Who of us would not be glad to lift the veil behind which the future lies hidden: to cast a glance at the next advances of our science and at the secrets of its development during future centuries? What particular goals will there be toward which the leading mathematical spirits of coming generations will strive? What new methods and new facts in the wide and rich field of mathematical thought will the new centuries disclose?

These words were spoken by David Hilbert, the greatest mathematician of his day, at possibly the most famous mathematical lecture of all time - a plenary presentation at the Second International Congress of Mathematicians, held in Paris in August 1900. At this lecture Hilbert posed 23 unsolved problems, designed to set the mathematical agenda for the next hundred years. One of Hilbert's problems was the famous Riemann hypothesis, which remains unsolved. We'll meet this problem, together with some of Hilbert's other problems, shortly.

One hundred years later, also in Paris, the Clay Mathematics Institute, set up in the US in 1998 with the mission 'to increase and disseminate mathematical knowledge', hosted a meeting called A celebration of the universality of mathematical thought. The central theme of this meeting was the announcement of seven so-called 'millennium problems', with a prize of one million dollars for the solution to each - seven million dollars in all - in order to celebrate mathematics in the new millennium.

Later in this lecture we’ll look at some of these. But first let's return to the Hilbert problems and see what happened to them.

The Hilbert problems

Here's a list of Hilbert's 23 problems from the write-up of his lecture, although only ten of them were included in the lecture itself. The ones that appear in italic are the ones I'll talk about.

1. The Continuum Hypothesis: there is no set whose size is strictly between that of the integers and that of the real numbers
2. Prove that the axioms of logic are consistent
3. Given two polyhedra of equal volume, can we always cut the first into finitely many pieces that can be reassembled to yield the second
4. Construct all metrics whose lines are geodesics
5. Are continuous groups automatically differential groups?
6. Axiomatise all of physics
7. Is always transcendental for algebraic a ≠ 0, 1 and irrational algebraic b?
8. The Riemann hypothesis: prove that the real part of any non-trivial zero of the Riemann zeta function is ½
9. Find the most general law of reciprocity theorem in any algebraic number field
10. Find an algorithm to determine whether a given polynomial Diophantine equation with integer coefficients has an integer solution
11. Solve quadratic fields with algebraic numerical coefficients
12. Extend the Kronecker-Weber theorem on abelian extensions of the rational numbers to any base number field
13. Solve all 7th-degree equations using functions of two variables
14. Prove the finiteness of certain systems of functions
15. Provide a rigorous foundation for Schubert's enumerative calculus
16. Study the topology of algebraic curves and surfaces
17. Express a definite rational function as a quotient of sums of squares
18. Is there a non-regular space-filling polyhedron? What is the densest sphere packing?
19. Are the solutions of Lagrangians always analytic?
20. Do all variational problems with certain boundary conditions have solutions?
21. Prove the existence of linear differential equations with a prescribed monodromy group
Problem 1: The continuum hypothesis

This asks us to prove that there is no set whose size is strictly between that of the integers and that of the real numbers (or continuum). What does this mean?

A famous result of Georg Cantor in the 19th century was the truly revolutionary idea that some infinities are larger than others. This idea had its origins in the musings of Galileo over 300 years earlier, who had noticed that there are far fewer perfect squares \(1, 4, 9, 16, \ldots\) than natural numbers \(1, 2, 3, \ldots\), and yet we can match them up exactly: \(1 \leftrightarrow 1, \ 2 \leftrightarrow 4, \ 3 \leftrightarrow 9, \ 4 \leftrightarrow 16, \ldots\)

We can also match the natural numbers with some larger sets, such as the integers. We simply list them in the order: \(0, 1, -1, 2, -2, 3, -3, 4, \ldots\) Notice that every integer occurs somewhere in this list - none is omitted.

Cantor's first amazing discovery was that we can similarly list all the fractions in order: so the set of all fractions has the same size as the set of all natural numbers, even though it seems so much bigger. However, as Cantor proved, the set of all real numbers cannot be made to match up with the natural numbers, and so must be a 'bigger' infinite set. It follows that some infinite sets are genuinely bigger than others: not all infinities are the same. Indeed, to take his idea further, we find that there are infinitely many infinities, all of different sizes.

We can construct an 'arithmetic of infinities'. The size of the set of all integers (or of all fractions) is called \(\aleph_0\) (aleph is a Hebrew letter), and we can write 'infinite equations' such as \(\aleph_0 + 5 = \aleph_0\), and \(\aleph_0 + \aleph_0 = \aleph_0\). The size of the set of all real numbers is called \(\aleph_1\) (which is larger than \(\aleph_0\)). The question arises: is there an infinite set whose size lies strictly between these two infinite numbers? This problem became known as Cantor's continuum problem.

