Squaring the circle and other impossibilities Transcript

Date: Wednesday, 16 January 2008 - 12:00AM

Location: Barnard's Inn Hall

SQUARING THE CIRCLE AND

OTHER IMPOSSIBILITIES

Professor Robin Wilson

In my final set of Gresham lectures I'd like to consider a wide range of mathematical problems. It is often said that solving problems is the life-blood of mathematics: it is by doing so, rather than building up elaborate theories, that the subject stays alive and vigorous — at least that was the view of David Hilbert, whose 23 celebrated problems I'll survey in my third lecture. In my next lecture, I'll be looking at a wide range of mathematical puzzles covering a thousand years or more. But today, I'm going back two-and-a-half millennia, to just three problems, the so-called 'Three Classical Problems of Ancient Greece' — *doubling the cube*, *trisecting the angle* and (most famous of all) *squaring the circle*.

These problems fascinated the Greeks, who suggested many ingenious lines of attack, but they were unable to solve them. Succeeding generations tried to settle them once and for all, but again in vain. It was not until the 19th century, with the development of algebra by Galois and others, that these three problems were finally *proved* to be impossible — indeed, the phrase 'squaring the circle' is widely used for to describe a task that is impossible to carry out.

So what exactly do these problems ask? The first, that of *doubling the cube*, asks whether, given a cube of given size, we can construct another cube with exactly double the volume. The second, *trisecting the angle*, asks whether, given an arbitrary angle, we can divide it into three equal parts. And the third problem, *squaring the circle*, asks whether, given a circle of given size, we can construct a square enclosing exactly the same area. All three problems date from about the fifth century BC.

In this talk I'll first set the scene by explaining what the Greeks meant by 'construct', with a number of simple examples. Then I'll take each problem in turn and describe how they, and succeeding generations of mathematicians, tried to solve it.

Early Greek mathematics

The earliest Greek mathematician that we know anything about (and we don't know much) is Thales of Miletus, who lived around 600 BC, about fifty years before Pythagoras. Rigorous deductive proof is our gift from the Greeks, and Thales is said to have been the first to prove that circles are bisected by their diameters.

Thales is mentioned in the following extract from Aristophanes' comedy *The Birds*, of about 414 BC, which show how the problem of squaring the circle had already become part of every-day parlance. Enter Meton, an enthusiastic astronomer:

METON: These are my special rods for measuring the air. You see, the air is shaped — how shall I put it? — like a sort of extinguisher; so all I have to do is to attach this flexible rod at the upper extremity, take the compasses, insert the point here, and — you see what I mean?

PEISTHETAERUS: No.

METON: Well, I now apply the straight rod — so — thus squaring the circle; and there you are . . .

PEISTHETAERUS: Brilliant — the man's a Thales.

We need to explain what we mean by 'construct'. The first three postulates at the beginning of Euclid's *Elements* enable us to *draw a straight-line segment between any two points*; *extend a straight-line segment*; and *draw a circle with any given centre and radius*. These are the only constructions we allow. We call them *line-and-circle constructions*, or *ruler-and-compass constructions*, because they can be carried out using a ruler and a pair of compasses, though the term *straight-edge* is often used instead of *ruler*, to emphasise that we do not allow any measuring.

Euclid's *Elements*, from about 300 BC, contains many such constructions. The first proposition of the *Elements* (Book I, Proposition 1) shows us how to *construct an equilateral triangle on a given line segment AB*, using ruler and compasses. We simply draw two circles, with centres *A* and *B* and radius *AB*. These circles meet at two further points *C* and *D*, and joining either of them to *A* and *B* gives the equilateral triangle. Moreover, by joining these points *C* and *D*, we can construct the mid-point of *AB*, thereby demonstrating how to *bisect any given line segment*. There are also simple constructions for *trisecting a line segment*, or more generally, dividing it into any desired number of equal parts.

