The Continuum and the Evolution of the Concept of Real Number

John L. Bell
Department of Philosophy
The University of Western Ontario
jbell@uwo.ca

Abstract

This chapter traces the historical and conceptual development of the idea of the continuum and the allied concept of real number. Particular attention is paid to the idea of infinitesimal, which played a key role in the development of the calculus during the 17th and 18th centuries, and which has undergone a revival in the later 20th century.

Keywords

Continuum, real number, arithmetization, infinitesimal, constructive mathematics, intuitionistic mathematics, non-standard analysis, smooth infinitesimal analysis

Contents

1. The Origins of the Real Number Concept.

The concept of number has undergone a long evolution, leading to the emergence of several types of "number". The fundamental numbers are, of course, the *natural or whole numbers* 1, 2, 3, ... associated with the process, temporal in nature, of *counting*. Counting is applied to *discrete* collections of well-distinguished objects such as flocks of sheep, piles of stones, etc. Discrete collections are not always exactly divisible into numerically equal parts of a prescribed size. Thus, for example, a flock of 31 sheep, not being divisible into an exact number of pairs or triplets, cannot be exactly "counted by twos", or "counted by threes". Once the unit, in this case "individual sheep" is fixed, it cannot be arbitrarily changed and used to provide an exact count of the flock. This "rigidity of the unit" is a characteristic feature of the use of natural numbers in counting discrete collections. In this sense counting is *absolute.*

Natural numbers can be *compared.* The basic standard of comparison of numbers is, of course, *greater* (>) and *less* (<). A subtler standard of comparison is *proportionality* or *ratio.* Intuitively, given two numbers $m < n$, the ratio or proportion $m : n$ represents the *relative size* of m to n. In cases where m divides n exactly, $m: n$ can be regarded as a natural number, the number of times m divides n. For example 4 : 24 may be identified with the number 6. But when m is not a divisor of n, the ratio $m: n$ cannot be represented by a natural number in this way. Such ratios cane to be known as *fractions*, and later as *rational numbers.*

Such is the case for counting. In the case of *measuring,* additional problems arise from the fact that measuring is concerned not with discrete collections ultimately composed of distinct indivisible "atoms", but with *infinitely divisible, continuous entities,* in a word, with *continua.* In the case of discrete collections, like flocks of sheep, we are automatically provided with a natural "intrinsic" unit for counting purposes, so generating a natural number. Frege's brilliant idea of identifying the unit with a property or concept is illuminating here. In the case of a flock of 31 sheep, the number "31" is based , according to Frege, on the "concept-unit" "individual sheep". The "rigidity of the unit" in this case is illustrated by the fact that changing the concept-unit to "pair of sheep" renders it incapable of associating a natural number with the flock. Now because of the unlimited divisibility of continua, no such intrinsic units are, or can be provided. In measuring the length of a straight line, for example, we say that it is so many inches, or centimetres, long. Apart from the fact that these inches or centimetres represent lengths of straight lines, their lengths are entirely arbitrary "units", not determined in any further way by the extent of the given line. Measurement arises from a *comparison* of the length of the given line with the length of the linear unit, inch or centimeter as it may be. Measurement is thus *relative:* in measuring the length of a straight line an *additional* straight line must be specified as a unit of measurement. This fact further complicates the issue of defining the "ratio" of the lengths of straight lines, and divisible magnitudes generally. In the case of discrete collections, ratios are given directly in terms of the numbers of units constituting the collections: thus the ratio of a collection of 31 sheep to 68 sheep is 31: 68. But in the case of divisible magnitudes such as straight lines, it is not always immediately clear how the "ratio" between them should be defined. Of course, the are straightforward cases in the divisible case such as the following: suppose we are given a straight line L and we construct two straight lines L' and L'' by laying down the given line L 31 and 68 times, respectively. Then the ratio of (the lengths of) L' to L'' is 31: 68. In effect, the situation has been reduced to the discrete case by taking (the length of) L as a common unit. But when the lines L' and L'' are *arbitrary*, what guarantee is there that they are *commensurable,* that is, can a line L be found to play the role of common unit of measure for L' and L'' so enabling the ratio or proportion to be presented numerically? Let us call this the problem of *commensurability.*

Certain ancient cultures such as the Egyptians and Babylonians handled ratios in a pragmatic manner. Treating them as "fractions", which later became known as *rational numbers,* they formulated rules of calculation for them analogous to, if more complicated than, the familiar rules of calculation for natural numbers. While we do not know whether they regarded fractions as being true "numbers", for all practical purposes they treated them as such. And it seems likely that they treated the issue of commensurability, when it arose in practice, in a rough-and -ready way.

As far as we know, in antiquity only the mathematicians of the Greek school were exercised by the problem of commensurability, which had naturally arisen in their development of geometry - itself, as its name shows, a by-product of the practice of measurement. The Greek mathematicians had a strict conception of number- *arithmos,* defined by Euclid as a " multitude composed of units. – which resisted expansion to embrace proportions, ratios, or fractions. They accordingly regarded numerical proportions or ratios– that is, proportions or ratios *between* numbers - as being quite distinct from numbers *themselves*. By 500 B.C.E. they, more precisely the Pythagorean School, had developed a theory of numerical proportions.

It was a fundamental principle of the Pythagoreans that the world is explicable in terms of properties of, and relations between, whole numbers—that number, in fact, forms the very essence of the real. In geometry this doctrine led to the sweeping aside of the problem of commensurability by the affirmation of *universal commensurability.* In particular, given any pair of lines, they took it for granted that is possible to choose a unit of length sufficiently small so as to enable the proportion of the lengths of the lines to be presented as a numerical ratio.

It must then have come as a great shock to the Pythagoreans to find, as they did, that universal commensurability *cannot be upheld*. This followed upon the shattering discovery, probably made by the later Pythagoreans before 410 B.C., that ratios of whole numbers do not suffice to enable the diagonal of a square or a pentagon to be compared in length with its side. They showed that these line segments are *incommensurable*, that is, it is not possible to choose a unit of length sufficiently small so as to enable both lines to be measured by an integral number of the chosen units. Pythagorean philosophy was dealt a devastating blow by this discovery. Their geometry, too, was affected, since the demonstrations of certain basic propositions employed the principle of universal commensurability.

The discovery of incommensurability (which can be seen as an early instance of a failure to reduce the continuous to the discrete) had made it impossible for proportions between continuous magnitudes to be generally representable as numerical proportions. The response was to detach the general concept of proportion from the idea of numerical proportion and provide it with an axiomatic development independent both of numerical proportion and of commensurability. This is the theory presented in Book V of Euclid's *Elements.* The commentator Proclus attributes the general theory of proportion to Eudoxus of Cnidus (*c*.400–350 B.C.E.)

In Eudoxus's theory, we are given a collection of *similar magnitudes*, e.g. line segments, or planes , or volumes, or angles, etc., which can be compared in size, together with the notion of *ratio* of similar magnitudes satisfying certain *postulates*, of which the following, expressed in modern terms, are the most important:

P_1. Given any two similar magnitudes, there is an integral multiple of the one which exceeds the other.

This postulate, which is also known as *Archimedes' principle*, has the effect of excluding infinitely small or infinitely large quantities.

The second postulate is in effect a *definition of equality* for ratios:

P_2. The ratio *A:B* of two magnitudes *A, B* is equal to the ratio *C:D* of two other magnitudes *C, D* if, and only if, for any natural numbers *m,n*, *mA* is greater than, equal to, or less than *nB* according as *mC* is greater than, equal to, or less than *nD*.

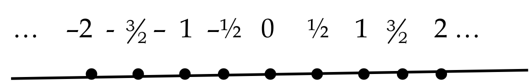
This postulate covers both the cases when *A* and *B* are commensurable, and when they are incommensurable. In the former case, numbers *p, q* can be found to satisfy $pA = qB$. It follows that $pC = qD$ and it is then easily shown that the remaining conditions of the definition are satisfied. In this case equality between the ratios *A: B* and *C:D* is established in a finite number of steps, so the ratios are numerical, and correspond to *rational numbers.*

In the case of *incommensurable A* and *B*, the content of **P_2** has been interpreted by historians of mathematics as an anticipation of the concept of *irrational number,* and so, by extension, of the general concept of *real number.* There is indeed a strikingly close relationship (Fine 1917) between the idea of the ratio of two incommensurables suggested in **P_2** and that of an irrational number as defined by Dedekind in the 19th century (discussed in the next section). The key element in **P_2** is the fact that when *A* and *B* are incommensurable the definition of the ratio *A : B* demands comparing *mA* and *nB* for *all* pairs (*m, n*) of all natural numbers *m* and *n,* and thus, implicitly at least, the assembling of these pairs into two separate classes, the class $\{(m, n): mA > nB\}$ and the class $\{(m, n): mA < nB\}$. This separation of pairs of numbers into two classes is, remarkably, identical with the Dedekind cut defining the irrational number which expresses the ratio *A : B.* It follows that the collection of ratios, commensurable and incommensurable, as defined in Euclid Book V, is actually a *subset* of the set $\mathbb{R}$ of real numbers as defined by Dedekind. For this subset to be identifiable with the whole of $\mathbb{R}$ it is necessary to postulate, as Dedekind (and Cantor) did, the existence of a one-to-one correspondence between real numbers and points on a line. This amounts, in the present case, to the assertion that for every Dedekind cut there exists a pair of magnitudes *A* and *B* which will yield this cut in the manner just described.

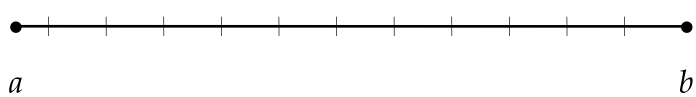
This is not, of course. to claim that Eudoxus or Euclid had literally "anticipated" Dedekiind's definition of irrational or real number. As already observed, they would have rejected the very idea that a numerical ratio could be identified as an actual number. They avoid the fraction and the irrational by basing the theory of proportion upon the equality of ratios instead of ratio itself. Nevertheless, there is no question that the basis of Dedekind's sound definition of the concept of real number millennia later was implicit in notions familiar to Eudoxus and Euclid.

Had incommensurability never been discovered, there would have been no necessity to extend the number system beyond the rational numbers.

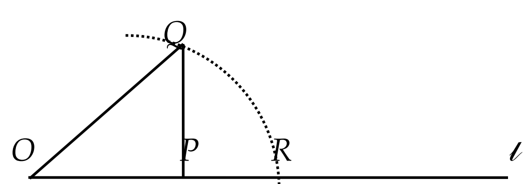
It was only very much later that fractions (along with negative numbers and 0, which, together with the natural numbers, constitute the *integers*) came to be regarded as genuine numbers. This essentially came about in the 16^{th} and 17^{th} centuries through an expansion of the Greek concept of number as discrete collections of units to numbers conceived as measuring arbitrary quantities such as geometric magnitudes. Numbers came to be associated with *points on a (continuous) line*. In particular the positive rationals (i.e. those of the form *m/n* with *m* and *n* positive integers) appear to the right of the origin 0 and the negative rationals (i.e those of the form *p/q* with *p* negative and *q* positive) to the left of 0.



Regarding the rationals as points on a line enables the line to be divided as finely as one pleases. For example, to represent all rationals of the form $m/10^9$ as points on the line, we divide the interval (0, 1) of the line between 0 and 1 into a billion equal pieces; similarly for all other intervals (1, 2), (2, 3),... and the points of subdivision then correspond to fractions of the form $m/10^9$. Since the denominator of these fractions can be made arbitrarily large, 10^{10}, 10^{100}, or whatever, thereby producing subdivisions of unlimited fineness, it would be natural to suppose that in this way one would capture as points of subdivision *all* the points on the line—in other words, that every point is represented by a *rational* number. Now it is certainly true that rational numbers suffice for all practical purposes of measuring; indeed, the Pythagorean doctrine of universal commensurability essentially elevated this practical fact to a general principle. Moreover, the rational points are *dense* on the line in the sense that they may be found in any interval (a, b), however small, with rational endpoints $a < b$ - we have only to observe that the rational number $(a + b)/2$ lies between a and b. It may be inferred from this fact that each such interval contains *infinitely* many rational points. For if the interval (a, b) contained only finitely many, n say, then we could mark them off as shown, below and then any interval between two adjacent points would be free of rational points, contradicting what we have already established.



All this would seem to lend support to the idea that every point on the line is represented by a rational number. But, of course, the Pythagorean discovery of the incommensurability of the side and diagonal of a square shows that this idea is incorrect. If we take two perpendicular lines OP and PQ of length 1 and use compasses to mark out on the line ℓ a line OR of the same length as OQ then the incommensurability of OP and OQ, and hence also OR, just means that R is not a rational point. In fact, if we designate



the "number" associated with the point *R*, then we have, by the Pythagorean theorem, $r^2 = 1^2 + 1^2 = 2$, so that, replacing *r* by the customary symbol $\sqrt{2}$, we conclude that *no rational number is equal to* $\sqrt{2}$, or that $\sqrt{2}$ is *irrational*. Here the term "irrational" is to be understood in the sense of "that which cannot be expressed as a ratio" as opposed to its more usual (but related) meaning "contrary to reason".

The fact that not every point on the line corresponds to a rational number means that, if the correspondence between points on a continuous line and "numbers" is to be maintained, the system of rational numbers has to be *enlarged still further*. It was essentially this observation, first made by the Dutch mathematician Simon Stevin (1548-1620), that led to the development of the system of *real numbers*. In his *L'Arithmétique* of 1585 Stevin explicitly introduces the idea of extending the discrete arithmetic of the Greeks to a continuous arithmetic. Stevin thinks of numbers as measures, and measures of continuous magnitudes are by their very nature continuous:

> *as well as to continuous water corresponds a continuous wetness, so to a continuous magnitude corresponds a continuous number.*

The term " continuous number" would accordingly have been more apposite than the term "real number". But history decreed otherwise. The term "real number" was introduced in the 17th century by Descartes to distinguish them from what had come to be termed imaginary" numbers—that is, square roots of negative numbers (such as $i = \sqrt{-1}$) arising formally as roots to algebraic equations. These latter were regarded as fictitious, the mere product of imagination. In this respect the term "imaginary"is to be contrasted with the term "irrational".

Stevin's most important innovation was the *decimal representation of real numbers,* which appears (although not of course in modern notation) in his 1585 work *De Thiende* ("the art of tenths"). This still today provides the most familiar way of presenting the continuum arithmetically.

In a major departure from the rational numbers which are given simply by pairs of integers, under the decimal representation each real number takes the form of an *infinite* sequence of integers

$$m.\, n_0 n_1 n_2 \ldots.$$

A real number presented in this way is then a rational number precisely when it is *periodic*, i.e., displays a repeating pattern indefinitely after a certain stage. For example,

$$6\tfrac{5}{8} = 6.6250000\ldots \qquad \tfrac{3}{7} = 0.428571\;428571\ldots..$$

It follows that just a *finite* amount of information is needed to represent a rational number as a decimal fraction.

Any *nonperiodic* decimal must then represent an *irrational* number; it is easy to furnish examples of these, for instance the following decimal containing an increasing number of zeros:

$$0.101001000100001000001\ldots.$$

Surprisingly, perhaps, no simple rule of this kind exists for constructing the decimal representation of familiar irrational numbers such as $\sqrt{2}$.

The system of *real numbers* can be formally defined to be the set of all finite or infinite (positive or negative) decimals: considered geometrically the real numbers constitute the *geometric continuum* or *real line.* The operations of addition and multiplication can be naturally extended to the real numbers so that, like the rational numbers, they constitute a *field*, which we shall denote by the symbol ℝ. Thus the real numbers resemble the rational numbers insofar as they are subject to the same operational laws. On the other hand, if we regard the integers as the basic ingredients from which the other numbers are constructed, then, as pointed out above, while each rational number can be defined in terms of just two integers, in general a real number requires *infinitely many* integers to define it. The fact that infinite processes play an essential role in the construction of the real numbers places them in sharp contrast with the rationals.

The definition of real numbers as infinite decimals is not entirely satisfactory since, for one thing, there is no compelling mathematical reason to choose the number 10 as a base for them. Moreover, the real numbers are supposed to correspond exactly to points on a line: but how do we know that every such point corresponds to a real number thus defined as a decimal? To establish this it is necessary to show that there are no "gaps" in our set of real numbers, and to define with precision what is to be understood by this assertion. This was carried out in the latter half of the nineteenth century and resulted in the modern theory of real numbers.

