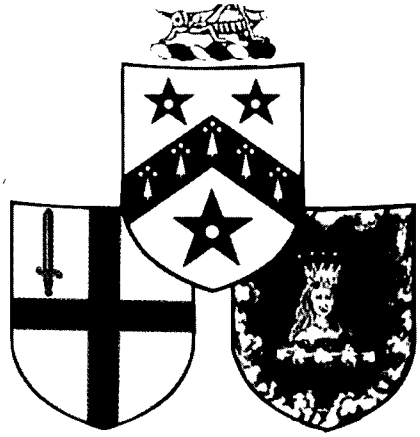
G R E S H A M

COLLEGE



QUEEN DIDO'S HIDE

A Lecture by

PROFESSOR IAN STEWART MA PhD FIMA CMath
Gresham Professor of Geometry

10 March 1998

GRESHAM COLLEGE

Policy & Objectives

An independently funded educational institution, Gresham College exists

- to continue the free public lectures which have been given for 400 years, and to reinterpret the 'new learning' of Sir Thomas Gresham's day in contemporary terms;

- to engage in study, teaching and research, particularly in those disciplines represented by the Gresham Professors;

- to foster academic consideration of contemporary problems;

- to challenge those who live or work in the City of London to engage in intellectual debate on those subjects in which the City has a proper concern; and to provide a window on the City for learned societies, both national and international.

Gresham College, Barnard's Inn Hall, Holborn, London EC1N 2HH
Tel: 020 7831 0575 Fax: 020 7831 5208
e-mail: enquiries@gresham.ac.uk

Gresham Lecture

Queen Dido's Hide

Ian Stewart 10 March 1998

Why are soap bubbles spherical? The energy in a film of soap depends on its area. The smaller the area, the smaller the energy. Nature is fundamentally lazy, and does everything using the least energy possible. So a soap film always has the smallest area possible — consistent with doing its job. The job of a bubble is to contain a given quantity of air. The surface of smallest area that contains a given quantity of air is a sphere. That's why soap bubbles are round. But *why* is the surface of smallest area that contains a given quantity of air a sphere? That's a question for mathematics.

Mathematics can help us find which thing of a given kind is the longest, the shortest, the best, the biggest, the smallest, or the cheapest. What shape should a piece of card be in order to make the largest box? You can imagine that a company that sells groceries, say, would be interested in such questions. Similarly, an airline company wishing to run services between a number of cities would be interested in finding out how to schedule the flights so as to maximise their profits. Problems of this type come under the general heading of 'optimization' — finding the best solution.

Dido's Hide

The earliest recorded example of a mathematical solution to an optimization problem is the ancient Greek legend of Queen Dido. Dido, it is said, was given a bull hide and told that she could take possession of whatever land she could enclose with it. She cut the hide into an enormously long, thin strip, and arranged it in a huge circle, thereby enclosing the largest possible area of land. On it, she founded the city of Carthage.

Dido found the answer to the two-dimensional version of the soap bubble problem — to enclose a given area with the shortest curve. Our job is to prove that her answer was right. To do the same for a three-dimensional bubble is beyond our powers — it *can* be done, but only with a lot of mathematical technique. Even the two-dimensional version is far from easy.

We'll assume that we are presented with a *fixed* length of hide, and ask whether a circle is indeed the largest area that it can enclose. We model the long, thin hide by a mathematical curve of zero thickness. Now we've got a problem that can be subjected to mathematical analysis. Given a curve of fixed length, what shape should it be to enclose the greatest area?

Existence?

It took mathematicians quite a long time to realise that questions like this come in two parts:

- Show that an answer to the question *exists*,
- Find out what it is.

Answers to mathematical questions do not always exist, even if the question looks reasonable. Here's an example. It is known that the shortest path between two given points (in the plane) is a straight line. We might ask 'what is the shortest *non-straight* path between two given points?' In effect, this asks what the next shortest path, after the straight line, is. However, there is no such beast. Given any path that's not straight, you can always find a shorter path that is also not straight, by taking a short cut across some bend. So the shortest non-straight path *does not exist.*

In Queen Dido's problem, it turns out that a solution does exist. In 1838 the great geometer Jacob Steiner found a beautiful argument to show that *once you know that an answer exists*, you can see that it has to be a circle. His basic idea is that if you take any curve that is not a circle, then you can change it to increase the area that it contains. This is a line of attack that will be familiar to Sherlock Holmes fans: 'Once you have eliminated the impossible, then whatever remains, however improbable, must be the truth.'

Without a proof of existence, Steiner's style of argument can lead to fallacies. Probably the simplest such fallacy is this statement:

- The largest non-zero whole number is 1.

