Comments on the Difficulty and Validity of Various Approaches to the Calculus

DAVID TALL

With the introduction of new infinitesimal methods in the last two decades, there are now available a number of different approaches to the calculus. In her perceptive review essay on "Infinitesimal Calculus" [3] Peggi Marchi raised a number of important points worthy of comment. The first concerned the validity of the "old infinitesimal calculus". Her implication was that this is a flawed theory, a judgment which has been made by most mathematicians over the last three centuries. My major aim in this article is to show that this judgment is a matter of interpretation of the *meaning* of the ideas and the context in which they are used and that, given a suitable context, the methods are perfectly satisfactory. A possible source of error in the method which arises in university analysis does not arise in the more restricted context of school or college calculus, nor did it arise in the original theory of Leibniz. It can be eliminated in a simple way.

My second aim is to go some way towards answering Peggy Marchi's questions:

"Why are infinitesimal simpler and more intuitive than epsilon-deltas" (page 38).

"Calculus proofs are easy in the hyperreals but hard in the reals — why is this the case?" (page 42).

"Why is the intuitive picture of the hyperreals easy and the rigorous picture of the hyperreals difficult?" (page 41)

My final aim is to show how all these considerations arise out of old-style school calculus in a natural way and to put the case that this form of calculus, suitably interpreted, is mathematically correct and forms the best basis for beginning calculus, especially when allied to modern numerical and pictorial devices available on the computer.

Let us begin by considering some of the many approaches to the calculus and the different meanings attributed to the concepts.

(i) The "old, intuitive, infinitesimal method"

To differentiate a function $y = f(x)$ in the original Leibniz method, the variable x is incremented by an infinitesimal quantity dx and the dependent variable y is then incremented by $y+dy = f(x+dx)$. The derivative $f'(x)$ (using modern notation) is then the quotient dy/dx, and in computing this derivative, any "higher order" infinitesimal quantities are neglected. For instance, if $f(x) = x^3$ then

$$dy/dx = (f(x+dx)-f(x))/dx = ((x+dx)^3-x^3)/dx$$
$$= 3x^2+3xdx+dx^2$$

The quantity $3xdx+dx^2$ is infinitesimal and is neglected, leaving the derivative dy/dx to be $3x^2$.

(ii) The "dynamic limit method"

Here the calculations are similar to (i). We let h be a variable real number and compute the ratio $(f(x+h)-f(x))/h$, then allowing h to get closer and closer to zero. In this process, if the ratio approaches a limiting value, we take this limiting value to be the derivative $f'(x)$. Often the symbol δx is used instead of h and δy instead of $f(x+h)-f(x)$, so that we think of the ratio $\delta y/\delta x$ getting closer and closer to the limit $f'(x)$. By blending the notation of this method with that of Leibniz we arrive at the idea that dy/dx is the limit of the ratio $\delta y/\delta x$ as δx approaches zero. Often the ration dy/dx is now interpreted as a compound symbol. It is *not* to be thought of as the quotient of dy by dx and the symbols dy and dx have no meaning in themselves. Even so, the symbolism is used to advantage in remembering formulae like

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

though the student may be told that it is only a convenient memory device. He may even be told that he can cancel the "du" mentally, but must not mark the cancellation on paper!

Matters are further complicated by the school of thought that *defines* dx and dy by $dx = \delta x$ and $dy = f'(x)dx$. This school considers dy/dx to be a quotient once more, but arising out of the definition of dy in terms of $f'(x)$ and dx, rather than the computation of $f'(x)$ as a quotient of given quantities.

It is one of these variants of the dynamic limit method that is used in old-style school calculus. In England the teaching of analysis at universities is complicated by the fact of the students coming from school backgrounds where a variety of subtle shades of interpretation of meanings are current.

(iii) The "numerical method"

The derivative $f'(x)$ can be computed for a specific numerical value of x by tabulating values of h against $(f(x+h)-f(x))/h$. For instance, when $x = 1$ we have:

h	$(f(x+h)-f(x))/h$
1	7
0·1	3·31
0·001	3·003001
10^{-10}	3·000000003000000001
. . .	. . .
−1	1
−0·1	2·71
−0·001	2·997001
-10^{-10}	−2·999999997000000001

From this table we see that as h tends to zero from either side the values in the second column tend to 3.

