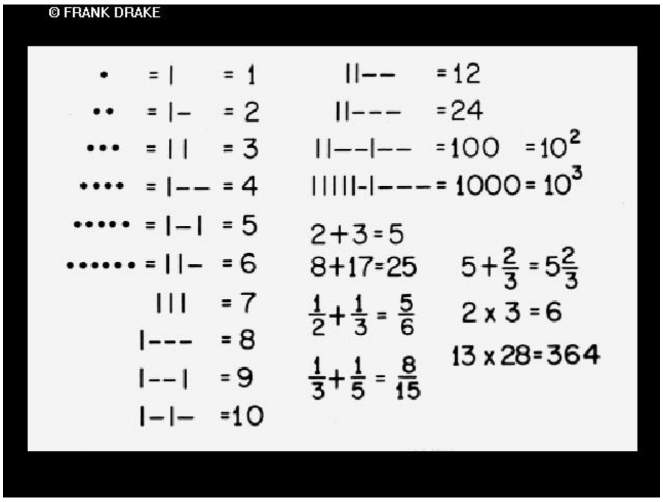
Where do mathematical symbols come from?

Professor Sarah Hart

27 April 2021

How do you communicate with aliens? When the Voyager 1 space probe was launched on September 5th 1977, it contained a "golden record" – the aim was to give any alien civilization that discovered it a snapshot of humanity – what we look like, something about our biology, what our culture is, and most importantly where we are. The contents of the record were chosen for NASA by a team led by Carl Sagan. It's a fascinating story told in the book *Murmurs of Earth* (1), which is well worth a look if you can get hold of a copy. The disc contained music by Bach and other composers, as well as human voices, and over a hundred pictures. But how can we send information like "our planet is 93 million miles from our star"? Aliens don't know what miles are, and they don't know the symbols we use for numbers. The first few pictures are as follows. Picture 1 is a circle – this is for calibration so that the aliens, in trying to get data off this disc, know they are doing it right (after all, circles are a universal mathematical concept – it's inconceivable that an intelligent lifeform would not have that concept). Picture 2 is an image of the region of space with our Sun in it. Picture 3 is this.



In subsequent pictures, units are defined in reference to things like hydrogen atoms and the wavelength of light, using these very basic numerical building blocks and symbols.

In Picture 3, we see the digits 1 to 9, then the decimal place system with 10, 12, 24 as examples, then 100, 1000 and the exponentials 10^2 and 10^3. The vertical and horizontal lines (| and −) are used as the two symbols for the binary expression of numbers. Also defined are +, × and fractions. Decimals, interestingly, are not defined or used anywhere, perhaps because of concerns that decimal points would be missed, or interpreted as damage to the pictures.

In this lecture, we are going to look at mathematical symbols and notation. One could quite reasonably ask why this is worth doing. After all, mathematics is universal. Surely whatever symbols you use to write the equation $a^2 + b^2 = c^2$, there can be no effect on the underlying mathematics? The response to that is "yes, but". Mathematics is done by people, and people need to understand concepts. Our choice of notation can make life much easier or much harder. It can pave the way to new understanding, or it can obscure underlying patterns and make some generalisations almost unthinkable.

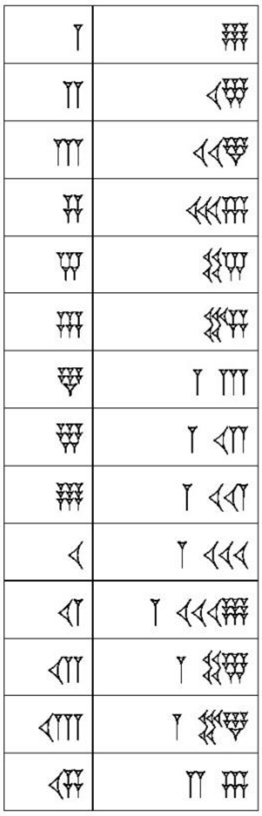
For anyone interested in language, the naming of concepts is already worthy of exploration. But the story of mathematical words and symbols is also the story of how mathematical ideas have spread around the world. On the Voyager image, we have what we call Arabic numerals, for example – but why? That's a misnomer, as we'll see, but the bigger question is: why have we settled on these particular number symbols out of all the many ancient notations? What about these symbols for addition and multiplication, where did they come from? And exponentiation? And what about zero? We'll tell these stories and more, but there is one other symbol here that is such an elemental part of the way we write mathematics that we haven't even mentioned it. It's the equals sign. One of the great surprises when you start learning about these symbols is that the equals sign, which seems so fundamental, is not, like the numerals, something invented so long ago that we cannot say who thought of it, passed down by nameless scribes over millennia. It was invented by a Welshman working in Tudor England, Robert Recorde. We'll come back to him later, but he had a real gift for clarity, and introduced many symbols and words into English, one of which was the equals sign, which he decided to represent as a pair of parallel lines, because "no two things can be more equal". But we are getting ahead of ourselves. I want to take a chronological tour through the origins of mathematical notation, from the beginnings of the representation of number, through to when actual mathematics started to be written with symbols rather than words, and then the choice of symbols and how they affect our understanding. We'll finish with a few fun numerolingistical diversions.

Counting and Numbers

Symbols for numbers have been with us since the beginnings of writing. To begin with, this would simply have been lines |, ||, ||| and so on, often grouped, for example into fives as we still do with our tally system. As agriculture developed, it became important to understand when to plant your crops, when the river would flood, but also when to sacrifice your ox to appease the sky-gods and make sure the sun continues to rise. An understanding of the calendar, and excellent record keeping and astronomical observations, became a cultural imperative. Babylonian tablets show us impressively sophisticated mathematical and astronomical calculations, using a base 60 number system. This undoubtedly comes from the "360 day year" idea (itself dating back at least to the Sumerians). Now, 360 is a very convenient number – it's where we get the number of degrees in a circle as well – because it can be divided up so neatly – into four quarters, twelve months, and so on. This should not be interpreted as saying that the Babylonians thought there were really 360 days in a year – they were much better astronomers than that. They, and several other ancient civilizations, would have twelve 30-day months, supplemented either with five extra days each year or with an intercalary month every six years. The ancient Egyptian myth of Ra and Nut explains it like this: Ra, the sun god, decreed that Nut (the sky goddess) "shall not give birth on any day of the year". So, Nut played a gambling game with Khonsu (the moon god). Every time Khonsu lost, he had to give Nut some of his moonlight. He lost so many times that Nut got enough moonlight to make five extra days, which were not really parts of the year. She could then have five children: Orisiris (god of the dead), Horus (god of war), Set (god of evil), Isis (goddess of magic), and Nephthys (goddess of water). And that's why there are 365 days in the year.

