Parabola Volume 41, Issue 1 (2005)

History of Mathematics: The Foundations of Calculus

Michael A B Deakin[1]

One of the most significant advances in the entire history of mathematics was the development of Calculus in the latter years of the 17th Century. This is usually seen as proceeding from the work of Newton in England and of Leibniz in what is now Germany. Whether their work was entirely independent, to what extent either of them appropriated material from the other, and who thought of the idea first are all questions that have long and vigorously been debated. However, that is not my concern here; rather I will outline another debate, one which (unlike the other) led to fruitful mathematical advances in both the 19th and 20th centuries.

I will set the stage by quoting from a popular exposition: W W Sawyer's *Mathematician's Delight*, first published in 1943. On pages 206–207, he has this to say:

> If a mathematical method gave the correct answer to a practical problem, people did not bother much whether it was logical or not. In dealing with small changes, Δx, mathematicians followed their own convenience: at one moment they said, Δx is very small, it will be convenient to regard it as being equal to 0.' A little later they wanted to divide by Δx, so they said, 'If Δx is 0 we cannot divide by it: we will suppose it to be very small, but not quite 0.' Whichever was the more convenient, that they supposed to be true. . . . . [T]his rough and ready method worked quite well. [But after] about 150 years of carefree mathematics, difficulties began to be felt. . . . . [B]y doing things that look reasonable, we [can reach] an untrue result. . . . . [During the 19th century there was] a reaction to the earlier carefree times, and a spirit of caution spread. Mathematicians became rather like lawyers, very concerned about the exact use of words, very suspicious of arguments which merely 'looked reasonable'.

Sawyer goes on to demonstrate cases in which plausible reasoning leads to untrue, even absurd, results. The examples he gives all come from the theory of infinite series (sums of discrete, or separate, terms) rather than from any confusion as to the status of a small number Δx. As we shall see, there is a reason for this. But let us not get ahead of the story.

For the present, consider a specific example. It is one of the simplest available and many authors (including Sawyer) discuss it in considerable detail. Take the function $y = x^2$, which, as readers will know, graphs as a parabola. At some point (x, x^2) we seek to determine the slope of this parabola. (This question closely relates to another: the determination of the speed of a falling object.)

[1]Michael Deakin is a lecturer in Mathematical Sciences, Monash University.

To do this, consider a second point located to the right of the vertical axis at a distance of, not x, but $x+\Delta x$. At this second point the value y will be not x^2, but $(x+\Delta x)^2$. We may think of the symbol Δ as a shorthand for 'the change in $\ldots$'. So the second value of one quantity (say x) is its first value plus the change. We may now compute the change in the value of the second quantity from the effect of a change in the first. Specifically:

$$\Delta y = \text{ the change in } y = (x + \Delta x)^2 - x^2 = 2x\Delta x + (\Delta x)^2. \tag{1}$$

Thus the slope of a straight line joining the two points of the graph (which is also regarded as the average, or mean, slope of the parabola itself) is given by

$$\frac{\Delta y}{\Delta x} = 2x + \Delta x. \tag{2}$$

And now that we have done all this, and determined the average slope for any value of Δx whatsoever, set $\Delta x = 0$, and so find that the slope of the tangent at the point (x, x^2) is $2x$.

This (in essence) was the approach adopted by Newton and Leibniz. Newton described his work in terms that are now quite obsolete. Generally, he referred to y as the *fluent* and to the slope as the *fluxion*, words that have now disappeared. (However, his notation $\dot{y}$ for this latter is still used.) There has actually been some confusion as to the precise meanings of these and other terms, but most contemporary dictionaries tell us that the word 'fluxion' carries the meaning of today's term 'derivative'.

However, it is Leibniz' notation that has achieved wider acceptance. If we set $\Delta x = 0$ in the right-hand side of equation (2), we cannot also use this convention on the left. For that, Leibniz wrote $\frac{dy}{dx}$, and so would have it that the correct equation for the slope is

$$\frac{dy}{dx} = 2x. \tag{3}$$

Among the many attempts to make these various operations logically consistent, we may perhaps single out the interpretation of dy and dx as 'infinitesimal quantities'. These were supposed to be numbers smaller than any real number we could imagine, but nonetheless larger than zero. This seems to be an impossibility, but computations using these infinitesimals, as they came to be called, somehow always seemed to yield correct answers! The understanding was that if, in equation (2), we replaced Δy and Δx respectively by dy and dx, then we could ignore the final term, as it would add nothing to the (finite) total. Alternatively, we could make this same change in equation (1) and regard dx as a quantity, which, although not itself quite zero, has a square which is.

