Infinitesimals and the Continuum

John L. Bell

The opposed concepts of *continuity* and *discreteness* have figured prominently in the development of mathematics, and have also commanded the attention of philosophers. Continuous entities may be characterized by the fact that they can be *divided indefinitely* without altering their essential nature. So, for instance, the water in a bucket may be indefinitely halved and yet remain water. (For the purposes of illustration I ignore the atomic nature of matter which has been established by modern physics.) Discrete entities, on the other hand, typically cannot be divided without effecting a change in their nature: half a wheel is plainly no longer a wheel. Thus we have two contrasting properties: on the one hand, the property of being indivisible, separate or discrete, and, on the other, the property of being indefinitely divisible and continuous although not actually divided into parts.

Now one and the same object can, in a sense, possess both of these properties. For example, if the wheel is regarded simply as a piece of matter, it remains so on being divided in half. In other words, the wheel regarded as a wheel is discrete, but regarded as a piece of matter, it is continuous. From examples such as these we see that continuity and discreteness are complementary attributes originating through the mind's ability to perform acts of abstraction, the one arising by abstracting an object's divisibility and the other its self-identity.

In mathematics the concept of whole number provides an embodiment of the concept of pure discreteness, that is, of the idea of a collection of separate individual objects, all of whose properties—apart from their distinctness—have been refined away. The basic mathematical representation of the idea of continuity, on the other hand, is the geometric figure, and more particularly the straight line. By their very nature geometric figures are continuous; discreteness is injected into geometry, the realm of the continuous, through the concept of a point, that is, a discrete entity marking the boundary of a line. In modern mathematics correspond to *real numbers*. Now it has been held by a number of thinkers that, while a continuum is an inexhaustible source of points, these points cannot be “reconstituted” to form the continuum from which they sprang. In short, they assert that the continuous is not explicable in terms of the discrete. Here are some examples.

Aristotle

> *No continuum can be made up of indivisibles, as for instance a line out of points, granting that the line is continuous and the point indivisible.*

Leibniz

A point may not be a constitutive part of a line.

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.*

Peirce

The very word continuity *implies that the instants of time or the points of a line are everywhere welded together.*

The continuum does not consist of indivisibles, or points, or instants, and does not contain any except insofar as its continuity is ruptured.

Poincaré

Between the elements of a continuum there is 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.

Weyl

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

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

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.

René Thom

A true continuum has no points.

To these should be added the following observation of **Brentano**, who was convinced of the primacy of the continuous in intuition:

Thus I affirm that… the concept of the continuous is acquired not through combinations of marks taken from different intuitions and experiences, but through abstraction from unitary intuitions…Every single one of our intuitions—both those of outer perception as also

> *their accompaniments in inner perception, and therefore also those of memory—bring to appearance what is continuous.*

These views are much at variance with the contemporary set-theoretical formulation of mathematics in which all mathematical entities are assemblages of individuals and so 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 oxymoronically named "arithmetical continuum". Set theory harbours no true continua, but a goodly number of sham ones!

Closely allied to the concept of a continuum is that of *infinitesimal.* An infinitesimal may be thought of as what remains after a (genuine) continuum has been subjected to an exhaustive metaphysical analysis—it is a continuum viewed, as it were, *im kleinen*. In this sense an infinitesimal may be taken as an "ultimate part" of a continuum: in a similar sense, mathematicians of the seventeenth century took the "ultimate parts" of curves to be infinitesimal straight lines. On the set-theoretical or discrete account, however, infinitesimals can be nothing other than points (or singletons). But if continua are truly continuous and do not have points as parts, then an infinitesimal in the above sense, as a part of a continuum, cannot be a point. Such an infinitesimal may be considered *nonpunctiform* or *continuous.*

A related concept is that of *infinitesimal quantity*—a quantity 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—such quantities are *nilpotent*, or, to be precise, *nilsquare.*

Infinitesimals have a long and turbulent history. They make an early appearance in the mathematics of the Greek atomist philosopher *Democritus* (c. 450 B.C.), only to be banished by the mathematician *Eudoxus* (c. 350 B.C.) in what was to become official "Euclidean" mathematics. Taking the somewhat obscure form of "indivisibles", they resurface 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, *et al.*, 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 calculus (*Analyse des Infiniment Petits pour l'Inntelligence des Lignes Courbes*, 1696) invokes the concept in postulating that "a curved line may be regarded as being composed of infinitely many 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 once and for all supplanted by the limit concept which by the end of the end of the 19th century had assumed a rigorous and final form.

