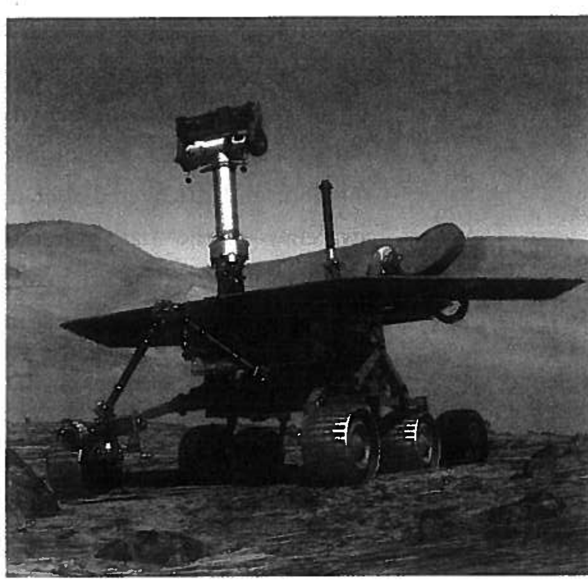
Ian Stewart

Behind the Scenes: The Hidden Mathematics That Rules Our World

My story begins with a strange event, which took place on 4 January 2004, on Mars. A Martian wandering around near Gusev Crater on that particular day would have undergone a life-changing experience. First, a streak of fire high in the sky would have heralded the arrival of an alien artefact, descending rapidly beneath a hemisphere of fabric. Then, as the artefact neared the ground, the fabric would have torn away, allowing it to fall the final hundred metres. And bounce. In fact, it bounced twenty-seven times before finally coming to rest. It would certainly have been a sight to remember.

The bouncy visitor was Mars Exploration Rover A, otherwise known as *Spirit*. After a journey of 487 million kilometres it entered the Martian atmosphere at a speed of 19,000 kilometres per hour. It was still travelling at a healthy 50 kilometres per hour a few seconds before impact when its airbags inflated and it made its touchdown. *Spirit* and its companion *Opportunity* have now spent more than four years exploring the surface of Mars, nearly twenty times as long as originally planned, leading to a wealth of new scientific information about Earth's sister planet. They may not have finished yet.

Much of the credit for this stunning success must go to NASA's engineers and managers, but other disciplines were also essential – among



Artwork of the Mars Exploration Rover *Spirit* landing on and exploring the surface of the planet.

them, mathematics. The spacecraft's trajectories were calculated using Newton's laws of motion and gravity; Einstein's later refinements were not needed. Isaac Newton was elected a Fellow of the Royal Society in 1672, twelve years after the Society was founded. His role in the development of space travel is not hard to identify, even though he died 240 years before the first Moon landing. Less obvious is the influence of a Fellow from the Victorian era, George Boole, whose pioneering ideas in logic and algebra proved fundamental to computer science. His influence can be detected in the error-correcting codes that made it possible for the Rovers (and most other space missions) to send images and scientific data back to Earth. Mathematics, both ancient and modern, is deeply embedded in today's science, and makes vital contributions on a daily basis to many aspects of human society.

The importance of mathematics in the space programme should be evident even to a casual observer. Yet when the Rovers landed, and the American mathematician Philip Davis pointed out that the mission 'would have been impossible without a tremendous underlay of mathematics' – so tremendous, in fact, that 'it would defy the most knowledgeable historian of mathematics to discover and describe all the mathematics that was involved' – he found it necessary to add that 'The public is hardly aware of this.'

This remark was an understatement. In 2007 two Danes with postgraduate mathematics degrees, Uffe Jankvist and Björn Toldbod, decided to uncover the hidden mathematics in the Mars Rover programme. They visited NASA's Jet Propulsion Laboratory at Pasadena, which ran the mission, and discovered that it is not only the general public that lacks awareness of the mathematics used in the Rover mission. Many of the scientists most intimately involved were also unaware of the mathematics being used. Some denied that there was any.

