01 Zero

At a young age we make an unsteady entrance into numberland. We learn that 1 is first in the 'number alphabet', and that it introduces the counting numbers 1, 2, 3, 4, 5,. . . Counting numbers are just that: they count real things – apples, oranges, bananas, pears. It is only later that we can count the number of apples in a box when there are none.

Even the early Greeks, who advanced science and mathematics by quantum leaps, and the Romans, renowned for their feats of engineering, lacked an effective way of dealing with the number of apples in an empty box. They failed to give 'nothing' a name. The Romans had their ways of combining I, V, X, L, C, D and M but where was 0? They did not count 'nothing'.

How did zero become accepted?

The use of a symbol designating 'nothingness' is thought to have originated thousands of years ago. The Maya civilization in what is now Mexico used zero in various forms. A little later, the astronomer Claudius Ptolemy, influenced by the Babylonians, used a symbol akin to our modern 0 as a placeholder in his number system. As a placeholder, zero could be used to distinguish between examples (in modern notation) such as 75 and 705, instead of relying on context as the Babylonians had done. This might be compared with the introduction of the 'comma' into language – both help with reading the right meaning. But, just as the comma comes with a set of rules for its use – there have to be rules for using zero.

The seventh-century Indian mathematician Brahmagupta treated zero as a 'number', not merely as a placeholder, and set out rules for dealing with it. These included 'the sum of a positive number and zero is positive' and 'the sum of zero and zero is zero'. In thinking of zero as a number rather than a placeholder, he was quite advanced. The Hindu-Arabic numbering system which included zero in this way was promulgated in the West by Leonardo of Pisa – Fibonacci – in his Liber Abaci (The Book of Counting) first published in 1202. Brought up in North Africa and schooled in the Hindu-Arabian arithmetic, he recognized the power of

the condensed idea
Nothing is quite something

timeline

700BC	AD628	830	1100	1202
The Babylonians use zero as a placeholder in their number system	Brahmagupta uses zero and states rules for its use with other numerals	Mahavira has ideas on how zero interacts with other numerals	Bhaskara uses 0 as a symbol in algebra and attempts to show how it is manipulated	Fibonacci uses the extra symbol 0 added to the Hindu-Arabic system of numerals 1, . . . , 9 but not as a number on a par with them

number and the letter E stands for 'exponential'. Sometimes we might want to use bigger numbers still, for instance if we were talking about the number of hydrogen atoms in the known universe. This has been estimated as about $1.7\times10^{\mathbf{77}}$. Equally 1.7×10^{-77}, with a negative power, is a very small number and this too is easily handled using scientific notation. We couldn't begin to think of these numbers with the Roman symbols.

Zeros and ones

While base 10 is common currency in everyday life, some applications require other bases. The binary system which uses base 2 lies behind the power of the modern computer. The beauty of binary is that any number can be expressed using only the symbols 0 and 1. The tradeoff for this economy is that the number expressions can be very long.

Powers of 2	**Decimal**
2^0	1
2^1	2
2^2	4
2^3	8
2^4	16
2^5	32
2^6	64
2^7	128
2^8	256
2^9	512
2^{10}	1024

How can we express **394** in binary notation? This time we are dealing with powers of 2 and after some working out we can give the full expression as,

$$\mathbf{394} = \mathbf{1}\times256+\mathbf{1}\times128+\mathbf{0}\times64+\mathbf{0}\times32+\mathbf{0}\times16+\mathbf{1}\times8+\mathbf{0}\times4+\mathbf{1}\times2+\mathbf{0}\times1$$

so that reading off the zeros and ones, **394** in binary is **110001010** .

As binary expressions can be very long, other bases frequently arise in computing. These are the octal system (base 8) and the hexadecimal system (base 16). In the octal system we only need the symbols 0, 1, 2, 3, 4, 5, 6, 7, whereas hexadecimal uses 16 symbols. In this base 16 system, we customarily use 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F. As 10 corresponds to the letter A, the number **394** is represented in hexadecimal as 18A. It's as easy as ABC, which bear in mind, is really 2748 in decimal!

the condensed idea
Writing numbers down

timeline

30,000BC	2000BC	AD600	[illegible]	1600
Palaeolithic peoples in Europe make number marks on bones	The Babylonians use symbols for numbers	The forerunner of our modern decimal notation is used in India	The Hindu-Arabic system of writing numerals 1, ..., 9 and a zero, spreads	The symbols of the decimal system take their recognizable modern forms

numerator is bigger than the denominator. Dividing 14 by 5 we get 2 with 4 left over, which can be written as the 'mixed' number $2\frac{4}{5}$. This comprises the whole number 2 and the 'proper' fraction $\frac{4}{5}$. Some early writers wrote this as $\frac{4}{5}2$. Fractions are usually represented in a form where the numerator and denominator (the 'top' and the 'bottom') have no common factors. For example, the numerator and denominator of 8/10 have a common factor of 2, because $8 = 2 \times 4$ and $10 = 2 \times 5$. If we write the fraction $8/10 = {}^{2\times4}/_{2\times5}$ we can 'cancel' the 2s out and so $8/10 = \frac{4}{5}$, a simpler form with the same value. Mathematicians refer to fractions as 'rational numbers' because they are ratios of two numbers. The rational numbers were the numbers the Greeks could 'measure'.

Adding and multiplying

The rather curious thing about fractions is that they are easier to multiply than to add. Multiplication of whole numbers is so troublesome that ingenious ways had to be invented to do it. But with fractions, it's addition that's more difficult and takes some thinking about.

Let's start by multiplying fractions. If you buy a shirt at four-fifths of the original price of £30 you end up paying the sale price of £24. The £30 is divided into five parts of £6 each and four of these five parts is $4 \times 6 = 24$, the amount you pay for the shirt.

Subsequently, the manager of the shop discovers that the shirts are not selling at all well so he drops the price still further, advertising them at ½ of the sale price. If you go into the shop you can now get the shirt for £12. This is $\frac{1}{2} \times \frac{4}{5} \times 30$ which is equal to 12. To multiply two fractions together you just multiply the denominators together and the numerators together:

$$\frac{1}{2} \times \frac{4}{5} = \frac{1\times4}{2\times5} = \frac{4}{10}$$

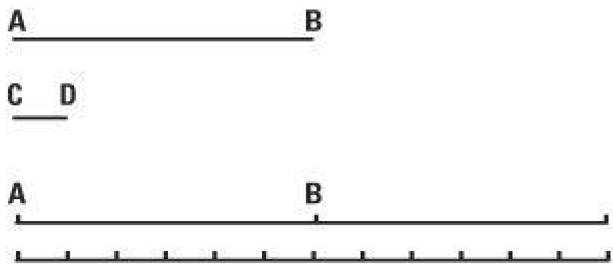
If the manager had made the two reductions at a single stroke he would have advertised the shirts at four-tenths of the original price of £30. This is $4/10 \times 30$ which is £12.

Adding two fractions is a different proposition. The addition $\frac{1}{3} + \frac{2}{3}$ is OK because the denominators are the same. We simply add the two numerators together to get 3/3, or 1. But how could we add two-thirds of a cake to fourfifths of a cake? How could we figure out $\frac{2}{3} + \frac{4}{5}$? If only we could say $\frac{2}{3} + \frac{4}{5} = {}^{2+4}/_{3+5}$

These numbers all have square roots, but they are not equal to whole numbers. Virtually all calculators have a √ button, and using it we find, for instance, $\sqrt{7} = 2.645751311$.

Let's look at $\sqrt{2}$. The number 2 had special significance for the Pythagoreans because it is the first even number (the Greeks thought of the even numbers as feminine and the odd ones as masculine – and the small numbers had distinct personalities). If you work out $\sqrt{2}$ on your calculator you will get 1.414213562 assuming your calculator gives this many decimal places. Is this the square root of 2? To check we make the calculation 1.414213562 × 1.414213562. This turns out to be 1.999999999. This is not quite 2 for 1.414213562 is only an approximation for the square root of 2.

What is perhaps remarkable is that all we will ever get is an approximation! The decimal expansion of $\sqrt{2}$ to millions of decimal places will only ever be an approximation. The number $\sqrt{2}$ is important in mathematics, perhaps not quite as illustrious as π or e (see pages 20–27) but important enough to gets its own name – it is sometimes called the 'Pythagorean number'.



Are square roots fractions?

Asking whether square roots are fractions is linked to the theory of measurement as known to the ancient Greeks. Suppose we have a line AB whose length we wish to measure, and an indivisible 'unit' CD with which to measure it. To make the measurement we place the unit CD sequentially against AB. If we place the unit down m times and the end of the last unit fits flush with the end of AB (at the point B) then the length of AB will simply be m. If not we can place a copy of AB next to the original and carry on measuring with the unit (see figure). The Greeks believed that at some point using n copies of AB and m units, the

Range	1–100	101–200	201–300	301–400	401–500	501–600	601–700	701–800	801–900	901–1000	1–1000
Number of primes	25	21	16	16	17	14	16	14	15	14	168

In 1792, when only 15 years old, Carl Friedrich Gauss suggested a formula P(n) for estimating the number of prime numbers less than a given number n (this is now called the prime number theorem). For n = 1000 the formula gives the approximate value of 172. The actual number of primes, 168, is less than this estimate. It had always been assumed this was the case for any value of n, but the primes often have surprises in store and it has been shown that for n = 10^{371} (a huge number written long hand as a 1 with 371 trailing 0s) the actual number of primes exceeds the estimate. In fact in some regions of the counting numbers the difference between the estimate and the actual number oscillates between less and excess.