The Vexed Issue of Infinitesimals. Closely associated with the concept of an infinitely divisible continuum is that of an *infinitesimal.* According to the *Oxford English Dictionary* the term *infinitesimal* was originally

> *an* ordinal, *viz. the "infinitieth" in order; but, like other ordinals, also used to name fractions, thus* infinitesimal part *or* infinitesimal *came to mean unity divided by infinity* ($\frac{1}{\infty}$), *and thus an infinitely small part or quantity.*

An *infinitesimal magnitude* has been, traditionally, somewhat hazily conceived as a continuum "viewed in the small", an "ultimate part" of a continuum. In something like the same sense as a discrete entity is made up of its individual units, its “indivisibles”, so, it was maintained, a continuum is “composed” of infinitesimal magnitudes, its ultimate parts. (It is in this sense, for example, that mathematicians of the 17th century held that continuous curves are "composed" of infinitesimal straight lines.) Now the “coherence” of a continuum entails that each of its (connected) parts is also a continuum, and, accordingly, divisible. Since individual points are indivisible, it follows that no point can be part of a continuum. Points are, in fact, just *locations* in a continuum. Infinitesimal magnitudes, as parts of continua, cannot, of necessity, be points: they are, in a word, *nonpunctiform.*

Magnitudes are normally taken as being *extensive* quantities, like mass or volume, which are defined over extended regions of space. By contrast, infinitesimal magnitudes have been conceived as *intensive* magnitudes resembling locally defined intensive quantities such as temperature or density. The effect of “distributing” or “integrating” an intensive quantity over an infinitesimal magnitude is to convert the former into an infinitesimal extensive quantity: thus temperature is transformed into infinitesimal heat and density into infinitesimal mass. When the continuum is the trace of a motion, the associated infinitesimal/intensive magnitudes have been identified as *potential* magnitudes—entities which, while not possessing true magnitude themselves, embody a *tendency* to generate magnitude through motion, so manifesting “becoming” as opposed to “being”.

An infinitesimal *number* has been conceived as a number so small that, while not coinciding with zero, is in some sense smaller than any finite positive number. An infinitesimal number is, so to speak, “greater than nothing but less than anything” (Pyle 1997, p. 208). An infinitesimal numbers has been construed as a “number’ which fails to satisfy *Archimedes’ Principle,* that is, as a nonzero “number” *a* such that, for any integer *n, n.a* is less than any finite nonzero number. We have already pointed out that Eudoxus and Euclid explicitly exclude the existence of such “numbers”.

The concept of infinitesimal was beset by controversy from its beginnings. The idea makes an early appearance in the mathematics of the Greek atomist philosopher Democritus c. 450 B.C., only to be banished c. 350 B.C. by Eudoxus in what was to become official "Euclidean" mathematics. Taking the form of "indivisibles", infinitesimals resurfaced in the 16th and 17th centuries and were systematically employed by Kepler, Galileo's student Cavalieri, the Bernoulli clan, and a number of other mathematicians. In the guise of the delightfully named "linelets" and "timelets", infinitesimals played an essential role in Barrow's "method for finding tangents by calculation", which appears in his *Lectiones Geometricae* of 1670. As "evanescent quantities" infinitesimals were instrumental (although later abandoned) in Newton's development of the calculus, and, as "inassignable quantities", in Leibniz's. The Marquis de l'Hôpital, who in 1696 published the first treatise on the differential calculus (entitled *Analyse des Infiniments Petits pour l'Intelligence des Lignes Courbes*), invokes the concept in postulating that "a curved line may be regarded as being made up of infinitely small straight line segments," and that "one can take as equal two quantities differing by an infinitely small quantity."

However useful infinitesimals may have been in practice, they could scarcely withstand logical scrutiny. Derided by Berkeley in the 18th century as "ghosts of departed quantities", in the 19th century execrated by Cantor as "cholera-bacilli" infecting mathematics, and in the 20th roundly condemned by Bertrand Russell as "unnecessary, erroneous, and self-contradictory", the use of infinitesimals in the calculus and mathematical analysis was believed to have been finally supplanted by the limit concept which took rigorous and final form in the latter half of the 19th century. By the beginning of the 20th century, the concept of infinitesimal had become, in analysis at least, essentially an "unconcept".

Nevertheless, the proscription of infinitesimals did not succeed in extirpating them; they were, rather, driven further underground. Physicists and engineers, for example, never abandoned their use as a heuristic device for the derivation of correct results in the application of the calculus to physical problems. Differential geometers of the stature of Lie and Cartan relied on their use in the formulation of concepts which would later be put on a "rigorous" footing. Even in mathematical analysis they survived in Du Bois-Reymond's "orders of infinity" And, in a mathematically rigorous sense, they lived on in the algebraists' investigations of nonarchimedean fields (Ehrlich 2009).

A new phase in the saga of infinitesimals has opened in the past few decades with the refounding on a solid basis of the concept of the infinitesimal in analysis and differential geometry. This has been achieved in two essentially different ways, *Nonstandard Analysis* and *Smooth Infinitesimal Analysis.* These will be discussed in the concluding sections of this survey.

2. The Cantor-Dedekind Theory of Real Numbers and the Arithmetization of the Continuum

By the beginning of the 19th century the conception of real numbers as corresponding to points on a line, and so forming a continuum, was firmly established in mathematical practice. But solid definitions of the concept of continuum, and indeed the concept of real number itself, were lacking. Mathematical arguments involving real numbers frequently made appeal to geometric - even spatiotemporal- intuition and the intuitive idea of continuity. (Respectability on the use of spatiotemporal intuition in mathematical arguments had been conferred by the central role such intuition played in Kant's philosophy of mathematics. Kant had claimed that arithmetic and geometry were grounded in the intuition of time and space, respectively. The influence of Kant's philosophy on mathematicians waned after the emergence of non-Euclidean of geometry early on in the 19th century.) In addition, the concept of infinitesimal, which had played such an important role in the development of the calculus over the previous two centuries, but whose logical foundations were worryingly shaky, had begun to come under scrutiny.

Cauchy, Bolzano and Hamilton had all attempted to clarify the concepts of real number, but it was **Karl Weierstrass** (1815–97) who made the first systematic attack on the problem. He was determined to expel spatiotemporal intuition, and the infinitesimal, from the foundations of analysis. To instill complete logical rigour Weierstrass proposed to establish mathematical analysis on the basis of natural number alone - to "arithmetize" it. According to Hobson (1907, p. 22), "the term 'arithmetization' is used to denote the movement which has resulted in placing analysis on a basis free from the idea of measurable quantity, the fractional, negative, and irrational numbers being so defined that they depend ultimately upon the conception of integral number."

In pursuit of this goal Weierstrass had first to formulate a rigorous "arithmetical" definition of real number. He did this by defining a (positive) real number to be a countable set of positive rational numbers for which the sum of any finite subset always remains below some preassigned bound, and then specifying the conditions under which two such "real numbers" are to be considered equal, or strictly less than, one another.

Weierstrass was also concerned to make precise the concept of *continuous function.* The concept of function had by this time been greatly broadened: in 1837 Dirichlet suggested that a variable y should be regarded as a function of the independent variable x if a rule exists according to which, whenever a numerical value of x is given, a unique value of y is determined. (This idea was later to evolve into the set-theoretic definition of function as a set of ordered pairs.) Dirichlet's definition of function as a correspondence from which all traces of continuity had been purged, made necessary Weierstrass's independent definition of continuous function.

Weierstrass formulated the familiar (ε, δ) definition of continuous function: a function $f(x)$ is continuous at a if for any $\varepsilon > 0$ there is $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ for all x with $|x - a| < \delta$. He also proved his famous *approximation theorem for continuous functions*: any continuous function defined on a closed interval of real numbers can be uniformly approximated to by a sequence of polynomials.

The notion of *uniform continuity* for functions was later introduced (in 1870) by Heine: a real valued function f is uniformly continuous if for any $\varepsilon > 0$ there is $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for all x and y in the domain of f with $|x - y| < \delta$. In 1872 Heine proved the important theorem that any continuous real-valued function defined on a closed bounded interval of real numbers is uniformly continuous.

Following Weierstrass's efforts, another attack on the problem of formulating rigorous definitions of continuity and the real numbers was mounted by **Richard Dedekind**

(1831–1916). We learn from the introductory remarls to *Continuity and Irrational Numbers* (1872), that Dedekind was inspired to embark on his investigations by his belief that, in presenting the differential calculus, "geometric intuition", while "exceedingly useful from a didactic standpoint", can "make no clam to being scientific". Accordingly he made "the fixed resolve to keep meditating on the question till I should find a purely arithmetic and perfectly rigorous foundation for the principles of infinitesimal analysis". He went on to observe

> *The statement is so frequently made that that the differential calculus deals with continuous magnitude, and yet an explanation of this continuity is nowhere given; even the most rigorous expositions of the differential calculus do not base their proofs upon continuity but, with more or less consciousness of the fact, they either appeal to geometric notions or those suggested by geometry, or upon theorems which are never established in a purely arithmetic manner.*

Dedekind considered the concept of number in general to be a part of *logic,* and not, as Kant had claimed, dependent on spatiotemporal intuition. In the Preface to his later book *The Nature and Meaning of Numbers* (1888) he says

> *In speaking of arithmetic (algebra, analysis) as a part of logic I mean to imply that I consider the number-concept entirely independent of the notions or intuitions of space and time, that I consider it an immediate result from the laws of thought.*

In his investigations of continuity Dedekind focussed attention on the question: exactly what is it that distinguishes a continuous domain from a discontinuous one? He seems to have been the first to recognize that the property of density, possessed by the ordered set of rational numbers, is insufficient to guarantee continuity. In *Continuity and Irrational Numbers* he remarks that when the rational numbers are associated to points on a straight line, "there are infinitely many points [on the line] to which no rational number corresponds" (Ewald 1999, p. 770 *et seq*.), so that the rational numbers manifest "a gappiness, incompleteness, discontinuity", in contrast with the straight line's "absence of gaps, completeness, continuity." He goes on:

> *In what then does this continuity consist? Everything must depend on the answer to this question, and only through it shall we obtain a scientific basis for the investigations of* all *continuous domains. By vague remarks upon the unbroken connection in the smallest parts obviously nothing is gained; the problem is to indicate a precise characteristic of continuity that can serve as the basis for valid deductions. For a long time I pondered over this in vain, but finally I found what I was seeking. This discovery will, perhaps, be differently estimated by different people; but I believe the majority will find its content quite trivial. It consists of the following. In the preceding Section attention was called to the fact that every point p of the straight line produces a separation of the same into two portions such that every point of one portion lies to the left of every point of the other. I find the essence of continuity in the converse, i.e., in the following principle:*
>
> *'If all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions.'*

Dedekind regards this principle as being essentially indemonstrable; he ascribes to it, rather, the status of an axiom "by which we attribute to the line its continuity, by which we think continuity into the line." It is not, Dedekind stresses, necessary for *space* to be continuous in this sense, for "many of its properties would remain the same even if it were discontinuous." And in any case, he continues,

> *if we knew for certain that space were discontinuous there would be nothing to prevent us ... from filling up its gaps in thought and thus making it continuous; this filling up would*

> *consist in a creation of new point-individuals and would have to be carried out in accordance with the above principle.*

The filling-up of gaps in the rational numbers through the "creation of new point-individuals" is the key idea underlying Dedekind's construction of the domain of real numbers. He first defines a *cut* to be a partition (A_1, A_2) of the rational numbers such that every member of A_1 is less than every member of A_2. After noting that each rational number corresponds, in an evident way, to a cut, he observes that infinitely many cuts fail to be engendered by rational numbers. The discontinuity or incompleteness of the domain of rational numbers consists precisely in this latter fact. That being the case, he continues,

> *whenever we have a cut* (A_1, A_2) *produced by no rational number, we* create *a new number, an* irrational *number* α, *which we regard as completely defined by this cut* (A_1, A_2); *we shall say that the number* α *corresponds to this cut, or that it produces this cut. From now on, therefore, to every definite cut there corresponds a definite rational or irrational number, and we regard two numbers as* different *or* unequal *if and only if they correspond to essentially different cuts.*

It is to be noted that Dedekind does not identify irrational numbers with cuts; rather, each irrational number is newly "created" by a mental act and remains quite distinct from its associated cut.

Dedekind goes on to show how the domain of cuts, and thereby the associated domain of real numbers, can be ordered in such a way as to possess the property of *continuity*, viz.

if the system R *of all real numbers divides into two classes* A_1, A_2 *such that every number* a_1 *of the class* A_1 *is less than every number* a_2 *of the class* A_2, *then there exists one and only one number* α *by which this separation is produced.*

Dedekind notes that this property of continuity is actually equivalent to two principles basic to the theory of limits; these he states as:

If a magnitude grows continually but not beyond all limits it approaches a limiting value

and

if in the variation of a magnitude x we can, for every given positive magnitude δ, *assign a corresponding interval within which x changes by less than* δ, *then x approaches a limiting value.*

Dedekind's definition of real numbers as cuts was to become an essential part of the rigorous analysis of the continuum.

The most visionary "arithmetizer" of all was **Georg Cantor** (1845–1918). Cantor's analysis of the continuum in terms of infinite point sets led to his theory of transfinite numbers and to the eventual freeing of the concept of set from its geometric origins as a collection of points, so paving the way for the emergence of the concept of general abstract set central to today's mathematics· It is of interest to note that neither Dedekind nor Cantor too real numbers to be actual *sets*. This step was in fact first taken by Bertrand Russell in 1903 in *The Principles of Mathematics* (Russell 1964).

At about the same time that Dedekind published his researches into the nature of the continuous, Cantor formulated his theory of the real numbers. This was presented in the first section of a paper (Cantor 1872) on trigonometric series. Like Weierstrass and

Dedekind, Cantor aimed to formulate an adequate definition of the irrational numbers which avoided the presupposition of their prior existence, and he follows them in basing his definition on the rational numbers. Following Cauchy, Cantor calls a sequence a_1, a_2, ..., a_n, ... of rational numbers a *fundamental sequence* (nowadays called a *Cauchy sequence*) if there exists an integer N such that, for any positive rational ε, $|a_{n+m} - a_n| < \varepsilon$ for all m and all $n > N$. Any sequence $<a_n>$ satisfying this condition is said to have a *definite limit b.* Dedekind had taken irrational numbers to be "mental objects" associated with cuts, so, analogously, Cantor regards these definite limits as nothing more than *formal symbols* associated with fundamental sequences (Dauben 1979, p. 38).

The domain B of such symbols may be considered an enlargement of the domain A of rational numbers, since each rational number r may be identified with the formal symbol associated with the fundamental sequence $r, r, ..., r, ...$. Order relations and arithmetical operations are then defined on B: for example, given three such symbols b, b', b'' associated with the fundamental sequences $<a_n>$, $<a'_n>$, $<a''_n>$, the inequality $b < b'$ is taken to signify that, for some $\varepsilon > 0$ and N, $a_n - a'_n > \varepsilon$ for all $n > N$, while the equality $b + b' = b''$ is taken to express the relation $\lim(a_n + a'_n - a''_n) = 0$.

Having imposed an arithmetical structure on the domain B, Cantor is emboldened to refer to its elements as (real) *numbers.* Nevertheless, he still insists that these "numbers" have no existence except as representatives of fundamental sequences: in his theory,

the numbers (above all lacking general objectivity in themselves) appear only as components of theorems which have objectivity, for example, the theorem that the corresponding sequence has the number as limit. (Dauben 1979, p. 39).

Cantor next considers how real numbers are to be associated with points on the linear continuum. If a given point on the line lies at a distance from the origin O bearing a rational relation to the point at unit distance from that origin, then it can be represented by an element of *A*. Otherwise, it can be approached by a sequence $a_1, a_2, \ldots, a_n, \ldots$ of points each of which corresponds to an element of *A*. Moreover, the sequence <a_n> can be taken to be a fundamental sequence; Cantor writes:

The distance of the point to be determined from the point O (the origin) is equal to b, where b is the number corresponding to the sequence (Dauben 1979, p. 40).

In this way Cantor shows that each point on the line corresponds to a definite element of *B*. Conversely, each element of *B* should determine a definite point on the line. Realizing that the intuitive nature of the linear continuum precludes a rigorous proof of this property, Cantor simply assumes it as an axiom, just as Dedekind had done in regard to his principle of continuity:

Also conversely, to every number there corresponds a definite point of the line, whose coordinate is equal to that number (Dauben 1979, p. 40).

For Cantor, who began as a number-theorist, and throughout his career cleaved to the discrete, it was numbers, rather than geometric points, that possessed objective significance. Indeed, the isomorphism between the discrete numerical domain *B* and the linear continuum was regarded by Cantor essentially as a device for facilitating the manipulation of numbers.