Opposite: Portrait of George Boole.

'We don't do any of that,' said one. 'We don't really use any abstract algebra, group theory, and that sort.'

'Except in the channel coding,' one of the Danish mathematicians pointed out.

'They use abstract algebra and group theory in that?'

'The Reed–Solomon codes are based on Galois fields.'

'That's news to me. I didn't know that.'

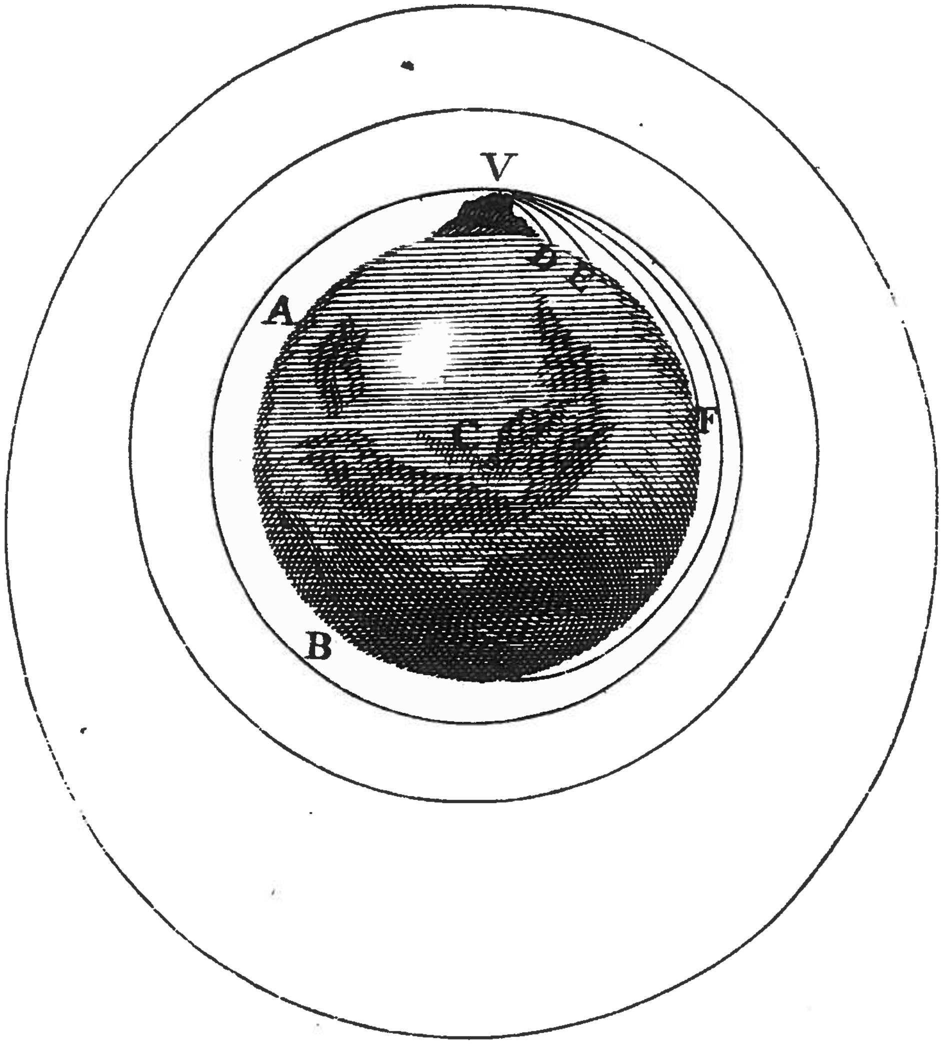
This story is fairly typical. Few people are aware of the mathematics that makes their world work. Indeed, few are aware that mathematics is involved in their world *at all*. But – as the history of the Royal Society exemplifies – mathematics has long been central to science, and science has long been a major driving force for social change.