Anyway, back to Babylon. Their sexagesimal system extended to more than just days in a year. There were 60 minutes in an hour and also 60 shekels in a mina (these were units of weight as well as units of currency). Using 60, not 10, makes many more numbers easily writable as fractions: in fact, 60 is the smallest number divisible by 1, 2, 3, 4, 5 and 6.

Let's see some Babylonian numbers. They wrote in cuneiform script on clay tablets, using a stylus, so the marks were a mix of straight lines and wedges. Here is my slightly neatened up version of the front of a particular tablet from the Hilprecht Collection at the University of Jena, with the catchy name of HS 0217a. It's Middle Babylonian, from between 1400 and 1100 BC, found in Nippur.



We can decipher these marks with a mix of common sense and mathematical intuition. The first column surely must be the numbers 1 to 14. Then in the second column, we get 9, 18, 27 and so on – it's clearly the nine times table. But something odd happens after 54. This number must, from context, be 63. But it is written as 𒁹 followed by 𒁹𒁹𒁹, with a small space inbetween. The first 𒁹 must mean "60". This positional system is exactly analogous to how we write 63, where the 6 means six tens, because of its position. It is an extremely good idea but with, at this stage, a fatal flaw. How can we tell if 𒁹 means 1, 60, or 3600? Initially, the Babylonians would leave a space, but this is very hard to spot, especially where the "space" would be required at the end of the number. On the reverse side of this tablet, we see the problem – twenty times nine, or 180, is written the same way as 3 would be, and it is only from context that we can tell it must represent three 60s, not three 1s. Eventually, a placeholder symbol was introduced – this is in some sense a "zero", but it was not considered a number at this point, just a notational device. But the ambiguities of the notation perhaps explain why this otherwise very efficient method, requiring as it did only ten different symbols (plus, later, a placeholder symbol) was not more widely adopted.

Since the remaining number systems we'll discuss are decimal (that is, base 10), we mention here that there are also instances of other bases being used. Base 10 is natural because we have ten fingers, but some cultures have used base 5 and some base 20. There are also vestiges of 12 being important in number words – eleven and twelve mean one "extra" and two "extra", before we revert to the three-and-ten of thirteen and the other teens. The Mayans had a base 20 system. You can see remnants of a base 20 system in number systems of some European languages, for example the French for 80 is quatre-vingts.

The Ancient Egyptians had a different system from the Babylonians, based on powers of ten. They had symbols for 1, 10, 100, 1000, 10,000, 100,000 and 1 million. The number 1 is a stick (𓏺), 10 a bent stick (∩), 100 a coiled rope (𓍢), 1,000 a lotus plant (𓆼), and 10,000 a pointing finger (𓂭). They also had symbols for several unit fractions – these are still sometimes known as Egyptian fractions, along with $\frac{2}{3}$. You could write $\frac{3}{8}$, for instance, but as $\frac{1}{4}+\frac{1}{2}$. Numbers would be written additively, as shown.

1,234: 𓆼 𓍢𓍢 𓎆𓎆𓎆 𓏺𓏺𓏺𓏺 40,905: 𓂭𓂭𓂭𓂭 𓍢𓍢𓍢𓍢𓍢𓍢𓍢𓍢𓍢 𓏺𓏺𓏺𓏺𓏺 495: 𓍢𓍢𓍢𓍢 𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆 𓏺𓏺𓏺𓏺𓏺 200: 𓍢𓍢 2000: 𓆼𓆼

As we can see, there is no ambiguity and no need for placeholders, a great advantage over the Babylonian system. But it is less efficient, as more symbols are needed.

The Ancient Greeks began by using a so-called acrophonic system where number symbols were taken from the first letters of the words representing them. Thus, I was 1 (for iota), 5 was Π (Pente), 10 was Δ (Deka), 100 was H (Ekaton/Ηκατόν). The number 302 is HHHII.

To write 5 times a number, place the number under the symbol Γ. So, in acrophonic writing, Δ Δ Δ Δ added to Δ Δ would be Δ Δ.

Later, the Greeks switched to an alphabetic representation. Here are the Greek numbers – note that a couple of the symbols may be unfamiliar, such as the digamma ς, which was dropped from the Greek alphabet several hundred years BC.

α	β	γ	δ	ε	ς	ζ	η	θ
1	2	3	4	5	6	7	8	9
ι	κ	λ	μ	ν	ξ	ο	π	ϙ
10	20	30	40	50	60	70	80	90
ϱ	σ	τ	υ	φ	χ	ψ	ω	ϡ
100	200	300	400	500	600	700	800	900

This was a dreadful solution for large numbers, but more convenient for small numbers, because each number from 1 to 9 had a separate symbol, which reduced the need for counting or mental arithmetic in the reading and writing of numbers. Instead of HHHH, we can write υ (upsilon), and instead of Δ Δ Δ Δ added to Δ Δ becoming Δ Δ, we now have μ(mu) added to κ (kappa) is ξ (xi).

The number 47 would be δζ (delta zeta), and there is no chance of getting it confused with 470 (υο – upsilon omicron). But there are problems with the Greek alphabetic system. The most obvious is that you simply run out of letters. More subtly, it impedes the development of symbolic algebra, as we'll see later. The use of normal letters as numbers had consequences in Hebrew writing (which used an alphabetic system like the Greek one): if letters are also numbers, then any word is also a number – this means there is the temptation to indulge in numerology. More seriously, you weren't allowed to write down the name of God. To write 14 in Hebrew, you would write 10 (Yod י) and then 4 (Daletד), from right to left of course. Hence, 14 isיד. Similarly, to write 15, you "should" write 10 (Yod י) and then 5 (He ה). But unfortunately, these are also the first two letters of the name of God. Thus, 15 has to be written instead as 9 (Tet ט) + 6 (Vav ו), or טו.