How many?

There are infinitely many prime numbers. Euclid stated in his Elements (Book 9, Proposition 20) that 'prime numbers are more than any assigned multitude of prime numbers'. Euclid's beautiful proof goes like this:

> Suppose that P is the largest prime, and consider the number N = (2 × 3 × 5 × . . . × P) + 1. Either N is prime or it is not. If N is prime we have produced a prime greater than P which is a contradiction to our supposition. If N is not a prime it must be divisible by some prime, say p, which is one of 2, 3, 5, . . ., P. This means that p divides N − (2 × 3 × 5 × . . . × P). But this number is equal to 1 and so p divides 1. This cannot be since all primes are greater than 1. Thus, whatever the nature of N, we arrive at a contradiction. Our original assumption of there being a largest prime P is therefore false. Conclusion: the number of primes is limitless.

Though primes 'stretch to infinity' this fact has not prevented people striving to find the largest known prime. One which has held the record recently is the enormous Mersenne prime $2^{24036583} - 1$, which is approximately $10^{7235732}$ or a number starting with 1 followed by 7,235,732 trailing zeroes.

prime. But it is those Mersenne numbers that are also prime that can be used to construct perfect numbers.

Mersenne knew that if the power was not a prime number, then the Mersenne number could not be a prime number either, accounting for the non-prime powers 4, 6, 8, 9, 10, 12, 14 and 15 in the table. The Mersenne numbers could only be prime if the power was a prime number, but was that enough? For the first few cases, we do get 3, 7, 31 and 127, all of which are prime. So is it generally true that a Mersenne number formed with a prime power should be prime as well?

Many mathematicians of the ancient world up to about the year 1500 thought this was the case. But primes are not constrained by simplicity, and it was found that for the power 11 (a prime number), $2^{11} - 1 = 2047 = 23 \times 89$ and consequently it is not a prime number. There seems to be no rule. The Mersenne numbers $2^{17} - 1$ and $2^{19} - 1$ are both primes, but $2^{23} - 1$ is not a prime, because

Just good friends

The hard-headed mathematician is not usually given to the mystique of numbers but numerology is not yet dead. The amicable numbers came after the perfect numbers though they may have been known to the Pythagoreans. Later they became useful in compiling romantic horoscopes where their mathematical properties translated themselves into the nature of the ethereal bond. The two numbers 220 and 284 are amicable numbers. Why so? Well, the divisors of 220 are 1, 2, 4, 5, 10, 11, 20, 22, 44, 55 and 110 and if you add them up you get 284. You've guessed it. If you figure out the divisors of 284 and add them up, you get 220. That's true friendship.

Mersenne Primes

Finding Mersenne primes is not easy. Many mathematicians over the centuries have added to the list, which has a chequered history built on a combination of error and correctness. The great Leonhard Euler contributed the eighth Mersenne prime, $2^{31} - 1 = 2{,}147{,}483{,}647$, in 1732. Finding the 23rd Mersenne prime, $2^{11213} - 1$, in 1963 was a source of pride for the mathematics department at the University of Illinois, who announced it to the world on their university postage stamp. But with powerful computers the Mersenne prime industry had moved on and in the late 1970s high school students Laura Nickel and Landon Noll jointly discovered the 25th Mersenne prime, and Noll the 26th Mersenne prime. To date 45 Mersenne primes have been discovered.

11 Fibonacci numbers

In The Da Vinci Code, the author Dan Brown made his murdered curator Jacques Saunière leave behind the first eight terms of a sequence of numbers as a clue to his fate. It required the skills of cryptographer Sophie Neveu to reassemble the numbers 13, 3, 2, 21, 1, 1, 8 and 5 to see their significance. Welcome to the most famous sequence of numbers in all of mathematics.

The Fibonacci sequence of whole numbers is:

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, 2584, . . .

The sequence is widely known for its many intriguing properties. The most basic – indeed the characteristic feature which defines them – is that every term is the addition of the previous two. For example 8 = 5 + 3, 13 = 8 + 5, . . ., 2584 = 1587 + 987, and so on. All you have to remember is to begin with the two numbers 1 and 1 and you can generate the rest of the sequence on the spot. The Fibonacci sequence is found in nature as the number of spirals formed from the number of seeds in the spirals in sunflowers (for example, 34 in one direction, 55 in the other), and the room proportions and building proportions designed by architects. Classical musical composers have used it as an inspiration, with Bartók's Dance Suite believed to be connected to the sequence. In contemporary music Brian Transeau (aka BT) has a track in his album This Binary Universe called 1.618 as a salute to the ultimate ratio of the Fibonacci numbers, a number we shall discuss a little later.

Origins

The Fibonacci sequence occurred in the Liber Abaci published by Leonardo of Pisa (Fibonacci) in 1202, but these numbers were probably known in India before that. Fibonacci posed the following problem of rabbit generation:

Mature rabbit pairs generate young rabbit pairs each month. At the beginning of the year there is one young rabbit pair. By the end of the first month they will have matured, by the end of the second month the mature pair is still there and they will have generated a young rabbit pair. The process of maturing and generation continues. Miraculously none of the rabbit pairs die.

The value of £4, as we draw the coins out of the purse, can be any of the following ways, 1 + 1 + 1 + 1; 2 + 1 +1; 1 + 2 + 1; 1 + 1 + 2; and 2 + 2. There are 5 ways in all – and this corresponds to the fifth Fibonacci number. If you take out £20 there are 6,765 ways of taking the £1 and £2 coins out, corresponding to the 21st Fibonacci number! This shows the power of simple mathematical ideas.

The golden ratio

If we look at the ratio of terms formed from the Fibonacci sequence by dividing a term by its preceding term we find out another remarkable property of the Fibonacci numbers. Let's do it for a few terms 1, 1, 2, 3, 5, 8, 13, 21, 34, 55.

1/1	2/1	3/2	5/3	8/5	13/8	21/13	34/21	55/34
1.000	2.000	1.500	1.333	1.600	1.625	1.615	1.619	1.617

Pretty soon the ratios approach a value known as the golden ratio, a famous number in mathematics, designated by the Greek letter Φ. It takes its place amongst the top mathematical constants like π and e, and has the exact value

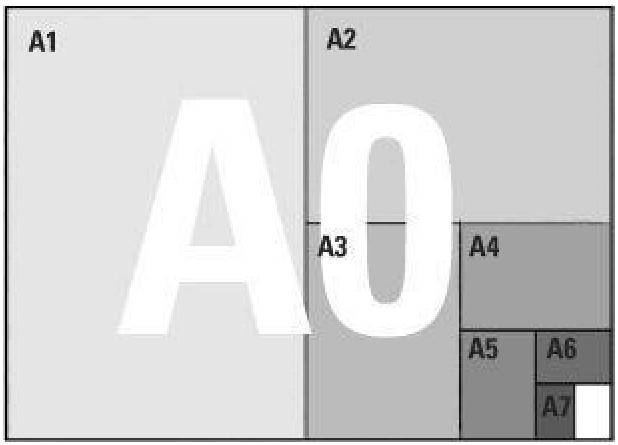
$$\phi = \frac{1+\sqrt{5}}{2}$$

and this can be approximated to the decimal 1.618033988. . . With a little more work we can show that each Fibonacci number can be written in terms of Φ.

12 Golden rectangles

Rectangles are all around us – buildings, photographs, windows, doors, even this book. Rectangles are present within the artists' community – Piet Mondrian, Ben Nicholson and others, who progressed to abstraction, all used one sort or another. So which is the most beautiful of all? Is it a long thin 'Giacometti rectangle' or one that is almost a square? Or is it a rectangle in between these extremes?

Does the question even make sense? Some think so, and believe particular rectangles are more 'ideal' than others. Of these, perhaps the golden rectangle has found greatest favour. Amongst all the rectangles one could choose for their different proportions – for that is what it comes down to – the golden rectangle is a very special one which has inspired artists, architects and mathematicians. Let's look at some other rectangles first.



Mathematical paper

If we take a piece of A4 paper, whose dimensions are a short side of 210 mm and a long side of 297 mm, the length-to-width ratio will be 297/210 which is approximately 1.4142. For any international A-size paper with short side equal to b, the longer side will always be 1.4142 × b. So for A4, b = 210 mm, while for A5, b = 148 mm. The A-formulae system used for paper sizes has a highly desirable property, one that does not occur for arbitrary paper sizes. If an A-size piece of paper is folded about the middle, the two smaller rectangles formed are directly in proportion to the larger rectangle. They are two smaller versions of the same rectangle.

In this way, a piece of A4 folded into two pieces generates two pieces of A5. Similarly a piece of A5-size paper generates two pieces of A6. In the other direction, a sheet of A3 paper is made up of two pieces of A4. The smaller the number on the A-size the larger the piece of paper. How did we know that the particular number 1.4142 would do the trick? Let's fold a rectangle, but this time let's make it one where we don't know the length of its longer side. If we take the breadth of a rectangle to be 1 and we write the length of the longer side as x, then the length-to-width ratio is $x/1$. If we now fold the rectangle, the length-to-width ratio of the smaller rectangle is $1/\frac{1}{2}x$, which is the same as $2/x$. The point of A sizes is that our two ratios must stand for the same proportion, so we get an equation $x/1 = 2/x$ or $x^2 = 2$. The true value of x is therefore $\sqrt{2}$ which is approximately by 1.4142.