If Steiner's line of argument — ignoring the question of existence — is valid, then we can easily establish that 1 is the largest nonzero whole number. To do so we merely show that, given any non-zero whole number that is not equal to 1, we can find a bigger one. To do this is easy. If you take any non-zero whole number that is not equal to 1 and multiply it by itself, you get something even bigger. *Only* for the number 1 does this step not produce a bigger number, because $1 \times 1 = 1$. Conclusion: any non-zero whole number other than 1 cannot be the biggest non-zero whole number. But does this permit us to conclude that 1 is the biggest? No. All we can legitimately conclude is that *either* 1 is the biggest, *or the biggest does not exist.*

In this case we know, on other grounds, that it is the second statement that is true: there is no biggest non-zero whole number. Proof: if there were such a number, then it would be greater than or equal to any nonzero whole number — for example itself plus one. But *no* number is greater than or equal to itself plus one.

Steiner would have seen the fallacy in this numerical argument, but he didn't seem able to grasp that his own 'proof' for Queen Dido's problem suffered from the same potential fault. The difference, perhaps, was that on this occasion his answer was correct. It is so intuitive that the answer is a circle that it's hard to consider the possibility that no answer exists. But in mathematics, intuition and proof are not the same — and you may get the right answer by faulty reasoning, which is what Steiner did.

Circular Reasoning

Other mathematicians quickly filled the gap in Steiner's proof, however, by showing that in this case an answer does exist. Bearing that in mind, we can now appreciate how clever the rest of Steiner's proof was. It goes like this.

Suppose we have, by some method, laid hands on a curve of the chosen length that really does contain the largest possible area. We will now infer, from the maximal area property, various other features of that curve. The aim is to show that it has to be a circle, and we'll reach that deduction in several easy stages. Here's the first step:

Step 1: The curve is convex.

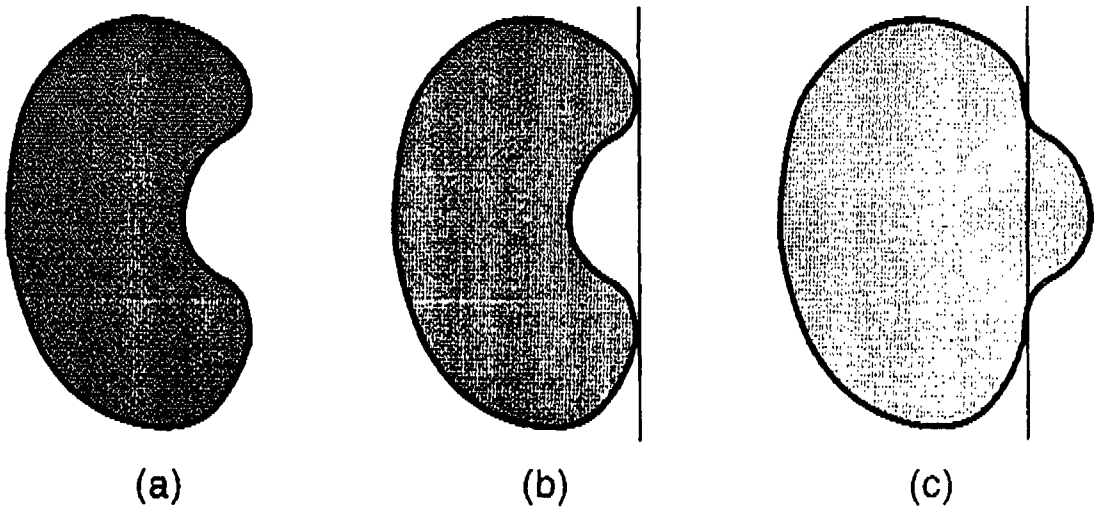


Fig.1 Proof of convexity.

By 'convex' I mean that given any two points inside the curve, the line segment that joins them also lies inside the curve. In other words, there are no 'dents' where the curve

bulges inwards as in **Fig.1a**. Well, suppose there is such a dent. Then we can find a line that touches the curve at two points, as in **Fig.1b**. Then we form a new curve by reflecting part of the old one in a mirror that lies along that line, as in **Fig.1c**. Clearly the resulting curve has the same length as the old one — because the length of the reflected part doesn't change. However, the new curve encloses a larger area — the shaded part in the figure. But the original curve enclosed the *largest* possible area! The only way out of the logical impasse is that no such enlargement is possible. This carries the inevitable conclusion that the original curve must have been convex all along.

An important feature of a convex curve is that if a line cuts across it, then it divides the region inside the curve into exactly *two* parts. Steiner needed this property for his second step. Before describing how that goes, it's useful to have some terminology. Say that a line is a *diameter* of the curve if it divides the perimeter into two equal parts.