But in fact the proscription of infinitesimals did not succeed in eliminating them altogether. Physicists and engineers continued to use them for quick applications of the calculus to physical problems. The differential geometers Lie and Cartan relied on their use in formulating concepts that would later be put on a "rigorous" footing. Hermann Weyl saw the infinitesimal as playing an essential role in our understanding of nature:

> *Only in the infinitely small may we expect to encounter the elementary and uniform laws*[*of nature*], *hence the world must be comprehended through the infinitely small.*

One of the most committed champions of the infinitesimal was Charles Sanders Peirce, who saw the concept of the continuum as arising from the subjective grasp of time and the subjective "now" as a continuous infinitesimal. Here are a few of his observations on the matter.

> *It is singular that nobody objects to* ν–1 *as involving any contradiction, nor, since Cantor, are infinitely great quantities objected to, but still the antique prejudice against infinitely small quantities remains.*
>
> *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.*

A new phase in the long struggle between the continuous and the discrete has opened in the past few decades with the refounding of the concept of infinitesimal on a solid basis. This has been achieved in two essentially different ways.

The first of these developments took place in the nineteen sixties when *Abraham Robinson* (1918–1974), using methods of mathematical logic, devised *nonstandard analysis,* an extension of mathematical analysis embracing both "infinitely large" and infinitesimal numbers in which the usual laws of the arithmetic of real numbers continue to hold, an idea which in essence goes back to Leibniz. Here by an infinitely large number is meant one which exceeds every positive integer; the reciprocal of any one of these is infinitesimal in the sense that, while being nonzero, it is smaller than every positive fraction $1/n$. Nonstandard analysis is an extension of classical set theory to embrace infinitesimal quantities; as such, it is perfectly compatible with that theory's discrete account of mathematical objects.

The concept of infinitesimal has been refounded in a second, and strikingly different way through the emergence in the nineteen seventies of *smooth infinitesimal analysis* (**SIA**). Employing the methods of category theory, this is a rigorous framework for mathematical analysis in which the use of limits in defining the basic notions of the calculus is replaced by the use of nilsquare infinitesimals. Smooth infinitesimal analysis provides an image of the world in which the continuous is an autonomous notion, not explicable in terms of the discrete. In **SIA** *all* functions or correlations between mathematical objects are smooth—that is, differentiable arbitrarily many times, and so in particular continuous. Accordingly **SIA** realizes in a very strong way Leibniz's principle of continuity: *natura non facit saltus.*

The correctness of Leibniz's principle in **SIA** induces a subtle, but significant change of logic there: from classical to *intuitionistic.* For, in the first place, we have only to observe that, if the law of excluded middle held without restriction, then each real number x would either be equal to 0 or unequal to 0, in which case the correlation $0 \mapsto 0$, $x \mapsto 1$ for $x \neq 0$ (the well-known "blip" function) would define a map from the space $\mathbb{R}$ of real numbers to the set $2 = \{0, 1\}$; but it is evidently discontinuous, contradicting Leibniz's principle. From this we see that Leibniz's principle implies that the law of excluded middle cannot be universally affirmed. To be precise, this argument shows that the statement

For any real number x, either $x = 0$ *or* $x \neq 0$

is *refutable* in **SIA**. (Also refutable in **SIA**, by the way, is the assertion *for any real number x,* $\neg x \neq 0 \rightarrow x = 0$.)