What causes this lack of awareness of the importance of mathematics in the modern world? One of the main reasons, as the NASA story shows, is that you don't have to know any mathematics, or even be aware of its existence, to use the technology that it enables. This is entirely sensible – you don't need to understand computer programming to buy CDs over the Internet, and you don't need a degree in engineering to drive a car. However, most computer users are aware that someone had to write the software, and most drivers realise that someone had to design and build the car. With mathematics, it seems to be different.

Why? The story of the Mars Rovers is instructive. JPL scientists did not realise how deeply mathematics was involved in the Rover mission because the mathematical techniques were built into dedicated computer chips and programs. The resulting hardware and software carried out the necessary calculations without human intervention. Moreover, most of the chips and software were designed and manufactured by external subcontractors.

In actual fact, the Rover mission rested on a huge variety of mathematical techniques. These included dynamical systems and numerical analysis to calculate and control the spacecraft's trajectory on its way to Mars, signal processing methods to compress data and eliminate transmission errors caused by electrical interference, even the design and deployment of the

Engraving published in Newton's *A Treatise of the System of the World*, 1728, showing the effects of gravity.



airbags. These techniques did not come into being overnight, and they were not, initially, developed with the space programme in mind. The work of Newton makes this very clear.

Newton's father was a Lincolnshire farmer, who died three months before his son was born. The boy did not impress some of his schoolteachers, who reported that he was idle and inattentive, but he did impress his headmaster, who persuaded Isaac's mother to send him to university. At Cambridge he studied law, but he also read books on physics, philosophy and mathematics. In 1665 the university was closed because of plague, and he returned to Lincolnshire. There, in a few years, he made huge advances in several areas of mathematics and physics, which led to his election as a Fellow of Trinity College.

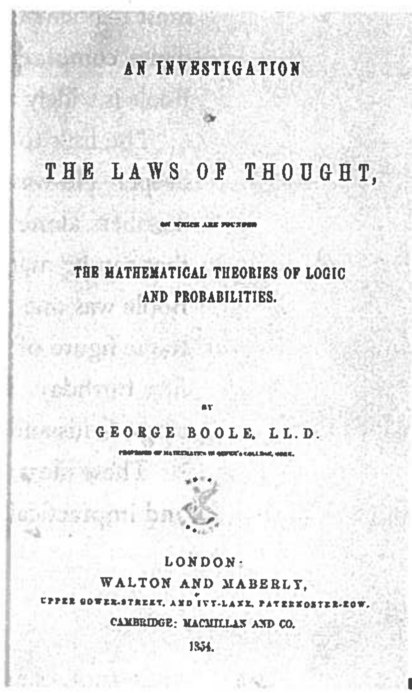
Newton is famous for many things – his laws of motion, calculus (also discovered by Gottfried Wilhelm Leibniz), the beginnings of numerical analysis. All of this work leaves fingerprints on the Mars Rover mission, but the most significant is the law of gravitation. Every body in the universe, Newton declared, attracts every other body with a force that is proportional to their masses and inversely proportional to the square of the distance between them. When coupled to his laws of motion, the law of gravitation provided accurate descriptions of the motion of the assorted planets and moons of the solar system, and much more. It explained the curious way in which the Moon wobbled on its axis, and the paths of comets. It made the future of the solar system predictable, millions of years ahead.

Newton's motivation was 'natural philosophy', the scientific study of Nature. If he had practical objectives in mind, they were related to things like navigation, and were secondary to understanding what he called 'the system of the world', which was the subtitle to his epic *Principia Mathematica* (*Mathematical Principles of Natural Philosophy*). At that time, the idea that humans might travel to the Moon was considered absurd, when anyone considered it at all. Yet such is the power of mathematics that when spacecraft began to leave the Earth in the 1960s, the tools needed to calculate their orbits and plan their re-entry trajectories through the atmosphere were those developed by Newton and his successors. In particular, since the law of gravitation applies to every particle of matter in the universe, it must apply to spacecraft.

