Next to Nothing – a Single Paradigm

Why infinitesimals and limits are two aspects of the same thing (and always have been)

Version 2 – 5 December 2021, Mark C Marson
(Original published – 11 January 2019)

Preface

Since posting *Next to Nothing – a Single Paradigm* almost three years ago I have been able to compose a more complete version of the paper. I here address the main criticism of the original, elaborate on one of its themes, and offer some thoughts on a related matter. Improvements have also been made throughout the original text (while retaining the general narrative), while the Abstract and many of the References are new. The proof itself remains unchanged. The additional Continuation section follows the main text.

Abstract

I here tackle the most enduring controversy in mathematics, namely the question of what is the correct foundation for calculus. This has been taken to be either infinitesimals or limits at different times in history. I here give a novel proof (the nilsquare-limit theorem) that these concepts are two aspects of the same thing – indefinite precision. This contrasts with the prevalent opinion that the two methodologies are incompatible. The infinitesimals considered are nilpotent – a property uncontroversially possessed by infinitesimals before the 20th century. These are also the infinitesimals of smooth infinitesimal analysis (SIA) which I contrast with the more widely known discipline of non-standard analysis (NSA). I argue that these schools are equivalent in effect but that the former is more convenient. I give a graphical demonstration of the proof and a corollary which explicates the old idea of 'degrees of smallness'; and then use the new perspective offered by the proof to reinterpret the history of calculus, placing particular emphasis on Leibniz's efforts to justify his notation for calculus and Lagrange's later efforts to do the same. I mention the ancient antecedents of calculus (the Methods of Exhaustion and of Mechanical Theorems) in context. I then cover the crisis of foundations in mathematics in the late 19th and early 20th centuries with emphasis on the role (or lack thereof) of Cauchy, pathological functions, and the philosophies of Cantor and

formalism (also mentioning their antitheses – namely intuitionism and constructivism).

In conclusion I clarify the role of series in calculus, discuss how to reconcile the new paradigm with the law of excluded middle (LEM), and explain the close connection between this approach and finite difference calculus (FDC). In the Continuation I respond to a criticism of Version 1 by explaining how 'microlinearity' originally justified calculus, I elaborate on an 'increment free' approach to calculus pioneered by Carathéodory, and I address a related issue – namely how the absence of a properly algebraic approach to calculus for most of the 20th century led to widespread confusion about how calculus actually works i.e. I explain Leibniz's higher derivative notation and discuss a mistaken attempt to reformulate it. I then give an example of the use of differentials in ratios with an illustration. I finally conclude by appealing that the philosophical rift between most mathematicians and the rest of science be remedied. Other topics covered include: the attempts of the formalists to free mathematics from contradiction while also admitting the Axiom of Choice (**ref** 23); the need for FDC together with a simple numerical example (**refs** 24 to 26); and, one of the consequences of the period of hegemony enjoyed by formalism – namely the independent rediscovery of various aspects of its antithetical philosophies by various researchers (**ref** 37).

Introduction

To gain true understanding of a subject it can help to study its origins and how its theory and practice changed over the years – and the mathematical field of calculus is no exception to this. But calculus students who do read accounts of its history encounter something strange – the claim that the theory which underpinned the subject for two centuries after its creation was wrong and that it was then corrected, *in spite of the fact that the original theory never produced erroneous results*. I argue here that both this characterization of the original theory (infinitesimals) and this interpretation of the paradigm shift to its successor (limit theory) are false.

The paradigm shift in question took place in the late nineteenth and early twentieth centuries, and was accompanied by heated debate on the merits of the two approaches. They were supposed to have been reconciled in the 1960s with the invention of non-standard analysis (NSA), but this is a misrepresentation. The infinitesimals employed in NSA are a simplified version of infinitesimals as they were used until their near replacement with limits. Original infinitesimals always had one crucial property missing from those of NSA, namely they were nilpotent i.e. their higher powers were set to zero as they arose in derivations [1]. Since this property is, for our purposes, equivalent to being ‘nilsquare’ I mostly use that term here. Unfortunately, nilpotency was not adopted as an explicit rule from the start even though a contemporary of Leibniz advocated for this. The rule consequently became informal and often came under suspicion for being ‘non-rigorous’, and by implication liable to cause error. Mathematicians could however always claim that calculus by its nature builds curves from infinitesimal linear segments, and that therefore one cannot assume that the so-called law of excluded middle (LEM) applies to those curves (nilpotency is a corollary of this). But as the supporters of LEM gained influence in the late nineteenth century this position became less tenable; and to *their* mind the limit concept, sold as a complete alternative to infinitesimals, could finally make calculus rigorous.

I argue here that this development in the philosophy of mathematics was misguided – that original infinitesimals were only non-rigorous in the sense that they incorporated a number of deductive steps. Furthermore, those steps constitute a proof that *the criterion of the existence of a limit is met in general by expressions employing nilsquare infinitesimals*. That is to say, limits did not make calculus rigorous *per se*, rather they could have made original infinitesimals rigorous by exposing the deductive steps which had remained hidden. This concealment had not been deliberate – it was simply not assumed that LEM applies to the continuum. Consequently, sufficiently small (i.e. nilpotent) expressions could be ‘neglected’, an approach today known as smooth infinitesimal analysis (SIA). I now prove that original infinitesimals *are* in fact compatible with limit theory, contrary to the common opinion that the two approaches represent irreconcilable philosophical positions.

The Nilsquare-Limit Theorem

We begin by deriving the gradient equation. Since 1 = 1 and $y = y$ we have:

$$y + dy = y + dy$$

$$y + dy = y + dx\,\frac{dy}{dx} \qquad (1)$$

Note that we do not assume that *dy* and *dx* are anything other than variables i.e. the gradient equation is simply a property of the plane. If desired we can convert it from the Leibniz to the Lagrange notation (with ε instead of *dx* as the increment) thus:

$$f(x+\varepsilon)=f(x)+\varepsilon f'(x) \tag{2}$$

In this form [2] the equation is used as a starting point for the derivation of the theorems of calculus. However, on its own it is insufficient. Since ε is a finite variable the equation will yield the gradients of secants, resulting in finite difference calculus. What if we want to do standard calculus? One of our options is to 'take the standard part' (as in NSA) at the end of derivations by neglecting (i.e. setting to zero) the increment ε; remembering that it is not the case that ε is both equal and unequal to zero, it is simply indefinitely small. But a simpler way of achieving the same result is to employ the nilsquare rule as in SIA. For example, to derive the power rule from the gradient equation we do this:

$$(x+\varepsilon)^n=x^n+\varepsilon x^{n'}$$

$$x^n+nx^{(n-1)}\varepsilon=x^n+\varepsilon x^{n'}$$

$$x^{n'}=nx^{(n-1)} \tag{3}$$

The second equation results from applying the binomial theorem and then SIA's nilsquare rule i.e. $\varepsilon^{n>1} \rightarrow 0$. But although the two methodologies differ in their main technique they do both have a cancellation by ε to isolate *f'*(*x*) from ε near the end – this normalizes the associated term in SIA and saves it from nullification in NSA. But the nilsquare rule is not just a more convenient alternative to taking the standard part. It seems to imply that all the higher power incremental terms are indefinitely small in comparison with the first power term – otherwise how can it be justified in its *own* right? Let us test this conjecture, first we express the two sets of terms as a ratio:

$$r=\frac{\pm b\varepsilon^2 \pm c\varepsilon^3 \pm d\varepsilon^4 \pm \ldots}{\pm a\varepsilon}$$

The letters *a*, *b*, *c* and so on here represent terms involving the normally variable *x* – but since our proof works for arbitrary *x*, we here hold it constant and vary (i.e. indefinitely minimize) ε. Cancelling by ε yields:

$$r = \frac{\pm b\varepsilon \pm c\varepsilon^2 \pm d\varepsilon^3 \pm ...}{\pm a}$$

Any reduction in ε will now only affect the numerator, but we cannot assume that *r* will be reduced by a given reduction in ε because there are both positive and negative terms present – if the magnitude of the negative terms decreases more than that of the positive terms the value of *r* will increase. Does there always exist a smaller ε to overcome such increases? Since only the difference between the positive and negative sums is relevant to this question we can simplify the last equation thus:

$$r = \frac{p - n}{a} \tag{4}$$

where *p* is the sum of the positive terms and *n* is the sum of the negative terms. Since ε is indefinitely small we know that the magnitudes of both *p* and *n* can be as small as we like. (The denominator is made positive for simplicity i.e. multiplying either level of the ratio by minus would simply reverse the numerator terms without affecting the logic of the proof.) The effect of this is that the *range* defined by *p* and *n*, which contains (*p* – *n*), is always decreasing with ε, or algebraically:

$$p - -n = p + n$$

Subtracting amounts *j* and *k* from *p* and *n* respectively (to model unknown reductions in the sums of the positive and negative terms) gives us:

$$(p - j) - -(n - k) = (p + n) - (j + k) \tag{5}$$

which shows that the range always decreases with ε. So to get (*p* – *n*) and therefore *r* below given values we simply reduce the range until it is less than the (*p* – *n*) target value. Since by definition the range must include zero and (*p* – *n*), (*p* – *n*) will then be less than its target value. Consequently the ratio *r* can be made indefinitely small, which justifies neglecting higher power incremental (i.e. infinitesimal) terms. This process, in which the steady absolute increase or (as in this case) decrease of all individual terms overcomes any reversals in the change of their sum, can be termed *inexorable*.