The answer was unexpected. In 1902 Bertrand Russell upset the set theory apple-cart by producing a famous logical paradox that seemed to have no answer. One version of this paradox is that of the barber in a village who shaves everyone who doesn't shave themselves, but who obviously doesn't shave those who do shave themselves; the question is: who shaves the barber? (Whatever answer you give must be false, as you'll see after a moment's thought.)

The paradox was eventually overcome by Zermelo and Fraenkel who produced a system of logic that could deal with it - and this logical system became universally accepted and used. Then, in the 1930s, Kurt Gödel produced another bombshell: in any axiomatic system that one can construct, there are true results that cannot be proved, and there are statements that are undecidable - they cannot be proved true, and they cannot be proved false.

In 1963 Paul Cohen stunned the mathematical world with his contribution to Hilbert's first problem: he proved that Cantor's continuum problem cannot be decided within the Zermelo-Fraenkel logical system - it cannot be proved either true or false - it's undecidable.

Problem 2: Consistency

Problem 2, also on logic, asks us about whether the axioms of arithmetic are consistent - or whether, somewhere out there, we'll find a contradiction that we never expected. Is the theory of the integers, or of the real numbers, consistent? Can we be sure that contradictions will never occur?

Hilbert assumed that the answer is yes - but here Gödel's bombshell was produced by Gödel in 1931, who proved that the consistency of any theory containing that of the integers cannot be proved within the theory itself. Worse, no such consistent theory can be complete in the sense that every mathematical truth that we can express in its language can be proved within the theory - Gödel's so-called completeness theorem. One would have thought that this would have finished off mathematics, but mathematicians gritted their teeth and carried on, more or less regardless. If you want to read more about Gödel's theorem, a couple of novels feature it: Martínez's The Oxford murders and Doxiadís's Uncle Petros and the Goldbach conjecture - we'll see
the Goldbach conjecture again later.

Our next couple of Hilbert problems are geometrical.

**Problem 3** tries to generalise an old problem of the Hungarian mathematician János Bolyai, who proved in 1833 that if we have any two polygons with the same area, then they can be cut up into a finite number of equivalent triangles - in other words, either of the polygons can be cut up into triangles that can be reassembled to give the other, and in particular, every polygon can be 'squared' - cut up into triangles that can be reassembled to give a square of the same area?

Hilbert asked whether this can be done in three dimensions: can any two polyhedra with the same volume be cut up into a finite number of equivalent tetrahedral? - in other words, can either of the polyhedra be cut up into tetrahedral bits that can be reassembled to give the other? and in particular, can every polyhedron be 'cubed'?

This was the first Hilbert problem to be solved - by his brilliant student Max Dehn in 1901/2. The answer is no - and he proved this by presenting a regular tetrahedron that cannot be cut up into bits that can be reassembled to make a cube.

The other geometrical problem was **Problem 18**, which comes in three parts. The first part contains geometrical symmetries. It has been known for some time that different letters of the alphabet have different symmetries: for example, Fs have no symmetries, whereas Es have horizontal symmetry, As have vertical symmetry, and Ns have rotational symmetry and Hs have all of these. In total, there are exactly seven types of symmetry exhibited by a row of similar letters - the so-called frieze patterns.

If we now move up to two dimensions, we obtain the so-called wallpaper patterns that are exhibited in the Alhambra in Grenada. How many of these are there? We find that there are exactly 17 types of symmetry.

What happens in higher dimensions? In 1890, Fedorov proved that in three dimensions there are 230 patterns, and Hilbert asked whether there is always a finite number in any dimension. This was answered in the affirmative by Ludwig Bieberbach in 1910, but it was not until 1970 that the number for four dimensions was found - it's 4783. But no general formula for \( n \) dimensions - or even any further results - are known.

The second part of Problem 18 concerned tilings of the plane. We know that the plane can be tiled exactly with equilateral triangles, or with squares, or with regular hexagons, but no other type of regular polygon will work.

We can now head in two directions. We can either ask for tilings with more than one type of tile - these are the so-called semi-regular, or Archimedean, tilings - or we can stick to a single type of tile which must then be irregular. Hilbert asked essentially whether one could tile the plane with just one type of non-symmetrical tile, and the answer was yes, as illustrated in 1935 by Heinrich Heesch, who came to play an important role later on in the proof of the four-colour theorem.