Mathematical gold

The golden rectangle is different, but only slightly different. This time the rectangle is folded along the line RS in the diagram so that the points MRSQ make up the corners of a square.

The key property of the golden rectangle is that the rectangle left over, RNPS, is proportional to the large rectangle – what is left over should be a mini-replica of the large rectangle.

As before, we'll say the breadth $MQ = MR$ of the large rectangle is 1 unit of length while we'll write the length of the longer side MN as x. The length-to-width ratio is again $x/1$. This time the breadth of the smaller rectangle RNPS is $MN - MR$, which is $x - 1$ so the length-to-width ratio of this rectangle is $1/(x-1)$. By equating them, we get the equation

$$\frac{x}{1} = \frac{1}{x-1}$$

14 Algebra

Algebra gives us a distinctive way of solving problems, a deductive method with a twist. That twist is 'backwards thinking'. For a moment consider the problem of taking the number 25, adding 17 to it, and getting 42. This is forwards thinking. We are given the numbers and we just add them together. But instead suppose we were given the answer 42, and asked a different question? We now want the number which when added to 25 gives us 42. This is where backwards thinking comes in. We want the value of x which solves the equation 25 + x = 42 and we subtract 25 from 42 to give it to us.

Word problems which are meant to be solved by algebra have been given to schoolchildren for centuries:

My niece Michelle is 6 years of age, and I am 40.
When will I be three times as old as her?

We could find this by a trial and error method but algebra is more economical. In x years from now Michelle will be 6 + x years and I will be 40 + x. I will be three times older than her when

$$3 \times (6 + x) = 40 + x$$

Multiply out the left-hand side of the equation and you'll get $18 + 3x = 40 + x$, and by moving all the xs over to one side of the equation and the numbers to the other, we find that $2x = 22$ which means that $x = 11$. When I am 51 Michelle will be 17 years old. Magic!

What if we wanted to know when I will be twice as old as her? We can use the same approach, this time solving

$$2 \times (6 + x) = 40 + x$$

to get $x = 28$. She will be 34 when I am 68. All the equations above are of the simplest type – they are called 'linear' equations. They have no terms like x^2 or x^3, which make equations more difficult to solve. Equations with terms like x^2 are called 'quadratic' and those with terms like x^3 are called 'cubic' equations. In past times, x^2 was represented as a square and because a square has four sides the term quadratic was used; x^3 was represented by a cube.

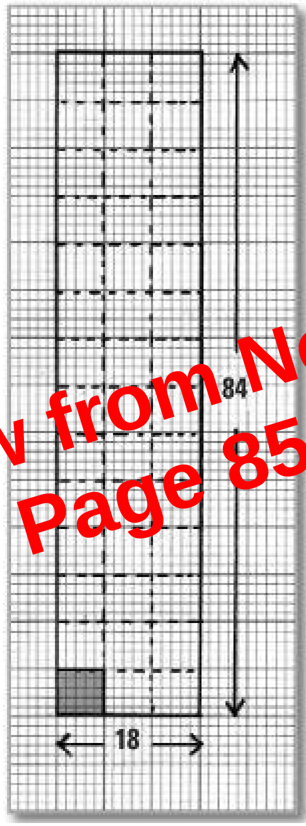
Mathematics underwent a big change when it passed from the science of arithmetic to the science of symbols or algebra. To progress from numbers to letters is a mental jump but the effort is worthwhile.

15 Euclid’s algorithm

Al-Khwarizmi gave us the word ‘algebra’, but it was his ninth-century book on arithmetic that gave us the word ‘algorithm’. Pronounced ‘Al Gore rhythm’ it is a concept useful to mathematicians and computer scientists alike. But what is one? If we can answer this we are on the way to understanding Euclid’s division algorithm.

Firstly, an algorithm is a routine. It is a list of instructions such as ‘you do this and then you do that’. We can see why computers like algorithms because they are very good at following instructions and never wander off track. Some mathematicians think algorithms are boring because they are repetitious, but to write an algorithm and then translate it into hundreds of lines of computer code containing mathematical instructions is no mean feat. There is a considerable risk of it all going horribly wrong. Writing an algorithm is a creative challenge. There are often several methods available to do the same task and the best one must be chosen. Some algorithms may not be ‘fit for purpose’ and some may be downright inefficient because they meander. Some may be quick but produce the wrong answer. It’s a bit like cooking. There must be hundreds of recipes (algorithms) for cooking roast turkey with stuffing. We certainly don’t want a poor algorithm for doing this on the one day of the year when it matters. So we have the ingredients and we have the instructions. The start of the (abbreviated) recipe might go something like this:

- Fill the turkey cavity with stuffing
- Rub the outside skin of the turkey with butter
- Season with salt, pepper and paprika
- Roast at 335 degrees for 3½ hours
- Let the cooked turkey rest for ½ hour



All we have to do is carry out the algorithm in sequential steps one after the other. The only thing missing in this recipe, usually present in a mathematical algorithm, is a loop, a tool to deal with recursion. Hopefully we won't have to cook the turkey more than once.

In mathematics we have ingredients too – these are the numbers. Euclid's division algorithm is designed to calculate the greatest common divisor (written gcd). The gcd of two whole numbers is the greatest number that divides into both of them. As our example ingredients, we'll choose the two numbers 18 and 84.

The greatest common divisor

The gcd in our example is the largest number that exactly divides both 18 and

If we have another proposition **b** such as 'Ethel is a cat' then we can combine these two propositions in several ways. One combination is written **a** V **b**. The connective V corresponds to 'or' but its use in logic is slightly different from 'or' in everyday language. In logic, **a** V **b** is true if either 'Freddy is a spaniel' is true or 'Ethel is a cat' is true, or if both are true, and it is only false when both **a** and **b** are false. This conjunction of propositions can be summarized in a truth table.

a	b	a → b
T	T	T
T	F	F
F	T	T
F	F	T

Implies truth table

We can also combine propositions using 'and', written as **a**∧**b**, and 'not', written as ¬**a**. The algebra of logic becomes clear when we combine these propositions using a mixture of the connectives with **a**, **b** and **c** like **a** ∧ (**b** V**c)** . We can obtain an equation we call an identity:

$$\mathbf{a} \wedge (\mathbf{b} \vee \mathbf{c}) = (\mathbf{a} \wedge \mathbf{b}) \vee (\mathbf{a} \vee \mathbf{c})$$

The symbol ≡ means equivalence between logical statements where both sides of the equivalence have the same truth table. There is a parallel between the algebra of logic and ordinary algebra because the symbols Λ and V act similarly to × and + in ordinary algebra, where we have x × (y + z) = (x × y) + (x × z). However, the parallel is not exact and there are exceptions.

Other logical connectives may be defined in terms of these basic ones. A useful one is the 'implication' connective **a**→**b** which is defined to be equivalent to ¬ **a** ∧ **b** and has the truth table shown.

Now if we look again at the newspaper leader, we can write it in symbolic form to give the argument in the margin:

$$\begin{array}{c} C \rightarrow P \\ C \vee S \\ S \rightarrow H \\ \neg H \\ \hline P \end{array}$$

or, rewriting using brackets,

$$6 \times 6 = 2 \times (3 + 3 + 3 + 3 + 3 + 3)$$

This means 6×6 is a multiple of 2 and, as such, is an even number. But in this argument there is nothing which is particular to 6, and we could have started with $n = 2 \times k$ to obtain

$$n \times n = 2 \times (k + k + . . . + k)$$

and conclude that $n \times n$ is even. Our proof is now complete. In translating Euclid's Elements, latter-day mathematicians wrote 'QED' at the end of a proof to say job done – it's an abbreviation for the Latin quod erat demonstrandum (which was to be demonstrated). Nowadays they use a filled-in square ■. This is called a halmos after Paul Halmos who introduced it.

The indirect method

In this method we pretend the conclusion is false and by a logical argument demonstrate that this contradicts the hypothesis. Let's prove the previous result by this method.

Our hypothesis is that n is even and we'll pretend $n \times n$ is odd. We can write $n \times n = n + n + . . . + n$ and there are n of these. This means n cannot be even (because if it were $n \times n$ would be even). Thus n is odd, which contradicts the hypothesis. ■

This is actually a mild form of the indirect method. The full-strength indirect method is known as the method of reductio ad absurdum (reduction to the absurd), and was much loved by the Greeks. In the academy in Athens, Socrates and Plato loved to prove a debating point by wrapping up their opponents in a mesh of contradiction and out of it would be the point they were trying to prove. The classical proof that the square root of 2 is an irrational number is one of this form where we start off by assuming the square root of 2 is a rational number and deriving a contradiction to this assumption.

