Introduction

According to the *Encyclopedia Britannica* (11th edition, 1913, Volume 14, p. 535, emphasis added),

> The *infinitesimal calculus* is the body of rules and processes by which *continuously varying magnitudes* are dealt with in mathematical analysis. The name '*infinitesimal*' has been applied to the calculus because most of the leading results were first obtained by means of arguments about 'infinitely small' quantities; the 'infinitely small' or 'infinitesimal' quantities were vaguely conceived as being neither zero nor finite but in some intermediate, nascent or evanescent state.

In this passage attention has been drawn to two important, and closely related, mathematical concepts: *continuously varying magnitude* and *infinitesimal*. The first of these is founded on the traditional idea of a *continuum*, that is to say, the domain over which a continuously varying magnitude *actually varies*. The characteristic features of a (connected) continuum are, first, that *it has no gaps* – it 'coheres' – so that a magnitude varying over it has no 'jumps' and, secondly, that it is *indefinitely divisible*. Thus it has been held by a number of prominent thinkers that continua are *nonpunctate*, that is, not 'composed of' or 'synthesized from' discrete points. Witness, for example, the following quotations:

> *Aristotle*: . . . no continuum can be made up out of indivisibles, granting that the line is continuous and the point indivisible.
>
> (*Aristotle, 1980, Book 6, Chapter 1*)

> *Leibniz*: A point may not be considered a part of a line.
>
> (*Quoted in Rescher, 1967, p. 109*)

> *Kant*: Space and time are *quanta continua* . . . points and instants mere positions . . . and out of mere positions viewed as constituents capable of being given prior to space and time neither space nor time can be constructed.
>
> (*Kant, 1964, p. 204*)

Poincaré: . . . between the elements of the continuum [there is supposed to be] a sort of intimate bond which makes a whole of them, in which the point is not prior to the line, but the line to the point.

(*Quoted in Russell, 1937, p. 347*)

Weyl: Exact time- or space-points are not the ultimate, underlying, atomic elements of the duration or extension given to us in experience.

(*Weyl, 1987, p. 94*)

A true continuum is simply something connected in itself and cannot be split into separate pieces; that contradicts its nature.

(*Weyl, 1921: quoted in van Dalen, 1995, p. 160*)

Brouwer: The linear continuum is not exhaustible by the interposition of new units and can therefore never be thought of as a mere collection of units.

(*Brouwer, 1964, p. 80*)

René Thom: . . . a true continuum has no points.

(*See Cascuberta and Castellet (eds), 1992, p. 102*)

We note that these views are much at variance with the generally accepted set-theoretical formulation of mathematics in which all mathematical entities, being synthesized from collections of individuals, are ultimately of a *discrete* or *punctate* nature. This punctate character is possessed in particular by the set supporting the 'continuum' of real numbers – the 'arithmetical continuum'. As applied to the arithmetical continuum 'continuity' is accordingly not a property of the collection of real numbers *per se*, but derives rather from certain features of the additional structures – order-theoretic, topological, analytic – that are customarily imposed on it.

Closely associated with the concept of continuum is the second concept, that of 'infinitesimal'. Traditionally, an infinitesimal quantity is one which, while not necessarily coinciding with zero, is in some sense smaller than any finite quantity. In 'practical' approaches to the differential calculus an infinitesimal quantity or number is one so small that its square and all higher powers can be neglected, i.e. set to zero: we shall call such a quantity a *nilsquare infinitesimal*. It is to be noted that the property of being a nilsquare infinitesimal is an *intrinsic* property, that is, in no way dependent on comparisons with other magnitudes or numbers. An infinitesimal magnitude may be regarded as what remains after a (genuine) continuum has been subjected to an exhaustive analysis, in other words, as a continuum 'viewed in the small'. In this sense an infinitesimal[1] may be taken to be an 'ultimate part' of a continuum: in this same sense,

[1] Henceforth, the term 'infinitesimal' will mean 'infinitesimal quantity', 'infinitesimal number' or 'infinitesimal magnitude', and the context allowed to determine the intended meaning.

mathematicians have on occasion taken the ‘ultimate parts’ of curves to be infinitesimal straight lines.