The third question that Hilbert asked under the heading of Problem 18 concerned an ancient three-dimensional problem of Harriot, Kepler and others: what's the most efficient way of stacking cannon-balls, or oranges, so that the amount of empty space between them is as small as possible?

One way to stack them is the so-called cubic stacking, and another way is hexagonal stacking, but neither of these is the most efficient. It turns out that the way your greengrocer stacks oranges is the most efficient - the proportion of empty space is about 0.36, which is less than the 0.48 and 0.40 of the other two. But to prove this rigorously was horrendous. A computer-aided proof was eventually given by Thomas Hales in 1998 and involved 3 gigabytes of computer power.

**Problem 7: Transcendental numbers**

Before we come on to the greatest problem of them all, let's look briefly at transcendental numbers. What are these? Just as we can split numbers into rational ones (the fractions) and irrational ones, so we can further split the irrational ones into those that are algebraic and those that are transcendental.

The algebraic ones are those that are solutions to equations with whole numbers as coefficients. For example, \( \sqrt{2} \) and \( 3\sqrt{7} \) are algebraic since they are the solutions of the equations \( x^2 = 2 \) and \( x^3 = 7 \). On the other hand, it was proved in the 19th century,
after a lot of trouble, that $e$ and $\pi$ are transcendental. But classifying numbers in this way is extremely difficult, and in Problem 7 Hilbert asked whether $2^{\sqrt{2}}$ is transcendental (a result eventually proved by Siegel in 1930) - and more generally, whether $a^b$ is transcendental if $a$ is an algebraic number different from 0 or 1 and $b$ is an irrational algebraic number. Even now, we know very little: we now know that $e^\pi$ is transcendental, but we don't know about $e + \pi$, or near $\pi^e$.

We turn now to number theory and two very important unsolved problems - the Goldbach conjecture and the Riemann hypothesis. What are these?

**Problem 8: The Riemann hypothesis**

The Riemann hypothesis is named after the German mathematician Bernhard Riemann, who died at the age of 40 while a professor at Göttingen University where he followed in the footsteps of Gauss and Dirichlet. On being elected a member of the Berlin Academy in 1859, Riemann was required to contribute to the Academy's Proceedings; the result was his only paper in number theory: 'On the number of primes less than a given magnitude', and it is now regarded as one of the classics of mathematics.

So it's about prime numbers. In very general terms it asks: *do all the solutions of a certain equation have a particular form?* The answer to this very hard problem is simply 'yes' or 'no' - but we don't know which. The full statement is: *do all the non-trivial zeros of the zeta function have real part $\frac{1}{2}$?* But what does that mean, and what's its connection with prime numbers?

So let's start with prime numbers. A *prime number* is a number, larger than 1, whose only factors are itself and 1: so 13, 17 and 19 are prime numbers, while 15 (= $3 \times 5$) and 18 (= $2 \times 3 \times 3$) are not. The primes up to 100 are: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.

Prime numbers are central to mathematics because they're the building blocks for numbers - every whole number can be built up from them. So, for example, $60 = 2 \times 2 \times 3 \times 5$ and $100 = 2 \times 2 \times 5 \times 5$, and more generally *every whole number splits into primes in only one way*, apart from the order in which we write down the factors.

Primes are things that were designed to be multiplied, and we get into real trouble if we try to add or subtract them: in fact, some of the most intriguing problems in mathematics arise in this way. For example, if we add two primes other than 2 we always get an even number: $6 = 3 + 3$, $10 = 7 + 3$, $20 = 13 + 7$, and so on. But can we get all even numbers in this way? The belief that we can is called the *Goldbach conjecture*, but no-one knows for sure - Hilbert mentioned it when discussing his Problem 8. It's been checked for all even numbers up to 4 trillion - but this isn't good enough for mathematicians who need *absolute* proof, true in all cases. The best knowledge we have is a celebrated result of Chen Jin-run who proved in 1966 that every large enough even number is the sum of a prime and another number with at most two prime factors.

So twin primes seem to occur forever, and yet we can find arbitrarily large gaps between primes. For example, from 90 to 96 is a string of seven non-primes, and if we want a string of 99 non-primes, we consider the number 100!, which is $100 \times 99 \times \ldots 2 \times 1$, and add 2 (giving a number divisible by 2), 3 (that's divisible by 3), and so on, up to $100! + 100$ (which is divisible by 100).