(iv) The "computer drawing method"

To find the derivative $f'(x)$ for a specific numerical value of

x, tell a computer to plot the graph of f over a small interval $[x-a, x+a]$ and draw it to a highly magnified scale. Figure 1 gives computer printouts of $f(x) = x^3$ over the specified intervals centred on $x = 1$. To a suitable scale the computer drawing is (within its limits of accuracy) a straight line. The derivative $f'(x)$ is the gradient of this straight line. The key is to get the size of interval small enough to give a straight line, yet not so small that the computer's error in calculation distorts the picture. My colleague John Mills at the University of Warwick and I are developing computer graphics for this approach. Its value is that it exhibits part of the graph as being indistinguishable from a straight line, in particular the tangent to the curve at x is indistinguishable from the graph itself. This gives strong support to the notion that when δx is small, δy and dy are very close in value, with the difference between them being an error of higher magnitude. The method clearly shows that dy/dx *can* have a perfectly satisfactory interpretation as a ratio.

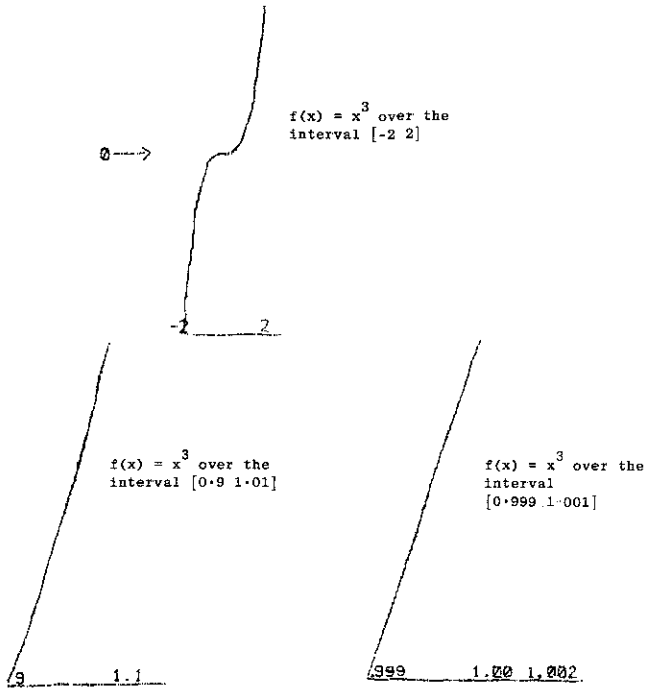

Figure 1

(v) The "epsilon-delta method"

To show that the quotient $(f(x+h)-f(x))/h$ tends to the real number $f'(x)$ as h tends to zero, the epsilon-delta method requires us to first specify how close the quotient is required to be to $f'(x)$ (say within a positive quantity epsilon), then we must be able to compute how close h must be to zero (within a positive quantity delta) so that $(f(x+h)-f(x))/h$ *is then within epsilon of* $f'(x)$. For a given epsilon the mental gymnastics required to find delta, even for a simple function like $f(x) = x^3$ are fairly prodigious, involving the manipulations of general inequalities.

During the last century the epsilon-delta method has led to fruitful advances in analysis and a clarification of the meaning of certain concepts for the professional mathematician, but it is too complex and intractable for the beginning student in calculus.

(vi) The "modern infinitesimal method"

In essence the approach follows that of Leibniz, but with modern logic to support the precise nature of infinitesimal and the "neglecting" of infinitesimal quantities. An infinitesimal quantity is now an element in an ordered field which is smaller in size than any positive rational number. The description of infinitesimals can be given in clear algebraic terms (given in detail in [1] and [2])

The derivative of $f(x) = x^3$ in these terms is found by taking a non-zero infinitesimal e and computing

$$(f(x+e)-f(x))/e = 3x^2+3xe+e^2.$$

The standard part of this expression (that is, the real number infinitesimally close to it) is

$$f'(x) = \mathrm{st}(3x^2+3xe+e^2) = 3x^2.$$

Sometimes e is denoted by δx, now considered as an infinitesimal, and δy is the infinitesimal quantity $f(x+\delta x)-f(x)$. Then we have

$$f'(x) = \mathrm{st}(\delta y/\delta x)$$

The infinitesimal dx, dy are defined to be $dx = \delta x$, $dy = f'(x)dx$, so that

$$dy/dx = \mathrm{st}(\delta y/\delta x)$$

In this way dy/dx is a quotient once more, but a quotient of infinitesinal quantities.