We observe that the ‘coherence’ of a genuine continuum entails that any of its (connected) parts is also a continuum, and accordingly, divisible. A point, on the other hand, is by its nature not divisible, and so (as asserted by Leibniz in the quotation above) cannot be part of a continuum. Since an infinitesimal in the sense just described is a part of the continuum from which it has been extracted, it follows that it cannot be a point: to emphasize this we shall call such infinitesimals *nonpunctiform*.

Infinitesimals have a long and somewhat turbulent history. They make an early appearance in the mathematics of the Greek atomist philosopher Democritus (*c.* 450 BC), only to be banished by the mathematician Eudoxus (*c.* 350 BC) in what was to become official ‘Euclidean’ mathematics. Taking the somewhat obscure form of ‘indivisibles’, they reappear in the mathematics of the late middle ages and were systematically exploited in the sixteenth and seventeenth centuries by Kepler, Galileo’s student Cavalieri, the Bernoulli clan, and others in determining areas and volumes of curvilinear figures. As ‘linelets’ and ‘timelets’ they played an essential role in Isaac Barrow’s ‘method for finding tangents by calculation’, which appears in his *Lectiones Geometricae* of 1670. As ‘evanescent quantities’ they were instrumental in Newton’s development of the calculus, and as ‘inassignable quantities’ in Leibniz’s. De l’Hospital, the author of the first treatise on the differential calculus (entitled *Analyse des Infiniment Petits pour l’Intelligence des Lignes Courbes*, 1696) invokes the concept in postulating that ‘a curved line may be regarded as made up of infinitely small straight line segments’ and that ‘one can take as equal two quantities which differ by an infinitely small quantity’. Memorably derided by Berkeley as ‘ghosts of departed quantities’ and roundly condemned by Bertrand Russell as ‘unnecessary, erroneous, and self-contradictory’, these useful, but logically dubious entities were believed to have been finally supplanted by the limit concept which took rigorous and final form in the latter half of the nineteenth century. By the beginning of the twentieth century, most mathematicians took the view that – in analysis at least – the concept of infinitesimal had been thoroughly exploded.

Now in fact, the proscription of infinitesimals did not succeed in eliminating them altogether but, instead, drove them underground. Physicists and engineers, for example, never abandoned their use as a heuristic device for deriving (correct!) results in the application of the calculus to physical problems. And differential geometers as reputable as Lie and Cartan relied on their use in formulating concepts which would later be put on a ‘rigorous’ footing. And, in a technical sense, they lived on in algebraists’ investigations of non-archimedean fields. The concept of infinitesimal even managed to retain some public champions, one of the most active of whom was the philosopher–mathematician Charles Sanders Peirce, who saw the concept of the continuum (as did Brouwer) as arising from the subjective grasp of the flow of time and the subjective ‘now’ as a nonpunctiform infinitesimal. Here are a few of his observations on these matters:

> It is singular that nobody objects to $\sqrt{-1}$ as involving any contradiction, nor, since Cantor, are infinitely great quantities objected to, but still the antique prejudice against infinitely small quantities remains.
>
> (*Peirce, 1976, p. 123*)

> It is difficult to explain the fact of memory and our apparently perceiving the flow of time, unless we suppose immediate consciousness to extend beyond a single instant. Yet if we make such a supposition we fall into grave difficulties, unless we suppose the time of which we are immediately conscious to be strictly infinitesimal.
>
> (*ibid., p. 124*)

> [The] continuum does not consist of indivisibles, or points, or instants, and does not contain any except insofar as its continuity is ruptured.
>
> (*ibid., p. 925*)

In recent years, the concept of infinitesimal has been refounded on a solid basis. First, in the 1960s Abraham Robinson, using methods of mathematical logic, created *nonstandard analysis*, in which Leibniz’s infinitesimals – conceived essentially as infinitely small but nonzero real numbers – were finally incorporated into the real number system without violating any of the usual rules of arithemetic (see Robinson, 1966). And in the 1970s startling new developments in the mathematical discipline of category theory led to the creation of *smooth infinitesimal analysis*, a rigorous axiomatic theory of nilsquare and nonpunctiform infinitesimals. As we show in this book, within smooth infinitesimal analysis the basic calculus and differential geometry can be developed along traditional ‘infinitesimal’ lines – with full rigour – using straightforward calculations with infinitesimals in place of the limit concept.

