of other lands across the sea.

What’s even more amazing is that the finite human mind has to some extent been able to explore the infinite and has discovered detailed facts about infinity. As amazing as if that isolated island race had detailed knowledge of the cathedrals of Europe.

Yet that is the case. Mortal man has notions of immortality, earth-bound man conceived of other worlds long before space travel began, and finite man has had a word for “infinity”. But what exactly is infinity?

I once asked a group of students what they thought ‘infinity’ meant. Some said, “it’s the biggest number there is”. Others said, “it’s something you can approach but never reach.” Yet others said, “the ultimate”.

All of these answers have captured a little of the mystery of the infinite but they’re notions far too vague on which to build any knowledge.



“Infinity is the biggest number there is.” Well of course there is no biggest number in the sense of a number in the sequence 1, 2, 3, ... So we’ll have to invent a new number that goes after all the others:
1, 2, 3, ..., ∞

How does that sound? Any set or collection that isn’t finite is infinite, and the number of elements in an infinite set is denoted by ∞.

That’s a perfectly good state of affairs if we don't want to be discriminating about the infinite. We’re saying that all infinite sets are to be regarded as having equivalent size – there’s only one infinity.

There’s nothing wrong with this – except that it’s a bit like the tribe of Tasmanian aborigines who are supposed to have had no words for numbers after three. Counting in their language went “one”, “two”, “three”, “many”. Any more than three is a crowd. That certainly keeps arithmetic very simple! “Two plus two is many”.

Anthropologists have dismissed this story as false and indeed we’re discovering that Aboriginal culture was rather more sophisticated than we thought.

Elsewhere I describe how the kinship rules of one tribe anticipated the discovery of Group Theory.

Georg Cantor discovered, in the latter part of the nineteenth century, that it's possible to distinguish between different sizes of infinity in a very natural way and this has proved a useful tool in both mathematics and computer science.

Not just one infinity, but many. Now if you're hearing about this for the first time you're perhaps a little sceptical. It's quite a radical idea, even though it's been around for over a hundred years. At least keep an open mind on the question. Simply to automatically lump all infinite sets under the one heading is to make up your mind in advance.

Well if we're to proceed and to ask the question whether or not all infinite sets have the same size, we need to develop some concept of size or "same size". In chapter 1 we saw that we could define two sets of things as having the same size if they can be paired off exactly.

Two sets have the same size if we can pair the elements of one exactly with the elements of the other.

§3.3. Counting Couples

One of the greatest hindrances to social harmony in a society is held to be an imbalance between the sexes. Nowhere is this more evident than at a formal dance.

Now I know that the problem of insufficient men at a dance has often been overcome by women dancing with women, and in modern times the necessity or desirability of dancing with a member of the opposite sex, or indeed having a partner at all, has been called into question. But for the old-fashioned formal balls, for which the Strauss brothers wrote their waltzes and polkas, it was taken as an axiom that dancers were couples and each couple came from opposite sides of the biological tracks.

Imagine then that you're in Vienna at a ball and that you cast your eyes around the many dancing couples. You notice that nobody is sitting out – all are dancing. You'd be justified in concluding that the number of men was the same as the number of women, that is, if you exclude yourself. If, on the other hand, there were a few female wallflowers, and no men, you'd conclude that there were more women than men.

These conclusions would have been reached without counting the men and counting the women and then doing the necessary arithmetic comparison.

Are there more left legs or right legs on the dance floor? A brief inspection reveals no one-legged dancers hobbling on crutches, so the number of left legs is the same as the number of right legs. Again, no counting was involved. Just the realisation that left legs are paired with right legs, just as male dancers are linked to their female partners.

§3.4. The Biggest Number There Is

We tend to think of counting as the most basic of all mathematical activities. Yet more primitive still is the notion of one-to-one correspondences, or pairings.

When, as kindergarten children, we counted out loud as we pointed to each object in turn, we were setting up a one-to-one correspondence between the things we were counting and a certain set of counting numbers. We pointed to a yellow duckling and said “one”. The next one was called “two”. We may have thought, at one stage in our conceptual development, that we were giving names to the fluffy creatures.

Gradually it would have dawned on us that these ‘names’ have nothing to do with what we were counting as we abstracted the concept of number from the things themselves. Soon we felt very proud that we could count to a hundred and beyond. As we learnt to write down longer and longer numbers we began to realise that there was no end in sight. We might not have known what words to use after ‘trillions’ and ‘quadrillions of quadrillions’ but we knew that we could keep adding zeros to make larger and larger numbers.

Big numbers fascinate little children and a favourite pastime is to think of a description of a bigger number than other children.

“I bet you a trillion, trillion, quadrillion dollars that ....”

“I bet you all that and a trillion dollars more!”

“Alright, I bet you all the money in the universe.”

“I bet you a hundred times all the money in the universe.”

It was fortunate that none of these childish bets ever had to be paid. The next stage was the concept of ‘infinity’.

“I bet you infinity dollars.” This was supposed to be a winning move because infinity is the biggest number there is.

“I bet you infinity times infinity dollars!”

§3.5. Dancing To The Music of Schröder and Bernstein

The infinite world is in many ways an extension of the familiar finite world. But in other ways it is quite different. The concept of pairing as the basis for same-number-as works just as well for the infinite as it does for the finite. Where the difference comes is that a finite set gets smaller if you take one thing out. An infinite set does not.

This may seem paradoxical but that's because we're to some extent imprisoned by our experience of the finite world. Remember we've agreed to say that two sets have the same size if they can be paired off exactly with nothing left over.

Dancers in competitions often have numbers pinned to their backs. Imagine a competition with *infinitely* many men and infinitely many women. The dance floor may get a little crowded but with a bit of effort we can read the numbers pinned to them: 1, 2, 3, ... They go on forever.

Number '1' gentleman dances with number '1' lady, '2' dances with '2' and so on. Everyone's happy because the number of men is exactly equal to the number of women. But lady number '1' feels poorly and goes home so number '1' man is without a partner. The numbers of men and women are no longer the same. Right?

Wrong? All it needs is a little reorganisation. Number ‘1’ man can now dance with number ‘2’ lady. Number ‘2’ man, having lost his partner taps number ‘3’ man on the shoulder and takes over his partner. The dance becomes an ‘excuse me’ dance as each man changes to the next numbered lady.

In a finite world, the last man misses out. But in an infinite world, there is no last man! Nobody misses out. Soon everyone has his or her new partner and the dancing goes on. Everyone is happy and so we’re forced to conclude that the number of men and women has remained the same.

So the fact that there are some men, and no ladies, sitting down not dancing doesn’t mean that there are more men than woman. Not at an infinite dance anyway. And if at another dance there are only lady wallflowers it needn’t be the case that there are more ladies than men. The numbers may in fact be the same in each case and it may just need a bit of reorganisation of partners to get everyone on the dance floor.

Of course with finite sets of dancers this can’t happen. Only women left on the side? There must be fewer men. But with infinitely many it’s possible for this apparent disparity to occur with equal numbers.

If lady number ‘1’ had returned after the above excuse-me dance had reorganised the couples, she’d be without a partner, notwithstanding the fact that the

numbers of men and women would still be the same. All very strange, but you can't dabble with the infinite without getting a few shocks.

Schröder and Bernstein may sound like a pair of musicians but in fact they were a couple of mathematicians who discovered, and proved, what is known as the Schröder-Bernstein Theorem. If they had been musicians in an infinite ballroom they might have had a conversation like this:

"I say, Schröder, did you notice that in the last dance there were only ladies left over."

"Of course my dear Bernstein. Such poor organisation. In the dance before that there were only men sitting out."

"That surely means that there are equal numbers of ladies and gentlemen."

"Probably, but can you prove it?"

"I'll think about during the next dance."

Schröder and Bernstein did in fact prove this fact, though not while playing at an infinite ball. In less colourful terms the Schröder-Bernstein Theorem goes something like this (to the tune of "*The Number Rhumba*")

If all the elements of J can be paired
with some of the elements of K,
and all the elements of K can be paired

with some of the elements of J,
then it follows as surely as dead cats have flies
that J and K must have the same size.

Would you like to see a proof of the Schröder-Bernstein Theorem? No, then you'd best make a detour and go straight to the next chapter, but make sure you go via the story *Pam and the Prime Minister*.

So, you're still reading? Well, don't say I didn't warn you. Here's a story that contains the basic idea of the proof of the Schröder-Bernstein Theorem.

We are going to look at imaginary family trees. In this story we have the ability to know the future, so that we can project family trees into the future. Suppose that all men and women who have ever lived, or will live in the future, have, or will have, exactly one son and one daughter.

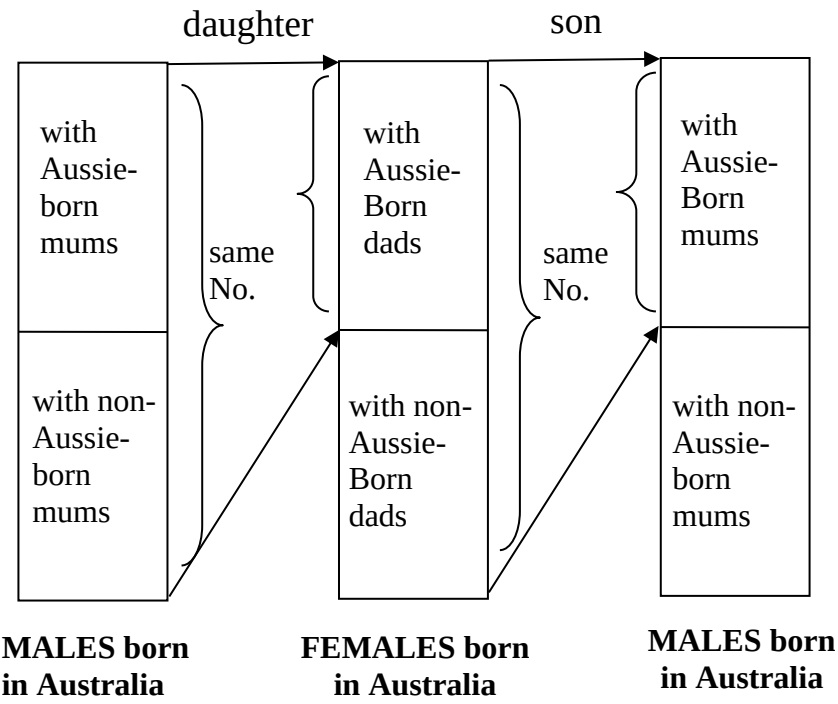
Suppose too, that the daughter of every male born in Australia was, or will be, herself born in Australia and the son of every female born in Australia was, or will be, also born in Australia.

This doesn't seem to be a very good model for the human race but, as I said, this is just a story. We have two sets that we'll call M and F. The set M is the set of all males who were, or will be, born in Australia and F is the set of all females who were born, or will be born in Australia.

Now every male has, or will have, exactly one daughter, also born in Australia and so the males born in Australia will be paired exactly with their fathers, and so there will be the same number of each.

Let FAF be the set of all females with an Australian born father and let MAM be the set of all males with an Australian born mother. These will be paired off exactly with their mother and so there will be the same number of each. So we have 6 sets of people:

- Males born in Australia
- Females born in Australia
- Males born in Australia with Australian mothers
- Males born in Australia with non-Australian mothers
- Females born in Australia with Australian fathers
- Females born in Australia with non-Australian fathersb



The males born in Australia will pair off exactly with their mothers. The females born in Australia will pair off exactly with their fathers.

We want to show that there are as many males born in Australia as there are females born in Australia (under these rather unrealistic assumptions).

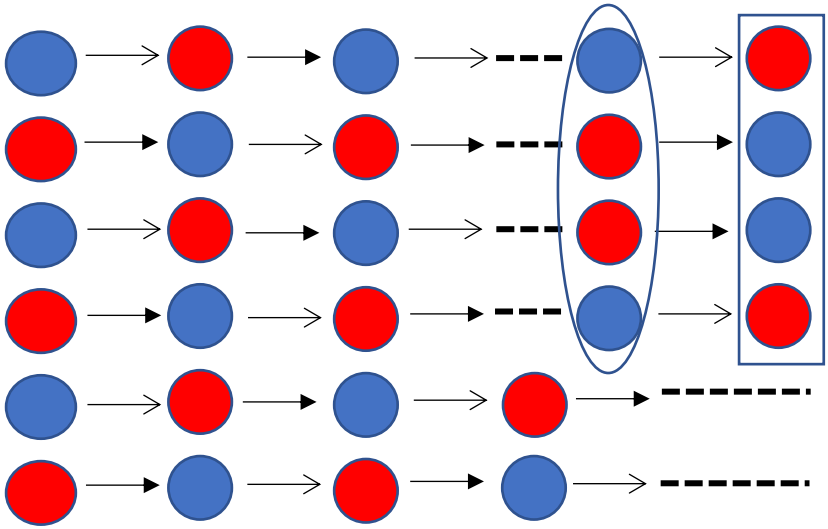
Now many males born in Australia had/have/will have Australian-born mothers. Many of these mothers will have Australian-born fathers, but some will have fathers born overseas. If we trace back, alternately through the mother and the father, we may

eventually strike someone who was born overseas. On the other hand, in this tale of infinitely many generations, both backward and forward, this alternating chain of ancestors may go on forever.

If, for a male, this alternating chain (mother, maternal grandfather, mother of maternal grandfather etc) reaches someone born overseas we call his last Australian born ancestor in this chain his **Ultimate Australian Ancestor** or just **UAA**.

If, for a female, this alternating chain will go father, paternal grandmother, father of paternal grandmother etc. If it reaches someone born overseas we call her last Australian born ancestor in this chain her **Ultimate Australian Ancestor**, or **UAA**.

If such an alternating chain never ends, and there's no UAA for a person, we'll call that person **aboriginal**. Historically, indigenous Australians are descended from people who came from Indonesia over forty-thousand years ago and so they would have UAA's just like the Europeans and Asians in Australia whose UAA was much more recent. But here I'm reserving the term 'Aboriginal' for those mythical people who can trace their ancestry through infinitely many generations, all born in Australia. (I hope I don't offend any indigenous people by this use of the word, but I believe that they prefer the word 'indigenous' to 'Aboriginal' anyway.)



KEY: = male = female

→ **points to father** → **points to mother**

= **born overseas** = **Ultimate Australian Ancestors**

The last two rows depict Aboriginals i.e. with no Ultimate Australian Ancestor.

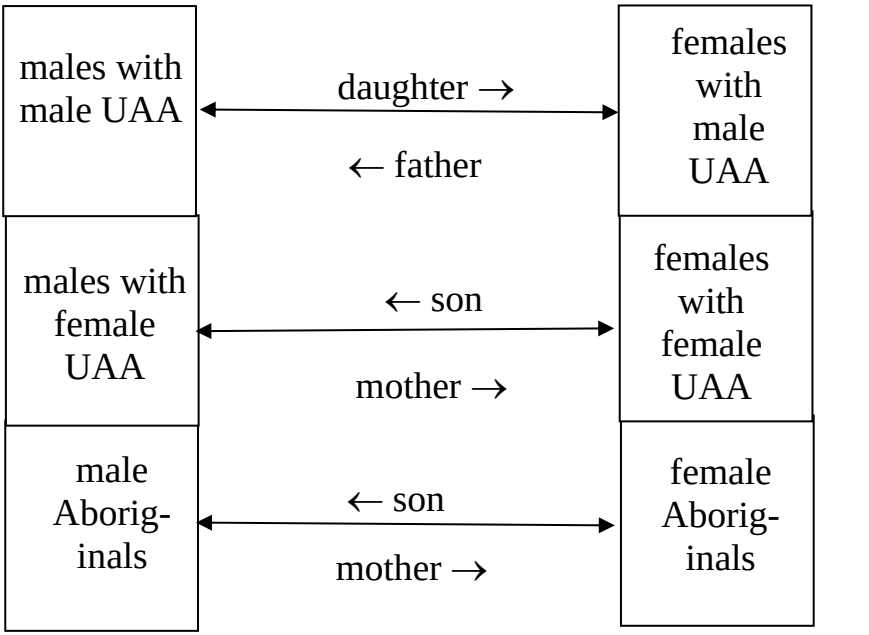
Take young David. He was born in Australia. His mother, Louise, was also born in Australia. Louise's father, Christopher was born in Australia, and Christopher's mother Sue was also born in Australia. But suppose that her father, George, was born in England. Then Sue will be the **Ultimate Australian Ancestor** (**UAA**) of David. She will also be the UAA

of Christopher, and Sue herself. Sue will also be the UAA of David's daughter and his daughter's son.

Now consider Alice. She was born in Hobart. Her father, Bruce, was born in Adelaide. Bruce's mother Connie was born in Sydney and Connie's father was Donald and he was born in Brisbane. So all of these ancestors were born in Australia. Finally, Donald's mother was Eleanor and she was born in London. So Donald is the UAA of Alice. He is also the UAA of Bruce, Connie and Donald himself. Donald will also be the UAA of Alice's son and her son's daughter.

Little Bindi was born in Alice Springs, and her father, Jimba was born in Central Australia. Tracing back Jimba's ancestry in this alternating fashion (mother, father, mother …) we find that they were all born in Australia. In that case we can call Bindi an **Aboriginal** in the rather special meaning of the word that I am using in this story. Jimba will also be Aboriginal and Bindi's son and his daughter will also be Aboriginal. Bindi's daughter may be aboriginal, but that depends on who Bindi marries because the alternating lineage that matters in this story will pass back through her husband.

Now we divide the males born in Australia, and the females born in Australia, each into three subsets:



The males with a male UAA are paired exactly with females with a male UAA.

The males with a female UAA are paired exactly with females with a female UAA.

The male Aboriginals (with no UAA) are paired exactly with female Aboriginals.

So each of the three subsets of males born in Australia has the same size as the corresponding subset of females born in Australia. Hence the set of males born in Australia has the same size as the set of females born in Australia. This example reflects the general proof.

§3.6. The Infinite Manifesto

A political manifesto is a sequence of statements that are believed to be true about the way society should be run. Of course there will only be finitely many statements in such a manifesto. But consider the following *infinite* manifesto:

INFINITE MANIFESTO

[1] At least one of the following statements is FALSE
[2] At least one of the following statements is FALSE
[3] At least one of the following statements is FALSE
[4] At least one of the following statements is FALSE
[5] At least one of the following statements is FALSE
.....

This appears to be a single statement repeated infinitely many times, but each one refers to a different collection of statements and so they're subtly different. Notice that no statement refers to itself, either directly, or indirectly. There is no circular self-referentiality.

Yet there is a paradox hidden in this seemingly innocuous manifesto. Suppose that one of the statements is TRUE. Let's suppose statement [*n*] is

TRUE. Then, by what it says, there is a FALSE statement below it. Let's suppose this is statement [m], where $m > n$.

So statement [m] is FALSE, that is, it is FALSE that there is a FALSE statement below it. This must mean that all the statement after [m[are TRUE.

In particular statement [$m + 1$] is TRUE. So there is a FALSE statement below statement [$m + 1$]. Yet we said that all the statements after statement [m] are TRUE. We have a contradiction.

But we don't yet have a paradox because we assumed that one of the statements is TRUE. We have therefore proved, by contradiction, that all the statements are FALSE. But in each case that would mean that there's a TRUE statement below it, which can't be if all the statements in the list are FALSE. Now we have a real paradox!

Go through the argument slowly a few times until you can see that we cannot assign truth values to these statements in any consistent manner. But note – this example doesn't show that logic is nonsense. It merely shows that the artificially constructed infinite list of, what *appear* to be statements, don't contain any genuine statements at all.

§3.7. The Largest Prime Number

How can you prove that there are infinitely many golden eggs in a magic goose? Just waking up each morning to a new gleaming golden egg is no proof. Perhaps tomorrow there will be none, or the next day. No, the only way to be certain that there will always be a new egg each morning is to cut the goose open and find out. But, you all know the story!

A prime number is a number bigger than 1 that has no factors other than 1 and itself. The list of prime numbers starts with 2, 3, 5, 7, 11, 13, 17, 19, 23, ... Are there infinitely many prime numbers or is there, somewhere out there, a largest prime number? You can't settle it as easily as showing that there is no biggest number. You can't add 1, or even 2, to a prime number and expect to get a prime number. Prime numbers have fascinated mathematicians for thousands of years because in a certain sense they're as unpredictable as random numbers.

What is certain about them is their statistical distribution. While there's no known formula for the *n*'th prime, the probability that a random number of a certain size is prime is known. This probability falls off as the size increases. Primes get rarer and rarer. Could they, in fact, dry up altogether? Euclid proved that they

do go on forever, even though they become scarcer and scarcer. This is a famous example of a proof by contradiction.

Theorem: There are infinitely many prime numbers.
Proof: Suppose to the contrary that there are finitely many prime numbers.

Multiply them altogether and you get a number which is divisible by them all.

Add one more and you get a number that's not divisible by any of them (a prime number can't divide two successive numbers).

Being bigger than every prime it can't be prime itself, yet it must factorise into primes and so is divisible by at least one prime number.

This is a contradiction and so there are infinitely many prime numbers.

INTERLUDE: STORY
Pam and the Prime Minister

The pure voices of the boy sopranos floated up to the lofty recesses of St Mersennes. "... primes without end ... a-men". The service was over.

Elisabeth turned to her friend and said, "Isn't he a dream – those eyes!".

But Pamela said, "I was more interested in his sermon. It seemed very persuasive but I'm sorry, Elisabeth, I'm still an agnostic. I just can't believe in your doctrine of the Infinitude of Primes. I mean, perhaps it *is* true that there are infinitely many prime numbers. I can't see how you could ever know for sure".

"But Pam, you can see here in *Primes Ancient and Modern* and here in *The Book of Common Primes* that there *are* primes for ever and ever into eternity. Look there's no sign of them petering out."

They had reached the church door and the young curate held out his hand.

"I trust you enjoyed the service, Elisabeth?" He greeted them, while looking at Pamela with his penetrating blue eyes.

"Oh yes", gushed Elisabeth, "I found the primes *so* inspiring. But I'm afraid my friend here is an unbeliever".

Pamela smiled sheepishly. “It’s just that I can’t see how you can be so certain. I admit that it seems very unlikely that the list of primes will ever come to an abrupt halt but ... I mean ... it *is* possible. After all primes become rarer as you go among the larger numbers. Is it inconceivable that they eventually give out altogether?”

She pointed vaguely in the direction of the churchyard, but there were too many people behind them waiting to shake the curate’s hand to continue the conversation.

“How about if you and Elisabeth come to the rectory next Sunday afternoon? We could talk some more over tea and scones.”

* * * * * * * * * * * * *

Reverend Matthews poured the tea and passed round the excellent scones that Mrs Duffy had made.

“I’m sure Pam would like to believe that there are infinitely many primes but she doesn’t seem to have enough faith.”

“If only there was some way you could prove it to me,” sighed Pamela, “but of course that’s impossible. Even if I spent from now till the end of the world factorising numbers I’d only be considering a finite number of possibilities. There’s no way the question can ever be settled.”

“Well,” said the curate, “you *do* believe that there are infinitely many numbers altogether don’t you?”

"Oh yes, of course, that's obvious. I mean you just keep on adding one to get bigger and bigger numbers."

"And if I claimed that there was a biggest number?" he asked.

"Then I'd say what about that number plus one?"

"Exactly. I'd be forced to admit that my claim was false."

"But that wouldn't work for prime numbers," protested Pam, "because all primes are odd … except for the number two of course. And so the largest-prime-plus-one would be an even number so it couldn't be prime. And the next number after that mightn't be prime either."

At this Reverend Matthews took a handful of cubes from the sugar bowl and laid them neatly in a row on the damask tablecloth. "Suppose," he said, "that each of these sugar cubes represents a prime number. Here's two and three, five, seven, eleven and thirteen, seventeen, nineteen and twenty-three. Now just suppose, for argument's sake, that there *does* exist a largest prime."

He scooped up the glistening white cubes and put them back into the sugar bowl. "Just suppose that this bowl contains every prime number up to the largest prime."

"Well, alright then," agreed Pamela, "just for the sake of argument. But don't forget that I maintain that believing in a largest prime is just as illogical as believing in the Infinitude of Primes. You'd need infinite time to prove it one way or the other."

“I hope it won’t take *that* long,” he said looking at his watch and smiling, “I have to conduct Evensong at six o'clock!” He picked up the pot containing all-the-primes-in-the-world and said, looking earnestly at Pam, “we have here every prime number that exists and, we’re supposing, there are only finitely many of them.”

“But a very large finite number,” said Elisabeth helpfully.

“Now we can multiply all these numbers together to get an exceedingly large number.”

“What if there’s not enough paper in the whole world to write it down?” asked Elisabeth.

“That's of no consequence”, he assured her, “we can conceive and discuss numbers bigger than the number of atoms in the cosmos. Don’t forget, a number’s existence doesn’t depend on the vital statistics of our universe.”

“But I don’t see what you're getting at”, said Pam as she took another scone. “The result of multiplying all the prime numbers won’t be a prime number itself, so where’s the contradiction, if there is one?”

“But would you agree that this product-of-all-primes will be divisible, exactly, by all prime numbers?”

“Yes Pam, don’t you see,” said Elisabeth excitedly, “every prime number will go into it exactly because every prime will be one of its factors!”

Pam did see. She was more concerned about where the argument would go from there.

"Well the product-of-all-primes will be divisible by all primes so the product-of-all-primes plus one can't be divisible by *any* prime." Reverend Matthews leaned over towards Pamela to make sure she got the point.

"You mean because no two consecutive numbers can have a common factor?", said Pam thoughtfully.

"Exactly. So we're brought to a number which has *no* prime factors. Now this product-of-all-primes-plus-one is too big to be a prime itself."

He put his hand on Pam's head to steady her from the impact of the contradiction that was about to follow. "But every number, if not prime itself, can be factorised into prime factors, so it *must* be divisible by at least one prime and hence we reach a contradiction. And remember Pam that contradiction only came about because we were foolish enough to contemplate a largest prime."

Pam appeared to recover quickly from the shock of the contradiction, if she felt it at all. But in case of an aftershock his hand across the table steadied her arm.

Pam, in fact, was so deep in thought that she forgot for a moment that she even had an arm. She

screwed up her pretty, little nose, trying to make sense of it all. At last she discovered the arm, drew it away from the young curate's grasp and picked up an unused sugar cube which had lain unnoticed on her saucer.

"Well all that means," she said, "is that since it isn't divisible by any of the primes already in our pot it must itself be prime, one we overlooked. So we just pop this extra prime in the pot." And she dropped the sugar cube into the bowl.

"But", protested Reverend Matthews, "you'll just get the same contradiction all over again."

Pamela picked up a handful of the cubes and dropped them one by one back into the bowl. "So, as fast as you keep getting a contradiction I just keep adding more and more primes to the pot. I can always keep one step ahead of you." She grinned, confident that she had him beaten.

But Reverend Matthews yielded no ground. In fact he must have been about to deliver another intellectual earthquake because he felt the need to steady her arm again. "The point is Pam, you agreed that we had *all* primes in our finite pot, and now that I contradict you, you want to add another. That's hardly fair."

But Elisabeth came to her rescue. "Is it such a sin to change one's mind?"

"Look if it were a game of chess I'd be only too glad to let her change her mind to correct an oversight. But she can't claim to be always one step ahead of me just because I let her keep changing her move every time she lands in trouble. Besides, finding a prime

that's not in the prime pot isn't an oversight. We agreed to *define* the contents of the pot to consist of all primes. It's just an inescapable contradiction. And any assumption that leads to a contradiction must be false. Q.E.D."