These various methods of viewing the derivative by no means exhaust all the possibilities. Cauchy's original epsilon-delta method included the use of infinitesimals, the modern constructive analysis of Bishop restricts the system to numbers and functions which can be explicitly computed, and so on. However, they are sufficient for this present discussion, for they give a wide variety of interpretations of the process of differentiation of which the formal epsilon-delta method is but one among many. When we judge the validity of any of these approaches, wittingly or unwittingly, we usually do so in a context which assumes one or more of them to be the valid one. Most mathematicians nowadays happily accept the epsilon-delta method as the touchstone of validity. Some also accept the modern infinitesimal method but may be nervous about the use of the axiom of choice in constructing the hyperreals.

There is a basic cognitive problem here involving the meaning that we assign to the processes and concepts of the calculus. As can be seen, the various methods outlined can have radically different meanings. It is often differences of *meaning* which cause controversy in the calculus.

For instance, Cantor's celebrated attack on infinitesimals was largely motivated by his interpretation of infinite numbers as cardinals or ordinals. Since neither of these types of number can be divided, he asserted that infinitesimals, which should be produced as the reciprocals of infinite numbers, cannot exist. A further boost to the attack on infinitesimals resulted from the formalization of the real number concept in the second half of the nineteenth century. A positive infinitesimal cannot be a real number, smaller than all positive quantities, since half of it is still positive but smaller. This did not eliminate infinitesimals from mathematics. Far from it. They remained, but with a distinct shift in meaning. An infinitesimal ceased to be regarded as a very small quantity and came to be considered as a function which tends to zero. This interpretation had arisen in the work of Cauchy and it continued to pervade textbooks well into the twentieth century.

The resolution of controversies involving the meaning of mathematics is rightly one of the provinces of the mathematics educator in the widest sense It concerns not only mathematics but also the cognitive processes involved in the historical development, in our current culture, and in the form of mathematical meaning which we pass on to the next generation It cannot be left to chance An educator must do more than just pass on the current mathematical culture, he must analyse it and modify it to make it appropriate for the learning and for its future use and development.

If we apply this wide frame of reference to the calculus we find a possible source of error in the old intuitive method which leads to incorrect proofs when misapplied in epsilon-delta calculus.

Cauchy defined an infinitesimal to be a "function which tends to zero". In the dynamic limit method (ii), if we write $u(h) = 3xh+h^2$, then we find that $u(h)$ tends to zero as h tends to zero. In Cauchy's terminology, $u(h)$ is an infinitesimal! Furthermore, if we write $v(h) = h$, then $v(h)$ is also an infinitesimal. In essence the dynamic limit method reduces to the old infinitesimal method but with a vital difference in meaning In the old infinitesimal method the infinitesimals are *quantities*, but in the dynamic limit method they are *functions*. As quantities they were usually considered to be variable and to grow arbitrarily small without actually reaching zero.

It is here that a fundamental source of error can arise In the Cauchy sense we define a function f to be an *infinitesimal* if $f(x)$ tends to 0 as x tends to 0 (Implicit in the definition is that f must be defined for all values of x near, but not necessarily equal, to zero) An infinitesimal will be said to be *proper* if $f(x)$ is non-zero for x near, but not equal, to zero, otherwise it will be called *improper*.

Examples $f(x) = x(x-1)(x-2)$ is a proper infinitesimal because $f(x)$ tends to zero as x tends to zero, but $f(x)$ is not equal to zero for x near the origin. On the other hand $u(x) = x \sin(1/x)$ $(x \neq 0)$ is an improper infinitesimal because it tends to zero as x tends to zero, but it is zero at a sequence of points arbitrarily close to the origin.

It is improper infinitesimals which are a source of error in the calculus Specifically, in the proof of

$$\frac{dv}{dx} = \frac{dv}{du}\frac{du}{dx}$$

when writing

$$\delta u = u(x+\delta x)-u(x), \quad \delta v = v(u+\delta u)-v(u) = v(u(x+\delta x))-v(x)$$

and considering the limit of

$$\frac{\delta v}{\delta x} = \frac{\delta v}{\delta u}\frac{\delta u}{\delta x}$$

we must make sure that δu is a proper infinitesimal. Otherwise δu will be taking on the value zero infinitely often as δx tends to 0 and we shall be dividing by zero If δu is a proper infinitesimal, then this proof is perfectly valid.