What remains is to show that the above line of reasoning is equivalent to the limit criterion. Limits are a part of so-called real analysis, a more modern version of which is non-standard analysis. One advantage of NSA is that it clearly expresses the derivative as the 'standard part' of the gradient equation. Thus with dx as the increment of x:

$$f'(x)=st\left(\frac{f(x+dx)-f(x)}{dx}\right) \tag{6}$$

dx, and below dy, are used instead of the customary Δx and Δy because in this case dy/dx does itself refer to the finite version of the derivative. The limit criterion now requires that:

a limit exists if for every dx_2 in $|dy_2/dx_2-f'(x)|$ a smaller dx_1 can be found such that $|dy_1/dx_1-f'(x)|<|dy_2/dx_2-f'(x)|$

The 'smaller error' offered by a sufficiently smaller dx_1 is equivalent to the inexorable decrease of ($p-n$), and this justifies nilpotency. So the above statement applies equally to both NSA and calculus with original infinitesimals (SIA). This should not be surprising since the nilsquare rule and standard part operation have the same ultimate effect, the former is just concerned with taking an intermediate limit in a given derivation. (Incidentally, the above syntax also supports the idea that limit theory actually *presupposes* a value for the derivative $f'(x)$.) But can we express *that* action specifically in the language of limit theory? Using our previous notation for the constituent terms, the limit criterion now requires that:

a limit of zero exists if for every p_2 and n_2 in $|(p_2-n_2)/a|$ smaller p and n can be found such that $|(p_1-n_1)/a|<|(p_2-n_2)/a|$

This is what the proof shows for polynomials or functions that can be expressed in polynomial form; note that these are polynomials in terms of both the variable *and* the increment. This includes analytic functions – those which can be expressed using convergent power series. The result can be given symbolically as $(\Sigma\ b_n\varepsilon^{n>1})/(b_1\varepsilon) \to 0$; with ε as the increment, *b* as terms involving an arbitrary *x* value, and → meaning 'goes as close as desired to [zero]'. This completes the proof.

In summary, limit theory can be seen as a justification for neglecting nilpotent infinitesimal terms. But this immediately raises a question – how can the two approaches be in different philosophical camps? Maybe the dichotomy of LEM or not-LEM is too simplistic, as the use of LEM itself often is. The philosophy of this is discussed later, but for now it should be noted that the device of neglecting incremental terms has a simple geometrical interpretation. Recall that incremental terms are used if we are doing finite difference calculus and allow us to determine the properties of secants. For example, the length of a secant from a given point on a curve is a function of the increment i.e. the increment implies the length of the secant. The basic logical principle of contraposition now states that if there is no discernible increment there can also be no discernible secant length. What do we call a line defined on a curve by a secant with no discernible length? A tangent!

Addendum 1: The question naturally arises that if the sum of higher power incremental terms can be made an indefinitely small proportion of the first power incremental term, can we also make the sum of the higher-than-*n* power incremental terms an indefinitely small proportion of the *n*-or-lower power incremental terms? Yes, we can. First we divide both levels of the ratio by ε^n:

$$\frac{\pm d\varepsilon^{n+1} \pm e\varepsilon^{n+2} \pm \ldots}{\pm a\varepsilon \pm b\varepsilon^2 \pm \ldots \pm c\varepsilon^n} / \frac{\varepsilon^n}{\varepsilon^n} = \frac{\pm d\varepsilon \pm e\varepsilon^2 \pm \ldots}{\pm a\varepsilon^{1-n} \pm b\varepsilon^{2-n} \pm \ldots \pm c\varepsilon^{n-n}} \qquad (7)$$

Then we note that the numerator can be made indefinitely small by invoking inexorable reduction. Also note that the denominator consists of one constant and a series of negative power incremental terms. Negative power terms increase as their variable decreases, but since they are here part of the denominator this inexorably reduces the value of the ratio. Therefore, since both levels are changing inexorably with the effect that the ratio is decreasing we can say that the numerator is infinitesimal. Note that this idea is the same as that found in some older textbooks regarding 'degrees of smallness'; but since a presumably arbitrary increment with its associated terms would remain after a higher-than-*n* nullification this technique cannot be considered as important as the nilsquare rule.

Addendum 2: An example of the reduction of an incremental term only producing a better approximation of the tangent after further reductions, can be found on the graph of $y = 24x^3 + 8x^2$ between -⅓ and ⅓. The standard derivative of this is $72x^2 + 16x$ while the finite derivative is $72x^2 + 16x + 72\varepsilon x + 8\varepsilon + 24\varepsilon^2$ (with ε as the increment). Taken from -⅓ the value of the finite derivative is 2⅔ with an increment of ⅔, and 0 with an increment of ⅓. Since the value of the standard derivative at -⅓ is also 2⅔ we can see that the derivative *gets worse as the increment decreases from ⅔ to ⅓*.

Figure 1: Graphs of $y = 24x^3 + 8x^2$ (purple) and $y = 8/3\ x + 8/9$ (orange). The latter is the tangent of the polynomial at $x = -1/3$ and also a secant. The secant for the increments of 1/6 and 3/6 ($y = 2/3\ x + 2/9$, broken orange) is also shown.

Calculus as it Began

Realizing that much of the controversy over the foundations of calculus was caused by a misunderstanding allows us to re-evaluate various episodes in the history of mathematics – here is a brief attempt to do that.

Calculus as we know it was created in the seventeenth century by Gottfried

Leibniz and Isaac Newton and was a consequence of various preceding mathematical innovations. In particular the invention of Cartesian coordinates (named after Rene Descartes) naturally led to a new focus by mathematicians on functions and their graphs; and although this can be done geometrically [3] calculus makes it simpler. Descartes simplified things further by advocating that mathematicians focus on algebraic not mechanical curves [4]; and Pierre de Fermat coined the word ‘adequality’ for the relationship between infinitesimals and their proximate points. But the immediate precursor to calculus proper was quite clearly the work of Isaac Barrow, particularly the idea of the differential (or ‘characteristic’) triangle.

Leibniz was initially more open than Newton about his innovation, saying:

> What is best and most convenient about my new (infinitesimal) calculus is that it offers truths by a kind of analysis and without any effort of imagination, which often only succeeds by chance, and that it gives us over Archimedes all the advantages which Vieta and Descartes had given us over Apollonius [5].

Leibniz’s notation became standard, but the logical basis of his method was criticized. One of his replies was:

> For instead of the infinite or the infinitely small, one takes quantities as large, or as small, as necessary in order that the error be smaller than the given error, so that one differs from Archimedes' style only in the expressions, which are more direct in our method and conform more to the art of invention [6].

When Leibniz refers to “Archimedes’ style” he is almost certainly referring to the Method of Exhaustion (also used by Euclid) not the Method of Mechanical Theorems, since the latter had been lost in antiquity and an account of it was only rediscovered in 1906. The former method is considered to be equivalent to limits whereas the latter is considered to be equivalent to infinitesimals. So Leibniz is saying that his method is equivalent to limits but is more convenient [7], which could easily be said about infinitesimals. So do the *dy* and *dx* in his notation actually represent infinitesimals? If so then one of his contemporaries thought they could be improved on – Bernard Nieuwentijt suggested that terms with higher powers of infinitesimal increments should be neglected in derivations as they arise. In response Leibniz replied:

> it is rather strange to posit that a segment *dx* is different from zero and at the same time that the area of a square with side *dx* is equal to zero.

As John L Bell notes [8] Leibniz could be accused of contradiction here since the nilsquare property is a consequence of the principle of ‘microlinearity’, which Leibniz *did* accept. (Consider $y = x^2$ around $x = 0$. If the curve is microlinear there must be a small straight segment around zero containing small ‘non-zero nilsquare values’.) What we can now say is that he could also be accused of contradiction because, as proven here, the nilsquare property *is entirely compatible with his own conception of limits* (his clarification in the second quote is equivalent to the nineteenth century definition).

This must be considered one of the great missed opportunities in the history of science, because although the use of nilsquare infinitesimals soon became standard practice, they were considered informal. This was despite the fact that they always yielded correct results. The issue came to a head around the year 1900 when a majority of mathematicians decided to reject such perceived informalities; and from then on infinitesimals were subject to a self-imposed prohibition by academia (though they were still ‘unofficially’ used in physics). This would have been inconceivable if Leibniz had explicitly advocated for their use. But why didn’t he?

Probably for the same reason as other mathematicians – a reticence to accept ideas that violate or seem to violate the law of excluded middle [9]. A variable which by definition is smaller than any value you can state manifestly cannot be distinct from zero or necessarily equal to zero. Nieuwentijt was not therefore saying that the increment’s square is *literally* zero (simplistically $x^2 = 0$ implies that $x = 0$) so how could it be right to set it to zero? As mentioned before, one justification that could be used for nilpotency in the two centuries after Leibniz and Newton was microlinearity [10]. But the example given of a linear segment on $y = x^2$ around $x = 0$ implying that $dx^2 \to 0$ suggests a more general result. Namely that the curves of smooth functions consist everywhere of indefinitely small linear segments and that we can use nilpotent infinitesimals to ‘generate’ those linear segments. This was the kind of thinking evident in what is considered to be the first textbook on differential calculus, published by Guillaume de L’Hopital in 1696:

> For, as curves are nothing but polygons with an infinity of sides, and are only distinguished from each other by the difference of the angles that these infinitely small sides form with

> each other; only the Analysis of the infinitely small can determine the position of these sides and so obtain the curvature which they form, which is to say the tangents of these curves [11]

Now though ‘indefinitely’ should be used instead of ‘infinitely’ for the sake of convention. One recent commentator summarized the method used in the book thus:

> The basic differential formulas for algebraic functions – sums, products, quotients, powers, and roots – are derived by L'Hopital in the customary manner, infinitesimals of higher order being neglected [12].