He appeared to think that this final blow would require a little extra support and it didn't matter that he upset the sugar bowl in the process because those little crystal cubes had served their purpose. Pamela displayed her discomfort at her intellectual position by blushing brightly. She looked down at the spilt cubes on the table as if pleading with them to deliver her the inescapable conclusion of her argument.

At last she looked up into Reverend Matthews' deep blue eyes and sighed, "I suppose you're right."

But Elisabeth, who had becoming more and more agitated while all this was going on, said tersely, "I'm not so sure now. If believing in the Infinitude of Primes stops people from changing their minds I think I'd rather be an agnostic!"

4. THE UNCOUNTABLE

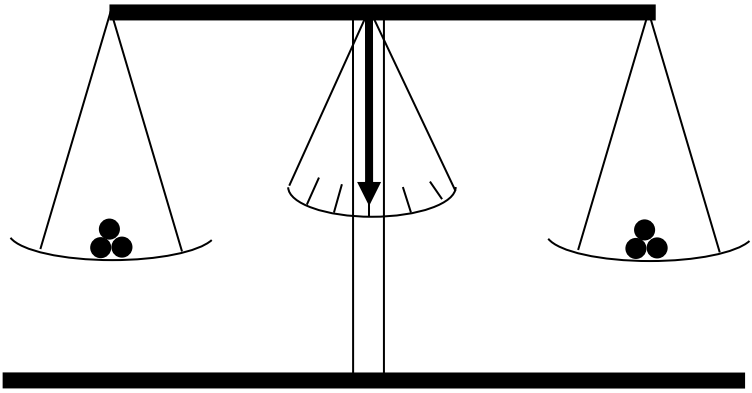
§4.1. The Same-Number Balance

One might think that counting is the most fundamental concept in all of mathematics. Yet, as we have seen, it is a complex idea built on the even more fundamental one of same-number-as. The one-to-one pairing that defines same-number-as can play a similar role as the old-fashioned beam balance.

This was a device that can only compare the weights of two objects. By itself it can't weigh things absolutely. It merely shows you whether or not the weights are equal. Pairing off in a one-to-one correspondence is the balance we use for counting.

Two sets have the same-number-as each other if it is possible to pair the elements of one exactly with the elements of the other exactly.

The reason for the hyphens in 'same-number-as' is because it's a single concept, like 'balance'. As yet we haven't given an independent meaning to the word 'number'. Once we do, we will be able to identify 'same-number-as' with 'same number as' in the sense of each set having a number and those numbers being equal.



§4.2. Standard Sets

A beam balance can be used for weighing things absolutely, as distinct from comparing weights, only if we have a set of standard weights. We need some 1 gram weights and 5 gram weights, and so on, perhaps up to 1 kilogram weights. We put combinations of these into one pan of the scales until they balance exactly with the unknown weight. This enables us to associate a number with the object that we call its 'weight'.

Before we can count, that is, associate a number with a set to represent its size, we need some standard sets to use in the comparisons. In kindergarten we were introduced to a system of symbols 1, 2, 3, ... and

associated words. These ‘objects’ were initially meaningless things that had a defined ordering. ‘Two’ comes after ‘one’ and then comes ‘three’ and so on. We learnt to recite this list “one two three” as we would a nursery rhyme.

What we were setting up in our brains was a nested collection of standard sets (each fitting inside the other for convenience by just stopping at different places). These standard sets are:

STANDARD SET	**SIZE**
{ } (empty set)	0
{1}	1
{1, 2}	2
{1, 2, 3}	3
{1, 2, 3, 4}	4
.....	...

What we are doing when we count a set is to select a standard set which pairs off exactly with it. The **size** of the set is just the number associated with it. (For finite sets it’s the last symbol in the list but when we come to infinite standard sets we’ll need to invent new symbols.)

Perhaps as adults we learnt to count in sophisticated ways, grouping things together for convenience. But if we go back to the primitive act of kindergarten counting we point to each object in turn and call out the next number in the sequence. The last

number we reach will automatically be the answer to the counting.

It's important we get it quite clear what the act of counting really means before we introduce our first infinite number.

To find the number of elements in a set:
Find a standard set which can be put in one-to-one correspondence with it.
The associated number is the answer.

§4.3. The Smallest Infinite Number $\aleph_0$

Are you ready for your first infinite number? We need a standard set and then a symbol to represent its size. What better standard set than the set of all finite numbers

{1, 2, 3,....}?

Now for a symbol. You see, we can't use the last element in the list because there isn't one. We could have used the standard 'infinity symbol', ∞, but that would suggest that this is the only infinite number we're going to get. Besides it's not the symbol used by Georg Cantor who first investigated infinite counting around the end of the nineteenth century. He chose the first letter of the Hebrew alphabet, $\aleph$, and because it was the smallest infinite number he added the subscript '0'. So our list of standard sets has been extended to the following:

STANDARD SET	SIZE
{ } (empty set)	0
{1}	1
{1, 2}	2
{1, 2, 3}	3
.....	..
{1, 2, 3, 4, 5, 6, ...}	$\aleph_0$

§4.4. In Search of a Bigger Infinite Number (Adding)

Now we begin our long journey, in search of an infinite number bigger than $\aleph_0$. With finite numbers we were always able to get a bigger number by adding one.

"My dad's played footy a trillion, trillion times!"

"My dad's played it trillion, trillion plus one times!"

Let's see if $\aleph_0 + 1$ is a bigger number than $\aleph_0$. Well it's certainly not smaller. But could it be just as big? Before we can answer that we must say what we mean to add one to a number, in a way that makes sense for infinite numbers.

When we were learning how to add such finite numbers as 2 and 3 we possibly had a picture of two ducks and three rabbits. Count the ducks. Two. Count the rabbits. Three. How many animals altogether? Before we learnt to add we would have had to count the entire menagerie. One, two, three, four, five. The whole

collection of animals matches exactly with our standard set {1, 2, 3, 4, 5} and so its size is 5. We've demonstrated that 2 + 3 = 5.

As time went on we learnt ways of adding without counting. But if pressed for what it means for 37 plus 63 to equal 100 we would have to say something like: "if you take 37 of one type of thing and combine it with 63 of something else we get 100 things altogether".

Addition corresponds to combining two sets of things together. But it's important that the two sets have nothing in common, otherwise we're double counting. So here's our definition of the sum of two numbers.

To add the numbers *m* and *n*:

(1) Take a set of size *m*.
(2) Take a set of size *n*.
(3) Ensure that these sets are disjoint (have no common elements).
(4) Combine them into one set (take the union of these disjoint sets).
(5) Put this union into 1-1 correspondence with a standard set.
(6) The number of elements in this combined set is defined to be *m* + *n*.

Let's use this to calculate $\aleph_0 + 1$. First we take a set of size $\aleph_0$. The standard set {1, 2, 3, ...} will do.

Now a set of size 1. The standard set of size 1 is {1}, but these two sets have '1' in common. So let's change the second set to {0}.

The union of these two sets is {0, 1, 2, 3, ...}.

Now this certainly appears to be bigger than the set {1, 2, 3, ...} but is it? No. We can match {0, 1, 2, 3, ...} off exactly with {1, 2, 3, ...}. Just write out these sets in rows and each number in the top row pairs off exactly with the one below it:

0	1	2	3	4	5	...
↕	↕	↕	↕	↕	↕	
1	2	3	4	5	6	...

Since neither set has a last element, there is nothing in one row without a mate in the other. According to our definition, therefore, these two sets have the same number of elements. In other words $\aleph_0 + 1 = \aleph_0$.

"But that's absurd. If you add something extra of course you make it bigger!" Careful, you're revealing your parochialism. It's just like someone who's lived all his life in some small outback country town. "Of course, if you go into a bank they'll know your name!"

You're no longer in the finite backwoods you've been in all your life. This is the big city of the infinite. Some facts you've accepted as having universal

application, you now find are just curiosities that only work for finite numbers. Other things you've learnt *do* extend to the infinite. What can you trust in this strange new world? Just the definitions and logic.

So, contrary to naive intuition, you don't make an infinite number bigger by adding one to it. Our search for a number bigger than $\aleph_0$ has so far failed. What about $\aleph_0 + \aleph_0$?

For this we need two disjoint sets of size $\aleph_0$. The standard set {1, 2, 3, ...} will do for one of them and we can take the negative numbers for the other:

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

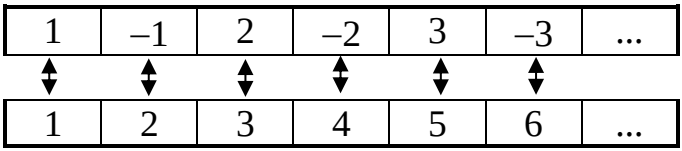
We can set these out in a table with two infinite rows:

1	2	3	4	5	...
–1	–2	–3	–4	–5	...

Surely these can't be paired off with our standard set for $\aleph_0$. To do that we'd have to squeeze both infinite lists into a single one. But that's not difficult. Simply take from each row alternately:

$$1, -1, 2, -2, 3, -3, ...$$

Nothing is left out, but now that they're in a single infinite list we can pair them off with our standard set {1, 2, 3, ...}.



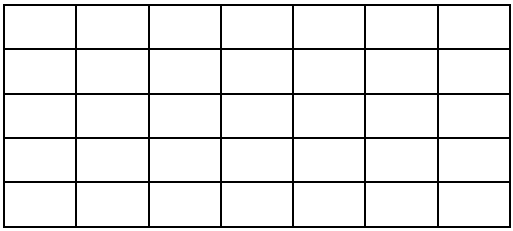
Note that any infinite set which can be listed in a single list has size $\aleph_0$. We just pair the first thing in the list with 1, the second with 2, and so on. Another word that's used for this is **countable**. A set is **countable** if its elements can be listed. Countable sets include the finite ones, as well as those sets which can be put in an infinite list. Our goal is to find an **uncountable** set, whose size will therefore be bigger than $\aleph_0$. So far we've failed.

§4.5. In Search of a Bigger Infinite Number (Multiplication)

We've not yet been successful in finding a number bigger than $\aleph_0$. But we were only using addition up till now – a much more powerful operation is multiplication. Perhaps we'll find that $\aleph_0 \times \aleph_0$ is bigger than $\aleph_0$.

What do we mean by multiplication? Repeated addition? But that won't work with infinite numbers for it would mean that $\aleph_0 \times \aleph_0$ is $\aleph_0 + \aleph_0 +$ ..... with infinitely many terms. Instead we use the idea of ordered pairs.

A table with 5 rows and 7 columns has 35 cells.



Each cell corresponds to a pair (r, c) where r is the number of the row and c is the number of the column in which it lies. It's an ordered pair, that is, for example, $(3, 5) \neq (5, 3)$ because they refer to different cells. So here's the basis for a recipe for multiplying infinite numbers.

To multiply two numbers *m* and *n*

(1) Take a set of size *m*.

(2) Take a set of size *n*.

(3) Form the set of all ordered pairs, with the first item in the pair coming from the first set and the second coming from the second set.

(5) Put this union into 1-1 correspondence with a standard set.

(6) The number of elements in this combined set is defined to be $m \times n$.

Let's use it to find 2×3 and see if we get the answer 6. Take a set of size 2, such as the standard set {1, 2}. Now take a set of size 3, such as the standard set {1, 2, 3}. These sets aren't disjoint, but that doesn't

matter for multiplication. The ordered-ness of the pairs will keep them apart.

Now take all ordered pairs with the first item in each pair coming from {1, 2} and the second from {1, 2, 3}. Here they are:

(1, 1)	(1, 2)	(1, 3)
(2, 1)	(2, 2)	(2, 3)

and as you can see there are 6 of them. So we've proved, using our definition of multiplication, that $2 \times 3 = 6$, which is just as well! Our extended definition of multiplication agrees with the way we've always multiplied numbers *but* it gives us a way of multiplying *infinite* numbers.

Now before we tackle $\aleph_0 \times \aleph_0$, let's first try $2 \times \aleph_0$.

First take a set of size 2. The standard set {1, 2} will do but for a change we'll take {+, −}.
Take a set of size $\aleph_0$. The standard set {1, 2, 3, ...} will do.
The pairs (x, y) where x is a "+" or a "–" and y is in {1, 2, 3, ...} can be put in a table as follows:

(+, 1)	(+, 2)	(+, 3)	...
(−, 1)	(−, 2)	(−, 3)	...

Obviously this is very little different to what we had above and so

$$2 \times \aleph_0 = \aleph_0 + \aleph_0 = \aleph_0$$

as we would expect. So we haven't yet broken the $\aleph_0$ barrier. But we still have $\aleph_0 \times \aleph_0$ up our sleeve!

Take two sets of size $\aleph_0$. Since they don't have to be disjoint we may as well take the standard set {1, 2, 3, ...} for both. Now form all ordered pairs. These can be set out in a two-way infinite table:

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	...
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	...
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	...
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	...
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	...
...	...	...	...	...	...

Can we squeeze this into a single infinite list? All we have to do is to list them by going along the diagonals, starting in the top left-hand corner:

First comes (1, 1), then (2, 1) and (1, 2). Now across to (1, 3) and down the next diagonal and so on.

As a single list this two-way infinite table can be written as a single row:

(1, 1), (2, 1), (1, 2), (1, 3), (2, 2), (3, 1), (4, 1), (3, 2), (2, 3), (1, 4), (1, 5), (2, 4), (2, 3), ……

$\aleph_0 \times \aleph_0$ elements, all written in a single infinite list, means that $\aleph_0 \times \aleph_0 = \aleph_0$. We still haven't succeeded in finding a number bigger than $\aleph_0$. Notice, by the way, that fractions can be represented by pairs of whole numbers so the above diagonal process would give us a way of listing all fractions. So while there appear to be more fractions than whole numbers, in fact the set of all fractions has size $\aleph_0$, just as the set of all whole numbers. Notice that a set can seem to be very much bigger than another, but using the concept of 'same size' that we've adopted they can still have the same size.

Perhaps this apparent paradox disturbs you. Perhaps you say that this definition of 'same size' is the wrong one. Feel free to make up your own definition if you like. However you won't be able to develop the very rich theory of infinite numbers that Cantor did and you'll miss out on a large chunk of the mystery of the mathematical infinite.

§4.6. The Search for a Bigger Infinite Number (Powers)

If we can't find a number bigger than $\aleph_0$ we've made a lot of fuss for nothing. But in fact we're just about to reach our goal. Raising numbers to powers is much more powerful an operation than either addition or multiplication. For example $10 + 10 = 20$, $10 \times 10 = 100$, but $10^{10} = 10000000000$.

You might like to try $\aleph_0^{\aleph_0}$, but instead we'll settle for $2^{\aleph_0}$, which is easier to discuss and is just as big.

How can we give a meaning to 2^n for any counting number n. Multiplying 2 by itself n times is satisfactory for finite n but not if n is infinite. The secret to the correct definition lies in the concept of subsets.

One set is a **subset** of another if everything in the first set is an element, or member, of the second set. For example the set of all women in the world is a subset of the set of all people.

We allow a set to be a subset of itself, so the set of all people is another subset of the set of all people. We even include the empty set as a subset. The set of all people who are over 1000 years old is a subset of the set of all people. It's just that it happens to be empty.

Take a set with two elements, say {1, 2}. How many subsets does it have? Well, what are the subsets of {1, 2}? First there's the empty set { }. then the subsets {1}, and {2}, and finally the set itself {1. 2}. There are 4 subsets. This will be true of any set with 2 elements.

Take a set with 3 elements such as {1, 2, 3}. What are the subsets? They are:

{ }, {1}, {2}, {3}, {1, 2}, {1, 3}, {2, 3} and then finally {1, 2, 3} itself. How many? Eight.

So a set with 2 elements has 4 subsets, a set with 3 elements has 8. Is there a pattern? Yes, a set with n elements has 2^n different subsets, at least if n is finite.

A quick way to see this is to consider that each subset corresponds to a decision for each element whether or not it is to be in the subset. Imagine a sergeant lining up his men and asking for volunteers for latrine duty. In true military fashion it's the sergeant who does the volunteering. As he goes along the row of men he says, "you're in, not you, nor you, yes I want you, no, no, no, yes, ...}. There are n choices, each a choice from two alternatives, so altogether there are 2^n possible subsets. Now this, which is a fact for finite numbers, can be taken as the definition of 2^n for infinite numbers.

To raise 2 to the power *n*

(1) Take a set of size *n*.

(2) Form the set of all its subsets.

(3) Put this union into 1-1 correspondence with a standard set.

(4) The number of elements in this combined set is defined to be 2^n.

§4.7. The Number $2^{\aleph_0}$ is bigger than $\aleph_0$

Powers of 2 grow quickly and it is a simple fact that 2^n is bigger than n, for finite n. But we'll show, by a cunning argument, that 2^n is bigger than n for *all* numbers, n, finite or infinite.

Showing that 2^n is bigger than n involves two steps. We'll first prove that $2^{\aleph_0}$ is bigger than $\aleph_0$.

$2^{\aleph_0} \geq \aleph_0$

"At least as big as" means finding a way of pairing off all the elements of a set with *some* of its subsets. That's easy – you just pair off each element in the set with the corresponding set with one element. The elements of {1, 2, 3} can be paired off with some of its subsets, namely $1 \leftrightarrow \{1\}$, $2 \leftrightarrow \{2\}$, $3 \leftrightarrow \{3\}$. The fact that there are subsets left over, such as {1, 2} etc, shows that 2^n is bigger than n, for finite n, but, as we have seen, having things left over after a pairing doesn't necessarily mean 'bigger' because there could be another pairing that leaves nothing over.

$2^{\aleph_0} \neq \aleph_0$

The proof that $2^{\aleph_0}$ and $\aleph_0$ are different runs along very familiar lines. We suppose that $2^{\aleph_0} = \aleph_0$ and get a contradiction. So suppose then that $2^{\aleph_0} = \aleph_0$.

Let $\mathbb{N}$ be the set {1, 2, 3, ...}. This has size $\aleph_0$. To say that $2^{\aleph_0} = \aleph_0$ means that there must be an exact pairing off of the elements of $\mathbb{N}$ with its subsets. Every element has a corresponding subset and vice versa.

For a given number *n*, one of two things will be true. Either *n* belongs to the subset that it corresponds to, or it does not.

For example one of the subsets of $\mathbb{N}$ will be $\mathbb{N}$ itself, and of course the corresponding element belongs to it. At the other end of the scale, one of the subsets is the empty set and the corresponding element cannot belong to it.

If, for example, 3 corresponds to the set {1, 4, 5} then 3 will not belong to the set that it corresponds to. If 7 corresponds to the set of all odd numbers then 7 will belong to the set it corresponds to.

Suppose we call those elements which belong to the subset they correspond to, **internal** elements. Those which lie outside their corresponding subset will be called **external** elements. So in the hypothetical examples above, 3 will be an external element and 7 will be an internal one.

In symbols, if we denote the subset that corresponds to the element *x* by S(*x*), and use the symbol '$\in$' to denote 'is a member of' and '$\notin$' to

denote ‘is not a member of’, then we can describe these properties of being internal and external as follows:

$$x \text{ is } \textit{internal} \text{ if } x \in S(x)$$
$$x \text{ is } \textit{external} \text{ if } x \notin S(x)$$

Of course whether an element is internal or external would depend on the particular one-to-one correspondence. But if somebody claimed to have a way of pairing off all the elements of a set with all of its subsets (rash claim!) it’s perfectly reasonable to expect that they could tell us whether any given element is internal or external.

Suppose, for argument sake, that somebody claimed to have paired off all the elements of {1, 2, 3, ...} with all of its subsets. Then, in principle, they must have a list such as the following:

1	↔	{11, 32, 117}
2	↔	set of powers of 2
3	↔	empty set
4	↔	set of all multiples of 3
5	↔	set of prime numbers
...	...	.....
3427	↔	set of all numbers
...	...	.....
185367	↔	set of even numbers
...	...	.....
673867	↔	set of all external numbers
...	...	.....

If this was indeed such a list then 1, 3, 4 and 185367 would be external. They would lie outside their corresponding set. The elements 2, 5, 3427 would be internal. If somebody claimed to have such a list it would also be reasonable, in principle, for us to ask where the set of all external elements appears in the list - what number does this set correspond to? There is such a subset and so if the pairing is exact, as claimed, there's a corresponding element. In the above example we are supposing that it's 673867.

Is 673867 itself an internal number or an external one? It has to be one or the other.

If it's internal then it belongs to the set that it corresponds to, that is, it belongs to the set of all external numbers which would make it external. That's nonsense. But, if it's internal, then it's external. So it can't be internal.

But wait! **If it is external**, it's a member of the set of external numbers. So it *does* belong to the set it corresponds to. But this would make it internal! That's nonsense too!

If it's internal then it's external. If it's external, it's internal. One big resounding contradiction! And that contradiction all rests on the assumption that we started with, that the elements could be paired off with the subsets. Therefore they can't be. That is, the

number of elements of any set **cannot** be paired off exactly with the subsets.

This argument can be used for any set.

THE ELEMENTS OF A SET CANNOT BE PAIRED EXACTLY WITH ITS SUBSETS

Suppose the elements of a set are paired off exactly with its subsets.
Let S(x) denote the subset that corresponds to x.
Let Y be the set of all x such that $x \notin$ S(x).
Let y be the corresponding element.
So S(y) = Y.
If $y \in$ Y then by the definition of Y, $y \notin$ S(y), that is, $y \notin$ Y.
And if $y \notin$ Y then $y \notin$ S(y) and so $y \in$ Y.
This is a contradiction, which tells us that such a 1-1 pairing is impossible.

§4.8. The Universe of Infinite Numbers

So $2^{\aleph_0}$ is bigger than $\aleph_0$. We'll call it $\aleph_1$. Actually $\aleph_1$ is usually defined to mean the next infinite number after $\aleph_0$. But nobody knows whether that is $2^{\aleph_0}$ or not. So it seems reasonable to define $\aleph_1$ to be $2^{\aleph_0}$. But if we do that, what if somebody finds an infinite number between $\aleph_0$ and $2^{\aleph_0}$? We'd then have to call it $\aleph_{½}$ or something like that. Relax! That will never happen. Nobody will ever find any numbers between $\aleph_0$ and $2^{\aleph_0}$. How can we be so sure? Because

it has been *proved* that the existence of something between the two is unprovable. Surely that means there aren't any! Not exactly, because nobody has been able to prove that the next number after $\aleph_0$ is indeed $2^{\aleph_0}$. What's more, nobody ever will because a proof exists that shows that it is impossible to prove the next number after $\aleph_0$ is $2^{\aleph_0}$!

Amazing stuff, but all quite logical. We can prove that the statement that *there's no number between* $\aleph_0$ *and* $2^{\aleph_0}$ can never be proved. We can also prove that the statement *there* ***is*** *a number between* $\aleph_0$ *and* $2^{\aleph_0}$ can never be proved. The question is **undecidable**.

The statement that nothing exists between $\aleph_0$ and $2^{\aleph_0}$ is called the **Continuum Hypothesis**. It's an hypothesis, not a fact. But it isn't a conjecture that will be settled one day. It will forever remain an hypothesis. You could say that whether it is true or not is a matter of faith.

"I believe in the Continuum Hypothesis," your creed might run. Fine. That's perfectly consistent with everything else we know about mathematics. But the opposite view is equally logical. I suppose the proper stance to take would be that of an agnostic.

On the other hand, even though it can never be proved, there's a metalogical argument in favour of believing in the Continuum Hypothesis. Since nobody will ever find an actual example of a number between the two (for if they did the matter would be decidable) then for all practical purposes there *isn't* one. Though this falls short of an actual rigorous proof of non-existence, it seems a reasonable position to take and that is the position that most mathematicians take. We are believers in the Continuum Hypothesis.

So taking $\aleph_1$ to be $2^{\aleph_0}$ we can then use the same argument as above to show that $2^{\aleph_1}$ is bigger than $\aleph_1$ and so on. That means there is a whole infinity of infinite numbers:

$\aleph_0, \aleph_1, \aleph_2, \aleph_3, \aleph_4, \ldots$ each bigger than the one before.

If we set out to construct a catalogue of numbers we would start with two rows in our table:

0	1	2	3	4	…
$\aleph_0$	$\aleph_1$	$\aleph_2$	$\aleph_3$	$\aleph_4$	…

But, as they say in the TV advertisements for steaks knives, "there's more!" If you take a whole collection of sets, one for each of the infinite numbers in the second row of this table, and combine all these sets into one huge set (we call this "taking the union") the size of that set will be at least as big as any number in the row, and hence must actually be *bigger* than anything in the row (think about it!). This will then give

us a number bigger than anything in these two rows, so we can use it to start a third row.

But then by taking successive powers of 2 we can work our way along the third row to produce a third infinite sequences of infinite numbers. But wait, there's more. In the same way we got from the second row to the third we can get from the third to a fourth row, and a fifth and so on.

So our catalogue of numbers, all but the first row being infinite, now covers an entire infinite page, with infinitely many infinite rows. But there's still more. There exists a number bigger than any number on the page and so we can start a second page, and a third, and so on until our catalogue occupies infinitely many pages, each with infinitely many infinite rows.

But why stop at one such volume. We can have infinitely many volumes on an infinitely long shelf, and infinitely many such shelves .... The human mind is a wonderful thing to be able to conceive, and even think logically about, such expansive concepts.

Is there any practical use to all this? Such a question brings us back to earth with a thud, even though the answer is "yes". Mathematicians have a real use for knowing about $\aleph_0$, $\aleph_1$ and to some extent about $\aleph_2$. We could live without the others. The number $\aleph_1$ is the number of points on a line, or the number of real

numbers. The number $\aleph_2$ is the number of functions from the set of real numbers to itself.

Where does $\aleph_0^{\aleph_0}$ fit into all this? Is it bigger than $2^{\aleph_0}$? No, in fact it can be shown that it's just the same as $2^{\aleph_0}$.

INTERLUDE: RADIO FEATURE "Beyond the Finite"

MUSIC: *Also Sprach Zarathustra* by Richard Strauss

MALE VOICE: Beyond the familiar numbers 1, 2, 3, ... of the kindergarten, beyond the hundreds of the cricket scoreboard, beyond the millions, tens of millions, millions of millions of economic statistics, beyond the billions of billions, billions of billions of billions of astronomy ... beyond all finite numbers, lies ... the infinite!

Man, imprisoned though he is in a finite world, is able to glimpse the infinity beyond, through the tiny barred windows of religion, philosophy and mathematics.

FEMALE VOICE: Infinity is an ideal that one can approach but never reach.

PRESENTER: This popular view of the infinite has grown out of the mathematical concept of infinite limits which underpins the calculus, but it's not the only insight that mathematics can give us into the nature of the infinite. A somewhat more recent development, though known to mathematicians for

over a hundred years, has yet to make its imprint on the popular mind.

In the 1890s, Georg Cantor extended the concept of counting to infinite collections and came up with a theory of transfinite numbers. Not just a single unattainable infinity, but a whole infinity of bigger and bigger infinite numbers.

To appreciate this mind-boggling concept I want you to come with me on a journey – a journey of the imagination – a journey beyond the finite to the infinite world of Infinland.

MUSIC: *Enigma Variations* by Edward Elgar

****

Once upon a time there was, in a far-off place, a kingdom called Infinland in which there lived a race of creatures, rather similar to men only much smaller, called Infins. There was nothing very remarkable about these Infins except that there were infinitely many of them. I don't mean that their population was exploding at an ever-increasing rate, approaching infinity. It was infinite and always had been.

