Contents

1

Most Extreme

Where do mathematicians get their problems from? How do they go about solving them? Where do they get new ideas from? Do they work in isolation, compete, or cooperate? Why do they seek proofs of statements that seem obvious? Why do many of them devote time and energy to 'useless' areas such as number theory, instead of tackling 'practical' problems head on? What, for that matter, *is* mathematics? What are mathematicians? Have computers made them obsolete? If not, will artificial intelligence shortly administer the coup de grâce?

To investigate these questions, we'll look at typical examples of great mathematics and try to get inside the minds of the mathematicians who worked on them.

Attempts to define mathematics in a few words are numerous and largely unsuccessful, in part because of its rapid growth – millions of pages per year, all of it new. It's hard to define something when it's changing under your feet. 'The science of significant form' is one of the better definitions. 'What mathematicians do' is closer to the truth, but not very helpful. Similarly, when it comes to defining mathematicians, the most accurate definition is 'people who do mathematics', which isn't terribly helpful either.

Also, I feel, missing the point.

Consider the same questions for business. What is business? What business people do. What is a business person? Someone who does business. Also missing the point. A business person doesn't just *do* business: they see *opportunities* to do it where most of us don't. It's the same for mathematics: a mathematician is a person who sees opportunities to do mathematics where most of us don't. They can do that because they're immersed in a long tradition of seeking such opportunities, and collectively they've amassed a huge toolkit of methods for exploiting them.

Mathematics is such a vast subject that any discussion has to be selective.[1] In this book I'm going to discuss a variety of problems and their solutions, linked by a common theme: the search for extremes. Mathematicians don't just find solutions to problems; they seek the best solutions, the simplest proofs, the ideas that give the most insight. They seek more informative *problems*, too.

Every problem is an opportunity.

Every opportunity is a problem.

. . .

One of the main aims of this book is to provide insights into how mathematicians *think*. Where they get their problems from. How they get original ideas. How, over the course of time, contributions from innumerable people, some of them intellectual colossi, others bread-and-butter workers, blend together to create vast, deep theories that transcend their humble origins. There's no unique 'mathematical mentality'. Mathematicians go about their work each in their own distinctive way. But, because they all belong to the same global network and contribute to the same subject, they have many thought patterns in common.

Mathematics is a natural human activity, already in action at least 20,000 years ago when someone living in what is now the Democratic Republic of the Congo scratched a series of marks on a bone. We don't know what animal this bone came from, because it was heavily worked, but it seems to have been some kind of counting tool. A tally stick, with notches like |, ||, |||, |||| to record numbers. A lunar calendar. A device used by a woman to track her menstrual cycle. We don't know for sure. Maybe the marks were just made to provide a better grip and the numerology is just coincidence. It's called the Ishango bone, because it was discovered in a small fishing village of that name.

There are more contentious bones that might be even earlier signs of mathematical thought. The Lebombo bone, radiocarbon dated to around 42,000 years ago, was found in the Lebombo mountains, which separate South Africa from Eswatini (formerly Swaziland). It looks like a tally stick, and has twenty-nine notches that might be evidence for a lunar calendar (twenty-nine days in a lunar month). But the bone is broken and there might have been more notches.

The origins of mathematical thinking go back a long way, but – not surprisingly – they're shrouded in mystery and lost in the mists of time, to compound the clichés. But it does look like humans were starting to think symbolically and numerically tens of thousands of years ago.

By the time of the Old Kingdom of ancient Egypt, around 3000 BC, basic arithmetic was well established. By 2600 BC the Sumerians were incising multiplication tables in clay tablets. The inhabitants of Babylon were also writing numbers by this time, and went on to become the master mathematicians of the ancient world. They excelled in both theory and practice. Surviving clay tablets include solutions of quadratic equations, partial results on cubic equations, an impressive calculation of the square root of 2 written along the diagonal of a square, approximations of π. Astronomy and mathematics have always been closely linked, and the Babylonians were impressive astronomers too. The *Venus Tablet of Ammisaduqa* records the risings of the planet Venus over a twenty-one-year period. It is one tablet from the *Enûma Anu Enlil*, a compilation of about seventy tablets – mostly astrological: 'If the moon at its appearance wears a crown: the king will reach the highest rank.'