The Romans had a juxtapositional system with different letters for 1, 5, 10, 50, 100, 500, 1000. They introduced the innovation of subtraction on the left – if you see a smaller number to the left of a larger one, you subtract it. Thus, IX means 9, XI means 11, MDXCVII means M + D + (XC) + V + I + I, which is 1000 + 500 + (100-10) + 5 + 1 + 1 = 1597 (the year Gresham College was founded). This is all very well for writing down years, but try doing long multiplication in Roman numerals!

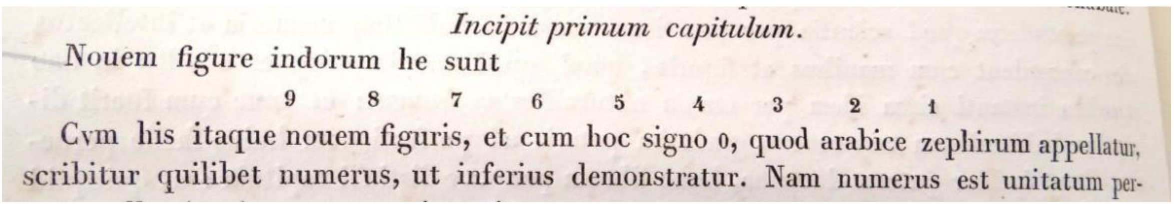
The Chinese system also had symbols for each power of ten, avoiding the need for a zero placeholder. But their system was much better than the Roman or Greek one, and closer to a true decimal place-value system, because they have symbols for each number 1 to 9, then symbols for 10, 100, 1,000 and so on, and to write something like 4,070, they would (to borrow Roman numerals) write 4M 7X. This is ingenious, avoiding all the problems with placeholders as well as the need to count lots of tiny symbols.

Why didn't the Romans and Greeks develop a place-value system? The Greeks might have got there eventually, because numbers above a thousand, due to running out of letters, had to start again from the numbers for 1, 2, 3 and so on. For example, 4,400 would be written δυ (δ is 4, and υ is 400). Why didn't they realise they could use δ for 4, 40, and 400 if they just added one more symbol for an empty place? It's important to remember that they were not doing calculations with these written numbers. Both the Greeks and Romans used counting boards. They didn't have bits of scrap paper lying around to do long multiplication on. Calculations were done with *calculi* – little stones or counters, on counting boards or the abacus (hence the term calculation, and later calculus). We don't have extant specimens of Roman counting boards, but we do have other evidence, such as clues from idioms in Latin. For example, the phrase *vocare aliquem ad calculos*, meaning to settle up with someone, literally means "to call someone to the counting table". Meanwhile the Greek historian Polybius, in the second century BC, wrote that "the courtiers who surround kings are exactly like the counters on the lines of a counting board, for, depending on the will of the reckoner, they may be valued either at no more than a mere *chalkós*, or else at a whole talent!" We do have an ancient Greek counting board – the Salamis tablet, which, a bit like our hundreds, tens and units columns that are used when we first learn counting, has columns for talents, minae, drachmas, obols, and fractions of obols. A talent is 60 minae, a mina is 100 drachmas, a drachma is six obols, and a chalkós is $\frac{1}{8}$th of an Obol. Thus, by moving your counter from the chalkós column to the talent column, you can increase the value by a factor of 288,000. This was how calculation was done, but because of the lack of a symbol for an empty column, it could not be how the finished answers were written. And so, for a thousand years, the Western World struggled with Roman numerals.

The numbers we used today began in India as a development of an ancient Brahmi number system, rather like the Hebrew and Greek one, with symbols for 1 to 9, then 10, 20, 30 and so on. The revolutionary idea probably came at some point in India in the first few centuries AD when all but the first nine symbols were dropped, and we see in Sanskrit/Hindu writing the use of just the numerals 1 to 9 and then place-values for powers of ten, with a dot as a place-holder if a particular power of ten was absent. This dot evolved into our zero. Initially, it was not a number in its own right. One early instance of zero being treated as a number is in a book by the Indian mathematician Brahmagupta, around 630AD. He gives rules for using zero (still written as a filled dot). For example, he says that a number subtracted from itself is zero. The Indian numbers were adopted in the Arab world within a century or two. The renowned Arab mathematician Al-Khwarizmi learned of them from the Arabic translation of Brahmagupta's work, and he then wrote a text book on arithmetic using these new Indian numbers. Then, in about 820AD, he wrote *Hisab aljabr w'almuqabala*, or the *Book of restoration and equalization*, a book on how to solve equations, again using these Indian numbers. That "aljabr" is the origin of our word "algebra". The Indian system spread through the Arab world, and thence to Europe, which is why we wrongly refer to Arabic numerals.

But how did the numbers spread to Europe? Yes, Latin translations of al-Khwarizmi's work did eventually become available, but it was through trade routes that the numbers really spread, and in particular through the popularising work of Leonardo of Pisa, better known as Fibonacci.

As a young man, Fibonacci travelled widely with his father, learning the mathematics used in trade. He was struck by the utility of the Arabic/Indian number system, not only its efficiency but in the fact that one could actually perform calculations directly with the numbers, rather than having to do everything with an abacus or counting board first. He wrote a book called Liber Abbaci (or Liber Abaci) – the Book of Calculation, in 1202, describing the new numbers and how to use them to calculate. Here is the moment when he introduces the numbers (this is from a 19th century edition of the book).


Incipit primum capitulum.
Nouem figure indorum he sunt
9 8 7 6 5 4 3 2 1
Cvm his itaque nouem figuris, et cum hoc signo 0, quod arabice zephirum appellatur, scribitur quilibet numerus, ut inferius demonstratur. Nam numerus est unitatum per-

A rough translation is: These are the nine figures of the Indians 9, 8, 7, 6, 5, 4, 3, 2, 1. With these nine figures, and with this sign 0, which is called zephyr by the Arabs, any number can be written, as is shown below."