Infins were happy little creatures. And so they should be for their infinite land was ideally suited to cope with an infinite population. There were none of the annoying shortages that we experience in our overcrowded world.

Take housing for example. Every Infin had his own house. When a pair of young Infins got married

and left home, they were allowed to choose any house in the kingdom. Never mind that it was already occupied. The family living there, and all those beyond that point on that side of the infinitely long street, were obliged to move up to the next house to make room. Because the street was infinitely long there was no last house in the street to be pushed off the end. A marvellous system! You and I might resent the frequent moving, but the Infins had never known any other way.

They all lived on both sides of two roads – East Road and West Road. Each of these stretched for ever, east or west. To the north and south of these roads were the royal gardens belonging to King Aleph II. These, too, stretched for ever to the north and to the south. Situated right between East Road and West Road was the castle of King Aleph. So that the inhabitants East Road could have contact with those on West Road the king very kindly provided a right of way across the castle grounds – for which he charged a modest toll.

Now the Infins had not always lived on East Road or West Road. In fact once the Infins' houses had completely covered what were now King Aleph's fields. Then, the castle in the middle

enclosed a modest garden within its walls. Around this ran the road called “The Circle”. From this stretched the North Road, the South Road, the East Road and the South Road – all going on forever out from the castle.

These were the major highways of the kingdom, but the whole area was crossed by a network of minor roads as well. Some ran north and south and others east and west. All roads stretched on forever in both directions.

One day, King Aleph decided that his modest garden was not big enough. He decreed, therefore, that henceforth Infins must only live on either East Road or West Road.

In any ordinary kingdom this would have created a severe housing problem. However, as the Infin kingdom was infinite it was possible to rehouse everybody on either East Road or West Road. On the day appointed, the King sent a messenger around the kingdom.

Starting with the houses nearest the castle, he travelled in a spiral fashion around the castle, moving further out all the time. Calling at each house (including those already on East Road or West Road) he gave out their new addresses in order, alternating between East and West Roads.

Travelling in this spiral fashion, the messenger was able to call upon every house. In an ordinary

kingdom this would have taken him forever. In fact he would have never got to the end. But Infins can move infinitely quickly if they have a mind to and so the job was done in next to no time at all. And because there were infinitely many houses on East Road and West Road he never ran out of new addresses to give the families.

The traffic chaos was unimaginable as families moved to their new houses. Even families already living on East or West Roads were unhappy about the change for they had to move much further out.

All the houses left unoccupied were demolished and the land became gardens for the king's private use. Since that time traffic on East Road and West Road has been in a permanent state of chaos and Infins who had once been close neighbours now lived vast distances apart.

This mean and despotic act was just one of the many carried out by King Aleph II. As you may have gathered, the King was very unpopular. Yet he was allowed to rule as the people respected the ancient charter laid down by the much-loved grandfather of King Aleph II, Aleph Zero. In this charter it was laid down that his descendants would be entitled to rule so long as they carried out their duties as Lord of Committees.

That's another peculiar thing about the Infins – their love of committees. They liked nothing better than forming committees. They formed committees at work and committees at school. Every Infin family was

organised into committees and subcommittees and sub-sub committees.

Infins waiting at bus-stops would immediately elect a chairman and ask for the minutes of the last meeting to be read out.

Another curious feature of the Infin committees is that the identity of the committee depended solely on the collection of Infins present. If any Infin was absent from a regular meeting the committee was deemed to be a different one. This made the call for apologies redundant because, by definition, every member was automatically present. But it did complicate the reading of the minutes because they had to recall when and where and under what circumstances that exact collection of Infins last met. A committee that met by chance at a bus-stop one day may have had exactly the same membership as the one that happened to be in the same laundromat on the same day many months previously and so constituted the same committee.

The King, as Lord of Committees, had the statutory right and duty to choose a chairman for every committee. However he had to respect the rule, laid down by Aleph Zero:

No Infin may be chairman to more than one committee.

So long as he carried out this task the Infins allowed him to continue to rule. But if he ever defaulted he lost the right to rule.

Now Infins are notorious for their poor memories, the King included. So it often happened that he forgot that he had chosen a certain Infin previously and made the mistake of choosing him to chair a totally different committee. The trouble is that although Infins often *thought* they could remember that someone had doubled up, they couldn't quite be sure. And although they always took minutes of their meetings, they were so disorganised that they could never find them later when they needed them.

So the King continued to get away with his ineptitude. He made out that he consulted a large volume in which he had written down all possible committees and his chosen chairmen but the truth was that he just chose the first name that came into his head or, if he couldn't remember any name he just pointed to someone and said, "you there, I appoint you".

The chairman could be, and often was, chosen by the King from within the committee. Such a chairman was called an **internal** chairman. At other times the King chose a chairman from outside, known as an **external** chairman.

An external chairman was not actually co-opted because changing the composition would change the committee into a quite different one which would of course require a different chairman. No, the Infins were not so stupid as to allow themselves to be caught in a

recursive trap like that. An external chairman chaired but always from the outside.

Now although he usually made a random choice, King Aleph was occasionally put in the position of having to be very crafty in his choice. Once, in an attempt to overthrow the King, the Infins called together a committee consisting of everybody except the King. The King clearly could not suffer the indignity of being the only one excluded and so he chose himself as external chairman of that committee.

Another attempt was made to overthrow the King by Count Able, one of the noblemen of the kingdom. Count Able maintained that King Aleph constituted a committee of one and asked the King to select a chairman. Now King Aleph remembered that he had nominated himself as external chairman of the Every-One-Except-The-King committee, and just in case Count Able remembered that too, he thought it safest to select an external chairman. So he chose Count Able himself as that external chairman. And, for his insolence he cast Count Able into solitary confinement in the dungeons. As he was being dragged off he screamed for the King to appoint a chairman of the Solitary Confinement Committee that consisted of just Count Able. You see, he couldn't be named as an internal chairman – he was already external chairman to the Kings-Of-Infinland Committee and he hoped that whoever got to be external chairman might be able to help him escape.

Of course the King merely appointed the soldier who kept guard outside Count Able's cell as external chairman so such hopes of escape came to naught.

Nothing ever lasts for ever, not even in Infinland, and eventually Count Able was released. But during his confinement he had hit upon a cunning plan to trap the King.

He called a meeting of the Every-One-Except-The-King Committee. Of course, as external chairman, the king had to come too. So the whole infinite population of Infinland crowded into the Great Meeting Hall of the castle. Count Able respectfully asked leave to put a question to the King, and leave was granted.

"Oh noble King, Lord of Committees, you have the royal privilege of choosing chairmen for all committees both actual and potential."

"Indeed I do. I have it all written down in my Book of Chairmen here."

“And most noble King, you may not choose the same Infin to chair more than one committee.”

“Quite right. That's why I have it all written down.”

“So nobody can possibly be both an external and an internal chairman.”

“Certainly not for that would violate my grandfather’s charter and I would lose the right to rule.”

“So you would be able to consult your book and tell us who among us you have chosen to be external chairmen.”

“That is so. At the back of the book I have an index that lists the name of every Infin and next to those whom I have honoured by choosing them as chairmen I have recorded the letter ‘E’ to denote that they are an external chairman or ‘I’ to denote an internal chairman.”

Of course the King was making all this up. The great book was completely blank but nobody was permitted to look inside. However the more he said the more poor King Aleph was playing into Count Able’s hands.

“Then I wish to call a meeting of the Committee of all External Chairmen. Would your highness please read out the names.”

The King should have insisted that this would take too long but, eager to demonstrate his power as Lord of Committees he foolishly co-operated far too readily.

“Certainly,” he said lifting up the great book and indicating a line of division. “You my people on my

right are the external chairmen of Infinland – oh, plus myself and Count Able. All others are dismissed."

Half of the Infins present filed out muttering. Many of them half-remembered having been appointed external chairman of some committee in the past but their memories were not sufficiently strong to contradict the King. Someone else remarked on the extreme coincidence of the external chairmen happening to be all standing on one side. But even an extreme coincidence is not a water-tight proof of fraud, not if it's the King who is supposed to be guilty.

"So your Infinite Majesty we here comprise all the external chairmen of your kingdom. You're sure?"

"Of course I'm sure. It's all written down in my book."

"Each one of us is the external chairman of some committee?"

"That's what I said."

"And those who've left are either internal chairmen ...".

"... or they're not chairmen of anything". The King finished that sentence but he had no idea of the next one!

"Then I ask you, as Lord of Committees to appoint a chairman of this present committee."

The king pretended to consult his book while he thought this out carefully. He sensed a trap but he knew he was safe because of the incredibly bad memories of Infins. If he couldn't remember whether this present collection of Infins had ever assembled before nor could anyone else. So even if he was inconsistent with

what he had done in the past nobody would be able to remember. No he was quite safe.

He was just about to point to the nearest Infin and announce that he was chairman when the thought struck him. That would make him an internal chairman but no internal chairman remained. He'd confirmed that just a moment ago. The memory of an Infin is bad, but it's not that bad! They could all remember that only external chairmen remained.

That was close. He'd nearly put his foot in it. But fortunately he was a match for Count Abel. All he had to do was remember the name of someone who'd left. He consulted his completely blank book.

"I appoint the last Infin who left as the chairman of this committee."

Count Able had him, and King Aleph knew it as soon as he had said this. The King hid his face behind the book to hide his blushes.

"So if he's not here, that would make him an external chairman of this committee. But all external chairmen are present in this hall. Q.E.D."

Immediately there was an uproar and the Infin Revolt had begun. Very few Infins knew what the letters Q.E.D. stood for but they had been told that those three letters would be a sign that the revolution had begun. Perhaps Q.E.D. meant "Quick, everyone destroy" because they did just that. The castle was destroyed and Count Abel was declared the next King.

At his induction he made just two conditions to his accepting the crown.

“I ask two things as your new King. Firstly I wish not to assume the duties of Lord of Committees and secondly I ask that throughout the rest of my life I be granted ownership of any house on West Road which is situated next door to one that I already own.”

The Infins thought that both of these were very reasonable requests and the Count became King.

What the Infins had forgotten was that, although Count Abel would now be living in the castle, as king Aleph III, he already owned number 31761 West Road. So now at his induction he was now to granted ownership of number 31763 West Road. But that then meant he had to be granted ownership of number 31765, and so on.

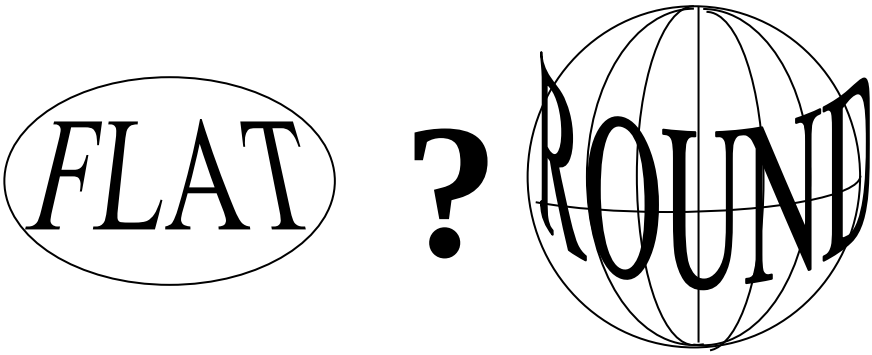
Somebody said something about this being the Principle of Induction. But King Aleph III was much more gracious than the deposed Alep II. He said that all those living on the odd side of West Road didn’t have to move. They could rent the house they were living in. Moreover the rent was infinitesimal, so everyone was happy.

5. THE UNIMAGINABLE

§5.1. How Do We Know That The Earth Isn't Flat?

"The world is flat and the greatest hoax of history is the belief that it's round." I remember hearing this many years ago when I was at university. The claim was made at a lunchtime lecture given by a representative of the Flat Earth Society. Everyone in my physics class went along to heckle this 'nut'. But we were stunned by the fact that he appeared to know far more physics than we did and every objection that we raised was answered by the most convincing and authoritative of explanations.

The belief that light travels in straight lines is the illusion, he said. Ships appear to disappear over the horizon because the light is bending. And the fact that nobody has ever reached the edge of the world is because the closer you come to it, the smaller you become and the more slowly you travel while maintaining the illusion of constant speed. We'd heard of the Theory of Relativity and the lecturer's explanations seemed to be consistent with the very vague understanding we had of that theory.



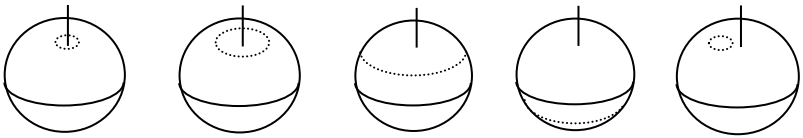
We began to believe that he might just be right! No doubt this was partly due to the heavy atmosphere in the lecture theatre and to his charisma. As we walked out we felt that he was probably wrong but we were no longer sure we could prove that he was.

Now of course nobody who has ever walked up and down mountains believes that the world is *quite* flat. Nor is it as perfectly round as a mathematical sphere. It is, after all, slightly flattened at the poles and its surface is somewhat distorted by mountains and valleys. When the flat-earthist says that the world is flat he means that it is *essentially* a flat disk, but one that may be distorted in some way like a piece of rubber that has been stretched and rippled. We round-earthists likewise assert that the surface of the earth is *essentially* a sphere but concede that it is actually somewhat distorted. The difference between a disk and a sphere is not simply one of shape or curvature. It's 'topological' — it has to do with the way the surface is connected.

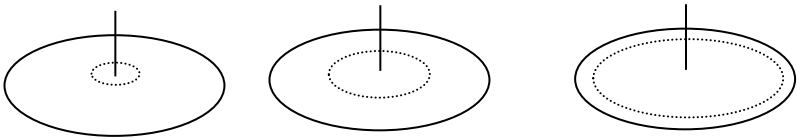
The surface of a sphere can be distorted into many different shapes but without tearing it can't become a disk. A disk can be bent to form a hemisphere or even stretched till it becomes a sphere with a little round hole. But only by sewing up the hole (the reverse of tearing) could it become a complete sphere.

The difference between a 'flat' earth and a 'round' earth is a topological one. Consider the following conceptual experiment. Place a rubber band around the base of the North Pole (assuming it to be an actual pole hammered into the ice – or if there is no ice left there we might need to float a stick on a buoy). Now imagine that this rubber band is enormously elastic and can be stretched as much as we want. Is it possible, by stretching the band, but without breaking it, and keeping it at all times in contact with the earth's surface, to free the band from the pole?

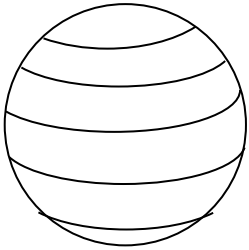
The answer depends on which topological model you accept for the surface of the earth. If it's topologically a sphere, the answer is "yes". All you have to do is to stretch the band over the surface until it runs right around the equator. Then continue moving it south, keeping it in contact with the surface of the sphere at all times, and let it shrink again as it moves towards the South Pole. Now back to its original size it can be slid back north till it lies right beside the North Pole – no longer enclosing it.



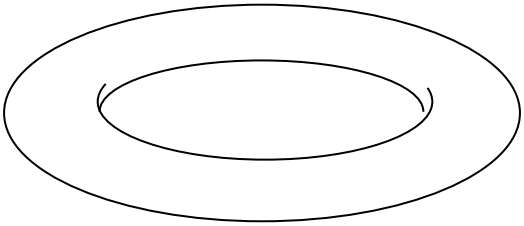
But if the earth is topologically flat then there's no way it could be freed from the pole. No matter how much the band is stretched, the pole will remain 'inside'. It's tempting to say that we could stretch the band till it runs right around the boundary of the disk and then roll it onto the other side. But remember that if the earth is really flat there *is* no other side, or at least it doesn't belong to the surface of the earth. So this is a topological way of distinguishing a sphere from a disk.



To decide which model fits the earth we just have to carry out this experiment. But there's no need to have an actual band that can be stretched so much. The circles of latitude represent the successive positions of such an elastic band moving continuously over the surface. Our flat-earthist might question the validity of the circles of latitude and so remain unconvinced. However the aim of this introduction is not to settle the geographic question but to ask topological ones.



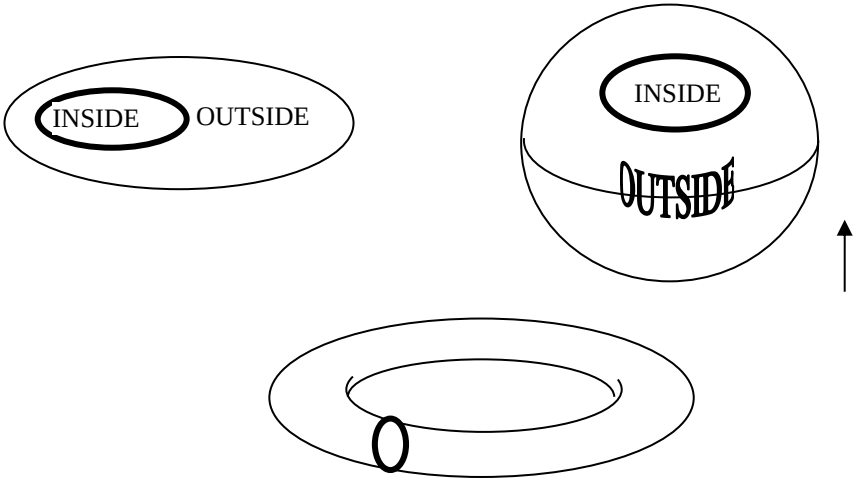
On the other hand the earth might be neither a topological disk nor a sphere. Perhaps it's really a doughnut (or to use the more mathematical word, a 'torus'). Let's leave aside the objection that if so then one part of the world would cast a shadow on the other. This depends on certain assumptions about light and leads us away from topology and back into physics. If we lived on the surface of a torus and had no experience of anything above or below the surface, how could we tell that it wasn't the surface of a sphere? After all you can circumnavigate both a sphere and a torus by travelling in what appears to be a straight line.



In other words we're asking for a *topological* difference between a torus and a sphere or a disk. The infinite elastic band experiment works for the sphere but not for the torus (remember that every part of the band must always remain in contact with the surface at

all times). But it can't distinguish a torus from a disk. This calls for a different conceptual experiment — the Great Wall Experiment.

Build a great wall on the surface of the earth so that its two ends meet. This amounts to drawing a closed curve on the surface. Those inhabitants *inside* the wall are safe from the savage hordes *outside* … or are they? What if the surface of the earth is a torus (doughnut shape) and the wall is built around the smaller radius? The enemy is safely on the other side of the wall, until they wake up to the fact that all they have to do is to travel around the larger radius. Here we have a closed curve that doesn't separate the surface into an inside and an outside. This can happen on a torus but it can't happen on a disk or a sphere.



With the torus, which side of the wall is inside?

§5.2. Do Parallel Lines Exist?

Parallel lines are lines that don't intersect. Of course it's easy to have line segments that don't meet but, if we extended them far enough, they might. Here we're talking about *whole* lines, extending indefinitely in both directions.

In the normal Euclidean plane there are indeed pairs of lines that don't intersect. We just have to draw them so that they're exactly the same distance apart all the way along. It might be difficult to do this precisely. But, of course, we're talking about an imaginary ideal Euclidean plane where things can be as exact as we want them to be. So the answer to that question is an easy "yes" – parallel lines *do* exist in the Euclidean Plane. Or, to put it another way, there *are* pairs of lines that have no point of intersection.

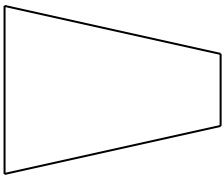
On the one hand it's true that through any two distinct points there's exactly one line. There are no exceptions here – any two distinct points determine a line. But it is not true if we reverse the role of lines and points. It is not true that any two distinct lines intersect in exactly one point. There are exceptions – parallel lines.

We have a similar situation here to what existed with the real number system. It isn't true that every real number has square roots. At least it wasn't true until we invented imaginary numbers. Perhaps we can invent

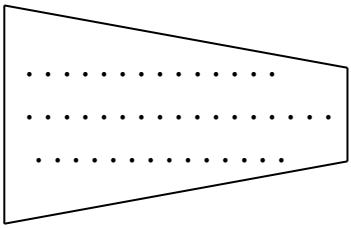
imaginary points where parallel lines can meet. Indeed we can. We enrich the Euclidean plane by inventing extra imaginary points – only they're called 'ideal points'.

For every direction we invent an ideal point and decree that all lines in that direction pass through the corresponding ideal point. But are we allowed to make such a decree? Indeed we are, provided that the geometry we produce is consistent – that is, provided it doesn't lead to any contradictions. So the concept of a line passing through a point will be just the ordinary one if the line and the point are ordinary points on the Euclidean plane. But if the point is an ideal point, 'passing through' will mean that the line is in the direction that corresponds to the ideal point.

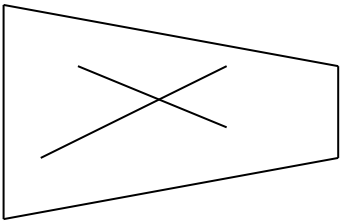
It would help if we could see what is going on in pictures. We can't draw the infinite Euclidean plane on a finite sheet of paper, but we can represent it by a rectangle drawn in perspective.



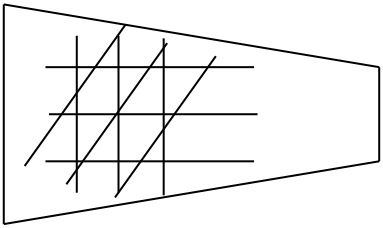
We call the points on the Euclidean plane **ordinary points**.



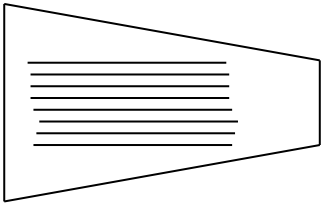
We call the lines on the Euclidean plane **ordinary lines**.



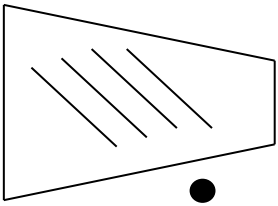
We sort these ordinary lines into parallel classes.



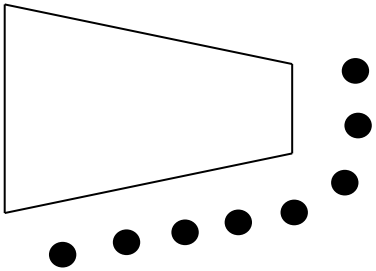
A **parallel class** consists of a line together with all lines parallel to it.



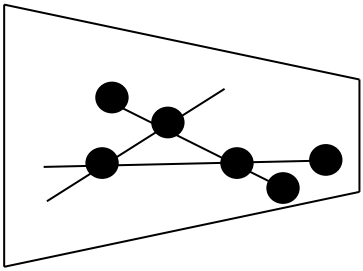
For each parallel class we invent a new point, called an **ideal point**.



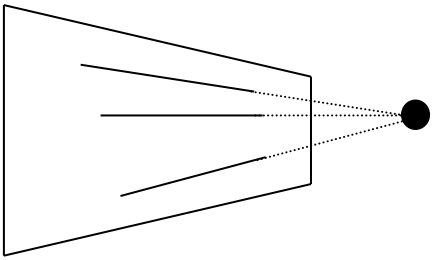
These ideal points don't lie on the Euclidean plane. Where are they then? The answer is simply "in our minds". However, to assist our imagination, we can put these ideal points on our diagram outside of the shape that represents the Euclidean plane.



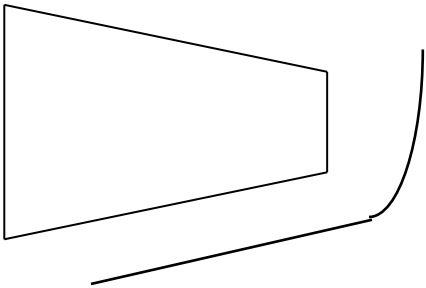
As well as ordinary points lying on ordinary lines in the usual way



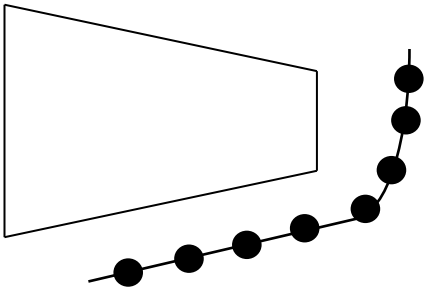
we decree that all lines in a given parallel class (and no others) pass through the corresponding ideal point.



We also invent a new line called the **ideal line**



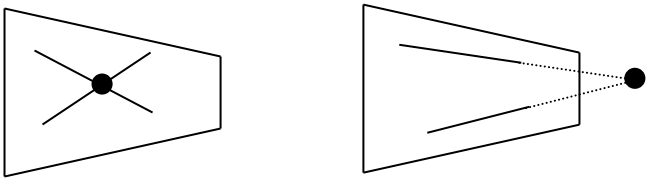
and decree that this line passes through all the ideal points (and no others).



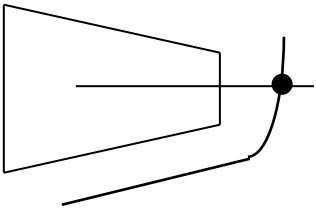
The resulting geometry is called the P**rojective plane**. It contains all of the Euclidean plane, as well as the extra ideal points and the ideal line. Any theorem that we can prove for the Projective plane will be true for the Euclidean plane simply by taking the points and lines to be ordinary ones.

§5.3. The Projective Plane Has No Parallel Lines

Now that we've invented the ideal points and lines, our Projective plane has no parallel lines. Any two distinct lines meet in exactly one point. There are no exceptions. If the two lines are ordinary lines they meet in an ordinary point in the usual way, provided they're not parallel in the Euclidean plane. But if they *are* parallel there we've invented an ideal point in which they can meet.

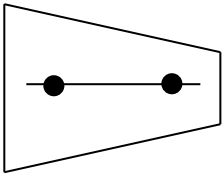


But what if one line is ordinary and the other is the ideal one. No problem. The ordinary line has a certain direction and passes through the ideal point that corresponds to that direction. And of course both lines can't be ideal as there's only one ideal line.

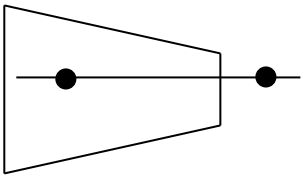


But before we get too excited we might have lost the first property in trying to fix up the second. Is it still true that through any two distinct points there is exactly one line. Let's think it through, case by case.

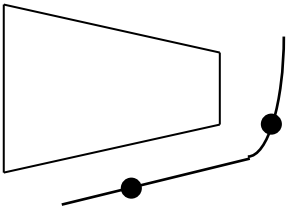
If the two points are ordinary they lie on exactly one ordinary line, in the usual way. They can't also lie on the ideal line (that would make a second line passing through both) because the ideal line has only ideal points.



What if one point is ordinary and one is ideal? The ideal point will correspond to a certain direction. And through any point in the Euclidean plane there is exactly one line in any given direction.

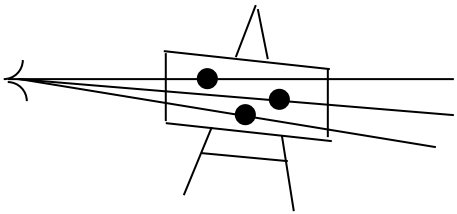


Finally, if both points are ideal then they lie on the ideal line. Could they lie on an ordinary line as well? Well, no. The two distinct ideal points would correspond to two distinct directions and that would mean that the ordinary line would go in two directions at once.



