Parabola Volume 49, Issue 2 (2013)

History of Mathematics: Proving the glaringly obvious

Michael A B Deakin[1]

Sometimes mathematicians seem to be overcautious in their assertions, overcautious even to the point of pedantry, requiring proofs of things that seem so clearly obvious that surely no proof is required. In this column, I discuss two such cases, but hope to show that perhaps those critics who lay the charge of pedantry may perhaps themselves be missing part of the point.

One of these cases comes from antiquity, the other is relatively recent. Other examples, very many of them, could be adduced; indeed have been, but these two have acquired especial notoriety.

1. Euclid I.20

Euclid's *Elements* is arranged in thirteen books, each of which contains a number of Propositions (theorems or constructions). The result I discuss here is Proposition 20 from the first book. It is a theorem, and it asserts that:

Any two sides of a triangle are together greater than the third.

If we label the triangle ABC, then the theorem asserts that the length of AB plus the length of BC exceeds the length of AC. In symbols:

$$|AB| + |BC| > |AC| .$$

Euclid saw fit to offer a proof of this theorem, and you can find it at

http://www.themathpage.com/abooki/propI-20.htm

Sir Thomas Heath, who produced the English translation of Euclid, not only reproduces Euclid's proof, but also offers two others. The first is one presented by the fifth century mathematician Proclus, who wrote an influential commentary on Euclid's classic. We also learn from Proclus that the Epicureans (followers of the fourth century philosopher Epicurus) ridiculed this theorem as being evident even to an ass and requiring no proof. They argued that if fodder is placed at one vertex of a triangle and the ass at another, then the ass would proceed directly toward the fodder rather than take the long way round via the other two sides. Lawrence Spector, author of the website given above comments:

[1]Dr Michael Deakin is an Adjunct Senior Research Fellow in the School of Mathematical Sciences at Monash University.

> If anyone wanted to ridicule mathematics for its insistence on the axiomatic method of orderly proof, then this theorem offers a wide target.

However, the mathematician Sir Henry Savile (1549-1622) offered a caustic comment on the Epicureans, saying that they "were worthy indeed to share the hay with the ass". [Perhaps a better way of putting things would have been to remark that evidently the Epicureans placed more faith in asses than in geometers!] Proclus himself commented that a mere perception of the truth of a theorem is a different thing from a scientific proof of it and a knowledge of the reason *why* it is true.[2] Heath also quotes the Scottish geometer Robert Simson (1687-1768) to the effect that the number of axioms should not be increased without necessity. [One can very cheaply reach the desired conclusion by invoking as an axiom, "A straight line is the shortest distance between two points". But that's cheating! It imports as an axiom a very sophisticated way of looking at the matter.]

However, suppose we adopted a different approach to the question, and used coordinate geometry. Let A be the point (x_1, y_1), B the point (x_2, y_2) and C the point (x_3, y_3).

Then

$$
\begin{aligned}
|AB| &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} = L_1 \text{ (say)},\\
|BC| &= \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2} = L_2 \text{ (say) and}\\
|AC| &= \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2} = L_3 \text{ (say)}.
\end{aligned}
$$

We need to prove that $L_1 + L_2 > L_3$, that is to say, that for all sets of three pairs $(x_1, y_1)\,,\ (x_2, y_2)\,,\ (x_3, y_3)\,,$

$$
\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} + \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2} \\
\geq \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2}
$$

All of a sudden things are looking a lot harder!

[Note that the $\geq$ sign replaces $>$; this covers the case where $A,\ B,\ C$ are collinear.] This inequality is known as the *triangle inequality* for obvious reasons. At first, I tried to prove it directly via a pencil and paper approach, but things got very messy, and so I resorted to the computer algebra software that comes with the package I use to prepare these columns.