Step 2: Every diameter divides the *area* into two equal parts as well.

Suppose that some diameter does *not* divide the area into two equal parts. Then we can take the piece with the larger area (**Fig.2a**), reflect it across the line (**Fig.2b**), and thereby create a new curve (**Fig.2c**) with the same perimeter as before, but larger area. Again, since the original curve enclosed the largest possible area, no such enlargement is possible. Therefore the assumption that some diameter does not divide the area into two equal parts must be false. Therefore every diameter divides the area into two equal parts, as required.

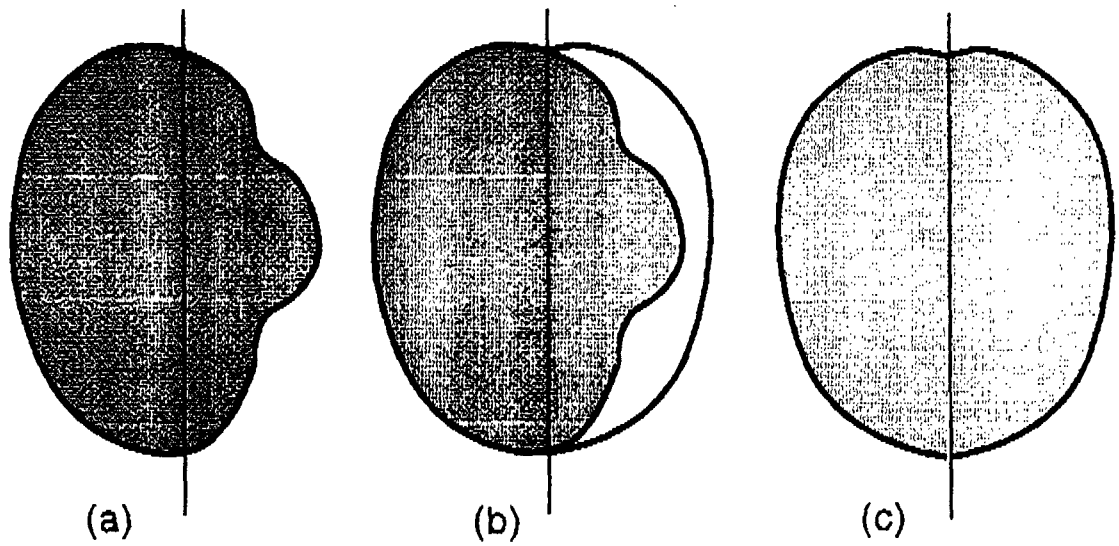


Fig.2 Any diameter must split the area in two.

With this established, we can simplify the problem by looking at just half of the curve. Choose some diamater, and cut the curve in half. Its area also halves. If we can prove that this half-curve must be a semicircle, then it follows — by doing the same for the other half — that the whole curve must be a circle. And that's how Steiner proceeded. First he proved a very neat geometrical property:

Step 3: The angle subtended (in the halved curve) by its diameter is always a right angle.

'Subtended by' means draw a line from each end of the diameter to a point on the curve: then see what angle they meet at (**Fig.3a**). Suppose that the angle subtended by some diameter is not a right angle. The it is either more than a right angle, or less than a right angle. If it is less than a right angle, we can increase the area of the curve by 'spreading the angle out' as in **Fig.3b**. The same goes if the angle is more than a right angle (**Fig.3c**).