But unfortunately, since neglecting higher order infinitesimals *seemed* to be an approximation, the doubts persisted. One notable attempt to clarify the issue was made by Joseph-Louis Lagrange a century later:

> A strikingly new treatment of the fundamental conceptions of the calculus is exhibited in J.L. Lagrange's [Theory of Analytical Functions], 1797. Not satisfied with Leibniz's infinitely small quantities, nor with Euler's presentation of dx as 0, nor with Newton's prime and ultimate ratios which are ratios of quantities at the instant when they cease to be quantities, Lagrange proceeded to search for a new foundation for the calculus in the processes of ordinary algebra. Before this time the derivative was seldom used on the European Continent; the differential held almost complete sway. It was Lagrange who, avoiding infinitesimals, brought the derivative into a supreme position. Likewise, he stressed the notion of a function [13].

He consequently introduced the now eponymous Lagrange notation. But although this was seemingly part of an attempt to de-emphasize infinitesimals, Lagrange later said:

> I have kept the ordinary notation of the differential calculus because it fits the system of infinitesimals adopted in this treatise. Once the spirit of this system has been grasped well and the accuracy of its results established by either geometrical methods or by the analytical method of derived functions, the infinitesimal calculus can then be applied as a certain and manageable tool to shorten and simplify the demonstrations. It is in this way by using the method of indivisibles that the demonstrations of the Ancients are shortened [14].

Yet the controversy was still not over. Toward the end of the nineteenth century doubts began to re-emerge – and the way mathematicians had explained their thought processes was also disputed. Could it really be true that they had not genuinely known what they were doing?

Calculus Under Scrutiny

The late nineteenth century witnessed a dispute over the foundations of mathematics where previously uncontroversial notions, such as the nature of the continuum, were challenged; and by the early twentieth century this had cast doubt on calculus. As William Osgood said in 1907:

> Thus mathematicians have necessarily discarded the differentials of Leibniz as the elements out of which the calculus can be built up, and some are more than doubtful about the advisability of retaining them in any form… We sometimes hear it said that hardly a theorem in our textbooks on the calculus is true as stated there [15].

He does however go on to defend the careful use of infinitesimals for practical purposes. What exactly caused this crisis of confidence? Three possibilities are covered here.

A common narrative is that Augustin Cauchy in the first half of the nineteenth century found an alternative to the informality of infinitesimals by clarifying the limit concept. This claim must however be questioned since Cauchy seems to have been perfectly comfortable using infinitesimals, saying:

> When the successive numerical values of such a variable decrease indefinitely, in such a way as to fall below any given number, this variable becomes what we call *infinitesimal*, or an *infinitely small quantity*. A variable of this kind has zero as its limit [16].

As previously noted, today they would be described as *indefinitely* small quantities – Carl Gauss himself insisted on this distinction [17]. Cauchy goes on to define infinite numbers in a similar way. But although Cauchy was comfortable with infinitesimals he did make use of expressions such as [18]:

$$\frac{f(x_1)-f(x_0)}{x_1-x_0} \tag{8}$$

which *can* be used to derive standard calculus in an algebraic fashion *without an incremental or infinitesimal term* [19]. An obvious drawback of this method is that it is no good for finite difference calculus, but arguably it is also not as intuitive as using infinitesimals, and its more overt use of the ratio 0/0 may have hindered its adoption. Cauchy’s general approach to calculus was however very influential. Austrian mathematician Otto Stolz had this to say about it in 1881:

> Cauchy relied on infinitesimal calculus, abandoning the limits of the method of Lagrange, believing that only infinitesimal methods provide the necessary rigor. [The] clarity and elegance of its presentation facilitated the widespread and universal adoption of his course. Even significant shortcomings [when] found, as time has shown, can be eliminated by the adoption of consistent principles based on Cauchy's arithmetic considerations. A few years before Cauchy these same views [were] sometimes substantially more fully developed by Bernard Bolzano [20]

Since circa 1907 the initial assertion in this account has clearly contradicted the 'official' story of Cauchy's thought processes, a story which also casts Bolzano as one of the progenitors of limit theory. But Stolz, speaking decades before the supporters of LEM gained hegemony, would have disagreed with the later narrative and casts both mathematicians as trying to develop a rigorous theory of calculus based on infinitesimals.

A second possible cause of the crisis was investigations into so-called pathological functions. These are functions that could not be analyzed with the usual techniques. Henri Poincare had this to say about them in 1899:

> Logic sometimes makes monsters. For half a century we have seen a mass of bizarre functions which appear to be forced to resemble as little as possible honest functions which serve some purpose… Indeed, from the point of view of logic, these strange functions are the most general; on the other hand those which one meets without searching for them, and which follow simple laws, appear as a particular case which does not amount to more than a small corner… In former times when one invented a new function it was for a practical purpose; today one invents them purposely to show up defects in the reasoning of our fathers and one will deduce from them only that [21].

The most well known of these is the Weierstrass function (introduced by Karl Weierstrass in 1872) which is continuous but not differentiable. Infinitesimals, suited as they are to polynomials, would have seemed inadequate in this new terrain; and so limits, being more general, would have gained favor. But unfortunately Weierstrass' epsilon-delta limit criterion (which was essentially the same as that of Leibniz) began to take on a more exclusionary role. Since this change was seen (by those advocating it) as a continuation of Cauchy's approach there may have been a temptation to 'backdate' the new philosophy so as to imbue it with more authenticity [22].

The third possibility is that the crisis was a side effect of the introduction of Georg Cantor's theory of sets and cardinal (i.e. actual infinite or transfinite) numbers beginning in 1874 (motivated by his take on 'points' on the number line). Cantor initially justified his thinking by invoking the so-called Axiom of Archimedes. This axiom states that if we have two positive numbers a and b for which $a < b$ then there exists a number n such that $na > b$. It could be argued that this axiom precludes both infinitesimal and infinite numbers, but Cantor seems initially to have ignored the second possibility. Later though (after 1883) he sought to base cardinal numbers on his so-called well-ordering principle; and in 1908 after much dispute Ernst Zermelo introduced the Axiom of Choice (AC) which can be used to justify the well-ordering principle, but also implies LEM for the continuum. Within this framework then original infinitesimals had to be disallowed; indeed Cantor seems to have always assumed that this was the case and had conducted a long and angry anti-infinitesimal campaign, calling them the "infinitesimal Cholera bacillus of mathematics". Meanwhile AC was proving useful to mathematicians of the era, although this may have only been because more time was needed to develop a better approach. For example, in 1942 Paul Bernays introduced the so-called Axiom of Dependent Choice, which apparently can be used to develop most of real analysis while *not* implying LEM for the continuum. Serious drawbacks to AC also emerged. For example, it implies results such as the Banach-Tarski paradox (1924) which, put simply, states that a sphere can be decomposed into a finite number of pieces which can then be assembled into two spheres each identical to the original. It could be argued that Cantor's philosophy should have been sidelined after this revelation (since it is not actually relevant to any other branch of natural philosophy) and in 1925 Hermann Weyl advocated this:

> Mathematics attains with Brouwer the highest intuitive clarity; his doctrine is idealism in mathematics thought through to the end. But with pain the mathematician sees the larger part of his towering theories fall apart [23].

Weyl is here saying that an alternative philosophy (the intuitionism of Brouwer, see refs 2 and 37) clarifies the foundations of mathematical analysis better than the philosophy of formalism, which by then had fully incorporated Cantor's work; but that if we embrace it some previously obtained results have to be abandoned so as to preclude contradiction. Some mathematicians did take this advice, and in time analysis was reconstructed on firmer ground, culminating in the emergence of SIA in the 1970s. But by then a large majority of mathematicians and mathematics departments had embraced AC, together with all of its consequences.

Conclusion

Whatever the reason, the use of infinitesimals came to be in effect discontinued within academic mathematics soon after 1900, and it became obligatory to refer to limits in relevant proofs. It was assumed that limits are *without qualification* compatible with LEM. But what would that really mean? Why should LEM even be an appropriate condition for calculus? It is worth noting that while some types of limit in mathematics consist of terms to be summed in theory simultaneously (such as decimal numbers), expressions in calculus are evaluated in theory repeatedly – each iteration is an 'improved' version of the previous calculation and is independent of it. In other words, calculus is not *about* infinite series (although these are often utilized), it is about indefinite processes. But what 'improved' means here must be clarified. If we are trying to calculate the gradient of a tangent of a smooth curve as a limiting value then we are not guaranteed continual improvement with a decreasing incremental term. Instead what we are guaranteed is *inexorable* improvement. This ability to posit, but not specify the value of, an increment with a desired property (namely that it yields an error less than an arbitrary value) justifies the use of infinitesimals in calculus and explains what limits are actually limits of [24]. The question then becomes – does LEM prohibit us from subsequently neglecting those small inappreciable values as it does those which are finite?

The answer *should* depend on exactly how the condition is phrased. One condition that cannot be violated it that of non-contradiction – a number cannot be both equal and unequal to another number. This would imply that the answer is No – neglecting infinitesimals is not prohibited because they are *indistinguishable* from their proximate values, equality is not the issue. Some mathematicians though have extrapolated from this to assert that even though we do not in practice distinguish infinitesimals, 'in theory' we could and that therefore the answer is Yes – we must prohibit the technique of neglecting them. The problem here is the use of the word ‘therefore’. We could just as easily say that since infinitesimals are indefinitely small we *are* allowed to neglect them, provided we can cancel any remaining infinitesimals. This unique capability of calculus, together with regular algebra, gives us a widely applicable set of tools for obtaining useful theorems.