§5.4. Can We Describe The Projective Plane Precisely?

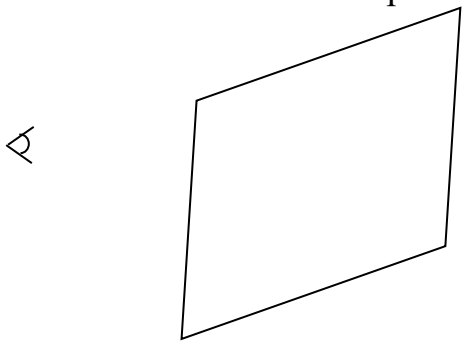
When we observe a long straight railway line, receding into the distance, it looks as if they meet on the horizon. Renaissance artists had no problem with the concept of parallel lines meeting a point. This happens all the time in a perspective drawing.



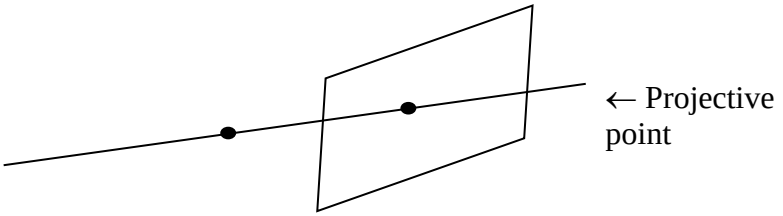
Consider what an artist does when he sketches a scene. You might think that he represents points in the scene by points on the canvas, but it would

be more accurate to say that he represents rays not points. Every ray emanating from his eye corresponds to a single point on his canvas. This leads to the next way of thinking about the real projective plane.

We start with 3-dimensional space and choose a plane, which we call the ‘canvas’ and a viewpoint lying away from the ‘canvas’ where we place an ‘eye’.

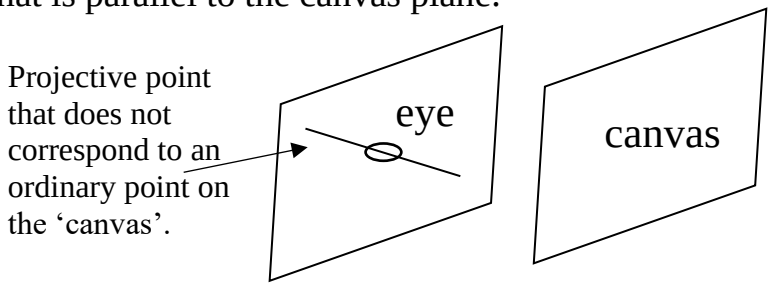


The plane is a complete Euclidean plane (rather larger than the average canvas!). We now define a **projective point** to be a line through the eye. In practice the artist can only see what’s in front of his eye but it suits us to use whole lines rather than rays.



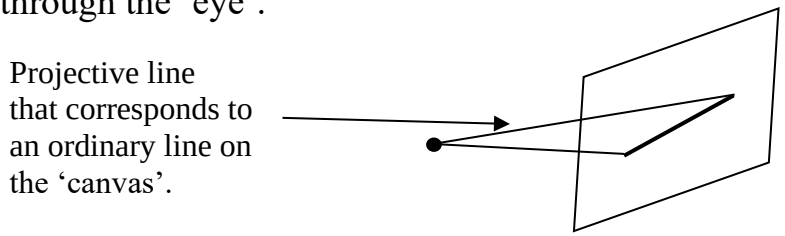
Every point on the canvas corresponds to a projective point but there are some projective points left over that don’t correspond to points on the canvas. These are the lines through the eye that are parallel to

the canvas. These will lie on the plane through the eye that is parallel to the canvas plane.



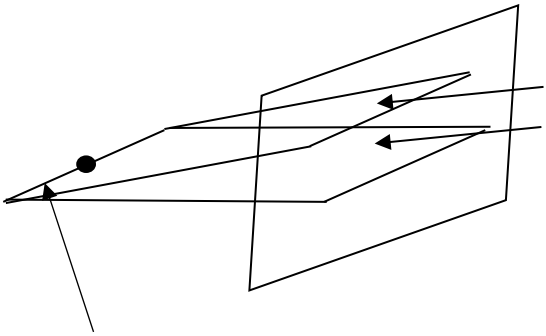
It may seem strange to call something a ‘point’ that we’d normally call a line. That’s why we add the prefix ‘projective’. It bumps things up by one dimension. The rationale behind it is that a projective point (line through the artist’s eye) would appear as a single point on the artist’s canvas.

We now define a **projective line** to be any plane through the ‘eye’.



If the artist paints two parallel lines on his canvas they represent two projective lines in space (that is, planes through the artist’s eye). It’s a fact of 3-dimensional Euclidean space that any two planes through a single point must intersect in a whole line. Moreover that line will be parallel to the plane that contains the two lines.

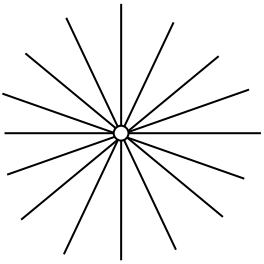
Just open this book and hold two consecutive pages so that the edges are parallel. These pages will intersect in a line that runs along the spine of the book.



This ordinary line in 3-dimensional space, being parallel to the canvas, will not intersect the canvas and so will not correspond to any ordinary point on the 'canvas'. So we can see that, while most of the projective points (lines through the eye) will correspond to ordinary points on the 'canvas', and all but one of the projective lines (planes through the eye) will correspond to ordinary lines on the 'canvas', we have here a model of the Projective Plane. The ideal projective line is the plane through the eye that is parallel to the canvas and the ideal projective points are the lines through the eye on this plane.

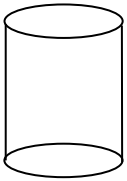
But once the canvas is removed the distinction between ordinary and ideal is removed. The Projective Plane can be considered as a single point O, all lines through O (the projective points) and all planes through O (the projective lines). This model looks rather like a porcupine. (The following picture should be viewed in

3 dimensions with lines pointing in and out of the page.)



§5.5. Can We Draw The Projective Plane on Paper?

There's a rather clever way of drawing the Projective Plane on a piece of paper. Let's begin by drawing the surface of a cylinder. The usual picture is something like this.



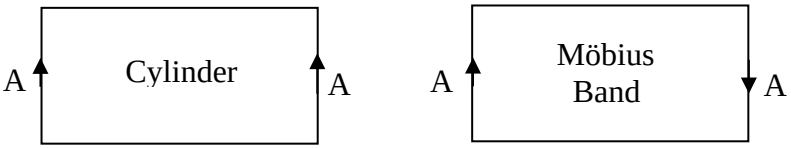
That's not a bad picture for those who can visualise 3 dimensions, but imagine it from the perspective of a tiny microbe moving around the surface, with no concept of up and down. Locally it looks no different to the Euclidean plane. A microbe who's lived on a flat piece of paper will probably notice no difference when it's transported to the surface of a cylinder. Until, that is, it goes on a long journey and circumnavigates the cylinder.

"That's strange," it might say to itself, "I travelled in a straight line and came back to where I started. That never happened on the sheet of paper." It might decide to draw a map of the cylinder as follows:

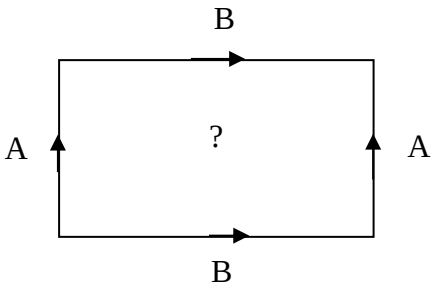
What this is supposed to mean is that a cylinder is basically a rectangle except that if you get to the right-hand side you are spontaneously transported to the left-hand side. The left and right are to be considered the same.

Indeed if you were to cut out the rectangle and bend it to make the left-hand and right-hand sides the same, you'd have a cylinder in 3 dimensions.

But if the rectangle was sufficiently long in relation to its width you could give it a half twist before joining the edges and you'd have a totally different surface called the Möbius Band. So perhaps we should add some arrows to our picture.



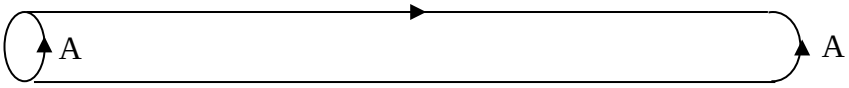
Now what do you think that this shape might be?



If you have played lots of computer games you might recognise this as a screen with ‘wraparound’. If you move off to the right you’re spontaneously transported to the corresponding position on the left edge and if you go off the top you come back to the bottom.

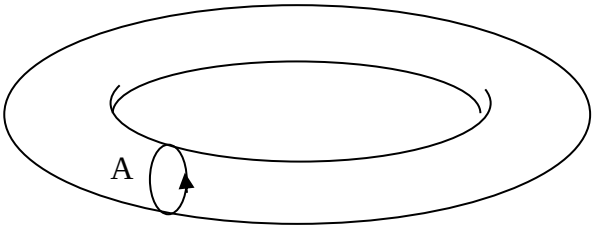
A microbe used to crawling around a sheet of postage stamps might think that this is where it was because if it goes off at the right hand edge it may conclude, not that it has been transported to the left, but rather it’s at the left-hand edge of another postage stamp identical to the one it has just left.

But can we wrap it up into a recognisable shape in 3 dimensions? First we can join up the two edges marked B into a cylinder.



Now here is where we must go a little topological. If we stretch the cylinder so that it

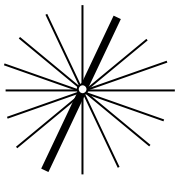
becomes a long hose the ends of this hose will be circles, each marked with the letter A. If we bring these ends together we'll find that the arrows will be going the same way so that we can join the edges to make … a hose with the ends joined!



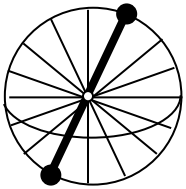
But what shape is that? Well if we lay the hose flat on the ground into a nice circle we discover that what we have is a rather distended doughnut. Actually in some countries doughnuts are not doughnut shaped so perhaps we'd better give the shape its correct mathematical name – the 'torus'.

The two pictures we have so far of the Projective Plane are quite unsatisfactory. The first was not a true picture and in the second we had to call lines points and planes lines. Isn't there an honest-to-goodness picture where points are points and lines are lines?

Suppose we take our porcupine model of the Projective plane where the projective points are really lines through a single point.

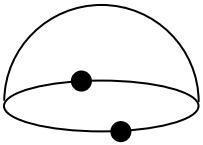


Now enclose this model in a sphere whose centre is that point. Each of our lines will cut the sphere in exactly two points.

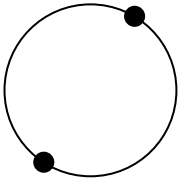


So now our projective points consist of pairs of antipodal points, that is, diametrically opposite points. Each such pair of points has to be considered as a single projective point. Not perfect, but at least a step up from having to consider a whole line as a single point.

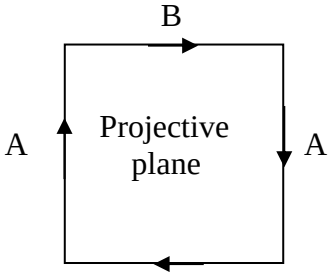
But we can cut this sphere in two. With each pair of points one will be in the ‘northern’ hemisphere and the other in the ‘southern’ – with the exception of the points actually on the equator. These will still be pairs of points representing single ones.



We can flatten the hemisphere so that it becomes a circle. Interior points represent projective points but on the perimeter we have to have pairs of opposite points representing single projective points.



Finally we can bend this surface so that it becomes a rectangle. To represent the fact that pairs of opposite points are to be considered as single projective ones we can place arrows and labels.



So here we have a 2-dimensional picture of the Projective Plane. Can you imagine it now? I doubt it. This business of having to consider pairs of points is still there and while mathematically it's quite sound it's not the sort of picture we might have hoped for. We could join up the A's but that would make a Möbius Band and there would be no way of joining up the B's – not in 3-dimensional space at least.

We're in a similar position to the disembodied angel who's highly intelligent but has no concept of space. We want to imagine the unimaginable, but we can't. Still, we can do lots of things with the Projective Plane. We can even prove theorems about it.

But who's interested in proving theorems about something that's totally imaginary? It sounds like someone writing a book about the biology of a mermaid! The amazing thing in mathematics is that there are problems about our real world that can only be solved by going outside it. There are facts about ordinary real numbers that can only be proved by temporarily going out among complex numbers. There are facts about the Euclidean plane that are best proved by temporarily considering the Projective Plane. There are problems about our familiar 3 dimensional world that we couldn't solve if we remained within 3-dimensions.

So the Projective Plane is a valuable mathematical space. But can we accurately imagine what the Projective Plane looks like? Not on this side of heaven! The Projective Plane is one of the many unimaginables in mathematics.

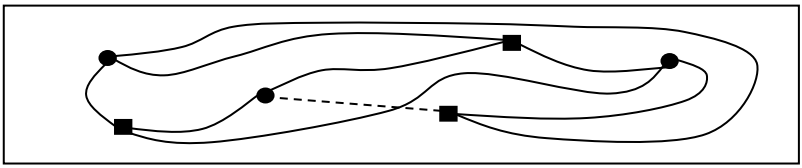
PUZZLE: GAS, WATER AND ELECTRICTY

Remember the Utilities Puzzle in chapter 1. We proved there that it's impossible. Your challenge here is to solve it with modified rules. Take a strip of clear plastic – the sort you can write on. If you don't have such a strip handy, don't worry. You should be able to solve this puzzle in your head,

Somewhere on this strip mark three points with small black circles. These represent the three utilities, a gasworks, a power station and a water reservoir. Now mark three more points with small squares. These represent three houses. These six buildings are supposed to be embedded inside the strip. You can repeat the marks on the other side if you wish but since you can see them from both sides it doesn't matter. The six buildings can be located anywhere you like. Just don't put them too close together or you'll have trouble drawing the lines between them.

Now each house has to receive gas, water and electricity by means of pipes and wires so your task is to draw lines from each of the round dots to each of the square ones. But here's the catch. Because this is essentially a 2-dimensional puzzle you must ensure that pipes and wires don't cross each other. Remember that, like the buildings themselves, the pipes and wires are supposed to be embedded inside the strip so it's no

good having a pipe on one side crossing over a wire on the other.



You'll find that the puzzle is no more possible than it was on a sheet of paper *unless* you take advantage of one extra thing you are allowed to do. You are allowed to bend the strip and join the two shorter edges together. Pipes or wires are permitted to cross this junction, but of course, not each other.

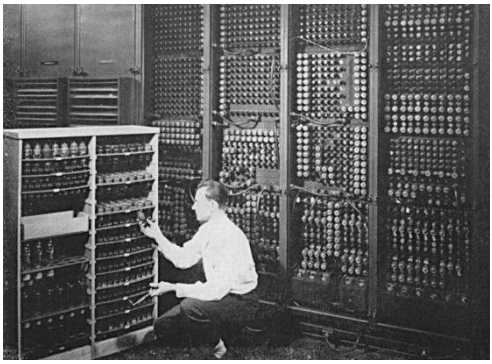
You'll find the answer at the back of the book.

6. THE UNSOLVABLE

§6.1. Are Computers Omnipotent?

Is there anything a computer can't do? Certainly we've witnessed some amazing developments during the eighty or so years computers have been around. Of course we can think of some things that computers can't do – yet. But sooner or later ...

Of course computers can't solve any problem that has no solution. They can't come up with a proof that 1 + 1 = 3, or a procedure which can trisect any angle exactly by ruler and compass. But surely if a solution to a problem exists a computer program can be written to find it. perhaps not today, or tomorrow, but at some time in the future.



In fact our popular belief in the intellectual omniscience of the computer is misplaced. There *are* problems which have solutions but which no computer has ever solved, *will* never solve and *can* never solve.

But wait! Aren't we limiting the ingenuity of man? People once said that man will never fly in

heavier-than air machines, that we will never be able to reach the moon, that smallpox will never be eradicated. How short-sighted is the person who declares that so and so will never happen. Yet that's what I'm saying. Problems exist, problems which have a solution, which man can never solve. And not just human man. No being whose thought processes are based on the same logic as ours can possibly solve these unsolvable problems.

§6.2. The Halting Problem

There's a dream that every novice programmer has. When a computer program is 'compiled' (this just means translating it into a form that the computer more readily understands) the compiler program generates error messages to say that you appear to have left out a comma here or you've misspelt the name of a variable there.

But despite this, usually the first time a novice writes his or her first really complex program the computer 'freezes'. Stupid machine – the keyboard doesn't work, the screen goes on strike. The program has to be aborted by using some emergency key-stroke combination or switching off the power. Even experienced programmers, like the ones who wrote the operating system of your computer, can't avoid having bugs that emerge from time to time – hopefully not too frequently.

When our own program ‘crashes’ our first thought may be to blame the operating system, or the hardware. Perhaps my computer has a virus. But soon the novice discovers that it wasn’t the computer that was at fault, but their program. There was an unforseen infinite loop in the program.

A very obvious case of such an infinite loop is:

```
10: GO TO 10
```

which in line 10 sends the machine back to the same instruction all the time.

An equally obvious case of a program failing to halt is:

```
10: N = 2
20: LET N = N + 1
30: IF N < 2 THEN GO TO 10
```

You might argue that we haven’t got into the sort of loop whereby the computer returns to a previous state. The value of N never repeats. Nevertheless we include such infinite paths as an infinite ‘loop’.

You’d have to be pretty stupid to write such programs, but the problem is that infinite loops can be very subtle and hard to find. Take the following program.

1: LET T = 0
2: LET N = 0
3: ADD 1 TO N
4: ADD $\frac{1}{N^2}$ TO T
5: IF T < 2 GO BACK TO STEP 3
6: OTHERWISE PRINT THE VALUE OF N AND HALT

We start with T = 0 and N = 0. Then we add 1 to N, giving N = 1, and add $\frac{1}{N^2}$ to T, giving T = 1. Since T is less than 2 we go back to step 3. Then we add 1 to N, so that now N = 2 and add $\frac{1}{4}$ to T, giving T = 1.25. Again T is less than 2 and so we go back to step 3. After 10 steps we will have:

$$T = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} + \frac{1}{49} + \frac{1}{64} + \frac{1}{81} + \frac{1}{100}$$

which is about 1.55.

Although T is increasing, it will never be bigger than 2. In fact it's possible to show that T will never be bigger than 1.645. So the program will continue running forever. It will never halt. The Swiss mathematician, Euler, proved that the sum to infinity of this series is $\pi^2/6$ which is 1.644934067 …

Now consider the following similar program.

1: LET T = 0
2: LET N = 1
3: ADD $\frac{1}{N}$ TO T
4: ADD 1 TO N
5: IF T < 10 GO BACK TO STEP 3
6: OTHERWISE PRINT THE VALUE OF N AND HALT

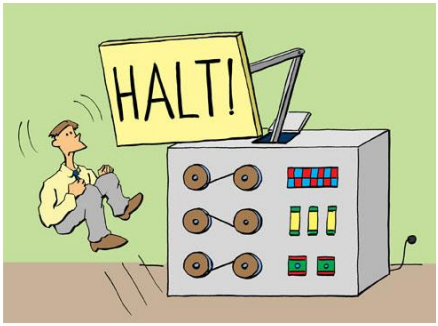
As with the previous program we are adding smaller and smaller numbers to the total. It will still be running after 10000 steps and it would be easy to think that it will go on forever. However, this time the program will eventually halt. When N = 12367, the value of T will become 10.00004301.

Merely deciding that if a program hasn't halted after a certain number of steps is no guarantee that it will never halt. If we changed the '10' in the above program to 1000 the program would, in principle, still halt. But if you ran this program on the fastest computer in the world it would still be running after a year. You might decide that surely it's going to run forever, but in principle it would halt. The only problem is that the universe, and all computers within it, would have decayed long before the program halted.

Now wouldn't it be wonderful if some piece of software

could examine a program, and the data that I plan to use as input, to see if it will get into any infinite loops *before* I actually run it. Such a program would examine the logical structure of my program and very cleverly predict whether or not my program would get itself in an infinite loop.

Such are the very simple specifications for a halting predictor program. "I can see how it could easily detect obvious bugs like 10: GO TO 10 but I'm not sure how it would detect the more subtle loops. But still I'm sure it could be done by some very clever programmer."



Alan designed the perfect computer

Not so! This dream will be forever a dream. Very clever programmers may be able to design something that picks up the more obvious loops. But no programmer will ever be able to write something that can pick up *all* of them. The reason is that doing so is a logical contradiction.

§6.3. Programs

A computer program is simply a list of instructions which the computer follows to solve a problem. Humans are often given instructions and

there's nothing fundamentally different between a computer and the human brain in this sense.

Recipes are simply programs for cooking. Knitting instructions use a set of symbolic abbreviations which one has to learn. In principle a human being armed with unlimited supplies of paper and pencils can do anything that a computer can do. It's just that the computer does it very much more quickly and accurately.

We can prove that the halting problem is unsolvable using any suitable programming language. One can even use the English language, provided we make our meaning sufficiently precise. This means you can follow this argument without knowing much about computers. Basically, all you need to know is that there are three ingredients in the computing process – input, the program and output.

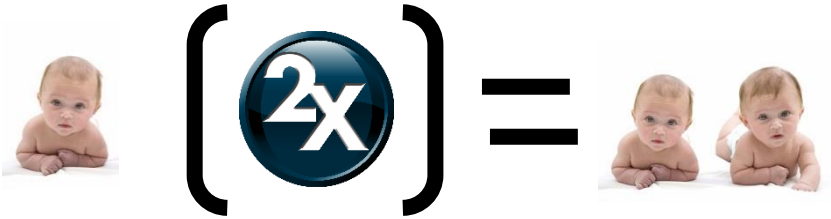
We'll use a shorthand to represent a diagram such as the above. We'll write:

INPUT[PROGRAM] = OUTPUT.

So if **TOASTER** is a program (list of instructions) for using a domestic toaster, we will write **BREAD[TOASTER] = TOAST** to indicate

that if the toaster instructions are applied to the input **BREAD**, the output is **TOAST**.

If **DOUBLE** is the program which describes how to double a number then **3[DOUBLE] = 6**.



Often the output of one program becomes the input of another.

So if **PROG A** and **PROG B** are two programs we'll write the program that consists of doing **PROG A** and then **PROG B** as **PROG A + PROG B**.

INPUT ➡ PROG A ➡ PROG B ➡ OUTPUT

We'll use a shorthand to represent a diagram such as the above. We'll write:

INPUT[PROG A + PROG B]
= INPUT[PROG A][PROG B]
= OUTPUT.

In the above example we could write the process as:

DOUGH[BAKE + CUT]
= DOUGH [BAKE] [CUT]
= BREAD [CUT]
= SLICED LOAF

But beware. PROG A + PROG B is not usually the same as PROG B + PROG A. The order in which you do things usually matters. Try following the steps in a recipe in random order!

If PROG A is the operation of putting on your socks and PROG B is the operation of putting on your shoes then PROG A + PROG B is the normal way of getting dressed, but PROG B + PROG A, where you put the shoes on first and then the socks will be quite different.

Often in mathematics the order of operations is not important. You can add numbers in any order or multiply them in any order. Adding 3 and then adding 2 is the same as adding 2 and then adding 3. But this is the exception. Most operations in mathematics are

sensitive to the order in which we carry them out. For example:

7 [DOUBLE] [ADD 1] = 14 [ADD 1] = 15

while

7 [ADD 1] [DOUBLE] = 8 [DOUBLE] = 16.

§6.4. Some Sample Programs

The input to a program could be a physical object, such as a slice of bread, or a number. But programs that can be processed by a computer can't handle input in the form of a slice of bread. When they were first built it was thought that computers could only accept numbers as their input, but this was wrong.

Fundamentally the input to any computer can only be a string of symbols, essentially strings of 0's and 1's. Such strings can represent numbers. But increasingly, as computers developed we began to realise that such strings can represent words, or pictures, or sounds.

In the examples that follow the input and output are strings of symbols. Fundamentally that's all computers can process. The strings might represent words, or numbers, or pictures but to the computer they're just meaningless strings to be manipulated according to certain rules.

We will assume that whenever our programs write something as output they write it on the same line as the input, coming immediately after it, and so the output includes the inpu, though often the program will explicitly instruct you to erase the input. And when the program runs out of things to do, it halts.

Although, for a real computer, the strings are strings of 0's and 1's, we'll use strings of letters of the alphabet. This will make it more interesting, and easier to understand, Our first program is called **REVERSE**. Very simply, it reverses the order of the letters in a string. The instructions that make up this program are as follows:

REVERSE
1. Write input backwards. 2. Erase input.

So **MESSAGE[REVERSE] = EGASSEM**.

A palindrome is a string which reads the same forward as backwards, like PUP so palindrome;s are those strings that **REVERSE** doesn't alter.

One of the most famous palindromes of all is what Napoleon is supposed to have said:

ABLE WAS I ERE I SAW ELBA. Another famous palindrome, this time without the spaces, is AMANAPLANACANALPANAMA.

Reversing a message twice of course brings the message back to the way it was. Thus we can write:

MESSAGE[REVERSE][REVERSE] = MESSAGE.

****

COUNT is a program which counts the number of symbols in a string.

COUNT

1. Count the symbols in the input.

2. Write this number in words.

3. Erase input.

Carefully examine the following examples and convince yourself that the output claimed is the correct one.

MESSAGE[COUNT] = SEVEN because MESSAGE has 7 letters.

MESSAGE[REVERSE][COUNT] =

MESSAGE[REVERSE][COUNT] =
EGASSEM[COUNT] = SEVEN

MESSAGE[COUNT][REVERSE] =
SEVEN[REVERSE] = NEVES

MESSAGE[COUNT][COUNT][COUNT]
= SEVEN[COUNT][COUNT]
= FIVE[SSAGECOUNT]
= FOUR

The first COUNT counts the number of letters in MESSAGE and the output is SEVEN. The second COUNT counts the number of letters in SEVEN, which is FIVE. The third COUNT counts the number of letters in FIVE and the output is FOUR.

In fact, if you start with any string and repeatedly apply the program COUNT, eventually you will reach FOUR. Why?

****

The next program doesn't erase the input. Instead it makes a second copy of the input.

REPEAT
1. Write "+".

2. Copy input.

MESSAGE[REPEAT] = MESSAGE+MESSAGE

BOO [REPEAT] [REPEAT] =
BOO+BOO [REPEAT]
BOO+BOO+BOO+BOO

MESSAGE[COUNT][REPEAT] =
SEVEN[REPEAT] = SEVEN+SEVEN

MESSAGE[REPEAT][COUNT]
= MESSAGE+MESSAGE[COUNT] = FIFTEEN.

****

The next program doesn't do much except halt. It does throw out an exclamation mark just to prove it's been run.

HALT
1. Write "!".

HELP[HALT] = HELP!

****

Now for a program which deliberately gets into an infinite loop.

LOOP

1. Copy the last letter of the input.

2. Go to step 1.

AGH[LOOP] = AGHHHHHHHHHHHHHHH.....

There's no real output because the program never halts. This program will loop, no matter what the input is.

****

The next program is more discriminating. In fact it will loop, but only if it is told to halt.

DISOBEY

1. If input = **LOOP** then **HALT**.

2. If input = **HALT** then **LOOP**.

3. Otherwise do nothing.

Of course **DISOBEY** doesn't really disobey its instructions. It only appears to do so.

LOOP[DISOBEY] = LOOP!

The machine actually halts after printing the exclamation mark.