NATURAL PHILOSOPHY HAS BORNE FRUIT AS TECHNOLOGY

Once pointed out, it's no great surprise that esoteric mathematics can be used in esoteric applications like Martian space probes, even if no one notices ... But what does that have to do with the everyday life of the ordinary citizen? Next time you listen to a CD while driving along the motorway in your car, and hit a bump, you may care to ask yourself why the CD player skips tracks only if it's a really *big* bump – big enough to risk damaging your wheel. After all, a CD player is an extremely delicate device, with a tiny laser that hovers a few millionths of a metre away from a plastic disc covered in tiny dots.

The answer goes back to George Boole and the other nineteenth-century mathematicians who founded modern abstract algebra. Boole also hailed from Lincolnshire, being born in Lincoln in 1815; his father was a cobbler who was also interested in making scientific instruments, and his mother was a lady's maid. He did not take a university degree, but his talent for mathematics attracted attention, and in 1849 he became Professor of



AN INVESTIGATION

OF

THE LAWS OF THOUGHT,

ON WHICH ARE FOUNDED

THE MATHEMATICAL THEORIES OF LOGIC
AND PROBABILITIES.

BY

GEORGE BOOLE, LL.D.

PROFESSOR OF MATHEMATICS IN QUEEN'S COLLEGE, CORK.

LONDON:
WALTON AND MABERLY,
UPPER GOWER-STREET, AND IVY-LANE, PATERNOSTER-ROW.
CAMBRIDGE: MACMILLAN AND CO.
1854.

The frontispiece to George Boole's *An Investigation of the Laws of Thought*, 1854.

Mathematics at Queen's College, Cork. His most significant work was his 1854 book *An Investigation of the Laws of Thought*. In it, he reformulated logic in terms of algebra – but a very strange kind of algebra. Most of the familiar algebraic rules, such as $x+y = y+x$, are valid in Boole's logical realm, but there are some surprises, such as $1+1 = 0$. Here 1 means 'true', 0 means 'false', and $x+y$ means what computer scientists now call 'exclusive or': either x is true, or y is true, *but not both*. The first formula says that this statement does not depend on the order in which the two statements x and y are considered. The second says that if x and y are both true, then $x+y$ is false – because the definition of + includes the requirement 'not both'. More elaborate algebraic laws, such as $(x+y)z = xz+yz$, are also true in Boole's system; now the product xy means 'x and y'. So Boole's algebraic rules follow from sensible logical ones.

It is a striking and surprising discovery. Logic, previously thought of as being more basic than mathematics, can actually be *reduced* to mathematics. And the reduction is so natural that the algebra of logic is almost the same as traditional algebra. The new rules do make a difference, but you soon get used to that. Boole knew he was on to something important, but it took a while for most mathematicians to appreciate it. 'Boolean algebra' really took off when digital computers started to appear. Computers are basically logic engines, and Boole is widely recognised as a founder of theoretical computer science.

The link to digital computation is natural, but Boole's influence runs deeper. He was one of the first to realise that algebra need not be about numbers alone: it can be about any mathematical concepts or structures that can be manipulated symbolically according to a fixed system of rules. Boole was one of the earliest thinkers in a long tradition that includes the tragic figure of Évariste Galois, killed in a duel shortly before his twenty-first birthday. Today's abstract algebra, with its key concepts of groups, rings, fields and vector spaces, represents the fruits of their early labours.

These ideas, if I were to explain them in any detail, would seem abstract and impractical – formal games played with symbols, to no clear purpose.

They look like that because they operate on a structural level and focus on deep generalities. But behind the scenes, the abstract algebra that Boole pioneered has taken over most areas of mathematics, because it organises concepts and provokes new ideas. The resulting mathematics can be found, embodied in computer chips, inside most of today's electronic gadgets: CDs, DVDs, digital TVs, mobile phones, iPods, Nintendo Wiis, BlackBerries, SatNav, digital cameras ...