The important point is that apart from when we are invoking the special properties of infinitesimals the normal rules of algebra apply, so we do not have to treat finite and standard calculus as radically different. As Felix Klein put it:

> I should like to remind you, first of all, that the bond which [Brook] Taylor established between difference calculus and differential calculus held for a long time. These two branches always went hand in hand, still in the analytical developments of [Leonhard] Euler, and the formulas of differential calculus appeared as limiting cases of elementary relations that occur in the difference calculus [25].

This connection is most apparent when physicists model phenomena using finite differences – the programs approximate equations with very small (but not infinitesimal) incremental terms to any required precision [26]. Not coincidentally physicists and engineers are largely responsible for ‘unofficially’ maintaining some of the original techniques of standard calculus throughout the twentieth century, in spite of the unwarranted accusation of lack of rigor from academic mathematics. One example of this is perturbation theory, which explicitly introduces infinitesimal increments to equations describing physical phenomena in order to study continuous change (more generally this technique can be called ‘microadditivity’). Unfortunately though, some of the more fundamental original techniques (which of course includes basic proofs) were seldom to be found in standard textbooks. Instead authors employed a sometimes awkward mixture of truncated infinitesimal algebra while also referencing the logic of limits [27]. This self-imposed censorship was only alleviated when the internet allowed alternative viewpoints to be widely expressed.

So what should be done about this situation? Perhaps, instead of dwelling on what was at the very least a pedagogical disaster, we should just correct it. Acknowledging that infinitesimals and limits are two aspects of the same thing, that they are both aspects of dealing with indefinite precision, would be a good place to start.

Continuation

Reply to Critique – Was Change Needed?

There have been no criticisms of the proof itself, the stating of which was the main purpose of the paper, but there was a criticism of one seeming implication of the narrative. The idea that before the twentieth century there *was* a "rigorous theory of infinitesimals" and that the advent of limit theory did not therefore represent any improvement in our understanding of calculus was challenged. This criticism is not wrong but *is* misleading – because it may not have been clear that such a theory was necessary and because mathematicians very nearly did have such a theory. Taking these two arguments in turn:

1. Before the twentieth century mathematicians used the techniques of neglecting higher order infinitesimals *and also* that of what would become known as taking the standard part. But as explained in Reference 27 (which was in Version 1 of the paper) these are in effect equivalent. So until the dispute over foundations restricted their options there would have been no reason for mathematicians to object to one technique while making use of the other, and whereas the former requires some theory the latter can be justified quite simply – if the infinitesimals of an expression can be seen as the subject of a polynomial then simple graphing shows that the sum of the terms near zero approaches zero. So to many mathematicians it may not have been clear that a "rigorous theory of infinitesimals" was actually needed.

2. Before formal limit theory was applied to calculus mathematicians mostly assumed that the subject was based on the concept of microlinearity. That approach was outlined by L'Hopital (see **ref** 11) and had been employed by, for example, Fermat in his technique of 'adequality'. This equates the values of a function before and after increasing it infinitesimally – if the curve being analyzed has a gradient of zero for a given *x*-value the points could be said to lie on a linear and horizontal segment of the curve. From this perspective tangents do not touch the

curve at a point, rather they coincide with the curve along an infinitesimal straight line. And this idea can be used in another simple proof of the so-called definition of the derivative. If two linear equations are formed with the same parameters but one with y_1 and x, while the other has y_2 and ($x + \varepsilon$), then we can easily show that the derivative is the slope of the line.

Is this a rigorous theory or a tautology? That line could of course be a secant *or* a tangent of the function – to be a tangent we would have to neglect and cancel the infinitesimal terms, and the nilsquare rule allows us to do it in that order. In the process it also makes the 'error' (i.e. the sum of the terms which cannot be standardized by cancellation near the end) indefinitely small by making the higher order terms indefinitely small as a proportion of any first order term. This is reminiscent of first order Taylor 'approximations' in which the relationship between microlinearity and the nilsquare rule is very clear [28].

So what good does limit theory offer then? Acknowledging 'linearization' as above means that we can retain the term 'taking the limit' because applying the nilsquare rule *is* taking a limit – the limit at which a curve 'becomes' a line. The ratio of terms neglected to the term kept in order to do that can be as low as we like, but *ideally* this should have been demonstrated. So the advocates of limit theory did correctly diagnose a minor problem, they just never bothered to come up with a solution to it. Instead they strove to create "a purely arithmetical and perfectly rigorous foundation for the principles of infinitesimal analysis" [29]. The term sometimes used for the mathematicians involved in that movement is 'arithmetizers', but it actually makes more sense to identify them with the formalist school. As David Hilbert, the chief formalist, said in 1926: "No one shall expel us from the paradise which Cantor has created for us". Formalism is normally considered to have suffered a fatal blow with the work of Gödel, but arguably it lingers on in the archaic insistence on the primacy of limits in calculus.

Elaboration – Alternative Algebras for Calculus

In the section Calculus Under Scrutiny I give an expression (Formula 8) used by Cauchy which "*can* be used to derive standard calculus in an algebraic fashion *without an incremental or infinitesimal term*." This does not however seem to have been done by Cauchy himself but mostly in the early twentieth century by Greek mathematician Constantin Carathéodory. In Reference 19 a proof of the power rule was given that used this method, and the other basic theorems of calculus can be derived from the same starting point. In his paper The Derivative á la Carathéodory [30] author Stephen Kuhn introduces his method as follows:

> Augustin-Louis Cauchy would be pleased. Each year we introduce our elementary analysis students to the notion of the derivative essentially as he gave it to us in 1823. But there is another, less well known, characterization of the derivative which appears in the last textbook [31] written by Constantin Carathéodory (1873-1950). This formulation is not only elegant but useful… The proofs of many important theorems… become significantly easier… [and later he continues] After the usual definitions and theorems about limits and continuity are presented in a standard elementary real analysis course, the definition of the derivative, essentially as given to us by Cauchy, is given for functions of a single variable. Typically it is presented in both the following forms…

He then gives the formulas given here as Formula 6 (but with a limit prefix instead of a standard part function) and Formula 8 (but including a subject). It should however be noted that: firstly, Kuhn fails to mention that Cauchy initially equates his derivative definition to $\Delta y/\Delta x$ (he thus obscures the fact that standard calculus ‘follows on’ from finite difference calculus); and secondly, that since the main proof in this essay begins with an utterly trivial proof of Taylor’s formula it is far from clear why it should only be stated “After the usual definitions and theorems… are presented in a standard elementary real analysis course”.

Kuhn is however correct when he says that Carathéodory’s method is “less well known”, and consequently it is not uncommon for students to stumble upon it while experimenting. As one anonymous internet contributor put it (in January 2019): “The limit actually does do something!” before giving the following proof of the derivative of $y = x^2$:

$$h(a)=10\,a^2$$
$$h(b)=10\,b^2$$

$$\frac{h(b)-h(a)}{b-a}=\frac{10\,(b^2-a^2)}{b-a}=\frac{10\,(b+a)(b-a)}{b-a}=10\,(b+a)$$

They then say that as *b* approaches the limit of *a*, 10(*b*+*a*) = 10(2*a*). Another contributor independently used this method (in December 2015) to derive the above result, the generalized power rule (as in **ref** 19), the addition rule, the product rule, and the chain rule as shown here [32]:

$$\frac{f(g(b))-f(g(a))}{b-a}=\frac{f(g(b))-f(g(a))}{g(b)-g(a)}\times\frac{g(b)-g(a)}{b-a} \tag{9}$$

The author, who only gives a pseudonym, presents their proofs in a different way to Carathéodory but the two do start working from the same premise, and arguably the above version is simpler. But as alluded to in Version 1 this method is (to quote Kuhn again) ”less practical from a computational point of view” since to do numerical analysis and utilize various algebraic techniques you do need a variable for the increment. It is also worth noting that the method’s apparent status as *the* algebraic manifestation of limit theory could also be claimed by the method of ‘tangent cones’ as described here by another online mathematician [33]:

> In calculus classes it is sometimes said that the tangent line to a curve at a point is the line that we get by "zooming in" on that point with an infinitely powerful microscope. This explanation never really translates into a formal definition… I seem to have found a way to obtain tangent lines (and more) by taking "zooming in" seriously… Take the curve $y = x(x - 1)(x + 1)$. I want to find an equation for the tangent line to this curve at the origin. So I zoom in on the origin with a microscope of magnification power c (i.e. I stretch both vertically and horizontally by a factor of c) to obtain

$$\frac{y}{c}=\frac{x}{c}\left(\frac{x}{c}-1\right)\left(\frac{x}{c}+1\right) \tag{10}$$

$$y=x\left(\frac{x}{c}-1\right)\left(\frac{x}{c}+1\right)$$

$$y=-x$$

The last line is obtained by “letting my magnification power go to infinity”; the result obtained is the equation of the tangent, not its gradient. He ends by asking “Do any books take this approach when developing the derivative?” They do, but as far as I can tell the above presentation is simpler than the usual textbook approach.

It should be apparent by now that there are several different *algebraic* ways of doing calculus. The main lessons I draw from the comparison of them given here are this: algebra and analysis are not distinct branches of mathematics, and that algebraic thinking is what characterizes mathematics itself. The algebraic versions of calculus should not contradict each other and, if treated carefully, can be used however we like.