HALT[DISOBEY] = HALTTTTTTTTTTTTT.....
This time it doesn't halt. For any other input nothing happens, except for halting.

STAY[DISOBEY] = STAY.

****

The last of our examples here combines **HALT** and **LOOP** with **COUNT**.

MAYBE
1. If the number of symbols in the input is even then **HALT**. 2. Otherwise **LOOP**.

NO[MAYBE] = NO!

YES[MAYBE] = YESSSSSSSSS.....

ANYTHING[REPEAT][MAYBE]
= ANYTHING+ANYTHING[MAYBE]
= ANYTHING+ANYTHINGGGGGG....

This is because ANYTHING+ANYTHING has odd length.

NO[MAYBE][MAYBE]
= NO![MAYBE]
= NO!!!!!!!!!!!!.....
Since NO has even length, **NO[MAYBE] = NO!**
Since NO! has odd length, **NO![MAYBE] = NO!!!!!!!!!!!!.....**

§6.5. Cannibalism

For convenience let's assemble all the programs we've discussed.

REVERSE	**COUNT**	**REPEAT**	**HALT**
1. Write the input backwards. 2. Erase input.	1. Count the symbols in the input. 2. Write this number in words. 3. Erase input.	1. Write +. 2. Copy input.	1. Write !.

LOOP	**DISOBEY**	**MAYBE**
1. Copy the last letter of the input. 2. Go to step 1.	1. If input = **LOOP** then **HALT**. 2. If input = **HALT** then **LOOP**. 3. Otherwise just halt.	1. If **COUNT⟩input** is even then **STOP**. 2. Otherwise **LOOP**.

Perhaps you may be thinking that it's confusing writing both programs and their input/output data with capital letters. Wouldn't it be better to use lower case for data and capitals for programs? The reason is that programs can be considered as data for other programs.

A compiler for a programming language is a very complicated program into which you feed a program to convert it to a form which is convenient for the computer. It's not uncommon for compilers to be written in the same language as the programs they're designed to compile. So you could feed a compiler into a second copy of itself!

Normally when feeding a program to itself we'd do this with the complete list of all the instructions – not just the name of the program. But for simplicity in this discussion let's just work with the names. Let's take each of our seven programs in turn and work out what would happen if their name was used as their own input.

REVERSE[REVERSE] = ESREVER

COUNT[COUNT] = FIVE

REPEAT[REPEAT] = REPEAT+REPEAT

HALT[HALT] = HALT!

LOOP[LOOP] = LOOPPPPPPPPPP.....

DISOBEY[DISOBEY] = DISOBEY

MAYBE[MAYBE] = MAYBEEEEEEE.....

§6.6. Predicting Loopiness

We now come to a program which doesn't exist, even though the following description suggests that it might.

PREDICT

1. If the input has the form data+program and the program would halt given that data as input, then erase the input, write **HALT** and halt.

2. If the input has the form data+program and the program would never halt given that data as input, then erase the input, write **LOOP** and halt.

3. If the input doesn't have the form data+program then erase everything and write **?**.

Although we've listed what we'd like the program to do we haven't said how it should decide whether the program would halt given that data as input. Of course the fact that we can't think how to do it doesn't of itself make **PREDICT** an . impossibility. That is something we've yet to prove. But just *suppose* for the moment that such a **PREDICT** existed.

MESSAGE+COUNT[PREDICT] = HALT
because MESSAGE[COUNT] = SEVEN, and so stops.

LOOP[PREDICT] = ?
because the input doesn't have the required form with a "+" separating two parts.

MESSAGE+LOOP[PREDICT] = LOOP
because MESSAGE[LOOP] = MESSAGEEEEEE...., going into an infinite loop.

YES+MAYBE[PREDICT] = LOOP
because YES[MAYBE] = YESSSSSSSS....., which doesn't stop.

NO+MAYBE[PREDICT] = HALT
because NO[MAYBE] = NO! which stops.

YES+NO[PREDICT] = ?
because although YES+NO has a + separating YES and NO, the program NO hasn't been defined.

Notice that in all these cases our human brain was ingenious enough to work out what would happen — halt or loop. How did we do it? Did we have a systematic procedure? If so, we're well on the way to bringing **PREDICT** into existence. But no, we predicted the behaviour of our programs on an *ad hoc* basis. As we shall see this is the best we can ever hope for.

Now the specifications for PREDICT include the requirement that it always give an answer. So PROGRAM+DATA[PREDICT] will always be either HALT or LOOP.

§6.7. Cannibal Programs

We'll call a program a **cannibal** if it halts when fed a copy of itself as input. Let's see how many cannibals we've bred.

REVERSE[REVERSE] = ESREVER
This halts, so **REVERSE** is a cannibal.

COUNT[COUNT] = FIVE
This halts. It, too, is a cannibal.

REPEAT[REPEAT] = REPEAT+REPEAT

HALT[HALT] = HALT!

Both REPEAT and HALT are cannibals.

DISOBEY[DISOBEY] = DISOBEY
Although DISOBEY will sometimes loop forever (if the input is HALT), when fed its own description it halts and so it too is a cannibal.

LOOP[LOOP] = LOOPPPPPPP ……

MAYBE[MAYBE] = MAYBEEEEE ……

These are not cannibals because they loop when fed their own description.

§6.8. The Final Showdown

We're now going to build our final program. I call it **MEPH**, short for MEPHISTOPHELES. In the description that follows **THIS** represents any possible input and **THAT** represents any valid program.

Now to do this we'll need to assume that a program called **PREDICT**, as described above, *does* exist. If it's fed some data of the form **THIS+THAT** it assumes that **THIS** is data and **THAT** is a program. It then works in some clever way whether the program that we are calling **THAT** would halt, if given **THIS** as input, and prints out **HALT** or **LOOP** accordingly.

If the input doesn't have the right format, or if **THAT** is not in our list of programs, then it halts with **?** as output.

Now of course it can't do this by running the program **THAT**, starting with input **THIS**, because if **THAT** would never halt if given this input then the **PREDICT** program would never reach a final answer.

In a world where some clocks could go forever you could never predict that a given clock will never stop just by waiting for it to stop.

If some humans were immortal, and others weren't, you couldn't decide who was which by reading the death notices in the newspaper. The fact that a name doesn't appear for over a thousand years (of course we're assuming that all deaths are reported in the death notices) might mean that the person is really an immortal or simply that he's a very long living mortal. Computer programs are the same.

If **PREDICT** is going to exist it will have to be exceedingly cunning and examine the structure of the program whose behaviour it is trying to predict. And is it impossible that a clever programmer might one day be able to do it? Frankly, yes, it is impossible. We will *prove* that it is impossible.

MEPH is built up from the programs that we've constructed, which certainly do exist, together with the one we have merely described, **PREDICT**. If things go wrong, as they will, it will mean that **PREDICT** doesn't really exist.

MEPH

1. REPEAT.

2. PREDICT.

3. DISOBEY

Let's check out MEPH with certain inputs.

THIS [MEPH]
= THIS [REPEAT] [PREDICT] [DISOBEY]
= THIS+THIS [PREDICT] [DISOBEY]
= ? [DISOBEY]
= ?

Here **THIS** is not representing arbitrary input, but the specific four letter word. Since we have not defined a program called **THIS**, **PREDICT** simply prints **?** and halts. So unless the input to MEPH is a valid program the output will simply be ?

DOUBLE [MEPH]
= DOUBLE [REPEAT] [PREDICT] [DISOBEY]
= DOUBLE+DOUBLE [PREDICT] [DISOBEY]
= HALT [DISOBEY]
= HALT [LOOP]
= HALTTTTTT …..

LOOP [MEPH]
= LOOP [REPEAT] [PREDICT] [DISOBEY]
= LOOP+LOOP [PREDICT] [DISOBEY]
= LOOP [DISOBEY]

= LOOP [HALT]
= LOOP!

Since **LOOP [LOOP]** = **LOOPPPPPPP ……** **PREDICT** will detect this and merely print out **LOOP**. Then **DISOBEY** will run the **HALT** program.

The big question we are now going to ask is this:

IS MEPH A CANNIBAL?

Of course the answer has to be either "yes" or "no". Let's examine each possibility in turn. The logic of the argument requires a little tenacity to follow. Just hang in there and follow it slowly, step by step.

CASE 1: Suppose **MEPH** is a cannibal.
What does that mean? It means that **MEPH** will halt if it feeds upon itself, that is:

Now **MEPH[MEPH]**
= MEPH[REPEAT+PREDICT+DISOBEY]
= MEPH+MEPH[PREDICT][DISOBEY]
= HALT[DISOBEY]
= HALTTTTTTTT ……..

MEPHISTOPHELES+MEPHISTOPHELES[PREDICT] [DISOBEY]
= HALT [DISOBEY]
= HALT!!!!!! …

But this says that **MEPH** *doesn't* halt when fed its own description, contradicting our assumption for this case.

CASE 2: Suppose **MEPH** is not a cannibal, that is **MEPH+MEPH[PREDICT] = LOOP.**

Now **MEPH [MEPH]**
= MEPH [REPEAT][PREDICT][DISOBEY]
= MEPH+MEPH [PREDICT] [DISOBEY]
= LOOP [DISOBEY]
= HALT

But this says that **MEPH** *does* halt when fed its own description. Again this contradicts our assumption.

Only two possibilities and neither of them true. Each alternative leads to a contradiction. We're in a maze and there's no way out except the door by which we came in. Everything we did was conditional on our assumption that a program satisfying the specifications of **PREDICT** can exist. Therefore it cannot! The halting problem is unsolvable!

INTERLUDE: PLAY
"It's Got To Stop Sometime"

Scene: *A classroom with a long, wide blackboard at the front. The professor is standing at the front, asking for volunteers.*

Prof: Come on now, I need five volunteers to be "people programs". All you need to do is to hold up one of these cards and, when I say, you just perform the instructions on the card to whatever is written on the board.

Noel: I'll have a go but I'm not very good at this sort of thing. I'm sure I'll get it all back-to-front.

Prof: That's exactly what I want you to do. Your program is called REVERSE.

He hands Noel a card on which is written the words:

REVERSE
Reverse what's on the board.

Now whenever I call on you, all you have to do is to rewrite whatever is on the board backwards.

Peter: If it's as easy as that then I'm your man and as my mum always says if you want someone to do a job

properly and not give up half-way through then ask me because I'm your man and as my mum always says ...

Prof: I'm sure you are, Peter. Your program is called REPEAT.

He hands him a second card bearing the instruction:

REPEAT
While what is written on the board ends in "T" put another "T" at the end of it.

The Bubble Twins: *(in chorus)* We'd like to help too, but only if we can do it together.

Prof: Oh, then you'll like your job.

He gives June Bubble a card on which is written:

DOUBLE
Make a second copy of what's on the board, separated by a space

When I call on you, all you have to do is to make a second copy of whatever appears on the blackboard.
Jane: *(to her sister)* Ooh, I'll do the copying because I've got the steadier hand. You can hold up the instructions in case I forget them.

Prof: Right, let's practise those three programs.

Mary: What about me? I knew you'd forget me. It's just not fair!

Prof: You'll get your chance, Miss Contrary, I've got just the job for you. But we'll just practice these first three. Now when I call out the name of your program you have to perform the instructions on your card to whatever is on the blackboard. If I say REVERSE that's your cue, Leon.

Noel: Do you mean me?

Prof: Sorry, Noel, yes it's you I mean. And if I say REPEAT its over to you Peter. And your cue girls is DOUBLE.

He writes the letters RAH on the board.

OK it's DOUBLE first.

*The Bubble sisters write a second **RAH** next to the first to get RAH RAH.*

Now REVERSE**.**

*Noel rubs out the message **RAH RAH** and replaces it with **HAR HAR**.*

And DOUBLE again.

*The message now becomes **HAR HAR HAR HAR**.*

And finally REPEAT.

*Peter was about to start tacking a row of **R**'s on the end of the data but the Prof caught him just in time.*

No Pete. Your instructions are to add T's and only when what is already there ends in T. When it ends in anything else you do nothing.

Peter, somewhat disappointed, sits down again.

Now we'll try another one.

*He cleans the board and writes the word **EXIT**.*

REVERSE.

*Noel changes **EXIT** into **TIXE**.*

Peter: Isn't ENTRANCE the reverse of EXIT?
Prof: No Pete, Noel's right. I said REVERSE, not OPPOSITE. OK, now DOUBLE.

*Jane Bubble adds a second **TIXE**.*

REVERSE

*Noel replaces the **TIXE TIXE** with **EXIT EXIT**.*

And now REPEAT.

*Peter excitedly writes **T** after **T**, getting **EXIT EXITTTTTTTTTTTTTT.....** until he runs out of blackboard. The Prof has to restrain him from continuing across the wall.*

Mary: That's stupid! Whenever Pete takes off nobody else can follow him.

Prof: No, Mary, its not stupid. It's just what happens when a computer program crashes because it gets into a loop.

Mary: Well it's stupid ever to get into a loop. The computer should be clever enough to know that it's being told to get into a never-ending loop and spit out the offending program.

Prof: But Mary, it's not always so easy to ensure that a program will go on forever.

Mary: 'Course it is! Any fool could see what was going to happen when Pete took over. A clever computer would be able to examine any programs it had to run and refuse any which would make it crash.

Prof: But that would need another program to work out what would happen.

Mary: So what! It might be a complicated program but I'm sure someone smart like Tim could come up with one. You just get Tim's program to look at the one you're going to run and if it's OK it rings a bell and if it would loop forever it rings a buzzer. Then you'd know not to let the computer run any program that sets off the buzzer.

Prof: But this program would have to be able to work on every possible program.

Mary: Sure, and what's wrong with that?

Prof: Well, it would even have to be able to work on itself.

Mary: Well any dum dum can see that Tim's program would always halt so if you ran it on itself you'd get the bell, of course. Now when are you going to give me my program, or had you forgotten?

Prof: OK Mary Contrary, I've got just the program for you. It's called DISOBEY.

He gives her a card with the following instructions:

DISOBEY
If what's written on the board
is HALT then REPEAT.
If it is LOOP then
REVERSE.
Otherwise print "?"

Mary: But that's silly. If I'm told to HALT I go on forever writing HALTTTTTTT.....
and if, for example, I'm told LOOP, I write POOL and then halt. I'll always be doing the opposite to what I'm told.

Prof: That's why it's called DISOBEY, Miss Contrary! Let's try it out.

*He writes **POTS** on the board.*

Now REVERSE.

Noel changes it to STOP.

And now DISOBEY.

Mary: Well the data isn't HALT so I do the "otherwise" bit. That means getting POTS again.

She picks up the duster but the professor gently restrains her.

What's the matter, I've got to do a REVERSE, don't I?

Prof: Not you, your job is to activate Leon as a subroutine. He does the actual reversing.

Mary: Oh, all right then. Go on Noel. (I suppose that's who you meant.)

*Noel reverses **LOOP** and once again the word **POOL** is written on the board.*

Prof: Now again.

*He cleans the board and writes **HALT**.*

OK Mary DISOBEY.

*Mary gives Peter a hard thump and Peter starts writing dozens of **T**'s until the Professor gives Peter a nudge to break him out of his infinite loop.*

Now has it ever occurred to you that a program can be made to operate on itself?

Tim: Well I suppose I could write a program called COUNT which counts the number of words in a piece of text and I could run it on a copy of the COUNT program itself.

Prof: Exactly. So June, if DOUBLE acted upon itself, what would happen?

June: DOUBLE DOUBLE toil and trouble – well just DOUBLE DOUBLE I suppose.

Prof: And, Leon, what if you REVERSE REVERSE?

Noel: You'd get ESREVER.

Prof: Pete, would you mind doing REPEAT on REPEAT.

Peter: What do you mean?

Prof: I mean write REPEAT on the board as your data and carry out the REPEAT program on it.

Peter writes REPEAT on the board and then, after scratching his head for a minute, he turns it into REPEATTTTTTTTTTTTTTTT.....

Prof: So if DOUBLE acts upon itself it will halt. The same is true of REVERSE. But if REPEAT acts on its own description as data it will never halt.

Jane: It's just like it gets indigestion. It can't digest a copy of itself.

Mary: Sounds like a cannibal. What a positively disgusting idea!

Prof: That's a good analogy. How about if we call a program a "cannibal" if it halts when it feeds on itself. So DOUBLE and REVERSE are cannibals. But REPEAT isn't. As Jane says, it gets indigestion if it tries to eat a copy of itself. What about DISOBEY Mary?

Mary: DISOBEY isn't HALT so once again I do the "otherwise". Go on Noel, REVERSE.

And Noel proceeds to turn DISOBEY into YEBOSID.

Prof: So DISOBEY is a cannibal program.
Now Tim, the last program is yours. It's called PREDICT.

Tim: I knew you'd say something like that. You're going to tell me that my program predicts whether or not any program will halt, or whether it will go into an infinite loop.

Prof: Exactly, and because the answer will depend on what data it's given it needs to be given the program plus the data.

He hands Tim the last card with the program:

PREDICT
If the program will halt when given the data, print out HALT but if the program will loop, print out LOOP

Noel: That's not very difficult. All Tim's program has to do is just run the given program and if it halts then it prints out HALT and if it doesn't halt ...

Prof: ... then you'd never be able to break into it to print out the message LOOP.

Peter: Well can't you just break it out of its loop if it seems to be going on too long?

Prof: How long is too long? A program might take a very long time and still halt. Even if you waited a hundred years you wouldn't know for certain that it's not going to halt some time in the future.

Noel: Well how's Tim going to do it?

Prof: He can't. It's impossible.

Mary: That's rubbish. Tim's a computer whiz. And even if Tim can't, someone will one day. It makes me mad when people say that something is impossible just because they're not clever enough to do it themselves! Someone clever can examine the program and work out whether it will halt, without actually running it.

Prof: Well, we're supposing for the sake of argument that Tim has done it and PREDICT is that program. Let's try it out.

*He writes the word **TEST** followed by the word **DOUBLE**.*

Prof: OK Tim, PREDICT.

Tim: Well it's obvious that if you ran the program DOUBLE on the TEST data you're just going to get TEST TEST.

Prof: So, carry out your program.

Tim: If I ran DOUBLE on TEST the program would halt so I write the word **HALT**.

*He erases **TEST DOUBLE** and replaces it by **HALT**. The Prof now writes **REPEAT** to the right of **HALT** to get HALT REPEAT.*

Right Tim, here's another example, go ahead and PREDICT.

Tim: Clearly I predict that REPEAT will loop in this case.

*He writes the word **LOOP** in place of **HALT REPEAT**.*

Prof: Well Tim, is PREDICT a cannibal? Will it halt if it feeds upon its own description?

Tim: I guess so. It is supposed to print either HALT or LOOP, but in either case it, itself, *has* to halt so that you can read its answer.

Prof: Now if I was to attach DOUBLE to PREDICT you'd get a program which tells you whether or not any given program is a cannibal. But I want to give it a twist. Here is a program I've called MONSTER.

The professor holds up the last card displaying the four words:

MONSTER
DOUBLE
PREDICT
DISOBEY

Prof: Do you think MONSTER is a cannibal?

Peter: Well it sounds like a pretty uncivilised, pagan program so I guess it is.

Prof: Guessing isn't good enough. We must have certainty.

Jane: Well, one thing's for certain, either it is a cannibal or it isn't.

Mary: Stupid girl. Where do you think that inane remark will get us?

Prof: Further than you might think. Let's follow up each possibility in turn. Suppose Pete is right and it is a cannibal. Let's feed MONSTER its own description to digest. What happens first?

June: Well first we do DOUBLE and get MONSTER MONSTER.

Tim: Then PREDICT examines the structure of MONSTER and decides whether it will halt when it feeds on MONSTER.

Noel: And because we're at the moment assuming that it's a cannibal it will be able to digest its own description, so PREDICT will spit out HALT.

Mary: Then I come along and upset the applecart, because as soon as I see the word HALT, my instructions in DISOBEY tell me to turn this into HALTTTTTTTTTTTTTT...

Prof: Not quite. You have to ask Peter to run the program REPEAT. But it amounts to the same thing.

Peter: But that will give MONSTER indigestion. It'll never get to the end.

Mary: So MONSTER is not a cannibal after all. That's dumb. We assumed it was.

Prof: So all that means is that that assumption has to be rejected.

Tim: Oh, I see, that contradiction proves that MONSTER is not a cannibal.

Prof: Well, as that seems to be the only possibility remaining, let us assume that MONSTER is not a

cannibal, that is, it will go on for ever if it feeds on a copy of itself.

Mary: We don't need to assume that, we know that.

Prof: So lets follow through MONSTER again as it attempts to digest MONSTER. First step gets us MONSTER MONSTER.

Tim: Then my PREDICT program interprets this as the program MONSTER acting on the data MONSTER and PREDICT must predict whether it will halt.

Noel: And since we know that MONSTER is not a cannibal, the answer LOOP will come out of the PREDICT part of MONSTER.

Mary: And then I come along and DISOBEY, which means that since I don't see the word HALT I simply turn the LOOP into a POOL and halt. But that's dumb too because that means that MONSTER is a cannibal. It fed upon itself and finished. Didn't you say that MONSTER couldn't be a cannibal?

Prof: Well we do appear to be in a bit of a fix. If we suppose that MONSTER is a cannibal we can prove he isn't and if he isn't we can prove he is.

Mary: That's the dumbest thing I ever heard. If he is, he isn't and if he isn't he is!

Prof: So we've reached a blank wall again. But remember, we're still making an assumption.

Noel: What's that?

Prof: Well Tim hasn't actually got a PREDICT program.

Peter: So ... ?

Prof: If ever he, or anyone else for that matter, ever came up with a PREDICT program that can decide in advance whether or not any given program will halt, the contradiction we reached a moment ago must inevitably follow. So no such program could ever be written. The Halting Problem is insoluble!

Mary: My "Halting Problem" is the fact that this stupid lesson seems to be going on forever. Tim, do you predict it will ever HALT?

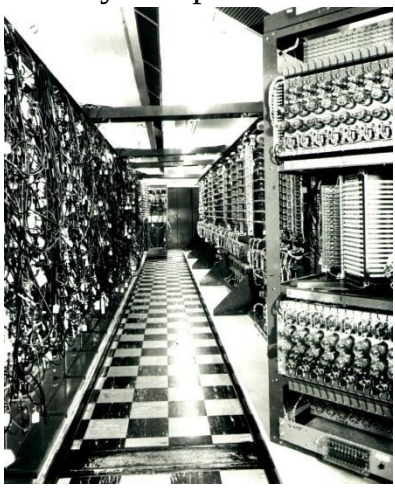
At that moment the end-of-lesson bell was heard.

Tim: Indeed I do.

7. THE UNCOMPUTABLE

§7.1. Conceptual Models for the Computing Process

If we want to investigate in detail what computers can or cannot do, we need a precise conceptual model for the computing process. Computers, their operating systems and their programming languages can be very complex. But this complexity has to do with practicality and efficiency, not possibility. One can use a very primitive computing device and still, given enough time and patience, be able to do anything that the most advanced 'state-of-the-art' computer can do. So, in setting up a model for computability we should set up an abstract machine which is as simple as possible.

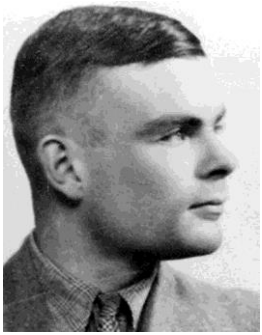


But what we must insist on, with our conceptual model computer, is unlimited memory. Those who drive powerful computers with terabytes of memory

are still conscious of the limitations placed upon them by how much computer memory they have. They'd always like more. Having a fixed amount of storage places artificial limitations on computability, even with something as straightforward as multiplying two whole numbers.

No computer in the world will ever be able to multiply any two arbitrarily large numbers. The process isn't difficult, and computers can be programmed to do this. But with limited memory, even if that limit is huge, we may not even be able to store the input, and even if we just managed to store the two numbers we may not have any memory left over to store the intermediate calculations. Yet we know, in principle, how to multiply any two numbers no matter how large they are. So we say that the multiplication function is computable. The abstract computer that's usually used to explore computability is the **Turing Machine**.

§7.2. Turing Machines



So what is a Turing Machine? Alan Turing was a mathematician who worked in Cambridge in the 1940's. He was fascinated by the concept of computability. Remember this was at a time just before the first actual computers were built. His

conceptual model of a computing machine was based on the English public service.

He imagined a large room filled with clerks. These clerks would make marks on a paper tape or erase marks with an eraser, according to certain instructions. Being clerks in the English public service they were not expected to show any initiative – they had to simply follow orders.

The paper tape was infinitely long (so avoiding any artificial limitations due to limited memory). And the tape came in one hatch and out another. Input was written on the tape, the tape was pulled through the hatch and when the process terminated the tape would be pushed out with the output written on it.

Dispensing with the unnecessary imagery of a room filled with public servants, we can describe a Turing Machine as having an infinitely long paper tape, ruled up into squares. A small device runs up and down this tape, capable of writing and reading marks on the tape. At each moment of time this read/write head is scanning a single square.

There’s only one symbol that can be written onto the tape. It doesn't much matter what it is. We’ll

represent this mark by the symbol 1. Those squares which don't contain a 1 are said to be blank. Throughout the calculation the head writes 1's or erases them.

Now because of its invisibility, a blank is a difficult symbol to represent. For convenience we'll use the symbol 0 to represent a blank. When the machine begins, we assume that there are only finitely many 1's on the tape, representing the input data. Sometimes we begin with the tape completely blank, represented by a two-way infinite sequence of 0's.

In addition to this infinite external memory a Turing Machine has a finite amount of internal memory. There's a gear wheel that can rotate and it can be in any one of a finite number of positions. We call these various positions the 'states' of the machine. So, if the machine has n states, they're labelled 0, 1, 2, ... , $n-1$.

At any given moment the machine is in one of its states and the head is scanning one of the squares. There's a program, or set of instructions, which regulates the behaviour of the machine. Depending on the current state of the machine and the symbol being scanned, the machine writes to the square, moves either left or right one square, and the gear wheel rotates to a new state (or perhaps it stays in the same state). Then the process starts all over again.

The instructions in a Turing program are written in a table. The table has two columns, one labelled 0 and the other labelled 1. The symbol that the head is currently scanning, either a 1 or a blank, that is a 0, determines which column we take the next instruction from.

The rows of the table are labelled 0, 1, 2, ... $n-1$ and represent the states. The current state of the machine determines the row for the next instruction. So if the machine is currently in state 3 and the head is reading a 1, the next instruction comes from the row labelled 3 and the column labelled 1.

Now what do these instructions look like? There's just one type of instruction, which is why it's so easy to learn the Turing Machine language. Suppose that the machine is in state 3, reading a 1 and that the instruction in the appropriate cell of the table is 0L5. This highly cryptic instruction says "print 0, move left and go to state 5". The symbol on the current square is erased (print 0), the read/write head moves one square left and the gear wheel rotates to position 5. One step in the calculation has just occurred.

Just two more comments are needed to complete the description of the Turing Machine – how does it start and how does it stop? The Turing machine always begins in state 0. It stops whenever it's told to go to a non-existent state. For an n-state machine, with states 0, 1, 2, ... , $n-1$, an instruction which tells the machine

to go to state *n* has the effect of halting the machine, indicating that the computation has been completed. What appears on the tape at this stage represents the output of the machine. So if a machine has 5 states, numbered 0, 1, 2, 3, 4 the instruction 1R5 has the effect of printing a 1, moving the head one square to the right, and then halting.