This leads us to one of the main questions about primes - *how are they distributed?* They don't seem to be well distributed, as we can see if we look at the numbers just below and just above 10 million. The hundred just below contain nine primes (that's quite a lot), while the hundred just above contain just two.

In his inaugural lecture at the University of Bonn, the number-theorist Don Zagier said: *There are two facts of which I hope to convince you so overwhelmingly that they will permanently be engraved in your hearts.*

The first is that the prime numbers belong to the most arbitrary objects studied by mathematicians: they grow like weeds, seeming to obey no other law than that of chance, and nobody can predict where the next one will sprout.

The second fact is even more astonishing, for it states just the opposite: that the prime numbers exhibit stunning regularity, that there are laws governing their behaviour, and that they obey these laws with almost military precision.
To see what he meant by this, we'll look at the prime-counting function \( \pi(x) \), which counts the number of primes up to any number \( x \) - it's nothing at all to do with the circle number \( \pi \). So \( \pi(10) = 4 \), because there are exactly four primes \( (2, 3, 5, 7) \) up to 10; and \( \pi(20) = 8 \), because there are now a further four \( (11, 13, 17, 19) \). Continuing, we find that \( \pi(100) = 25 \); \( \pi(1000) = 168 \); \( \pi(10,000) = 1229 \); and so on. If we plot the values up to 100 on a graph we get a jagged pattern - each new prime creates a jump. But if we stand further away and plot the primes up to 100,000, we get a lovely smooth curve - the primes do indeed seem to increase very regularly. It turns out, as Gauss conjectured, that \( \pi(x) \) behaves very much like \( x/ \log x \) as \( x \) becomes large. This is a celebrated result in mathematics known as the prime number theorem.

Let's now leave prime numbers for a while and look at some series, where we add lots of terms together. The series \( 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots \) is called the harmonic series, since the Greeks connected it with the harmonic series of music. But it has no sum - if we add enough terms we can make it as large as we wish. To see why, let's bracket it as \( 1 + \frac{1}{2} + \left( 1\frac{1}{3} + \frac{1}{4} \right) + \left( 1\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} \right) + \ldots \). But this is greater than \( 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \ldots \) which increases without limit, so the harmonic series cannot converge to any fixed number. And amazingly, even if we throw away most of the terms and leave only those that correspond to primes \( 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \ldots \), then there's still no sum.

In the 18th century one of the big challenges was to add up all the reciprocals of the squares - namely, to find the sum of the series \( 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \ldots \) This was eventually solved by the Swiss mathematician Leonhard Euler, who found (by somewhat dubious means) that the sum is the number \( \pi^2/6 \), a remarkable result. In the same way, he found that the sum of the reciprocals of the fourth powers is \( \pi^4/90 \), for the sixth powers it's \( \pi^6/945 \), and so on up to the 26th powers. He called the sum of the \( k \)th powers \( \zeta(k) \), naming it the zeta function: so,

\[
\zeta(1) \text{ (the harmonic series) is not defined; } \zeta(2) = \frac{\pi^2}{6}; \quad \zeta(4) = \frac{\pi^4}{90}; \quad \text{and so on.}
\]

It turns out that, \( \zeta(k) \) is defined for each number \( k \) larger than 1.

What have these series and the zeta function to do with prime numbers? Seemingly nothing - but Euler spotted a crucial link. It can be used to give a proof that there are infinitely many primes.

We can write the series for \( \zeta(1) \) as follows:

\[
1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots
= (1+\frac{1}{2}+\frac{1}{4} + \ldots ) \times (1+\frac{1}{3}+\frac{1}{9} + \ldots ) \times (1+\frac{1}{5}+\frac{1}{25} + \ldots ) \times (1+\frac{1}{7}+\frac{1}{49} + \ldots ) \times \ldots,
\]

where each bracket involves the powers of just one prime. Summing the separate series gives \( \frac{2}{2-1} \times \frac{3}{3-1} \times \frac{5}{5-1} \times \ldots = 2 \times 1\frac{1}{2} \times 1\frac{1}{4} \times \ldots \) Since \( \zeta(1) \) is not defined, the right-hand side must go on for ever, and so there must be infinitely many primes.

Euler further extended these ideas to prove that, for any number \( k \), larger than 1: \( \zeta(k) = \frac{2^k}{2^k-1} \times \frac{3^k}{3^k-1} \times \frac{5^k}{5^k-1} \times \ldots \times \frac{7^k}{7^k-1} \times \ldots \).

So Euler spotted an unexpected link between the zeta function, which involves adding reciprocals of powers of numbers and which seems to have nothing to do with primes, and a product that intimately involves the prime numbers. This was a stunning result and a major breakthrough.