Mathematics as we know it began to take off with the geometry of ancient Greece, the arithmetic of India, the algebraic advances of the Persians, the anticipations of calculus by the Chinese… The full story here is complex, and is still being disentangled from the Eurocentric views that dominated the history of mathematics until very recently. But it shows that human beings have been developing mathematics for millennia, and that it's fundamentally a collective activity in which advances small or big, made by innumerable individuals, accumulate over time to create one of the most powerful thinking tools that humanity has devised.

. . .

Many of the deepest and most important areas of mathematics have arisen from questions about extremes – shortest lines, smallest areas, densest packings, fewest colours. Mathematicians have been grappling with extremes for centuries – even thousands of years. The isoperimetric problem, which asks for the shortest path enclosing a given area, can be traced back to a myth about the foundation of the city of Carthage in 814 BC. Other problems are of more recent vintage. The astonishing results of Maryna Viazovska in 2016, only the second woman to win the prestigious Fields Medal, tell us the densest ways to pack identical spheres in spaces of eight and twenty-four dimensions. The corresponding question for ordinary three-dimensional space was posed by Johannes Kepler in 1611 in an essay about snowflakes, and finally given a rigorous answer by Thomas Hales in 1998. Viazovska's work arose from a mathematical reflex: *generalise*. Use analogies to extend problems into broader areas, in this case 'spaces' of dimension bigger than the familiar three. The concept of higher-dimensional spaces has been around for less than 200 years, and it revolutionised mathematics and science. Generalisation, a major source of new mathematics, has proved remarkably fruitful. The underlying urge is *improvement*: a more general theorem is a better theorem.

Many of the problems I'll discuss were originally motivated by real-world issues, another common source of new questions. The 'Plateau problem', about the geometry of soap bubbles, led to the notion of a minimal surface – the surface of smallest area enclosing a given volume – which now has applications to topics as diverse as biology and cosmology. The origin of the 'travelling salesperson problem' – find the shortest route that visits a given set of cities – is self-explanatory. It was, in fact, first raised explicitly in 1832 by a German travelling salesman, back in the days when such people were exclusively men. There were reasons for this. At that time, the job required being on the road for most of the year, on horseback, and the prevailing culture discouraged women from taking on such a role.[2] The theory of geodesics – find the shortest paths on a curved surface or higher-dimensional space – applies to the paths followed by light rays, in both classical and Einsteinian physics. Its origins, however, trace back to very different questions, such as land measurement in antiquity, the quest to find the shape of the Earth, and more abstract issues in geometry.

Other problems have come from speculative questions asked by mathematicians, with no obvious practical use. The 'four-colour problem' asks whether any map on a plane or sphere can be coloured with four colours so that no two countries that share a border have the same colour. Its applications to real map-making are negligible, but its lure for pure mathematicians proved to be immense because the simplicity of its statement belies the tremendous difficulties involved in its solution. The search for a single shape of tile that covers the infinite plane, but cannot tile it 'periodically', that is, like a wallpaper pattern that repeats the same arrangement infinitely often, finally succeeded in 2023. It has no practical application yet, aside from covering a wall or floor, but it has intriguing links with logic. Tilings in general have a strong aesthetic appeal, and unusual patterns intrigue our jaded senses.

Reaching for the Extreme examines the workings of mathematical minds by telling the stories of this quest for extremes – where problems come from, why mathematicians find them interesting, their struggles to solve them, how progress has often come from unexpected directions, and the uses that can be made of the results. It emphasises the importance of pushing the boundaries of mathematics as far as they can possibly go, and finding out exactly where those boundaries lie.

. . .

Mathematicians are never satisfied.