This is clearly an unsatisfactory state of affairs, and it drew trenchant criticism from a rather unlikely source. George Berkeley (1685–1753) was a philosopher and cleric, indeed a bishop. However, his criticism of infinitesimals ('evanescent increments') was forcefully put:

> And what are these fluxions? The velocities of evanescent increments. And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them ghosts of departed quantities?

He recognised the *effectiveness* of Calculus, and offered his own analysis of why it worked. Equations like equation (1), he thought, contained two errors: in the first place they did not end up calculating the slope of the tangent, but rather that of a chord (or "secant"), and secondly they conveniently overlooked the existence of higher powers of Δx. The success of Calculus, in Berkeley's view, came about because these two errors happened to cancel one another out.

His motive in writing was not mathematical, but rather to make his own discipline of Theology seem simple and lucid by comparison.

> He who can digest a second or third fluxion, a second or third difference, need not, we think, be squeamish about any point of divinity.

(Berkeley believed that his 'cancellation of errors' argument gave a satisfactory basis for the first derivative, but he baulked at second, third and higher ones!)

The lasting effect of Berkeley's criticism was, however, its influence on mathematics, rather than on the foundations of religious faith. The point that an increment, even an 'evanescent' one, could not be both zero and non-zero, was telling, and indeed troubling. (Leibniz is supposed to have described infinitesimals as 'fictions, but useful fictions'.)

The avoidance of logical error, when it came, involved the scrapping of the concept of infinitesimals altogether. Our modern understanding was developed in the 19th century, and there were many contributors. One, however, is outstanding. He is Augustin Louis Cauchy (1789-1857). We can get a taste for Cauchy's approach by considering the relationship between equations (2) and (3). Equation (2) is valid for all values of Δx *except* for $\Delta x = 0$. So, we carefully refrain from using this value. But we can make Δx as small as we please, just not zero. The smaller we make Δx, the closer the right-hand side of equation (2) approximates the value $2x$. We can make the discrepancy smaller than any tiny amount we care to specify.

It is rather like the engineering concept of a *tolerance.* A production-line filling bottles, say, aims to put 750 ml into each. This ideal will never be *exactly* achieved, but if we are within say half a ml of the nominal volume, we agree that we will be satisfied. If greater accuracy is needed (perhaps in a medical application), the allowable tolerance is reduced. It will take more effort (and cost more!) to make the improvement, but in principle it can be done.

So it is with our Calculus example. We abandon the quest for exactitude as such, but replace it with another goal, which, as it turns out, is just as good. We say that, in equation (2), the right-hand side may be made to differ from $2x$ by an amount we can reduce below any given allowance for error – merely by making Δx sufficiently small. Moreover, there is *no value other than* $2x$ that has this property. If we were told of an alternative, not equal to $2x$, then we could disprove the claim that it shared this property. Thus the unique value $2x$ is referred to as the limit of the right-hand side as Δx tends to zero. We are now in a position to define $\frac{dy}{dx}$. It is to be the limit of $\frac{\Delta y}{\Delta x}$ as Δx tends to zero. In symbols:

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \tag{4}$$

This equation holds good not only for our particular example of $y = x^2$, but generally. Just as long as such a limit can be defined, equation (4) can be applied.

It is this account that enabled Cauchy and later mathematicians to avoid falling into self–contradiction, and in consequence it is this definition of the derivative that is now taught in our schools and universities. Sometimes students see this roundabout approach as needlessly complicated, and it is perhaps for this reason that the old way of thinking never entirely disappeared. It is too convenient to treat dy and dx as being separate quantities with $\frac{dy}{dx}$ being their ratio. A viewpoint rather like this is particularly useful when y is a function of not merely the single variable x, but of two variables, say x and t.

There is no strict necessity to take such a step; it is merely most convenient. Nor does it seem to lead to any troubles. Of course, the old interpretation of these quantities as 'infinitesimals' was thrown out. It was hard to see how these could survive the withering criticism of Bishop Berkeley. The new term was 'differential', and indeed Calculus, once called 'the infinitesimal calculus', became instead 'the differential calculus'. Moreover the differential was not a mysteriously 'evanescent' quantity, but was well and truly finite.

In the figure that follows, one popular explanation is illustrated. The original function is represented by a curve PQ. Its tangent at the point P is shown as PR. Because PR is straight, its slope is constant, and it is moreover equal to the slope of the curve PQ at the point P. Let A and B be points to the right (typically) of P, and let the length of PA be Δx. Then the length of AQ is Δy and the slope of the curve at P is approximated, but only approximated, by $\frac{\Delta y}{\Delta x}$ with the approximation becoming better and better as Δx gets smaller and smaller. The length of PB is taken to be dx and that of

BR to be dy. The slope of the line PR is then $\frac{dy}{dx}$ and indeed dx is arbitrary because of the constancy of the slope. Thus these differentials are perfectly ordinary quantities with nothing mysterious about them at all! They do, however, refer not to the original function but rather to the tangent.