Cantor and Dedekind had offered different ways of "constructing" the arithmetical continuum of real numbers: the former, using Cauchy sequences of rational numbers, the latter, "cuts" in the rationals. Since each claimed that his system of real numbers corresponded exactly to the points on a geometric line, it would seem to follow that Cantor's and Dedekind's systems of real numbers were essentially "the same", i.e. isomorphic. However, neither provided an explicit proof of this assertion. That the two systems are indeed isomorphic later came to be rigorously proved by showing that they are both *complete ordered fields* and that any two complete ordered fields are isomorphic An *ordered field* is an algebraic field on which a linear ordering is defined which is compatible with the algebraic operations. *Completeness* is the condition that any bounded subset has a greatest lower bound and a least upper bound (McShane and Botts 1959, Ch. 1). It can be shown that every complete ordered field has the *Archimedean property:* for any *a, b* > 0, there is a natural number *n* such that *na* > *b.* This is essentially the condition $\mathbf{P_1}$ laid down in Eudoxus' theory of proportions.

Cantor's arithmetization of the continuum had another important consequence. It had long been recognized that the sets of points of any pair of line segments, even if one of them is infinite in length, can be placed in one-one correspondence. This fact was taken to show that such sets of points have no well-defined "size". But Cantor's identification of the set of points on a linear continuum with a domain of numbers enabled the *sizes* of point sets to be compared in a definite way, using the well-grounded idea of *one-one correspondence* between sets of numbers. Thus in a letter to Dedekind written in November 1873 Cantor notes that the totality of natural numbers can be put into one-one correspondence with the totality of positive rational numbers, and, more generally, with the totality of finite sequences of natural numbers. It follows that these totalities have the same "size"; they are all *denumerable.* Cantor next raises the question of whether the natural numbers can be placed in one-one correspondence with the totality of all positive *real* numbers (Ewald 1999, p. 844).

He quickly answers his own question in the negative. In letters to Dedekind written during December 1873. Cantor shows that, for any sequence of real numbers, one can define numbers in every interval that are not in the sequence. It follows in particular that the whole set of real numbers is nondenumerable. Another important consequence concerns the existence of *transcendental numbers*, that is, numbers which are not algebraic in the sense of being the root of an algebraic equation with rational coefficients. In 1844 Liouville had established the transcendentality of any number of the form

$$\frac{a_1}{10} + \frac{a_2}{10^{2!}} + \frac{a_3}{10^{3!}} + \cdots$$

where the a_i are arbitrary integers from 0 to 9 (Kline 1972, p. 981).

In his reply to Cantor's letter of November 1873, Dedekind had observed that the set of algebraic numbers is denumerable; it followed from the nondenumerability of the real numbers that *there must be many transcendental numbers* (in fact nondenumerably many, although Cantor did not make this fact explicit until some time later).

By this time Cantor had come to regard nondenumerability as a necessary condition for the continuity of a point set, for in a paper of 1874 he asserts:

Moreover, the theorem...represents the reason why aggregates of real numbers which constitute a so-called continuum (say the totality of real numbers which are ≥ 0 and ≤ 1), cannot be uniquely correlated with the aggregate (ν); thus I found the clear difference between a so-called continuum and an aggregate like the entirety of all real algebraic numbers (Dauben 1979, p. 53).

Cantor next became concerned with the question of whether the points of spaces of different dimensions—for instance a line and a plane—can be put into one-one correspondence. In a letter to Dedekind of January 1874 he remarks:

It still seems to me at the moment that the answer to this question is very difficult – although here too one is so impelled to say no *that one would like to hold the proof to be almost superfluous* (Ewald 1999, p. 850).

Nevertheless, three years later Cantor, in a dramatic *volte-face*, established the existence of such correspondences between spaces of different dimensions. He showed, in fact, that (the points of) a space of any dimension whatsoever can be put into one-one correspondence with (the points of) a line. This result so startled him that, in a letter to Dedekind of June 1877 he was moved to exclaim: *Je le vois, mais je ne le crois pas* ("I see it, but I don't believe it." (Ewald 1999, p. 860).

Cantor's discovery caused him to question the adequacy of the customary definition of the dimension of a continuum. For it had always been assumed that the determination of a point in an *n*-dimensional continuous manifold requires *n* independent coordinates, but now Cantor had shown that, in principle at least, the job could be done with just a *single* coordinate. For Cantor this fact was sufficient to justify the claim that

... all philosophical or mathematical deductions that use that erroneous presupposition are inadmissible. Rather the difference that obtains between structures of different *dimension-number must be sought in quite other terms than in the number of independent coordinates – the number that was hitherto held to be characteristic* (Ewald 1999, p. 860).

In his reply to Cantor, Dedekind conceded the correctness of Cantor's result, but balked at Cantor's radical inferences therefrom. Dedekind maintained that the dimension-number of a continuous manifold was its "first and most important invariant" and emphasized the issue of continuity:

For all authors have clearly made the tacit, completely natural presupposition that in a new determination of the points of a continuous manifold by new coordinates, these coordinates should also (in general) be continuous *functions of the old coordinates, so that whatever appears as continuously connected under the first set of coordinates remains continuously connected under the second* (Ewald 1999, p. 863).

Dedekind also noted the extreme discontinuity of the correspondence Cantor had set up between higher dimensional spaces and the line:

... it seems to me that in your present proof the initial *correspondence between the points of the* ρ-*interval (whose coordinates are all irrational) and the points of the unit interval (also with irrational coordinates) is, in a certain sense (smallness of the alteration), as continuous as possible; but to fill up the gaps, you are compelled to admit, a frightful, dizzying discontinuity in the correspondence, which dissolves everything to atoms, so that every continuously connected part of the one domain appears in its image as thoroughly decomposed and discontinuous* (Ewald 1999, pp. 863 -4).

Dedekind avows his belief that no one-one correspondence between spaces of different dimensions can be continuous:

If it is possible to establish a reciprocal, one-to-one, and complete correspondence between the points of a continuous manifold A of *a dimensions and the points of a continuous manifold B of b dimensions, then this* correspondence itself, *if a and b are* unequal, *is necessarily* utterly discontinuous (Ewald 1999, p. 863).

In his reply to Dedekind of July 1877 Cantor clarifies his remarks concerning the dimension of a continuous manifold:

...I unintentionally gave the appearance of wishing by my proof to oppose altogether the concept of a ρ-*fold extended continuous manifold, whereas all my efforts have rather been intended to clarify it and to put it on the correct footing. When I said: "Now it seems to me that all philosophical and mathematical deductions which use that erroneous presupposition – " I meant by this presupposition* not *"the determinateness of the dimension-number" but rather the determinates of the independent coordinates, whose number is assumed by certain authors to be in all circumstances equal to the number of dimensions. But if one takes the concept of coordinate*

generally*, with no presuppositions about the nature of the intermediate functions, then the number of independent, one-to-one, complete coordinates, as I showed, can be set to any number* (Ewald 1999, p. 863).

Nevertheless he agrees with Dedekind that if "we require that the correspondence be continuous, then only structures with the same number of dimensions can be related to each other one-to-one." (Ewald 1999, p. 863). In that case, an invariant *can* be found in the number of independent coordinates, "which ought to lead to a definition of the dimension-number of a continuous structure" (Ewald 1999, p. 863).

The problem is to correlate that dimension-number, a perfectly definite mathematical object, with something as elusive as an arbitrary continuous correspondence. Cantor writes:

> *However, I do not yet know how difficult this path (to the concept of dimension-number) will prove, because I do not know whether one is able to limit the concept of* continuous correspondence in general. *But everything in this direction seems to me to depend on the possibility of such a limiting.*
>
> *I believe I see a further difficulty in the fact that this path will probably fail if the structure ceases to be thoroughly continuous; but even in this case one wants to have something corresponding to the dimension-number – all the more so, given how difficult it is to prove that the manifolds that occur in nature are thoroughly continuou*s (Ewald 1999, p. 863).

In rendering the continuous discrete, and thereby admitting *arbitrary* correspondences "of [a] frightful, dizzying discontinuity" between geometric objects "dissolved to atoms", Cantor grasps at the same time that he has rendered the intuitive concept of spatial dimension a hostage to fortune.

In 1878 Cantor published a fuller account of his ideas (Cantor 1878). Here he explicitly introduces the concept of the *power* (*Mächtigkeit*). of a set of points: two sets are said to be of equal power if there exists a one-one correspondence between them. Cantor presents demonstrations of the denumerability of the rationals and the algebraic numbers, remarking that "the sequence of positive whole numbers constitutes...the least of all powers which occur among infinite aggregates." (Dauben 1979, p. 59).

The central theme of Cantor's 1878 paper is the study of the powers of continuous *n*-dimensional spaces. He raises the issue of invariance of dimension and its connection with continuity:

Apart from making the assumption, *most are silent about how it follows from the course of this research that the correspondence between the elements of the space and the system of values* x_1, x_2, ..., x_n *is a* continuous *one, so that any infinitely small change of the system* x_1, x_2, ..., x_n *corresponds to an infinitely small change of the corresponding element, and conversely, to every infinitely small change of the element a similar change in the coordinates corresponds. It may be left undecided whether these assumptions are to be considered as sufficient, or whether they are to be extended by more specialized conditions in order to consider the intended conceptual construction of n-dimensional continuous spaces as one ensured against any contradictions, sound in itself* (Dauben 1979, p. 60).

The remarkable result obtained when one no longer insists on continuity in the correspondence between the spatial elements and the system of coordinates is described by Cantor in the following terms:

As our research will show, it is even possible to determine uniquely and completely the elements of an n-dimensional continuous space by a single real coordinate t. If no assumptions are made about the kind of correspondence, it then follows that the number of independent, continuous, real *coordinates which are used for the unique and complete determination of the elements of an n-dimensional continuous space can be brought to any arbitrary number, and thus is* not *to be regarded as a unique feature of the space* (Dauben 1979, p. 60).

Cantor shows how this result can be deduced from the existence of a one-one correspondence between the set of reals and the set of irrationals, and then, by means of an involved argument, constructs such a correspondence.

Cantor seems to have become convinced by this time that the essential nature of a continuum was fully reflected in the properties of sets of points—a conviction which was later to give birth to abstract set theory. In particular a continuum's key properties, Cantor believed, resided in the range of powers of its subsets of points. Since the power of a continuum of any number of dimensions is the same as that of a linear continuum, the essential properties of arbitrary continua were thereby reduced to those of a line. In his investigations of the linear continuum Cantor had found its infinite subsets to possess just two powers, that of the natural numbers and that of the linear continuum itself. This led him to the conviction that these were the only possible powers of such subsets—a thesis later to be enunciated as the *continuum hypothesis.* The question of the truth or falsity of the continuum hypothesis was to become one of the most celebrated problems of set theory. In the 20th century its independence of the axioms of set theory was established by Kurt Gödel and Paul Cohen.

The problem of establishing the invariance of dimension of spaces under continuous correspondences remained a pressing issue. Soon after the publication of Cantor's 1878 paper, a number of mathematicians, for example Lüroth, Thomae, Jürgens and Netto attempted proofs, but all of these suffered from shortcomings which did not escape notice (Dauben 1979, pp. 70-72). In 1879 Cantor himself published a proof which seems to have passed muster at the time, but which also contained flaws that were not detected for another twenty years (Dauben 1979, pp. 72-76).

Satisfied that he had resolved the question of invariance of dimension, Cantor returned to his investigation of the properties of subsets of the linear continuum. The results of his labours are presented in six masterly papers published during 1879–84, *Über unendliche lineare Punktmannigfaltigkeiten* ("On infinite, linear point manifolds"). Remarkable in their richness of ideas, these papers contain the first accounts of Cantor's revolutionary theory of infinite sets and its application to the classification of subsets of the linear continuum. In the third and fifth of these are to be found Cantor's observations on the nature of the continuum.

In the third article, that of 1882, which is concerned with multidimensional spaces, Cantor applies his result on the nondenumerability of the continuum to prove the startling result that continuous motion is possible in discontinuous spaces. To be precise, he shows that, if M is any countable dense subset of the Euclidean plane R^2, (for example the set of points with both coordinates algebraic real numbers), then any pair of points of the discontinuous space $A = R^2 - M$ can be joined by a continuous arc lying entirely within A[1]. In fact, Cantor claims even more:

After all, with the same resources, it would be possible to connect the points...by a continuously running line given by a unique analytic rule *and completely contained within the domain* A.

(Spaces like A are today called *arcwise connected*.) Cantor points out that the belief in the continuity of space is traditionally based on the evidence of continuous motion, but now it has been shown that continuous motion is possible even in discontinuous spaces. That being the case, the presumed continuity of space is no more than a hypothesis. Indeed, it cannot necessarily be assumed that physical space contains every point given by three real number coordinates. This assumption, he urges,

---

.

...must be regarded as a free act of our constructive mental activity. The hypothesis of the continuity of space is therefore nothing but the assumption, arbitrary in itself, of the complete, one-to-one correspondence between the 3-dimensional purely arithmetic continuum (x, y, z) and the space underlying the world of phenomena. (Dauben 1979, pp. 72-76).

These facts, so much at variance with received views, confirmed for Cantor once again that geometric intuition was a poor guide to the understanding of the continuum. For such understanding to be attained reliance must instead be placed on arithmetical analysis.

In the fifth paper in the series, the *Grundlagen* of 1883, is to be found a forthright declaration of Cantor's philosophical principles, which leads on to an extensive discussion of the concept of the continuum. Cantor distinguishes between the *intrasubjective* or *immanent* reality and *transsubjective* or *transient* reality of concepts or ideas. The first type of reality, he says, is ascribable to a concept which may be regarded

> *as actual in so far as, on the basis of definitions, [it] is well distinguished from other parts of our thought and stand[s] to them in determinate relationships, and thus modifies the substance of our mind in a determinate way* (Ewald 1999, p. 895).

The second type, transient or transsubjective reality, is ascribable to a concept when it can, or must, be taken

> *as an expression or copy of the events and relationships in the external world which confronts the intellect.*

Cantor now presents the principal tenet of his philosophy, to wit, that the two sorts of reality he has identified invariably occur together,

> *in the sense that a concept designated in the first respect as existent always also possesses, in certain, even infinitely many ways, a transient reality.*

Cantor's thesis is tantamount to the principle that correct thinking is, in its essence, a reflection of the order of Nature (or, perhaps, vice-versa. In a footnote Cantor places his thesis in the context of the history of philosophy. He claims that "it agrees essentially both with the principles of the Platonic system and with an essential tendency of the Spinozistic system" and that it can be found also in Leibniz's philosophy. But philosophy since that time has come, in Cantor's eyes, to deviate from this cardinal principle:

> *Only since the growth of modern empiricism, sensualism, and scepticism, as well as of the Kantian criticism that grows out of them, have people believed that the source of knowledge and certainty is to be found in the senses or in the so-called pure form of intuition of the world of appearances, and that they must confine themselves to these. But in my opinion these elements do not furnish us with any secure knowledge. For this can be obtained only from concepts and ideas that are stimulated by external experience and are essentially formed by inner induction and deduction as something that, as it were, was already in us and is merely awakened and brought to consciousness.*

This linkage between the immanent and transient reality of mathematical concepts—the fact that correct mathematical thinking reflects objective reality—has, in Cantor's view, the important consequence that

> *mathematics, in the development of its ideas has* only *to take account of the* immanent *reality of its concepts and has* absolutely no *obligation to examine their* transient *reality.*

It follows that

mathematics in its development is entirely free and is only bound in the self-evident respect that its concepts must both be consistent with each other and also stand in exact relationships, ordered by definitions, to those concepts which have previously been introduced and are already at hand and established. In particular, in the introduction of new numbers it is only obliged to give definitions of them which will bestow such a determinacy and, in certain circumstances, such a relationship to older numbers that they can in any given instance be precisely distinguished.

For these reasons Cantor bestows his blessing on rational, irrational, and complex numbers, which "one must regard as being every bit as existent as the finite positive integers"; even Kummer's introduction of "ideal" numbers into number theory meets with his approval. But not infinitesimal numbers, as we shall see.

Cantor begins his examination of the continuum with a tart summary of the controversies that have traditionally surrounded the notion:

The concept of the 'continuum' has not only played an important role everywhere in the development of the sciences but has also evoked the greatest differences of opinion and event vehement quarrels. This lies perhaps in the fact that, because the exact and complete definition of the concept has not been bequeathed to the dissentients, the underlying idea has taken on different meanings; but it must also be (and this seems to me the most probable) that the idea of the continuum had not been thought out by the Greeks (who may have been the first to conceive it) with the clarity and completeness which would have been required to exclude the possibility of different opinions among their posterity. Thus we see that Leucippus, Democritus, and Aristotle consider the continuum as a composite which consists ex partibus sine fine divisilibus, *but Epicurus and Lucretius construct it out of their atoms considered as finite things. Out of this a great quarrel arose among the philosophers, of whom some followed Aristotle, others Epicurus; still others, in order to remain aloof from this quarrel, declared with Thomas Aquinas that the continuum*

> *consisted neither of infinitely many nor of a finite number of parts, but of* absolutely no *parts. This last opinion seems to me to contain less an explanation of the facts than a tacit confession that one has not got to the bottom of the matter and prefers to get genteely out of its way. Here we see the* medieval-scholastic origin *of a point of view which we still find represented today, in which the continuum is thought to be an unanalysable concept, or, as others express themselves, a pure* a priori *intuition which is scarcely susceptible to a determination through concepts. Every arithmetical attempt at determination of this* mysterium *is looked on as a forbidden encroachment and repulsed with due vigour. Timid natures thereby get the impression that with the 'continuum' it is not a matter of a* mathematically logical concept *but rather of* religious dogma (Ewald 1999, p. 903).