Leibniz's principle also implies that propositional functions, or predicates, cannot be taken as being merely "bipolar" in Wittgenstein's sense, that is, representable in terms of assuming just two "truth values" within the set 2 = {true, false} = {1, 0}. For let Ω be the domain of truth values in a world in which Leibniz's principle holds. Then, as usual, for any object X, parts of X correspond to predicates on X, that is "propositional functions" on X, in other words maps $X \rightarrow \Omega$. Now if X is a (connected) continuum, it presumably does have proper nonempty parts. But there are only two *continuous* maps $X \rightarrow 2$, namely the constant ones corresponding to the whole of X and the empty part of X, because a nonconstant continuous such map on X would yield a "splitting" of X into two nontrivial disconnected pieces. Thus: X has more than two parts; these correspond to maps $X \to \Omega$, so there are more than two of these; but there are just two maps $X \to 2$; whence $\Omega \neq 2$.

It is of interest to note in this connection Peirce's awareness, even before Brouwer, of the fact that a faithful account of the truly continuous would involve abandoning the unrestricted applicability of the law of excluded middle. In a note written in 1903, he says:

> *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 law 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.*

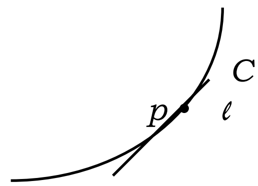
The prescience shown by Peirce here is all the more remarkable since in **SIA** the law of excluded middle does, in a certain sense, apply to individuals. This follows from the fact that, despite its failure for arbitrary predicates, the law of excluded middle can be shown to hold in **SIA** for arbitrary closed *sentences.* So if P is any predicate and a any particular real number, $P(a) \vee \neg P(a)$ will be true. Also for any particular real numbers a, b the statement $a = b \vee a \neq b$ holds. Note, however, that in SIA the truth of this statement for each pair of particular real numbers does not imply the truth of the universal generalization

$$\forall x \in \mathbb{R}\ \forall y \in \mathbb{R}\ \ x = y\ \vee\ x \neq y.$$

Indeed, we have seen that this is refutable in **SIA**: in a word, *equality on* $\mathbb{R}$ *is undecidable.* We may take this as indicating that, in **SIA** $\mathbb{R}$ is, unlike an ordinary discrete set, more than the mere "sum" of its elements.

In fact in **SIA** $\mathbb{R}$ and, indeed, all connected continua, have the stronger property of *indecomposability,* a space S being indecomposable if it cannot be expressed as a union $U \cup V$ of proper disjoint parts U, V. Thus continua in SIA are *true* continua in Anaxagoras's sense of not having parts which can be "chopped off as if by an axe".

Perhaps most remarkably, **SIA** embodies a concept of continuous infinitesimal in the form of *infinitesimal tangent vectors* to curves. A tangent vector to a curve C at a point p on it is a short straight line segment ℓ passing through p and pointing along C. In **SIA** we may take ℓ



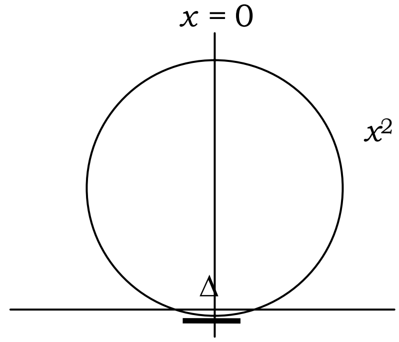
actually to be an infinitesimal *part*—a *microsegment*—of *C:* thus curves in **SIA** are "locally straight" and accordingly may be conceived as being "composed of" infinitesimal straight lines in de L'Hospital's sense.**SIA** embodies the *principle of microstraightness for smooth curves*—for any smooth curve *C* and any point on it, there is a nondegenerate infinitesimal segment—a microsegment— of the curve at that point which is straight.