First, I used the **define** command to set

[2] I made essentially this same point in relation to Pythagoras's Theorem in my column *Volume* **46**, *Number* **1**.

$$L_1 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$
$$L_2 = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2}$$
$$L_3 = \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2}$$

Then I used the **expand** command to calculate $(L_1 + L_2)^2 - L_3^2$. This resulted in the expression

$$\begin{aligned} 2x_2^2 + 2y_2^2 + 2\sqrt{x_1^2 - 2x_1x_2 + x_2^2 + y_1^2 - 2y_1y_2 + y_2^2} \\ \times \sqrt{x_2^2 - 2x_2x_3 + x_3^2 + y_2^2 - 2y_2y_3 + y_3^2} - \\ 2x_1x_2 + 2x_1x_3 - 2x_2x_3 - 2y_1y_2 + 2y_1y_3 - 2y_2y_3 \end{aligned}$$

The aim is to show that this expression is positive, or at very least zero. The squarings involved in the last calculation have reduced the number of square roots involved from three to (essentially) one. So, although the expression looks horrendous, there has been progress. We can further somewhat simplify the task ahead by cancelling out the common factor of 2, which is simply a nuisance. So the object now is to prove that

$$\begin{aligned} &\sqrt{x_1^2 - 2x_1x_2 + x_2^2 + y_1^2 - 2y_1y_2 + y_2^2} \\ &\times \sqrt{x_2^2 - 2x_2x_3 + x_3^2 + y_2^2 - 2y_2y_3 + y_3^2} \\ &\qquad \geq x_1x_2 - x_1x_3 + x_2x_3 + y_1y_2 - y_1y_3 + y_2y_3 - x_2^2 - y_2^2 \end{aligned}$$

We can now entirely eliminate the square root by means of a further squaring, which leaves us needing to show that

$$\begin{aligned} &\left(x_1^2 - 2x_1x_2 + x_2^2 + y_1^2 - 2y_1y_2 + y_2^2\right)\left(x_2^2 - 2x_2x_3 + x_3^2 + y_2^2 - 2y_2y_3 + y_3^2\right) \\ &\qquad \geq \left(x_1x_2 - x_1x_3 + x_2x_3 + y_1y_2 - y_1y_3 + y_2y_3 - x_2^2 - y_2^2\right)^2 \end{aligned}$$

Next I rearranged things to put everything on the left of the inequality sign, giving there the expression

$$\begin{aligned} &\left(x_1^2 - 2x_1x_2 + x_2^2 + y_1^2 - 2y_1y_2 + y_2^2\right)\left(x_2^2 - 2x_2x_3 + x_3^2 + y_2^2 - 2y_2y_3 + y_3^2\right) - \\ &\qquad \left(x_1x_2 - x_1x_3 + x_2x_3 + y_1y_2 - y_1y_3 + y_2y_3 - x_2^2 - y_2^2\right)^2. \end{aligned}$$

And now the task is to show that this horrible expression is positive (or at very least zero). To this end, I invoked the **simplify** command to find

$$(x_1y_2 - x_2y_1 - x_1y_3 + x_3y_1 + x_2y_3 - x_3y_2)^2 .$$

And now we are done, because a square cannot be negative. (I leave to the reader the task of proving that the case in which this expression reduces to zero is that of collinearity.)[3]

A book that I can thoroughly recommend to *Parabola*'s readers is *The Mathematical Experience* by Philip Davis[4] and Reuben Hersh. There they distinguish between two different types of proof: the 'Analog' and the 'Analytical', and note that the two approaches may lead to vastly different levels of difficulty. In the case of the Triangle Inequality, the *Analog* approach yields the obviousness noted by the Epicureans; the *Analytical* version was hard slog even with the help of a computer algebra package!

This mismatch is even more obvious in one of their most telling examples.

2. The Jordan Curve Theorem

The so-called Jordan Curve Theorem is notorious as an example of something that is "obvious" but extremely difficult to prove. One statement of it reads:

> A simple closed curve in the plane separates the plane into two regions, one finite and one infinite.

To understand what it says, we need to become clear on what is meant by "a simple closed curve". Put in straightforward lay terms, it just means a curve that ends up where it began without intersecting itself on the way. When this understanding of the concept is adopted, then we can readily agree with Davis and Hersh that the *Analog* proof is "visually obvious". They go on to say that the *Analytical* proof is "Very difficult, the difficulty deriving from the fact that an excessive degree of analytic generality has been introduced into the problem".