Yes, they celebrate the latest breakthrough like anyone else – once they're convinced it's correct, which can be a long and tiresome process – but a breakthrough is hardly ever the end of the matter. A really good breakthrough in mathematics is a discovery that starts a whole new pursuit. One that raises new questions, with a realistic hope that answers can eventually be found.

In, or very close to, the year 1637, the lawyer Pierre de Fermat had an idea. In the margin of a classic textbook, the *Arithmetica* of the ancient Greek Diophantus, he wrote a brief note: 'It is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any power higher than the second, into two like powers. I have discovered a truly marvellous proof of this, which this margin is too narrow to contain.' Fermat stated many of his discoveries in letters to his contemporaries, but hardly ever wrote down the proofs. Over the years almost all of his claims were proved, but this particular statement stubbornly refused to cave in. It occupied the minds of numerous mathematicians for over three centuries; it became known as Fermat's Last Theorem, because it was the last of his statements that no one had succeeded in proving. Technically, it wasn't actually a theorem: theorems are statements with logically correct proofs. It was a 'conjecture', something that some mathematicians believe to be true but whose proof or disproof is elusive. But in Fermat's time the distinction between theorems and conjectures was a bit fuzzy, and anyway, it's an intriguing name, so it stuck.

Fermat's Last Theorem did in fact turn out to be a theorem, famously proved by Andrew Wiles 358 years later in 1995, using deep and difficult methods that couldn't possibly have been available to Fermat. New theories, new concepts, new techniques: Fermat would have had to invent a large part of the pure mathematics of the seventeenth to twentieth centuries even to get started.

Once the world's number theorists had recovered from the shock of Wiles's tour de force, the usual game played out. Congratulations were in order, and duly offered, but the next step was a reality check.

Is the proof correct? All too often, claims of major breakthroughs founder on this simple question. Problems that have hung around unanswered for centuries don't give in easily. Subtle errors can creep into any proof, and the more complicated the proof is, the more scope there is for such errors. And, to no great surprise, an error duly emerged. Ninety-nine times out of a hundred, that proves fatal. The wonderful new idea dissolves into smoke and mirrors in the harsh searchlight of logic. In this case, however, the story had a fairytale ending. After a desperate struggle, with the aid of a former student Richard Taylor, Wiles repaired his proof.

And they all lived happily ever after...? No, something much better happened.

It would be easy to assume that since everyone was trying to find the answer to Fermat's claim, the *answer* had to be very important.

Not at all.

Very little hinges on whether Fermat was right or wrong. Knowing which would be nice, but of itself the answer would do very little to advance the cause of mathematics. In 1816 Carl Friedrich Gauss, the leading mathematician of his day and an expert in number theory, wrote: 'Fermat's Last Theorem, as an isolated proposition, has very little interest for me, because I could easily lay down a multitude of such propositions, which one could neither prove nor dispose of.' Fair comment, to be honest, but like so many such pronouncements by people in authority, the overall thrust was fine but the chosen example turned out to be dead wrong.

Over the decades, other mathematicians did their best to rise to the challenge, spurred on by the feeling (which apparently eluded Gauss) that if you can't answer such a natural question, there's a big gap in your understanding. Entire new theories emerged from the quest, and mathematics was enriched beyond measure, even though Fermat's Last Theorem itself remained an enigma.

What made Wiles's solution so important wasn't the answer. It was the *method*. Even though Wiles worked alone for seven years, in secret,[3] he built on the collaborative efforts and discoveries of a slowly evolving group of like-minded colleagues. Starting from crucial insights by Yves Hellegouarch, Gerhard Frey, and Ken Ribet, Wiles linked Fermat's problem to a rich and deep area of mathematics: elliptic curves and modular functions. (You don't need to know what those are, but see this note.[4]) Nothing excites mathematicians more than a new method, especially if it's linked to powerful techniques that have already been developed. You can explore a new territory using tried-and-tested methods. It's like finding important new uses for an existing toolkit: you can jump right in without leaving your comfort zone.