It is not Cantor's intention to "conjure up these controversial questions again". Rather, he is concerned to "develop the concept of the continuum as soberly and briefly as possible, and only with regard to the *mathematical* theory of sets". This opens the way, he believes, to the formulation of an exact concept of the continuum—nothing less than a demystification of the *mysterium.* Cantor points out that the idea of the continuum has heretofore merely been presupposed by mathematicians concerned with the analysis of continuous functions and the like and has "not been subjected to any more thorough inspection."

Cantor next repudiates any use of temporal intuition in an exact determination of the continuum:

> *...I must explain that in my opinion to bring in* the concept of time *or* the *intuition of time in discussing the much more fundamental and more general concept of the continuum is* not *the correct way to proceed;* time *is in my opinion a representation, and its clear explanation presupposes the concept of continuity upon which it depends and without whose assistance it cannot be conceived either objectively (as a substance) or subjectively (as the form of an* a priori *intuition), but is nothing other than a* helping and linking

> concept, *through which one ascertains the relation between various different motions that occur in nature and are perceived by us. Such a thing as* objective *or* absolute time *never occurs in nature, and therefore* time *cannot be regarded as the measure of motion; far rather motion as the measure of* time—*were it not that* time, *even in the modest role of a* subjective necessary a priori *form of intuition, has not been able to produce any fruitful, incontestable success, although since Kant the time for this has not been lacking*.

These strictures apply, *pari passu*, to spatial intuition:

> *It is likewise my conviction that with the so-called* form of intuition *of* space *one cannot even begin to acquire knowledge of the* continuum. *For only with the help of a conceptually already* completed *continuum do* space *and the structure thought into it receive that content with which they can become the object, not merely of aesthetic contemplation or philosophical cleverness or imprecise comparisons, but of sober and exact mathematical investigations.*

Cantor now embarks on the formulation of a precise arithmetical definition of a continuum. Making reference to the definition of real number he has already provided (i.e., in terms of fundamental sequences), he introduces the n-dimensional arithmetical space G_n as the set of all n-tuples of real numbers $(x_1 | x_2 | \ \dots | x_n)$, calling each such an *arithmetical point* of G_n. The distance between two such points is given by

$$\sqrt{(x_1' - x_1)^2 + (x_2' - x_2)^2 + \dots (x_n' - x_n)^2} \ .$$

Cantor defines an *arithmetical* point-set in G_n to be any "aggregate of points of the points of the space G_n that is given in a lawlike way".

After remarking that he has previously shown that all spaces G_n have the same power as the set of real numbers in the interval (0, 1), and reiterating his conviction that any infinite point set has either the power of the set of natural numbers or that of (0, 1) - the continuum hypothesis - Cantor turns to the definition of the general concept of a continuum within G_n.. For this he employs the concept of *derivative* or *derived set* of a point set introduced in his 1872 paper on trigonometric series. Cantor had defined the derived set of a point set *P* to be the set of *limit points* of *P,* where a limit point of *P* is a point of *P* with infinitely many points of *P* arbitrarily close to it. A point set is called *perfect* if it coincides with its derived set (in the terminology of general topology, a set is perfect if it is closed and has no isolated points). Cantor observes that this condition does not suffice to characterize a continuum, since perfect sets can be constructed in the linear continuum which are dense in no interval, however small: as an example of such a set he offers the set consisting of all real numbers in (0 , 1) whose ternary expansion does not contain a "1". This set later became known as the *Cantor ternary set* or the *Cantor discontinuum.*

Accordingly, an additional condition is needed to define a continuum. Cantor supplies this by introducing the concept of a *connected* set. A point set *T* is connected in Cantor's sense if for any pair of its points *t, t′* and any arbitrarily small number ε there is a finite sequence of points $t_1, t_2, ..., t_n$ of *T* for which the distances $\overline{tt_1}, \overline{t_1t_2}, \overline{t_2t_3}, \ldots, \overline{t_nt'}$ are all less than ε. Cantor now observes:

> *all the geometric point-continua known to us fall under this concept of* connected *point-set, as it easy to see; I believe that in these* two *predicates 'perfect" and 'connected' I have discovered the necessary and* sufficient *properties of a point-continuum. I therefore define a point-continuum inside G_n as a* perfect-connected set. *Here 'perfect' and 'connected' are not merely words but completely general predicates of the* continuum; *they have been conceptually characterized in the sharpest way by the foregoing definitions* (Ewald 1999, p. 906).

Cantor points out the shortcomings of previous definitions of continuum such as those of Bolzano and Dedekind, and in a note dilates on the merits of his own definition:

> *Observe that this definition of a continuum is free from every reference to that which is called the* dimension *of a continuous structure; the definition includes also continua that are composed of connected pieces of different dimensions, such as lines, surfaces, solids, etc....I know very well that the word 'continuum' has previously* not *had a precise meaning in mathematics; so my definition will be judged by some as too* narrow, *by others as too* broad. *I trust that that I have succeeded in finding a proper* mean *between the two.*
> *In my opinion, a* continuum *can* only *be a* perfect *and* connected *structure. So, for example, a straight line segment lacking one or both of its end-points, or a disc whose boundary is excluded, are not* complete *continua; I call such point-sets* semi-continua.

It will be seen that Cantor has advanced beyond his predecessors in formulating what is in essence a *topological* definition of continuum, one that, while still dependent on metric notions, does not involve an order relation. It is interesting to compare Cantor's definition with that given in modern general topology. There we find a *continuum* defined as a compact connected subset of a topological space. Now within any *bounded* region of Euclidean space it can be shown that Cantor's continua coincide with continua in the sense of the modern definition. While Cantor lacked the definition of compactness, his requirement that continua be "complete" (which led to his rejecting as continua such noncompact sets as open intervals or discs) is not far away from the idea.

Cantor's analysis of infinite point sets had led him to introduce transfinite numbers (Hallett 1984) and he had come to accept their objective existence as being beyond doubt. But throughout his mathematical career he maintained an unwavering, even dogmatic opposition to infinitesimals, attacking the efforts of mathematicians such as du Bois-Reymond and Veronese to formulate rigorous theories of actual infinitesimals. As far as Cantor was concerned, the infinitesimal was beyond the realm of the possible; infiinitesimals were no more than "castles in the air, or rather just nonsense", to be classed "with circular squares and square circles"(Fisher 1981). His abhorrence of infinitesimals went so deep as to move him to outright vilification, branding them as "cholera-bacilli of mathematics" (Dauben 1979, p. 233).

Cantor believed that the theory of transfinite numbers could be employed to explode the concept of infinitesimal once and for all. Cantor's specific aim was to refute all attempts at introducing infinitesimals through the abandoning of the *Archimedean principle*—i.e. the assertion that for any positive real numbers $a < b$, there is a sufficiently large natural number n such that $na > b$. (Domains in which this principle fails to hold are called *nonarchimedean*). In a paper of 1887, Cantor attempted to demonstrate that the Archimedean property was a necessary consequence of the "concept of linear quantity" and "certain theorems of transfinite number theory", so that the linear continuum could contain no infinitesimals. He concludes that "the so-called Archimedean axiom is not an axiom at all, but a theorem which follows with logical necessity from the concept of linear quantity (Fisher 1981). Cantor's argument (in amended form) seems to have convinced a number of influential mathematicians, including Peano and Russell, of the untenability of infinitesimals.

Cantor's argument relied on the claim that the product of a positive infinitesimal, should such exist, with one of his transfinite numbers could never b finite. But since a proof of this claim was not supplied, Cantor's alleged demonstration that infinitesimals are impossible must be regarded as inconclusive (Fisher 1981). In any case consistent theories of infinitesimals were later constructed (see sections 5 and 6 below).

Cantor's rejection of infinitesimals stemmed from his conviction that his own theory of transfinite ordinal and cardinal numbers exhausted the realm of the numerable, so that no further generalization of the concept of number, in particular any which embraced infinitesimals, was admissible. For Cantor, transfinite numbers were grounded in *transient reality*, while infinitesimals and similar chimeras (in his estimation) could not be accorded such a status. Recall Cantor's assertion:

> *In particular, in the introduction of new numbers it is only obliged to give definitions of them which will bestow such a determinacy and, in certain circumstances, such a relationship to older numbers that they can in any given instance be precisely distinguished.*

Accordingly Cantor could not, or would not, grant the infinitesimal an immanent reality which was compatible with "older" numbers—among which he of course included his transfinite numbers— for had he done so he would perforce have had, in accordance with his own principles, to grant the infinitesimal transient reality. This seems to be the reason for Cantor's determination to demonstrate the inconsistency of the infinitesimal with his concept of transfinite number.

It has to be said (Fisher 1981) that concerning infinitesimals Cantor displayed a dogmatic attitude, arguing, in effect, that the admission of infinitesimals was self-contradictory. While Cantor's intolerant attitude towards the infinitesimal is not, strictly speaking, inconsistent with the "freedom" of mathematics he proclaimed, it does seem to reflect his deep-seated conviction (Dauben 1979, pp. 288–91) that his transfinite set theory was the product of "divine inspiration", so that anything in conflict with it must be anathematized.

3. Dissenting Voices: Divergent Conceptions of the Continuum in the 19th and Early 20th Centuries

Despite the remarkable success of Weierstrass, Dedekind and Cantor in constructing the continuum from arithmetical materials, a number of thinkers of the late 19th and early 20th centuries remained opposed, in varying degrees, to the idea of explicating the continuum concept entirely in discrete terms, whether arithmetical or set-theoretic. These include the mathematicians *du Bois-Reymond, Poincaré, Brouwer* and *Weyl,* and the philosophers *Brentano* and *Peirce*.

Paul du Bois-Reymond (1831–1889), against whose theory of infinities and infinitesimals Cantor fought so hard, was a prominent mathematician of the later 19th century who made significant contributions to real analysis, differential equations, mathematical physics and the foundations of mathematics. While accepting many of the methods of the Dedekind-Cantor school, and indeed embracing the idea of the actual infinite, he rejected its associated philosophy of the continuum on the grounds that it was committed to the reduction of the continuous to the discrete. In 1882 he writes:

The conception of space as static and unchanging can never generate the notion of a sharply defined, uniform line from a series of points however dense, for, after all, points are devoid of size, and hence no matter how dense a series of points may be, it can never become an interval, which must always be regarded as the sum of intervals between points (Ehrlich 1994, p. x).

Du Bois-Reymond took a somewhat mystical view of the continuum, asserting that its true nature, being beyond the limits of human cognition, would forever elude the understanding of mathematicians. This echoes his older brother Emile's famous *ignorabimus* decaration in 1880 concerning "world riddles," certain of which, such as the ultimate nature of matter and force, and the origin of sensations, would always elude explanation.

Nevertheless, Du Bois-Reymond's view of the continuum as a *mysterium* did not prevent him from developing his own theory of the mathematical continuum, a *continuum of functions*, during the 1870s and 80s. This was introduced in an article of 1870-1 as the *calculus of infinities*. Here du Bois-Reymond considers "functions ordered according to the limit of their quotients" (this, and all subsequent quotations of Du Bois-Reymond are tahen from Fisher 1981).

.

The orderings of functions, in du Bois-Reymond's notation,

$$f(x) \succ \varphi(x),\ f(x) \approx \varphi(x), f(x) \prec \varphi(x),$$

are defined respectively by

$$\lim_{x\to\infty} f(x)/\varphi(x) = \infty,\ \lim_{x\to\infty} f(x)/\varphi(x) \text{ is finite and } \neq 0,\ \lim_{x\to\infty} f(x)/\varphi(x) = 0.$$

Thus, for example, $e^x \succ x \succ \log(x)$, $x^p \succ x$ for any $p > 1$, while $cx^r \approx x^r$ for any c and r. When $f(x) \succ \varphi(x)$, , $f(x)$ is said to have an "infinity greater than $\varphi(x)$"; when $f(x) \prec \varphi(x)$, $f(x)$ may be thought of (although du Bois-Reymond does not say this explicitly) as being infinitesimal in comparison with $\varphi(x)$. Du Bois-Reymond considers sequences of functions linearly ordered under $\prec$ or $\succ$. Such "scales of infinity" can be caused to become arbitrarily complex by the continued interpolation of new such sequences between terms. Du Bois-Reymond draws an analogy with the ordered set of real numbers:

Just as between two functions two functions as close with respect to their infinities as one may want, one can imagine an infinity of others forming a kind of passage from the first function to the second, one can compare the sequence F [a scale of infinity] to the sequence of real numbers, in *which one can also pass from one number to a number very little different from it by an infinity of other ones.*

While du Bois-Reymond uses the term "infinities" in connection with his classification of functions, he does not at this point speak of infinite numbers or actual infinities. But in an article of 1875 he drops his reservations on the matter, and boldly begins by asserting:

> *I decided to publish this continuation of my research on functions becoming infinite in German after I overcame my aversion to using the word 'infinite* (unendlich)' *as a substantive, like the French their 'infini'. I even flatter myself that, by this 'infinite* (unendlich)', *I have enriched our mathematical vocabulary in a noteworthy way.*

He goes on to say:

> *In earlier articles I have distinguished the different infinities of functions by their different magnitudes so that they form a domain of quantities (the infinitary) with the stipulation that the infinity of $\varphi(x)$ is to be regarded as larger than that of $\psi(x)$ or equal to it...according as the quotient $\varphi(x)/\psi(x)$ is infinite or finite. Thus in the infinitary domain of quantities the quotient enters in place of the difference in the ordinary domain of numbers. Between the two domains there are many analogies... I can add further that the most complete symmetry exists between functions becoming zero and becoming infinity, in such a way that everywhere the positive numbers correspond in the most striking way to becoming infinity, the negative numbers to becoming zero, zero to remaining finite. Instead of numbers as fixed signs in the domain of numbers, one has in the infinitary domain of quantities an unlimited number of simple functions; the exponential functions, the powers, the logarithmic functions, that likewise form fixed points of comparisons, and between whose arbitrarily close infinities a limitless number of infinities different from each other can be inserted.*

In a paper of 1877 du Bois-Reymond compares his system of “infinities” and that of “ordinary” numbers. He introduces the concept of “numerical continuity”, an idea which he suggests underlies the introduction of irrational numbers. To illustrate the idea, du Bois-Reymond offers as a metaphor the distribution of the stars on a great circle in the sky. The readily identified brighter stars he compares to rational numbers with small numerators and denominators. Use of telescopes reveals the presence of new stars in any region, however small, but patches of darkness are always found between them. And then

> *our imagination, or speculation, peoples this as it were asymptotically uniform nothingness which always remains, with matter whose radiation or our observation can no longer make accessible. In our thought, we may believe there is no end, and we admit no empty spot in the sky.*

This is analogous to the generation of rational and irrational numbers:

> *Thus through more precise consideration the rational numbers always approach more closely to one another, yet in our minds gaps are always left between them, which mathematical speculation then fills with the irrationals.*

According to du Bois-Reymond this is essentially the way in which “numerical continuity” has arisen. He sees mathematical intuition as assigning equal authenticity to geometric and numerical quantity, but the attainment of complete equality between the two can only be attained through the use of the limit concept in introducing the irrationals. And the insertion of the irrationals between the rationals is an extension of the primitive concept of number to an equally primitive, but more comprehensive, concept of continuous quantity. That being the case, the comparison between numerical and geometric quantities may conceal further subtleties.

One such subtlety is brought to light in connection with continuous families of curves. When these are allowed to increase with growing rapidity their approximative behaviour is quite different from that associated with ordinary spatial continuity. Du Bois-Reymond writes:

> *If we think of two different quickly increasing functions, then all the transitions from one to the other are spatially conceivable and present in our minds. We cannot conceive anywhere a gap between two curves increasing to infinity or in the neighbourhood of one such curve, which could not be filled with curves; on the contrary, each curve is accompanied by curves which proceed arbitrarily close to it, to infinity.*

Now, unlike the points on a line segment, the curves which run between two such curves do not form a “simple infinity”, that is, they do not depend on just a single parameter. Du Bois-Reymond shows that this infinity is “unlimited” in the sense that it is not n-fold for any finite n. He continues:

> *...just as in the ordinary domain of quantities we can only express quantities numerically exactly by means of rational numbers, since the other numbers are not actual numbers but only limits of such numbers: so we can only express infinities with well-defined functions, of which we only have at our disposal up to now those belonging to the family of logarithms, powers, exponential functions*.