This then is the Turing Machine. It's a wonderful tool in theoretical computing science, but it only exists in the mind. Nobody has ever built such a machine. Infinitely long paper tapes are hard to come by! But since it would be highly impractical for practical purposes this is no loss. The mind is the appropriate place for it.

That's not to say that Turing Machines haven't been simulated on actual computers. It's a very easy exercise to program an actual computer to act like a Turing Machine with a very long tape, which is the nearest one can get in reality to the infinitely long tape.

You may wonder why we need an infinitely long tape if, in the course of a finite number of steps between starting and halting, only finitely many squares are visited. The reason is not that we need infinitely many squares. But we *do* need an arbitrarily large number. We may not know in advance how many squares will be visited so we have to have infinitely many to be on the safe side. We want our uncomputability results to

be absolute, and not simply because we've run off the end of the tape.

Some descriptions of Turing Machines use finite tapes which can be extended if the head is about to fall off the end. But since the real purpose of these machines is conceptual, not practical, we may as well have infinitely many squares and be done with it. After all, an infinitely long tape is no more difficult to imagine than an infinitely long line in geometry or an infinite collection of numbers in arithmetic.

§7.3. Turing Programs

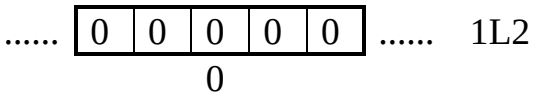
We're now ready for our first Turing Machine program. This one has 3 states and doesn't do anything particularly useful.

	0	**1**
0	1L2	1R1
1	1L0	0L3
2	0R2	1R1

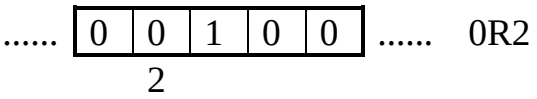
Let's run this program on our Turing Machine. The best way to describe what happens, step by step, is to draw a picture of the tape, or at least a portion of it – as long as the portion includes all the 1's on the tape. So we can assume that everything to the left or the right of the portion that's shown, is blank. We then mark the position of the head and the state of the gear wheel by putting the number of the state underneath the square

being scanned. Finally, as an extra aid to following what is going on, we put the next instruction to be performed at the right of the picture.

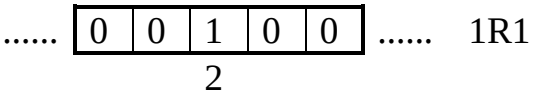
Suppose we start the machine with a blank tape. Being in state 0, reading a 0, the first instruction is 1L1.



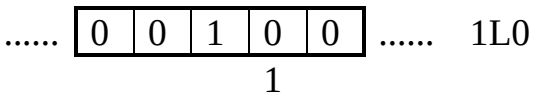
Carrying out this instruction writes a 1, moves the head one square left, and causes the machine to go into state 2. A description of the machine at the end of this machine cycle is:



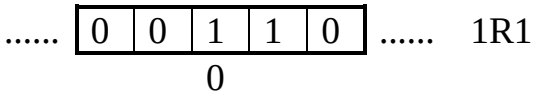
After the next step we have:



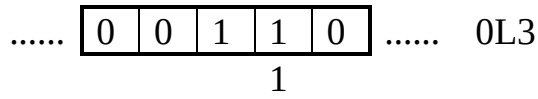
And then:



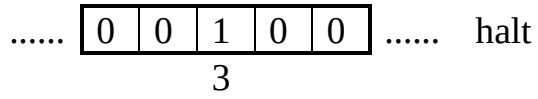
And then:



And then:



And finally:



The machine halts in state 3 with a single 1 on the tape.

Here’s another Turing Machine.

	0	**1**
0	0L1	1R0
1	0R2	1L1

Suppose we start with the data 11111 on the tape, with 0’s to the left and right of these five 1’s, and suppose the head starts scanning the left-most 1. The first instruction to be obeyed is 1R0. This leaves the symbol 1 as it is, but moves right. The machine stays in state 0. The second 1 is encountered, and the same thing happens. We have a loop, with instruction 1R0 being performed over and over again, until the first 0 is reached to the right of all the 1's.

At this stage the instruction 0L1 is encountered. Nothing is changed on the tape, but the head moves left and the gear wheel changes to state 1. Now it is the turn of 1L1 to be performed over and over, while the head moves progressively left and the gear wheel stays in state 1. The head returns, past all five 1's until the 0 to their left is reached. The instruction now changes to 0R3. This moves the head back to where it started, and being sent to the non-existent state 3, the machine halts. The net effect is to return to exactly the same situation as existed at the beginning.

This machine hasn't resulted in any useful computation, but it has performed a little bit of mildly amusing animation, simulating a train which starts at a station, goes down the line till it reaches a blank, returns, overshoots the station, backs up and finally stops at the station.

The next machine behaves in a fundamentally different way to any machine so far.

	0	**1**
0	0L1	1R0
1	0R0	1L1

A quick examination of the instructions will show you that no matter what the initial data is, this machine will never halt. There is simply no halting instruction in the whole table.

The fundamental problem in the Theory of Computation is to find a way of deciding whether or not a given Turing program will halt when we use certain input data. Now we did solve the problem very easily in this particular case (no halting instruction). We don't need to run the machine to see that it will never halt. But the problem is to devise a method which will work in *all* cases.

Certainly if, when you scan the instructions in a Turing program you find nowhere for it to halt, then you can say "it doesn't halt!" But the problem is that the converse doesn't work. Here's a program which provides a halting instruction in the bottom right hand corner. But, if we start it with a blank tape the blighter just ignores it!

	0	**1**
0	0R1	0R0
1	0R1	1L2

After one step, the machine finds itself in state 1, reading 0's, with the head moving continually to the right, for ever and ever.

If we started the above program with input consisting of a finite string of 1's with the head commencing on the one at the left, the behaviour of the machine is easy to predict. It moves to the right, wiping out each 1 as it goes until it reaches a 0. By now the tape is completely blank and the machine then behaves

as before, moving forever to the right. This time the only one of the four instructions not to be reached is the halting one.

§7.4. A Program For Locating a 1

Have you ever run out of petrol on an isolated road in the country and had the difficult job of deciding whether you should walk back the way you came, or walk on. You may remember how far it is back to the last town but what if the next town is just around the corner?

There's an interesting problem like this with Turing machines. Suppose you had the usual infinitely long tape, and you were told that there is a single 1 on the tape – all the other squares are blank. The machine starts somewhere, but you don't know whether the 1 is to the left or the right. The problem is to design a Turing Machine program that will locate this 1 and halt on that square.

It would be no good going left until you hit the 1 because the 1 might be to the right and this strategy would have you going left forever. Similarly it would not do to just go right. The only strategy is to alternately search to the left and to the right. Each time you move to the left you'll need to go further than you did last time and similarly when you search to the right.

So you might go one square to the right, then back to where you started and go one square to the left. Then back to where you started and go two squares to the right. In this way you alternate between left and right searches, and each time your search goes one square further than last time. Sooner or later you will reach the 1 and you then halt.

Such a strategy is fairly obvious, but the primitive nature of the Turing Machine means that we must employ a bit of ingenuity to implement that strategy. All blank squares look the same. You would need some mechanism of counting so that you knew when you had reached the point you had reached previously in that direction so that you could go just one step further.

States can be used for counting in Turing Machines in a limited way. Each time we move to the right we could go to a new state. The problem is that every Turing Machine, by definition, has a *finite* number of states and the number of squares we might need to move might exceed this. Remember the one program has to work in all cases, no matter how far away the 1 is from our starting position.

If you decided to adopt this alternating left/right strategy on the long, straight, featureless road across the Nullarbor Plains in Australia, you might hit on the idea of marking the furthest point you have reached in each direction with a chalk mark on the road. You

wouldn't need to mark the starting point, just the furthest point in each direction.

So, you move east and west alternately. When moving east you continue till you find the chalk mark, erase it, walk a certain distance further and mark the road. Then you walk west till you come to the chalk mark, erase it, walk a certain distance further west and then mark the road. This way you'll eventually reach a petrol station!

How do we adapt this to the Turing Machine problem? A chalk mark would simply be a 1 that we write on a blank square. But we have to be careful we don't confuse the 1's we are writing with the 1 we're looking for. On the Nullarbor we're not likely to confuse a chalk mark with a petrol station, but on a paper tape any 1 looks exactly like any other.

Here's a solution to the problem. Note that the beginning is a little different to the subsequent steps in that we have to put down the two 1's to begin with. Note too that if we are just about to write a 1 on a square that already has a 1 we know that we've found what we were looking for. Oh, and being tidy programmers, we go back and erase the 1's we made and halt on the located 1, leaving the rest of the tape completely blank.

	0	1	
0	1R1	1L7	Put down left marker for the first time
1	1L2	1L8	Put down right marker
2	0L2	0L3	Move left - erase left marker
3	1R4	1R5	Put down left marker in new position
4	0R4	0R1	Move right – erase right marker
5	0R5	0L6	Tidy up right marker
6	0L6	1L7	Return to the found “1”
7	0R10		Halt on the “1”
8	0L8	0R9	Tidy up left marker
9	0R9	1L7	Return to the found “1”

§7.5. The Longest Running Turing Machines

The Turing Machine Olympics is a great occasion. Turing Machines from all round the world compete in many events. But as with any Olympics the star event is the marathon. Actually the Turing Marathon is an endurance race. Speed is a non-issue because all Turing machines run at the same speed. The gold medal goes to the one which runs longest when started with a blank tape.

Now larger Turing machines have the potential for running longer so there are marathon events for each size of machine. So the winner in the 20 state division will be the 20 state Turing machine that runs longest when started with a blank tape.

Imagine the excitement of this great event. Countless Turing Machines are all lined up around a huge stadium. Each of them is loaded with a blank tape. The starting pistol fires and these machines spring into action. Heads fly left and right across each infinite tape as the machines operate furiously.

The machines all run at the same speed so that after a while each machine has run 1000 steps. By now many machines have halted and so are out of the race. Attention focuses on those still running. After a time there are only a few machines left, each showing no sign of tiring. Finally there is just one competitor left, and eventually he halts. Or perhaps there are several who halt at the same time – they are declared joint winners.

But what if a particular program runs forever? It's easy to write a Turing machine program that doesn't halt when started with a blank tape. Apart from tying up the stadium and preventing the next event from taking place, it isn't fair. Programs which loop have to be disqualified at the outset. Only those that will eventually halt are allowed to compete. The stewards have to examine the machines before the race begins and disqualify those that will never halt.

So for each number of states the prize goes to the Turing Machine with that number of states that for the longest number of steps *and eventually halts*, when started with a blank tape. Of course, since the number

of states in a Turing program can be arbitrarily large, there are infinitely many programs competing. But there are only finitely many in each division because there's only a finite number of ways you can fill out any specific table.

The fact that there are only finitely many competitors in each race is important because if there were infinitely many the race may still never finish, even if each competitor does. Suppose there were infinitely many competitors C1, C2, C3, … and that C1 halted after one step, C2 after 2 steps and so on. Even though each competitor eventually halts there would never be a stage when they had *all* halted. But as there are only finitely many Turing machines with a given number of states this problem never arises.

§7.6. The Busy Beaver

When the problem we're about to discuss was first described it was called the Busy Beaver Problem. Beavers are industrious little animals, found in North America, who chop down small trees with their huge teeth and use the timber to construct small dams. The apparent tirelessness of the beavers has inspired such phrases as "as busy as a beaver" and the thought of Turing

Machines "beavering away" suggested the name "Busy Beaver Problem".

As we've seen, with our fanciful Turing Marathon Race, for a given number of states, *n*, there are only finitely many *n*-state machines. Of these some will never halt when started with a blank tape and so are disqualified. If the remaining ones are run, there will eventually be a winner, or at least some joint winners. There will be a certain number of steps at which these winners finally stop.

Let's call this number B(*n*). It's a function of *n* in that you need to know the value of *n* before you can work out B(*n*). It's much like the function f(*n*) = n^2 + *n*, except that we don't have a neat formula for it.

The Busy Beaver Function
B(*n*) is the largest number of steps that an *n*-state Turing Machine can run for, starting with a blank tape, and still halt.

The Busy Beaver Problem
The Busy Beaver Problem is to write a program that will calculate the Busy Beaver Function.

If there was a formula for B(*n*), even a complicated one, it would be a simple matter to write such a program. But programs can be written even when there isn't a formula. If there is *any* systematic

procedure for working out B(n) in all cases, one can write such a program. In what computer language do we want such a program to be written? It doesn't matter because any such program, in any computer language, can be converted to a Turing program.

In what format do we want our answer? Do we want the B(n) to be written in normal notation, or in binary or in some other form. Binary notation is the system used for expressing numbers with just with 0's and 1's and it's the way numbers are actually stored inside a real computer.

But again it doesn't matter because it's a routine programming task to convert from one system of notation to another. Binary might seem to be one that's very suitable for Turing Machines, but don't panic if you don't know about binary notation. There's a much simpler system we can use.

§7.7. Unary Notation for Numbers

There have been many systems of notation for the numbers 1, 2, 3, ... The Babylonians had a system based on counting in 60's. The Romans had their Roman numerals. The Arabic system we use is quite efficient. Binary is particularly suitable for computers. But the simplest system of notation is **unary**.

In unary we represent the number 7 by 1111111. This is the system prisoners use to mark off the number of days of their imprisonment. It's a system which is also used to score in various sports like cricket. Sometimes the strokes are grouped into 5's or 10's to make them easier to count, but the basic system just has 1's. What's attractive about it is the simplicity of adding 1. None of this fuss about carrying 1 that you get when you have to add 1 to 99. Just put down an extra stroke.

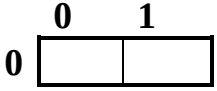
Of course the unary system would be totally impractical. Imagine the date 13-5-21 expressed in unary:

1111111111111-11111-111111111111111111111.

But we're not concerned with practicalities here, just whether or not something is possible.

§7.8. Turing Programs with One or Two States

We'll calculate the value of B(1) to illustrate what would be needed in a program for calculating the Busy Beaver Function. If a Turing Program has just one state, the initial state 0, the whole program can be written in a table with one row and two columns:



Each of the two cells in this table will contain an instruction. In this case there are only eight possible instructions: 0L0, 0L1, 0R0, 0R1, 1L0, 1L1, 1R0, 1R1. With eight possibilities for each cell, there are $8 \times 8 = 64$ possible Turing Machines. In terms of our story of the Turing Marathon, there will be 64 possible entrants in the 1-state division. But some of these will be disqualified.

Let's focus our attention on the first instruction. We can sort these 64 programs into eight groups of eight according to their first instruction.

A	0	1
0	0L0	

B	0	1
0	0L1	

C	0	1
0	0R0	

D	0	1
0	0R1	

E	0	1
0	1L0	

F	0	1
0	1L1	

G	0	1
0	1R0	

H	0	1
0	1R1	

Each of these partially completed tables represents eight programs corresponding to the eight different possibilities for the second instruction.

Now let's examine these eight tables in turn. Starting with a blank tape the machines in groups A, C, E, G will loop. The second instruction will never be reached. Machines in group A, for example, move continually to the left, leaving the tape blank. Machines in group E move continually to the left, leaving behind a trail of 1's. Machines in groups C and G exhibit similar behaviour to the right.

The remaining four groups, representing 32 programs in all, will halt at the very first step. So the longest number of steps that a 1-state program can run for, starting with a blank tape, and still halt, is 1. Thus B(1) = 1.

It's much more difficult to calculate B(2). For a start there are 12 possible instructions now: 0L0, 0L1, 0L2, 0R0, 0R1, 0R2, 1L0, 1L1, 1L2, 1R0, 1R1, 1R2. And there are now four cells in which to put them. So there are $12 \times 12 \times 12 \times 12$ such programs in all. That's 20736 programs to consider.

I once set this as a problem for a post-graduate course on the Theory of Computation. With the help of a computer program that they had to write, they analysed these cases and concluded that B(2) = 6. No 2-state Turing machine program will halt after more than 6 steps, but there are some 2-state programs that halt after exactly 6 steps. They are the joint winners in the 2-state division of the Turing Machine Marathon. Here is one of them.

	0	**1**
0	0L1	1L2
1	1R0	1L1

Not only did my students able to come up with 2-state programs that halted after 6 steps, they also had to show, by an analysis of all the others, that any

program that was still going after six steps would go on forever.

A two-state analysis was difficult enough so it would appear that it would be very difficult to write a program that would handle the general case. Very difficult, but is it actually impossible? To say that it is would seem to limit the ingenuity of man (or woman). Yet the ingenuity of the human mind has limits – a rather humbling thought. At least we are ingenious enough to recognise our own limitations.

We can never write a program to compute the Busy Beaver Function, but at least we can prove that we can't, which is perhaps the next best thing.

The Busy Beaver Function can never be computed. It might be possible to find the values of B(3), B(4) and so on, but the methods would be forever changing. No one set of ideas can handle all B(*n*)'s. Why not? Read on!

§7.9. Why B(*n*) is uncomputable

The busy-beaver function is uncomputable. That is, there is no Turing Machine which computes B(*n*) for all *n*. Nor could one ever be found. It's a logical impossibility. What's more, the fact that no such Turing Machine can exist means that no program can ever be written in *any* computer language on *any* computer – not now, not ever.

If anyone is ever clever enough to do so, such a program can be converted to a Turing Machine and he or she will have created a logical impossibility. Our whole world of logical reasoning will collapse!

Our proof will be a proof by contradiction. We suppose that there *is* a Turing program **BEAVER** which computes B(*n*). That is, if we input the number *n* by writing *n* 1's on the tape, the output will be B(*n*) 1's. Both input and output will be in unary notation.

With this supposedly-existing program, together with two other programs that *do* exist, we construct another program. The two auxiliary programs are **INCREMENT** and **DOUBLE**.

INCREMENT is a 2-state program that computes the function F(*n*) = *n* + 1. In unary notation, this is very easy to do. We simply put down one extra 1.

INCREMENT	0	**1**
0	1 L 1	1 L 0
1	0 R 2	

The other auxiliary program is **DOUBLE**. It's a 9-state program that computes the function G(*n*) = 2*n*. It takes a string of 1's, representing the input *n*, and joins a second copy onto it, making a string of twice the length. This is quite tricky, because after we've copied a 1 we have to mark it in some way to avoid copying it

again. We do this by temporarily changing the 1 to a 0. After the head has move across to put down the copy and comes back, it can recognise where the 1 came from. It then reinstates the 1, moves to the right and proceeds to copy the next 1.

If you have the patience it's interesting to work through this program, say with an input of 3. That is, the tape consists of 111 on an otherwise blank tape and the head begins on the left-most 1.

If you can't be bothered working through it you can just accept that such a program is possible.

DOUBLE	**0**	**1**
0	0 L 5	0 R 1
1	0 R 2	1 R 1
2	1 L 3	1 R 2
3	0 L 4	1 L 3
4	1 R 0	1 L 4
5	0 R 6	1 L 5
6	0 R 9	0 R 7
7	1 L 8	1 R 7
8	0 R 9	1 L 8

We now take as many copies of **DOUBLE** as we like and build up a Turing program called **OMEGA**.

OMEGA
INCREMENT
DOUBLE
DOUBLE
.....
DOUBLE
BEAVER

How many states will this program have? Well, that depends on how many copies of **DOUBLE** we're taking. Suppose we take n copies. Each copy has 9 states, so that's $9n$ states, plus 2 for **INCREMENT** plus however many states this mythical **BEAVER** has. Since we don't have a **BEAVER** program, we can't count them, but if **BEAVER** exists its number of states must exist. Let's suppose there are b states in **BEAVER**. So **OMEGA** with n **DOUBLE**'s will have $9n + b + 2$ states.

Now what will **OMEGA** do with a blank tape? Well, first it will add 1, to get 1, and then it will double that n times. At this point there will be 2^n 1's on the tape. If $n = 4$ we'll have doubled the 1 four times to get $2^4 = 16$.

At this stage **OMEGA** hands over control to **BEAVER**, which will take the 2^n as input and proceed to compute $B(2^n)$. So at the end of the day, starting with a blank tape, **OMEGA** will halt, leaving $B(2^n)$ 1's on the tape.

.steps because it takes one step to put down each 1. Suppose it runs for *s* steps. Then *s* is at least as big as B(2^n). So $\boldsymbol{s \geq B(2^n)}$.

Now **OMEGA** is itself a Turing program, with $9n + b + 2$, states, and it halts. So it can't run longer than the maximum for all programs of its size. Hence *s*, the number of steps that **OMEGA** runs for must be less than or equal to B($9n + b + 2$), the maximum for programs in the same class as **OMEGA**. This means that $\boldsymbol{s \leq B(9n + b + 2)}$.

Perhaps you need a breather at this point. We're establishing a number of inequalities which are probably more easily considered using symbols. Let's recap.

We have:

s = number of steps that **OMEGA** (with *n* doubles) runs for

B(2^n) = number of 1's that **OMEGA** prints

B($9n + b + 2$) = maximum number of steps that any program as big as **OMEGA** can run for (and still halt).

$B(2^n) \leq s$

OMEGA must run for at least as many steps as the number of 1's it prints.

$s \leq B(9n + b + 2)$

B($9n + b + 2$) is the maximum for programs in the same class as **OMEGA**

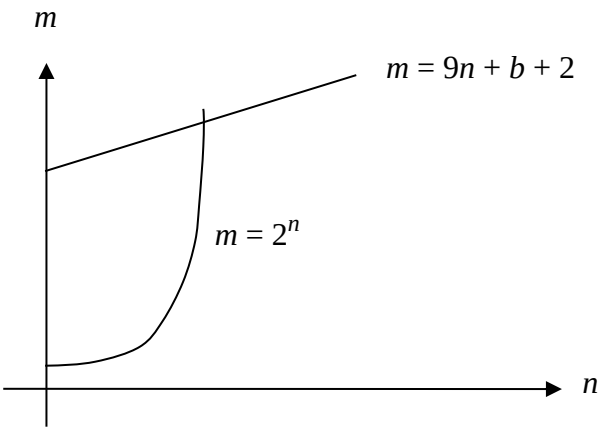
Combining these inequalities together we get:

$$\mathbf{B(2^n) \leq B(9n + b + 2)}$$

and the final contradiction is just around the corner.

Up till now we've not been particular about the size of n. Any n would have done. But now we want n to be large. How large? Well, we want n to be large enough so that 2^n is bigger than $9n + b + 2$. Unless we had a specific value for b we could never say explicitly how large we'd need n to be. But 2^n grows *exponentially*, and no matter how large b is, eventually 2^n would exceed $9n + b + 2$.

For example if $b = 100{,}000{,}000$ a value of $n = 27$ would be large enough. The important thing is not to calculate how big n would need to be, but in recognising that no matter how big b is, there will always be a suitably large value of n.



OK, so we choose a value of n that makes 2^n bigger than $9n + b + 2$. This will mean that $B(2^n)$ is bigger than $B(9n + b + 2)$. (Remember the more steps available the longer one can make the program run.)

$$\mathbf{B(2^n) > B(9n + b + 2)}$$

But, and here's the contradiction, we showed above that no matter how big n is,

$$\mathbf{B(2^n) \leq B(9n + b + 2)}.$$

These two inequalities contradict one another. The only way to resolve this contradiction is to deny the only unsubstantiated assumption we've made — the existence of **BEAVER**. Therefore no such program can possibly exist and so the Busy Beaver function is uncomputable. For each n there must be a value of $B(n)$ and we may be able to find out what some of these values are. But a uniform, systematic procedure, that will work for *all* n, has just been proved to be impossible.

§7. 10. Busy Beaver and the Halting Problem

We've given independent proofs of the Unsolvability of the Halting Problem and the Uncomputability of the Busy Beaver function. Actually, each could have been proved from the other.

If the Halting Problem did have a solution, that is if we had a program like **PREDICTOR**, we could calculate B(*n*) very simply, as follows:

(1). Go through the *n*-state programs, one by one and run **PREDICTOR**. This will tell us which machines to disqualify.

(2). Then we simulate the remaining ones, keeping a track of how long each one runs for. Since we can guarantee that these remaining candidates will all eventually halt, this procedure will terminate in a finite time.

(3). Finally we run through our record of how long each machine lasted, and take the maximum. This will be B(*n*).

The fact that we've shown that it's impossible to compute the Busy Beaver function shows that our assumption that we had solved the Halting Problem must be false.

Now suppose that we *did* have a program which could calculate the Busy Beaver function. We would then be able to solve the Halting Problem as follows:

(1). Given a program, count the number of states, *n*.

(2). Use **BEAVER** to compute B(*n*).

(3). Simulate the program for the first B(*n*) steps.

(4). If it halts within the first B(*n*) steps the answer is that the program will halt.

(5). If it hasn't yet halted by the B(*n*)'th step we'll know that it can never halt, because B(*n*) is the maximum number of steps for halting programs of that size.

The fact that we've shown that the Halting Problem is unsolvable gives us a contradiction and hence proves that **BEAVER** could not exist. So we've just shown that **BEAVER** exists if and only if **PREDICT** exists. Proving that either one cannot exist is sufficient to show that neither exists. In fact we've given independent proofs for each, so in a sense, each is doubly proved.

Not that a second proof increases the reliability of our claim. A proof is a proof is a proof. But the different methods employed in these two independent proofs are interesting and instructive.

There are many other computer programs you'd be wise not to waste time trying to write. A program that takes as input any two programs and determines whether or not they are equivalent, is such an impossibility. It would be nice to have such a program, particularly when marking students work in computing classes. If your program is equivalent to the tutor's then it's correct. We could leave it to the computer to

decide. Such computer marking of programs is actually used in practice, but they are all limited in their performance.

No program can possibly test equivalence in all circumstances. Why? Because it has been shown that such a program, if it existed, would lead to a solution of the Halting Problem. And since there is no solution to the Halting Problem there cannot exist a solution to the Equivalence Problem.

SCIENTIFIC ARTICLE:
AMITERMES LAURENSIS

[This scientific article begins as an accurate account of a species of termite but, towards the end, it becomes somewhat fanciful in order to tie in with the material of the previous chapter. The reader must decide where fact gives way to fantasy.]

Termites are found in many countries of the world, notably in Africa and Australia. The aboriginal word for termite is ‘ngartdan’.

The Amitermes Laurensis is a species of termite that builds mounds in the Northern Territory of Australia. They occur in Cape York Peninsula and eastern Arnham Land. In Queensland, north of the township of Laura (hence the name of the species), these mounds are built as thin flat plates, oriented in a north-south direction. South of Laura the mounds are conical.

Termites are sometimes referred to as “white ants” though ants and termites come from quite different insect groups. In fact termites are more closely related to cockroaches than to ants. Ants have elbowed antennae and a waist while termites have antennae like strings of beads and no waist.

However, like ants, termites are social insects and have a caste system. There are the reproductives,

the workers and the soldiers. The latter two castes have neither eyes nor wings.

The mounds, which can be up to ten metres tall, are highly organised “cities” with areas for different activities. The reason for the distinctive north-south shape of the Amitermes Laurensis species is to maintain a comfortable interior temperature. In the mornings the large, flat, eastern face gets the sun while the western face remains several degrees cooler. The majority of the colony is to be found on the cooler side in the mornings. In the hottest part of the day the sun shines directly only on the northern edge, helping to keep the mound cool.

Most termites eat wood. They can hollow out large branches and this is the source of the hollow tubes from which the Aborigines make didgeridoos. However the Amitermes Laurensis feed on grass and a single colony of them can process more grass than a large grazing animal. Moreover they are much more efficient in processing biomass than cows or sheep. They are probably the most efficient life-form on the planet for extracting energy from plant material. It is estimated that termites can turn a single sheet of paper into two litres of hydrogen.