Related Matter – Ratios of Differentials

Whether infinitesimals are used explicitly or are represented merely by the difference between two values we should remember that it is by subjecting them to indefinite (and inexorable) reduction that we get calculus to work. There was nothing arbitrary about Barrow's choice of the word 'differential' for his increments, or of Leibniz's decision to adopt the term in his work. We should therefore be able to understand other aspects of Leibniz's notation by referring back to such basic concepts; and as a corollary we might expect the formalists, who were dismissive of the Leibniz notation, to run into trouble with it. Or, considering the pervasive influence of limit theory on twentieth century calculus teaching, maybe regular people have got into trouble after trying to find coherent explanations of the topic. I quote alleged mathematician Jonathan Bartlett:

> However, when it came to the second derivative, I realized that not only is the notation unintuitive, there is literally no explanation for it in any textbook I could find… [the notation is of course d^2y/dx^2] I looked through 20 (no kidding!) textbooks to find an explanation for **why** the notation was the way that it was. Additionally, I found out that the notation itself is problematic. Although it is written as a fraction, the numerator and denominator cannot be separated without causing math errors… I would try to derive the notation myself. Well, when I tried to derive it directly, it turns out that the notation is simply wrong… I would argue that a fraction that can't be treated like a fraction *is* wrong [34].

This is symptomatic of a near-total philosophical train wreck, and it is entirely the fault of the formalists and their attempt to prohibit the various algebraic approaches to calculus. Admittedly, the author does have a motive to find fault (see **ref** 34) but he does have a point when he says that textbooks do not explain the notation. The reader should not be surprised to hear that *this* author has no objection to differentials being used in fractions. Furthermore, I worked out an explanation for Leibniz's second derivative notation in 2016, well before this essay was written:

> The d in the Leibniz notation means 'the difference between that value and the previous one.' It is not a factor, hence:
>
> $$\frac{dy}{dx}=\frac{dy_1}{dx}=\frac{y_1-y_0}{dx}$$
>
> Notice that we have to be careful with our subscripts – the first y value is subscripted zero because it corresponds to $dx = 0$ ([by definition] dx is just arbitrarily small). So for the second derivative we have:
>
> $$\left(\frac{dy_2}{dx}-\frac{dy_1}{dx}\right)/dx=\frac{dy_2-dy_1}{dx^2}=\frac{d(dy_2)}{dx^2}=\frac{d^2y}{dx^2} \tag{11}$$
>
> Notice that the superscript means 'apply the function twice' not square it! Seeing if this applies to higher derivatives we have:
>
> $$\left(\frac{d^2y_3}{dx^2}-\frac{d^2y_2}{dx^2}\right)/dx=\frac{d^2y_3-d^2y_2}{dx^3}=\frac{d(d^2y_3)}{dx^3}=\frac{d^3y}{dx^3}$$

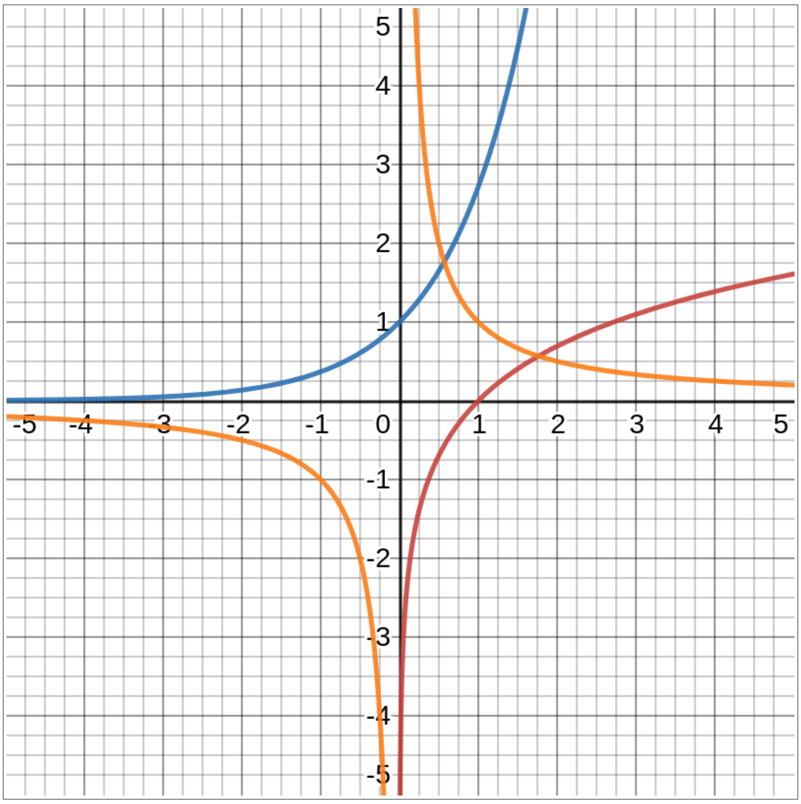
(This was posted on https://math.stackexchange.com at the time but was subsequently deleted by the moderators.) The easiest way to *visualize* how this notation works is to sketch adjacent differential triangles on a graph; the connection to finite differences could not be more obvious. After criticizing this notation Bartlett goes on to invent his own – for the second derivative this somehow consists of the conventional notation minus a complex ratio of differentials. To state the obvious, I consider such expressions invalid. The ratios in the conventional differential notation *are* genuine ratios, and considering them as such facilitates elegant arguments.

One example will be given here. For $y = e^x$ we have $dy/dx = y$. The inverse function of e^x (i.e. $ln|x|$) is its reflection in $y = x$ and obviously $ln|x|$ has the same relationship with x as e^x has with y. Therefore $dx/d(ln|x|) = x$ and $d(ln|x|)/dx = 1/x$. By FTC (as in **ref** 12) we then have $ln|x| = \int 1/x + c$ where ∫ signifies the integral. This proof may seem informal but we can achieve the same result by using the inverse function rule. If $f(x)$ is an invertible function we have $x = g(f(x)) = f(g(x))$:

$$\frac{d\,g(x)}{dx} = \frac{1}{d\,f(g)/d\,g(x)} \qquad (12)$$

This is a direct translation of the rule from the Lagrange to the Leibniz notation as done by the author – external sources gave inadequate representations. Note that the rule, as given above, can be derived using simple algebra (as was done for Formula 1) starting with the definition of inversion. The derivative of *ln*|*x*| can easily be found using the rule and is left as an exercise. The above example is historically important because mathematicians struggled for many years with the 'quadrature of the hyperbola', or as we would say, integrating the reciprocal [35]. It is also an example of a difficult proof being simplified by a clever symmetry argument, an occurence not uncommon in mathematics.

Figure 2: Graphs of $y = e^x$ (blue), $y = ln(x)$ (red), and $y = 1/x$ (orange). The first two are reflections in $y = x$, while the integral of $1/x$ is $ln|x| + c$.



The absence of any 'official' explanation of the Leibniz notation for higher derivatives (although there is now an attempted explanation on Wikipedia) is just one example of the obfuscation wrought by the formalists' insistence that limit theory is the one true lens through which to view calculus [36]. This doctrine downplays the close connection between standard and finite difference calculus, suppresses any explanations of calculus that (to their mind) seem too algebraic, and ignores versions of analysis which do not apply LEM to the continuum. Mathematicians who could not accept this approach embraced the philosophy of constructivism, which emerged in the early twentieth century, largely in response to formalism. Physicists meanwhile have mostly ignored the controversy [37], although various intuitive derivations were suppressed due to self-censorship. This stance is possible partly because the predominant method for deriving formulae in physics is to use variational principles, the most developed form of which is the Calculus of Variations (which employs the Euler-Lagrange equation). This largely superseded Newtonian mechanics, in which microadditivity is more explicit; the two methodologies are however equivalent [38]. Computer scientists meanwhile seemingly employ nilsquare infinitesimals as a matter of routine [39].

To conclude, there is a profound philosophical rift between most mathematicians and the rest of science, which obviously needs to be dealt with. The time has come for a thorough re-evalutation of the influence of the formalists – we must reject arguments claiming that calculus had previously been defective and then subsequently became better. The truth is the exact opposite of that.

References

(1) **p2** This rule actually predates calculus proper:

> Thus [Isaac] Barrow employs the concept of the 'characteristic triangle' – essentially the idea of the tangent line as the limiting position of the secant line as [side] *a* and [side] *e* approach 0 – and takes the limit by the expedient of neglecting 'higher order infinitesimals'.

The Historical Development of the Calculus, Charles H Edwards, 1994, p133.

(2) **p4** Anders Kock calls it Taylor's formula – obviously after Brook Taylor who introduced Taylor series in 1715 as a way of using derivatives to analyze a broader range of functions than was previously possible. The usage is perhaps due to FW Lawvere, one of the founders of synthetic differential geometry (developed in the 1970s) – a field that includes smooth infinitesimal analysis. The underlying philosophy of this approach is called constructivism, an early version of which was the so-called intuitionism of LEJ Brouwer (see also **ref** 37). **Synthetic Differential Geometry**, https://users-math.au.dk/kock/sdg99.pdf, Anders Kock, 2006, p7 (Second Edition).

(3) **p9** – **The Lost Calculus (1637-1670): Tangency and Optimization without Limits**, Mathematics Magazine, https://www.maa.org/sites/default/files/pdf/upload_library/22/Allendoerfer/suzuki339.pdf, Jeff Suzuki, 2005 (Volume 78). Suzuki mainly credits Dutch mathematician Jan Hudde with this development:

> in the years between 1637 and 1670, very general algorithms were developed that could solve virtually every "calculus type" problem concerning algebraic functions. These algorithms were based on the theory of equations and the geometric properties of curves and, given time, might have evolved into a calculus entirely free of the limit [or infinitesimal?] concept… Rather than being heralds of the calculus that was to come, Hudde's results are instead the ultimate expressions of a purely algebraic and geometric approach to solving the tangent and optimization problems.