Du Bois-Reymond next notes the difference between the approximative behaviour of real numbers and that of “infinities” associated with functions. While one can approximate a number, say ½, by many sequences in such a way that any number, however close to ½, will fall between two members of any such sequence, the situation is quite different for the functions associated with infinities. For example, consider the sequence of functions

$$x^{1/2}, x^{2/3}, \ldots, x^{p/p+1}, \ldots .$$

The exponents of the members of this sequence approach 1, but it is not hard to establish the existence of functions whose infinities fall between *all* of the infinities of the members of this sequence and the function x to which the sequence converges in an appropriate sense. For example, $x^{\log\log x / \log\log x + 1}$ is readily shown to be such a function.

Du Bois-Reymond next proceeds to demonstrate the generality of this phenomenon:

> *One cannot approximate a given infinity $\lambda(x)$ with any sequence of functions $\varphi_p(x)$, $p = 1, 2, \ldots$ in such a way that one could not always specify a function $\psi(x)$ which satisfies for arbitrarily large values of p $\lambda(x) \gtrless \psi(x) \gtrless \varphi_p(x)$.*

He continues:

> *Now the fact that we can with no conceivable sequence of functions approach without limit a given infinity, certainly has something strange about it. For it would be ... completely counter to our intuition to suppose that there is necessarily a gap, for example around the line $y = x$. We can always fill this gap in our thoughts with curves which accompany the line $y = x$ to infinity.*

However, he finds here

> *no irreconcilable conflict of the results of different forms of thought, but only one of the idea of a perhaps not very familiar but still not inaccessible spatial behaviour.*

For du Bois-Reymond this only indicates the presence of "a gap between in the analogy between ordinary and infinitary quantities", the manifestation of "a behaviour peculiar to the infinitary domain".

In his book *Die allgemeine Functionentheorie* of 1882 du Bois-Reymond presents his views on the nature and existence of infinitesimals. He begins by stating that in the analysis of "continuous mathematical quantities", one begins with a "geometric quantity" and tries to relate other quantities to it. So the finite decimals are assigned correlates on a segment, that is, "points". This correlation between finite decimals and points is then extended to infinite decimals by a limit process. But the totality of such points can never form a complete segment, since

> *points are just dimensionless, and therefore an arbitrarily dense sequence of points can never become a distance*.

Here we see once again a rejection of the idea that the continuous is reducible to the discrete. Consequently, a geometric segment must contain something other than finite and infinite decimals. These "others", according to du Bois-Reymond, are infinitesimal segments: there are infinitely many of these in any line segment, however short.

Du Bois-Reymond provides just a few rules of calculation for infinitesimal segments, reminiscent of those used by 17^{th} and 18^{th} mathematicians. To wit:

> *A finite number of infinitely small segments joined to one another do not form a finite segment, but again an infinitely small segment ... no upper bound can be specified either for the finite or for the infinitely small.*

> *I say two finite segments are equal when there is no* finite *difference between them ... Two finite quantities whose difference is infinitely small are equal to one another ... A finite quantity does not change if an infinitely small quantity is added to it or taken away from it.*

While we may be incapable of forming a mental image of the *relation* of the infinitesimal to the finite, according to du Bois-Reymond we can visualize the infinitesimal *in itself*, and when we do so we find that it behaves just like the finite:

> *The infinitely small is a mathematical quantity and has all its properties in common with the finite.*

Du Bois-Reymond took a dim view of some mathematicians' conception of infinitesimals as being ordinary magnitudes continually in a state of flux towards zero, remarking sarcastically

> *As long as the book is closed there is perfect repose, but as soon as I open it there commences a race of all the magnitudes which are provided with the letter d towards the zero limit.* (Ehrlich 1994, pp. 9-10.)

The admission of the infinitesimal in relation to the *finite* opens the way to the infinitesimal in relation to the *infinitesimal* (so entailing, reciprocally, the presence of the infinitely large):

> *In this way, there arises a series of types of quantities, whose successive relation always is that a finite number of quantities of one kind never yields a quantity of the preceding kinds.*

Such quantities accordingly form a nonarchimedean domain. Moreover,

> *If within one and the same of these types of quantities, the properties of ordinary mathematical quantities hold, hence the same types of calculation as in the finite, then the comparison of the different types of quantities with each other is the object of the so-called infinitary* calculus. *This calculus reckons with the relations of the infinitely large or infinitely small from type to type, and these types show connections with each other that do not fall under the ordinary*

concept of equality. The passages of one type into another do not show, for example, the continuity of change of mathematical quantities, although no jump changes result.

Du Bois-Reymond concludes his musings on the infinitesimal with the observation that there is an imbalance between belief in the infinitely large and belief in the infinitely small. A majority of educated people, he says, will admit an "infinite" (i.e., actual infinite) in space and time, and not just an "unboundedly large" (i.e., potential infinite). But only with difficulty will they accept the infinitely small, despite the fact that it has the same "right to existence" as the infinitely large. In sum,

A belief in the infinitely small does not triumph easily. Yet when one thinks boldly and freely, the initial mistrust will soon mellow into a pleasant certainty.... Were the sight of the starry sky lacking to mankind; had the race arisen and developed troglodytically in enclosed spaces; had its scholars, instead of wandering through the distant places of the universe telescopically, only looked for the smallest constituents of form and so were used in their thoughts to advancing into the boundless in the direction of the unmeasurably small: who would doubt that then the infinitely small would take the same place in our system of concepts that the infinitely large does now? Moreover, hasn't the attempt in mechanics to go back down to the smallest active elements long ago introduced into science the atom, the embodiment of the infinitely small? And don't as always skilful attempts to make it superfluous for physics face with certainty the same fate as Lagrange's battle against the differential?

Du Bois-Reymond was, indeed, a doughty champion of the infinitesimal.

In his later years the Austrian philosopher **Franz Brentano** (1838-1917) became preoccupied with the nature of the continuous. Much of Brentano's philosophy has its starting-point in Aristotelian doctrine, and his conception of the continuum constitutes no exception. Aristotle's theory of the continuum, it will be recalled, rests upon the assumption that all change is continuous and that continuous variation of quality, of quantity and of position are inherent features of perception and intuition. Aristotle considered it self-evident that a continuum cannot consist of points. Any pair of unextended points, he observes, are such that they either touch or are totally separated: in the first case, they yield just a single unextended point, in the second, there is a definite gap between the points. Aristotle held that any continuum—a continuous path, say, or a temporal duration, or a motion—may be divided *ad infinitum* into other continua but not into what might be called "discreta"—parts that cannot themselves be further subdivided. Accordingly, paths may be divided into shorter paths, but not into unextended points; durations into briefer durations but not into unextended instants; motions into smaller motions but not into unextended "stations". Nevertheless, this does not prevent a continuous line from being divided at a point constituting the common border of the line segments it divides. But such points are, according to Aristotle, just *boundaries,* and not to be regarded as actual *parts* of the continuum from which they spring. If two continua have a common boundary, that common border unites them into a single continuum. Such boundaries exist only *potentially*, since they come into being when they are, so to speak, marked out as connecting parts of a continuum; and the parts in their turn are similarly dependent as parts upon the existence of the continuum.

In its fundamentals Brentano's account of the continuous is akin to Aristotle's. Brentano regards continuity as something given in perception, primordial in nature, rather than a mathematical construction. He held that the idea of the continuous is a fundamental notion abstracted from sensible 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 (Brentano 1988, p. 6).

Brentano suggests that the continuous is brought to appearance by sensible intuition in three stages. First, sensation presents us with objects having parts that coincide. From such objects the concept of *boundary* is abstracted in turn, and then one grasps that these objects actually *contain* coincident boundaries. Finally, it becomes clear that this is all that is required in order to understand the concept of a continuum.

Continuity is manifested in sensation in a variety of ways. In *visual* sensation, we are presented with *extension*, something possessing length and breadth, and hence with something such that between any two of its parts, provided these are separated, there is a third part. Every sensation possesses a certain *qualitative* continuity in that the object presented in the sensation could have a given manifested quality (colour, for example) in a greater or less degree, and between any two degrees of that quality lies still another degree of that quality. Finally, each sensation manifests *temporal* continuity: this is most evident when we perceive something as moving or at rest.

Brentano recognizes that continua have *qualities* which cause them to possess *multiplicity*—a continuum may manifest continuity in several ways simultaneously. This led him to classify continua into *primary* and *secondary*: a secondary continuum being one whose manifestation is dependent upon another continuum. Here is Brentano himself on the matter:

Imagine, for example, a coloured surface. Its colour is something from which the geometer abstracts. For him there comes into consideration only the constantly changing manifold of spatial differences. But the colour, too, appears extended with the spatial surface, whether it manifests no specific colour-differences of its own—as in the case of a red colour which fills out a surface uniformly—or whether it varies in its colouring—perhaps in the manner of a rectangle which begins on one side with red and ends on the other side with blue, progressing uniformly through all colour-differences from violet to pure blue in between. In both cases we have to do with a multiple continuum, and it is the spatial continuum which appears thereby as primary, the colour-

continuum as secondary. A similar double continuum can also be established in the case of a motion from place to place or of a rest, in which case it is a temporal continuum as such that is primary, the temporally constant or varying place that is the secondary continuum. Even when one considers a boundary of a mathematical body as such, for example a curved or straight line, a double continuity can be distinguished. The one presents itself in the totality of the differences of place that are given in the line, which always grows uniformly, whether in the case of straight, bent, or curved lines, and is that which determines the length of the line. The other resides in the direction of the line, and is either constant or alternating, and may vary continuously, or now more strongly, now less. It is constant in the case of the straight line, changing in the case of the broken line, and continuously varying in every line that is more or less curved. The direction-continuum here is to be compared with the colour-continuum discussed earlier and with the continuum of place in the case of rest or motion of a corporeal point in time. In the double continuum that presents itself to us in the line it is this continuum of directions that is to be referred to as the secondary, the manifold of differences of place as such as the primary continuum (Brentano 1988, p. 21f).

It is worth mentioning that Brentano's distinction of primary and secondary continua can be neatly represented within *category theory*: to put it succinctly, a *primary continuum* is a *domain,* a *secondary continuum* a *codomain.* We form a category C —the *category of continua*— by taking continua as objects and *correlations* between continua as arrows. Then, given any arrow $f\colon A \to B$ in C, the domain A of f may be taken as a "primary" continuum and its codomain B as a "secondary" continuum. In Brentano's example of a coloured surface, for instance, the primary continuum A is the given spatial surface, the secondary continuum B is the colour spectrum, and the correlation f assigns to each place in A its colour as a position in B. In the case of a corporeal point moving in space, the primary continuum A is an interval of time, the secondary continuum B a region of space, and the correlation f assigns to each instant in A the position in B occupied by the corporeal point. Finally, in the case of the varying direction of a curve the primary continuum A is the curve itself, the secondary continuum is the continuum of measures of angles, and the correlation *f* assigns to each point on the curve the slope of the tangent there: thus *f* is nothing other than the *first derivative* of the function associated with the curve.

For Brentano the essential feature of a continuum is its inherent capacity to engender boundaries, and the fact that such boundaries can be grasped as coincident. Boundaries themselves possess a quality which Brentano calls *plerosis* ("fullness"). Plerosis is the measure of the number of directions in which the given boundary actually bounds. Thus, for example, within a temporal continuum the endpoint of a past episode or the starting point of a future one bounds in a single direction, while the point marking the end of one episode and the beginning of another may be said to bound doubly. In the case of a spatial continuum there are numerous additional possibilities: here a boundary may bound in all the directions of which it is capable of bounding, or it may bound in only some of these directions. In the former case, the boundary is said to exist in *full plerosis*; in the latter, in *partial plerosis.* Brentano writes:

…the spatial nature of a point differs according to whether it serves as a limit in all or only in some directions. Thus a point located inside a physical thing serves as a limit in all directions, but a point on a surface or an edge or a vertex serves as a limit in only some direction. And the point in a vertex will differ in accordance with the directions of the edges that meet at the vertex… I call these specific distinctions differences of plerosis. Like any manifold variation, plerosis admits of a more and a less. The plerosis of the centre of a cone is more complete than that of a point on its surface; the plerosis of a point on its surface is more complete than that of a point on its edge, or that of its vertex. Even the plerosis of the vertex is the more complete the less the cone is pointed. (Brentano 1988, p. xvii).

Brentano believed that the concept of plerosis enabled sense to be made of the idea that a boundary possesses "parts", even when the boundary lacks dimensions altogether, as in the case of a point. Thus, while the present or "now" is, according to Brentano, temporally unextended and exists only as a boundary between past and future, it still possesses two "parts" or aspects: it is both the end of the past and the beginning of the future. It is worth mentioning that for Brentano it was not just the "now" that existed only as a boundary; since, like Aristotle he held that "existence" in the strict sense means "existence *now*", it necessarily followed that existing things exist only as boundaries of what has existed or of what will exist, or both.

Brentano ascribes particular importance to the fact that points in a continuum can *coincide.* On this matter he writes:

Various other thorough studies could be made [on the continuum concept] *such as a study of the impossibility of adjacent points and the* possibility of coincident points*, which have, despite their coincidence, distinctness and full relative independence.* [This] *has been and is misunderstood in many ways. It is commonly believed that if four different-coloured quadrants of a circular area touch each other at its centre, the centre belongs to only one of the coloured surfaces and must be that colour only. Galileo's judgment on the matter was more correct; he expressed his interpretation by saying paradoxically that the centre of the circle has as many parts as its periphery. Here we will only give some indication of these studies by commenting that everything which arises in this connection* follows from the point's relativity as involves a continuum and the fact that it is essential for it to belong to a continuum. *Just as the possibility of the coincidence of different points is connected with that fact, so is the existence of a point in diverse or more or less perfect* plerosis. *All of this is overlooked even today by those who understand the continuum to be an actual infinite multiplicity and who believe that we get the concept not by abstraction from spatial and temporal intuitions but from the combination of fractions between numbers, such as between 0 and* 1 (Brentano 1974, p. 357).

Brentano's doctrines of plerosis and coincidence of points are well illustrated by applying them to the traditional philosophical problem of the initiation of motion: if a thing begins to move, is there a last moment of its being at rest or a first moment of its being in motion?

The usual objection to the claim that both moments exist is that, if they did, there would be a time between the two moments, and at that time the thing could be said neither to be at rest nor to be in motion—in violation of the law of excluded middle. Brentano's response would be to say that both moments do exist, but that they *coincide*, so that there are no times between them; the violation of the law of excluded middle is thereby avoided. More exactly, Brentano would assert that the temporal boundary of the thing's being at rest—the end of its being at rest—is the same as the temporal boundary of the thing's being in motion—the beginning of its being in motion—, but the boundary is *twofold* in respect of its plerosis. The boundary is, in fact, in half plerosis at rest and in half plerosis in motion.

Brentano took a dim view of the efforts of mathematicians to construct the continuum from numbers. His attitude varied from rejecting such attempts as inadequate to according them the status of "fictions". In a letter to Husserl drafted in 1905, Brentano asserts that "I regard it as absurd to interpret a continuum as a set of points." This is not surprising given his Aristotelian inclination to take mathematical and physical theories to be genuine descriptions of empirical phenomena rather than idealizations: in his view, if such theories were to be taken as literal descriptions of experience, they would amount to nothing better than "misrepresentations". Indeed, Brentano writes:

We must ask those who say that the continuum ultimately consists of points what they mean by a point. Many reply that a point is a cut which divides the continuum into two parts. The answer to this is that a cut cannot be called a thing and therefore cannot be a presentation in the strict and proper sense at all. We have, rather, only presentations of contiguous parts. … The spatial point cannot exist or be conceived of *in isolation. It is just as necessary for it to belong to a spatial continuum as for the moment of time to belong to a temporal continuum* (Brentano 1974, p. 354).

Concerning Poincaré's approach to the continuum (see below) Brentano has this to say:

Poincaré … follows extreme empiricists in the in the area of sensory psychology and therefore does not believe that there is granted to us an intuition of a continuous space. Poincaré's entire mode of procedure reveals that he also denies that we are in possession of an intuition of a continuous time. We saw how first of all he inserted between 0 and 1 fractions having a whole number as numerator and a whole power of 2 as denominator. In similar fashion, he then inserted all proper fractions whose denominator is a whole power of 3, and then also all those whose denominators are powers of every other whole number. He obtained thereby a series containing all rational fractions which, as he said, already has a certain continuity about it. He then inserted … a series of irrational fractions. To these one now adds the series of fractions involving transcendental ratios… . Poincaré was prepared to admit that this process will never come to an end… . But he believed that he could be satisfied with the insertions already made. And nothing is more self-evident than that we have here a confession that the attempt to obtain a true continuum in this way has broken down (Brentano 1988, p. 39).