Just as with nonEuclidean geometry, the consistency of smooth infinitesimal analysis is established by the construction of various *models* for it[2]. Each model is a mathematical structure (a category) of a certain kind containing all the usual geometric objects such as the real line and Euclidean spaces, together with transformations or maps between them. Their key feature is that within each all maps between geometric objects are *smooth*[3] and *a fortiori*

[2] For a sketch of the construction of these models, see the Appendix.

[3] A map between two mathematical objects each supporting a differential structure is said to be smooth if it is differentiable arbitrarily many times. In particular, a smooth map and all its derivatives must be continuous.

continuous[4]. For this reason, any one of these models of smooth infinitesimal analysis will be referred to as a *smooth world*; we shall sometimes use the symbol $\mathbb{S}$ to denote an arbitrary smooth world.

Now in order to achieve universal continuity of maps within smooth worlds, and thereby to ensure the consistency of smooth infinitesimal analysis, it turns out that a certain logical price must be paid. In fact, one is forced to acknowledge that the so-called *law of excluded middle* – every statement is either definitely true or definitely false – cannot be generally affirmed within smooth worlds[5]. This stems from the fact that unconstrained use of the law of excluded middle legitimizes the construction of *discontinuous* functions, as the following simple argument shows. Assuming the law of excluded middle, each real number is either equal to 0 or unequal to 0, so that correlating 1 to 0 and 0 to each nonzero real number defines a function – the 'blip' function – on the real line which is obviously discontinuous. So, if the law of excluded middle held in a smooth world $\mathbb{S}$, the discontinuous blip function could be defined there (see Fig. 1). Thus, since all functions in $\mathbb{S}$ are continuous, it follows that the law of excluded middle must fail within it. More precisely, this argument shows that the statement

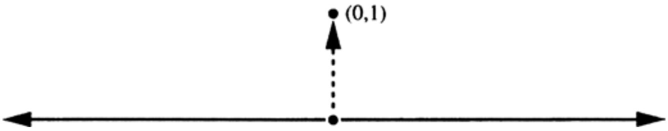

Fig. 1 The blip function

$$\text{for any real number } x, \text{ either } x = 0 \text{ or } \textit{not } x = 0$$

is *false* in $\mathbb{S}$.

Another way of showing that arbitrary statements interpreted in a smooth world cannot be regarded as possessing one of just two 'truth values' *true* or *false* runs as follows. Let Ω be the set of truth values in $\mathbb{S}$ (which we assume contains at least *true* and *false* as members). Then in $\mathbb{S}$, as in ordinary set theory,

[4] Thus each such model may be thought of as embodying Leibniz's doctrine *natura non facit saltus* – nature makes no jump.

[5] As the following quotation shows, Peirce was aware, even before Brouwer, that a faithful account of the truly continuous will involve jettisoning the law of excluded middle:

> Now if we are to accept the common idea of continuity . . . we must either say that a continuous line contains no points or . . . that the principle of excluded middle does not hold of these points. The principle of excluded middle applies only to an individual . . . but places being mere possibilities without actual existence are not individuals.
>
> (*Peirce, 1976, p. xvi: the quotation is from a note written in 1903*)

functions from any given object X to Ω correspond exactly to parts of X, proper nonempty parts corresponding to nonconstant functions. If X is a connected continuum (e.g. the real line), it presumably does have proper nonempty parts but certainly no nonconstant continuous functions to the two element set $\{$*true, false*$\}$. It follows that, in $\mathbb{S}$, the set of truth values cannot reduce to $\{$*true, false*$\}$. Thus logic in smooth worlds is *many-valued* or *polyvalent*.