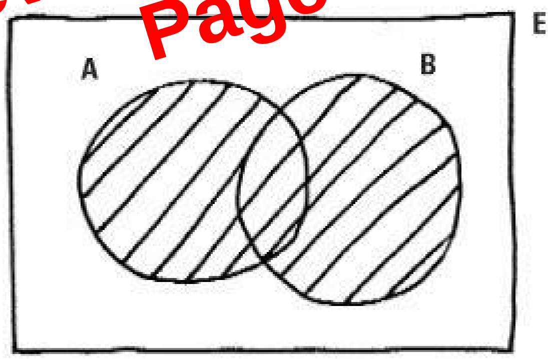
The method of mathematical induction

Mathematical induction is powerful way of demonstrating that a sequence of statements P_1, P_2, P_3, . . . are all true. This was recognized by Augustus De Morgan in the 1830s who formalized what had been known for hundreds of

18 Sets

Nicholas Bourbaki was a pseudonym for a self-selected group of French academics who wanted to rewrite mathematics from the bottom up in 'the right way'. Their bold claim was that everything should be based on the theory of sets. The axiomatic method was central and the books they put out were written in the rigorous style of 'definition, theorem and proof'. This was also the thrust of the modern mathematics movement of the 1960s.

Georg Cantor created set theory out of his desire to put the theory of real numbers on a sound basis. Despite initial prejudice and criticism, set theory was well established as a branch of mathematics by the turn of the 20th century.



The union of A and B

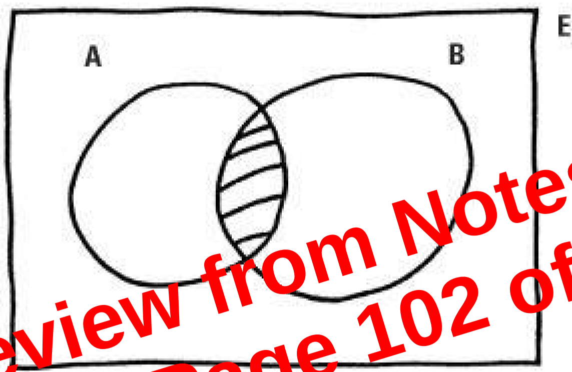
What are sets?

A set may be regarded as a collection of objects. This is informal but gives us the main idea. The objects themselves are called 'elements' or 'members' of the set. If we write a set A which has a member a, we may write $a \in A$, as did Cantor. An example is $A = \{1, 2, 3, 4, 5\}$ and we can write $1 \in A$ for membership, and $6 \in A$ for non-membership.

Sets can be combined in two important ways. If A and B are two sets then the set consisting of elements which are members of A or B (or both) is called the 'union' of the two sets. Mathematicians write this as $A \cup B$. It can also be described by a Venn diagram, named after the Victorian logician the Rev. John

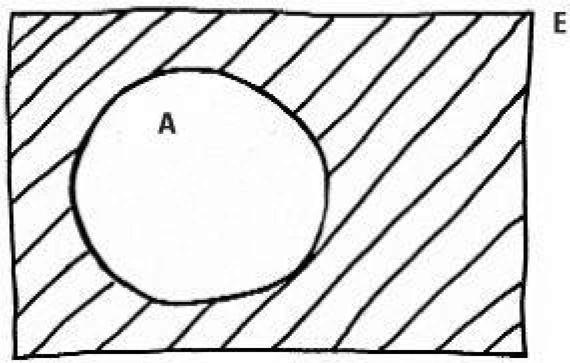
Venn. Euler used diagrams like these even earlier.

The set A $\cap$ B consists of elements which are members of A and B and is called the 'intersection' of the two sets.



The intersection of A and B

If A = {1, 2, 3, 4, 5} and B = {1, 3, 5, 7, 10, 21}, the union is A $\cup$ B = {1, 2, 3, 4, 5, 7, 10, 21} and the intersection is A $\cap$ B = {1, 3, 5}. If we regard a set A as part of a universal set E, we can define the complement set $\neg$A as consisting of those elements in E which are not in A.



The complement of A

The operations $\cap$ and $\cup$ on sets are analogous to $\times$ and + in algebra. Together with the complement operation $\neg$, there is an 'algebra of sets'. The Indian-born British mathematician Augustus De Morgan, formulated laws to show how all three operations work together. In our modern notation, De Morgan's laws are:

$$\neg(A \cup B) = (\neg A) \cap (\neg B)$$

and

19 Calculus

A calculus is a way of calculating, so mathematicians sometimes talk about the 'calculus of logic', the 'calculus of probability', and so on. But all are agreed there is really only one Calculus, pure and simple, and this is spelled with a capital C.

Calculus is a central plank of mathematics. It would now be rare for a scientist, engineer or a quantitative economist not to have come across Calculus, so wide are its applications. Historically it is associated with Isaac Newton and Gottfried Leibniz who pioneered it in the 17th century. Their similar theories resulted in a priority dispute over who was the discoverer of Calculus. In fact, both men came to their conclusions independently and their methods were quite different.

Since then Calculus has become a huge subject. Each generation bolts on techniques they think should be learned by the younger generation, and these days textbooks run beyond a thousand pages and involve many extras. For all these add-ons, what is absolutely essential is differentiation and integration, the twin peaks of Calculus as set up by Newton and Leibniz. The words are derived from Leibniz's differentialis (taking differences or 'taking apart') and integralis (the sum of parts, or 'bringing together').

In technical language, differentiation is concerned with measuring change and integration with measuring area, but the jewel in the crown of Calculus is the 'star result' that they are two sides of the same coin – differentiation and integration are the inverses of each other. Calculus is really one subject, and you need to know about both sides. No wonder that Gilbert and Sullivan's 'very model of a modern Major General' in The Pirates of Penzance proudly proclaimed them both:

With many cheerful facts about the square of the hypotenuse.
I'm very good at integral and differential calculus.

Differentiation

Scientists are fond of conducting 'thought experiments' – Einstein especially liked them. Imagine we are standing on a bridge high above a gorge and are

Integration

The first application of integration was to measure area. The measurement of the area under a curve is done by dividing it into approximate rectangular strips, each with width dx. By measuring the area of each and adding them up we get the 'sum' and so the total area. The notation S standing for sum was introduced by Leibniz in an elongated form ∫. The area of each of the rectangular strips is udx, so the area A under the curve from 0 to x is

$$A = \int_0^x u\,dx$$

If the curve we're looking at is u = x^2, the area is found by drawing narrow rectangular strips under the curve, adding them up to calculate the approximate area, and applying a limiting process to their widths to gain the exact area. This answer gives the area

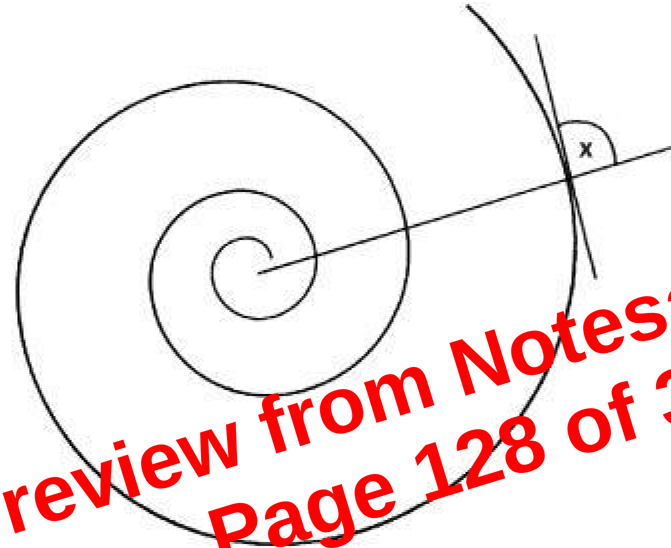
$$A = x^3/3$$

u	$\int_0^x u\,dx$
x^2	$x^3/3$
x^3	$x^4/4$
x^4	$x^5/5$
x^5	$x^6/6$
. . .	. . .
x^n	$x^{n+1}/(n+1)$

For different curves (and so other expressions for u) we could still calculate the integral. Like the derivative, there is a regular pattern for the integral of powers of x. The integral is formed by dividing by the 'previous power +1' and adding 1 to it to make the new power.

The star result

If we differentiate the integral A = $x^3/3$ we actually get the original u = x^2. If



The logarithmic spiral

Jacob Bernoulli of the famed mathematical clan from Switzerland was so enamoured with the logarithmic spiral that he wanted it carved on his tomb in Basle. The 'Renaissance man' Emanuel Swedenborg regarded the spiral as the most perfect of shapes. A three-dimensional spiral which winds itself around a cylinder is called a helix. Two of these – a double helix – form the basic structure of DNA.

There are many classical curves, such as the limaçon, the lemniscate and the various ovals. The cardioid derives its name from being shaped like a heart. The catenary curve was the subject of research in the 18th century and it was identified as the curve formed by a chain hanging between two points. The parabola is the curve seen in a suspension bridge hanging between its two vertical pylons.

are ‘simple’ (do not cross themselves) and ‘closed’ (have no beginning or end). Jordan’s celebrated theorem has meaning. It states that a simple closed curve has an inside and an outside. Its apparent ‘obviousness’ is a deception.

In Italy, Giuseppe Peano caused a sensation when, in 1890, he showed that, according to Jordan’s definition, a filled in square is a curve. He could organize the points on a square so that they could all be ‘traced out’ and at the same time conform to Jordan’s definition. This was called a space-filling curve and blew a hole in Jordan’s definition – clearly a square is not a curve in the conventional sense.