We can see from this that Fibonacci was not quite thinking of zero yet as a number like the rest. It was, in fact, treated with suspicion for a very long time. People hated that it meant nothing on its own, but suddenly if you wrote a number in front of it, it suddenly made that number bigger. Preposterous! This perception of zero as not really a number persisted for centuries. In Shakespeare's King Lear, for instance, the Fool makes plain to Lear that he is brought low by saying "thou art an O without a figure. I am better than thou art now. I am a fool, thou art nothing". The "O" is not a figure, it is something that can be attached to a figure, but nothing (literally) on its own.

We must remember, too, that it wasn't just new numbers that were being introduced. It was a wholesale revolution in the way calculation was done: calculation with the calculi of the counting board or abacus, versus this "algorismus" or "ciphering" done by manipulation of strange foreign symbols. In 1299, the City Council of Florence banned Arabic numerals from financial records, in the *Statuto Dell'Arte di Cambio*, on the grounds that it was too easy to tamper with records, turning 0 into 6 or 9. All numbers had to be written in letters (we do something similar today on the rare occasions when we still write cheques). Of course, this didn't stop people using the new numbers informally. There was a practical impediment though, in that number calculations have intermediate steps that need to be written down. Paper, after all, had only been in use in Europe since the 12th century, and was still very expensive. It would have felt like a needless expense to use it for jotting down calculations only to have to throw them away once the calculation was complete. Why not stick to your counting board? The battle between the two technologies lasted at least three hundred years. But, eventually, Arabic numerals won out over Roman numerals. We have seen that Roman numerals make calculation without a counting board impossible. But they also present another problem, which is implicit already in the work of early mathematicians like Diophantus. If your numbers are just normal letters, you can't also use letters for variables. This impeded the invention of symbolic algebra, as we'll see in our next section.

From Al-jabr to Algebra

What we now call algebra began with word problems. Babylonian tablets contain such problems as this:

> *I have multiplied length and width, thus obtaining the area. Then I added to the area, the excess of the length over the width: [the result was] 183. Moreover, I have added length and width: 27. Required: length, width, and area.*

The tablet goes on to give a step-by-step method for finding the solution. Nowadays we would probably convert it into a quadratic equation and solve by factorization or using the formula. Discussion of what we would now call equations continued to be done purely in words for many centuries. This is known as rhetorical algebra. A much later, and more lyrical, example, is in the work of the twelfth century Indian mathematician and poet Bhaskara, who wrote all his mathematical works in verse, as was the convention at that time in India. One of his poems in a book dedicated to his daughter Lilivati, reads (in translation of course):

A fifth part of a swarm of bees came to rest on the flower of Kadamba, a third on the flower of Silinda.
Three times the difference between these two numbers flew over a flower of Krutaja, and one bee alone remained in the air, attracted by the perfume of a jasmine and a bloom. Tell me, beautiful girl, how many bees were in the swarm?

The solution to such a problem would also be expressed in words.

One of the great achievements of early mathematics is Diophantus's *Arithmetica*. Diophantus of Alexandria was Greek mathematician working in the third century AD, best known for his work, in *Arithmetica*, on whole number equations, now known as Diophantine equations. The book was very influential. It was in a Latin translation of *Arithmetica* that Fermat wrote his infamous marginal note to the effect that $x^n + y^n = z^n$ has no positive integer solutions when $n > 2$, adding that the margin was too small to contain his marvellous proof.

Diophantus would not have recognised an equation like $x^2 + 4x + 3 = 0$. Apart from anything else, he had no symbols for $+$, $=$, or 0. He wrote in words, but he did use some abbreviations, and with hindsight we can see in this the seeds of modern symbolic notation. It would be anachronistic, however, to say that this is true abstraction – these were abbreviations rather than abstract symbols. This next step is known, rather jazzily, as syncopated algebra. The problem for Diophantus is as follows. The words for the number, the unknown, the square, and the cube, in Greek, are "arithmos", "monadon", "dynamis", and "kubos". Naturally, then, you might wish to abbreviate them as α, μ, δ and κ. But unfortunately, these letters already represent the numbers 1, 40, 4, and 20.

To get round this, how about adding the second letter as well, to get αϱ, μο, δύ, and κύ? Unfortunately, these also represent numbers. For example, μο means 4,070. The solution was to use the abbreviations μ°, δ^υ, κ^υ for the unknown, the square, and the cube, and a symbol a bit like a sigma, which may be a corruption of α^{ρ}, for "the number". Diophantus does not have an equals sign or a plus sign, but he does use a symbol, roughly this: ⋔ for minus. What makes this very hard to read is that all the abbreviations, including those for numbers and operators, were cut from the same cloth: the letters of the alphabet. It's really hard to pull out, say x^2 and x^3, which would be represented by δ^υ and κ^υ, and recognise those as fundamentally dealing with the same unknown x.

It would be well over a millennium before anything approaching what we might recognise as symbolic algebra came into being. To modern eyes it is amazing how advanced the rhetorical algebra managed to get. Until the sixteenth century, negative numbers and zero were at best treated with suspicion. A problem that resulted in solving what we would now write as an equation like $x^2 + 7x = 15$ was viewed as a fundamentally different kind of problem from one resulting in $x^2 = 7x + 15$, because the statements of such problems could only use positive coefficients. The square added to some multiple of the root equalling the number, is different from the square equalling some multiple of the root added to the number. Something like the formula for solving all quadratic equations in one go could not have been expressed in the notation of the time, even though any individual quadratic equation could be solved (if it had any positive real number solutions, of course). Even by 1545, when the Italian mathematician Girolamo Cardano gave rules for solving cubic equations in his *Ars Magna*, he steered clear of zero, and of negative coefficients. This meant that there was a long list of "different" cubic expressions instead of writing, as we now do, that every cubic equation is equivalent to one of the form $ax^3 + bx^2 + cx + d = 0$. The mathematics in *Ars Magna* is mostly still in words, with a few symbols like ℞ for square root, an abbreviation of "radix". The arguments used were geometrical, which made it almost impossible to give justifications of any results about polynomials higher than cubics. In fact, Cardano wrote that "it would be foolish to go beyond this point", because "nature does not permit it". In order to solve cubics, there is no choice but to involve things like the square roots of negative numbers, and you have to do this even in the case where the cubic equation has three real roots (the imaginary parts cancel out, but they are a necessary part of the calculation). That Cardano managed to do this, even at arm's length, is truly remarkable. All this goes to show that algebra was being done before the symbols were invented. The introduction of symbols, though, made the things that were already known much more understandable and routine. This frees up the mind to make the next steps, much in the same way that we don't really get to grips with reading until the short words like "and" and "the" are second nature, and we can't get good at long division until we have instantaneous recall of times tables. Another benefit of symbols is that they make new ideas and connections possible.