The first edition of Hardy's "Pure Mathematics" had a celebrated error in which this was not taken into account.

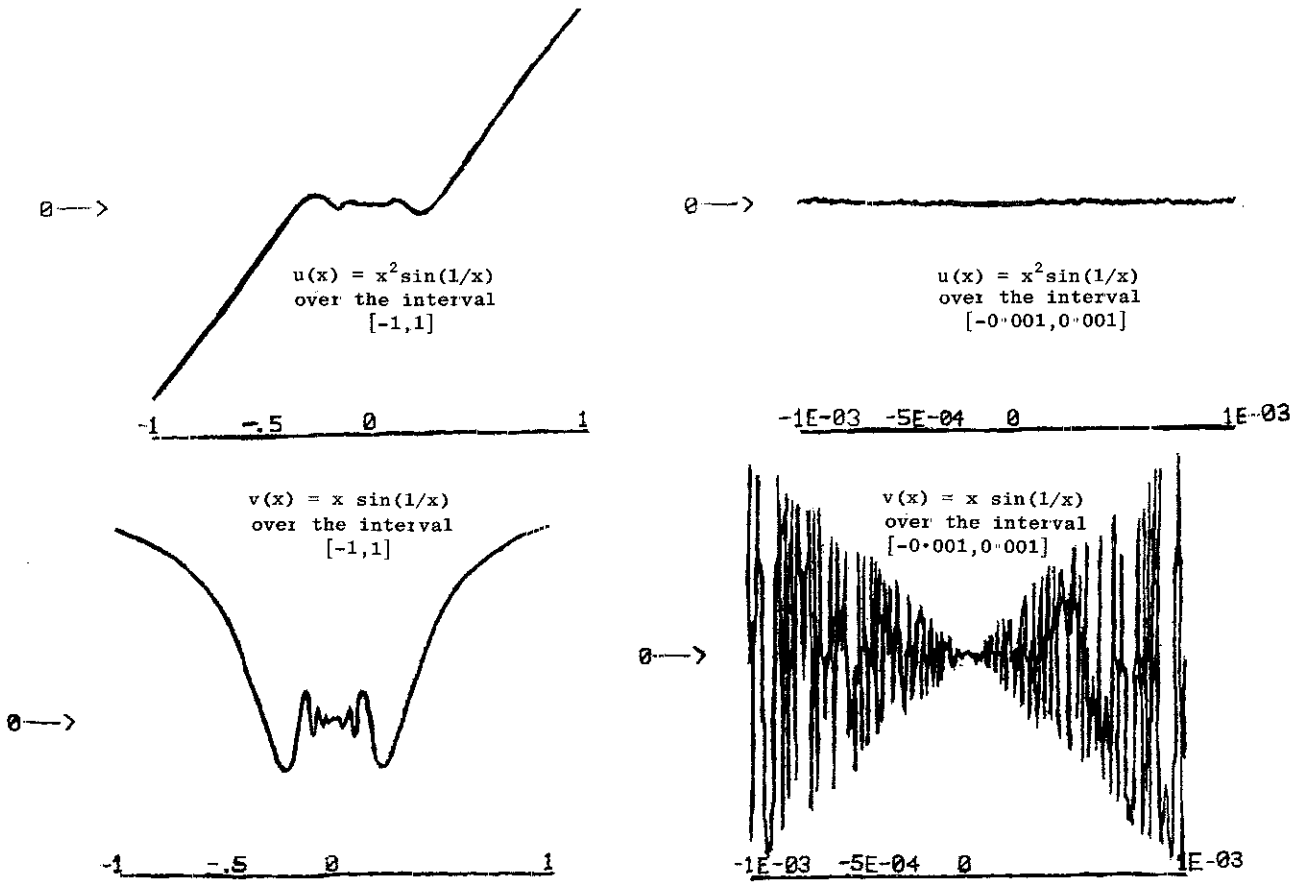

Figure 2

Nowadays, when this proof is dealt with in epsilon-delta analysis it is often made to seem unintuitive and hard. The fact is, however, that if δy is a proper infinitesimal and dy/dx exists, then $dy/dx = 0$ (To prove this, simply note that if $dy/dx = k$ then this means that $\delta y/\delta x - k$ tends to zero as δx tends to zero. Since δy is improper, it keeps on taking the value zero during this process whilst k is constant; the only way this can happen is if $k = 0$.) Now if δu is improper and $\delta v = v(u+\delta u) - v(u)$, then δv is also improper Thus the chain rule (1) is true *because both sides are zero*.