Examples of space-filling curves and other pathological examples caused mathematicians to go back to the drawing board once more and think about the foundations of curve theory. The whole question of developing a better definition of a curve was raised. At the start of the 20th century this task took mathematics into the new field of topology.

the condensed idea
Going round the bend

timeline

c.300BC	c.250BC	c.225BC	AD1704	1890	1920s
Euclid defines the conic sections	Archimedes investigates spirals	Apollonius of Perga publishes *Conics*	Newton classifies the cubic curves	Peano proves a solid square is a curve (the space-filling curve)	Menger and Urysohn define curves as part of topology

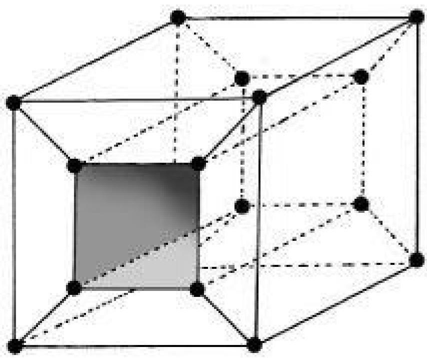
Archimedean solids which are semi-regular. Examples can be generated from the Platonic solids. If we slice off (truncate) some corners of the icosahedron we have the shape used as the design for the modern soccer ball. The 32 faces that form the panels are made up of 12 pentagons and 20 hexagons. There are 90 edges and 60 vertices. It is also the shape of buckminsterfullerene molecules, named after the visionary Richard Buckminster Fuller, creator of the geodesic dome. These 'bucky balls' are a newly discovered form of carbon, C_{60}, with a carbon atom found at each vertex.

Euler's formula

Euler's formula is that the number of vertices V, edges E and faces F, of a polyhedron are connected by the formula

$$V - E + F = 2$$

For example, for a cube, V = 8, E = 12 and F = 6 so V – E + F = 8 – 12 + 6 = 2 and, for buckminsterfullerene, V – E + F = 60 – 90 + 32 = 2. This theorem actually challenges the very notion of a polyhedron.



The cube with a tunnel

If a cube has a 'tunnel' through it, is it a real polyhedron? For this shape, V = 16, E = 32, F = 16 and V – E + F = 16 – 32 + 16 = 0. Euler's formula does not work. To reclaim the correctness of the formula, the type of polyhedron could be limited to those without tunnels. Alternatively, the formula could be generalized to include this peculiarity.

Classification of surfaces

is scaled up by a factor of 3 its area is 9 times its previous value or 3^2 and so the dimension is 2. If a cube is scaled up by this factor its volume is 27 or 3^3 times its previous value, so its dimension is 3. These values of the Hausdorff dimension all coincide with our expectations for a line, square, or cube.

If the basic unit of the Koch curve is scaled up by 3, it becomes 4 times longer than it was before. Following the scheme described, the Hausdorff dimension is the value of D for which $4 = 3^D$. An alternative calculation is that

$$D = \frac{\log 4}{\log 3}$$

which means that D for the Koch curve is approximately 1.262. With fractals it is frequently the case that the Hausdorff dimension is greater than the ordinary dimension, which is 1 in the case of Koch curve.

The Hausdorff dimension informed Mandelbrot's definition of a fractal – a set of points whose value of D is not a whole number. Fractional dimension became the key property of fractals.

The applications of fractals

The potential for the applications of fractals is wide. Fractals could well be the mathematical medium which models such natural objects as plant growth, or cloud formation.

Fractals have already been applied to the growth of marine organisms such as corals and sponges. The spread of modern cities has been shown to have a similarity with fractal growth. In medicine they have found application in the modelling of brain activity. And the fractal nature of movements of stocks and shares and the foreign exchange markets has also been investigated. Mandelbrot's work opened up a new vista and there is much still to be discovered.

the condensed idea
Shapes with fractional dimension

26 Chaos

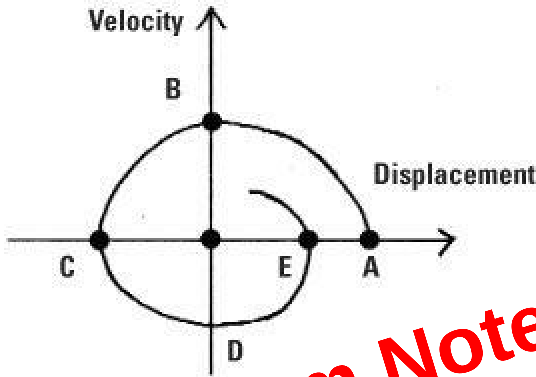
How is it possible to have a theory of chaos? Surely chaos happens in the absence of theory? The story goes back to 1812. While Napoleon was advancing on Moscow, his compatriot the Marquis Pierre-Simon de Laplace published an essay on the deterministic universe: if at one particular instant, the positions and velocities of all objects in the universe were known, and the forces acting on them, then these quantities could be calculated exactly for all future times. The universe and all objects in it would be completely determined. Chaos theory shows us that the world is more intricate that that.

In the real world we cannot know all the positions, velocities and forces exactly, but the corollary to Laplace's belief was that if we knew approximate values at one instant, the universe would not be much different anyway. This was reasonable, for surely sprinters who started a tenth of a second after the gun had fired would break the tape only a tenth of a second off their usual time. The belief was that small discrepancies in initial conditions meant small discrepancies in outcomes. Chaos theory exploded this idea.

The butterfly effect

The butterfly effect shows how initial conditions slightly different from the given ones, can produce an actual result very different from the predictions. If fine weather is predicted for a day in Europe, but a butterfly flaps its wings in South America then this could actually presage storms on the other side of the world – because the flapping of the wings changes the air pressure very slightly causing a weather pattern completely different from the one originally forecast.

We can illustrate the idea with a simple mechanical experiment. If you drop a ball-bearing through the opening in the top of a pinboard box it will progress downwards, being deflected one way or the other by the different pins it encounters on route until it reaches a finishing slot at the bottom. You might then attempt to let another identical ball-bearing go from the very same position with exactly the same velocity. If you could do this exactly then the Marquis de Laplace would be correct and the path followed by the ball would be exactly the same. If the first ball dropped into the third slot from the right, then so would

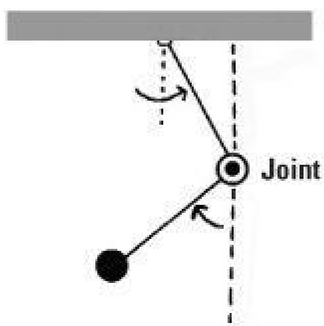


Phase diagram for the simple pendulum

This is not the case for the double pendulum in which the bob is at the end of a jointed pair of rods. If the displacement is small the motion of the double pendulum is similar to the simple pendulum, but if the displacement is large the bob swings, rotates, and lurches about and the displacement about the intermediate joint is seemingly random. If the motion is not forced, the bob will also come to rest but the curve that describes its motion is far from the well-behaved spiral of the single pendulum.

Chaotic motion

The characteristic of chaos is that a deterministic system may appear to generate random behaviour. Let's look at another example, the repeating, or iterative, formula $a \times p \times (1 - p)$ where p stands for the population, measured as a proportion on a scale from 0 to 1. The value of a must be somewhere between 0 and 4 to guarantee that the value of p stays in the range from 0 to 1.



Strange attractors

Dynamic systems can be thought of possessing 'attractors' in their phase diagrams. In the case of the simple pendulum the attractor is the single point at the origin that the motion is directed towards. With the double pendulum it's more complicated, but even here the phase portrait will display some regularity and be attracted to a set of points in the phase diagram. For systems like this the set of points may form a fractal (see page 100) which is called a 'strange' attractor that will have a definite mathematical structure. So all is not lost. In the new chaos theory, it is not so much 'chaotic' chaos that results as 'regular' chaos.

the condensed idea
The wildness of regularity

timeline

AD1812	1889	1961	1971	2004
Laplace publishes his essay on a deterministic world	Poncaré encounters chaos in his work on the three-body problem for which he is awarded a prize by King Oscar of Sweden	Lorenz observes the butterfly effect	Robert May investigates chaos in the population model	Chaos theory reaches popular culture in the film *The Butterfly Effect*

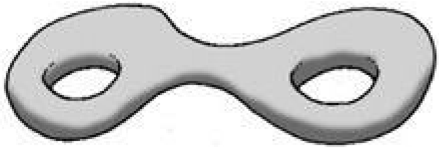
The western states of America

Let’s look at the map of the western states of America. If only blue, green and red were available we could start off by colouring Nevada and Idaho. It does not matter which colour we begin with so we’ll choose blue for Nevada and green for Idaho. So far so good. This choice would mean that Utah must be coloured red, and in turn Arizona green, California red, and Oregon green. This means that both Oregon and Idaho are coloured green so cannot be distinguished. But if we had four colours, with a yellow as well, we could use this to colour Oregon and everything would be satisfactory. Would these four colours – blue, green, red and yellow be sufficient for any map? This question is known as the four-colour problem.

The spread of the problem

Within 20 years of De Morgan recognizing the problem as one of significance, it became known within the mathematical community of Europe and America. In the 1860s, Charles Sanders Peirce, an American mathematician and philosopher, thought he had proved it but there is no trace of his argument.