The existence of this approach, together with classical geometry, may shed some light on Newton's sparing use of calculus proper in *Principia Mathematica*. The fact that calculus represented a simpler alternative for solving those problems soon became undeniable though. Suzuki also credits John Wallis with providing groundwork for the fundamental theorem of calculus (FTC) by reviving a technique developed by Apollonius of Perga (in connection with conic sections) which implies “a useful link between the derivative and the integral”. One of Wallis’ other claims to fame is that he introduced the ∞ symbol for infinity and 1/∞ for infinitesimals. (If ∞ is replaced with 1/ε in the definition of *e* we get, after rearranging, the first order ‘approximation’ of the Taylor series of *e.*)

(4) **p9** Descartes wrote that mechanical curves:

> belong only to mechanics, and are not among those curves that I think should be included here, since they must be conceived of as described by two separate movements whose relation does not admit of exact determination.

Geometry Book II – On the Nature of Curved Lines, https://www.gutenberg.org/files/26400/26400-pdf.pdf, Rene Descartes, 1637, p11-12. Quoted in **Mathematics and Its History**, John Stillwell, 2010, p256. Note that by using computers mechanical curves (and mechanical devices) can be simulated to any required precision.

(5) **p9** Quoted in **G.W. Leibniz: Critical Assessments**, R.S. Woolhouse, 1993, p428 (Volume 2).

(6) **p9** – **Justification of the Calculus of Infinitesimals**, Gottfried Leibniz, 1701. Quoted in **The Calculus in the Eighteenth Century**, Henk JM Bos, 1975, p56.

(7) **p9** Newton said something similar:

> But I premised [with] these lemmas to avoid the tediousness of deducing long demonstrations to an absurdity, according to the methods of the ancient geometers. For demonstrations are rendered more concise by the method of indivisibles. But because the hypothesis of indivisibles is somewhat harsh, and therefore that method is esteemed less geometrical, I chose rather to reduce the demonstrations to the prime and ultimate sums and ratios of nascent and evanescent quantities; that is, to the limits of those sums and ratios… For hereby the same thing is performed, as by the method of indivisibles; and those principles being demonstrated, we may now use them with more safety.

Principia Mathmatica, Isaac Newton, 1687, p17. Newton goes on to say that infinitesimals are not really indivisible, as they had previously been termed, but are rather “evanescent divisible quantities*”.*

(8) **p10** – **The Continuous and the Infinitesimal**, John L Bell, 2005, p91.

(9) **p10** Another possible reason is that in the Leibniz notation higher derivatives retain higher power infinitesimal terms. For example, the second derivative is expressed d^2y/dx^2 which means ‘the increment of the increment of *y* divided by the square of *dx*’, and naively this represents a division by zero. However, if dx^2 is indefinitely small then we can also ascribe that property to d^2y – so the derivative *could* be set to 0/0 which is simply undetermined until the other side of the equation has been evaluated. L’Hopital (see below) says in his textbook that Niewentijt rejected both the Leibniz notation and higher differentials.

(10) **p10** Microlinearity also offers a satisfying proof that $sin(\varepsilon)/\varepsilon = 1$. An angle θ casting a triangle in the unit circle with an opposite side *O* and hypotenuse *H* gives us $sin(\theta) = O/H$. The angle θ in radians is the arc length α divided by *H*, but if θ is an infinitesimal ε then due to microlinearity $\alpha = O$. So we have:

$$\frac{\sin(\varepsilon)}{\varepsilon}=\frac{O}{H}/\frac{\alpha}{H}=1 \tag{13}$$

This result is used in differentiating *sin*(θ).

(11) **p11** – **Analysis of the Infinitely Small, for the Understanding of Curved Lines**, Guillaume de L’Hopital, 1696, p1. This book was published anonymously by L’Hopital but was largely the work of Johann Bernoulli. It contains the first appearance of L’Hopital’s rule regarding indeterminate forms. Here is a simple (but probably not original) proof of this rule by this author: if $dy = y'dx$ and $dz = z'dx$ then $dy/dz = y'/z'$. However, *dy* is $y_1 - y_0$ and if $y_0 = 0$, *dy* is simply *y*, and the same logic applies to *z*. So we get $y/z = y'/z'$. But note that indeterminate forms can often be circumvented through algebraic manipulation.

(12) **p11** – **The First Calculus Textbooks**, The European Mathematical Awakening, Carl B Boyer, 2013, p200. This method can also be used to prove the fundamental theorem of calculus (FTC). If the area under the curve of a function $f(x)$ is signified by $A(x)$ then:

$$A(x+\varepsilon)=A(x)+\varepsilon A'(x)$$

Employing Barrow's characteristic triangles (see **ref** 1) the area under an indefinitely small segment of $f(x)$ is $\varepsilon f(x) + \frac{1}{2}\varepsilon \cdot \varepsilon f'(x)$, and the second term can be neglected. So we have:

$$\varepsilon A'(x)=\varepsilon f(x)$$

Intuitively, since the area is changing between x and $y = f(x)$, the dependent variable *should* determine the area's rate of change. If we cancel by ε and use ∫ for the inverse of differentiation we have:

$$A(x)=\int f(x)+c \qquad (14)$$

This illustrates the fundamental relationship between area and rate of change. Note that we add a constant c because in order to use the result we take the difference between two distinct values of the antiderivative (this is a corollary of FTC), which eliminates the constant.

(13) **p11** – **The History of Notations of the Calculus**, Annals of Mathematics, https://www.jstor.org/stable/pdf/1967725.pdf, Florian Cajori, 1923, p6 (Second Series, Volume 25). Cajori also says that Leibniz sometimes omitted the differential after integrals:

> If integration is conceived as the converse of differentiation, then no serious objection can be raised to the omission of the differential.

ibid, p36. And from Reference 12, if we take $dA(x) = f(x)\,dx$ and then integrate we get $A(x) = \int f(x)\,dx$, which also suggests that the dx at the end of integrals is purely customary.

(14) **p11** – **Analytical Mechanics**, https://link.springer.com/content/pdf/bfm%3A978-94-015-8903-1%2F1.pdf, Joseph-Louis Lagrange, 1811, p8.

(15) **p12** – **The Calculus in our Colleges and Technical Schools**, The Teaching of the Calculus, William F Osgood, 1907, p449.

(16) **p12** – **Cours d'Analyse**, Augustin Cauchy, 1821, p7.

(17) **p12** Gauss' opinion on this topic has been debated, but it does reflect the widespread and long held distrust of *actual* infinities:

> I protest against the use of infinite magnitude as something completed, which is never permissible in mathematics. Infinity is merely a way of speaking, the true meaning being a limit which certain ratios approach indefinitely close, while others are permitted to increase without restriction.

From a letter to Schumacher, Carl F Gauss, 1831. Quoted in **Men of Mathematics**, Eric T Bell, 1937, P556.

(18) **p12** – **Summary of the Lessons Given at the Royal Polytechnic School on Infinitesimal Calculus**, Augustin Cauchy, 1823, p36.

(19) **p12** Here is an example of this method from an old textbook. First, their definition of the derivative of a polynomial:

$$f'(x)=\frac{a\left(x_1{}^n-x_0{}^n\right)}{x_1-x_0}$$

which yields:

$$f'(x)=a\left(x_1{}^{n-1}+x_0x_1{}^{n-2}+x_0{}^2x_1{}^{n-3}+\ldots+x_0{}^{n-2}x_1+x_0{}^{n-1}\right)$$

This step does however require further algebraic justification (omitted in the textbook) using the difference of powers technique i.e. if p is a positive integer then $x^{p+1} - y^{p+1} = (x - y)(x^p + x^{p-1}y + \ldots + xy^{p-1} + y^p)$. Letting $x_1 \to x_0$ we obtain $f'(x) = nax^{n-1}$.

(20) **p13** – **The Importance of B Bolzano in the History of Calculus**, Otto

Stolz, 1881, p255. Quoted in **On the History of Epsilontics**, https://arxiv.org/abs/1502.06942, Galina Sinkevich, 2015, p18.

(21) **p13** Quoted in **Proofs and Refutations: the Logic of Mathematical Discovery**, Imre Lakatos, 1976, p25 (2015 edition). It should however be noted that even functions that look innocuous may actually be 'pathological'. For example $y = n^{1/x} / (1 + n^{1/x})$ where n is a constant has a discontinuity at zero.

(22) **p13** – **Ten Misconceptions From the History of Analysis and Their Debunking**, https://arxiv.org/abs/1202.4153, Piotr B Laszczyk, Mikhail G Katz, and David Sherry, 2012.

(23) **p14** – **Philosophy of Mathematics and Science**, Hermann Weyl, 1925. Quoted in **Hermann Weyl's Intuitionistic Mathematics**, The Bulletin of Symbolic Logic, https://www.jstor.org/stable/421038, Dirk Van Dalen, 1995, p145-169 (Volume 1, Number 2). Not all mathematicians were willing to join Weyl in making that sacrifice though; instead they attempted to 'get around' the contradictions of Cantorianism (which had by then been fully subsumed into the broader philosophy of formalism):

> For instance, about forty-five years ago, Whitehead and [Bertrand] Russell seemed to have all but succeeded in reducing mathematics to an absolute logic; but within a decade of the publication of their *Principia Mathematica*, Ramsey and Chwistek exposed a number of contradictions in the *Principia* logic. Unfortunately, neither Ramsey's application of Wittgenstein's ideas, nor Chwistek's theory of [misnamed?] constructive types, both of which were designed to save the *Principia* system from shipwreck, had any better luck. So mathematicians began to devise newer and still more ponderous logics. Such were the logics of Curry and Church which, in their turn, were proved inconsistent by Kleene and Rosser. Of the five different systems of logics enumerated by Lewis and Langford, in their *Symbolic Logic*, not one was found by the authors sufficiently 'precise' to embody 'acceptable principles of deduction.'