Essentially the same argument shows that, in a smooth world, a connected continuum X is continuous in the strong sense that its only detachable parts are X itself and its empty part: here a part U of X is said to be *detachable* if there is a *complementary* part V of X such that U and V are disjoint and together cover X. For, clearly, detachable parts of X correspond to maps on X to $\{$*true, false*$\}$, so since all such maps on X are constant, and they in turn correspond to X itself and its empty part, these latter are the sole detachable parts of X[6].

Now at first sight in the failure of the law of excluded middle in smooth worlds may seem to constitute a major drawback. However, it is precisely this failure which allows nonpunctiform infinitesimals to be present. To get some idea of why this is so, we observe that since the law of excluded middle fails in any smooth world $\mathbb{S}$, so does its logical equivalent the *law of double negation*: for any statement A, *not not* A implies A. If we now call two points a,b on the real line *distinguishable* or *distinct* when they are not identical, i.e. *not* $a = b$ – which as usual we shall write $a \neq b$ – and indistinguishable in the contrary case, i.e. if *not* $a \neq b$, then, in $\mathbb{S}$, indistinguishability of points will not in general imply their identity. As a result, the ‘infinitesimal neighbourhood of 0’ comprising all points indistinguishable from 0 – which we will denote by I – will, in $\mathbb{S}$, be nonpunctiform in the sense that it does not reduce to $\{0\}$, that is,

it is not the case that 0 is the sole member of I.

If we call the members of I *infinitesimals*, then this statement may be rephrased:

it is not the case that all infinitesimals coincide with 0.

Observe, however, that we evidently cannot go on to infer from this that

there exists an infinitesimal which is $\neq 0$.

[6] In this connection it is worth drawing attention to the remarkable observation of Weyl (1940), who realized that the essential nature of continua can only be given full expression within a context resembling our smooth worlds:

> A natural way to take into account the nature of a continuum which, following Anaxagoras, defies ‘chopping off its parts with a hatchet’ would be by limiting oneself to continuous functions.

For such an entity would possess the property of being both distinguishable and indistinguishable from 0, which is clearly impossible[7]. What this means is that, while in $\mathbb{S}$, it would be incorrect to assume that all infinitesimals coincide with 0, it would be no less incorrect to suppose that we can single out an actual nonzero infinitesimal, i.e. one which is distinguishable from 0. In other words, nonzero infinitesimals can, and will, be present only in a 'virtual' sense[8]. Nevertheless, as we shall see, this virtual existence will suffice for the development of 'infinitesimal' analysis in smooth worlds.

In traditional mathematics two distinct, but closely related, conceptions of nonpunctiform infinitesimal can be discerned. Both may be considered as resulting from the attempt to measure continua in terms of discrete entities. The first of these conceptions stems from the idea that, just as the perimeter of a polygon is the sum of its finite discrete collection of edges, so any continuous curve should be representable as the 'sum' of an (infinite) discrete collection of infinitesimally short linear segments – the 'linear infinitesimals' of the curve. This conception was formulated by l'Hospital, and also advanced in some form two millenia earlier by the Greek mathematicians Antiphon and Bryson (*c.* 450 BC)[9]. The second concept arises analogously from the idea that a continuous surface or volume can be conceived as the sum of an indefinitely large, but discrete assemblage of lines or planes, the so-called *indivisibles*[10] of the surface or volume. This idea, exploited by Cavalieri in the seventeenth century, also appears in Archimedes' *Method*.

