Next to Nothing – a Single Paradigm

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

Mark C Marson, 11 January 2019

Abstract

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. But calculus students who do read accounts of its history encounter something strange – the claim that the theory which underpinned the subject for long after its creation was wrong and that it was corrected several hundred years later, *in spite of the fact that the original theory never produced erroneous results*. I argue here that both this characterization of the original theory and this interpretation of the paradigm shift to its successor are false. Infinitesimals, used properly, were never unrigorous and the supposed rigor of limit theory does not imply greater correctness, but rather the (usually unnecessary) exposition of hidden deductive steps. Furthermore those steps can, if set out, constitute a proof that original infinitesimals work in accordance with limit theory – contrary to the common opinion that the two approaches represent irreconcilable philosophical positions. This proof, demonstrating that we can adopt a unified paradigm for calculus, is to my knowledge novel although its logic may have been employed in another context. I also claim that non-standard analysis (the most famous previous attempt at unification) only partially clarified the situation because the type of infinitesimals it uses are critically different from original infinitesimals.

Introduction

The two main concepts which have underpinned calculus since its invention in the seventeenth century are infinitesimals and limits. For the following two centuries the former idea was held to be the justifying principle of calculus, for the next century the latter idea mostly took that role. The paradigm shift between the two methods 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' we mostly use that term here. Unfortunately nilpotency was not adopted as an explicit rule from the start even though a contemporary of Leibniz advocated 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 they were not assuming that the so-called law of excluded middle (LEM) applies to the continuum, and that 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.

As mentioned above I argue here that this development in the philosophy of mathematics was misguided – that nilsquare infinitesimals are only non-rigorous in the sense that they imply a number of deductive steps. Furthermore, those steps constitute a proof that *the criterion for 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 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 and that consequently sufficiently small expressions could be 'neglected', an approach today known as smooth infinitesimal analysis (SIA). We now prove that original infinitesimals *are* actually compatible with limit theory, before discussing how LEM can be qualified to allow this.

The nilsquare-limit theorem

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

$$y + dy = y + dy$$
$$y + dy = y + dx\frac{dy}{dx}$$

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:

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

In this form 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 regular 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. 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 would do this:

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

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 separate *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 would seem to imply that all higher power incremental terms are indefinitely small in comparison with the first power term – otherwise how could it be justified in its *own* right? Let us test this conjecture, first we express the two sets of terms as a ratio:

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

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. Canceling 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 it has both positive and negative terms – 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 these increases? Since only the difference between the positive and negative sums is relevant to this we can simplify the last equation thus:

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

Where *p* is the sum of the positive terms and *n* is the sum of the negative terms. This may seem like an unwarranted simplification, but since ε is indefinitely small we know the magnitudes of both *p* and *n* can be as small as we like. (The denominator is made positive for simplicity i.e. multiplying both levels 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* 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:

$$(p - j) - -(n - k) = (p + n) - (j + k)$$

Which shows that the range always decreases with ε. To get (*p* - *n*) and therefore *r* below given values simply reduce the range until they are less than those values. Since by definition the range must include zero, (*p* - *n*) must then be less than its given 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*:

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

We use *dx*, and below *dy*, instead of the customary Δ*x* and Δ*y* 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** $|dy_1/dx_1 - f'(x)| < |dy_2/dx_2 - f'(x)|$

Note that this shows that both NSA and calculus with original infinitesimals (SIA) are compatible with limit theory, because nilpotency is an alternative to taking the standard part. But is the nilsquare step itself compatible with limit theory? Using our previous notation for the constituent terms the limit criterion would now require that:

a limit of zero exists if for every p_2 and n_2 in $\left|\frac{p_2 - n_2}{a}\right|$ **smaller p and n can be found such that** $\left|\frac{p_1 - n_1}{a}\right| < \left|\frac{p_2 - n_2}{a}\right|$

Which is what we have shown, at least 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). In other words $\Sigma(b_n\varepsilon^{n>1})/(b_1\varepsilon) \rightarrow 0$ where *b* represents terms involving powers of arbitrary *x* values. 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 they be in two different philosophical camps? Maybe the dichotomy of LEM and not-LEM is too simplistic, as the use of LEM itself often is. The philosophy of this is discussed below, 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 increment there can also be no secant length. What do we call the line defined on a curve by a secant with no length? A tangent!

Addendum 1: 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*.