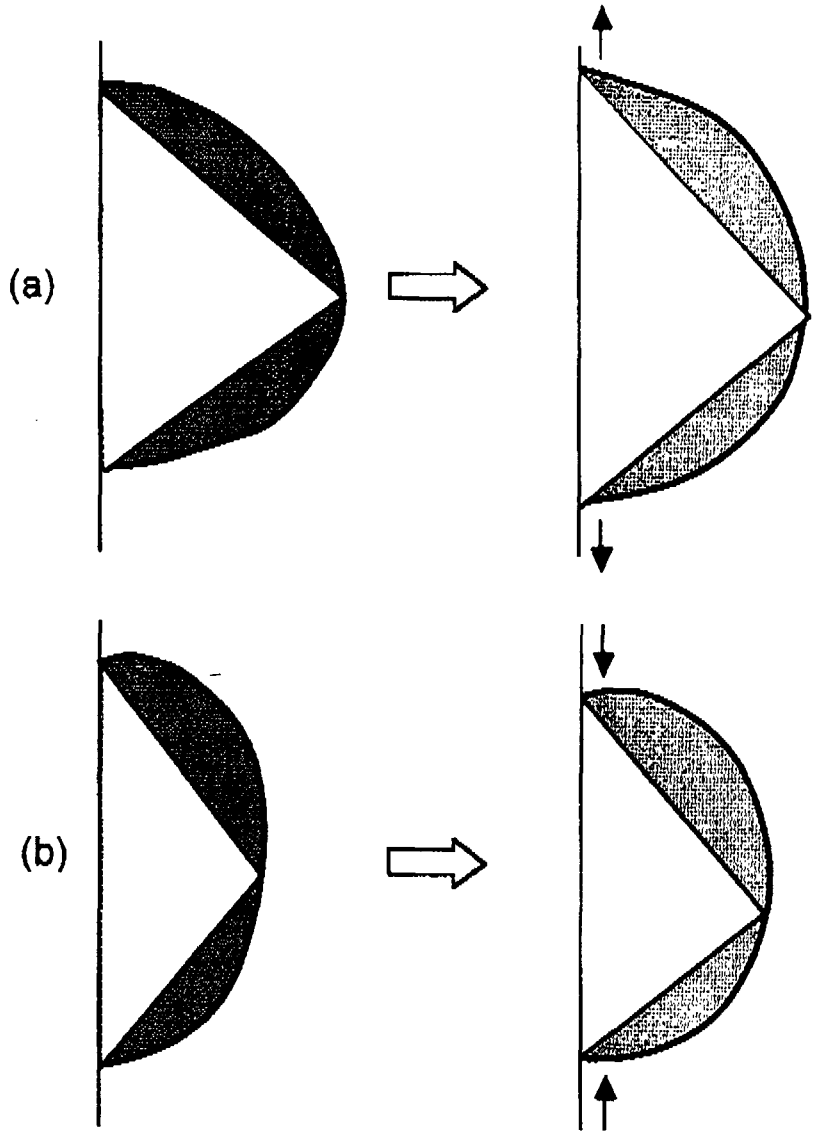


Fig.3 Bending the curve to get a right angle.

TECHNICAL EXTRA

How do we know that these changes to the angle really do increase the area?

The way we change the curve is this: we leave the shaded regions as they were, apart from moving them around a bit, and we open up or squash the white triangle until the top angle is 90°. What I'm claiming is that this angle makes the triangle's area maximal. Notice that when we open up or squash the triangle, the two sloping sides don't change in length: all that changes is the base. So what I'm really claiming is this: *if you are given two sides of a triangle, then its area is greatest when the angle between them is a right angle.*

Fig.4 shows various possible positions for such a triangle, including one where the angle is a right angle, one where it is less, and one where it is more. Now the area of such a triangle is half the base times the height (h). So we want to move the end marked X to the highest possible position. Since X wanders round a circle as we vary the angle, we want to determine the point of the circle that is highest. This is of course the point that lies along a line at right angles to the base of the triangle, and that's exactly what we want to establish.

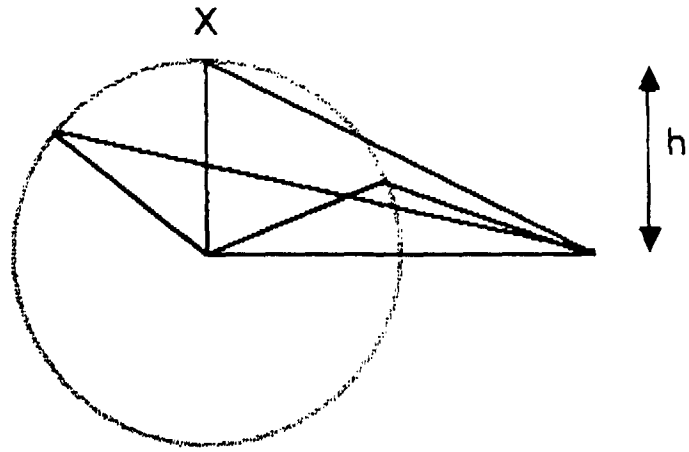


Fig.4 Maximising the area of a triangle.

Step 4: The curve must be a circle.

Choose a diameter, and let its ends be A and B. Choose any point C on the curve. We know that angle ACB is 90°. It is a general theorem in geometry that the angle in a semicircle is 90°. Less well known, but also true, is the converse: if all such angles are 90° then the curve is a semicircle.

TECHNICAL EXTRA

The fact that the angle in a semicircle is 90° is so well known that I won't prove it. Unfortunately, what we need is the converse: if all such angles are 90° then the curve is a semicircle. That is, *if* the curve is a circle, then the part lying above the line AB is a semicircle, and then angle ACB must be 90°. Now this is very similar to what we already know, but the *if...then* is the wrong way round. We know that 'if it's a circle then angle ACB is 90° for any C'. Unfortunately what we have to prove is that 'if angle ACB is 90° for any C then it's a circle.'