Dedekind's account of the continuum does not fare much better (Brentano 1988, pp. 40-41):

> *Dedekind differs from Poincaré already in the fact that he does not wish to deny that we have an intuition of a continuum – he simply does not want to make any use thereof. … Dedekind's and Poincaré's constructions share in common that they fail to recognise the essential character of the continuum, namely that it allows the distinguishing of boundaries, which are nothing in themselves, but yet in conjunction make a contribution to the continuum. Dedekind believes that either the number ½ forms the beginning of the series ½ to 1, so that the series 0 to ½ would thereby be spared a final member, i.e. an end point which would belong to it, or conversely. But this is not how things are in the case of a true continuum. Rather it is the case that, when one divides a line, every part has a starting point, but in half plerosis. …*

Here Brentano appears to be saying that when one divides a closed interval [*a, b*] at an intermediate point *c*, one necessarily obtains the closed intervals [*a, c*], [*c, b*], with the common point *c* (in half plerosis). In that case, Brentano have probably have regarded a continuous line as indecomposable into disjoint intervals.

He goes on:

If a red and a blue surface are in contact with each other then a red and a blue line coincide, each with different plerosis. And if a circular area is made up of three sectors, a red, a blue and a yellow, then the mid-point is a whole which consists to an equal extent of a red, a blue and a yellow part. According to Dedekind this point would belong to just one of the three colour-segments, and we should have to say that it could be separated from this while the segment in question remained otherwise unchanged. Indeed, the whole circular surface would then be conceivable as having been deprived of its mid-point, like Dedekind's number-series from which only the number ½ has fallen away. One sees immediately that this is absurd if one keeps in mind that the true concept of the continuum is obtained through abstraction from an intuition, and thus also that the entire conception has missed its target.

That Brentano considered "absurd" the idea of removing a single point from a continuum seems to indicate that his continuum has the same "syrupy" property as those of intuitionistic and smooth infinitesimal analysis: see sections 4 and 6 below.

In conclusion,

One sees that in this entire putative construction of the concept of what is continuous the goal has been entirely missed; for that which is above all else characteristic of a continuum, namely the idea of a boundary in the strict sense (to which belongs the possibility of a coincidence of boundaries), will be sought after entirely in vain. Thus also the attempt to have the concept of what is continuous spring forth out of the combination of individual marks distilled from intuition is to be rejected as entirely mistaken, and this implies further that what is continuous must be given to *us in individual intuition and must therefore have been extracted therefrom* (Brentano 1988, pp. 4f).

Brentano's analysis of the continuum centred on its phenomenological and qualitative aspects, which are by their very nature incapable of reduction to the discrete. Brentano's rejection of the mathematicians' attempts to construct it in discrete terms is thus hardly surprising.

The American philosopher-mathematician **Charles Sanders Peirce**'s (1839–1914) view of the continuum was, in a sense, intermediate between that of Brentano and the arithmetizers. Like Brentano, he held that the cohesiveness of a continuum rules out the possibility of it being a mere collection of discrete individuals, or points, in the usual sense:

> *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* (Peirce 1976, p. 925).

And even before Brouwer (see below) Peirce seems to have been aware that a faithful account of the continuum will involve questioning 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).

But Peirce also held that any continuum harbours an *unboundedly large* collection of points—in his colourful terminology, a *supermultitudinous* collection—what we would today call a *proper class*. Peirce maintained that if "enough" points were to be crowded together by carrying insertion of new points between old to its ultimate limit they would—through a *logical* "transformation of quantity into quality"—lose their individual identity and become fused into a true continuum. Here are his observations on the matter:

It is substantially proved by Euclid that there is but one assignable quantity which is the limit of a convergent series. That is, if there is an increasing convergent series, A *say, and a decreasing convergent series,* B *say, of which every approximation exceeds every approximation of* A, *and if there is no rational quantity which is at once greater than every approximation of* A *and less than every approximation of* B, *then there is but one surd quantity so intermediate…There is one surd quantity and only one for each convergent series, calling two series the same if their approximations all agree after a sufficient number of terms, or if their difference approximates toward zero. But this is only to say that the multitude of surds equals the multitude of denumerable sets of rational numbers which is… the primipostnumeral multitude.* [Peirce assumed what amounts to the generalized continuum hypothesis in supposing that each possible infinite set has one of the cardinalities $\aleph_0, 2^{\aleph_0}, 2^{2^{\aleph_0}}, \ldots$. These he termed denumerable, primipostnumeral, secundipostnumeral, etc.]

> …*We remark that there is plenty of room to insert a* secundipostnumeral *multitude of quantities between [a] convergent series and its limit. Any one of those quantities may likewise be separated from its neighbours, and we thus see that between it and its nearest neighbours there is ample room for a* tertiopostnumeral *multitude of other quantities, and so on through the whole denumerable series of postnumeral quantities.*

But if we suppose that all *such orders of systems of quantities have been inserted, there is no longer any room for inserting any more. For to do so we must select some quantity to be thus isolated in our representation. Now whatever one we take, there will always be quantities of higher order filling up the spaces on the two sides.*

We therefore see that such a supermultitudinous collection sticks together by logical necessity. Its constituent individuals are no longer distinct and independent subjects. They have no existence—no hypothetical existence—except in their relations to one another. They are not subjects, but phrases expressive of the properties of the continuum.

…Supposing a line to be a supermultitudinous collection of points, … to sever a line in the middle is to disrupt the logical identity of the point there and make it two points. It is impossible to sever a continuum by separating the connections of the points, for the points only exist by virtue of those connections. The only way to sever a continuum is to burst it, that is, to convert what was one into two (Peirce 1976, p. 95).

There is some resemblance between Peirce's conception of the continuum and John Conway's system of *surreal numbers* (Ehrlich 1994a). Conway's system may be characterized as being an η_α-field for every ordinal α, that is, a real-closed ordered field S which satisfies the condition that, for any pair of subsets X, Y for which every member of X is less than every member of Y, there is an element of S strictly between X and Y.

In their Introduction to Peirce (1992), Ketner and Putnam characterize Peirce's conception of the continuum as "a possibility of repeated division which can never be exhausted in any possible world, not even in a possible world in which one can complete [nondenumerably] infinite processes."(This description would seem to apply equally well to Conway's conception.) It is not hard to show that, between any pair of members of S there is a *proper class* of members of S—in Peirce's terminology, a supermultitudinous collection. Nevertheless, S is still discrete: its elements, while supermultitudinous, remain distinct and unfused (were it not for this fact, Conway would scarcely be justified in calling the members of *S* "numbers"). On the face of it the discreteness of *S* would seem to imply that the presence of superabundant quantity in Peirce's sense is not enough to ensure continuity. Of course, Brentano would have dismissed this idea altogether, in view of his critical attitude towards any construction of the continuum by repeated insertion of points.

Peirce's conception of the number continuum is also notable for the presence in it of an abundance of *infinitesimals*, a feature it shares with du Bois-Reymond's nonarchimedean number systems. In defending infinitesimals, Peirce remarks that

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

Peirce held the view that the conception of infinitesimal is suggested by introspection— that the *specious present* is in fact an infinitesimal:

> *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* (Peirce 1976, p. 124).

> *We are conscious of the present time, which is an instant, if there be any such thing as an instant. But in the present we are conscious of the flow of time. There is no flow in an instant. Hence, the present is not an instant.* (Peirce 1976, p. 925).

Peirce also championed the retention of the infinitesimal concept in the foundations of the calculus, both because of what he saw as the efficiency of infinitesimal methods, and because he regarded infinitesimals as constituting the "glue" causing points on a continuous line to lose their individual identity.

The idea of continuity played a central role in the thought of the great French mathematician **Henri Poincaré** (1854–1912). But sorting out his views on the continuum, concerning which he made numerous scattered remarks, is by no means an easy task. Indeed, there seems to be an inconsistency in his attitude towards the set-theoretical, or arithmetized, continuum. On the one hand, he rejected actual infinity and impredicative definition—both cornerstones of the Cantorian theory of sets which underpins the construction of the arithmetized continuum. And yet in his mathematical work he employs variables ranging over all the points of an interval of the set-theoretical continuum, and he "accepts the standard account of the least upper bound, which is impredicative" (Folina 1992, p. xv). (Impredicativity is a form of circularity: a definition of a term is impredicative if it contains a reference to a totality to which the term under definition belongs. See, e.g., Fraenkel, Bar-Hillel and Levy 1973, pp. 193-200.)

Beneath this apparent inconsistency lies his belief that what ultimately underpins mathematics, creating its linkage with objective reality, is *intuition*—that "intuition is what bridges the gap between symbol and reality" (Folina 1992, p. 113).

His view of the continuum, in particular, is informed by this credo. For Poincaré the continuum and the range of points on it is grasped in intuition in something like the Kantian sense, and yet the continuum cannot be treated as a completed mathematical object, as a "mere set." (Folina 1992, p. xvi).

Of the arithmetical continuum Poincaré remarks:

The continuum so conceived is only a collection of individuals ranged in a certain order, infinite to one another, it is true, but exterior *to one another. This is not the ordinary conception, wherein* *is supposed between the elements of the continuum a sort of intimate bond which makes of them a whole, where the point does not exist before the line, but the line before the point. Of the celebrated formula "the continuum is unity in multiplicity", only the multiplicity remains, the unity has disappeared. The analysts are none the less right in defining the continuum as they do, for they always reason on just this as soon as they pique themselves on their rigor. But this is enough to apprise us that the veritable mathematical continuum is a very different thing from that of the physicists and the metaphysicians.* (Poincaré 1946, pp. 43-44).

But despite Poincaré's apparent acceptance of the arithmetic definition of the continuum, he questions the fact that (as with Dedekind and Cantor's formulations) the (irrational) numbers so produced are mere symbols, detached from their origins in intuition:

But to be content with this [fact] would be to forget too far the origin of these symbols; it remains to explain how we have been led to attribute to them a sort of concrete existence, and, besides, does not the difficulty begin even for the fractional numbers themselves? Should we have the notion of these numbers if we had not known a matter that we conceive as infinitely divisible, that is to say, a continuum? (Poincaré 1946, pp. 45-46).

That being the case, Poincaré asks whether the notion of the mathematical continuum is "simply drawn from experience." To this he responds in the negative, for the reason that our sensations, the "raw data of experience", cannot be brought under an acceptable scheme of measurement:

> *It has been observed, for example, that a weight A of 10 grams and a weight B of 11 grams produce identical sensations, that the weight B is just as indistinguishable from a weight C of 12 grams, but that the weight A is easily distinguished from the weight C. Thus the raw results of experience may be expressed by the following relations:*

$$A = B,\ \ B = C,\ \ A < C,$$

which may be regarded as the formula of the physical continuum.

According to Poincaré it is the "intolerable discord with the principle of contradiction" of this formula which has forced the invention of the mathematical continuum. (Actually, the formula ceases to be contradictory if the identity relation = is replaced by a symmetric, reflexive, but *nontransitive* relation ≈: here $x \approx y$ is taken to assert that the sensations or perceptions x and y are indistinguishable.) The continuum is then obtained in two stages. First, formerly indistinguishable terms are distinguished and a new term, indistinguishable from both, inserted between them. Repeating this procedure indefinitely gives rise to what Poincaré calls a *first-order continuum*, in essence the rational number line. A second stage now becomes necessary because two first-order continua, for example the diagonal of a square and its inscribed circle, need not intersect. This second stage, in which are added all possible "boundary" points between first-order continua leads to the second-order or mathematical continuum. Here is how Poincaré describes the process:

But conceive of a straight line divided into two rays. Each of these rays will appear to our imagination as a band of a certain breadth; these bands moreover will encroach one on the other, since there must be no interval between them. The common part will appear to us as a point which will always remain when we try to imagine our bands narrower and narrower, so that we admit as an intuitive truth that if a straight line is cut into two rays their common boundary is a point; we recognize here the conception of Dedekind, in which an incommensurable number was regarded as the common boundary of two classes of rational numbers.

Such is the origin of the continuum of second order, which is the mathematical continuum so called.

Poincaré goes on to discuss continua of higher dimensions. To obtain these he considers *aggregates* of sensations. As with single sensations, any given pair of these aggregates may or may not be distinguishable. He remarks that, while these aggregates, which he terms *elements*, are analogous to mathematical points, they are not in fact quite the same thing, for

> *we cannot say that our element is without extension, since we cannot distinguish it from neighbouring elements and it is thus surrounded by a sort of haze. If the astronomical comparison may be allowed, our 'elements' would be like nebulae, whereas the mathematical points would be like stars* (Poincaré 1946, p. 49).

This leads to a definition of a physical continuum:

> *a system of elements will form a* continuum *if we can pass from any one of them to any other, by a series of consecutive elements such that each is indistinguishable from the preceding. This* linear series *is to the* line *of the mathematician what an isolated* element *was to the point.*

Poincaré defines a *cut* in a physical continuum *C* to be a set of elements removed from it "which for an instant we shall regard as no longer belonging to this continuum." Such a cut may happen to subdivide *C* into several distinct continua, in which case *C* will contain two distinct elements *A* and *B* that must be regarded as belonging to two distinct continua. This becomes necessary

> *because it will be impossible to find a linear series of consecutive elements of C, each of these elements indistinguishable from the preceding, the first being A and the last B,* without one of the elements of this series being indistinguishable from one of the elements of the cut.

On the other hand, it may happen that the cut fails to subdivide the continuum *C*, in which case it becomes necessary to determine precisely which cuts will subdivide it. Poincaré calls a continuum *one-dimensional* if it can be subdivided by a cut which reduces to a finite number of elements all distinguishable from one another (and so forming neither a continuum nor several continua). When *C* can be subdivided only by cuts which are themselves continua, *C* is said to possess several dimensions:

> *If cuts which are continua of one dimension suffice, we shall say that C has two dimensions; if cuts of two dimensions suffice, we shall say that C has three dimensions, and so on.*

Thus is defined the concept of a multidimensional physical continuum, based on "the very simple fact that two aggregates of sensations are distinguishable or indistinguishable."

Unlike Cantor, Poincaré accepted the infinitesimal, even if he did not regard all of the concept's manifestations as useful. This emerges from his answer to the question: "Is the creative power of the mind exhausted by the creation of the mathematical continuum?". He responds:

> *No; the works of Du Bois-Reymond demonstrate it in a striking way. We know the mathematicians distinguish between infinitesimals and that those of second order are infinitesimal not only in an absolute way, but also in relation to those of first order. It is not difficult to imagine infinitesimals of fractional and even irrational order, and thus we find again that scale of the mathematical continuum which has been dealt with in the preceding pages.*
>
> *Further, there are infinitesimals which are infinitely small in relation to those of the first order, and, on the contrary, infinitely great in relation to those of order* $1 + \varepsilon$, *and that however small*

> *ε may be. Here, then, are new terms intercalated in our series ... I shall say that thus has been created a sort of continuum of the third order.*
>
> *It would be easy to go further, but that would be idle; one would only be imagining symbols without possible application, and no one would think of doing that. The continuum of the third order, to which the consideration of the different orders of infinitesimals leads, is itself not useful enough to have won citizenship, and geometers regard it as a mere curiosity. The mind uses its creative faculty only when experience requires it.*

Poincaré's attitude towards the continuum resembles in certain respects that of the intuitionists (see below): while the continuum exists, and is knowable intuitively, it is not a "completed" set-theoretical object. It is geometric intuition, not set theory, upon which the totality of real numbers is ultimately grounded.

The Dutch mathematician **L. E. J. Brouwer** (1881–1966) is best known as the founder of the philosophy of (neo)*intuitionism*. Brouwer's highly idealist views on mathematics bore some resemblance to Kant's. For Brouwer, mathematical concepts are admissible only if they are adequately grounded in intuition, mathematical theories are significant only if they concern entities which are constructed out of something given immediately in intuition, and mathematical demonstration is a form of construction in intuition. Brouwer's insistence that mathematical proof be constructive in this sense required the jettisoning of certain received principles of classical logic, notably the *law of excluded middle*: the assertion that, for any proposition *p*, either *p* or not *p*. Brouwer maintained, in fact, that the applicability of the law of excluded middle to mathematics

was caused historically by the fact that, first, classical logic was abstracted from the mathematics of the subsets of a definite finite set, that, secondly, an a priori existence independent of *mathematics was ascribed to the logic, and that, finally, on the basis of this supposed apriority it was unjustifiably applied to the mathematics of infinite sets (*Kneebone 1963,p. 246).