Our naive idea of a "simple closed curve" is one that is topologically equivalent to a circle, which is to say that, without stretching it or passing it over itself, it can be pushed around until it adopts a circular shape. And then it really does seem trivial to notice as obvious that some points lie inside it and others outside. The suggestion that

[3] One approach (not the best in this case) is to recognize that the expression in the parentheses is in fact the area of the triangle ABC. However, this was exactly the route used by Elsie Cerutti and P J Davis in their first-ever computer proof of a (different) mathematical theorem. See "FORMAC meets Pappus: Some Observations on Elementary Analytic Geometry by Computer" (*American Mathematical Monthly* **76** (1969), pp. 895-705). These authors prefaced their proof by saying that coordinate geometry may be thought of as "a 'machine' into which one feeds the hypotheses of certain geometric situations and which is guaranteed to 'grind out' the desired conclusions given sufficient patience on the part of the problem solver. However, it ... has long been known that many elementary situations give rise to impossibly long and tedious computations, and hence the universal method [i.e. coordinate geometry]... founders upon the rock of limited human patience and endurance." They suggested that this is a place for the computer to be called on to act.

So I have just given an example of their principle!

[4] This is the same Davis as mentioned in the previous footnote.

this fact actually requires proof first came from Bernard Bolzano.[5] In this connection, he wrote:

> If a closed line lies within a plane and if by means of a connected line one joins a point of the plane which is enclosed within the closed line with a point which is not enclosed within it, then the connected line must cut the closed line. [6]

Bolzano himself offered no proof of this, but he did realize that the notions of "curve" current at his time needed more exact definition. His entry in the *Dictionary of Scientific Biography* supplies some very relevant background, not only for this particular example, but also for the theme of my whole article.

> [Bolzano's] interest in mathematics was stimulated by reading G. Kāstner's *Anfangsgrūnde der Mathematik*, because Kāstner took care to prove statements which were commonly understood as evident in order to make clear the assumptions on which they depended.

It was the later mathematician Camille Jordan (1838-1922) who took the matter further. In a textbook written in 1887 (and much reprinted thereafter), he stated the theorem and gave a proof. (A curve that exhibits this property has since been named a "Jordan curve" in his memory.) His proof, however, has been criticized as incomplete or even erroneous, and you will find such a judgment on many of the websites devoted to the theorem. However, it was defended in 2007 by the US mathematician Thomas Hales, who claimed that Jordan's proof was perfectly legitimate. Hales recast Jordan's proof in the course of a 16-page paper. Others credit the first correct proof to the US mathematician Oswald Veblen (1880-1960); but this too has attracted criticism.

The inference to be drawn from all this is that the proof is far from simple! However, it is not too difficult to prove the result in the special case of a polygonal curve. This special case of the theorem is discussed at some length in the popular text *What is Mathematics?* by Richard Courant and Herbert Robins, a classic I can thoroughly recommend. See pp. 267-269. Alternatively look at the website:

http://www-cgrl.cs.mcgill.ca/~godfried/teaching/cg-projects/97/Octavian/compgeom.html#proofthm

[5]Bernard Bolzano (1781-1848) is a most interesting figure. As well as being a mathematician, he was a philosopher and a theologian, indeed an ordained Catholic priest. However his commitment to the spirit of the Enlightenment, with its insistence on rigorous demonstration and hard evidence, meant that his Christian theology was of a non-standard sort. He doubted the literal truth of much of the Bible, adopted a utilitarian stance on Ethics and embraced both pacifism and a form of socialism. His strongly held methodological and moral principles got him into trouble with the Church authorities. They stripped him of his academic post (he had been Professor of Philosophy at Charles University in Prague), banished him from the city and censored his writings. It was only well after his death that the significance of his work was recognized.

[6]A closely related result is now known as "Bolzano's Theorem": If $f(x)$ is a continuous function, and if $f(a) < 0$ and $f(b) > 0$, where $a < b$, then there exists a c such that $a < c < b$ and $f(c) = 0$. This also apparently obvious theorem finds wide use in numerical analysis and elsewhere.

Here is the gist of the proof in this special case. Let Γ be a closed non-self intersecting polygon and let P be a point in the plane but not on Γ. Starting at P, draw a horizontal (half-)line (which I will call a "ray") to the right, extending to infinity. Because the polygon Γ has only a finite number of sides, this ray will intersect it in only finitely many places. If the number of such intersections is even, then P lies outside the polygon; if it is odd, then it lies inside. (Remember that 0 is an even number.)