That's not the same statement. Compare 'if it's raining, then my garden gets wet' with 'if my garden gets wet, then it's raining.' The first is true. The second could be false — for example, I may be watering the garden with a hose. So the two statements aren't logical equivalents of each other.

To fix things up, we compare our (half) curve with a (semi) circle and show that there's no difference. Suppose, then, that our half curve is not a semicircle. Then we can find a diameter AB and a point C such that C does not lie on the circle with diameter AB. This means that AC cuts the circle at a point D *different from* C. Now, we know that angle ACB is 90° because we've proved that our curve has that property. We *also* know that angle ADB is 90° because that's how circles behave. So the line CB must be parallel to DB, since both cut the same line at right angles. However, the two lines CB and DB meet at B, whereas parallel lines don't meet at all.

There's nothing wrong with the logic, so our initial assumption has to be at fault. What was it? That the half curve isn't a semicircle. Conclusion: actually, it *is* a semicircle.

Step 5: Go for the jugular.

Since each half curve is a semicircle, and they adjoin along a common diameter, the whole curve is a circle.

Done!

Arithmetic and Old Lace

Our next optimisation problem is about shoelaces. There are at least three common ways to lace shoes, shown in **Fig.5**: American zig-zag lacing, the European straight lacing, and quick-action shoe-store lacing. From the point of view of the purchaser, styles of lacing can differ in their aesthetic appeal and in the time required to tie them. From the point of view of the shoe manufacturer, a more pertinent question is which type of lacing requires the shortest — and therefore cheapest — laces. Here I'll side with the shoe manufacturer, but you might care to assign a plausible measure of complexity to the lacing patterns illustrated, and decide which is the simplest to *tie*.