Reed–Solomon codes are a typical example. These are the codes that NASA used to detect and correct potential errors in the Rovers' images of the Martian surface as they were beamed across the vastness of the solar system to planet Earth. More familiar devices, such as CD players, also would not work without Reed–Solomon codes. These codes hinge on, and were motivated by, the algebraic legacy of Boole and Galois. They transform the digital data that represents music in a way that makes it easy to spot, and put right, any errors that occur when the CD is being played. Virtually all of today's digital communications are wholly reliant on sophisticated and very modern mathematical coding methods. None of it would work without them. And that turns out to be just the tip of a very large iceberg.

A few weeks ago I looked through a randomly chosen issue of *New Scientist* magazine. Of the fifty or so stories reported, there were a dozen that – to my sensitive eye – involved a significant amount of mathematics. Not one story mentioned this, though a few hinted about 'models' of the process under study. When the contribution of mathematics is hidden that far behind the scenes, it is hardly surprising that the media and the public have little idea of what mathematics is, or what it is good for.

Sometimes mathematics should be kept behind the scenes. When I listen to music in my car, I don't want to have to think about the intricacies of Galois fields. When NASA engineers are firing a space probe's rockets to nudge it into the right entry trajectory to prevent it burning up in the Martian atmosphere, they don't want to be worrying about differential equations. But *someone* has to do the sums, write the program, design the algorithm, invent the concept, or prove the theorem. Someone has to provide the tools for the job and make sure they are reliable. If neither the media, nor the public, nor even practising scientists realise that this hidden mathematics exists, we will stop training mathematicians, and the necessary people will cease to exist too.

To most of us, 'mathematics' is something we did at school, and promptly forgot. Curiously, many of us also think that what we did at school was *the whole of mathematics*: all done and dusted. And pointless, now that we've got computers to do the sums for us. Some of us discover there is more to it than that. Some go on to university, take a science degree, and come to grips with statistics (in biology or medicine), differential equations (physics and engineering), or mathematical logic (computer science). And the mental picture that we get is that there's a certain amount of genuinely *useful* stuff (statistics, differential equations, mathematical logic ...) plus a lot of highbrow intellectual fun and games that never has been and never will be useful to anyone living in the 'real world'.

Both of these views of mathematics are caricatures; real mathematics is quite different. Today's mathematics is intimately bound up with two key areas of human knowledge and activity: the natural world, and the society in which we live. Human understanding of our planet, and our universe, rests heavily on the shoulders of mathematics. So does the day-to-day working of our world. Take the hidden mathematics away, and today's world would fall to pieces. That statement applies to a lot of the apparently esoteric parts of the subject, as well as the more obviously applicable ones – partly because mathematics is an interconnected whole, but also because the esoteric concepts are often very general and very powerful. New and unexpected applications are common.

The 'classical' areas of mathematics are mainly those that led up to, or developed from, calculus – continuous mathematics, where everything can be subdivided into pieces that are as tiny as you wish. Most core mathematical physics and classical applied mathematics, such as acoustics or aerodynamics or elasticity theory, are of this kind. An important newcomer is discrete mathematics, which is suited to the digital age. Here the basic ingredients come in indivisible packets; essentially, anything whose natural description uses whole numbers or finite lists of symbols. Straddling both areas is the theory of probability, a mathematical description of uncertainty.

Geometry is also crucial. Despite appearances to the contrary, mathematics is primarily visual, and the formal symbolism tends to be closely related to some kind of mental image. Today's geometric thinking, however, takes a variety of forms, few of them resembling the traditional geometry of Euclid. Modern mathematics rightly places value on generality, when appropriate. That naturally leads to a degree of abstraction, because the focus of attention has to shift from 'what objects are we looking at?' to 'what properties are we assuming?' Logical proof remains central to the enterprise; it's how mathematicians keep themselves and their subject honest. Computers now play an increasingly central role. They seldom solve problems without further thought, but they can create a huge improvement in our understanding when they are used intelligently.