And that's exactly what happened. I'm not trying to suggest that further exploration of Wiles's methods was easy. Many vital questions remain unanswered, even now. But progress was rapid, despite all that. For example, the deepest result used in Wiles's proof was limited to a special type of elliptic curve, leaving an obvious question: is this limitation necessary? Within six years the experts had shown that it wasn't. First Fred Diamond made a significant advance, then Brian Conrad and Taylor joined in to extend the ideas, and finally Christophe Breuil completed both the collaboration and the proof.

This is how mathematicians react to breakthroughs. 'Great, well done.' Pause. '*Can we reach further?*'

Usually, we can.

. . .

Spectacular breakthroughs aren't the only way to advance mathematics. They're very rare – though also very welcome – and most advances in the subject are incremental. Successive attacks chip little bits and pieces off the problem. Sometimes this process ends up solving it completely; often it grinds to a halt when the methods being used start to get so complicated and cumbersome that it becomes apparent that this isn't the right way to proceed. But incremental advances have their own value too. They provide psychological reinforcement that the problem will eventually be solved, and they add evidence in favour of conjectures. When you're struggling to see your way through a difficult mathematical problem it's easy to become disillusioned or just worn out. A bit of encouragement can go a long way. Moreover, each tiny step forward adds to the common toolkit. You never know what might prove useful in the future.

In the 1770 edition of his *Meditationes Algebraicae*, the English mathematician Edward Waring wrote (in Latin): 'Every integer is a cube or the sum of two, three, 4, 5, 6, 7, 8, or nine cubes; every integer is also the square of a square, or the sum of up to nineteen such; and so on.' Squares, you'll recall, are obtained when we multiply a number by itself, so the square of 2 is 4, the square of 3 is 9, and so on. Cubes arise when we multiply three copies of a number together: the cube of 2 is 8, the cube of 3 is 27, and so on. The 'square of a square' is the fourth power.

Where did Waring get this curious problem from? He doesn't say, but it seems pretty clear that he'd been exploring a natural generalisation of a beautiful (and difficult) theorem of Joseph-Louis Lagrange: every (positive) integer is a sum of four squares (including zero if required). For example, $2{,}025 = 42^2 + 14^2 + 8^2 + 1^2$. The question goes back to the Greek Diophantus, who wrote a book on number theory around AD 250; Lagrange answered it in 1770. Unsatisfied, we ask whether the number 'four' can be reduced to 'three'. The answer is 'no', and the proof is easy, because the only ways to write 7 as a sum of squares (excluding zeros) are

$$7 = 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 = 2^2 + 1^2 + 1^2 + 1^2$$

requiring either seven squares or four. So the number of squares in the second sum, namely four, is 'best possible'.

'Great, well done… *Can we reach further?*'

Not if we stick to the original question. But we can change the question. Mathematicians do this a lot when they get stuck, or when they've solved one problem and are looking for another one. With amazing rapidity, given how slow communications were in 1770, Waring promptly wondered what would happen with cubes or fourth powers in place of squares. What led him to formulate his conjecture? He presumably did lots of 'experimental' calculations. For example it's not hard to see that 23 requires nine cubes:

$$23 = 2^3 + 2^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3$$

So the answer is at least nine cubes. Further experiments fail to uncover anything that needs ten, so presumably the answer should be nine. Similar experiments with fairly small numbers show that 79 needs nineteen fourth powers, and no bigger number seems to need more.

Waring didn't prove anything, and 139 years passed before David Hilbert proved that whichever n we choose, every positive integer is a sum of some specific finite number of nth powers. 'Great, David. Well done… *Can we reach further?*'