Let us show, by means of an example, how these two concepts of infinitesimal are related, and how they give rise to the concept of nilsquare infinitesimal. Given a smooth curve AB, suppose we want to evaluate the area of the region $ABCO$ by regarding it as the sum of thin rectangles $XYRS$ (Fig. 2). If X and S are distinguishable points then so are Y and R, so that the 'area defect' ∇ under the curve is nonzero; in this event the figure $ABCO$ cannot literally be the sum of such rectangles as $XYRS$. On the other hand, if X and S coincide, then ∇ is zero but $XYRS$ collapses into a straight line, thus failing altogether

[7] Although the law of excluded middle has had to be abandoned, the law of noncontradiction – a statement and its negation cannot both be true – will of course continue to be upheld in $\mathbb{S}$.

[8] The virtual infinitesimals of smooth worlds resemble both the virtual displacements of classical dynamics and the virtual particles of contemporary particle physics. Each has no more than only a transitory presence, and vanishes at the completion of a calculation (in the first two cases) or an interaction (in the last case).

[9] See Boyer (1959), Chapter II.

[10] The use of this term in connection with continua, although traditional, is a trifle unfortunate since no part of a continuum is 'indivisible'. This fact seems to have contributed to the general confusion – which I hope is not compounded here – surrounding the notion. See Boyer (1959), especially Chapter III.

Cambridge University Press
978-0-521-88718-2 - A Primer of Infinitesimal Analysis, Second Edition
John L. Bell
Excerpt
More information

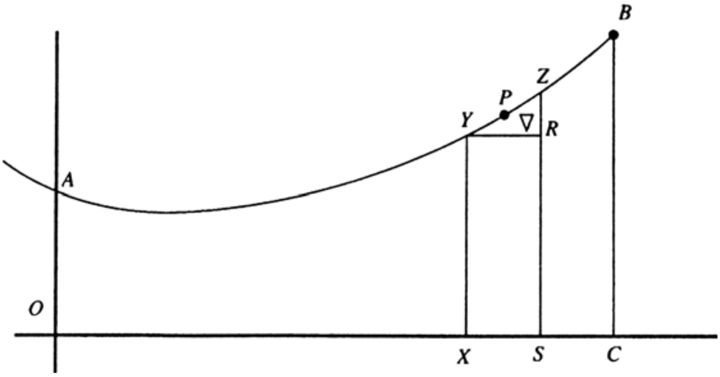

Fig. 2

to contribute to the area of the figure. In order, therefore, for *ABCO* to be the sum of rectangles like *XYRS*, we require that their base vertices *X, S* be indistinguishable without coinciding, and yet the area defect ∇ be zero. This desideratum (which is patently incompatible with the law of excluded middle) necessitates that the segment *XS* be a nondegenerate[11] linear infinitesimal of a special kind: let us appropriate Barrow's delightful term and call it a *linelet*.

Now to achieve our object we want *YRZ* to be a nondegenerate triangle of zero area. For this to be the case we clearly require first that

(a) the segment *YZ* of the curve around the point *P* is actually straight and nondegenerate (in particular, does not reduce to *P*).

In this event, the area ∇ of *YRZ* is proportional to the square of the length of the line *XS*, so that, if this area is to be zero, we must further require that

(b) *XS* is nondegenerate of length ε with $\varepsilon^2 = 0$, that is, ε is a nilsquare infinitesimal.

If, for any point *P*, a segment *YZ* of the curve exists such that the corresponding conditions (a) and (b) are satisfied, then the rectangles *XYRS* may be regarded as indivisibles whose sum exhausts the figure. Accordingly, an indivisible of the figure may be identified as a rectangle with a linelet as base.

[11] Here and throughout we term 'nondegenerate' any figure not identical with a single point.

If this procedure is to be performable for any curve, (a) needs to be extended to the following principle:

I. For any smooth curve C and any point P on it, there is a (small) nondegenerate segment of C – a *microsegment* – around P which is straight, that is, C is *microstraight* around P.

And (b) must be extended to the following principle:

II. The set Δ of magnitudes ε for which $\varepsilon^2 = 0$ – the *nilsquare infinitesimals* – does not reduce to $\{0\}$.