Mathematics, embodied in digital devices, has made technologies possible that seem to verge on magic. In February 2008 my wife and I spent two weeks exploring the private tombs of the Egyptian nobility, from Cairo down to Aswan. We took more than 1,400 photographs with two digital cameras; the whole lot were recorded on three 1-gigabyte memory cards, each the size of a postage stamp. The engineering feats involved are amazing, and they rest on all sorts of advances in materials science, photolithography, even quantum mechanics. Those advances required a lot of mathematics, as it happens, but I want to focus on just one aspect of digital cameras: data compression. The quantity of raw information required to specify 1,400 high-resolution colour pictures is far larger than those three cards can hold. Despite huge advances in miniaturisation, you simply cannot get that amount of data into such a small space.

A Secure Digital (SD) memory card, widely used in digital cameras for storing JPEGs.

Yet the pictures exist. I can print them out, or put them on the computer screen. How do the camera manufacturers cram so much information into so little memory? It may seem like magic, but the magic is mostly invisible mathematics. The clue lies in the names of the image files, which on my camera look something like P1000565.JPG. This tells the computer that the file is formatted using the JPEG standard, issued by the Joint Photographic Experts Group in 1992. This format uses various features of human vision, and typical images, to 'compress' the image data substantially.

In general terms, a computer represents a picture as a list of numbers. The list represents a rectangular array of tiny picture elements, called pixels, and the numbers describe the colour and the brightness of each pixel. If you do the sums, however, you find that there's nowhere near enough space in a memory card to hold all the pictures that undeniably are in there. It's not just like trying to get a quart into a pint pot: more like getting a tanker-load of milk into a pint pot.

This problem is a common one in the digital world, and it is usually tackled by compressing the data – reducing the quantity of information while retaining enough of it to do the job. Just as you can get more luggage into the car if you load it in the right way, so you can get more of the important data into a computer file if you leave out stuff that's not really relevant, or take advantage of certain inbuilt redundancies. For instance, many photographs have a large area of blue sky. Instead of repeating the code for 'blue' thousands of times, once for each pixel, we could tell the computer 'colour everything in this rectangle blue', and specify the rectangle by listing its corners. Suddenly thousands of numbers collapse to a few dozen. That's not how JPEG works, but it shows how redundancies in a list of numbers may make the list compressible. The actual procedure is carefully tailored to what can be done efficiently inside a small camera. The details don't really matter for my main point, but I want you to appreciate that there *are* details, which use several different mathematical ideas. So please indulge me while I tell you just how cunning the process is.

JPEG starts by splitting the data into three separate arrays. One lists how bright each pixel is. The other two take advantage of the fact that the colours perceived by the eye can be specified as points in a plane, the 'colour triangle'. A plane is two-dimensional, so each point can be defined using just two numbers, its horizontal and vertical coordinates. These 'colour components' form the other two lists. The human eye is more sensitive to variations in brightness than in colour, so the two lists of colour components can be shortened – usually they are reduced to one quarter of their original size – by using a coarser list of colours.

The next step uses a trick introduced by the French mathematician Joseph Fourier in 1824 – a year after his election to the Royal Society, as it happens – who at the time was working on the flow of heat. In general terms, Fourier's idea was to represent a pattern of numbers by combining specific patterns with different frequencies – much as the note played by a clarinet is made up from a fundamental 'pure' note and various higher-pitched 'harmonics', all added together in suitable proportions. JPEG uses a similar trick for spatial patterns of numbers, treating each of its three arrays in the same way. First, the array is broken up into 8x8 blocks of pixels. Then each block is transformed into a list of its spatial frequencies in the horizontal and vertical directions. Roughly, this splits the pattern into black-and-white stripes of various thicknesses, and works out how much of each stripe you need to reconstruct the actual image. This step employs a fast Fourier transform, exploiting number-theoretic features of binary numerals to speed up a difficult computation; this is why 8x8 blocks are used, eight being a power of two. The Fourier transform does not compress the data, but rewrites it in a compressible form. The eye is fairly insensitive to high-frequency stripes, so these can be ignored. Medium-frequency stripes can be specified using smaller numbers, which occupy less space on the memory card.