Of course. A better theorem gets us closer to the extreme; the best theorem is the one that reaches it. Here, the extreme that we should seek is the *smallest* number that works, because that gives the strongest possible result. Eleven years later, in 1920, G. H. (Godfrey Harold) Hardy and J. E. (John Edensor) Littlewood came up with a completely new method to prove that nineteen fourth powers always work. They used complex analysis – calculus with numbers involving the square root of minus one – to express Waring's problem in terms of certain infinite series, and then wielded the existing toolkit in that area, together with some new devices of their own, to pull the series apart into pieces they could understand. Waring's problem, they realised, isn't really about number theory. It just looks like it. They also suggested modifying the question. With cubes, for instance, the *only* numbers that need nine of them are 23 and 239, as proved by Leonard Dickson in 1939. Any 'sufficiently large' number needs at most eight cubes. So Hardy and Littlewood introduced a subtler number, $G(n)$: the smallest number of nth powers needed to represent *all but a finite number* of positive integers as a sum of nth powers. The best result here would be an exact formula for $G(n)$, but that's currently out of reach. Instead, progress has been limited to small values of n, placing increasingly stringent limits on the size of $G(n)$. For instance, between 1925 and 1995, seventy years of snail-paced progress[5] reduced the smallest known value for $G(5)$ – fifth powers – from 41 to 17. But there's probably some way to go yet: the conjectured value is 6.

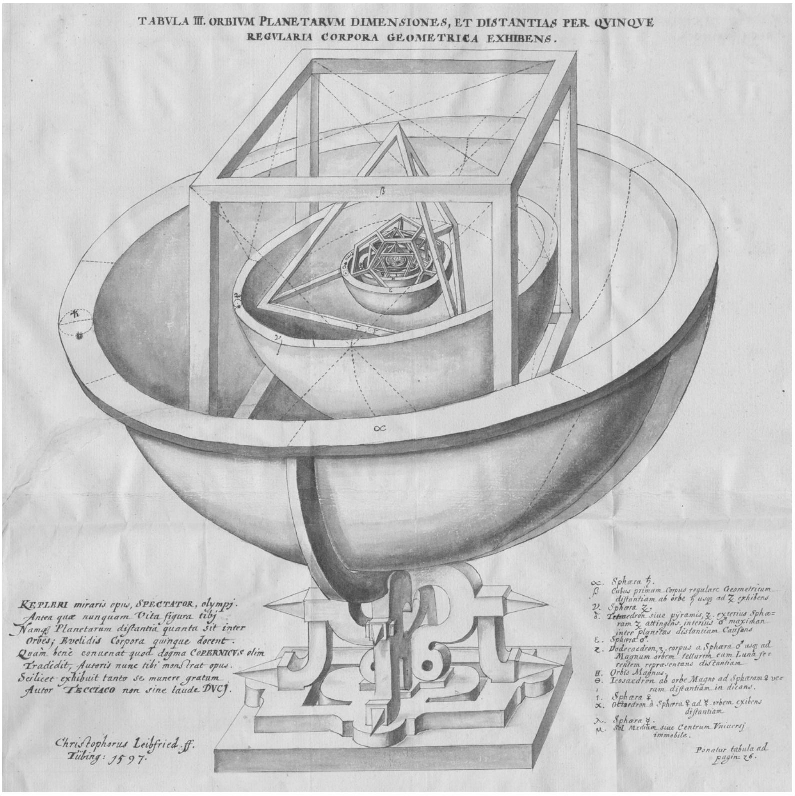
Does it *matter* whether the answer here is 17 or 6? Or something in between? Nothing of practical importance depends on it. The answer won't change the world. But it does matter that *we don't know*. And we have no idea what might emerge from the quest to find out, even if it doesn't succeed.

. . .

New mathematics stems from three distinct sources.

The first, and most obvious, is the desire to understand the natural world. Humans are pattern-seeking animals, finding apparent significance in random alignments such as tea leaves in a cup or a scatter of stars in the night sky. The constellations are a good example. Cygnus looks like a swan, flying along the Milky Way, but the component stars are at different distances and it wouldn't look remotely like a swan when viewed from another direction. Even accidental patterns such as this have their uses: they helped ancient navigators to remember the night sky and identify key stars, for instance. But accidental patterns don't offer deep insights into the natural world. Mathematics seeks – and finds – *significant* patterns. They can be visual, numerical, structural, conceptual or so esoteric that only a mathematician would consider them worthy of the name. The quest for extremes focuses attention on those patterns that give the deepest insights into nature.