Addendum 2: 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 (i.e. a product of ε^1) 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:

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

Then 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 decreases 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 does remain after this operation the technique cannot be considered as important as regular calculus.

Calculus as it began

Realizing that much of the controversy over the foundations of calculus was caused by a misunderstanding allows us to reevaluate 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[2] calculus makes it simpler. Descartes simplified things further by advocating that mathematicians focus on algebraic not mechanical curves[3]; and Pierre de Fermat coined the word ‘adequality’ for the relationship between infinitesimals and their proximate points.

Leibniz’ 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.[4]

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 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,[5] 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 higher power terms 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[6] 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 a 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 shown above, the nilsquare property *is entirely compatible with his own conception of limits* (his clarification quoted above is equivalent to the nineteenth century definition). This must be considered one of the great missed opportunities in the history of science. The use of nilsquare infinitesimals soon became standard practice but they were considered informal, despite the fact that they always yielded correct results. This 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 self-imposed prohibition by academia (though they were still used in physics). This would have been inconceivable it 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 which violate or seem to violate the law of excluded middle.[7] Surely Nieuwentijt was not 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, the justification most often used for nilpotency in the two centuries after Leibniz and Newton was microlinearity.[8] The example given of the segment on $y = x^2$ around $x = 0$ implying that $dx^{n>1} \to 0$ works because every higher power of the increment is in effect a product of dx^2. But *dx* is just an indefinitely small portion of *x*, so it is natural to ask whether allowing there to be indefinitely small segments of a polynomial's curve allows us to treat those segments as linear as a rule. We know that indefinitely small values are nilsquare and therefore nilpotent, and a quick look at the binomial theorem shows that adding a nilpotent increment to the variable of a polynomial results in a linear equation in terms of the remaining increment, which is what we want in a gradient equation. 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[9]

Although now 'indefinitely' should be used instead of 'infinitely' for the sake of convention. One recent commentator summarized the method 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.[10]

But unfortunately, since neglecting higher order infinitesimals *seems* to be an approximation, after several hundred years of success doubts began to reemerge – and the way mathematicians had explained their thought processes was also criticized. 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:

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.[11]

He did 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 qualified[12] 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.[13]

Cauchy goes on to define infinite numbers in a similar way. Today these would be referred to as *potentially* infinitesimal and infinite numbers (and as before 'indefinitely' should be used instead of 'infinitely').[14] But although Cauchy was comfortable with infinitesimals he did make use of expressions such as:

$$\frac{f(x_1)-f(x_0)}{x_1-x_0}$$ [15]

which *can* be used to derive standard calculus in an algebraic fashion *without an incremental or infinitesimal term*.[16] An obvious drawback of this method is that it cannot be used for finite difference calculus, but arguably it is also not as intuitive as infinitesimals, and its implicit use of the ratio 0/0 may have hindered its widespread adoption. Cauchy's general approach to calculus was however very influential. 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[17]

Since 1900 the initial assertion in this account has clearly contradicted the 'official' story of Cauchy's thought processes and of how he influenced subsequent mathematics.

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.[18]

The most well known of these is the Weierstrass function (named after Karl Weierstrass) which is continuous but not differentiable. Infinitesimals, suited as they are to polynomials, would have seemed inadequate in this new terrain; so limits, being more general, would have gained favor. Consequently Weierstrass' epsilon-delta limit criterion (which was essentially the same as that of Leibniz) became a standard definition. Since this was seen as a continuation of Cauchy's approach there may have been a temptation to 'backdate' any new ideas. The third possibility is that the crisis was a side-effect of the introduction of Georg Cantor's theory of transfinite numbers. The theory depends on the Axiom of Choice, which implies LEM for the continuum, so there may have been a temptation to over-apply its attendant restrictions irrespective of whether or not they were relevant.