Over the hundred years or so following Cardano, great strides in notation were made, though it still took a very long time for the notation to be standardised. Plus and minus, for example were often abbreviated to *p* and *m*. But in Germany in 1544, Michael Stifel used symbols $+$ and $-$, that he had probably seen used in warehouses to indicate over- and under-weight cargoes, in his treatise *Arithmetica Integra*. We can see a transition in different editions of books like Christoff Rudolff's *Coss* (this is a corruption of the Italian "cosa" or "thing" used to refer to the unknown in an equation, and meant that in Germany and sometimes in England, algebra was sometimes called the "Cossic Art"). An expression that we would nowadays write $x^2 + 5x - 6$ would have been written as $1Zp.\ 5Rm.\ 6$. That is, 1 lot of the Zensus ("number") plus 5 lots of R (the root of Z) minus 6. This notation R implicitly restricts you to a consideration only of positive square roots. Within a few years this would become $1Q + 5N - 6$, (now N is the number and Q is its square, and then finally $1AA + 5A - 6$. This last expression now finally makes explicit the relationship between A and AA. It also leaves the way clear for thinking about AAA and $AAAA$, and even higher powers, as well as at least not ruling out the possibility that A could be negative.

Once we start to write expressions like AAA, it's only a matter of time before someone thinks to shorten this to something like $A\underbrace{}_{3}$. In 1579, Rafael Bombelli did something like this. An expression

4

we would now write as $3x^4$, he would have written 3˘ . He then gave a lot of individual rules 4 7 11

$^4x^7 = x^{11}$) which we would nowadays encapsulate with the

like uia fa , (that is, x˘ ˘ ˘ single rule $x^a x^b = x^{a+b}$. This was an important step, but by neglecting any notation for the variable, which is implicit in the "⌣", it is rather difficult to manipulate it sensibly. Matters were finally resolved by Descartes in La Géométrie (1637), a ground-breaking work that made marvellously visible the links between algebra and geometry. It is surprisingly readable. It's written in French, not Latin, and it uses our modern exponents for the first time, as well as other nowstandard conventions, like using letters at the end of the alphabet, like x, y and z, for variables, and letters at the beginning of the alphabet, like a, b and c, for constants. (Earlier, François Viète had used vowels for variables, and consonants for constants, but this didn't catch on.)

The brilliance of the exponential notation is that not only does it make the mathematics clearer, but it opens up new avenues for the mind to play with. For example, once we write x^2 and x^3, the dimensional barrier we trap ourselves with by having words for "square" and "cube" is removed. It's much more natural to then think easily of x^4, x^5, and even x^n. We then perhaps find that there are general rules like $x^a x^b = x^{a+b}$. The fact that exponents turn multiplication into addition is what underlies the mathematics behind logarithms (though that connection would not be made explicit for a long time). Once you have x^n, you might ask what happens if n is not a positive integer? It's a short step from noticing that $(\sqrt{x})^2 = x$ to defining $x^{1/2} = \sqrt{x}$, and from

(1) × $x^2 = x$, we can get $x^{-1} = $ 1, and $x^0 = 1$. The fact that this all works so nicely even though

x x

it goes far beyond our original conception is one of the wonderful things about mathematics. As Ernst Mach said in 1895, "The student of mathematics often finds it hard to throw off the uncomfortable feeling that his science, in the person of his pencil, surpasses him in intelligence".

Descartes was not always entirely scrupulous about crediting others for their influence on his work. He claimed he knew nothing of Viète's writing, for instance, but this has been disputed. Others before him had used versions of coordinate systems, and even as far back as Menaechmus in the fourth century BC, mathematicians had made links between geometrical curves and roots of polynomials – Menaechmus showed that the cube root of two is the intersection of a parabola and a hyperbola. One example of this, in modern notation, is that $\sqrt[3]{2}$ is the x-coordinate of the

– $= \frac{1}{}$

1 2 y. Descartes' work is of

course much intersection of the parabola $y = x$ and the hyperbola

2 x

more sophisticated. It has to be said that he doesn't give the reader as much help as he might, with statements like "it already wearies me to write so much", and leaving lots of things for us to do, apparently to give us "the pleasure of discovering" it for ourselves, and definitely not because he was too lazy to fill in the gaps, honest.

One of Descartes's aims was to start thinking about geometry more algebraically; the Cartesian coordinate system was not just about being able to say where you are, but about being able to describe curves in terms of equations, as dynamic relationships between variables, rather than static entities existing in space. This was the huge advance, not coordinates per se; without it, calculus could not have developed. Of course, to do these things, you need to be able to write expressions like $y = x^2$ and $x^2 + y^2 = 1$, using more than one variable, something that Bombelli's notation would have made impossible.

We won't discuss calculus, except to note that Newton and Leibniz came up with separate notation, and Leibniz's is clearly superior. Just like the exponent notation, it adds to our understanding, has greater flexibility, and makes the link between differentiation and integration more apparent. Newton used $\dot{x}$, $\ddot{x}$, and so on, where the number of dots above the x tells you how many times to differentiate (with respect to t), whereas Leibniz used $^{d}{}_{d}{}^{y}{}_{x}$, $^{d}{}_{d}{}^{2}{}_{x}{}^{y}{}_{2}$, and so on. This notation is more flexible because you can differentiate with respect to anything you like, not just

t, and you can differentiate with respect to different things in the same equation (those familiar with the chain rule of differentiation will appreciate that this is extremely useful, and also that the

Leibniz notation makes it intuitively very natural). You can also write things like $\frac{d^8y}{dx^8}$ much more clearly than $\overset{......}{x}$, and there is scope for general expressions like $\frac{d^ny}{dx^n}$.