Johannes Kepler, a German polymath who flourished around 1600, analysed decades of observations of the planet Mars, trying to find patterns in its apparently erratic motion across the sky. He eventually extracted three simple patterns, now called Kepler's laws of planetary motion. None of them can be detected by a glance at



Kepler's elegant but flawed model of the solar system,
from *Mysterium Cosmographicum*.

the night sky; they emerged from sophisticated analysis of observational data. The best known of these is: planets orbit the Sun in ellipses (a geometric pattern). The other two relate the size of that ellipse to the planet's orbital period (a numerical pattern) and describe how its speed changes as the planet repeatedly cycles round the Sun – faster near the Sun, slower in the orbit's distant reaches (both numerical *and* geometric).

Continuing the mathematician's natural urge to seek extremes, Kepler tried to explain not just the shapes of planetary orbits, but their sizes. He thought that he'd found another mathematical pattern, a link between the six then-known planets and the five regular solids known to Euclid. In the end, this particular pattern proved no more illuminating than Cygnus's resemblance to a swan, if only because more planets were found.[6] Later, Isaac Newton reformulated Kepler's three laws in terms of a mysterious force: gravity. Albert Einstein's subsequent improvement to general relativity notwithstanding, we still use Newton's laws of motion and gravity today to figure out how to put satellites in orbit, direct probes to far planets and put humans on the Moon. These laws let us predict the positions of the planets millions of years ahead (subject to some issues with chaotic dynamics), but we still have very little idea why the orbits are spaced the way they are. Most known exoplanet systems, round other stars, have very different spacings.

The second source of new ideas is the equally practical – indeed, often more practical – demands of human affairs. Taxation, ownership, trade, money. Buildings, transportation. Healthcare. Politics, even: some mathematicians are currently paying a lot of attention to making democratic voting systems fair and detecting gerrymandering; that is, drawing political boundaries to distort voting patterns. The aim is to understand the human world rather than the natural one. Mathematicians mainly use the same approach that they use for the natural world: seek patterns, understand them, exploit them. But they apply those ideas in a different context. A well-developed sense of structure, of how to formulate a question in an insightful way, is vital here. Opportunity can then be translated into action.

The third source of inspiration is far more mysterious: the internal musings of the human mind. Monkey curiosity, if you wish – though a very focused kind of curiosity. Human beings seem to have an innate fascination with patterns. Patterns of numbers, patterns of shapes. Kepler's ellipse was known to the geometers of ancient Greece. Some of their interest in geometry was practical; it's very useful for land measurement and surveying, for instance. With simple tools from geometry you can build a tunnel through a mountain by starting at both ends and meeting in the middle, thereby halving the time needed. Which they did. You can also use geometry to find the shape and size of the Earth, and the distance to the Moon and the Sun; not quite as practical as building a tunnel, but at least related to the real world. The Greeks did those things too. But a major source of inspiration was the urge to understand geometry for its own sake; to prove the strongest, most insightful theorems, whether they have practical implications or not.

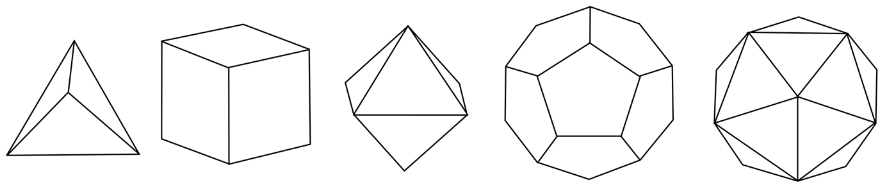
Extreme geometry.

Arguably, I should add a fourth source: the existing body of mathematics. There are always gaps in our knowledge, which mathematicians inevitably see as opportunities to fill. Mathematics seems to have a life of its own, an innate compulsion to extend its empire. Of course it's actually the collective minds of mathematicians that give this impression. I'll come back to this issue in Chapter 18. For now, suffice it to say that this fourth source ultimately traces back to the previous three.