But now consider a possible objection: What if we went off toward infinity in some different direction other than the very special one just specified? The insideness-outsideness property should not depend on the quite arbitrary choice we made in choosing the direction of the ray. Well, it turns out not to matter. The number of intersections can increase or decrease only as the result of an interaction with a corner of the polygon. There are various ways in which this can happen. If the newly positioned line now cuts through two sides of a previously bypassed corner, there will be two more intersections. On the other hand, if it now manages to avoid a corner that in another position it had previously cut through, it will decrease by two. In both these cases, the parity (evenness or oddness) remains unaffected. [If the ray passes through a vertex (corner point), then that counts as one intersection, exactly as it would if passed through a point to either side of that vertex; if the ray happens to coincide with a side for a while the same thing applies. Finally, if the ray just grazes a vertex, then this doesn't count as an intersection. All but one of these situations are illustrated by Courant and Robbins.]

Thus it follows that the new orientation of the ray never alters the parity of the number of intersections. In other words, it cannot affect the overall result. I leave to the reader the task of identifying the outside region as infinite and the inner one as finite.

The real difficulty comes when we try to extend the polygonal result to more general curves. One approach (one of a good many now available) is to note that a continuous curve may be approximated to arbitrary accuracy by a polygon (a claim that itself requires proof), and so hope to use the more specialized result in this context. This is where Jordan concentrated his efforts; indeed he did not bother to prove the polygonal result at all. It was on this ground that his proof has come in for criticism, but Hales suggests that he simply regarded this special case as obvious.

I will not here attempt to complete the proof. However one tries it, it involves one in mathematics beyond what we expect of *Parabola*'s target readers. But what I will do here is to point to some of the difficulties that still lie ahead. One of these, one of the least problematic, concerns an extension of the statement of the theorem as given above. The two regions separated by the curve are both "connected"; this means that any two points lying in the interior region for example can be joined by a curve all of whose points are also interior points. Indeed, we can assert more: this region is also "simply connected", which means that any closed curve lying entirely within it can be shrunk to a point via intermediate curves all lying entirely within the bounding curve Γ. The corresponding statements are also true of the exterior region. These properties of the regions require further proof.

Another, more difficult, aspect concerns classes of curves that would not have been envisaged by Jordan. There are, for example, fractal curves. Possibly the first and best-known of these is the "Koch snowflake". It is discussed and (in part) illustrated on many websites. Because this curve itself is actually produced as the limiting case of an infinite number of constructions, we can only display those approximating curves resulting from a finite number of these operations. Although fractals were discovered subsequent to Jordan's work, his theorem remains true for them.

And there are even more bizarre "curves". One of these is that devised in 1890 by Giuseppe Peano,[7] who constructed a curve that passes through every point inside a square of side 1. Again it is a limiting case of an infinite set of procedures. A graph of sorts could be given, but it would simply be a black square, and so not very informative! Again, more is to be learned from the approximating curves produced along the way. Peano's curve is now recognized as one of a family of "space-filling curves". This particular one is not closed, so the question of its relation to the Jordan Curve Theorem does not arise. However in 1903, the American mathematician William Osgood (1864-1943) found a space-filling Jordan curve, to which of course the theorem *did* apply.

For more on the Jordan Curve Theorem, I particularly commend to readers one website. Go to

https://facultystaff.richmond.edu/~wross/PDF/Jordan-revised.pdf

The article reproduced there is the result of a collaboration between a graphic artist and her mathematician husband. Not only does it provide an excellent background, but it also includes some clever and attractive illustrations. The first two curves illustrated there are indeed "simple closed curves" (Jordan curves), but the identification of their "insides" and their "outsides" is a rather tricky matter! These curves, however, are not of the more bizarre types (fractals and space-fillers) mentioned above. The reader is invited to work out how to determine, for example, the inside of their Figure 1. In a sense, it is not a difficult matter: tedious, yes, difficult, no! (Carrying this out, something I admit I haven't done, also provides a check on the claim that these curves really are of Jordan type.) The two authors have provided an example of that dichotomy that I have remarked in several of these columns: the logical versus the psychological. This dichotomy also underlies the comment made by Cerutti and Davis in Footnote 3 above.

If we compare the two cases discussed here, we see that the obviousness of Euclid's I.20 is rather more immediate than that of the Jordan Curve Theorem. Essentially this is because it is more straightforward to say what a triangle is than to nail down exactly what is meant by a "simple closed curve". Perhaps this is what Davis and Hersh had in mind when they spoke of "an excessive degree of analytic generality [being] introduced into the problem".

[7] Giuseppe Peano (1858-1932) was one of the very greatest mathematicians. Perhaps his best-known accomplishment was his axiomaticization of the laws of arithmetic.