The principle of microstraightness is closely related to what I shall call the *principle of microuniformity* of natural process. This is the assertion that process in nature not only take place continuously, but that they may, over a sufficiently short interval of time (a Barrovian "timelet", say) be considered as taking place at a constant rate. 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 given explicit enunciation, is freely employed in a heuristic capacity in classical mechanics and the theory of differential equations. The close relationship between the principles of microstraightness and microuniformity 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 mere degenerate "points" where, as Aristotle observed millenia ago, no motion is detectable (nor indeed even possible), but are rather, nondegenerate spatial segments just large enough for motion "over" each to be discernible. On this reckoning a state of motion is to be accorded an intrinsic status, and not merely identified with its result—the successive occupation of a series of distinct positions. Rather, a state of motion is represented by the smoothly varying straight microsegment, the infinitesimal tangent vector, 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, and so too short to indicate a rate of change of motion. It is accordingly an entity possessing (location and) direction *lacking magnitude*, intermediate in nature between a point and a Euclidean straight line.[1]

This analysis may also be applied to the mathematical representation of *time.* Classically, time is represented as a succession of discrete instants, isolated "nows" at which time has, as it were, stopped. The principle of microstraightness, however, suggests that time be instead 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, Ch. 1)* to refute Zeno's paradox of the arrow; it is also closely related to Peirce's ideas on time.

Let us turn to examine more closely the straight microsegment ℓ of a curve at a point. Since a curve is a continuous map f with domain a connected part of $\mathbb{R}$, it turns out that we may take ℓ to be the image under f of the intersection Δ of a circle with its tangent at its bottom point:



Δ is accordingly the part of $\mathbb{R}$ consisting of the points x for which $x^2 = 0$—the so-called *nilsquare infinitesimals.*[2] The axioms of **SIA**

---

[1] The idea of an infinitesimal as an entity itself lacking magnitude but possessing an intrinsic tendency to generate magnitudes (through motion) can be traced back as far as the Pythagoreans; the term *intensive infinitesimal* has been used for infinitesimals conceived in this way. See Mancosu, *Philosophy of Mathematics and Mathematical Practice in the Seventeenth Century.*

[2] It is of interest to note that, in his criticisms of Leibniz's account of the differential calculus, the 17th century theologian Bernard Nieuwentijt proposed just such a notion of infinitesimal. For good measure, he also "proved" that every curve is infinitesimally straight.

ensure that Δ is nondegenerate, i.e. does not reduce to $\{0\}$. It may be considered an *infinitesimal or micro-neigbourhood of* 0.

SIA, then, embodies the *principle of infinitesimal or micro-linearity*—Δ remains straight and unbroken under any map whatsoever; it is subject to rigid motions only—"too small to bend or break, but larger than a point". Δ is a pure synthesis of location and direction, without magnitude. Δ is sometimes called the *generic tangent vector* because it can be brought into coincidence with the infinitesimal tangent vector at any point on any curve. In **SIA** it is the archetypal "intensive" infinitesimal.

Within **SIA** the calculus and differential geometry can be developed in an elegant and intuitively appealing way, with no use of limits. For example, the derivative of a function $y = f(x)$ is given by the unique number A such that, for all nilsquare infinitesimal ε,

$$f(x + \varepsilon) - f(x) = A\varepsilon.$$

Defining the derivative in this way enables the basic rules and procedures of the differential calculus to be reduced to simple algebra.

Let me return once more to the refutability of the law of excluded middle in **SIA.** Its refutability leads to the refutability of an important principle of set theory, the *axiom of choice.* This is the assertion

> **(AC)** for any family $\mathscr{A}$ of sets, there is a function—a *choice function on* $\mathscr{A}$—$f\colon \mathscr{A} \to \bigcup\mathscr{A}$ for which $f(X) \in X$ whenever $X \in \mathscr{A}$ and $\exists x.\ x \in X$.