A Notational Miscellany

Before we wrap up this part of the talk, here is a brief guide to the origins of some more of our most commonly used pieces of notation. They are a lovely reminder of the international nature of mathematics. Words and symbols are imported, translated, adapted and adopted from everywhere, so that the language of mathematics, just like any language and society, is enriched by its interactions with many cultures.

We have seen that that the plus and minus signs appeared first in Germany. The English mathematician William Oughtred (inventor of the first slide rule) is credited with the first use, in 1631, of the symbol × for multiplication. Later, Oughtred would use a colon (:) for division; the Arab symbol for fraction used a line to divide two quantities. Our modern symbol ÷ fuses Oughtred's colon and the Arab line symbol. Oughtred is also thought to have been the first to use the abbreviations sin and cos for sine and cosine. The story of the origins of trigonometric words is another fascinating rabbit hole down which we don't have time to journey. But just for a taster, notice that an arc of a circle is like a bow. The corresponding chord is like the bowstring, and the arrow points along the radius. The sine of the angle produced is therefore related to the bowstring, and the cosine to the arrow. This is why the old Sanskrit word for sine, *jīvā*, means bowstring. It was transliterated into Arabic as *jība*, which is meaningless in Arabic so was read as the more familiar word *jayb,* meaning cavity or pocket. This was then translated into Latin as *sinus* (the latin word for cavity), and *sinus* became our *sine.* The whole story of the transmission of knowledge across cultures and times, in one little word!

Back to algebra, and the square root symbol arose in Germany in the early sixteenth century. Michael Stifel was probably the first to put what is recognisably our symbol in print, though it was a modification of earlier symbols dating back at least 50 years. For higher roots, early symbols were very confusing. In Christoff Rudolff's *Coss* of 1525, the cube root was the square root symbol with two extra wiggles thus: , but the fourth root was denoted by just one more wiggle ! It makes a kind of sense, because the fourth root is the square root of the square root, but it makes it very difficult to write down fifth, sixth or higher roots, and does not leave any headspace for general

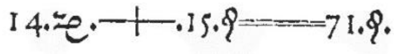
arguments about n^{th} roots. It is a notational dead end. Once people started writing things like $\sqrt[3]{x}$, it's an easier step to $\sqrt[n]{x}$.

We mentioned Robert Recorde earlier as being the inventor of the equals sign. He is an interesting character. Born in around 1510 in Tenby, Wales, he qualified first as a doctor (by virtue of achieving at least the minimum requirement of having "carried out two dissections and effected three cures"). But he was a true polymath. He could translate Latin, Greek, several European languages like German and Italian, and even a little Arabic. He was employed on the service of the Crown (for Edward VI and later Mary Tudor) in roles including comptroller of the Bristol Mint, as well as overseeing silver mines and iron foundries. He wrote several expository books on a wide range of topics, from *The Castle of Knowledge,* on astronomy, to *The Urinal of Physick*, on the diagnosis of ailments by studying urine. Nowadays we remember him most for his mathematics books. His *Ground of Artes* (1543) was the first home-grown (i.e., not just a translation) book on arithmetic in English. *The Pathway to Knowledg* (1551) was the first book expounding Euclid in English. *The Whetstone of Witte* (1557) deals with algebra. Recorde is not remembered because of mathematical theorems or discoveries, but because of his brilliance as a mathematical educator, which led to him introducing many mathematical terms into the English language (by importing them from writers in Germany, Italy and elsewhere, and also by inventing them). He made mathematical skills available to a wide group, by writing in English, and by helping the readers with clear language examples and explanations. He believed strongly in the importance of educating people in mathematics and science, especially lawmakers and politicians. Many of us would agree with these sentiments, from *The Ground of Artes:*

Oh in how miserable case is that realme where the ministers and interpreters of the lawes, are destitute of all good sciences, which [a]re the keyes of the lawes? How can they either make good lawes, or mayntayne them, that lack the true knowledg, whereby to iudge them?"

In reading a wide range of sources for his mathematics books, he had to make constant decisions with words that had no English equivalent, whether to adopt the foreign word (and if so which one) or to create a new English word. Many of his choices have not endured, but some have, including the words bimedial, binomial, commensurable, residual, and universal. Less successful terms include "absurde numbers", for negative numbers. He calls parallel lines "gemowe" lines (or "twin"). Isosceles triangles are "twilike", scalene are "nouelike", pentagons are "cinkangles". For powers, he introduces some astonishing words – from *Zensus*, from census or number as we saw earlier, he says about fourth powers, that these are commonly called squares of squares "and of some men they are named *Zenzizenzikes*, as square numbers are called *Zenzikes"*. And so, "by like reason, doe I cal the nexte numbers square cubes of cubes, or square cubike cubes: which other men doe cal zenzizenzizenzikes, that is, squares of squared squares." This gets him a mention in the Oxford English Dictionary for the word containing more Z's (six) than any other in the English language! But actually the historian Ulrich Reich has found that Recorde also used zenzizenzizenzizenzike for 16th power, which is eight Zs! We also see the zenzizenzike roote (4th root), as well as cubicubike roote (9th) and zenzizenzicubike roote (12th).

The *Whetstone of Witte* includes a section on the "Arte of Cossike numbers", or algebra. We mentioned earlier that Cossick comes from Cosa, the Italian word used for the unknown. But *cos* is also the Latin for Whetstone, so the title of the book is in fact a pun – the book is a tool for sharpening your wits, but via algebra. It's a "cossic cos". Recorde's notation uses different symbols for units, unknowns, squares of the unknown, cubes of the unknown etc, and it is here that he introduces the equals sign and give some examples of its use. We can therefore definitively state the first ever printed equation in history. Here it is:



In modern notation we would write this as $14x + 15 = 71$. If you solve it, you can claim to have solved the oldest equation in the world!

Tragically, Recorde's honesty was his downfall. He accused the Earl of Pembroke of corruption (a charge that was manifestly accurate), but the Earl, who had very powerful friends, in turn accused Recorde of slander. The Earl prevailed, Recorde was fined £1,000, which of course he could not pay, was therefore thrown into debtors' prison and died there in 1558.