Of course, this only happens if the derivatives exist If we consider

$$u(x) = x^2 \sin(1/x), \quad v(x) = x \sin(1/x)$$

for non-zero x and define $u(0) = v(0) = 0$, then for $x = 0$ the increments

$$\delta u = u(0+\delta x) - u(0), \ \delta v = v(0+\delta x) - v(0)$$

are both improper infinitesimals, but du/dx exists at 0 and dv/dx does not. On a small interval centred on zero the computer printout of u continues to oscillate but the graph of v flattens out, as in figure 2 The irregularities in the picture of v are due to the fact that the program being used at the time only computed a limited number of values and there just weren't enough computed to cope with the large number of oscillations. The computer printouts also suffer from being drawn in a small dot-matrix printer; the original television picture of u over $[-0.001, 0.001]$ appears as a flat straight line over the middle portion.

Cauchy defined an infinitesimal f to be of higher order than an infinitesimal g if $f(x)/g(x)$ tended to zero as x tends to zero (In this definition $g(x)$ needs to be proper to avoid dividing by zero.) It is clear that if δy is improper, then dy/dx exists if and only if δy is a higher order infinitesimal than δx In the two examples, δu is higher order than δx, but δv is not.

In general, if δy is not of higher order than δx then $\delta y/\delta x$ *must* oscillate as δx tends to zero, for it keeps on taking the value zero yet does not tend to zero

All this bother about improper infinitesimals is totally unnecessary in the theory of Leibniz as he conceived it since his infinitesimals were all proper For him an infinitesimal was a variable quantity which tended to zero without actually getting there. The class of functions he had in his repertoire all had this property (If one were to have suggested u and v to him as functions, he would likely as not have responded that they are not defined at the origin and they are perfectly well-behaved elsewhere At this early stage in the calculus, the term "continuous" meant "given by a single formula" and it would not be permissible to define $u(0) = v(0) = 0$, for that could be classified as using a different formula for the function in part of the domain.) Thus the fuss which is made about this possible source of error is irrelevant to Leibniz.

It is also largely irrelevant to beginning calculus students. In [5], [6] it was shown that the experiences of dynamic limits in English schools were such that $u(h) \to 0$ was interpreted as implying $u(h)$ got *close* to zero as h approached zero but that it never actually reached there In other words, the class of functions which beginning calculus students have in their repertoire leads to an implicit belief that all infinitesimals are proper In this context the old intuitive calculus is *perfectly correct*, and when the context is broadened to include a wider class of functions, the modification required which I demonstrated above is quite obvious and quite trivial.

Even when this wider class of functions is included there is no reason to forgo the old-style notations and meanings All one needs to do is to make sure that one never divides by an improper infinitesimal, so that in computing the limit of $\delta y/\delta x$ as δx tends to zero, the denominator δx is never allowed to be zero

However, there is a flaw with only dealing with proper infinitesimals. It may be that u, v are proper infinitesimals, yet $u+v$ is improper For instance:

$$u(x) = x + x \sin(1/x) \quad (x \neq 0), \quad u(0) = 0$$
$$v(x) = -x$$

Thus the arithmetic of proper infinitesimals is not a closed system. If one persists with the old-style calculus then one should check that this does not lead to difficulties. The best way out is to insist that *independent* variables must be proper but dependent variables need not be. In a formula like

$$\delta(u+v)/\delta x = \delta u/\delta x + \delta v/\delta x,$$

it does not matter if δu and δv are proper or improper because they are dependent variables. In elementary calculus we do not add dependent variables, so the flaw does not cause any problems