Principle II, which will be instrumental in reducing the differential calculus to simple algebra in our account of smooth infinitesimal analysis, is actually a consequence of Principle I. For, assuming I, consider the curve C with equation $y = x^2$ (Fig. 3). Let U be the straight portion of the curve around the origin: U is the intersection of the curve with its tangent at the origin (the x-axis). Thus U is the set of points x on the real line satisfying $x^2 = 0$. In other words, U and Δ are identical. Since I asserts the nondegeneracy of U, that is, of Δ, we obtain II.

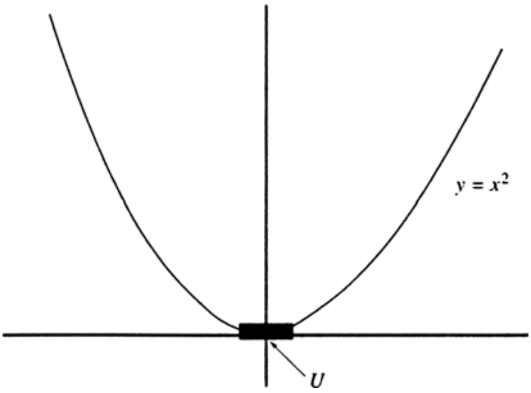

Fig. 3

Principle I, which we shall term the *Principle of Microstraightness* (for smooth curves) – and which will play a key role in smooth infinitesimal analysis – is closely related both to *Leibniz's Principle of Continuity*, and to what we shall call the *Principle of Microuniformity* (of natural processes). Leibniz's principle, in essence, is the assertion that processes in nature occur continuously, while the Principle of Microuniformity is the assertion that any such process may be considered as taking place at a constant rate over any sufficiently small period of time (i.e. over Barrow's 'timelets'). For example, if the process is the motion of a particle, the Principle of Microuniformity entails that over a timelet the particle experiences no accelerations. This idea, although rarely explicitly stated, is freely employed in a heuristic capacity in classical mechanics and the theory of differential equations. We observe in passing that the Principle of Continuity is actually a consequence of the Principle of Microuniformity.

The close relationship between the Principles of Microuniformity and Microstraightness becomes manifest when natural processes – for example, the motions of bodies – are represented as curves correlating dependent and independent variables. For then, microuniformity of the process is represented by microstraightness of the associated curve.

The Principle of Microstraightness yields an intuitively satisfying account of *motion*. For it entails that infinitesimal parts of (the curve representing) a motion are not degenerate 'points' where, as Aristotle observed millenia ago, no motion is detectable (or, indeed, even possible!), but are, rather, nondegenerate spatial segments just large enough to make motion over each one palpable. On this reckoning, states of motion are to be taken seriously, and not merely identified with their result: the successive occupation of a series of distinct positions. Instead, a state of motion is represented by the smoothly varying straight microsegment of its associated curve. This straight microsegment may be thought of as an infinitesimal 'rigid rod', just long enough to have a slope – and so, like a speedometer needle, to indicate the presence of motion – but too short to bend. It is thus an entity possessing (location and) *direction without magnitude*, intermediate in nature between a point and a Euclidean straight line.

This analysis may also be applied to the mathematical representation of *time*. Classically, time is represented as a succession of discrete instants, isolated 'nows', where time has, as it were, stopped. The Principle of Microstraightness, however, suggests rather that time be regarded as a plurality of smoothly overlapping timelets each of which may be held to represent a 'now' (or 'specious present') and over which time is, so to speak, still passing. This conception of the nature of time is similar to that proposed by Aristotle (*Physics*, Book 6, Chapter ix) to refute Zeno's paradox of the arrow.

Most important for our purposes, however, the Principle of Microstraightness decisively solves the problem of assigning a quantitative meaning to the concept of *instantaneous rate of change* – the fundamental concept of the differential calculus. For, given a smooth curve representing a physical process, the instantaneous rate of change of the process at a point P on the curve is given simply by the slope of the straight microsegment ℓ forming part of the curve at P: ℓ is of course part of the tangent to the curve at P. If the curve has equation $y = f(x)$ and P has coordinates (x_0, y_0), then the slope of the tangent