Brouwer held that much of modern mathematics is based on an illicit extension of procedures valid only in the restricted domain of the finite. He therefore embarked on the radical course of jettisoning virtually all of the mathematics of his day—in particular the set-theoretical construction of the continuum—and starting anew, using only concepts and modes of inference that could be given clear intuitive justification. In the process it would become clear precisely what are the logical laws that intuitive, or constructive, mathematical reasoning actually obeys, making possible a comparison of the resulting *intuitionistic,* or *constructive logic* with classical logic. This is not to say that Brouwer was primarily interested in *logic,* far from it: indeed, his distaste for formalization caused him to be quite dismissive of subsequent codifications of intuitionistic logic.

While admitting that the emergence of noneuclidean geometry had discredited Kant's view of space, Brouwer maintained, in opposition to the logicists (whom he called "formalists") that arithmetic, and so all mathematics, must derive from *temporal intuition.* In his own words*:*

Neointuitionism considers the falling apart of moments of life into qualitatively different parts, to be reunited only while remaining separated by time, as the fundamental phenomenon of the human intellect, passing by abstracting from its emotional content into the fundamental phenomenon of mathematical thinking, the intuition of the bare two-oneness. This intuition of two-oneness, the basal intuition of mathematics, creates not only the numbers one and two, but also all finite ordinal numbers, inasmuch as one of the elements of the two-oneness may be thought of as a new two-oneness, which process may be repeated indefinitely; this gives rise still further to the smallest infinite ordinal ω *. Finally this basal intuition of mathematics, in which the connected and the separate, the continuous and the discrete are united, gives rise immediately to the intuition of the* *linear continuum, i.e., of the "between", which is not exhaustible by the interposition of new units and which can therefore never be thought of as a mere collection of units. In this way the apriority of time does not only qualify the properties of arithmetic as synthetic a priori judgments, but it does the same for those of geometry, and not only for elementary two- and three-dimensional geometry, but for non-euclidean and n-dimensional geometries as well. For since Descartes we have learned to reduce all these geometries to arithmetic by means of coordinates. (*Benacerraf and Putnam 1977, p. 80).

Brouwer maintained that it is the awakening of awareness of the temporal continuum in the subject, an event termed by him "The Primordial Happening" or "The Primordial Intuition of Time", that engenders the fundamental concepts and methods of mathematics. In "Mathematics, Science and Language" (1929), he describes how the notion of number—the discrete—emerges from the awareness of the continuous:

Mathematical Attention *as an act of the will serves the instinct for self-preservation of individual man; it comes into being in two phases;* time awareness *and* causal attention. *The first phase is nothing but the fundamental intellectual phenomenon of the falling apart of a moment of life into two qualitatively different things of which one is experienced as giving away to the other and yet is retained by an act of memory. At the same time this split moment of life is separated from the Ego and moved into a world of its own, the world of perception. Temporal twoity, born from this time awareness, or the two-membered sequence of time phenomena, can itself again be taken as one of the elements of a new twoity, so creating temporal threeity, and so on. In this way, by means of the self-unfolding of the fundamental phenomenon of the intellect, a* time sequence of phenomena is created of arbitrary multiplicity (Mancosu 1998, p.45)

Brouwer took the elucidation of the continuum as his *raison d'être*. In his early thought he rejected out of hand the set-theoretic account of the continuum in terms of discrete points as put forward by Cantor and Dedekind, asserting that

…*The* continuum as a whole [*is*] *given to us by intuition; a construction for it, an action which would create from the mathematical intuition 'all' its points as individuals, is inconceivable and impossible.* (Mancosu 1998, p.45)

Later Brouwer was to modify this doctrine. In his mature thought, he radically transformed the concept of point, endowing points with sufficient fluidity to enable them to serve as generators of a "true" continuum. This fluidity was achieved by admitting as "points", not only fully defined discrete numbers such as $\sqrt{2}$, π, *e*, and the like—which have, so to speak, already achieved "being"—but also "numbers" which are in a perpetual state of becoming in that their the entries in their decimal (or dyadic) expansions are the result of free acts of choice by a subject operating throughout an indefinitely extended time. The resulting *choice sequences* cannot be conceived as finished, completed objects: at any moment only an initial segment of each is known (Fraenkel, Bar-Hillel and Levy (1973), pp. 255-261). In this way Brouwer obtained the mathematical continuum in a way compatible with his belief in the primordial intuition of time—that is, as an unfinished, indeed unfinishable entity in a perpetual state of growth, a "medium of free development". In this conception, the mathematical continuum is indeed "constructed", not, however, by initially shattering, as did Cantor and Dedekind, an intuitive continuum into isolated points, but rather by assembling it from a complex of continually changing overlapping parts.

The mathematical continuum as conceived by Brouwer displays a number of features that seem bizarre to the classical eye. For example, in the Brouwerian continuum the usual law of comparability, namely that for any real numbers *a, b* either $a < b$ or $a = b$ or $a > b$, fails. Even more fundamental is the failure of the law of excluded middle le in the form that for any real numbers *a, b,* either $a = b$ or $a \neq b$. The failure of these seemingly unquestionable principles in turn vitiates the proofs of a number of basic results of classical analysis, for example the Bolzano-Weierstrass theorem, as well as the theorems of monotone convergence, intermediate value, least upper bound, and maximum value for continuous functions. The failure of these important results of classical analysis in caused most mathematicians of the day to shun intuitionistic, and even constructive mathematics. It was not until the 1960s that adequate constructive versions were worked out (see section 4 below).

While the Brouwerian continuum may possess a number of negative features from the standpoint of the classical mathematician, it has the merit of corresponding more closely to the continuum of intuition than does its classical counterpart. Hermann Weyl pointed out a number of respects in which this is so:

In accordance with intuition, Brouwer sees the essential character of the continuum, not in the relation between element and set, but in that between part and whole. The continuum falls under the notion of the 'extensive whole', which Husserl characterizes as that "which permits a dismemberment of such a kind that the pieces are by their very nature of the same lowest species as is determined by the undivided whole (Weyl 1949, p. 52).

Far from being bizarre, the failure of the law of excluded middle for points in the intuitionistic continuum is seen by Weyl as "fitting in well with the character of the intuitive continuum":

For there the separateness of two places, upon moving them toward each other, slowly and in vague gradations passes over into indiscernibility. In a continuum, according to Brouwer, there can be only continuous functions. The continuum is not composed of parts (Weyl 1949, p. 54).
For Brouwer had indeed shown, in 1924, that every function defined on a closed interval of the continuum as he conceived it is uniformly continuous. Now one might be inclined

to regard this claim as impossible: is not a counterexample provided by, for example, the function f given by $f(0) = 0$, $f(x) = |x|/x$ otherwise? No, because from the intuitionistic standpoint, this function is not everywhere defined on the interval [–1, 1], being undefined at those arguments x for which it is unknown whether $x = 0$ or $x \neq 0$.

As a consequence, the intuitionistic continuum is *indecomposable* or *cohesive*, that is, *it cannot be split into two nonempty disjoint parts in any way whatsoever.* In contrast with a discrete entity, the cohesive Brouwerian continuum cannot be composed of its parts.

Hermann Weyl (1885–1955), one of most important and versatile mathematicians of the 20th century, devoted a great deal of thought to the nature of the continuum. In his *Philosophy of Mathematics and Natural Science* he reflects on what he calls the "inwardly infinite" nature of a continuum:

The essential character of the continuum is clearly described in this fragment of Anaxagoras: "Among the small there is no smallest, but always something smaller. For what is cannot cease to be no matter how small it is being subdivided." The continuum is not composed of discrete elements which are "separated from one another as though chopped off by a hatchet." Space is infinite not only in the sense that it never comes to an end; but at every place it is, so to speak, inwardly infinite, inasmuch as a point can only be fixed as step-by-step by a process of subdivision which progresses ad infinitum. *This is in contrast with the resting and complete existence that intuition ascribes to space. The "open" character is communicated by the continuous space and the continuously graded qualities to the things of the external world. A real thing can never be given adequately, its "inner horizon" is unfolded by an infinitely continued process of ever new and more exact experiences; it is, as emphasized by Husserl, a limiting idea in the Kantian sense. For this reason it is impossible to posit the real thing as existing, closed and complete in itself. The continuum problem thus drives one to epistemological idealism. Leibniz, among others, testifies that it was the search for a way out of the "labyrinth of the continuum" which first suggested to him the conception of space and time as orders of phenomena* (Weyl 1949, p. 41).

Weyl identifies three attempts in the history of thought "to conceive of the continuum as Being in itself". These are, respectively, *atomism, the infinitely small,* and *set theory*. In Weyl's view, despite atomism's brilliant success in unravelling the structure of matter, it had failed in that regard as to space, time, and mathematical extension because it "never achieved sufficient contact with reality." As for the infinitely small, it was not so much supplanted as rendered superfluous by the limit concept. Weyl saw the limit concept as providing the necessary link between the microcosm of the infinitely small and the realm of macroscopic objects. Without that link, the fact that the microcosm is governed by "elementary laws" making for ease of calculation, would remain entirely useless in drawing conclusions about the macrocosm (but in this connection see the remarks on the Constancy Principle in section 6).

In the case of set theory Weyl saw the danger in employing it a basis for explicating the continuum as lying in the fact that set theory bestows a finished, completed character on "the places in the continuum, i.e. to the possible sequences or sets of natural numbers." In Weyl's view this was a double error, for neither the aggregate of sets of natural numbers, nor (in general) individual such sets can be considered finished entities. Rather the continuum should be considered as an essentially incompletable "field of constructive possibilities" To suppose otherwise is to risk running up against set-theoretic paradoxes such as Russell's.

During the period 1918–1921 Weyl wrestled with the problem of providing the continuum with an exact mathematical formulation free of objectionable set-theoretic assumptions. As he saw it in 1918, there is an unbridgeable gap between intuitively given continua (e.g. those of space, time and motion) on the one hand, and the discrete exact concepts of mathematics (e.g. that of real number) on the other. For Weyl the presence of this gap meant that the construction of the mathematical continuum could not simply be "read off" from intuition. Rather, he believed at this time that the mathematical continuum must be treated as if it were an element of the realm immediately accessible to intuition, and so, in the end, justified in the same way as a physical theory. In Weyl's view, it was not enough that the mathematical theory of the continuum be *consistent*; it must also be *reasonable.*

Das Kontinuum (1918) embodies Weyl's attempt at formulating a theory of the continuum which satisfies the first, and, as far as possible, the second, of these requirements. In the following passages from this work he acknowledges the difficulty of the task:

> *... the conceptual world of mathematics is so foreign to what the intuitive continuum presents to us that the demand for coincidence between the two must be dismissed as absurd.*
>
> *... the continuity given to us immediately by intuition (in the flow of time and of motion) has yet to be grasped mathematically as a totality of discrete "stages" in accordance with that part of its content which can be conceptualized in an exact way.*
>
> *Exact time- or space-points are not the ultimate, underlying atomic elements of the duration or extension given to us in experience. On the contrary, only reason, which thoroughly penetrates what is experientially given, is able to grasp these exact ideas. And only in the arithmetico-analytic concept of the real number belonging to the purely formal sphere do these ideas crystallize into full definiteness.*
>
> *When our experience has turned into a real process in a real world and our phenomenal time has spread itself out over this world and assumed a cosmic dimension, we are not satisfied with replacing the continuum by the exact concept of the real number, in spite of the essential and undeniable inexactness arising from what is given.*

However much he may have wished it, in *Das Kontinuum* Weyl did not aim to provide a mathematical formulation of the continuum as it is presented to intuition, which, as the quotations above show, he regarded as an impossibility (at that time at least). Rather, his goal was first to achieve *consistency* by putting the *arithmetical* notion of real number on a firm logical basis, and then to show that the resulting theory is *reasonable* by employing it as the foundation for a plausible account of continuous process in the objective physical world.

Weyl had come to believe that mathematical analysis at the beginning of the 20th century would not bear logical scrutiny, for its essential concepts and procedures involved vicious circles to such an extent that, as he says, "every cell (so to speak) of this mighty organism is permeated by contradiction." In *Das Kontinuum* he tries to overcome this by providing analysis with a *predicative* formulation—not, as Russell and Whitehead had attempted in their *Principia Mathematica*, by introducing a hierarchy of logically ramified types, which Weyl seems to have regarded as too complicated—but rather by confining the basic principle of set formation to formulas whose bound variables range over just the initial given entities (numbers). Thus he restricts analysis to what can be done in terms of natural numbers with the aid of three basic logical operations, together with the operation of substitution and the process of "iteration", i.e., primitive recursion. Weyl recognized that the effect of this restriction would be to render unprovable many of the central results of classical analysis—e.g., Dirichlet's principle that any bounded set of real numbers has a least upper bound—but he was prepared to accept this as part of the price that must be paid for the security of mathematics.

In section 6 of *Das Kontinuum* Weyl presents his conclusions as to the relationship between the intuitive and mathematical continua. He poses the question: Does the mathematical framework he has erected provide an adequate representation of physical or temporal continuity as it is *actually experienced*? He begins his investigation by noting that, according to his theory, if one asks whether a given function is continuous, the answer is not fixed once and for all, but is, rather, dependent on the extent of the domain of real numbers which have been defined up to the point at which the question is posed. Thus the continuity of a function must always remain *provisional*; the possibility always exists that a function deemed continuous *now* may, with the emergence of "new" real numbers, turn out to be discontinuous *in the future*.

To reveal the discrepancy between this formal account of continuity based on real numbers and the properties of an intuitively given continuum, Weyl next considers the experience of seeing a pencil lying on a table before him throughout a certain time interval. The position of the pencil during this interval may be taken as a function of the time, and Weyl takes it as a fact of observation that during the time interval in question this function is continuous and that its values fall within a definite range. And so, he says,

> *This observation entitles me to assert that during a certain period this pencil was on the table; and even if my right to do so is not absolute, it is nevertheless reasonable and well-grounded. It is obviously absurd to suppose that this right can be undermined by "an expansion of our principles of definition"—as if new moments of time, overlooked by my intuition could be added to this interval, moments in which the pencil was, perhaps, in the vicinity of Sirius or who knows where. If the temporal continuum can be represented by a variable which "ranges over" the real numbers, then it appears to be determined thereby how narrowly or widely we must understand the concept "real number" and the decision about this must not be entrusted to* logical *deliberations over principles of definition and the like.*

To drive the point home, Weyl focuses attention on the fundamental continuum of *immediately given phenomenal time*, that is, as he characterizes it,

> *... to that constant form of my experiences of consciousness by virtue of which they appear to me to flow by successively.* (*By "experiences" I mean what I experience, exactly as I experience*

it. I do not mean real psychical or even physical processes which occur in a definite psychic-somatic individual, belong to a real world, and, perhaps, correspond to the direct experiences.)

In order to correlate mathematical concepts with phenomenal time in this sense Weyl grants the possibility of introducing a rigidly punctate "now" and of identifying and exhibiting the resulting temporal points. On the collection of these temporal points is defined the relation of *earlier than* as well as a congruence relation of *equality of temporal intervals,* the basic constituents of a simple mathematical theory of time. Now Weyl observes that the discrepancy between phenomenal time and the concept of real number would vanish if the following pair of conditions could be shown to be satisfied:

1. *The immediate expression of the intuitive finding that during a certain period I saw the pencil lying there were construed in such a way that the phrase "during a certain period" was replaced by "in every temporal point which falls within a certain time span OE.* [Weyl goes on to say parenthetically here that he admits "that this no longer reproduces what is intuitively present, but one will have to let it pass, *if it is really legitimate to dissolve a period into temporal points.*"]

2. If P is a temporal point, then the domain of rational numbers to which l belongs if and only if there is a time point L earlier than P such that OL = l.OE *can be constructed arithmetically in pure number theory on the basis of our principles of definition, and is therefore a real number in our sense.*

Condition 2 means that, if we take the time span *OE* as a unit, then each temporal point *P* is correlated with a definite real number. In an addendum Weyl also stipulates the converse.

But can temporal intuition itself provide evidence for the truth or falsity of these two conditions? Weyl thinks not. In fact, he states unequivocally that

> *... everything we are demanding here is obvious* nonsense: *to* these *questions, the intuition of time provides no answer—just as a man makes no reply to questions which clearly are addressed to him by mistake and, therefore, are unintelligible when addressed to him.*

The grounds for this assertion are by no means immediately evident, but one gathers from the passages following it that Weyl regards the experienced *continuous flow* of phenomenal time as constituting an insuperable barrier to the whole enterprise of representing this continuum in terms of individual points, and even to the characterization of "individual temporal point" itself. As he says,

> *The view of a flow consisting of points and, therefore, also dissolving into points turns out to be mistaken: precisely what eludes us is the nature of the continuity, the flowing from point to point; in other words, the secret of how the continually enduring present can continually slip away into the receding past. Each one of us, at every moment, directly experiences the true character of this temporal continuity. But, because of the genuine primitiveness of phenomenal time, we cannot put our experiences into words. So we shall content ourselves with the following description. What I am conscious of is for me both a being-now and, in its essence, something which, with its temporal position, slips away. In this way there arises the persisting factual extent, something ever new which endures and changes in consciousness.*

Weyl sums up what he thinks can be affirmed about "objectively presented time"—by which I take it is meant "phenomenal time described in an objective manner"—in the following two assertions, which he claims apply equally, *mutatis mutandis,* to every intuitively given continuum, in particular, to the continuum of spatial extension:

> 1. *An individual point in it is non-independent, i.e., is pure nothingness when taken by itself, and exists only as a "point of transition" (which, of course, can in no way be understood mathematically);*

> 2. *it is due to the essence of time (and not to contingent imperfections in our medium) that a fixed temporal point cannot be exhibited in any way, that always only an approximate, never an exact determination is possible*.