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 proofs. It was assumed that limits are *without qualification* compatible with LEM. But what would that really mean? Is LEM even an appropriate condition in 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 consisting of multiple terms 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 must be clarified. If we are trying to calculate a tangent of a smooth curve as a limiting value then we are not guaranteed continual improvement with a decreasing incremental term (i.e. an infinitesimal). Instead what we are guaranteed is *inexorable* improvement. This ability to posit, but not specify the value of, an increment with a desired property (i.e. which yields an error less than any given value) is how both the limits and the infinitesimals of calculus work.[19] The question then becomes – does LEM prohibit us from subsequently neglecting these small inappreciable values as it does with 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 we could in theory 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 cancel any remaining infinitesimals. Since the other algebraic manipulations are legitimate at any level of precision[20] we are perfectly entitled to do them before the nilsquare step and thus obtain useful theorems – and since the nilsquare step is a choice 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.[21]

This connection is most apparent when physicists model phenomena using finite difference calculus – the programs approximate equations with very small (but not infinitesimal) incremental terms. 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. These techniques include microadditivity – the introduction of infinitesimal increments to equations describing physical phenomena in order to study continuous change. 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.[22] This self-imposed censorship was only alleviated when the internet allowed alternative viewpoints to be widely expressed. But instead of dwelling on what was a pedagogical disaster at least, perhaps we should just correct it. Acknowledging that infinitesimals and limits are in essence the same thing would be a good way to start.

References

[1] "*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] Jeff Suzuki – **The Lost Calculus (1637-1670): Tangency and Optimization without Limits**, Mathematics Magazine (vol. 78), 2005.

[3] 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.*" (Rene Descartes, 1637, p44). 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, within reason.

[4] Quoted in **The Calculus in the Eighteenth Century**, Henk JM Bos, 1975, p56.

[5] 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 state that infinitesimals are not *actually* indivisible, as they had previously been termed, but are rather "*evanescent divisible quantities".*

[6] **The Continuous and the Infinitesimal**, John L Bell, 2005, p91.

[7] Another possible reason is that in the Leibniz notation higher derivatives retain higher power incremental (i.e. 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 we can also ascribe this 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.

[8] 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:

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

This result is used in differentiating $sin(\theta)$.

[9] **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 by the 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 have $y/z = y'/z'$.

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

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

Employing Barrow's characteristic triangles (see [1]) the area under an indefinitely small side 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:

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

If we cancel by ε and use ∫ for the inverse of differentiation we have:

$$A(x)=\int f(x)$$

[11] William F Osgood – **The Calculus in our Colleges and Technical Schools**, The Teaching of the Calculus, 1907, p449.

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

[13] **Cours d'Analyse**, Augustin Cauchy, 1821, p7.

[14] To quote Gauss "*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." (Carl F Gauss, 1831, a letter to Schumacher). Quoted in **Men of Mathematics**, Eric T Bell, 1937, P556.

[15] **Summary of the Lessons Given at the Royal Polytechnic School on Infinitesimal Calculus**, Augustin Cauchy, 1823, p36.

[16] Here is an example of this method from an old textbook. First, their definition of the derivative of a polynomial:

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

Which yields:

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

A correct step, but hardly intuitive. Letting $x_1 \rightarrow x_0$ we obtain $f'(x)=nax^{n-1}$

[17] Otto Stolz – **The Importance of B Bolzano in the History of Calculus**, 1881, p255. Quoted by Galina Sinkevich – **On the History of Epsilontics**, https://arxiv.org/abs/1502.06942, 2015, p18.

[18] Quoted in **Proofs and Refutations: the Logic of Mathematical Discovery**, Imre Lakatos, 1976, p25 (2015 edition).

[19] 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.*" (Lazare Carnot, 1832, p14). Quoted in **The Continuous and the Infinitesimal**, John L Bell, 2005, p105.

[20] Note that the indefinite precision of calculus is qualitatively different from

practical precision, which can always be given a physical value and is determined by necessity or the resolution of a particular device (no physical device can attain *indefinite* precision). Also note that finite difference calculus should not be confused with finite element analysis, where the behaviors of geometrically simple elements of shapes are modeled not with reference to an overarching function but rather based on how each interacts with its neighbors.

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

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

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

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

The LHS numerator is not *dh* as it is sometimes written. The reader may verify this notation by taking two arbitrary functions, compounding them and then taking the derivative of the result. It should be the same as the derivative obtained from applying the chain rule itself – but remember that Δ*h* (or *dh*) is the derivate of *h* multiplied by Δ*x* (or *dx*). Also note that even though the proof of the chain rule is the same in both branches of calculus the final derivatives it produces are different, because they do not use the power rule in the same way. 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 Δ*x* and then measuring predicted distances and/or angles. This is easier with finite differences because distinguishing 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.”

[22] 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 technique of taking the standard part 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 finding the limit with neglecting the higher power infinitesimal terms. Also notice that Granville uses Δx and Δy rather than dx and dy, thus making the point that the differentials and infinitesimals of calculus are real variables and can be treated as such.