However, the theory starts creaking at the edges when we consider

$$\frac{\delta v}{\delta x} = \frac{\delta v}{\delta u}\frac{\delta u}{\delta x}$$

for δu is a *dependent* increment in computing $\delta u/\delta x$, but an *independent* increment in computing $\delta v/\delta u$ The dynamic limit notion, which visualises δx moving "continuously" towards zero (in the cognitive sense that it passes through all intermediate values) enforces the improper infinitesimal δu to be zero infinitely often, causing a breakdown of the intuitive theory. It could be patched up by changing the conceptual imagery and allowing δx to move to zero in fits and starts, jumping over the "bad points" where $\delta u = 0$ But if the learner is using such sophisticated notions as improper infinitesimals, perhaps it is time to call it quits and move to a more appropriately subtle form of the calculus I believe it is for reasons like this that one may wish to criticise the Leibnizian notation and go on from the notion of dynamic limit to the epsilon-delta or modern infinitesimal methods

The reasons why infinitesimals are simpler and more intuitive than epsilon-deltas are two-fold Firstly infinitesimals rely on the dynamic idea of limit which gives a good cognitive feel for the limiting process. Secondly, the epsilon-delta method is not content with noting just that a variable tends to zero, it actually computes how fast this happens by laying down epsilon and asking one to compute delta. When this computation is carried through it can become amazingly complex

Contrast this with the infinitesimal definition of continuity:

> x, y infinitesimally close implies $f(x), f(y)$ are infinitesimally close

The fact that f is not continuous requires simply:

> there exist x, y infinitesimally close with $f(x), f(y)$ differing by a non-infinitesimal quantity.

Remarkably, this definition works for the dynamic limit version as well. To show that f is continuous at x one only need take any function u which is an infinitesimal and show that $f(x+u)-f(x)$ is infinitesimal. For instance, if $f(x) = x^3$ and $u(t) = t^2+5t^4$, then $u(t)$ is an infinitesimal, and so is

$$f(x+u(t))-f(x) = (x+t^2+5t^4)^3-x^3$$

because the latter, considered as a function of t, tends to zero as t tends to zero.

This illustrates why the infinitesimal method is so intuitive: it is essentially the same as the dynamic limit method which has such strong cognitive appeal. Proofs using the infinitesimal method also draw on this existing cognitive structure and so appear to be intuitive. This is why calculus proofs are easy in the hyperreals. They are equally easy in the dynamic limit method — in fact they are so easy that they are not regarded as "proofs" at all. The mathematical community gets drawn inexorably into the epsilon-delta mode of operation for formal proofs, and here the computations become far more difficult. There are quantifiers to manipulate, implicitly or explicitly, the method has to work backwards from the closeness of $f(x)$ and $f(y)$ (the given epsilon) to establish the closeness of x and y (the required delta), and *ad hoc* methods have to be called into play to do the computations.

However, all is not that straightforward in the modern infinitesimal method. First there is the problem that the more sophisticated is the learner, the more likely he is to have ideas in his cognitive structure which do *not* extend to the hyperreals. Ideas of dynamic limits carry over well but more complex statements (which involve quantification of sets) can break down. Students who have met such statements as the completeness of the reals (every non-empty subset of the real numbers which is bounded above has a least upper bound) initially lack the subtlety to distinguish between these statements, which do not extend to the hyperreals, and those of first-order predicate calculus which do. Thus to them the hyperreals are less intuitive because it is not clear which ideas in their cognitive structure give correct intuitions.

The statements in the modern infinitesimal method are also more subtle than is sometimes indicated. For instance, the earlier infinitesimal description of continuity I gave requires that f be extended to a larger hyperreal domain D^* and that when x and y are taken infinitesimally close we require x in D but y in D^*. The dynamic limit method gives an insight as to why this is so. There we have f is continuous at x in D if

$f(x+h)$ tends to $f(x)$ as h tends to 0

In this expression $y = x+h$ is a *function* (of h), whilst x is an element of D. Thus y is a function which gets infinitesimally close to x, indicating that it is a very different kind of animal from x in D. This is why we require y to be in D^*.