The problem gained greater prominence through the intercession of the Victorian man of science Francis Galton. He saw publicity value in it and inveigled



A torus with two holes

number of holes, h	1	2	3	4	5	6	7	8
sufficient number of colours, C	7	8	9	10	11	12	12	13

and in general, $C = [\frac{1}{2}(7 + \sqrt{1 + 48h})]$. The square brackets indicate that we only take the whole number part of the term within them. For example, when $h = 8$, $C = [13.3107...] = 13$. Heawood's formula was derived on the strict understanding that the number of holes is greater than zero. Tantalizingly the formula gives the answer $C = 4$ if the debarred value $h = 0$ is substituted.

The problem solved?

After 50 years, the problem which had surfaced in 1852 remained unproved. In the 20th century the brainpower of the world's elite mathematicians was flummoxed.

Some progress was made and one mathematician proved that four colours were enough for up to 27 countries on a map, another bettered this with 31 countries and one came in with 35 countries. This nibbling process would take forever if continued. In fact the observations made by Kempe and Cayley in their very early papers provided a better way forward, and mathematicians found that they had only to check certain map configurations to guarantee that four colours were enough. The catch was that there was a large number of them – at the early stages of these attempts at proof there were thousands to check. This checking could not be done by hand but luckily the German mathematician Wolfgang Haken, who had worked on the problem for many years, was able to enlist the services of the American mathematician and computer expert Kenneth Appel. Ingenious methods lowered the number of configurations to fewer than 1500. By late June 1976, after many sleepless nights, the job was done and in partnership with their trusty IBM 370 computer, they had cracked the great problem.

Four colours will be enough

timeline

AD1852	1879	1890	1976	1994
Guthrie, De Morgan's student, puts the problem to him	Kempe is believed to have solved the problem	Heawood exposes errors in Kempe's proof and proves a five-colour theorem	Appel and Haken give a computer-based proof for the general result	The computer proof is simplified but remains a computer-based proof

The monkey on a typewriter

Alfred is a monkey who lives in the local zoo. He has a battered old typewriter with 26 keys for the letters of the alphabet, a key for a full stop, one for a comma, one for a question mark and one for a space – 30 keys in all. He sits in a corner filled with literary ambition, but his method of writing is curious – he hits the keys at random.

Any sequence of letters typed will have a nonzero chance of occurring, so there is a chance he will type out the plays of Shakespeare word perfect. More than this, there is a chance (albeit smaller) he will follow this with a translation into French, and then Spanish, and then German. For good measure we could allow for the possibility of him continuing on with the poems of William Wordsworth. The chance of all this is minute, but it is certainly not zero. This is the key point. Let's see how long he will take to type the soliloquy in Hamlet, starting off with the opening 'To be or'. We imagine 8 boxes which will hold the 8 letters including the spaces.

T	o		b	e		o	r

The number of possibilities for the first position is 30, for the second is 30, and so on. So the number of ways of filling out the 8 boxes is 30 × 30 × 30 × 30 × 30 × 30 × 30 × 30. The chance of Alfred getting as far as 'To be or' is 1 chance in 6.561×10^{11}. If Alfred hits the typewriter once every second there is an expectation he will have typed 'To be or' in about 20,000 years, and proved himself a particularly long-lived primate. So don't hold your breath waiting for the whole of Shakespeare. Alfred will produce nonsense like 'xo,h?yt?' for a great deal of the time.

How has the theory developed?

When probability theory is applied the results can be controversial, but at least the mathematical underpinnings are reasonably secure. In 1933, Andrey Nikolaevich Kolmogorov was instrumental in defining probability on an axiomatic basis – much like the way the principles of geometry were defined two millennia before.

Probability is defined by the following axioms:

1. the probability of all occurrences is 1
2. probability has a value which is greater than or equal to zero
3. when occurrences cannot coincide their probabilities can be added

From these axioms, dressed in technical language, the mathematical properties of probability can be deduced. The concept of probability can be widely applied. Much of modern life cannot do without it. Risk analysis, sport, sociology, psychology, engineering design, finance, and so on – the list is endless. Who'd have thought the gambling problems that kick-started these ideas in the 17th century would spawn such an enormous discipline? What were the chances of that?

the condensed idea
The gambler's secret system

timeline

c.AD1650s	1785	1812	1912	1933
The foundations of probability are laid by Pascal and Huygens	Condorcet applies probability to the analysis of juries and electoral systems	Laplace publishes his two volume *Analytical Theory of Probabilities*	Keynes publishes his *Treatise on Probability* which influences his theories of economics and statistics	Kolmogorov presents probability in an axiomatic way

32 Bayes's theory

The early years of the Rev. Thomas Bayes are obscure. Born in the southeast of England, probably in 1702, he became a nonconformist minister of religion, but also gained a reputation as a mathematician and was elected to the Royal Society of London in 1742. Bayes's famous Essay towards solving a problem in the doctrine of chances was published in 1763, two years after his death. It gave a formula for finding inverse probability, the probability 'the other way around', and it helped create a concept central to Bayesian philosophy – conditional probability.

Thomas Bayes has given his name to the Bayesians, the adherents of a brand of statistics at variance with traditional statisticians or 'frequentists'. The frequentists adopt a view of probability based on hard numerical data. Bayesian views are centred on the famous Bayes's formula and the principle that subjective degrees of belief can be treated as mathematical probability.

Conditional probability

Imagine that the dashing Dr Why has the task of diagnosing measles in his patients. The appearance of spots is an indicator used for detection but diagnosis is not straightforward. A patient may have measles without having spots and some patients may have spots without having measles. The probability that a patient has spots given that they have measles is a conditional probability. Bayesians use a vertical line in their formulae to mean 'given', so if we write

prob(a patient has spots | the patient has measles)

it means the probability that a patient has spots given that they have measles. The value of prob(a patient has spots|the patient has measles) is not the same as prob(the patient has measles|the patient has spots). In relation to each other, one is the probability the other way around. Bayes's formula is the formula of calculating one from the other. Mathematicians like nothing better than using notation to stand for things. So let's say the event of having measles is M and the event of a patient having spots is S. The symbol $\bar{S}$ is the event of a patient not having spots and $\bar{M}$ the event of not having measles. We can see this on a Venn diagram.

addressed by Bayes in his essay. To work out the probabilities we need to put in some numbers. These will be subjective but what is important is to see how they combine. The probability that if patients have measles, they have spots, prob(S|M) will be high, say 0.9 and if the patient does not have measles, the probability of them having spots prob(S|$\bar{M}$) will be low, say 0.15. In both these situations Dr Why will have a good idea of the values of these probabilities. The dashing doctor will also have an idea about the percentage of people in the population who have measles, say 20%. This is expressed as prob(M) = 0.2. The only other piece of information we need is prob(S), the percentage of people in the population who have spots. Now the probability of someone having spots is the probability of someone having measles and spots plus the probability that someone does not have measles but does have spots. From our key relations, prob(S) = 0.9 × 0.2 + 0.15 × 0.8 = 0.3. Substituting these values into Bayes's formula gives:

$$prob(M \mid S) = \frac{0.2}{0.3} \times 0.9 = 0.6$$

The conclusion is that from all the patients with spots that the doctor sees he correctly detects measles in 60% of his cases. Suppose now that the doctor receives more information on the strain of measles so that the probability of detection goes up, that is prob(S|M) the probability of having spots from measles, increases from 0.9 to 0.95 and prob(S|$\bar{M}$), the probability of spots from some other cause, declines from 0.15 to 0.1. How does this change improve his rate of measles detection? What is the new prob(M|S)? With this new information, prob(S) = 0.95 × 0.2 + 0.1 × 0.8 = 0.27, so in Bayes's formula, prob(M|S) is 0.2 divided by prob(S) = 0.27 and then all multiplied by 0.95, which comes to 0.704. So Dr Why can now detect 70% of cases with this improved information. If the probabilities changed to 0.99, and 0.01 respectively then the detection probability, prob(M|S), becomes 0.961 so his chance of a correct diagnosis in this case would be 96%.

Modern day Bayesians

The traditional statistician would have little quarrel with the use of Bayes's formula where the probability can be measured. The contentious sticking point is probability that yet another person selected at random shares a birthday with the first two is 2/365 so the probability this person does not share a birthday with either of the first two is one minus this (or 363/365). The probability of none of these three sharing a birthday is the multiplication of these two probabilities, or $(364/365) \times (363/365)$ which is 0.9918.

Continuing this line of thought for 4, 5, 6, . . . people unravels the birthday problem paradox. When we get as far as 23 people with our pocket calculator we get the answer 0.4927 as the probability that none of them shares a birthday. The negation of 'none of them sharing a birthday' is 'at least two people share a birthday' and the probability of this is $1 - 0.4927 = 0.5073$, just greater than the crucial ½.

If $n = 22$, the probability of two people sharing a birthday is 0.4757, which is less than ½. The apparent paradoxical nature of the birthday problem is bound up with language. The birthday result makes a statement about two people sharing a birthday, but it does not tell us which two people they are. We do not know where the matches will fall. If Mr Trevor Thomson whose birthday is on 8 March is in the room, a different question might be asked.

How many birthdays coincide with Mr Thomson's?

For this question, the calculation is different. The probability of Mr Thomson not sharing his birthday with another person is 364/365 so that the probability that he does not share his birthday with any of the other $n - 1$ people in the room is $(364/365)^{n} - 1$. Therefore the probability that Mr Thomson does share his birthday with someone will be one minus this value.