We've now set the scene for the Riemann hypothesis. As we've seen, the zeta function \( \zeta(k) \) is defined for any number \( k \) larger than 1. But can we extend the definition to other numbers?

By way of analogy, we can prove that \( 1 + x + x^2 + x^3 + \ldots \) is equal to \( 1/(1-x) \).

But the left-hand side makes sense only when \( x \) lies between -1 and 1, whereas the right-hand side makes sense for any \( x \), apart from 1. So we can extend the formula on the left-hand side to all values of \( x \) (other than 1) by redefining it using the formula on the right-hand side.
In the same way, we can extend the definition of the zeta function to almost all of the complex plane. This consists of all symbols of the form \( a + bi \), where \( a \) and \( b \) are the \( x \)- and \( y \)-coordinates and \( i \) represents the 'square root of -1'. There's nothing particularly mystical about this - all it means is that in calculations, whenever you come across \( i^2 \) you replace it by -1.

Using something called contour integration, we can extend our definition of the zeta function to the whole plane, apart from the single point 1: this makes sense, since we know that \( \zeta(k) \) is not defined when \( k = 1 \). When \( k \) is a number larger than 1, we get the same value as before - for example, \( \zeta(2) = \pi^2/6 \). And for all other values of \( k \), including the complex ones (involving \( i \), the square root of -1), we get a value of the zeta function.

We've seen how Gauss attempted to explain why the primes thin out, by proposing the estimate \( x/\log x \) for the number of primes up to \( x \). Riemann's great achievement was to obtain an exact formula for the number of primes up to \( x \) and his formula involved in a crucial way the so-called zeros of the zeta function - the solutions of the equation \( \zeta(z) = 0 \). (For example, if the equation had been \( z^2 - 4 = 0 \), then the zeros (that is, the solutions) would be 2 and -2.)

It turns out that there are zeros of the zeta function at -2, -4, -6, -8, ... (known as the trivial zeros), and that all the other zeros (the non-trivial ones) lie within a vertical strip between 0 and 1 (the so-called critical strip), beginning at the following points: \( \frac{1}{2} \pm 14.1i \), \( \frac{1}{2} \pm 21.01i \), \( \frac{1}{2} \pm 25.01i \), \( \frac{1}{2} \pm 30.4i \), ... But these all have the form \( \frac{1}{2} + (\text{something times } i) \). So the question arises: do all the zeros in the critical strip have this form? do they all lie on the central vertical line?

That's the big question we now call the Riemann hypothesis. We know that the zeros in the critical strip are symmetrically placed, both horizontally and vertically, and that as we progress vertically up and down the line, the first few zeros do lie on this central line \( \frac{1}{2} \). In fact, the first billion and a half zeros lie on this line. Moreover, it's known that there are infinitely many zeros inside the critical strip, and in fact at least 40% of them lie on the central line. But do they all lie on this line? Or are the first billion and a half just a coincidence?

It's generally believed that all the non-trivial zeros do lie on the central line. Indeed, the appearance of just one of these zeros off the central line would cause major havoc in number theory and throughout mathematics. But no-one can prove it, even after a century and a half.

The Millennium problems

The Riemann hypothesis remains unproved, and it's one of the seven millennium problems I mentioned earlier. Let's now turn to these: they're the problems now considered by the mathematical community to be the most difficult and important in the subject - the 'Himalayas of mathematics', with the Riemann hypothesis as Mount Everest. They're all highly technical and some are extremely difficult to describe, but very roughly they're as follows:

- the **Riemann hypothesis** (on prime numbers)
- the **P = NP? problem** (on the efficiency of algorithms for solving problems)
- the **Poincaré conjecture** (relating to surfaces in four-dimensional space) - this is the only one of the seven that's been solved
- the **Navier-Stokes equation** (solving a partial differential equation arising from fluid motion)
- **Yang-Mills theory** (creating a mathematical theory of an area of quantum physics)
- the **Birch-Swinnerton-Dyer conjecture** (another problem from number theory)
- and finally the **Hodge conjecture** (a highly technical problem from geometry and topology)

These problems are all extremely difficult and definitely not for the amateur. Having already looked at the Riemann hypothesis, let's now consider one of the others.