Mathematical Ideas - THEIR NATURE AND USE, Jagjit Singh, 1961, p288. In contrast to this, the intuitionistic (and later 'constructive') remedy for the contradictions was to identify the troublesome axiom and not admit it as generally applicable. To put it another way – the reason the formalists were obssessed with axiomatic consistency was that subconsciously they knew that their *own* belief system was logically toxic.

(24) **p15** As Lazare Carnot wrote:

> We will call every quantity, which is considered as continually decreasing (so that it may be made as small as we please, without being at the same time obliged to make those quantities vary the ratio of which it is our object to determine), an infinitely small quantity… You ask me what infinitesimal quantities mean? I declare to you that I never by that expression mean metaphysical and abstract existences, as this abridged name seems to imply; but real, arbitrary quantities, capable of becoming as small as I wish, without being compelled at the same time to make those quantities vary whose ratio it was my intention to discover.

Reflections on the Metaphysical Principles of the Infinitesimal Analysis, Lazare Carnot, 1832, p17. Quoted in **The Continuous and the Infinitesimal**, John L Bell, 2005, p105.

(25) **p16** – **Elementary Mathematics from an Advanced Standpoint**, Felix Klein, 1908, p234 (Third Edition 1924). A good example of the close connection between finite and standard calculus is the proof of the chain rule – it is the same in both branches. If *f*(*x*) = *g*(*h*(*x*)) then:

$$\begin{aligned} f(x+\Delta x) &= g[h(x+\Delta x)] \\ &= g[h(x)+\Delta x h'(x)] \\ &= g(h(x))+\Delta x h'(x) g'(h(x)) \\ \frac{f(x+\Delta x)-f(x)}{\Delta x} &= h'(x) g'(h(x)) \end{aligned}$$

$$f'(x)=h'(x)g'(h(x)) \qquad (15)$$

Notice that Δ*x* could simply be replaced with *dx* in this proof. In the Leibniz notation the end result would read:

$$\frac{df}{dx}=\frac{dh}{dx}\cdot\frac{dg(h)}{dh}$$

The second RHS numerator is not stated as *df* as is normally done (see https://www.physicsforums.com/insights/demystifying-chain-rule-calculus/, PeroK, 2018). The reader may verify this notation by taking two arbitrary functions, compounding them (if possible), and then taking the derivative of the result. It should be the same as the derivative obtained by applying the chain rule itself. Also note that even though the proof of the chain rule is the same in both branches of calculus the final derivatives produced are different, because the power rule only applies to one branch. For further verification of this or any other theorem of calculus the reader can experiment with finite difference examples by graphing functions, substituting values for x and Δx, and then measuring predicted distances and/or angles. This is easier with finite differences because determining the intersection of a line with a curve can be more exact than drawing a tangent by sight. As Leibniz put it: "the whole matter can be always referred back to assignable quantities."

(26) **p16** This type of finite precision is subtly different from the indefinite precision of standard calculus; and both are qualitatively different from practical precision, which is determined by necessity or the resolution of a particular device (of course no *physical* device can attain indefinite precision). Also note that finite difference calculus is closely related to so-called finite element analysis; a comparison of the two methods is however beyond the scope of this essay. This field has become increasingly important as computing power has increased:

> When calculus finally gets taught, its limitations are rarely discussed, and when encountered, are often dismissed as special cases. The myth of calculus' power propagates into our research and industrial organizations and is responsible for considerable loss of time in fruitless attempts to apply it to real-life scientific or engineering problems... we can solve only a handful of very special nonlinear differential equations... That leaves out the vast majority of all interesting problems... Almost as a rule, we submit our closed-form solutions... to the computer for evaluation, plotting or some other transformation. But we must also recognize that modern computers are capable of doing a great deal more than that. They can accept and then apply the fundamental laws of physics to a variety of problems... In all these impressive successes no closed-form solutions are being used. We should therefore teach our future generations not to strive for closed-form solutions at all, but to accept the fact that we are unable to produce them for the majority of useful cases. As an alternative, we would teach children at an early age how to pose the problem to the computer... We teach them how to cast the problem into the form understood by the computer... how to express the problem in finite difference form

A Viewpoint on Calculus, Hewlett-Packard Journal, https://www.hpmemoryproject.org/timeline/zvonko_fazarinc/hpj1987_03_01.htm

Zvonko Fazarinc, 1987, (Volume 38, Number 3). Indeed, any calculus problem can be set out in finite difference form, and a classic calculus problem for students concerns falling ladders. Here is a finite difference solution to one such problem by this author (from 2013):

> This problem can be solved 'numerically' without resorting to other methods. First we have:
>
> $$w^2+g^2=L^2$$
>
> $$(w+f)^2+(g+p)^2=L^2 \tag{16}$$
>
> where *w* is the wall, *g* is the ground, *L* is the ladder... *f* is the fall rate and *p* is the pull rate. Note that the units of *f* and *p* are metres – time is implicit. Also note that we do not presume that the fall rate is negative – it should come out that way. Equating, canceling [sic] common terms and solving the quadratic in *f* yields:
>
> $$f=\sqrt{(w^2-2gp-p^2)}-w$$
>
> Now we can calculate a series of increasingly accurate approximations of the fall rate if the pull rate is 0.4m/s. Here is a Python function that does this:
>
> ```
> def fallrate(wall, ground, pullrate):
>     for i in [2**n for n in range(15)]:
>         print(i * ((wall**2 - 2 * ground * (pullrate/i) - (pullrate/i)**2)**.5 – wall) )
>     input("\nEnter to exit. ")
>
> fallrate(4, 3, 0.4)
> ```
>
> Note that if we halve the distance we use to calculate the pull rate we will get a rate for half the distance. We must therefore double the result for the purpose of comparison – hence the i * etc at the start of the approximation formula. The fall rate seems to converge on -0.3m/s. This approach can be seen as a type of calculus... and it agrees with the result given by Elf worked out using conventional methods [which, as Fazarinc says, cannot be used for many calculus problems].

Playing with finite difference simulations as in this example is a great way to ‘get a feel’ for how calculus works, but note that the word ‘approximation’ as used is a bit of a misnomer since we can compare the final result with that obtained using standard calculus, and they agree exactly (it may be more apt for problems where such a comparison cannot be made). Also note that these cases are qualitively different from the approximations mentioned in Reference 28 where the ‘working point’ is absolutely accurate, but the linear projections made from it are approximate.

(27) **p16** Here is an example of this from a popular textbook:

> **General Rule for Differentiation**
>
> **First Step**: In the function replace x by $x + \Delta x$, giving a new value of the function, $y + \Delta y$.
>
> **Second Step**: Subtract the given value of the function from the new value in order to find Δy (the increment of the function) by Δx (the increment of the independent variable).
>
> **Third Step**: Divide the remainder Δy (the increment of the function) by Δx (the increment of the independent variable).
>
> **Fourth Step**: Find the limit of this quotient, when Δx (the increment of the independent variable) varies and approaches zero. This is the derivative required.
>
> The student should become thoroughly familiar with this rule by applying the process to a large number of examples.

Elements of the Differential and Integral Calculus, William A Granville, 1904, p29 (1911 edition). The fourth step was later emulated by the standard part operation in NSA, while the third step is the inevitable cancellation by the increment. Since the only incremental terms remaining by step four would be those that were previously a higher power (than the first) the nilsquare rule allows us to reverse the order of the last two steps, and replaces taking the final limit with neglecting higher power infinitesimal terms. Also note that Granville uses Δx and Δy rather than dx and dy, thus making the point that the differentials and infinitesimals of calculus are extenuations of finite variables and can be treated accordingly.

(28) **p18** If a scientist is interested in a small but *finite* linear segment then the methodology for an infinitesimal segment may be applicable:

> If ε is sufficiently tiny (“sufficiently” is ambiguous and depends on how much accuracy some scientist is interested in), higher powers of ε decrease rapidly in magnitude and so higher-order terms in Eq. (6) can often be ignored as negligible, leading to a simple local approximation.

Tutorial on obtaining Taylor Series Approximations without differentiation, http://webhome.phy.duke.edu/~hsg/415/taylor-series-tutorial.pdf, Henry Greenside, 2018. The Taylor series can be written:

$$y=f(x)+\frac{f'(x)}{1!}\varepsilon+\frac{f''(x)}{2!}\varepsilon^2+\frac{f'''(x)}{3!}\varepsilon^3+\ldots \tag{17}$$

After taking the nilsquare limit the series takes the form of a linear equation i.e. the Taylor formula. With ε = 0 it yields the point (x, $f(x)$) but as ε increases or decreases a straight line is drawn with slope $f'(x)$ i.e. the tangent. To quote one anonymous internet commentator:

> That's what makes it so useful to engineers. You are able to break down a rather complex function (i.e. non-linear) into a linear function around some "working point" *x* of your choice. This way things become significantly easier to calculate while staying reasonably accurate (as long as you don't move too far from your working point).