If we compute this for $n = 23$ this probability is only 0.061151 so there is only a 6% chance that someone else will have their birthday on 8 March, the same date as Mr Thomson's birthday. If we increase the value of n, this probability will increase. But we have to go as far $n = 254$ (which includes Mr Thomson in the count) for the probability to be greater than ½. For $n = 254$, its value is 0.5005. This is the cutoff point because $n = 253$ will give the value 0.4991 which is less than ½. There will have to be a gathering of 254 people in the room for a chance greater than ½ that Mr Thomson shares his birthday with someone else. This is perhaps more in tune with our intuition than with the startling solution of the classic birthday problem.

timeline

AD1806	1809	1885–8	1896	1904
Adrien-Marie Legendre fits data by least squares	Carl Friedrich Gauss uses the least-squares method in astronomical problems	Galton introduces regression and correlation	Pearson publishes contributions to correlation and regression	Spearman uses rank correlation as a tool for psychological studies

intriguing mathematical properties.

the condensed idea
Uncertainty in the gene pool

timeline

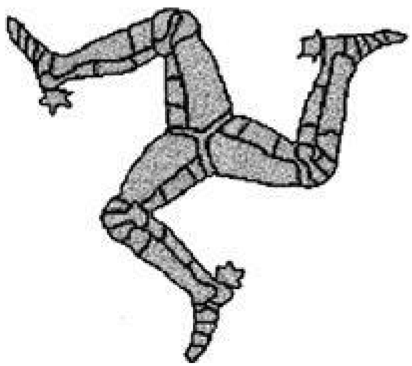
AD1718	1865	1908	1918	1953
Abraham de Moivre publishes the *Doctrine of Chances*	Mendel proposes the existence of genes and laws of inheritance	Hardy and Weinberg show why dominant genes do not supplant recessive genes	Fisher reconciles Darwin's theory with the Mendelian theory of heredity	The double helix structure of DNA is discovered

Can we set up a mirror so that an object looks the same in front of the mirror as in the mirror? The word MUM has mirror symmetry, but HAM does not; MUM in front of the mirror is the same as MUM in the mirror while HAM becomes MAH. A tripod has mirror symmetry, but the triskelion (tripod with feet) does not. The triskelion as the object before the mirror is right-handed but its mirror image in what is called the image plane is left-handed.

Rotational symmetry

We can also ask whether there is an axis perpendicular to the page so that the object can be rotated in the page through an angle and be brought back to its original position. Both the tripod and the triskelion have rotational symmetry. The triskelion, meaning 'three legs', is an interesting shape. The right-handed version is a figure which appears as the symbol of the Isle of Man and also on the flag of Sicily.

If we rotate it through 120 degrees or 240 degrees the rotated figure will coincide with itself; if you closed your eyes before rotating it you would see the same triskelion when you opened them again after rotation.



The Isle of Man triskelion

The curious thing about the three-legged figure is that no amount of rotation keeping in the plane will ever convert a right-handed triskelion into a left-handed one. Objects for which the image in the mirror is distinct from the object in front of the mirror are called chiral – they look similar but are not the same. The molecular structure of some chemical compounds may exist in both right-handed and left-handed forms in three dimensions and are examples of chiral objects.

A practical example

Suppose the matrix A represents the output of the AJAX company in one week. The AJAX company has three factories located in different parts of the country and their output is measured in units (say 1000s of items) of the four products it produces. In our example, the quantities, tallying with matrix A opposite, are:

	product 1	product 2	product 3	product 4
factory 1	7	5	0	1
factory 2	0	4	3	7
factory 3	3	2	0	2

In the next week the production schedule might be different, but it could be written as another matrix B. For example B might be given by

$$B = \begin{pmatrix} 9 & 4 & 1 & 0 \\ 0 & 5 & 1 & 8 \\ 4 & 1 & 1 & 0 \end{pmatrix}.$$

What is the total production for both weeks? The matrix theorist says it is the matrix A + B where corresponding numbers are added together,

$$A + B = \begin{pmatrix} 7+9 & 5+4 & 0+1 & 1+0 \\ 0+0 & 4+5 & 3+1 & 7+8 \\ 3+4 & 2+1 & 0+1 & 2+0 \end{pmatrix} = \begin{pmatrix} 16 & 9 & 1 & 1 \\ 0 & 9 & 4 & 15 \\ 7 & 3 & 1 & 2 \end{pmatrix}$$

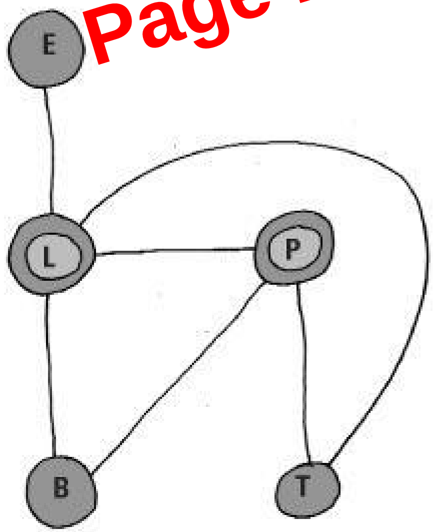
Easy enough. Sadly, matrix multiplication is less obvious. Returning to the AJAX company, suppose the unit profit of its four products are **3**, **9**, **8**, **2**. We can certainly compute the overall profit for Factory 1 with outputs 7, 5, 0, 1 of its four products. It works out as 7 × **3** + 5 × **9** + 0 × **8** + 1 × **2** = 68.

But instead of dealing with just one factory we can just as easily compute the total profits T for all the factories which does not arise in ordinary algebra where the order of multiplying two numbers together makes no difference to the answer.

Another difference occurs with inverses. In ordinary algebra inverses are easy to calculate. If a= 7 its inverse is 1/7 because it has the property that 1/7 × 7 = 1. We sometimes write this inverse as $a^{-1} = 1/7$ and we have $a^{-1} \times a = 1$.

An example in matrix theory is $A = \begin{pmatrix} 1 & 2 \\ 3 & 7 \end{pmatrix}$ and we can verify that $A^{-1} = \begin{pmatrix} 7 & -2 \\ -3 & 1 \end{pmatrix}$

because $A^{-1} \times A = \begin{pmatrix} 7 & -2 \\ -3 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 2 \\ 3 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$



where $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is called the identity matrix and is the matrix counterpart of 1 in ordinary algebra. In ordinary algebra, only 0 does not have an inverse but in matrix algebra many matrices do not have inverses.

Travel plans

$$74^5 = 74 \times 74 \times 74 \times 74 \times 74 = 2{,}219{,}006{,}624$$

and

$$2{,}219{,}006{,}624 = 8{,}983{,}832 \times 247 + 120$$

so dividing his huge number by 247 he gets the remainder 120. Sender's encrypted message is 120 and he transmits this to Receiver. Because the numbers 247 and 5 were publicly available anyone could encrypt a message. But not everyone could decrypt it. Dr R. Receiver has more information up his sleeve. He made up his personal number 247 by multiplying together two prime numbers. In this case he obtained the number 247 by multiplying $p = 13$ and $q = 19$, but only he knows this.

This is where the ancient theorem due to Leonhard Euler is taken out and dusted down. Dr R. Receiver uses the knowledge of $p = 13$ and $q = 19$ to find a value of a where $5 \times a \equiv 1$ modulo $(p - 1)(q - 1)$ where the symbol ≡ means equals in modular arithmetic. What is a so that dividing $5 \times a$ by $12 \times 18 = 216$ leaves remainder 1? Skipping the actual calculation he finds $a = 173$.

Because he is the only one who knows the prime numbers p and q, Dr Receiver is the only one who can calculate the number 173. With it he works out the remainder when he divides the huge number 120^{173} by 247. This is outside the capacity of a hand held calculator but is easily found by using a computer. The answer is 74, as Euler knew two hundred years ago. With this information, Receiver looks up word 74 and sees that J is back in town.

You might say, surely a hacker could discover the fact that $247 = 13 \times 19$ and the code could be cracked. You would be correct. But the encryption and decryption principle is the same if Dr Receiver had used another number instead of 247. He could choose two very big prime numbers and multiply them together to get a much larger number than 247.

Finding the two prime factors of a very large number is virtually impossible – what are the factors of 24,812,789,922,307 for example? But numbers much larger than this could also be chosen. The public key system is secure and if the might of supercomputers joined together are successful in factoring an encryption number, all Dr Receiver has to do is increase its size still further. In the end it is considerably easier for Dr Receiver to 'mix boxes of black sand and white sand together' than for any hacker to unmix them.

the condensed idea
Keeping messages secret

of the schoolgirls comes into its own.

Monday			Tuesday			Wednesday			Thursday			Friday			Saturday			Sunday		
a	A	V	b	B	V	c	C	V	d	D	V	e	E	V	f	F	V	g	G	V
b	E	D	c	F	E	d	G	F	e	A	G	f	B	A	g	C	B	a	D	C
c	B	G	d	C	A	e	D	B	f	E	C	g	F	D	a	G	E	b	A	F
d	f	g	e	g	a	f	a	b	g	b	c	a	c	d	b	d	e	c	e	f
e	F	C	f	G	D	g	A	E	a	B	F	b	C	G	c	D	A	d	E	B

It is called cyclic since on each subsequent day the walking schedule is changed from **a** to **b**, **b** to **c**, down to **g** to **a**. The same applies to the upper-case girls **A** to **B**, **B** to**C**, and so on, but **V**ictoria remains unmoved.