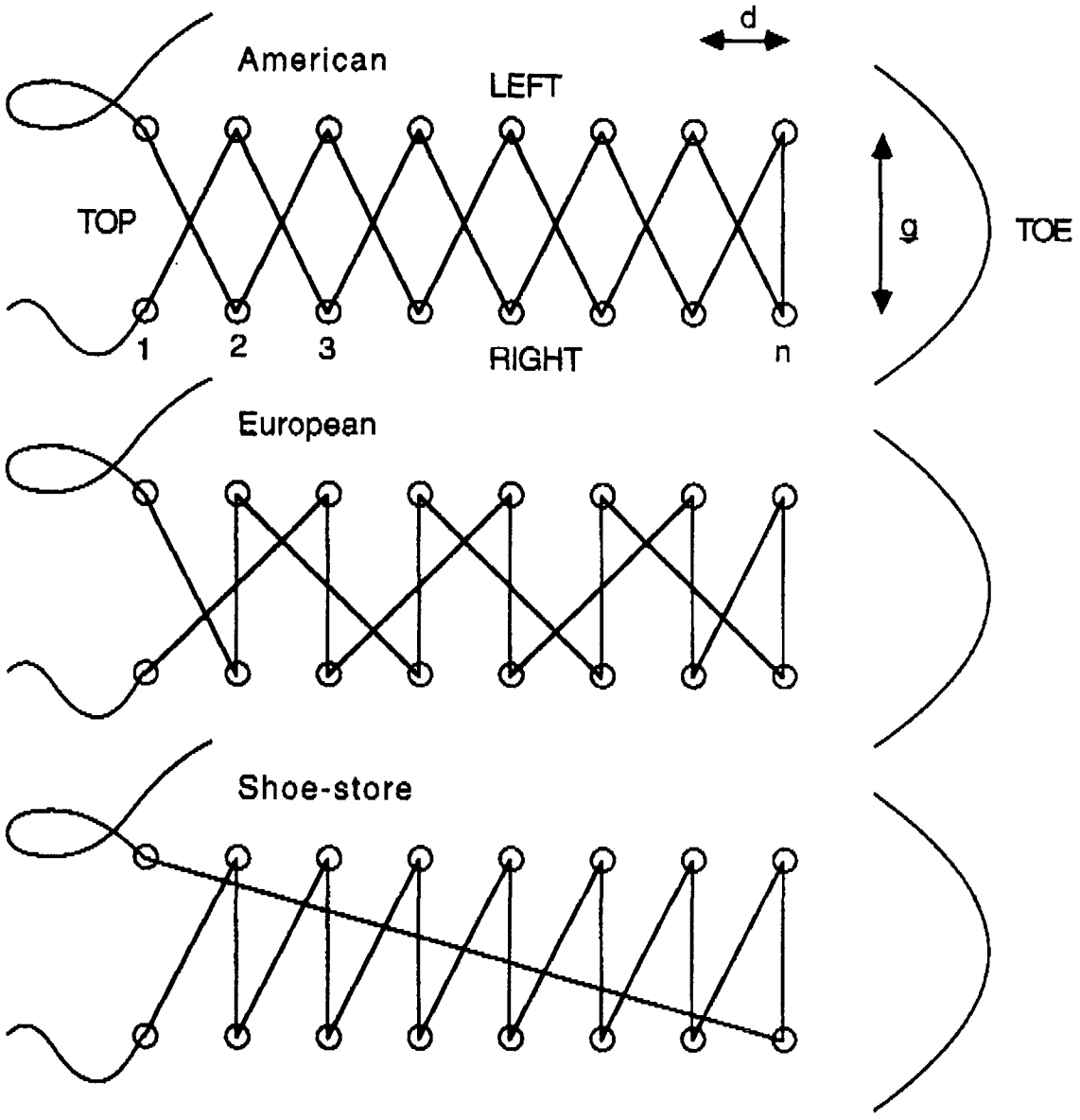


Fig.5 Three types of shoe lacing.

Of course, the shoemaker is not restricted to the three lacing patterns shown, and we can ask a more difficult question: which pattern of lacing, among *all* the possibilities, requires the shortest lace?

To keep the discussion simple, I'm going to assume that the lace moves alternately from the left row of eyelets to the right and back again. Some perfectly practical ways to

lace shoes don't do that, and some of them are shorter than anything I'm going to describe here. I'm choosing my context and I'm sticking to it — and my conclusions will be valid only within that context. I'll focus only on the length of shoelace that lies between the 'top' two eyelets of the shoe, on the left of the diagrams — the part represented by straight line segments. The amount of extra lace required is essentially that needed to tie an effective bow, and is the same for all methods of lacing, so it can be ignored.

My terminology will refer to the lacing as seen by the wearer (hence 'top' just now), so that the upper row of eyelets in the figure lies on the left side of the shoe, and the lower row on the right. I shall also idealise the problem so that the lace is a mathematical line of zero thickness and the eyelets are points. Using a brute force attack, the length of the lace can then be calculated in terms of three parameters of the problem:

- The number n of pairs of eyelets
- The distance d between successive eyelets
- The gap g between corresponding left and right eyelets.

With the aid of Pythagoras's Theorem (one wonders what the great man would have made of this particular application) it is not too hard to calculate the lengths for the lacings in Fig.5. The results are:

American: $g + 2n\sqrt{(d^2+g^2)}$

European: $ng + 2\sqrt{(d^2+g^2)} + (n-1)\sqrt{(4d^2+g^2)}$

shoe-store: $ng + n\sqrt{(d^2+g^2)} + \sqrt{(n^2d^2+g^2)}$.

Suppose, for the sake of argument, that $n = 8$ as in the figure, $d = 1$, and $g = 2$. Then the lengths are:

- *American*: 37.777
- *European*: 40.271
- *shoe-store*: 42.134.

The shortest is American lacing, followed by European, and finally by shoe-store. But can we be certain that this is always the case, or does it depend upon the numbers n, d and g?

Some careful algebra shows that if d and g are nonzero and n is at least 3 then the shortest lacing is always American, followed by European, followed by shoe-store. If $n = 2$ and d and g are nonzero then American is still shortest but European and shoe-store lacings are of equal length. (If $n = 1$, or $d = 0$, or $g = 0$, then all three lacings are equally long, but only a mathematician would worry about such cases!) However, the algebraic approach is complicated, and offers little insight into what makes different lacings more or less efficient.

Instead of usingd algebra, a mathematician called John Halton described a clever geometrical trick which makes it completely obvious that American lacing is the shortest of the three. With a little more work and a variation on that trick it also becomes clear that shoe-store lacing is the longest.

Fermat's Principle

Halton's idea owes its inspiration to optics, the paths traced by rays of light. Mathematicians discovered long ago that many features of the geometry of light rays can be made more transparent — if that is the word to use when discussing light — by applying carefully chosen reflections to straighten out a bent light-path, making comparisons simpler. For example, to derive the classical law of reflection — 'angle of incidence equals angle of reflection' — at a mirror, consider a light ray whose path is composed of two straight segments: one that hits the mirror, and one that bounces off. If you reflect the second half of the path in the mirror (**Fig.6**) then the result is a path that passes through the front of the mirror and enters Alice's mirror-world behind the looking glass. According to the Principle of Least Time, a general property of light rays enunciated some centuries back by Pierre de Fermat, such a path must reach its destination in the shortest time — which in this case implies that it is a straight line. Thus the 'mirror angle' marked in the figure is equal to the angle of incidence — but it is also obviously equal to the angle of reflection.