Doubling the cube

Another simple construction, which uses only a straight-edge and no compasses, is to *double the area of a given square*. This example appears in Plato's book *Meno*, in a dialogue between Socrates and a slave boy. The slave boy first guesses that we need to double the length of the side to double the area — but quickly realises that this gives four times the area. Eventually, he is led to the idea of drawing the square whose side is a *diagonal* of the original square. If the side of the original square is 2 and the area is 4, then the length of the diagonal is $2\sqrt{2}$ and the area of the new square is $2\sqrt{2} \times 2\sqrt{2} = 8$, as required.

Before turning to our first classical problem, that of doubling the cube, we introduce the idea of a 'mean proportional' of two numbers (or, to the Greeks, *lines*). Given numbers a and b, their *mean proportional* is the number x such that the ratios $a : x$ and $x : b$ are the same – for example, $4 : 6 = 6 : 9$, so 6 is the mean proportional of 4 and 9.

In Book IV of the *Elements*, Euclid gives a ruler-and-compasses construction for finding the mean proportional of a and b. Note that, since $x/a = b/x$, we have $x = \sqrt{(ab)}$. Since $x^2 = a \times b$, finding the mean proportional of a and b amounts to 'squaring' the rectangle with sides a and b – that is, finding a square with the same area as the rectangle.

Let us now turn to the problem of doubling the cube. We have two descriptions of how the problem might have arisen. The first comes from Theon of Smyrna (2nd century AD):

In his work entitled Platonicus*, Eratosthenes says that, when the god announced to the Delians* (the citizens of Delos) *by oracle that to get rid of a plague they must construct an altar double of the existing one, their craftsmen fell into great perplexity in trying to find how a solid could be made double of another solid, and they went to ask Plato about it. He told them that the god had given this oracle, not because he wanted an altar of double the size, but because he wished, in setting this task before them, to reproach the Greeks for their neglect of mathematics and their contempt for geometry.*

The second is later, from Eutocius in the 6th century AD:

The story goes that one of the ancient tragic poets represented Minos having a tomb built for Glaucus, and that when Minos found that the tomb measured a hundred feet on every side, he said: 'Too small is the tomb you have marked out as the royal resting place. Let it be twice as large. Without spoiling the form quickly double each side of the tomb.' This was clearly a mistake. For if the sides are doubled, the surface is multiplied four-fold and the volume eight-fold. Now geometers, too, sought a way to double the given solid without altering its form. This problem came to be known as the duplication of the cube*, for, given a cube, they sought to double it.*

The Greeks found themselves unable to solve this problem with ruler-and compasses. According to Eutocius, Hippocrates of Chios (from about 440 BC) found that we can double a cube by first finding *two mean proportionals* between given lines (or numbers), one twice the other — that is, if a and b ($= 2a$) are the given lengths, we must find new numbers x and y such that the ratios $a : x$, $x : y$ and $y : b$ are all equal. But if $x/a = y/x = b/y$, then $(x/a)^3 = (x/a) \times (y/x) \times (b/y) = b/a = 2$, and so $x^3 = 2a^3$. This corresponds to starting with a cube of side a and replacing it by a cube of side x, thereby doubling its volume.

The Greek mathematician Menaechmus, who is credited with introducing the *conic sections* — the parabola, ellipse and hyperbola, showed how conics arise in connection with mean proportionals, and thus with doubling the cube. From the equation $x/a = y/x$ we deduce that $ay = x^2$, which is a parabola; from the equation $y/x = b/y$ we deduce that $y^2 = bx$, which is another parabola; and from the equation $x/a = b/y$, we deduce that $xy = ab$, which is a hyperbola. Thus, we can double the cube by drawing these conics and finding where they intersect — but we can't do this with ruler and compasses only.

Over the centuries, many other methods were devised for doubling a cube — both accurate and approximate — but none of them solved the problem.