The lignocellulose polymers are firstly broken down into simple sugars and hydrogen by fermentation in the termite’s gut and then other bacteria transform these sugars into energy. Because of their efficiency in producing so much energy from a single kilogram of biomass they may one day help to solve the world’s energy problems.

The height that a termite mound can reach is determined by the number of termites in the colony, but it is not a simple linear relationship. A colony twice as large as another would not produce a mound twice as tall. Hence there is a mathematical function, called the Busy Termite Function. T(n) is the height in millimetres of a mound that a colony of n termites can build.

Though we can determine T(n) for specific values of n, by measuring termite heights and estimating the size of the colony, it has not been possible to obtain a mathematical formula for it.

This information is of interest not only to scientists studying termites, but also to the termite colonies themselves. They need to know, for example, whether they should continue to grow as one colony or to split into two. What is surprising is that termite colonies can compute the termite function, in a crude way.

Every termite mound is, in effect, used as a computer for this purpose just as Stonehenge was a computer for making simple astronomical predictions. Indeed a termite mound could be called ‘Sandhenge’ in view of the fact that the mound is constructed from particles of sand, held together by termite saliva. It is ironic that such silicon based computing was going on long before the invention of the silicon chip.

Each termite can hold eight grains of sand in its mandible, or bite, and this can represent a number from 0 to 255. The exact process by which the termites cooperate to perform the Busy Termite program has yet to be discovered. What we do know is that the steps, which must be genetically programmed in their DNA, cannot compute the Busy Termite function, T(*n*), for *all* values of *n* because it has been proved that this function is non-computable!

8. THE UNDECIDABLE

§8.1. Axiomatic Systems

As we have said, mathematical truth is established by logic, starting with some fundamental assumptions called axioms. One is obliged to accept the conclusions provided one accepts the logical principles used as well as the axioms. There is a real sense in which a set of axioms is a creed, like a religious creed.

Euclid is credited with devising the first set of axioms – the axioms for Geometry or, as we now consider it, the axioms for Euclidean Geometry. These axioms were considered to be 'self-evident'. Axioms such as "between any two distinct points there is exactly one straight line". Far from being self-evident, this is based on experimental evidence and has the same status as a scientific 'fact'.

Axioms for other mathematical systems were proposed in the late 19th century. The first were the axioms of Group Theory. Never mind what it is or what its axioms are. Rather than self-evident truths they were considered to simply make up a definition of a group.

These days there is much controversy about gay marriage. Some regard it as self-evident that 'marriage' means an arrangement between a man and a woman. In

fact, it is merely the definition of the word ‘marriage’. Certainly there’s no doubt that this is what was implied by the word over centuries. Others say the definition should be broadened. There’s a long history of the meaning of words being broadened. ‘Money’ once referred to what we now call ‘currency’ – coins and notes, but the meaning has been broadened to include electronic transactions. That doesn’t mean that the meaning of ‘marriage’ *should* be broadened. There are strong arguments on both sides. The point I’m making is that each person who uses the word ‘marriage’ should be prepared to state their definition.

The attitude towards Euclid’s axioms changed in the eighteenth century. They were no longer considered to be self-evident, but merely part of the definition of a particular geometry called Euclidean geometry. Other, slightly different, sets of axioms were set up for other geometries. From a mathematical point of view all of them are correct. It’s up to the scientist, the physicist, the cosmologist, to decide which is correct for our universe. And the jury is still out on that question.

A rather different state of affairs exists for Set Theory. A ‘set’ is a collection of things. In Axiomatic Set Theory these things are mathematical objects. Now unlike Group Theory, where there are lots of systems satisfying the axioms, in Axiomatic Set Theory we’re attempting to describe a concept that we hold intuitively.

§8.2. The Russell Paradox

Set theory has come to underlie all of mathematics, so in a sense it is the foundation for all mathematics. Up to the end of the 19th century it was considered that the truths of set theory were self evident, just as we don't fuss too much about the logic we employ. One of the assumptions is that for any property that things might have there is a corresponding set, consisting of all the things that have that property. This is the process of turning an adjective into a noun. 'Black' is an adjective, so there is the set of all black things. But the philosopher Bertrand Russell, who was interested in the foundations of mathematics, pointed out that the set of all sets that do not belong to themselves is self-contradictory.

Perhaps a bit of notation will help us to understand this. The fundamental property of sets is that things belong to them. We denote the fact that the thing x belongs to the set S by the notation $x \in S$.

If P is a property, like being black, and x is a thing, we denote the statement that x has the property P by Px. So if c = a crow and Bx = "x is black" then Bc is a true statement, while Bd is false if d = a dove. Crows are black but doves are not.

The set that corresponds to the property P is denoted by $\{x \mid Px\}$. Read it as "the set of all x such that

Px (or Px is true). The naïve assumption was that for all properties P there must be a set $\{x \mid Px\}$.

Russell considered the property of something not belonging to itself – in the sense of set belonging. Here the something is a set. A set can belong to another set because it is possible to have sets of sets, or sets of sets of sets ....

If T is the set of all pairs of distinct whole numbers then the set consisting of just 3 and 5 would belong to T. The symbol for "not belonging" is $\notin$, just like the symbol for "not equals" is obtained by crossing out the equals sign, as in $\neq$.

Now some sets clearly don't belong to themselves. The set of all positive numbers is not a positive number. But if there is such a thing as the set of all sets, then it belongs to itself.

So Russell said, what if $S = \{x \mid x \notin x\}$? This would be the set of all sets that are not members of themselves. This would be the case for most sets we might think of.

The set of all even numbers is not an even number. The set of all triangles is not a triangle. But the set of all infinite sets, if there is such a set, is itself an infinite set.

You may wonder why I keep saying, "if there is such a set". I will discuss this later.

The question is:

Does *S* belong to *S*?

Clearly the answer would have to be either "yes" or "no", but let's consider each possibility in turn.

SUPPOSE that $S \in S$.

Then it must satisfy the corresponding property, that is $S \notin S$. This is a contradiction.

SUPPOSE that $S \notin S$.

Then S satisfies the property that defines S and therefore $S \in S$. Again, a contradiction.

This seems like one of those logical paradoxes like the sentence "THIS SENTENCE IS FALSE". But we can't ignore it. Under our naïve concept of set theory such a set exists. If we want to ban it from being a set we'd better explain to it why it's being kicked out!

This may also remind you of the argument from the chapter on the uncountable. The difference is that in that case there was an assumption that led to the contradiction. If one can find a different chairman for every committee then we get a contradiction. Therefore it is impossible to provide a different chairman for every committee. It is false that there is the same number of subsets as elements.

But with the Russell Paradox there appears to be no such initial assumption, apart from the intuitively obvious 'fact' that for every property there's a set of all things with that property. Well, then, intuitively obvious or not, this assumption has to go.

Here we have a fundamental contradiction in set theory. And since we want to build our mathematics on the foundation of Set Theory, all of mathematics would fall to the ground if we didn't remove such a flaw. If you allow a single contradiction into mathematics you can prove anything.

I remember one of my lecturers telling me this and when someone asked him to prove that he was the Pope, assuming that 0 = 1, he said, "If 0 = 1 then, adding 1 to both sides, we conclude that 1 = 2. The Pope and I are two people, so therefore the Pope and I are the one person. QED."

Well, you can imagine the fuss that Russell's Paradox caused when it was first announced. At least it caused a fuss amongst those who were bothered about the foundations of mathematics. Ordinary working mathematicians just said, "oh, that's interesting" and went back to their work. They knew that someone would fix up the problem, and that they did.

The way of fixing up the problem was to set up a collection of axioms that made some restrictions on which properties *do* lead to a set. There have been

several formulations but they have all been proved to be equivalent to one another. The most well-known set of axioms are called the ZF axioms, after their proposers Zermelo and Fraenkel. I won't list them here because they're long and sound quite technical. Basically they mostly say that "if such and such is a set the so and so is a set". They are all dependent on already having some sets with which to make other sets – except for the first axiom, the existence of the **empty** set.

The empty set is the set with no elements. It doesn't matter what the no elements are. The set of unicorns is the same empty set as the set of elves or the set of whole numbers lying strictly between 1 and 2. Axiom 1 in the ZF system says: There exists a set corresponding to the property $x \neq x$, that is $\{x \mid x \neq x\}$ exists. The symbol for the empty set is $\varnothing$. Now you might be thinking that is silly to have a set with nothing in it.

"Oh, I have a collection of vintage Rolls Royce automobiles."
"Wow! How many have you got?"
"Oh, it's the empty set."

Stupid as it might seem, where would we be without the number zero? For centuries zero was never considered to be a number. Why have a number for something that doesn't exist. Yet, our modern system of notation for numbers relies on having zero. The

difference between my bank balance and that of Bill Gates is just a whole lot of zeros!

Now there's something rather delightful in the fact that all of mathematics can be manufactured from the empty set. First there's the set {∅} that contains just the empty set. It isn't the empty set itself because it *does* have something in it, even though that something is empty. Then there is {∅, {∅}}. This set contains two sets, the empty set itself, and the set consisting of the empty set. It might seem that we're splitting hairs here, but the distinction between ∅ and {∅} is important. In fact, when the number 2 is defined it is defined in this way of developing mathematics, it is the set {∅, {∅}} and 3 is {∅, {∅}, {∅, {∅}}. If this seems a rather esoteric way of defining the number 3, let me ask how you would define it. I'm sure what you might come up with would be more intelligible to a typical kindergarten pupil than {∅, {∅}, {∅, {∅}} but it wouldn't stand up to the high standard of rigour that professional mathematicians require.

You might say that this shows that God created mathematics. Just as God created the world from a void He created the whole of mathematics from the empty set! On the other hand, if you are somewhat of an atheist, at least you'll find a resonance between mathematics created from the empty set and the big-bang theory of how the universe began.

§8.3. Axioms for Mathematics

Almost all of mathematics can be built up from the following axioms. They are called the **Zermelo-Fraenkel Axioms**, or **ZF** for short. Other foundations have been suggested, but they are all equivalent to the ZF creed. For 'creed' it is – just as a religious creed. They are statements whose truths are taken without proof. One just has to believe in them. Remember that it is not possible to prove something from nothing.

In addition, there are assumptions about logic, we would be considering logical axioms as well. These will regulate the use of words such as 'and', 'or' and 'implies'.

Six of the eight ZF axioms are:

Equality: Two sets are equal if they have precisely the same elements.

Empty Set: There is a set with no elements.

Pairs: If S, T are sets there is a set with just S and T as elements.

Powers: If S is a set so is the set of all subsets of S.

Union: If S is a set so is the set of all elements of elements of S.

Specification: If S is a set and P is any property that can be expressed entirely in terms of set membership, then there is a set whose elements are precisely those elements of S for which the property holds.

The other two axioms are a bit more technical, so we'll omit them. A full discussion can be found in my notes on Set Theory on the website coopernotes.net. On the basis of these eight axioms virtually the whole of mathematics can be built.

So can we now be assured that no further contradiction, like Russell's Paradox will arise? This amounts to asking whether the ZF axioms are consistent. The slightly disturbing answer is that no, we do not know that they are consistent. Most mathematicians believe that they are, but most mathematicians also believe that we'll never be able to *prove* consistency.

§8.4. Consistency

A set of axioms is **inconsistent** if a contradiction can be validly derived from them. If it is not inconsistent then it is defined to be **consistent**. The easiest way to prove consistency is to come up with a model for the axioms, that is, an actual interpretation that satisfies all the axioms.

It's easy to come up with an inconsistent set of axioms. For example consider the following axioms for a *super number*. The set of super numbers has two operations, called addition and multiplication, such that the following axioms hold.

Axiom 1: There is a super number 0, such that:
$n + 0 = n$ for all super numbers, n.
Axiom 2: There is a super number 1 such that:
$1 + 1 \neq 1$.
Axiom 3: $(x + y)z = xy + xz$ for all super numbers x, y and z.
Axiom 4: There's a super number ∞ such that $0\infty = 1$.

This system of axioms is inconsistent. Here's a proof.

By axiom 1, $0 + 0 = 0$, and so $(0 + 0)\infty = 0\infty$.
By axiom 3, $0\infty + 0\infty = 0\infty$.
By axiom 4, $1 + 1 = 1$, contradicting axiom 2.

Here's another rather exotic axiomatic system that I've constructed to illustrate the concept of consistency. I call the system a **society**. In a society there's a set of undefined things called **persons** and three undefined relations:

father of,
mother of,
married to

Now the terminology suggests we're thinking of family relationships, and certainly that's what inspired these axioms. But it must be emphasized that these things called 'persons' are to be considered as undefined and so we must not make any use of what we know of actual family relationships.

We assume the following axioms:

Axiom 1: There exists a person.
Axiom 2: Each person has a unique mother and a unique father.
Axiom 3: If two people have the same mother then they have the same father.
Axiom 4: The mother and father of every person must be married.
Axiom 5: If two people have the same father they can't marry.

Suppose we define a **parent** to be a 'person' who is either a mother or a father and a **grandmother** to be the mother of a parent.

Theorem 1: Every person has exactly two grandmothers.
Proof: Let Peter be a person.
By axiom 2 Peter has exactly one father, who we'll call Frank, and exactly one mother, called Michelle. By axiom 4, Frank is married to Michelle.
Suppose Frank = Michelle. Then by axiom 4, Frank is married to himself, contradicting axiom 5.
Hence Frank ≠ Michelle.

By axiom 2, Frank has exactly one mother, denoted by Mildred and Michelle has exactly one mother, denoted by Mary.

Suppose Mildred = Mary. That is, suppose Frank has the same mother as Michelle. Then by axiom 3 Frank and Michelle have the same father, denoted by Phillip.

By axiom 5, Frank and Michelle can't marry, contradicting what we proved earlier.
Hence Mildred ≠ Mary and so Peter has exactly 2 grandmothers.

Notice that we proved the theorem only using the axioms, and without appealing to our intuition, or knowledge of society. Now are these axioms consistent? There's no point in proving theorems for a

non-existent system. To do this we need to devise a model – a concrete example in which these axioms hold.

Here's a different model for this system. A 'person' is one of the positive integers 1, 2, 3, … The father of n is $2n$ and the mother of n is $2n + 1$. Person m is married to person n if $m + n$ is odd. This system we shall call a 'society'.

Axiom 1 and **Axiom 2** are clearly true.

Axiom 3: If m and n have the same mother then $2n + 1 = 2m + 1$ and so $2n = 2m$, which means that they have the same father.

Axiom 4: The father and mother of person n are $2n$ and $2n + 1$. Since their sum is odd, they are married.

Axiom 5: If m and n have the same father the $2m = 2n$ and so $m = n$. Thus m + n is even and so they can't be married.

The fact that a model exists for a society, means that the axioms are consistent. But societies as described by these axioms can be very different to the model I had in mind when I devised the axioms. For, in the arithmetic model, any even number is married to every odd number, since their sum is odd. What an infinitely bigamous society! And every person has only one child!

In Axiomatic Set Theory we often consider extra 'optional axioms'. We could add optional axioms to make it more like the society of real people and their families. But we would have to be very flexible, because there some rather strange family relationships in today's society.

§8.5. The Axiom of Choice

Now, what's really interesting is that there a few things that can't be proved from the ZF axioms which most mathematicians believe are true. One of these is the Axiom of Choice, abbreviated to AXC. In a nutshell the AXC says that if you have a whole bunch of non-empty sets you can simultaneously choose one thing out of each of them. This seems an obvious enough statement but, remember that it says that this is possible, even if the sets are infinite and even if there are infinitely many of them.

Of course such a choice is impossible in practice because it would take infinite time, but we're not talking about 'in practice'. The question is, does such a choice exist and can they choices form a set? (The last question is not quite the one that is asked, but it's near enough for our purposes.)

The Axiom of Choice has been proved to be **consistent with**, and **independent of**, the ZF axioms. To show this you assume the ZF axioms and construct a model in which not only the ZF axioms hold, but also

the Axiom of Choice. That's the 'consistent with' part. Then you construct a different model, with a different definition of 'belonging to' that satisfies the ZF axioms but does *not* satisfy the Axiom of Choice. That's the 'independent of' half of the statement. Putting these halves together we come up with the statement:

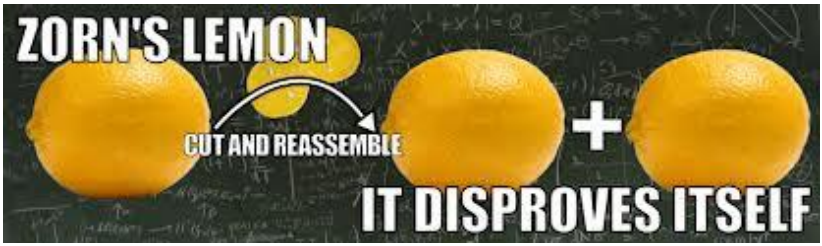
THE AXIOM OF CHOICE IS UNDECIDABLE.

This means that, assuming the ZF axioms are consistent, you'll never be able to prove that the AXC is true. But nor will you ever be able to prove that it's false. If ever a contradiction arises in mathematics when using the Axiom of Choice it won't be the fault of that axiom. It will mean that an inconsistency will have been found in the ZF axioms themselves. If ever a contradiction arises from denying the Axiom of Choice it will mean that the ZF axioms themselves are inconsistent, not the denial itself.

The bottom line is that you are free to choose! You can believe in AXC or not. Both positions are logically valid. Naturally, like most mathematicians, you will no doubt opt to believe in AXC. It sounds so plausible. But before you become a paid-up member of the Axiom of Choice religion, let me point out the following consequence of the Axiom of Choice.

It has been proved, assuming the ZF axioms, together with the AXC, that in principle it's possible to take a solid ball and dissect it into several pieces and to reassemble the pieces to make *two* solid balls *of the same size as the original one*!

Your reaction to this is probably to say that this proves that the AXC is false. After all, such a situation would contradict the law of conservation of volume,



surely. If you take a piece of wood its volume would remain constant no matter how you cut it up and reassembled the pieces. That is, ignoring the sawdust which, of course, we're doing.

However the law of conservation of volume only applies if the pieces have a defined volume. If a set of points is highly fragmented, like a cloud of infinitely small particles, then it's not possible to define its volume.

The way of dissecting the original sphere and reassembling them is not something one could replicate, even with precision tools. If it *was* possible to convert one ounce of gold into two with a laser

cutter, the price of gold would plummet! But the ‘pieces’ that are required to perform this magic are so highly fragmented that their volumes don’t exist.

Needless to say, while most mathematicians are happy to accept the Axiom of Choice because it simplifies the statements of many of their theorems, there’s a determined minority who reject it. A comforting thought, though, is that no bridge will ever fall down because its engineer believed or didn’t believe in the AXC.

The difference between believing or not believing the Axiom of Choice is more aesthetic than practical. In this sense it’s rather different to a religious belief. The Axiom of Choice believers will never wage war on the infidels, and no mathematician will become a martyr to his or her belief.

The general consensus is that one should try not to use the Axiom of Choice, but if necessary one uses it, and admits that it is “on the basis of the Axiom of Choice”.

§8.6. The Peano Axioms

The very first mathematical system we encountered was the system of the natural numbers:

$$0, 1, 2, 3, \ldots$$

When we did so, in kindergarten or even before, we were not interested in precise definitions. We learnt the many properties of natural numbers on the authority of our parents and teachers. Nowhere did we see a definition of the number 2, or a precise proof of the fact that $2 + 2 = 4$. We might have experimented with a few pairs of objects and observed that combining one with another we got a collection which, when we counted, gave us 4. Hence we learnt our mathematics as an experimental science.

Of course we did notice that sometimes it didn't work. Pour a litre of water into a bowl containing a litre of sugar and you'll find you get a whole lot less than a litre of sugar syrup. This can be explained, in part, by the air spaces between the grains of sugar, but to account for the reduction in volume completely you need to take the chemistry of solutions into account.

Nevertheless you understood that something different is going on here and that $1 + 1 = 2$ is still valid mathematically.

One approach to constructing the natural numbers, and their arithmetic, rigorously is to build them up as sets of sets of sets within axiomatic set

theory. Another approach is to define them by a set of axioms, the Peano Axioms.

We postulate a set of undefined things, together with an undefined function 'successor'. You can think of the 'successor' of n as $n + 1$, written n^+, but that interpretation isn't specifically part of the axioms.

Axiom 1: 0 is a natural number;

Axiom 2: If n is a natural number then so is its successor n^+;

Axiom 3: There is no natural number n for which

$$n^+ = 0;$$

Axiom 4: If S is any set of natural numbers that contains 0, and contains n^+ whenever it contains n, then S is the set of all natural numbers.

On the basis of these axioms we can define addition and multiplication and prove the basic properties of arithmetic.

§8.7. Gödel's Incompleteness Theorem

We've seen how mathematical systems, such as Set Theory, can be built up on the basis of a set of axioms. Provided that a set of axioms is consistent we can prove meaningful theorems about the system. But can we prove every true statement from the axioms? If

we left out one of the set theory axioms there would be true statements about arithmetic that couldn't be proved. A set of axioms is complete if every true statement about the system can be proved. Are the ZF axioms complete?

The answer is no. Well, then, we'd better add some extra axioms to make it complete. Unfortunately that's not possible. In 1931 Kurt Gödel proved that, not only are the ZF axioms incomplete. No set of axioms can be constructed for which they will be complete. What's more it is not possible for a finite set of axioms to exist for any formal system in which basic arithmetic can be formulated, such that the axioms are complete.

He did this by converting every statement in such a system into an arithmetic statement. He managed to express to express the statement "this statement cannot be proved from the axioms" as a statement about arithmetic. Such a self-referential statement cannot be proved from the axioms, yet it is a true statement and corresponds to a true statement about arithmetic.

Gödel's original proof is very long, and very hard to read. A much simpler proof by Nagel & Newman in 2001 converts the statement to one about computability, and uses the machinery of Turing Machine to show that completeness would imply that the halting problem could be solved, which we know is impossible.

So here we are left with this unsatisfactory state of affairs. The ZF axioms on which the whole of mathematics can be built, cannot be proved to be consistent, but it can be proved to be incomplete. So it is possible that a contradiction could be deduced from these axioms. But if, as we hope, they *are* consistent, they are still incomplete. There are truths about arithmetic (though not ones we'd be ever likely to meet) that cannot be proved from any finite set of axioms! Mathematics is very far from being cut and dried.

At the heart of Gödel's proof is a very clever method for converting statements *about* the system into arithmetic statements *within* the system. For a start, statements are expressed symbolically, such as:

$$\forall x(-(x = 0) \rightarrow \exists y(xy = 1))$$

which means "for all x, if x is not equal to zero then there exists y such that x times y is equal to 1".

Gödel devised a system for coding these statements as a number by assigning a code to each symbol and building up a number for each statement on the basis of that. So, given a number n one could, if that n indeed represented a statement, decode it and so obtain the corresponding statement G(n).

Every possible statement would have a code, but not every code would correspond to a valid statement. The numbers involved would be extremely large, but as this is an 'in principle' exercise, that isn't a worry.

Now consider the statement that a given statement S is provable. A proof is just a list of statements, where each one is an axiom, or a previously proved theorem, or a logical consequence of the previous ones, and where the statement of the theorem is the last in the list. There's a mechanical way of testing the validity of a proof and so one could, in principle, write a computer program for testing whether a given statement is provable from the axioms. It would be a case of generating all possible lists of statements that have S as the last statement, and then testing the 'proof' for validity.

Gödel showed how provability could be expressed as an arithmetic statement about natural numbers and so the statement P(n) = "the statement with Gödel number n is provable" can be expressed as an arithmetic statement and so will have a certain Gödel number. Similarly, the statement N(n) = "the statement with Gödel number n is not provable" has a Gödel number, say g.

Gödel then asked whether N(g) is true or false. The statement N(g) claims that it, itself, is unprovable. Thus we can obtain, as a purely arithmetic statement, within the language of arithmetic, a statement which claims "I am unprovable". Now such a statement can't be false because being false would mean it was provable and hence true. It must therefore be true and hence it's a true but unprovable statement in arithmetic. But wait! Haven't we just proved that it is true?

Certainly we gave a meta-mathematical proof. But this proof is not one which could be expressed as an arithmetic proof within the system. Our unprovable statement is not unprovable in any absolute sense. It might not even be meaningful to talk about absolute unprovability. N(g) is unprovable in the relative sense that no proof of it could ever be constructed which starts from the axioms and proceeds using the rules of inference. And even if the axioms and rules were supplemented by others, so long as they remained finite in number, the existence of unprovable statements would remain.

JOKE: PALINDROME

An Englishman, an Irishman and a Scotsman go into a bar. An American, who was already in the bar comes up to the Englishman and says, “Hey buddie, if you can tell me a good joke I’ll buy you a beer”.

So the Englishman clears his throat and says, “37”. At this the bar erupts into an uproar of laughter. The American looks puzzled, but says, “well it appears that was a great joke, so what’ll you have?”

A little later the American goes up to the Scotsman and says, “I’ll buy you a whisky if you can beat that last joke”. The Scotsman stands on a stool, adjusts his kilt and says, “42”. Once again the bar erupts into laughter, even louder than before. Several patrons are so carried away by their laughter that they roll around on the floor. So the American buys the Scotsman a Scotch.

A little later the American turns to the Irishman and says, “You Irish are renowned for your wonderful humour. I bet you can top that last joke – if you do, I’ll buy you a pint of Guinness”.

So the Irishman jumps up on the bar, adjusts his cap, and says, “93”. There’s deathly silence. Not even a murmur is heard. The American looks puzzled.

“I’ve worked out that you folks must number your jokes so that all you have to do is to give the joke number and you all know what the joke is. But back in the States we tell our jokes in full. Now I’m a little

puzzled why that last joke fell so flat. What went wrong?"

"Ah", says the Scotsman, "you know what the Irish are like. They're always getting things back to front."

"Well", said the American, "would you mind telling me that last joke in full".

"Och, aye", said the Scotsman, "but are ye sure ye want to hear it. As I said it's not very funny".

"Well, yes", said the American, "I'm fascinated by British humour".

"OK", says the Scotsman, "Joke number 93 goes like this. An Englishman, an Irishman and a Scotsman go into a bar. An American, who was already in the bar comes up to the Englishman and says, "Hey buddie, if you can tell me a good joke I'll buy you a beer". So the Englishman clears his throat and says, "37".

9. THE INEFFABLE

§9.1. Proofs Of The Existence of God

We've finished the mathematical content of this book. This final chapter consists of some philosophical/theological musings that arise in some minds as a result of encountering those parts of mathematics that deal with the edge of the rational universe. If you have no interest in the fundamental questions of life then it's best that you skip this chapter.

Now I wish to make it clear that my purpose in writing these notes is to communicate what I see as the nature of mathematics, not to talk about religion. I once received an email from an angry reader who believed that this chapter was the 'pill' and the previous ones were the 'sugar coating' and that my whole aim was to sneak religion in under the radar.

If you are one of those who get angry at the mere mention of anything religious then you'd better not

read on. But let me emphasise that I am not here talking about the ineffability of God, but the ineffability of mathematics. I present some 'proofs' of the existence of God, merely as a vehicle for discussing bad, or erroneous, logic. But, by the same token, the fact that I expose such faulty reasoning does not mean that I am arguing *against* the existence of God.