This is not the end: two more tricks are used to squash even more pictures into the same space. If you run through the resulting array of numbers in a zigzag order, from low frequency components to high ones, you typically find runs of repeated numbers, such as 7 7 7 7 7 7 7 7 7. Coding this as '9 consecutive 7's' converts it to 9 7, which is shorter. Finally, another coding method called Huffman coding is used on the resulting file, which compresses it even further.

So JPEG coding is quite complex, with sophisticated mathematical features. You don't need to know how it works to use your digital camera, but without the underlying ideas, that camera could never have been made. Now think of future developments, video cameras, cramming a camera into a mobile phone along with dozens of other applications ... We desperately need people who can understand that sort of mathematics.

At any rate, my wife and I were able to take lots of pictures without carrying sacks full of film because a lot of mathematically sophisticated engineers noticed that something that a nineteenth-century Frenchman invented for a completely different reason happened to have an unexpected use. But the hidden mathematics behind our holiday didn't stop there. Without a lot of other mathematics, often with similarly impractical or outmoded origins, we could never have got to Egypt to take the pictures.

Our flight was booked over the Internet and all Internet communications rely on error-correcting codes to ensure that messages are not garbled along the way by electrical interference. Like the codes used by the Mars Rovers, these techniques rely heavily on abstract algebra. The airline's schedules were designed using mathematical methods to improve efficiency – graph theory and linear algebra. Then there was radar, weather-forecasting, even the statistical analysis of different breeds of vegetables that governed the crops from which the airline food was made.

None of this is much use if the aircraft never gets to its intended destination. In the early days of navigation, when the great European explorers were mapping the globe in small wooden sailing ships, navigation was a major consumer of mathematics. Even finding the size and shape of the Earth involved mathematical calculations, as well as experimental observations.

Today we have GPS, the Global Positioning System, which comprises about fifteen satellites orbiting the Earth, sending out signals. A triumph of electronics and engineering, obviously. But mathematics?

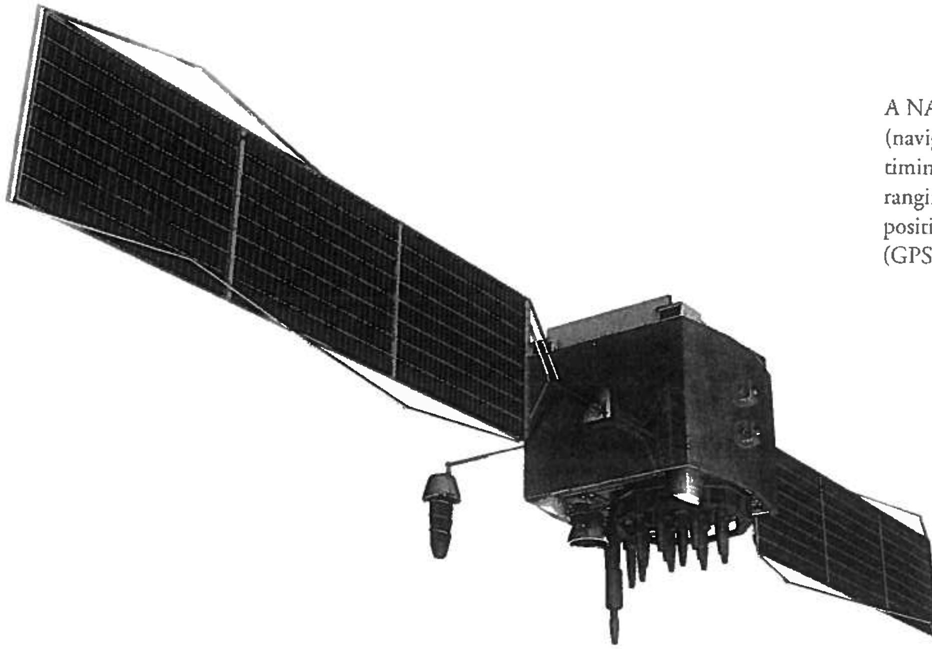
Leaving aside the heavy use of mathematics in designing and building launch vehicles and satellites, and in calculating orbital dynamics, let me focus solely on the signalling system that GPS uses. Each satellite transmits a signal, which can be used to work out how far away the satellite is from the GPS receiver (on board the aeroplane, ship, car, yacht, or inside someone's mobile phone). These distances, coupled with knowledge of the positions of the satellites, make it possible to calculate the location of the receiver on the surface of the Earth. That's another highly mathematical step, which I will also ignore.