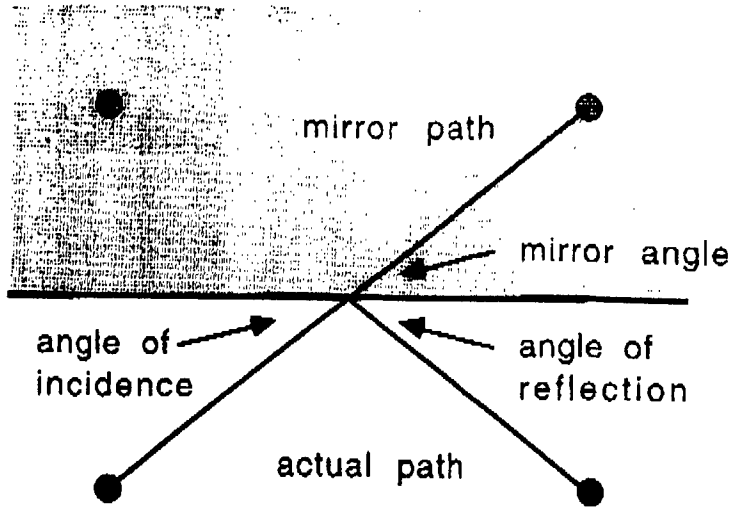


Fig.6 Reflection Principle.

Fig.7 shows geometric representations of all three types of lacing, which Halton derives by an extension of this optical reflection trick.

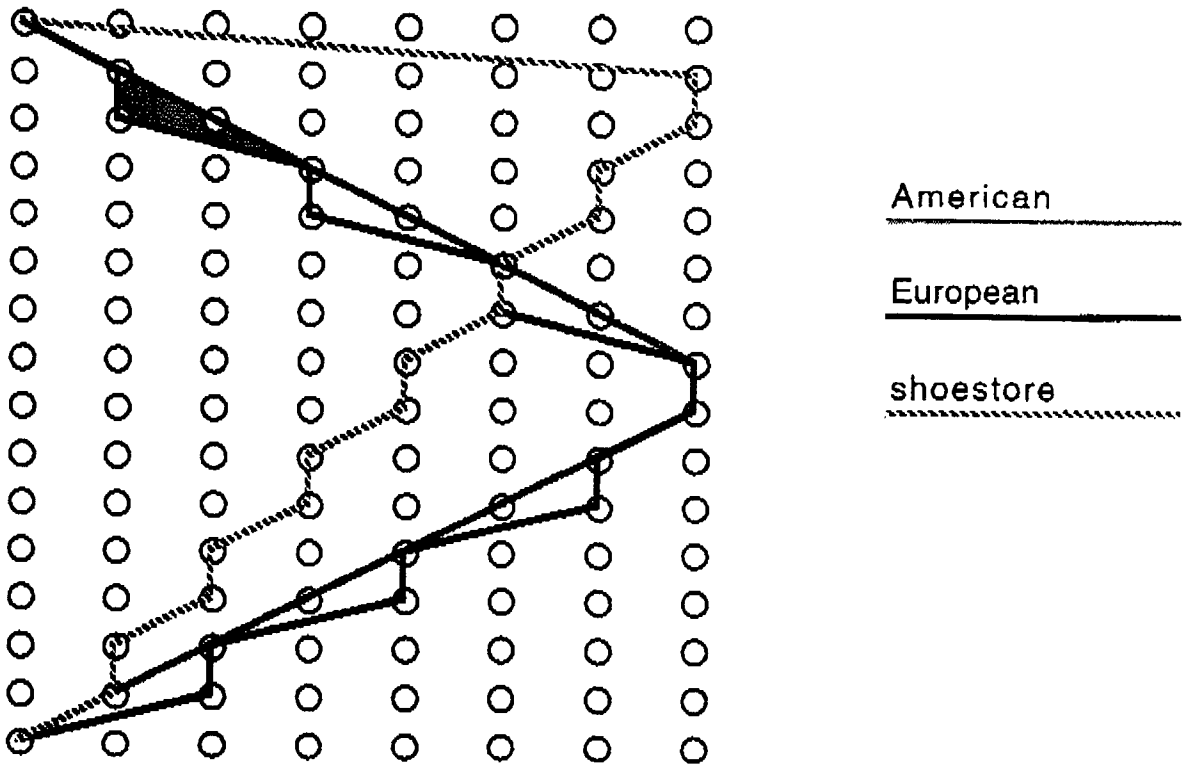


Fig.7 Representation of the lacings using reflections.

The figure requires a little explanation. It consists of $2n$ rows of eyelets, spaced distance d apart in the horizontal direction. Successive rows are spaced distance g apart vertically, and in order to reduce the size of the figure we have now reduced g from 2 (as it was in Fig.63) to 0.5. The method works for any values of d and g so this causes no difficulty. The first row of the diagram represents the left-hand row of eyelets. The

second row of the diagram represents the right-hand row of eyelets. After that, rows alternately represent the left-hand eyelets and the right-hand eyelets, so that the odd-numbered rows represent left-hand eyelets and even-numbered rows represent right-hand eyelets.