I lay no claim to any professional expertise in either philosophy or theology. But I can't help going beyond the mathematics of the ineffable to the ineffable itself. The word 'ineffable' means 'inexpressible in words'. It's a word that not only appears in hymns, describing God, but also in a large number of nineteenth century novels. We don't use the word today, yet there's as much interest in the transcendental as ever.

There's a fundamental contradiction in the desire to discuss the ineffable – to say something meaningful about something that can't be expressed in words. But by a little distortion of the meaning we can think of the ineffable as that which transcends logic.

Can one prove that God exists? There have been many attempts over the centuries. A very simplistic argument, at least in the Christian tradition, runs as follows. The Bible says that God exists. The Bible says

that everything in the Bible is the word of God and so must be true. Therefore God exists. Put more simply it says "GOD EXISTS AND THIS STATEMENT IS TRUE".

One need not spend too much time in refuting this feeble argument. Just one word is needed – self-referentiality. It would have been far better if the Bible had made just two claims:

verse 1:	**verse 2:**
God exists.	Everything in this Bible is false.

If verse 2 is true then both verse 1 and verse 2 are false. But this leads to a contradiction. Oh well then, verse 2 must be false. So it is false that everything in this little Bible is false. So something must be true. It can't be verse 2 because we're assuming that it's false. It must therefore be verse 1 that's true. Therefore God exists!

This might seem momentarily convincing until we realise that any statement could have taken the place of 'God exists' – even 'God does not exist'. The

problem lies in the fact that verse 2 is 'self-referential' – it refers to itself.

One must refrain from considering any sentence that refers to its own truth. Such self-referential statements are meaningless and meaningless statements used in a logical argument can lead to paradoxes such as the above.

I remember, when training for my accreditation as a lay preacher many years ago, that I had to study many of these arguments – mostly with big names like 'the ontological argument'. I mostly forget what they were. One of the ones I do remember went like this. We define God to be a being that's perfect in every way. Now existence is more perfect that non-existence. So if God didn't exist this would contradict our definition. Therefore God exists.

The problem with this argument is that it assumes the existence of a being that is perfect in every way but who does not exist. The contradiction comes from assuming simultaneously the existence and the non-existence and has nothing to do with perfection.

We might define 'infinity' as "a number that's bigger than every number" and ask the question, "Does this infinity exist?" Well a number that doesn't exist can't be bigger than every number. (In fact a non-existent number can't be bigger than *any* number.) Therefore infinity must exist. But, of course, such an

‘infinity’ must be bigger than itself, a situation that is clearly untenable.

The explanation for this paradox is that ‘not existing’ is not a property of something. It is the absence of something with a given property. We could say that a non-existent number can’t be even. But nor can it be odd. The statement ‘*n* is even’ is not true or false of a non-existent number – it is meaningless. In the same way “God is perfect” is not true or false of a non-existent God. It is meaningless.

§9.2. Proof by Design

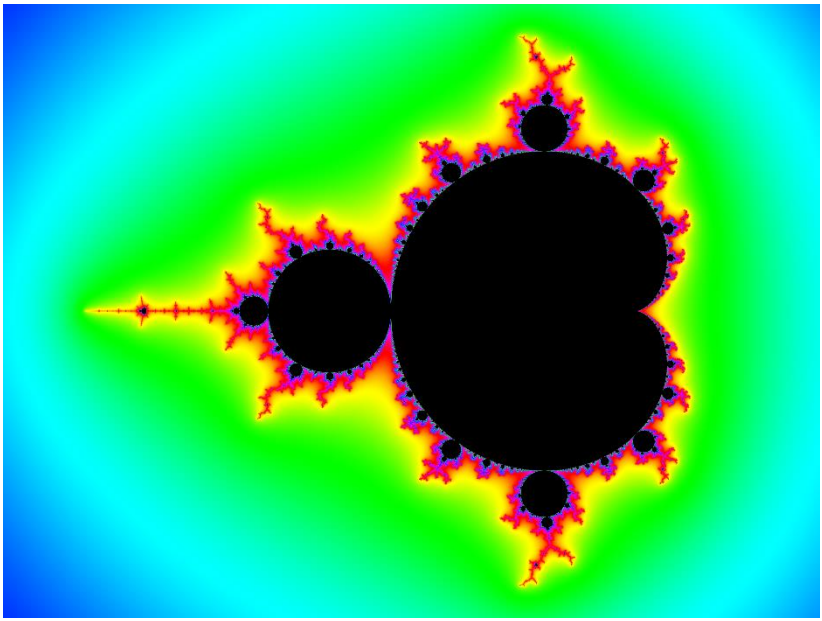
Another proof that God exists, one that was very popular in Victorian times, is Proof by Design. The world is a complex, finely balanced precision structure. If certain parameters were changed by only a small amount life would not be possible. It could not possibly have arisen by chance. There must have been an intelligent Creator. A watch could not come about by cogs just throwing themselves together. So the universe must have been created by a Divine Clockmaker.

But then along came Darwin and his Theory of Evolution. Then came chaos theory, and the theory of fractals and complexity. It *is* possible for complexity to arise from simple rules. This can occur in biology, with the amazingly complex variety of plant structure arising from a small number of biological rules. If God created flowers he didn’t do it the way an artist might

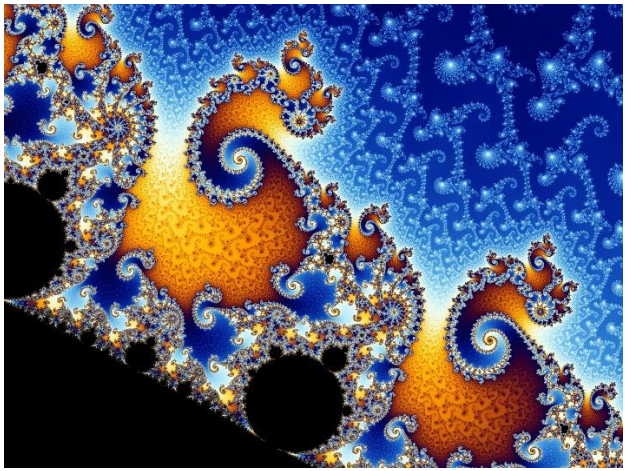
painstakingly paint a complicated picture. You can't argue if something is extremely complicated it must have been the result of an extremely brilliant maker. On the other hand you could argue, and many do, that to create life by a process of evolution that achieves complexity from a small set of simple rules is so brilliant that only a supreme mind could have thought of it.

§9.3. The Mandelbrot Set

You'll have to make up your own mind on this. Let me simply describe the most well-known example of complexity arising from simple rules. Perhaps you have seen pictures of the Mandelbrot set. It's a design, best seen in colour, which is startlingly beautiful. There is a boring bit in the middle, usually coloured black, and an outside that's almost as boring. It's the region between the two that is amazingly complex. If you zoom in to a part of the Mandelbrot set in this intermediate area you'll see complexity upon complexity. The more you zoom in, the more fascinating the pattern becomes. Parts of the pattern, when you zoom in, look very much like other parts at a lower magnification, but they're subtly different. Here is a view of the whole pattern.



Here is a magnified view of one portion.



This pattern is constructed by a computer program using a few very simple rules. The program

takes one point at a time on the ‘canvas’ and works out what colour to make it

It’s a very tedious process, and one would only attempt to do it with a computer program. One has to consider every point on the rectangular canvas. Well, actually there are infinitely many points in a given rectangle so one chooses a certain resolution and considers points in a rectangular grid, very close together. Usually one would make the distance between the points the same as the resolution of the computer screen, so that the pixels are considered one at a time.

Even so, this may involve many thousands of points. We also have to choose a scale, so that one unit, in each direction, is equal to so many pixels.

We choose at the outset, along with the scale, certain numbers N and R. You might, for example, choose N = 100 and R = 10. The number N represents the maximum number of steps we will perform for each point and R represents the radius of a circle whose centre is at the middle of the screen.

Each point is considered in turn. We move systematically so that at the end we’ll have considered every pixel on the screen. A certain calculation is carried out to determine what colour that pixel should be coloured and the image is built up in this way.

This calculation generates a sequence of points, though these points are not plotted. The sequence starts at the point whose colour we are determining. There's a very simple rule that calculates the next point in the sequence.

If a point in the sequence stays within the circle, of radius R, for N steps we colour that starting point black. If it breaks out of the circle we colour that original point some colour, depending on how long it took the sequence to 'escape' from the circle The points in the sequence don't get plotted, only the starting point for each step.

We decide on a certain palette of colours. We might choose red if the sequence breaks out of the circle after at most 10 steps, orange if it takes up to 20 steps, blue if it has broken out of the circle by the 30^{th} step, and so on.

Your decisions as to colours, as well as the choice of N and R, may make your picture look a bit different to mine, but the overall effect will be much the same.

A black and white version is far less interesting but still displays the enormous complexity of the Mandelbrot set. In this case, if the sequence breaks out of the circle after N steps it is coloured white. As with the coloured version, points whose sequence remains

within the circle for N steps, and quite probably will remain inside forever, are coloured black.

All that remains is to tell you how we go from one point to the next in generating each sequence.

If you are not frightened by a little bit of algebra, here's the simple rule to move a point one step to create the next point in each sequence. The centre of the canvas is the origin for the *x*-*y* plane, the point to which measurements of all other points are referred.

A horizontal axis through the origin is called the *x*-axis and the vertical axis through the origin is called the *y*-axis. A point has coordinates (*x*, *y*) if it is *x* units to the right of the origin and *y* units up. If *x* or *y* is negative then the point is to the left, or below (or both) of the origin.

The rule for getting the next point (X, Y), following the point (*x*, *y*) in the sequence, is:

$$X = x^2 - y^2 + a,\ Y = 2xy + b.$$

where (*a*, *b*) is the point whose colour we're trying to find.

We begin with (*x*, *y*) = (*a*, *b*).

If you know about complex numbers these equations can be expressed even more simply as:

$$Z = z^2 + a$$

where the points are represented by complex numbers and the complex number *a* represents the point we're starting the sequence with.

You can do these calculations on a spreadsheet, though to do a whole image would still be far too tedious. However there are websites on the internet where you can see the main image and where you can zoom in on a particular region, just like in Google Maps.

§9.4. Free Will

A fundamental prerequisite to having a religious faith is a belief in free will. If you don't have the freedom to choose to accept or reject the belief, what's the point? Mind you, this hasn't stopped people in some parts of the world forcing others to accept their faith at gunpoint.

A common argument against religion is to claim that you have discovered some psychological or biological cause for religious belief. Oppressed people, who have a miserable life, will believe in an after-life where pain and suffering and poverty will be no more. It's just wishful thinking.

Others claim to have discovered a God-gene, which explains why some of us have a religious belief and others don't. Still others claim that all emotions and all thought is purely a result of biochemical

processes. The human brain is influenced by its biochemistry and your belief in God can be explained by your diet, or your genetic makeup.

It's an interesting thought that a belief in a purely deterministic universe is a contradiction. If I assert that my thoughts are determined by the laws of physics, chemistry and biochemistry, then my assertion is also the result of such deterministic processes. In what sense could such a belief have any validity? The concept of truth presupposes that there's something beyond the material world. Otherwise what we might call true statements are just the babble of the mechanistic automata we call human beings.

But to say that there's something beyond the material world is a long way short of believing in any sort of God. It is intellectually respectable to be an atheist, but I fail to see how it could be considered intellectually respectable to believe that human beings are purely machines, with no free will.

Of course I can believe that *my* thoughts are valid truths, while *yours* are the result of deterministic processes. It's an interesting idea that I am the only reality, complete with free will, and the rest of the universe is simply an illusion.

Whatever we might claim to believe we all act as if we have free will. It might be a great illusion but if it is we'll never know it. But of course, as we all

know, free will is impacted upon by all sorts of external forces, and even internal, biochemical ones. No-one can claim to be completely free.

In *The Age of Reason* by Jean-Paul Sartre, the protagonist wants to be completely free. As a result he refuses to make any commitments, because that would limit his freedom. "If I marry her I'll remove, or at least reduce, my chance of marrying someone else." Every decision involves a certain reduction in freedom. Better not make any decisions.

So the person who's so determined to maximise his free will is forced to lock himself into a prison of indecision. He ends up with *less* freedom. Free will is a currency that must be spent or it becomes worthless.

I am now going to describe a demonstration that purports to prove that people don't have free will, even in a situation where there appear to be no external forces. Most people believe that they're completely free when they select their lotto numbers, although certain combinations get chosen less frequently than others because people believe that they're not random enough. Would you choose the numbers 1, 2, 3, 4, 5, 6? It's just as likely to come up as a more random sounding choice. The concept of random numbers is another area where there's an interface between mathematics and philosophy, but we'll not pursue this here.

This demonstration is designed to be performed in public, but you can just read about it and think about it. You have an audience from which you select five volunteers. You ask them to stand out the front in a line. Then you introduce the theme of free will as follows.

"Most people believe that they have free will. On being asked a certain question we might be influenced by certain facts, but if we have no facts, such as when choosing lotto numbers, we believe we can freely make up our minds." Check that your five volunteers agree with this. If any say no, you had better replace them.

"I'm about to give you each a card that asks a question about one of the other volunteers. The question won't even identify who that person is. At this stage they'll only be identified as person A, person B, and so on. You must freely choose an answer, YES or NO. It doesn't matter whether you're correct, or not, because after all you don't know yet to whom the question is referring. Oh, and you mustn't let anyone else know your question."

You give them cards, each of which has the same type of question:

Will person A give the correct answer to their question?

except that each person's question will refer to a different person: A, B, C, D, E.

You give each volunteer another card. On one side it reads YES and on the other side it reads NO, with these answers written large enough that the audience can read it. “Now I want you to display your answer by turning your card so everyone can see that answer.”

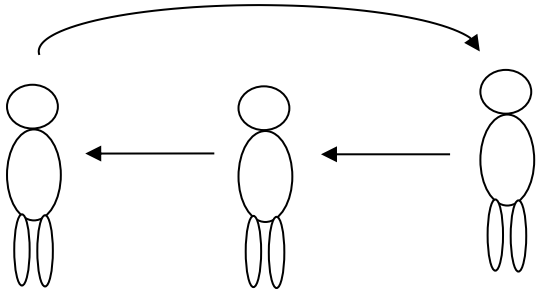
YES	**NO**

Your volunteers will display a sequence of YES’s and NO’s. Perhaps they will all be YES, or all NO. You must select three of the volunteers so that you have an **even number** of NO’s. Tell the other two to sit down. Do this as follows:

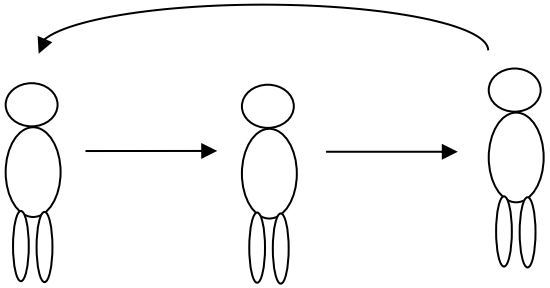
How many people say NO?	**0**	**1**	**2**	**3**	**4**	**5**
CHOOSE	YES YES YES	YES YES YES	YES YES YES	NO NO YES	NO NO YES	NO NO NO

You’ll notice that in the last case it is not possible to choose an even number of NO’s. Luckily the chance of this happening should be only 1 in 32. The demonstration I have in mind will not work in this case. We’ll discuss later what you might do when this happens.

Assuming that you have either no NO's or two NO's out of your three chosen volunteers you proceed as follows. Ask each of the three to display their question as well as their answer. You now announce who the questions refer to. It will look as if you have determined this in advance, but you choose people only at this time. You'll only choose someone who is in your chosen three, and never themselves. In fact, for best effect, your choice should be as follows, where the arrow shows who their question refers to.



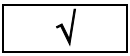
Now ask the two people on the right hand to swap places so that the people referred to are as shown.

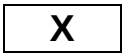


It doesn't really matter who is referring to whom, but it will seem a little less artificial if an adjustment is made. Now you ask any one of the three, "did you feel that you were completely free to choose your answer? You didn't feel constrained in any way?" They of course will insist that they were free.

"But I will show you that you were *forced* to choose as you did because of the answers of the others. If you had chosen otherwise there would have been a logical contradiction. Suppose you had chosen otherwise."

At this you instruct them to turn their card over. You then say, "suppose you were correct." Give them an other card marked with a tick, signifying that there answer was correct. Hopefully they can manage to hold all three cards without dropping any!





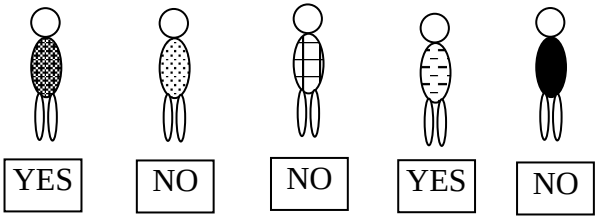
CORRECT **INCORRECT**

You then infer whether the person that their question relates to is correct or incorrect. You then give them a card with the appropriate word. Then you do this again so that the third person is holding a tick/cross card. Then you do this one more time so that you determine whether the first person is correct or incorrect. Because of the odd number of NO's this will

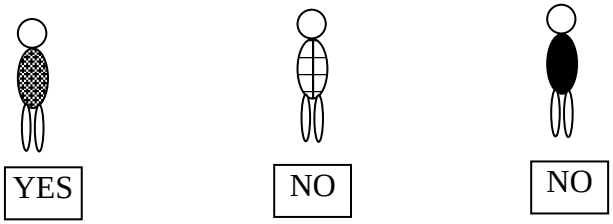
conflict with the card they are holding, and so you will have reached a contradiction.

"See, if you had answered differently to the way you did, your answer must have been incorrect, because a correct answer would lead to a contradiction. You then take away all the tick and cross cards and start all over again, this time giving the cross card to that person to display. You repeat the whole process, and discover that *again* you get a contradiction! "See, if you'd chosen other than what you did there would have been a logical contradiction. This proves that you *had* to choose the way you did!

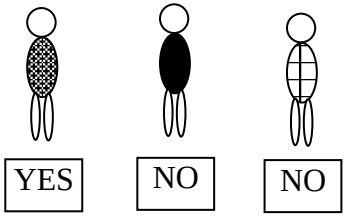
Here's an example of how this might work out. Remember each person's question refers to the person on *our* right, except the last, whose question refers to the first. Suppose the five answers are as follows.



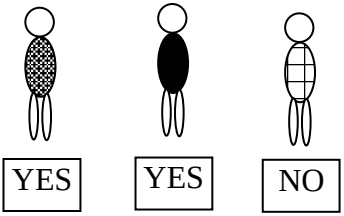
We choose two NO's and a YES as follows.



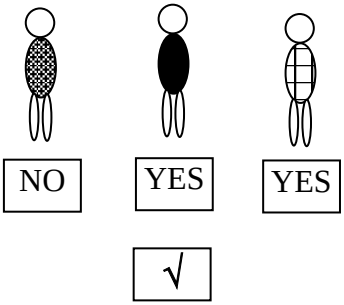
Now we swap the last two and move them together.



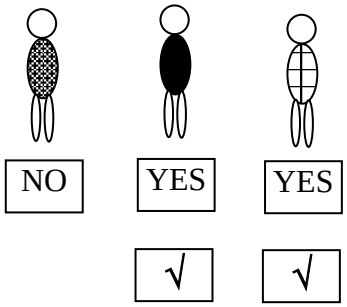
We get the middle person (it could have been any of them) to change his answer to see what logical implications this might have.



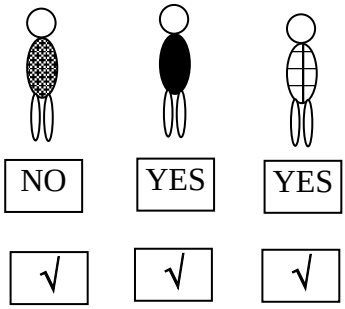
We suppose that the middle person was correct and you gave him a card containing a √.



The middle person said that the person on his left was correct (YES) and he was correct (√) so she must be correct. We give her a tick.

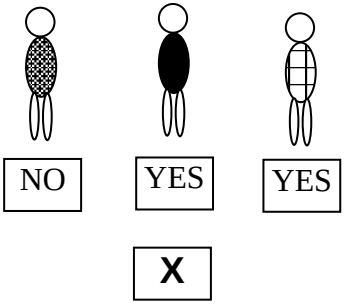


The third person says that the first would be correct, and the third person has a tick, so he is right. The first person gets a tick.

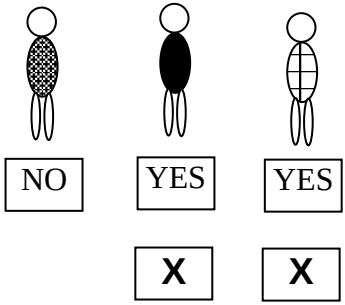


But the first person says that the middle person will say NO and he is supposed to be correct in saying this (we have had to give him a tick). But the middle person says YES. This is a contradiction.

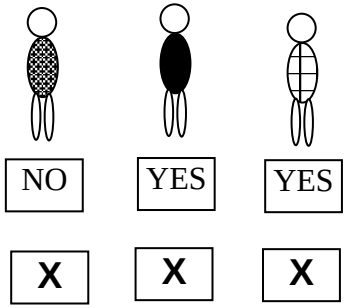
Now suppose the middle person, if he had said YES, would have been wrong. Give him a card with an X. We carry out the whole process afresh and again get a contradiction.



The middle person said that the person on his left was correct (YES) but he was wrong (X) so she must be incorrect. We give her a cross.



The third person says that the first will be correct and she is wrong, so he will be incorrect. Give the first person a cross.



But the first person says that the middle person will *not* be correct, and he is incorrect in saying this, so the middle person is correct. But the middle person is incorrect!

There's a problem if all of the original five people say NO. We'll not be able to select three people with an even number of NOs. This situation would be quite rare, but if it does arise you have the following options.

(1) Confess that the demonstration didn't work.

(2) Upbraid them for not having enough faith in their fellow volunteers and choose a fresh set of volunteers.

(3) Go through the above analysis to show that you get a contradiction, no matter whether the first person is correct or incorrect. Ask them how they have managed to defy logic and leave it at that. Hopefully nobody will ask "what has this got to do with free will?"

Remember that as you go round the circle of three (or all five) if someone has said YES then the next person will get the same tick or cross as they have, but every time you strike a NO the ticks and crosses swap over. Clearly with an odd number of such swaps around a circle with an odd number of people, there's bound to be a contradiction. Of course this only superficially has something to do with free will! It merely demonstrates the fact that the questions are indirectly self-referential.

So, after all this, does God exist? You certainly won't find the answer in this book or even the Bible, as useful as the Bible has been to many people. If there is a God and He chooses to reveal himself to you, then you'll know. Otherwise you have the free will to use the Axiom of Choice to not believe in God. (Actually that's not quite what the Axiom of Choice says but never mind.)

Oh, you don't believe in the Axiom of Choice. That's a logically valid position to take. What? You don't even believe in free will. Then why are you interested in proving things at all. You're simply a pre-programmed robot.

But let me remind you that this is a mathematics book, not a religious one. My goal is to explain the fact that logic has its limitations. As a mathematician I'm a great believer in it, but I'm fascinated to discover that there are impossible, uncountable, undecidable, unprovable things out there at the edge of the rational universe.

As the great bard once said:

There are more things in heaven and earth, Horatio,
Than are dreamt of in your philosophy.

[Hamlet Act 1, Scene 5]

THE MATHEMATICIAN'S CREED

I believe in the validity of standard logic, provided there is no self-referentiality, direct of indirect.

I believe that Mathematics was created from the empty set and that the ZF axioms are consistent.

I believe in Mathematical Intuition,
Informed by rigorous proof,
But inspired by Mathematical Imagination
Fuelled by countless cups of coffee.

I believe in the Axiom of Choice
And the Continuum Hypothesis
And whatever other axioms I might find convenient to use
Provided they've been proved consistent with ZF.

I believe that Mathematics contains no facts
But depends on definitions and sets of axioms.

I believe that mathematics is independent of the material world, so that it can be understood by a disembodied angel.

Yet I believe that mathematics is the one great tool for understanding the material world.

It guides and underpins the Kingdom of Science
And has brought great benefit to mankind.

I believe that great as Mathematics is
There are truths that lie beyond its reach.

Our minds can soar into realms unknown
But more truth lies beyond the edge of the rational universe.

POSTLUDE
Disclaimer

Having completed this work I now look back and contemplate it. Is it all true? I've tried hard to ensure that every statement which is claimed to be true, is indeed true.

Yet if I had followed the practice of artists in certain eastern cultures, such as the carpet weavers of Iraq, I'd have deliberately woven in one or two flaws. Only Allah is perfect and if I claim to have achieved perfection I'm setting myself up in opposition to Him, which may have dire consequences.

But it's not too late. So just to be on the safe side let me include the following disclaimer.

At least one assertion in this book is false.

There, that should do it. But wait a bit. That disclaimer itself can't help but be *true*. Why? Well, if it's false then *every* assertion in the book is true, including the disclaimer itself! But that would be a contradiction.

So the disclaimer is true and so there must be an error somewhere else in the book. But where? I've

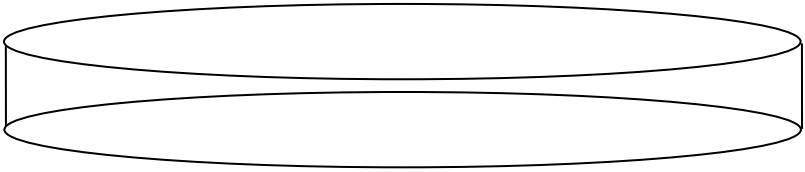
checked it most carefully, and I haven't found an error. Yet simply by adding this disclaimer it forces me to admit that I must have made a mistake! But for the life of me I simply can't find it.

Perhaps, if logic is forcing me to have made a mistake, maybe I don't have free will after all. What's that you say? Self-referentiality is not allowed? I see, it's wrong for me to make any statement about myself.

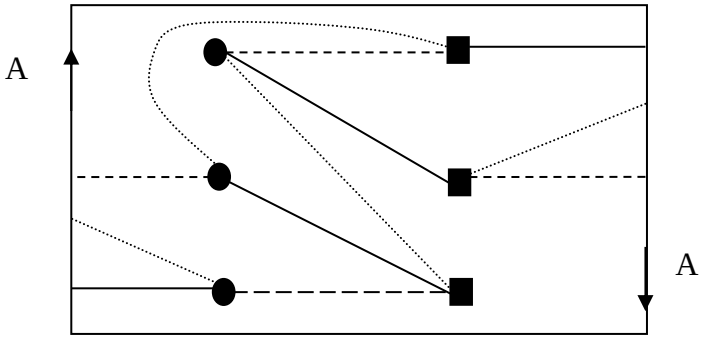
PUZZLE ANSWER

The secret to the puzzle lies in the way you join the two shorter ends.

If you bend the strip and join the ends so as to make a short cylinder with the longer edges becoming the circles at each end, you have no more chance of solving the puzzle than before.



But if you give the strip a half-twist before you join the ends then what you have is a Möbius Band. (If your strip is too short to make the half-twist use a longer strip.) And on a Möbius Band the puzzle can be solved.



BIBLIOGRAPHY

DUNNE, Paul E.
Computability Theory - concepts and applications
Ellis Horwood 1991

HOFSTADTER, Douglas R.
Gödel, Escher, Bach: an eternal golden braid
Harvester Press 1979

KILMISTER C. W.
Language Logic and Mathematics
English Universities Press 1967

NAGEL Ernest and NEWMAN, James R.
Gödel's Proof
Routledge & Kegan Paul 1959