This can be seen as an example of controlled informality based on a clear understanding of the subject. One danger of the formalists' practice of designating all uses of the nilsquare rule as informal, even though its use is unavoidable, is that some will adopt arbitrary mathematical rules simply because they 'feel right' (as witnessed by this author). The nilsquare-limit theorem can therefore be seen as one tool for deciding which supposed informalities are correct and which are not.

(29) **p18** – **Continuous and Irrational Numbers**, Richard Dedekind, 1872. Quoted in **Logicism and Neologicism**, https://plato.stanford.edu/entries/logicism/, Neil Tennant, 2013, (revised 2017).

(30) **p19** – **The Derivative á la Carathéodory**, The American Mathematical Monthly, Stephen Kuhn, 1991, p40-44 (Volume 98, Number 1).

(31) **p19** – **Calculus of Variations and Partial Differential Equations of First Order**, Constantin Carathéodory, 1935.

(32) **p20** – **Revisited Calculus**, https://nbviewer.jupyter.org/github/warsus/calculus/blob/master/calculus.ipynb, 2015.

(33) **p20** – **Taking "Zooming in on a Point of a Graph" Seriously**,

https://mathoverflow.net/questions/77175/, Steven Gubkin, 2011.

(34) **p21** – **Is Standard Calculus Notation Wrong?** https://uncommondescent.com/intelligent-design/is-standard-calculus-notation-wrong/, Jonathan Bartlett, 2019. The full paper is **Extending the Algebraic Manipulability of Differentials**, https://arxiv.org/pdf/1801.09553.pdf, Jonathan Bartlett and Asatur Khurshudyan, 2018. Bartlett could be accused of having an ulterior motive for his critique – he is part of the 'intelligent design' movement and implies that mathematics being wrong would cast doubt on other fields such as biology. In contrast, *this* author has no objection to biological evolution (except the sheer inefficiency of it) and does not see any need for other theories regarding the origin of living creatures. But by eschewing good explanations of mathematical concepts the formalists have left an 'open goal' for people who *do* want to attack natural philosophy.

(35) **p23** As author Eli Maor puts it:

> The problem of finding the area of a closed planar shape is known as *quadrature*, or squaring… Among the shapes that [had] stubbornly resisted all attempts at squaring was the hyperbola… As we recall, Archimedes tried unsuccessfully to square the hyperbola. When the method of indivisibles was developed early in the seventeenth century, mathematicians renewed their attempts to achieve this goal. [Maor recounts subsequent developments, such as Fermat's technique for finding the quadrature of simple polynomials – which later became the power rule of calculus.] Alas, there was one snag. Fermat's formula failed for the one curve from which the entire family derives its name: the hyperbola $y = 1/x = x^{-1}$. This is because for $n = -1$, the denominator $n + 1$ in [the early power rule] becomes 0… It remained for one of Fermat's lesser known contemporaries [Gregoire de Saint-Vincent] to solve the unyielding exceptional case… Thus the quadrature of the hyperbola was finally accomplished, some two thousand years after the Greeks had first tackled the problem.

e: the Story of a Number, Eli Maor, 1994 (chapter 7). Saint-Vincent's method was complicated – it involved indefinitely approximating the area under the hyperbola with equal rectangles. And of course soon after this result calculus as we know it was invented. The massively simplified procedure given in the text for integrating the reciprocal is testament to the efficacy of Leibniz's notation, and demolishes the argument that derivatives in his notation cannot be considered to be ratios of differentials. This author has not come across any cases (that bear scrutiny) of the Leibniz notation causing contradiction, and sees no problem with its continued usage, especially in conjunction with the Lagrange notation.

(36) **p24** Formalism was brought to bear on mathematical analysis by the publication of a number of books around the year 1900:

> Forsyth also seems to have played a part in stimulating the British interest in, and awareness of, continental work in analysis at the turn of the century… it was Forsyth's *Theory of Functions of a Complex Variable* (1893), which, according to Edmund Whittaker, "had a greater influence on British mathematics than any work since Newton's *Principia*"… [The book] was soon surpassed in its standard of rigor by Whittaker's *Course of Modern Analysis* (1902) and Hobson's *Theory of Functions of a Real Variable* (1907). But it was Hardy's *Course in Pure Mathematics* (1908)… which really marked the turning point in British university-level mathematical education. From then on, analysis would be a fundamental component.

The rise of British analysis in the early 20th century: the role of G.H. Hardy and the London Mathematical Society, Historia Mathematica, Adrian C Rice and Robin J Wilson, 2003, p173-194 (Volume 30, Issue 2). After real analysis had been smuggled in on the back of complex analysis the British mathematicians took to promoting the subject with the zealotry of a convert, which of course meant crushing all opposition; and this campaign influenced the entire Anglosphere. It was Bertrand Russell who became the movement's chief proselytizer saying in 1938 "Infinitesimals... must be regarded as unnecessary, erroneous and self-contradictory." **Principles of Mathematics**, Bertrand Russell, 1938, p345 (second edition). Apparently though, Russell never actually *checked* to see if infinitesimals were compatible with limit theory – and no one seems to have asked him why he failed to do that.

(37) **p24** This has had some interesting consequences, for example, the near rediscovery of a precursor to constructivism, namely the intuitionism of LEJ Brouwer, by physicist Nicolas Gisin:

> The above simple observation has the following important consequence: after the first bits, the next bits of almost all real numbers are random, they don't follow any structure. These bits are as random as the outcome of quantum measurements (on half a singlet, let's say), i.e. they are as random as possible. Accordingly, to name them "real number" is seriously confusing. A better terminology would be to call them "random numbers". Unfortunately,

> Descartes named them “real” to contrast them with the complex numbers, those numbers that include the square root of -1, traditionally denoted i. Hence: **Mathematical real numbers are physical random numbers.** [bold in original] If at school we had learned to name the so-called real numbers, using the more appropriate terminology of random numbers, we would be much less inclined to believe that they are at the basis of determinism.

Indeterminism in Physics, Classical Chaos and Bohmian Mechanics. Are Real Numbers Really Real? https://arxiv.org/abs/1803.06824v1, Nicolas Gisin, 2018 (Version 1). Now compare that with this description of a key concept in intuitionism namely Brouwer’s choice sequences:

> A choice sequence is an infinite sequence of natural numbers whose terms are generated in succession; in the process of generating them, free choices may play a part. At one extreme, the selection of each term may be totally determined in advance by some effective rule: a sequence generated by such a rule is a lawlike sequence. At the other extreme, we have a sequence the selection of each term of which is totally unrestricted: these are the lawless sequences. In between are those choice sequences the selection of whose terms is partially restricted in advance, but not completely determined.

Elements of Intuitionism, Michael Dummett, p418 (uncomfirmed). Comparing the above quotes, Gisin seems to be unaware of Brouwer, although he mentions him in later versions of the same paper. Did the suppression of alternatives to the formalists’ view of the continuum *force* a physicist to reinvent the main alternative out of frustration at the inadequacy of orthodox real analysis? Another scholer, from outside academia this time, independently arrived at a similar conclusion:

> I realized that central to Zeno’s argument was the assumption of the existence of a durationless instant in time at which a moving object could be said to have an exactly determined or instantaneous position. But for something to be in motion, its position has to be constantly changing and undetermined; if it weren’t, the body couldn’t be in motion. By wrongly assuming one could freeze and dissect motion at an instant — thus assigning it an exact position — the paradoxes were created. In the real world, the object’s motion is continuous. We can’t freeze the world at an instant, because nature is forever changing.

The Impossible Goal of Zeno’s Paradox, https://humanparts.medium.com/the-impossible-goal-of-zenos-paradox-64d8ff6ce4fa, Peter Lynds, 2019. Lynd’s work seemed to divide opinion – it did garner some positive responses, however the depth of animosity that the constructive approach to analysis can engender can be gauged by considering this response:

> I have only read the first two sections as it is clear that the author's arguments are based on profound ignorance or misunderstanding of basic analysis and calculus. I'm afraid I am unwilling to waste any time reading further, and recommend terminal rejection.

Presumably, the anonymous journal referee responsible for that comment would not object to a creationist reformulating the Leibniz higher derivative notation.

(38) **p24** As author Jennifer Coopersmith puts it:

> Isn't this a retrogade step, to move from the breathtaking abstraction of an infinite, eternal, empty space to a set of coordinates that are system-specific and only as extensive as the system requires? Well, it might have been a backward step if it wasn't for one surprising and outstanding advantage: the new Lagrangian formulation allows us to sacrifice a universal space in favour of a universal physical principle.

The Lazy Universe: An Introduction to the Principle of Least Action, Jennifer Coopersmith, 2017, p34.

(39) **p24** As applied mathematician Jack Couglin puts it:

> We're going to carry these ε through the computation to see how they affect the final result. To make that easier, we can use special rules of arithmetic to manipulate ε:
>
> - $\varepsilon^2 = 0$ ...
> - $(1 + \varepsilon)^2 = 1 + 2\varepsilon + \varepsilon^2 = 1 + 2\varepsilon$
> - $\sqrt{(1 + \varepsilon)} = 1 + \varepsilon/2 - \varepsilon^2/8 + ... = 1 + \varepsilon/2$
>
> [The last rule] comes from doing a Taylor expansion of $\sqrt{x}$ around the point 1. Another way to look at it is that, when x is very close to 1, $\sqrt{x} \approx x$, and the slope of the graph of $\sqrt{x}$ at $x = 1$ is ½ ... With these rules, we're ready to do the error calculation.

Taming Floating Point Error, https://www.johnbcoughlin.com/posts/floating-point-axiom/, John B Coughlin, 2020.

NB All graphs were drawn using https://www.desmos.com/calculator.