The underlying reason for the choice of notation is that the rows correspond to lines in the Fano geometry (see page 115). Kirkman's problem isn't only a parlour game but one that's part of mainstream mathematics.

the condensed idea
How many combinations?

timeline

c.1800BC	c.AD1100	1850	1930	1971
The Rhind papyrus is written in Egypt	Bhaskara deals with permutations and combinations	Kirkman poses the 15 schoolgirls problem	Frank Ramsey works in combinatorics	Ray-Chaudhuri and Wilson prove the existence of general Kirkman's systems

4	9	8
11	7	3
6	5	10

Remarkably, this is a magic square consisting of the consecutive numbers 3, 4, 5, up to 11. We also find that the number of letters of the magic sums of both 3×3 squares (21 and 45) is 9 and fittingly 3 × 3 = 9.

the condensed idea
Mathematical wizardry

timeline

c.2800BC	c.AD1690	1693	1770	1986
The legend of the Lo Shu square is born	de la Loubère produces the Siamese method for constructing magic squares	Bernard Frenicle de Bessy lists all the 880 possible 4x4 magic squares	Euler produces a squared square	Sallows creates his letter-based square

45 The diet problem

Tanya Smith takes her athletics very seriously. She goes to the gym every day and monitors her diet closely. Tanya makes her way in the world by taking part-time jobs and has to watch where the money goes. It is crucial that she takes the right amount of minerals and vitamins each month to stay fit and healthy. The amounts have been determined by her coach. He suggests that future Olympic champions should absorb at least 120 milligrams (mg) of vitamins and at least 880 mg of minerals each month. To make sure she follows this regime Tanya relies on two food supplements. One is in solid form and has the trade name Solido and the other is in liquid form marketed under the name Liquex. Her problem is to decide how much of each she should purchase each month to satisfy her coach.

The classic diet problem is to organize a healthy diet and pay the lowest price for it. It was a prototype for problems in linear programming, a subject developed in the 1940s that is now used in a wide range of applications.

	Solido	Liquex	Requirements
Vitamins	2 mg	3 mg	120 mg
Minerals	10 mg	50 mg	880 mg

At the beginning of March Tanya takes a trip to the supermarket and checks out Solido and Liquex. On a back of a packet of Solido she finds out it contains 2 mg vitamins and 10 mg minerals, while a carton of Liquex contains 3 mg vitamins and 50 mg minerals. She dutifully fills her trolley with 30 packets of Solido and 5 cartons of Liquex to keep herself going for the month. As she proceeds towards the checkout she wonders if she has the right amount. First she calculates how many vitamins she has in the trolley. In the 30 packets of Solido she has $2 \times 30 = 60$ mg vitamins and in the Liquex, $3 \times 5 = 15$. Altogether she has $2 \times 30 + 3 \times 5 = 75$ mg vitamins. Repeating the calculation for minerals, she has $10 \times 30 + 50 \times 5 = 550$ mg minerals.

As the coach required her to have at least 120 mg vitamins and 880 mg minerals, she needs more packets and cartons in the trolley. Tanya's problem is juggling the right amounts of Solido and Liquex with the vitamin and mineral requirements. She goes back to the health section of the supermarket and puts more packets and cartons into her trolley. She now has 40 packets and 15 case is to minimize the cost of transportation. In some linear programming problems the objective is to maximize (like maximizing profit). In other problems the variables only take integer values or just two values 0 or 1, but these problems are quite different and require their own solution procedures.

It remains to be seen whether Tanya Smith wins her gold medal at the Olympic Games. If so, it will be another triumph for linear programming.

The condensed idea
Keeping healthy at least cost

timeline

AD1826	1902	1945	1947	1984
Fourier anticipates linear programming; Gauss solves linear equations by Gaussian elimination	Farkas gives a solution of inequality systems	Stigler solves the diet problem by a heuristic method	Dantzig formulates the simplex method and solves the diet problem by linear programming	Karmarker derives a new algorithm for solving linear programming problems

simple principles of relativity and the whole theory unfolded. In particular, he showed that the energy of a particle E is determined by the equation

$E = \times mc^2$. For the energy of a body at rest (when v = 0 and so = 1), this leads to the iconic equation showing that mass and energy are equivalent:

$$E = mc^2$$

Lorentz and Einstein were both proposed for the Nobel Prize in 1912. Lorentz had already been given it in 1902, but Einstein had to wait until 1921 when he was finally awarded the prize for work on the photoelectric effect (which he had also published in 1905). That was quite a year for the Swiss patent clerk.

Einstein vs Newton

For observations on slow-moving railway trains there is only a very small difference between Einstein's relativity theory and the classical Newtonian theory. In these situations the relative speed v is so small compared with the speed of light that the value of the Lorentz factor ɑ is almost 1. In this case the Lorentz equations are virtually the same as the classical Galilean transformations. So for slow speeds Einstein and Newton would agree with each other. Speeds and distances have to be very large for the differences between the two theories to be apparent. Even the record breaking French TGV train has not reached these speeds yet and it will be a long time in the development of rail travel before we would have to discard the Newtonian theory in favour of Einstein's. Space travel will force us to go with Einstein.

The general theory of relativity Einstein published his general theory in 1915. This theory applies to motion when frames of reference are allowed to accelerate in relation to each other and links the effects of acceleration with those of gravity.

Using the general theory Einstein was able to predict such physical phenomena as the deflection of light beams by the gravitational fields of large objects such as the Sun. His theory also explained the motion of the axis of Mercury's rotation. This precession could not be fully explained by Newton's theory of gravitation and the force exerted on Mercury by the other planets. It was a problem that had bothered astronomers since the 1840s.

The appropriate frame of reference for the general theory is that of the four-dimensional space–time. Euclidean space is flat (it has zero curvature) but

Einstein's four-dimensional space–time geometry (or Riemannian geometry) is curved. It displaces the Newtonian force of gravity as the explanation for objects being attracted to each other. With Einstein's general theory of relativity it is the curvature of space–time which explains this attraction. In 1915 Einstein launched another scientific revolution.

the condensed idea
The speed of light is absolute

timeline

c.AD1632	1676	1687	1881	1887	1905	1915
Galileo gives the 'Galilean transformations' for falling bodies	Römer calculates the speed of light from observations of the moons of Jupiter	Newton's *Principia* describes the classical laws of motion	Michelson measures the speed of light with great accuracy	The Lorentz transformations are first written down	Einstein publishes *The field equations for gravitation*, describing general relativity	Einstein publishes *On the electrodynamics of moving bodies*, the paper that describes special relativity

50 The Riemann hypothesis

The Riemann hypothesis represents one of the stiffest challenges in pure mathematics. The Poincaré conjecture and Fermat's last theorem have been conquered but not the Riemann hypothesis. Once decided, one way or the other, elusive questions about the distribution of prime numbers will be settled and a range of new questions will be opened up for mathematicians to ponder.

The story starts with the addition of fractions of the kind

$$1+\frac{1}{2}+\frac{1}{3}$$

The answer is 1⅚ (approximately 1.83). But what happens if we keep adding smaller and smaller fractions, say up to ten of them?

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\frac{1}{9}+\frac{1}{10}$$

Number of terms	Total (approximate)
1	1
10	2.9
100	5.2
1,000	7.5
10,000	9.8
100,000	12.1
1,000,000	14.4
1,000,000,000	21.3

Using only a handheld calculator, these fractions add up to approximately 2.9 in decimals. A table shows how the total grows as more and more terms are added.

The series of numbers

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\ldots$$

is called the harmonic series. The harmonic label originates with the Pythagoreans who believed that a musical string divided by a half, a third, a quarter, gave the musical notes essential for harmony.

In the harmonic series, smaller and smaller fractions are being added but what happens to the total? Does it grow beyond all numbers, or is there a barrier somewhere, a limit that it never rises above? To answer this, the trick is to group the terms, doubling the runs as we go. If we add the first 8 terms (recognizing that $8 = 2 \times 2 \times 2 = 2^3$) for example

$$S_{2^3} = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right)$$

(where S stands for sum) and because $\frac{1}{3}$ is bigger than $\frac{1}{4}$ and $\frac{1}{5}$ is bigger than $\frac{1}{8}$ (and so on), this is greater than

$$1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) = 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}$$

So we can say

$$S_{2^3} > 1 + \frac{3}{2}$$

and more generally

$$S_{2^k} > 1 + \frac{k}{2}$$

If we take $k = 20$, so that $n = 2^{20} = 1{,}048{,}576$ (more than a million terms), the sum of the series will only have exceeded 11 (see table). It is increasing in an excruciatingly slow way – but, a value of k can be chosen to make the series total beyond any preassigned number, however large. The series is said to diverge to infinity. By contrast, this does not happen with the series of squared terms

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \ldots$$

We are still using the same process: adding smaller and smaller numbers together, but this time a limit is reached, and this limit is less than 2. Quite dramatically the series converges to $\pi^2/6 = 1.64493\ldots$