· · ·

The most fascinating feature of these three very different kinds of motivation is that they reinforce each other, combining to create something greater than any could manage alone. Most of the Greeks' interest in geometry wasn't practical at all. Euclid's *Elements*, one of the crowning glories of Greek geometry and the only one to be inflicted on generations of innocent schoolchildren in Victorian times, is firmly focused on the logic of mathematical deductions. *How do we know that a mathematical statement is true?* We can check the validity of assertions about the real world empirically. Make measurements, do experiments, compare the results with what theory predicts. This is the essence of the scientific method. Some of these things can be done in mathematics. Draw lots of triangles, see if their angles add up to 180°. But that kind of empirical test doesn't satisfy the mathematical urge. It tells us (within a certain level of experimental error) *that* something is probably true. But it doesn't tell us *why* it's true. Or even whether it's true.

Physicists, measuring (say) the speed of light, will be satisfied with an accurate answer. They'll be even happier with a *more* accurate



The five regular polyhedra: tetrahedron, cube, octahedron, dodecahedron and icosahedron.

answer. But they won't spend a great deal of effort wondering *why* that particular speed occurs – unless, perhaps, they have a philosophical turn of mind, and work in an area like cosmology, where such questions are considered to be worth pursuing. Mathematicians, in contrast, are much less interested in what the answer is than in why it's true. Yes, of course it's nice to know that the angles of a triangle add up to 180°, assuming that experimental errors don't cause any trouble. But – well, maybe it's really 180.00001°. Or it varies between 179.99999° and 180.00001°, depending on the triangle. Experiments won't distinguish that from the neat, tidy 180° unless they're extraordinarily precise, and accurate to boot (which is different). So a mathematician naturally wants to know whether the angle sum really is 180°, and if so, why that happens.

Euclid's answer to this was: *prove it*. State up front the assumptions you're going to make. (Nearly all of these were apparently obvious things like 'all right angles are equal'.) Then provide a logical argument that leads from those assumptions to the statement you wish to prove. In the thirteen books of the *Elements*, Euclid manages to go from his basic assumptions – today we call them 'axioms' – via a long series of theorems, some remarkable in their own right, to the existence and construction of the five regular solids: tetrahedron, cube, octahedron, dodecahedron and icosahedron. Of which more later.

Euclid's focus was on the pursuit of truth by logical means, and his systematic and precise approach to logic is another case of reaching for extremes. He was aiming not just at finding a plausible argument, but at finding one that could survive the most stringent criticism. His work wasn't perfect; in particular, he tacitly made assumptions that hadn't been stated up front.[7] But he still set a standard that was unsurpassed for two thousand years.

Euclid's motivation wasn't practical in any meaningful sense. There's no mention of real-world applications in the *Elements*. Indeed, there are few applications of a regular dodecahedron, and even if there were, you wouldn't need the *Elements* to make one.[8] But it often takes a long time for new mathematical concepts to pay off in the 'real world'. A few decades is common; it can take millennia. In the fullness of time, Euclid's ideas paid off handsomely. Apollonius of Perga extended Euclid's work in his *Conics*, which investigated the geometry of sections of a cone, distinguishing three types of 'conic section': ellipse, parabola and hyperbola. He established dozens of basic facts about their geometry. He had no particular applications of the conic sections in mind: he studied them because he thought they were interesting, and because they went beyond the lines, circles and spheres of Euclid. Mainly, I suspect, because he could prove things about them that weren't obvious. The fun of research. Others had been investigating these curves as well, and they were known to have some curious mathematical properties, leading to solutions of famous problems that seemed insoluble by Euclidean methods – such as trisecting an angle (splitting it into three equal parts) and duplicating the cube (constructing a cube with twice the volume of a given one). But these 'applications' of conics weren't terribly practical. They were just interesting.

And yet… some fifteen centuries later, it was Apollonius's ellipse that led Kepler to his three laws, leading Newton to this theory of gravity. Which, after another three centuries, were vital for the Moon landings, and for every artificial satellite or space probe ever launched.

(continued...)