**P = NP?**

This problem is concerned with algorithms, which are procedures for solving problems. It's known as the 'P versus NP problem', and it's concerned with how hard it is to solve problems. Its solution would be of great interest to both mathematicians and computer scientists, and also of great practical importance in operations research and business economics.
So what's an algorithm? It's basically a *finite step-by-step procedure for solving a problem* - a sort of recipe, like cake-making, where you input sugar, flour and eggs, carry out each step of the recipe, and the output is a perfect cake. For a general algorithm you input some data, carry out each step in turn, and produce an output that's the solution to your problem.

Let's look at some algorithms, starting with the *minimum connector problem* and the *travelling salesman problem*.

In the *minimum connector problem*, we wish to connect a number of cities by links, such as canals, railway lines, or air routes. We need enough links so that we can get from any city to any other (not necessarily in one step), but building canals is costly, so we must keep the total cost to a minimum. How can we do this?

Here's an example with five cities, A, B, C, D, E, where the links have the following lengths, or costs:

- AB: 6
- AC: 4
- AD: 8
- AE: 2
- BC: 5
- BD: 8
- BE: 6
- CD: 9
- CE: 3
- DE: 7

How do we find our 'minimum connector'?

We can experiment: taking BC, CD, DE and EA we link all the cities, with total length 2 + 7 + 9 + 5, which is 23; taking AB, AC, AD, and AE we again link all the cities, but with total length 6 + 4 + 8 + 2, which is 20 - quite an improvement.

But is this the best we can do?

Notice that each of these solutions is a *tree structure* - it contains no cycles: for if there were a cycle, we could decrease the total length by removing any one link from the cycle. So how do we find the right tree structure? There are 125 possible trees joining these cities, so we can't use trial and error. Worse still, the number of trees increases dramatically with extra cities: over 1000 trees for six cities, and 100 million for ten cities. Since most practical examples have hundreds of cities, we need an algorithm to solve this problem.

The algorithm we use is called the *greedy algorithm*, since at each stage we're 'greedy' and choose the shortest possible link available. So here we start with link AE (length 2). Then CE (length 3). We can't now choose AC (length 4), since that gives a cycle, so we use BC (length 5). What next? Each of the links of length 6, AB and BE, give a cycle, so we go to DE (length 7). This completes the tree, so we stop - the cities are now all linked by a tree of total length 2 + 3 + 5 + 7 = 17. This is a great improvement on what we had before, and is the best possible solution.

In fact, the greedy algorithm *always* gives us a minimum connector - however many cities there are. We just model the situation by a network, and program a computer to make the choices. Moreover, it's an *efficient algorithm*, taking very little computer running time - it's at worst a multiple of the *square* of the number of cities, so if the number of cities increases ten-fold, the computer time goes up by a factor of only 100, which is very efficient.

The other problem I mentioned is the *travelling salesman problem*. Here, a salesman wishes to sell his wares in a number of cities, so he visits them all and then returns to his starting point. How can he plan his route so as to minimise the total distance?

One route is to go round the outside: A→B→C→D→E and back to A, with total distance 29. Or he could follow the star-shaped route A→C→E→B→D→A - still distance 29. The route A→C→E→D→B→A is slightly better, with length 28. The best possible route is A→D→B→C→E→A, with length 26, but to find it I had to use trial and error: no known efficient algorithm is guaranteed to produce it - but neither has anyone ever proved that no such efficient algorithm can exist.

A problem that arises in computing science, and in many practical situations (such as when you're trying to pack the boot of your car) is the *bin-packing problem*. Here's an example of it:

*I want to place objects of the following sizes into bins of capacity 12.*

<table>
<thead>
<tr>
<th>bin</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>size</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

*How many bins do I need?*

There are several algorithms you might use to solve this problem. The simplest is the *next-fit algorithm*, where you go along the
list putting objects into the bins in turn until there's no room. Thus, into bin 1, place objects \( A, B, C \) (with size 11), and then:

\[
\begin{align*}
\text{bin 2: } & D, E, F \\
\text{bin 3: } & G, H \\
\text{bin 4: } & I, J \\
\text{bin 5: } & K \\
\text{bin 6: } & L
\end{align*}
\]

using six bins in total. This method always gives an answer that's at most twice the optimum one.

A slightly better method is the first-fit algorithm, where we go along the list as before, but put each object into the first bin available. Thus, the first three bins are as before, but when we come to object \( J \) we put it into bin 2, and we place object \( L \) into bin 4. This gives the following solution:

\[
\begin{align*}
\text{bin 1: } & A, B, C \\
\text{bin 2: } & D, E, F, J \\
\text{bin 3: } & G, H \\
\text{bin 4: } & I, L \\
\text{bin 5: } & K
\end{align*}
\]

using five bins in total.