In "Infinitesimal Calculus" [1] by Henle and Kleinberg, hyperreals are taken to be sequences of real numbers, and infinitesimals are those sequences which tend to zero. The extension D^* of D simply consists of sequences (x_n) where each x_n is in D, and if y denotes the sequence (x_n), then for y in D^* we have $f(y)$ is the sequence $(f(x_n))$. It is clear that if f is continuous at x in the usual sense, then $x_n \to x$ implies $f(x_n) \to f(x)$, so if the sequence (x_n-x) tends to zero, so does the sequence $(f(x_n)-f(x))$, whence we see that

x in D, y in D^* and $y-x$ infinitesimal implies $f(y)-f(x)$ infinitesimal. The problem with this seemingly naive piece of interpretation is that the arithmetic of sequences is not easy to make into a field. We can define addition and mutiplication in the obvious way by doing the sums term by term. The sequence 1,1,1, ... acts naturally as a unit and if a sequence has all non-zero terms then we can define its multiplicative inverse by taking the reciprocal of each term. But if a sequence has any terms equal to zero we cannot do this. We are in an analogous situation to that of proper and improper infinitesimals! If a sequence tends to zero without actually getting there, we can invert it, but if it has zero terms we can't.

To eliminate this problem a subtle equivalence relation is put on sequences, so that hyperreals are actually equivalence classes of sequences with the system set up in such a way that it has an ordered field structure. It is the setting up of this subtle equivalence relation which makes the rigorous picture so hard.

In [2] Keisler gives another hyperreal field construction which produces a different field from that given by Henle and Kleinberg. This construction ties in neatly with our earlier ideas of infinitesimals as functions which tend to zero. Here one attempts to take the hyperreals to be functions defined on subsets of $\mathbb{R}$ with values in $\mathbb{R}$. These are added and multiplied pointwise, so that if f is defined on D and g is defined on E, then the product p is given by

$$p(x) = f(x)g(x)$$

and this is only defined on the interaction of D and E. The unit element is clearly $e:\mathbb{R} \to \mathbb{R}$ where $e(x) = 1$ for all x, and now the problems start again. What is the inverse of f? The function $1/f$ given by

$$(1/f)(x) = 1/f(x)$$

is clearly a candidate, but it is only defined for those values of x in D with $f(x)$ non-zero. The product of f and $1/f$ is also only defined for these very values of x and, in general, it cannot equal e which is defined for all real numbers.

Once again the problem is solved by constructing a subtle equivalence relation on a certain collection of functions so that in this interpretation a hyperreal number is an equivalence class of functions. The subtleties are not obvious at all. Or to put it another way, the theory has only been around for a few years and mathematicians have not yet had enough time to work the ideas to make the subtleties *seem* obvious.

What is interesting about this particular construction is that infinitesimals are now equivalence classes of *functions which tend to zero*. The old duality of meaning raises its head again. An infinitesimal in the hyperreals may be regarded as a point on an extended number line which is infinitesimal close to the origin. It has another interpretation in terms of a function which tends to zero.

Returning to the calculus of Leibniz and restricting ourselves to analytic functions (where $f(x+h)$ is always expressible as a power series in h), we find a fundamental link with the earlier remark about infinitesimals. An analytic function which tends to zero is a power series which must start with a positive power of h, say with a term ch^k plus higher powers of h (where c is non-zero):

$$f(h) = ch^k+dh^{k+1}+ \ldots$$

By writing this as

$$f(h) = h^k(c + dh + \ldots)$$

and noting that the power series $c + dh + \ldots$ tends to c as h tends to zero, we find that $f(h)$ is not zero in a neighbourhood of the origin apart from at the origin itself. Thus an analytic function which tends to zero is a *proper* infinitesimal. By allowing the first power k to be a negative integer, if necessary, such power series form a field. I now realise that this is at the root of the construction that I described in the article immediately preceding that of Peggy Marchi in *For the Learning of Mathematics* [5]. It is because every analytic infinitesimal is proper that we get a decent field structure and this is why the system I described is such a good match for Leibniz's calculus where every infinitesimal is also proper.

It becomes almost invidious to add at this juncture that in [5] I pointed out that there is a correspondence between infinitesimals considered as points in this field structure and infinitesimals considered as functions which tend to zero. The way in which all the strands fit together seems truly amazing.

To sum up then, this article brings forward the following major points:

(i) Provided that improper infinitesimals are suitably handled, the old infinitesimal calculus is mathematically sound.

(ii) In the calculus of Leibniz improper infinitesimals did not occur.