The polygonal paths that zigzag across this diagram correspond to the lacings, but with an extra 'twist' — almost literally. Start at the top left eyelet of a lacing pattern and draw the first segment of lace, running from left to right of the shoe, in between the first two rows of the diagram. Draw the next segment of lace *reflected* to lie between rows 2 and 3 instead of going back from row 2 to row 1 as it does in a real shoe. Continue in this manner, reflecting the physical position of each successive segment whenever it encounters an eyelet. After two such reflections, the segment will be parallel to its original position but two rows lower, and so on. In effect, the two rows of eyelets are replaced by mirrors. So, instead of zigzagging between the two rows of eyelets, the path now moves steadily down the figure, one row at a time, while its horizontal motion along the rows repeats precisely the motion along the rows of eyelets of the corresponding segments of the lacing pattern.

Because reflection of a segment does not alter its length, this representation leads to a path that has exactly the same length as the corresponding lacing pattern. The added advantage, however, is that it is now easy to compare the American and European patterns. In a few places they coincide, but everywhere else the American pattern runs along one edge of a thin triangle (one such triangle is shown shaded) while the European one runs along two edges of the same triangle. Because any two sides of a triangle exceed the third side in length (that is, a straight line is the shortest path between two given points), the American lacing is obviously shorter.

TECHNICAL EXTRA

It is not quite so obvious that the shoe-store lacing is longer than the European. The simplest way to see this is to eliminate from both paths all vertical segments (which contribute the same amount to both lengths because each path has n-1 vertical segments) and also any sloping segments that match up. The result is shown in **Fig.8** (dark lines). If each V-shaped path is now straightened out by reflection about a *vertical* axis placed at the tip of the V (faint lines) it finally becomes easy to see that the shoe-store path is longer, again because two sides of a triangle exceed the third side.

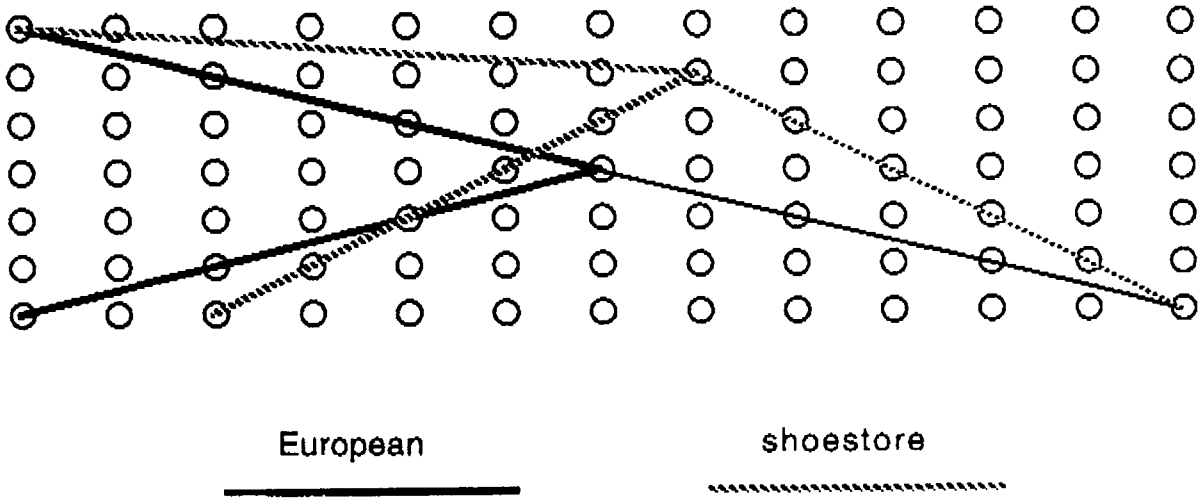


Fig.8 Comparing European and shoesore lacings.

FURTHER READING

Richard Courant and Herbert Robbins, *What is Mathematics?*, revised by Ian Stewart, Oxford University Press, New York 1996, pages 354, 392.

John H. Halton, The shoelace problem, *The Mathematical Intelligencer* **17** Number 4, pages 36-40.

Cyril Isenberg, *The Science of Soap Films and Soap Bubbles*, Dover, New York 1992.

Frank Morgan, The Double Bubble Conjecture, *Focus* [Newsletter of the Mathematical Association of America] volume **15** number 6, December 1995, pages 6-7.

Ian Stewart, *The Magical Maze*, Weidenfeld and Nicolson, London 1997.