These methods are rather simple-minded. After all, if you're trying to pack cases into your car, you'd probably fit in the largest ones first and then pack the others around them. This leads to the first-fit decreasing algorithm, where we first list the items in decreasing order of size before packing them. So here we first place object \( K \) (length 8) into bin 1. We then place the next largest one (\( A \)) into bin 2, and also the following one (\( H \)). We get the following solution:

\[
\begin{align*}
\text{bin 1: } & K, G \\
\text{bin 2: } & A, I \\
\text{bin 3: } & L, H, C \\
\text{bin 4: } & D, E, F, B \\
\text{bin 5: } & J
\end{align*}
\]

So again we use up five bins in total.

In this example, none of the above methods gives the optimum solution, which uses just four bins:

\[
\begin{align*}
\text{bin 1: } & A, C, D \\
\text{bin 2: } & B, E, F, G \\
\text{bin 3: } & H, K \\
\text{bin 4: } & I, J, L
\end{align*}
\]

There are many problems involving network-type diagrams, which we now call graphs: this is nothing to do with the usual meaning of that word. The travelling salesman problem involved visiting all the cities and returning to the starting point - and we can ask a similar problem for other graphs - for example, is there a cycle that visits all the corners of a cube and returns to the starting point? (Such a cycle is called a Hamiltonian cycle, after the Irish mathematician and physicist Sir William Rowan Hamilton.) A little experimentation shows that indeed there is. More generally, given a graph, is there a good algorithm for finding whether it has such a Hamiltonian cycle? - no such algorithm is known.

Another type of graph problem involves trying to draw a given graph diagram in the plane in such a way that none of the linking lines cross each other. For some graphs this can be done, whereas for others (such as the one in the minimum connector and travelling salesman problems) no such drawing exists. Is there an efficient algorithm for deciding whether a given graph or network can be drawn in the plane with no crossing lines?

Yet another problem, known as the graph isomorphism problem, involves trying to decide whether two graphs are the same. Given two different network drawings, how can we decide whether they are different drawings of the same graph, or whether they arise from different graphs? Again, no efficient algorithm is known - in general, we may need to check all possible ways of re-labelling the points of the graph, and that's extremely inefficient. Even for graphs with just ten points, there are over three million re-labellings we'd need to check.

It's time that we stepped back and looked at such problems from a more general point of view. In his famous Essay on population in 1798, Thomas Malthus contrasted the linear growth of food supplies with the exponential growth in population. He concluded that however well we might cope in the short term, the exponential growth would win in the long term, and there'd be severe food shortages - a conclusion that's been borne out in practice.

A similar distinction can be made for algorithms used to solve problems. Each problem has an input size, which we'll call \( n \) - it might be the number of cities in a network. The algorithm also has a running time, which we'll call \( T \), which may be the time that a computer needs to carry out all the necessary calculations, or the actual number of such calculations involved. In any case, the running time \( T \) depends on the size \( n \) of the input.

Particularly important, because they're the most efficient, are the polynomial-time algorithms, in which the maximum running time is proportional to a power of the input size - say, to \( n^2 \) or \( n^7 \). The collection of all polynomial-time algorithms is called \( P \).
In contrast, there are algorithms that are known not to take polynomial time - such as the exponential-time algorithms in which the running time may grow like $2^n$ or faster. The collection of all algorithms that are known to have no polynomial algorithm is called not-P. (This is not the same as NP, which we explain below.)

In general, algorithms in $P$ are efficient, and problems that can be solved using them are called tractable problems. Algorithms in not-P are not efficient, and the corresponding problems are called intractable problems.

To see the difference, let’s draw up a table comparing the running times for input sizes $n = 10$ and $n = 50$, for various types of polynomial and exponential algorithm and for a computer performing one million operations per second:

<table>
<thead>
<tr>
<th></th>
<th>$n = 10$</th>
<th>$n = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>0.00001 seconds</td>
<td>0.00005 seconds</td>
</tr>
<tr>
<td>$n^2$</td>
<td>0.0001 seconds</td>
<td>0.0025 seconds</td>
</tr>
<tr>
<td>$n^3$</td>
<td>0.001 seconds</td>
<td>0.125 seconds</td>
</tr>
<tr>
<td>$n^5$</td>
<td>0.1 seconds</td>
<td>5.2 minutes</td>
</tr>
<tr>
<td>$2^n$</td>
<td>0.001 seconds</td>
<td>35.7 years</td>
</tr>
<tr>
<td>$3^n$</td>
<td>0.059 seconds</td>
<td>$2.3 \times 10^{10}$ years</td>
</tr>
</tbody>
</table>

There are many polynomial algorithms around: their running time is at worst some power of the input - usually the square, or the cube. Examples include:

- the greedy algorithm for the minimal connector problem, whose running time is proportional to the square of the number of cities;
- various algorithms for determining whether a graph with $n$ points is planar - some of these have running time proportional to $n$.