(iii) In the intuitive approach to beginning calculus using the dynamic limit method, unless complicated examples such as $x \sin(1/x)$ are purposely introduced, the ideas of improper infinitesimals do not occur.

(iv) It is not necessary to introduce improper infinitesimals to beginners.

(v) When they are introduced, they can be explained simply and the dynamic limit method remains satisfactory, provided that the independent variables are restricted to taking only proper infinitesimal values.

(vi) The dynamic limit method provides a natural introduction to epsilon-delta techniques and also provides basic intuitions for the modern infinitesimal method.

(vii) The processes and proofs in modern infinitesimal calculus are easy because they mirror cognitive processes and proofs as in the dynamic limit method. They are hard in the epsilon-delta approach because of the complicated computations and the many quantifiers required to formalize the dynamic limit process in the real numbers without resorting to infinitesimals.

(viii) The rigorous concepts of modern infinitesimal calculus are hard because of the difficulty of setting up the ordered field structure of the hyperreal numbers. They are made worse when the approach demands a perceptive use of logical language and first-order predicate calculus as a pre-requisite.

(ix) Finally I should add that the infinitesimal method sometimes promises more than it can deliver because its construction is based on the axiom of choice and is therefore non-constructive. As an illustration: the theory promises an extension from a sequence $s_0, s_1, s_2, \ldots, s_n, \ldots$ to give $s_H \in \mathbb{R}^*$ for infinite hypernatural numbers $H \in \mathbb{N}^*$; now take s_n to be the n_{th} decimal place in the expansion of π ($s_0 = 3$, $s_1 = 1$, $s_2 = 4$, $s_3 = 1, \ldots$) and ask the \$64,000 question: what is $s(H)$? There are thus philosophical implications in the use of non-standard analysis which still require consideration. They may cause genuine problems for the learner.

My own personal belief is that the best introduction to the calculus is through the dynamic limit method supplemented by examples from the numerical method and computer drawing of graphs. As well as giving valuable spatial intuition, the computer drawings show graphically (in both senses of the word) why it is that, when the derivative exists, for small values of δx, the increments δy and dy differ by an error of higher order. Hence the symbol dy/dx can take its original meaning as a quotient of lengths.

A later study of epsilon-delta techniques for mathematics majors, though it is difficult, is a valuable grounding in clear logical thinking. If a theorem is sloppily stated in epsilon-delta terms, it is usually wrong.

The new infinitesimal method has distinct possibilities for a more intuitive understanding of the finer points. However, the setting up of the field structure on the hyperreals is abstruse and mathematicians are put off at the moment by the necessity of studying first-order predicate calculus before beginning the theory. Keisler [2] has shown the way by launching an axiomatic approach. Following this lead, a course has been given for two years at Warwick University based entirely on set-theoretic axioms without the need for any initial discussion of logic or any deep theory about different types of mathematical language. There is an allied geometrical interpretation which allows one to look through 'optical microscopes', in which the graphs of differentiable functions when magnified look straight. The picture that one sees are like the computer drawings mentioned earlier. Thus the circle closes with the esoteric idea of infinitesimals giving the same pictures as the practical efforts of a computer using small numbers.

I believe that it is the search for this fundamental unity of ideas between abstract theory and practical reality which will prove the most fruitful. And whatever the balance of future developments, a blend of the dynamic limit method with practical numerical computations and high magnification drawings is likely to provide the most suitable grounding for beginning calculus students, to prepare them for future refinements.

References

[1] Henle, J. and Kleinberg, E. M. *Infinitesimal calculus*. MIT Press, 1979.

[2] Keisler, J. *Foundations of infinitesimal calculus*. Prindle, Weber and Schmidt, 1976.

[3] Marchi, P. Can heuristic be taught? *For the Learning of Mathematics* 1, 2: 35-42, 1980.

[4] Schwarzenberger, R. L. E. and Tall, D. O. Conflicts in the learning of real numbers and limits. *Mathematics Teaching* 82, 1978.

[5] Tall, D. O. The anatomy of a discovery in mathematics research. *For the Learning of Mathematics* 1, 2: 25-34, 1980.

[6] Tall, D. O. and Vinner, S. Concept image and concept definition in mathematics, with particular reference to limits and continuity. *Educational Studies in Mathematics* (in press) 1981.