Let us mention here another Welshman's contribution to mathematical posterity. The Welsh mathematician William Jones was the first, in 1706, to use π for the ratio of the circumference of a circle to its diameter. The letter was chosen because it is the first letter of the word periphery. This wasn't the first time a letter was used to signify this ratio. That honour probably goes to one J. Christoph Sturm, professor at the University of Altdorf in Bavaria, who in 1689 used the letter *e* in a statement, "si diameter alicuius circuli ponatur *a*, circumferentiam appellari posse *ea* (quaecumque enim inter eas fuerit ratio, illius nomen potest designari littera *e*)". Speaking of mathematical constants known as e, the number we now know by that name was first studied by Christiaan Huygens in the context of a question about per annum interest rates being compounded at ever shorter intervals. The letter e was introduced by the mathematician Leonhard Euler, in a manuscript of 1727 or 1728, likely because that happened to be the first letter he'd not yet used in the paper he was writing, rather than, as some people have rather unfairly suggested, because he was trying to name it after himself. Leibniz had earlier used the letter b, but this did not catch on. Euler was also the first to use i for the square root of minus 1, where here i stands for "imaginary".

Symbols for sets of numbers were not needed until much more recently, but it's interesting just how recently consensus was reached on which symbols to use. Mathematicians who were taught at their mother's knee (fairly literally in my case as my mother was my A level maths teacher) that the set of integers (whole numbers) is $\mathbb{Z}$, the rational numbers (fractions) are $\mathbb{Q}$ and the real numbers are $\mathbb{R}$ may be surprised to learn that these were not fixed until at least the 1940s. The set of integers was called n by Peano in 1895, Γ by Helmut Hasse in 1926, G0 by Otto Haupt in 1929,

C by van der Waerden in 1930 and $\mathfrak{J}$ by Edmund Landau, also in 1930. The notation $\mathbb{Z}$ seems to be due to a group of French mathematicians writing under the pseudonym Nicolas Bourbaki, who wrote a series of books, Éléments de Mathématique, from the late 1930s onwards. The letter $\mathbb{Z}$, from the German Zahlen (number) was chosen by Bourbaki in around 1940 for the integers, with $\mathbb{Q}$, for Quotient, for the rational numbers. Around this time, $\mathbb{R}$ for the real numbers and $\mathbb{C}$ for the complex numbers, were also adopted.

There are many other symbols that we don't have time to discuss, but to end this section, just think about the equation $e^{i\pi} + 1 = 0$, which many mathematicians think is the most beautiful equation in mathematics. If things had gone differently, we might be writing it as $b^{ie} + 1 = 0$.

The letters e and i are due to Euler, who was Swiss. But π comes from William Jones, + comes from Michael Stifel, = is from Robert Recorde, and 0 and 1 are from India. So actually, this equation is, notationally at least, 28% Swiss, 28% Welsh, 28% Indian, and 14% German. The wonderful thing is, it belongs 100% to mathematics and always will.

The Mathematics of Language

Since this is my final Gresham lecture of the year, and therefore it's basically the last day of term, I thought we could spend the last few minutes playing with a couple of ideas related to language that can generate, in the first case, graphs, and the second, sequences, that we can have a bit of fun with. Of course, I also have to give you some homework to keep you going until the Autumn Term.

The incomprehensibility graph

When English speakers find something difficult to understand, we sometimes say "it's Greek to me" (like so much else, this phrase is first recorded in written English in Shakespeare). The obvious question is, what language do Greek speakers find impossible? A tongue-in-cheek paper *The Hardest Natural Languages* was written on this topic by Arnold Rosenberg in 1979, in which he attempts to construct an ordering of difficulty of languages based on idioms like this. In French, we can refer to Molière, for instance, who in *L'Étourdi* (1653) uses the expression "C'est de l'hébreu pour moi". He constructs graphs to show the relationships. There are caveats – some languages have more than one "difficult" comparator. Even in English, we talk about "double Dutch". But the Dutch find Latin difficult (Dat is Latijns voor mij), and in Latin we have *Graecum est; non potest legi* (It's Greek; it can't be read). We end up back at Greek, which has a phrase Μου φαίνεται κινέζικο (mou fainetai Kinesiko – it appears Chinese to me). Chinese is also hard in Hebrew. Is Chinese the hardest language? In Chinese, an incomprehensible thing is "Heavenly script to me". So, we can say that one pinnacle is Heavenly Script. Other parts of the graph have other endpoints, and there is even a cycle, from Turkish to Arabic to Persian and back to Turkish again! An updated version of this idea, called the "Directed graph of stereotypical incomprehensibility" is explored at the Language Log (2).

Numerolinguistics

My second bit of fun is a number game that you can play in your favourite language. Pick a number, any number. (Before you say π, e, or $9\frac{3}{4}$, it has to be one of the counting numbers 1, 2, 3, 4, ….) Say we choose 84. We're going to make a sequence. Starting with 84, the next number is

- 84 eighty-four

- 11 eleven
- 6 six
- 3 three
- 5 five
- 4 four

So, our sequence is 84, 11, 6, 3, 5, 4, 4, 4, … Four is clearly a fixed point of this iterative process.

Are there any others? Pick your favourite number and try it. We'll do one more: $77 \to 13 \to 8 \to 5 \to 4$. Again, we end up at 4. In fact, you always end up at four, wherever your starting point! I like this about English because my birthday is 4/4, and it makes up for the fact that 4 is said to be unlucky in Chinese. Speaking of Chinese: in Mandarin there is also one fixed point to which everything leads, but this time it is 1.

What about other languages? It seems as if perhaps there might be one or more fixed points, and every sequence will terminate in one of them. For example, you can check that in Italian, all roads lead to three (tre). In Danish, we always end up at 2 (to), 3 (tre) or 4 (fire). Let's try this in French. For a number n we'll write L(n) for the number of letters.

n	$L(n)$	n	$L(n)$	n	$L(n)$
		10 dix	3	20 Vingt	5
1 Un	2	11 onze	4	21 vingt et un	9
2 Deux	4	12 douze	5	22 vingt-deux	10
3 Trois	5	13 treize	6	23 vingt-trois	11
4 Quatre	6	14 quatorze	8	24 vingt-quatre	12
5 Cinq	4	15 quinze	6	25 vingt-cinq	10
6 Six	3	16 seize	5	26 vingt-six	9
7 Sept	4	17 dix-sept	8	27 vingt-sept	10
8 Huit	4	18 dix-huit	8	28 vingt-huit	10
9 Neuf	4	19 dix-neuf	8	29 vingt-neuf	10

For example, $1 \to 2 \to 4 \to 6 \to 3 \to 5 \to 4$. So, there is a cycle! Moreover, there are no fixed points in French. There are plenty of other languages with cycles. Hungarian has 2 (öt) ↔ 5(kettő), for instance, as well as the fixed point 4(négy). Finnish has one fixed point, 5 (viisi), and one cycle 8 (kahdeksan) ↔ 9 (yhdeksän).