This explanation had much currency during the middle years of the 20th century. One may see it influentially promulgated for example in the *American Mathematical Monthly* (1942, pp 110–112). However, it is not entirely satisfactory. It doesn't lead to a satisfactory derivation of the "chain rule": $\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$. Here we would like to 'cancel the du's', but there is no logical justification for this apparently natural step! Nor does it allow for a neat interpretation of the dx in integration, in expressions like $\int f(x)dx$.

A lively debate ensued in the pages of *American Mathematical Monthly* with several eminent mathematicians taking part. In the end, in 1952, the editor, C B Allendoerfer, intervened with his own definition:

> The presence of a differential in a mathematical expression is a sign to the reader that this expression is obtained by a limiting process from a second expression in which the differentials $dx,\ dy$, etc. are replaced by finite increments $\Delta x,\ \Delta y$, etc. The nature of this limiting process depends upon the particular expression and must be inferred from the context.

He concluded his editorial with the words:

> There is a discredited view that differentials are some sort of 'infinitely small quantities'. Of course, this is nonsense, but there is a germ of truth in it. For differentials do mark the places in mathematical expressions which are held by small, but finite, quantities while some type of limiting process is carried out in which these small quantities tend to zero. To the writer they have no other consistent meaning.

This was for its time very sound advice, but it now needs modification. In 1961, another mathematician, Abraham Robinson (1918–1974), succeeded in making the notion of an infinitesimal respectable after all. In effect, he showed that our usual system of numbers need not be the only one.

To get a flavour of the new viewpoint, consider a previous extension of the number system. Irrational numbers had been discovered by the ancient Greeks. (We first encounter them in the writings of Aristotle.) Yet it was not until the 19th century that their place in the number system was firmly and logically established. (Something of the same happened with the complex numbers, although their introduction came much later.)

It is actually quite a delicate matter to establish that the usual rules of arithmetic can be extended to cover the irrational as well as the rational numbers. We do not teach this material in our schools; it would be much too complicated to deal with at that stage. Nonetheless, this does not prevent our syllabuses from including material on irrational numbers and asking students to be able to calculate with them. That irrational numbers can be discussed without contradiction is taken for granted. Indeed very few students would seem to question their validity today.

Similarly, Robinson's 'Non-Standard Analysis' is an analogous extension of the number system in which there are, in addition to our familiar numbers, infinitesimals, which are smaller than any finite quantity, but are nonetheless non-zero.

The details are too complicated to explain here. (The best attempt to give a relatively elementary account is Lynn Arthur Steen's article in *Scientific American* in August 1971; even this, however, is incomplete, as well as being quite heavy going!) But the upshot is that Bishop Berkeley's objections are not the absolute 'killers' that previous generations thought them. Since we can indeed calculate with infinitesimals and never fall into error (as Robinson proved), we can use them secure in the knowledge that it is really OK after all. We now know why they work.

To make plausible the possibility of such new numbers, Steen uses probability theory. Suppose that we pick at random some number in the interval $0 \leq x \leq 1$. Now ask what the probability is that any *particular* number, $\frac{1}{2}$, say, is chosen. Of the infinitely many numbers that might have been chosen this one has emerged; the probability that it alone triumphed over all the others is zero. If we assigned to this probability any number larger than zero, we would soon be led into error. When we deal with *finite* sets, a zero probability corresponds to impossibility, but it is not so here. There are, to put it imprecisely, two subtly different sorts of zero involved. For it is not impossible for the number $\frac{1}{2}$ to be chosen, or any other number for that matter; in fact *some* number *must* be chosen.

The elaboration of these ideas leads to some difficult and intricate mathematics, and it was this area of the subject that Robinson was concerned to extend and apply. That his endeavour was successful has meant that we can now rest more easily with the concept of an infinitesimal than previously – even if we do not advert to the difficult and intricate mathematics that justifies the use of the concept.

Further reading

For those who would like to see more on this topic, begin with Steen's article (detailed above) and the references he gives. A Google search under 'Non-Standard Analysis' will also give more. Biographies of Newton, Leibniz, Berkeley, Cauchy and Robinson are all available on the MacTutor website and also in the *Dictionary of Scientific Biography*. The discussion in *American Mathematical Monthly* was reprinted in their anthology *Selected Papers in Calculus*. There are also good discussions in several books on the History of Mathematics in general.