Trisecting an angle

We next turn to the problem of trisecting an angle, but first let's see how to *bisect* it. If PAQ is the given angle, we draw a circle with centre A and any radius; this crosses AP and AQ at points B and C. Draw the line segment BC, and find its perpendicular bisector. This passes through A, bisecting the angle, as required.

When we come to the problem of *trisecting the angle*, it is important to realise that we are talking about an *arbitrary* angle, not a specific one. Certain angles, such as 90°, can readily be trisected, as demonstrated by the young Charles Dodgson (Lewis Carroll) at the age of only 12: essentially, we construct an equilateral triangle and thus an angle of 60°, which we then bisect to give an angle of 30°, which is one-third of 90°, as required.

One method of trisecting an angle, described by Hippias, was to introduce an auxiliary curve, called a *quadratrix* or *trisectrix*. Starting with a square, we place an object at the top-left hand corner and let it sweep out a quarter-circle, while the top of the square simultaneously moves downwards at a constant rate. The resulting path of the object is the quadratrix.

Now take the angle between the base of the square and a line from the bottom left-hand corner to the object at a given time. Using a simple argument based on the geometry of the quadratrix, we can trisect this angle — or indeed, divide it into any required number of parts. However, we cannot construct the quadratrix with ruler and compasses, so this is not a valid method.

Constructing regular polygons

As a change from the three classical problems, let us look briefly at the construction of regular polygons using only straight-edge and compasses. We have already seen how to construct an equilateral triangle, and constructing a square on a given line segment *AB* is almost as simple: we erect perpendiculars from *A* and *B*, and then use the compasses (with its point at *A* and *B* and radius *AB*) to locate the top two corners of the square. Similarly, we can construct a regular hexagon by starting with a triangle and bisecting each edge in turn. In Book IV of the *Elements*, Euclid also showed how to construct a regular pentagon and a regular 15-sided polygon using only straight-edge and compasses.

Not all polygons can be constructed — for example, it is impossible to construct a regular 7-gon, 9-gon or 11-gon with ruler and compasses. So, which regular polygons can be constructed? This question was not solved until the 1790s, when Carl Friedrich Gauss determined exactly when it can be done.

His answer involves the *Fermat primes* — these are prime numbers of the form $2^k + 1$, where k is itself a power of 2: the only ones known are $2^1 + 1 = 3$, $2^2 + 1 = 5$, $2^4 + 1 = 17$, $2^8 + 1 = 257$ and $2^{16} + 1 = 65,537$. Gauss proved that an n-sided regular polygon can be constructed by ruler and compasses if and only if n is equal to a power of 2, possibly multiplied by distinct Fermat primes; for example, one can construct a regular 60-sided polygon (since $60 = 2^2 \times 3 \times 5$) and a regular 170-sided polygon (since $170 = 2 \times 5 \times 17$). These are the only possibilities.

Squaring the circle

In Euclid's *Elements* I, 44 we learn that, given any triangle, we can construct a parallelogram of equal area. More generally, in *Elements* I, 45, we see that a similar statement applies also to a polygon with any number of sides. In *Elements* II, 14, we find that, given any polygon, we can construct a square of equal area. Thus, *we can square any polygon*.

But it is not only polygons that can be squared — we can also square certain regions with curved boundaries. More than a century before Euclid, Hippocrates of Chios gave a construction for squaring 'lunes' — the moon-shaped regions lying between two circles. And later, Archimedes showed that the area of a segment of a parabola is equal to $^4/_3$ times the area of an enclosed triangle (which can then be squared), thus providing the quadrature of a parabola.

Squaring the circle is much more difficult. Some people tried to do it using the quadratrix. Others, such as Antiphon and Bryson, tried to approximate the circle by regular polygons with ever-increasing numbers of sides: their idea was that since we can square any polygons, we can 'take the limit' and hence square the circle.