How high do we have to count in a language to be sure we have found all the cycles and fixed points? The structure of most languages means that, roughly speaking anyway, the words settle down to (number of tens) then (number of units) – and we can add hundreds, thousands and so on to that. In French, there is a flirtation with base 20 counting between 60 and 99, where instead of, say "95" being "word for nine tens" + "word for 5", it's actually "quatre-vingt-quinze" which literally means four twenties plus fifteen. But even then, it's pretty clear that beyond a fairly low point, we are always on a downward trajectory. We can write $L(n)$ for the number of letters in the word for n. Thus, in French, we have $L(4) = 6$. I suppose we ought to say $L_{\text{French}}(4) = 6$ and $L_{\text{English}}(4) = 4$. Since the fixed points are the values of n for which $L(n) = n$, and every cycle must contain at least one n for which $L(n) > n$ (otherwise we would keep decreasing until we reach a fixed point), we can find every fixed point and every cycle by considering just those numbers n with $L(n) \leq n$. Beyond a certain threshold point T in each language, we will have

$L(n) < n$ for every n. In French, this threshold is 4, so we can write $T_{\text{French}} = 4$. In general, we can ask what is a safe assumption to make about the threshold? How high do we have to go to capture all the fixed points and cycles? There are lots of language websites that teach you to count to ten in many languages. Is this safe?

Glad you asked. Let's look at Russian.

Counting in Russian is again based on the decimal system, and because the words for the tens, and the words for the units, are all relatively short (each at most 9), once we get to 20 and above, it's easy to verify that L(n) is always strictly less than n. Here are the numbers 1 to 19.

n		$L(n)$	n		$L(n)$
			10	Девять	6
1	Один	4	11	Одиннадцать	11
2	Две	3	12	Двенадцать	10
3	Три	3	13	Тринадцать	10
4	Четыре	6	14	четырнадцать	12
5	Пять	4	15	Пятнадцать	10
6	Шесть	5	16	Шестнадцать	11
7	Семь	4	17	Семнадцать	10
8	Восемь	6	18	восемнадцать	12
9	Девять	6	19	девятнадцать	12

Here, we can see that $T_{\text{Russian}} = 11$. So, it's not enough to look at only the numbers 1 to 10! However, to understand the numerolinguistics of Russian, it does suffice to test numbers only up to 11. For example, starting with 1 gives $1 \to 4 \to 6 \to 5 \to 4$, revealing a cycle. There are two fixed points, 3 and 11, and one cycle (4,6,5) of length 3.

Let's have some language prizes. Your homework is to try this out in your favourite languages, and to see if you can beat any of these records. In particular, I would love to know if there is a language with two or more cycles, or a language with a longer cycle than the (3, 5, 4, 6) of French. There's surprisingly little online about this – there's one website I found that analyses 22 languages, although it is wrong about Russian, I think because the author doesn't realise ы is a single character in Cyrillic script. I've researched a few others myself, including Latin, Greek, Igbo, and Zulu.

- *Language with the highest threshold*. The current leader is Zulu, with $T_{\text{Zulu}} = 29$. The number 29 in Zulu is amashumi amabili nesishiyagalolunye; it has 33 letters. The sequence it's in goes $29 \to 33 \to 25 \to 23 \to 23$ (a fixed point). Zulu is also current favourite for the *language with the highest fixed point* (27), and the *language with the most fixed points* (5). The fixed points are 4 (kune), 11 (ishumi nanye), 16 (ishumi nesithupha), 22 (amashumi amabili nambili) and 27 (amashumi amabili nesikhombisa).
- *Language with the longest cycle*. The current leader is French, with a cycle (3,5,4,6) of length 4.
- *Language with the most cycles*. Many languages have exactly one cycle (French, Russian, Latin and many more). All languages I've studied at most one cycle (English has none).

Can you do better? Let me know!

Further Reading

1. Sagan, Carl. *Murmurs of Earth: The Voyager interstellar record.* New York : Random House, 1978.
2. Liberman, Mark. The Directed Graph of Stereotypical Incomprehensibility. *The Language Log.* [Online] https://languagelog.ldc.upenn.edu/nll/?p=1024.
3. Kampf, Marijn. Everything is 4. [Online] [Cited: 21 March 2021.] http://www.marijn.org/everything-is-4/.
4. Of Languages and Numbers. [Online] [Cited: 21 March 2021.] https://www.languagesandnumbers.com.

5. Cajori, Florian. *A History of Mathematical Notations (Two Volumes Bound as One).* New York : Dover Publications, 1993.

6. Mazur, Joseph. *Enlightening Symbols: A Short History of Mathematical Notation and Its Hidden Powers.* Princeton, New Jersey : Princeton University Press, 2014.

7. Menninger, Karl. *Number Words and Number Symbols: A Cultural History of Numbers.* New York : Dover Publications, 1992.

8. Miller, Jeff. Earliest Uses of Various Mathematical Symbols. [Online] https://jeff560.tripod.com/mathsym.html.

9. Plofker, Kim. *Mathematics in India.* Princeton, New Jersey : Princeton University Press, 2009.

10. Yan Li, Du Shiran, J. N. Crossley (Translator), A. Lun (Translator). *Chinese Mathematics: A concise history.* Oxford : Clarendon Press, 1987.

11. *The Hardest Natural Languages.* Rosenberg, Arnold. 1979, Lingvisticae Investigationes, Vol. III: 2, pp. 323-399.

12. Roberts, Gareth and Smith, Fenny (Eds). *Robert Recorde: The Life and Times of a Tudor Mathematician.* Cardiff : University of Wales Press, 2012.