On the other hand, there are several problems for which no polynomial algorithm has ever been found. These include the travelling salesman problem, the bin-packing problem, the problem of deciding whether a given graph has a Hamiltonian cycle, and the graph isomorphism problem.

At this point we introduce NP. You’ll recall that $P$ is the set of ‘easy’ problems, those that can be solved with a polynomial-time algorithm. As far as we know, the Hamiltonian cycle problem is not one of these - but if you give me an attempted solution, then I can check in polynomial time that it is a solution. In general, NP is the set of ‘non-deterministic polynomial-time problems’ - those for which a solution, when given, can be checked in polynomial time.

Clearly $P$ is contained in NP, since if a problem can be solved in polynomial time, the solution can certainly be checked in polynomial time - checking a solution is far easier than finding it in the first place - but are they actually the same? That’s the big question: is $P = NP$? It seems very unlikely - indeed, few people believe that they are the same - but it’s never been proved.

We say that a problem is NP-complete if its solution in polynomial time means that every problem in NP can be solved in polynomial time. These include all your favourite problems. The travelling salesman problem is NP-complete - and so are the graph-isomorphism problem, the bin-packing problem, the Hamiltonian cycle problem, and thousands of others. If we could find a polynomial algorithm for just one of them, then polynomial algorithms would exist for the whole lot, and $P$ would be the same as NP. On the other hand, if we could prove that just one of them has no polynomial algorithm, then none of the others could have a polynomial algorithm either, and $P$ would then be different from NP.

The problems we’re dealing with here are of highly practical importance. There’s an enormous amount of money at stake in determining whether $P = NP$ - and it’s not just the million-dollar prize that we started with!

The Poincaré conjecture

I’d like to conclude with the one millennium problem that’s been solved - the Poincaré conjecture, named after the French mathematician Henri Poincaré.

Let’s start with a couple of surfaces in three-dimensional space - the sphere and the torus. They’re clearly not the same - you can’t deform either of them into the other, whereas you can deform a sphere into a cube or a torus into a teacup.
One way of seeing that they're not the same is to take a loop of thread, lay it on the surface, and try to shrink it to a point. With the sphere you can put the loop anywhere and it will shrink to a point. On the torus you sometimes can - but not always, because the hole might get in the way.

Let’s now restrict our attention to spheres, but let’s change the dimension. A circle is considered to be one-dimensional, even though it lives in 2-dimensional space, because it's just a bent-around line. We can join two circles by bending them and glueing corresponding points together; we get a sphere. The surface of this sphere is considered to be two-dimensional, even though it lives in 3-dimensional space, because it's just a bent-around couple of circles.

We can continue in this way. To get a three-dimensional sphere, we take two ordinary spheres and glue corresponding points together. This can’t be done in our three-dimensional world, but it can be imagined mathematically - and indeed, we can go on in this way and consider four-, five-, and even higher dimensional spheres.

Now to the Poincaré conjecture. With our original surfaces, the sphere is the only surface that has this loop-shrinking property - the torus doesn’t, as we saw, and nor does any other surface. Is this true in higher dimensions? Is the sphere the only surface with the loop-shrinking property. For the two-dimensional sphere the answer is yes - and in the 1960s Stephen Smale proved that the answer is yes for dimensions of 5 or more. But what about dimensions 3 and 4? In 1982, Michael Freedman proved it for dimension 4, but no-one could do dimension 3, until along came Grigory Perelman who succeeded around 2002.

Many of you will know that Perelman shuns publicity and seems to believe that mathematics should be done for its own sake alone and not for financial reward. Smale and Freedman received Fields medals (the mathematical equivalent of the Nobel prize) for their contributions to solving the Poincaré conjecture, but Perelman turned it down. He will also probably turn down the offer of the million dollars.

**Conclusion**

This concludes my four years of Gresham lectures in geometry. Although I'll be giving some lectures in the spring and summer, this concludes my fourth and final official series. May I thank you for coming along so regularly, and I hope that you will equally support my successor, whoever it is, over the coming years.

©Professor Robin Wilson, Gresham College, 27 February 2008