By considering polygons with 96 sides lying inside and outside a given circle, Archimedes had found that the value of π lies between $3^{10}/_{71}$ and $3^1/_7$, and later geometers improved this value by extended the method to polygons with many millions of sides. But however far we go, we never actually get the circle itself — we can indeed square all these polygons, but it does not follow that we can square the circle itself.

The middle period

For many years mathematicians tried to solve the three classical problems, using more and more ingenious methods. In particular, there was a brief flurry of interest in the 17th century, when Descartes, John Wallis and others tried to solve the problem, but progress was limited. It had to wait until the 19th century until the breakthrough could occur, with the algebraic ideas of Galois and others. Today, we usually prove the impossibility of solving the three problems by resorting to a branch of mathematics called Galois theory.

Some algebra

Starting with a unit length, we can use a ruler and compasses to mark out any of its whole number multiples, and we can also create any length that is a fraction — in other words, we can construct all the rational numbers. Can we construct any other numbers? More geometrically, while carrying out our constructions, points arise as the intersection of two lines, as the intersection of a line and a circle, and as the intersection of two circles. What are the coordinates of these points?

Suppose we take two lines whose coefficients are integers or fractions. Is the same true of their point of intersection? This is certainly true for the lines with equations $y = 2x + 3$ and $y = 5x – 1$, whose point of intersection is $x = {}^{4}/_{3}$, $y = {}^{17}/_{3}$, as we find on solving these equations. More generally, starting with two lines with equations $y = ax + b$ and $y = cx + d$, where a, b, c and d are fractions, their point of intersection always has the form (x, y), where x and y are also fractions.

Now let's look at the intersection of a line and a circle. If we take the line $x = 2$ and the circle $(x – 1)^2 + y^2 = 4$, we find that the points of intersection are $(2, \sqrt{3})$ and $(2, –\sqrt{3})$, so we now have to introduce square roots. More generally, if we take any line and circle with fractional coefficients, then we find the coordinates of the points of intersection by solving a quadratic equation, so again it will usually involve square roots.

Finally, what happens if we intersect any two circles? We again get integers, fractions and square roots, and nothing else besides.

It follows that whenever we carry out a ruler and compass construction, the numbers that arise as coefficients are always made up from adding, subtracting, multiplying, dividing and taking square roots. We say that such numbers are *constructible* — they are the only ones that can arise when we carry out our constructions. For example, when we double a square of side 2 and area 4, we are seeking a length x that satisfies $x^2 = 8$ — so $x = 2\sqrt{2}$, a number that can indeed be constructed.

We can now explain in general terms why the three classical problems cannot be solved — it is because their solutions would involve numbers that are not constructible by the above means. Let us look briefly at them in turn.

In order to *double a cube* (of side 1, say), we need to construct the number x that satisfies $x^3 = 2$. This number is $x = \sqrt[3]{2}$, a cube root that clearly cannot be constructed by repeatedly adding, subtracting, multiplying, dividing and taking square roots. So doubling the cube is impossible.

In order to *trisect an angle*, such as 60°, we use the trigonometrical formula

$$\cos 3\theta = 4 \cos^3 \theta – 3 \cos \theta.$$

Putting $x = \cos 20°$, the number we want to construct, and using the fact that $\cos 60° = ½$, we see that x satisfies the equation $½ = 4x^3 – 3x$, or $8x^3 – 6x – 1 = 0$. If we can factorise this into a linear and a quadratic factor (so that we get square roots at worst), then one factor must be $x \pm 1$, $2x \pm 1$, $4x \pm 1$ or $8x \pm 1$ — but none of these works. So we cannot split the equation up as we wanted: the solutions, in fact, involve cube roots. So trisecting this particular angle, and hence angles in general, is impossible.

Finally, in order to square a circle, we need to prove that the number π is not constructible — that is, it cannot be obtained by adding, subtracting, multiplying, dividing, and the taking of square roots. Although this seems obvious, it is difficult to prove, and was not obtained until 1882, by F. Lindemann. Thus, squaring the circle is also impossible.