Now the law of excluded middle can be derived merely from the assumption that any doubleton $\{U, V\}$ has a choice function. For let α be any proposition, define

$$U = \{x \in 2\colon x = 0 \vee \alpha\} \qquad V = \{x \in 2\colon x = 1 \vee \alpha\},$$

and let f be a choice function on $\{U, V\}$. Writing $a = fU$, $b = fV$, we have $a \in U$, $b \in V$, i.e.,

$$[a = 0 \vee \alpha] \wedge [b = 1 \vee \alpha].$$

It follows that

$$[a = 0 \wedge b = 1] \vee \alpha,$$

whence

(*) $$a \neq b \vee \alpha,$$

Now clearly

$$\alpha \Rightarrow U = V = 2 \;\Rightarrow a = b,$$

whence

$$a \neq b \Rightarrow \neg\alpha.$$

But this and (*) together imply $\neg\alpha \vee \alpha$.

Since the law of excluded middle is refutable in **SIA**, so is **AC**. The failure of **AC** under the conditions of universal smoothness in **SIA** is hardly surprising in view of the axiom's well-known "paradoxical" consequences. One of these is the famous *Banach-Tarski paradox* which asserts that any solid sphere can be decomposed into finitely many (as little as 5, in fact!) pieces which can themselves be reassembled to form *two* solid spheres of the same size as the original. Paradoxical decompositions of this sort become possible only when smooth geometric objects such as spheres are analyzed into discrete sets of points which the axiom of choice then allows to be rearranged in an arbitrary (discontinuous) manner. Such procedures are not admissible in **SIA**.

I conclude with some remarks on *order* on $\mathbb{R}$ in **SIA.** In **SIA** $\mathbb{R}$ carries an order relation < which, like =, differs in certain respects from its classical counterpart. For instance, while < is transitive and irreflexive, it fails to satisfy the law of trichotomy, that is, the assertion

$$\forall x \in \mathbb{R}\ \forall y \in \mathbb{R}.\ x < y \vee x = y \vee y < x$$

is *refutable* in **SIA**. The order relation behaves even more oddly on Δ, the microneighbourhood of 0. For, introducing ε, η as variables ranging over Δ, it can be shown, $\Delta \neq \{0\}$ notwithstanding, that

(1) $$\forall\varepsilon\ \forall\eta\ \ \neg(\varepsilon < \eta \vee \eta < \varepsilon).$$

If we define the "equal to or less than" relation $\leq$ by $x \leq y \equiv \neg y < x$, it follows from (1) that

$$\forall\varepsilon\ \forall\eta\ \ \varepsilon \leq \eta \wedge \eta \leq \varepsilon.$$

In particular, the members of Δ are all simultaneously ≤ 0 and ≥ 0, but cannot be shown to coincide with zero!

Finally, a speculation. In his recent (and admirable) book *Just Six Numbers* the astrophysicist Martin Rees makes some remarks concerning the microstructure of space and time, and the possibility of developing a theory of quantum gravity. In particular he says:

> *Some theorists are more willing to speculate than others. But even the boldest acknowledge the "Planck scales" as an ultimate barrier. We cannot measure distances smaller than the Planck length [about 10^{19} times smaller than a proton]. We cannot distinguish two events (or even decide which came first) when the time interval between them is less than the Planck time (about 10^{-43} seconds).*

It strikes me that, on Rees's account, Planck scales are very similar in certain respects to Δ. In particular, the sentence (1) above seems to be an exact embodiment of the idea that we cannot decide of two "events" in Δ which came first; in fact it makes the stronger assertion that actually neither comes "first".

Could Δ provide a good model for "Planck scales"? Well, it's certainly small enough! But if so, it would be remarkable, since because Δ inhabits a domain in which everything is smooth and continuous, while Planck scales live in the quantum world which, if not outright discrete, is far from being continuous. If Planck scales could indeed be assimilated to microneighbourhoods in **SIA**, this would suggest that the quantum microworld, the Planck regime—smaller, in Rees's words, "than atoms by just as much as atoms are smaller than stars"—is not, like the world of atoms, discrete, but instead continuous like the world of stars. This would be a considerable victory for the continuous in its long struggle with the discrete.