How do the signals convey distances?

Imagine that the satellite is playing a tune, and that you have access to a second 'copy' of that tune, being sent out from a known source that is in synchrony with the satellite. Because the satellite is further away than the reference source, the signal from the satellite is slightly delayed, by a time equal to the difference in distances divided by the speed of light. The time delay can be measured, very accurately, and the distance is obtained by multiplying that by the speed of light.

Instead of tunes, the signals are sequences of pseudo-random numbers – apparently patternless sequences generated by a fixed mathematical recipe. Both the satellite and the reference source know this recipe, so they can generate and recognise the same signals. So here we find a very practical application of the mathematics of pseudo-random numbers. If you use SatNav in your car, you are a major consumer of the hidden mathematics that runs our world.

Still pursuing the hidden mathematics that made my holiday possible, there is the small matter of designing an aircraft that stays up, one of the heaviest uses of mathematics in the whole enterprise. Nearly all of the analysis of airflow past an aircraft nowadays is done using 'numerical wind-



A NAVSTAR (navigational signal timing and ranging) global positioning system (GPS) satellite.

tunnels', which are mathematical simulations. They are much easier to use than physical wind-tunnels, and if anything, more accurate. They have innumerable other applications. They are essential to the design of Formula 1 and NASCAR racing cars, where effective aerodynamics is needed to keep the car on the track and reduce air resistance. If that's not green enough for you, the same techniques improve the fuel efficiency of ordinary road vehicles. Even the dynamics of a football has been analysed mathematically, with useful practical implications about how to make the ball behave unpredictably, which can help it get past the keeper into the goal. Computational Fluid Dynamics also has medical applications to blood flow and heart disease.

This makes the point that mathematics also saves lives. Have you had a medical scan recently? How do you think the scanner works out what's inside you? There's a whole branch of mathematics devoted to such questions. Are you concerned about crime? The FBI uses 'wavelets', a very recent piece of mathematics, to analyse and record fingerprint information to help catch criminals. Other police forces use similar techniques. Do you use oil or natural gas, for heating, cooking, or transport? The oil companies use powerful mathematical techniques to find out what the rocks miles underground look like, based on the echoes from explosions at the surface. Do you use anything with a spring in it – ballpoint pen, video recorder, mattress? The spring-making industry uses mathematics for quality control.

Another huge area that relies on mathematics is science, and science is our most successful method for understanding the natural world. The development of science, and that of mathematics, have gone hand in hand for about five hundred years. Newton invented calculus to understand the movements of the planets. Independently, Gottfried Leibniz developed much the same ideas for purely intellectual reasons. These two sources of mathematical inspiration can be roughly characterised as 'applied' and 'pure' mathematics. The main differences are motivation and attitude, rather than content. The same mathematical concept may appear in the solution of Fermat's last theorem (pure mathematics) or in the construction of a secure code for Internet banking (applied mathematics). Some areas are traditionally considered as being 'pure', others as 'applied', but these are convenient distinctions, not impassable barriers. Today's science is increasingly multi-disciplinary; so is mathematics.

Initially, the main beneficiaries of mathematical techniques were the physical sciences, and these are still the areas in which the use of mathematics is greatest. But the biological and medical sciences are catching up rapidly, and some of the most interesting new problems for research mathematicians are coming out of biology. A century or two from now we will look back at today's Newtons and Booles, and understand how vital their work has been to the development of our society. Provided we do not lose sight of the hidden mathematics that rules our world – because if we do, those advances will never happen.