The fact that single points in a true continuum "cannot be exhibited" arises, Weyl continues, from the fact that they are not genuine individuals and so cannot be characterized by their properties. In the physical world they are never defined absolutely, but only in terms of a coordinate system, which, in an arresting metaphor, Weyl describes as "the unavoidable residue of the eradication of the ego." This metaphor, which Weyl was to employ more than once (Weyl 1950, p. 8 and 1949, p. 123) reflects the continuing influence of Husserlian phenomenological doctrine: in this case, the thesis that the existent is given in the first instance as the contents of a consciousness.

By 1919 Weyl had come to embrace Brouwer's views on the intuitive continuum. The latter's influence looms large in Weyl's next paper on the subject, *On the New Foundational Crisis of Mathematics,* written in 1920. Here Weyl identifies two distinct views of the continuum: "atomistic" or "discrete"; and "continuous". In the first of these the continuum is composed of individual real numbers which are well-defined and can be sharply distinguished. Weyl describes his earlier attempt at reconstructing analysis in Das Kontinuum as atomistic in this sense:

> *Existential questions concerning real numbers only become meaningful if we analyze the concept of real number in this extensionally determining and delimiting manner. Through this conceptual restriction, an ensemble of individual points is, so to speak, picked out from the fluid paste of the continuum. The continuum is broken up into isolated elements, and the flowing-into-each other of its parts is replaced by certain conceptual relations between these elements, based on the "larger-smaller" relationship. This is why I speak of the* atomistic *conception of the continuum* (Weyl 1998, p. 91).

Weyl now repudiated atomistic theories of the continuum, including that of *Das Kontinuum*. He writes:

> *In traditional analysis, the continuum appeared as the set of its points; it was considered merely as a special case of the basic logical relationship of* element and set. *Who would not have already noticed that, up to now, there was no place in mathematics for the equally fundamental relationship of* part and whole? *The fact, however,* that it has parts, *is a fundamental property of the continuum; and so (in harmony with intuition, so drastically offended against by today's "atomism") this relationship is taken as the mathematical basis for the continuum by Brouwer's theory. This is the real reason why the method used in delimiting subcontinua and in forming continuous functions starts out from* intervals *and not* points *as the primary elements of construction. Admittedly a set also has parts. Yet what distinguishes the parts of sets in the realm of the "divisible" is the existence of "elements" in the set-theoretical sense, that is, the existence of* parts that themselves do not contain any further parts. *And indeed, every part contains at least one "element". In contrast, it is inherent in the nature of the continuum that every part of it can be further divided without limitation. The concept of a point must be seen as an idea of a limit, "point" is the idea of a* limit *of a division extending* in infinitum. *To represent the continuous connection of the points, traditional analysis, given its shattering of the continuum into isolated points, had to have recourse to the concept of a* neighbourhood. *Yet, because the concept of continuous function remained mathematically sterile in the resulting generality, it became necessary to introduce the possibility of "triangulation" as a restrictive condition* (Weyl 1998, p. 115).

Like Brentano, Weyl knew that to "shatter a continuum into isolated points" would be to eradicate the very feature which characterizes a continuum—the fact that its cohesiveness is inherited by every one of its (connected) parts.

Weyl welcomed Brouwer's construction of the continuum by means of sequences generated by free acts of choice, thus identifying it as a "medium of free Becoming" which "does not dissolve into a set of real numbers as finished entities". Weyl felt that Brouwer's intuitionistic approach had brought him closer than anyone else to bridging that "unbridgeable chasm" between the intuitive and mathematical continua. In particular, he found compelling the fact that the Brouwerian continuum is not the union of two disjoint nonempty parts—that it is *indecomposable* or *cohesive.* "A genuine continuum," Weyl says, "cannot be divided into separate fragments." In 1921 Weyl observed:

> *...if we pick out a specific point, say, $x = 0$, on the number line C (i.e., on the variable range of a real variable x), then one cannot, under any circumstance, claim either coincides with it or is disjoint from it. The point $x = 0$ thus does not at all split the continuum C into two parts C^-: $x < 0$ and C^+: $x > 0$, in the sense that C would consist of the union of C^-, C^+ and the one point 0 ... If this appears offensive to present-day mathematicians with their atomistic thought habits, it was in earlier times a self-evident view held by everyone: Within a continuum, one can very well generate subcontinua by introducing boundaries; yet it is irrational to claim that the total continuum is made up of the boundaries and the subcontinua. The point is, a genuine continuum is something connected in itself, and it cannot be divided into separate fragments; this conflicts with its nature* (Weyl 1921, p. 111).

In later publications he expresses this more colourfully by quoting Anaxagoras to the effect that a continuum "defies the chopping off of its parts with a hatchet."

There being only minor differences between Weyl's and Brouwer's accounts of the continuum, Weyl abandoned his earlier attempt at the reconstruction of analysis and "joined Brouwer." At the same time, however, Weyl recognized that the resulting gain in intuitive clarity had been bought at a considerable price:

> *Mathematics with Brouwer gains its highest intuitive clarity. He succeeds in developing the beginnings of analysis in a natural manner, all the time preserving the contact with intuition much more closely than had been done before. It cannot be denied, however, that in advancing to higher and more general theories the inapplicability of the simple laws of classical logic eventually results in an almost unbearable awkwardness. And the mathematician watches with pain the greater part of his towering edifice which he believed to be built of concrete blocks dissolve into mist before his eyes* (Weyl 1949, p. 54).

Although he later practiced intuitionistic mathematics very rarely, Weyl remained an admirer of intuitionism. And the "riddle of the continuum" retained its fascination for him: ;as is attested to by the observation he made one of his last papers, *Axiomatic and Constructive Procedures in Mathematics,* written in 1954,

> *... the constructive transition to the continuum of real numbers is a serious affair... and I am bold enough to say that not even to this day are the logical issues involved in that constructive concept completely clarified and settled* (Weyl 1985, p. 17).

4. The Continuum in Constructive and Intuitionistic Mathematics

As we have observed, Brouwer maintained that mathematical activity is a form of construction in intuition. If the philosophically loaded term "in intuition" is removed, we are left simply with the term "construction", a much more concrete term on whose meaning mathematicians have achieved some measure of agreement. This leads to the idea of *constructive mathematics.*

The central principle of constructive mathematics is that a problem is to be regarded as solved only if an explicit solution can, in principle at least, be produced. Thus, for example, "There is an x such that $P(x)$" *means* that, in principle at least, we can explicitly produce an x such that $P(x)$. If the solution to the problem involves parameters, we must be able to present the solution explicitly by means of some *algorithm* or *rule* when give values of the parameters. That is, "for every x there is a y such that $P(x, y)$ *means* that, we possess an explicit method of determining, for any given x, a y for which $P(x, y)$.

The practice of constructive mathematics requires, as Brouwer recognized, the abandoning of certain laws of classical logic. In constructive mathematical reasoning *an existential statement can be considered affirmed only when an instance is produced*, and *a disjunction can be considered affirmed only when an explicit one of the disjuncts is demonstrated.* Consequently, neither the classical law of excluded middle nor the law of strong reductio ad absurdum can be constructively admissible. Hermann Weyl said of nonconstructive existence proofs that "they inform the world that a treasure exists without disclosing its location."

Consider the existential statement *there exists an odd perfect number* (i.e., an odd number equal to the sum of its proper divisors) which we shall write as $\exists nP(n)$. Its contradictory is the statement $\forall n \neg P(n)$. Classically, the law of excluded middle then allows us to affirm the disjunction

$$\exists n P(n) \vee \forall n \neg P(n) \qquad (1)$$

Constructively, however, in order to affirm this disjunction we must *either* be in a position to affirm the first disjunct $\exists n P(n)$, i.e., to possess, or have the means of obtaining, an odd perfect number, *or* to affirm the second disjunct $\forall n \neg P(n)$, i.e. to possess a demonstration that no odd number is perfect. Since at the present time mathematicians have neither of these, the disjunction (1), and *a fortiori* the law of excluded middle is not constructively admissible.

The logical principles underlying constructive reasoning have been codified into what is known as *constructiive* or *intuitionistic* logic. Roughly speaking, constructive/intuitionistic logic is the body of arguments of classical logic which can be established without the use, explicitly or implicitly, of the law of excluded middle. Intuitionistic logic guides how one should *reason* in constructive mathematics. But how are mathematical objects to be given, or, as we shall say, *specified*, in constructive mathematics? To begin with, everybody knows what it means to specify an *integer*. For example, $7 \cdot 10^4$ is specified, while the number n defined to be 0 if an odd perfect number exists, and 1 if an odd perfect number does not exist, is not specified. The number of primes less than, say, $10^{1000000}$ is specified, in the sense intended here, since we could, *in principle at least,* calculate this number. Constructive mathematics as we shall understand it is not concerned with questions of feasibility, nor in particular with what can actually be computed in real time by actual computers.

In *constructive real analysis* (Bishop and Bridges 1985) a *rational number* may be defined as a pair of integers (a, b) without a common divisor (where $b > 0$ and a may be positive or negative, or a is 0 and b is 1). The usual arithmetic operations on the rationals, together with the operation of taking the absolute value, are then easily supplied with explicit definitions. Accordingly it is clear what it means to specify a rational number.

In constructive mathematics, a problem is counted as solved only if an explicit solution can, in principle at least, be produced. Thus, for example, "There is an x such that $P(x)$" *means* that, in principle at least, we can explicitly produce an x such that $P(x)$. If the solution to the problem involves parameters, we must be able to present the solution explicitly by means of some *algorithm* or *rule* when given values of the parameters. That is, "for every x there is a y such that $P(x, y)$ *means* that, we possess an explicit method of determining, for any given x, a y for which $P(x, y)$. This leads us to examine what it means for a mathematical object to be explicitly given.

To begin with, everybody knows what it means to give an *integer* explicitly. For example, $7 \cdot 10^4$ is given explicitly, while the number n defined to be 0 if an odd perfect number exists, and 1 if an odd perfect number does not exist, is not given explicitly. The number of primes less than, say, $10^{1000000}$ is given explicitly, in the sense intended here, since we could, *in principle at least,* calculate this number. Constructive mathematics as we shall understand it is not concerned with questions of feasibility, nor in particular with what can actually be computed in real time by actual computers. *Rational numbers* may be defined as pairs of integers (a, b) without a common divisor (where $b > 0$ and a may be positive or negative, or a is 0 and b is 1). The usual arithmetic operations on the rationals, together with the operation of taking the absolute value, are then easily supplied with explicit definitions. Accordingly it is clear what it means to give a rational number explicitly.

To specify exactly what is meant by giving a *real number* explicitly is not quite so simple. For a real number is by its nature an infinite object, but one normally regards only finite objects as capable of being given explicitly. This difficulty may be overcome by stipulating that, to be given a real number, we must be given a (finite) *rule* or *explicit* *procedure* for calculating it to any desired degree of accuracy. Intuitively speaking, to be given a real number r is to be given a method of computing, for each positive integer n, a rational number r_n such that

$$|r - r_n| < \frac{1}{n}.$$

These r_n will then obey the law

$$|r_m - r_n| \leq \frac{1}{m} + \frac{1}{n}.$$

So, given any numbers k, p, we have, setting $n = 2k$,

$$|r_{n+p} - r_n| \leq \frac{1}{n+p} + \frac{1}{n} \leq \frac{2}{n} = \frac{1}{k}.$$

One is thus led to *define* a (constructive) real number to be a sequence of rationals $(r_n) = r_1, r_2, \ldots$ such that, for any k, a number n can be found such that

$$|r_{n+p} - r_n| \leq \frac{1}{k}.$$[2]

Here we understand that to be given a *sequence* we must be in possession of a *rule* or explicit method for generating its members. Each rational number α may be regarded as a real number by identifying it with the real number $(\alpha, \alpha, \ldots)$. The set of all real numbers – the *constructive real line* or the *constructive continuum* - will be denoted, as usual, by $\mathbb{R}$.

Now of course, for any "given" real number there are a variety of ways of giving explicit approximating sequences for it. Thus it is necessary to define an equivalence relation,

[2] It will be observed that in defining a constructive real number in this way we are following Cantor's, rather than Dedekind's characterization.

"equality on the reals". The correct definition here is: $r =_{\mathbb{R}} s$ if and only if for any k, a number n can be found so that

$$|r_{n+p} - s_{n+p}| \leq \frac{1}{k} \text{ for all } p.$$

When we say that two real numbers are equal we shall mean that they are equivalent in this sense, and so write simply "=" for "$=_{\mathbb{R}}$". To assert the *in*equality $r \neq s$ of two real numbers r, s is to assert that the equality $r = s$ leads to a contradiction. Inequality in this sense is constructively weak. In constructive mathematics a stronger notion of inequality, that of *apartness*, is normally used instead. We say that r and s are *apart,* or *distinguishable*, written $r \# s$, if n and k can actually be found so that $|r_{n+p} - s_{n+p}| > \frac{1}{k}$ for all p. Clearly $r \# s$ implies $r \neq s$, but the converse cannot be affirmed constructively. (In fact the converse is equivalent to *Markov's Principle*, which asserts that, if, for each n, $x_n = 0$ or 1, and if it is contradictory that $x_n = 0$ for all n, then there exists n for which $x_n = 1$. This thesis is accepted by some, but not all schools of constructivism.)

Is it constructively the case that for any real numbers x and y, we have $x = y \lor x \neq y$? The answer is no. For if this assertion were constructively true, then, in particular, we would have a method of deciding whether, for any given rational number r, whether $r = \pi^{\sqrt{2}}$ or not. But at present no such method is known—it is not known, in fact, whether $\pi^{\sqrt{2}}$ is rational or irrational. We can, of course, calculate $\pi^{\sqrt{2}}$ to as many decimal places as we please, and if in actuality it is unequal to a given rational number r, we shall discover this fact after a sufficient amount of calculation. If, however, $\pi^{\sqrt{2}}$ is *equal* to r, even several centuries of computation cannot make this fact certain; we can be sure only that is very close to r. We have no method which will tell us, in finite time, whether $\pi^{\sqrt{2}}$ exactly coincides with r or not. This situation may be summarized by saying that equality on the reals is *not decidable.* (By contrast, equality on the integers or rational numbers is decidable.) Observe that this does *not* mean $\neg(x = y \lor x \neq y)$. We have not actually derived a *contradiction* from the assumption $x = y \vee x \neq y$; we have only given an example showing its implausibility. It is natural to ask whether it can actually be *refuted*. For this it would be necessary to make some assumption concerning the real numbers which contradicts classical mathematics. Certain schools of constructive mathematics are willing to make such assumptions; but the majority of constructivists confine themselves to methods which are also classically correct.

Despite the fact that constructive equality of real numbers is not a decidable relation, it is *balanced* in the sense of satisfying the *law of double negation* $\neg(r \neq s) \Rightarrow r = s$. In fact, we can prove the stronger assertion that $\neg(r \# s) \Rightarrow r = s$. For, given k, we may choose n so that $|r_{n+p} - r_n| \leq \frac{1}{4k}$ and $|s_{n+p} - s_n| \leq \frac{1}{4k}$ for all p. If $|r_n - s_n| \geq \frac{1}{k}$, then we would have $|r_{n+p} - s_{n+p}| \geq \frac{1}{2k}$ for all p, which entails $r \neq s$. If $\neg(r \# s)$, it follows that $|r_n - s_n| < \frac{1}{k}$ and $|r_{n+p} - s_{n+p}| \leq \frac{2}{k}$ for every p. Since for every k we can find n so that this inequality holds for every p, it follows that $r = s$. (From the facts that $\neg(r \# s) \Rightarrow r = s$ it follows easily that $r \neq s \Leftrightarrow \neg\neg(r \# s)$).

One should not, however, conclude from the stability of equality that the law of double negation $\neg\neg A \rightarrow A$ is generally affirmable. That it is not can be seen from the following example. Write the decimal expansion of π and below the decimal expansion $\rho = 0.333\ldots$, terminating it as soon as a sequence of digits 0123456789 has appeared in π. Then if the 9 of the first sequence 0123456789 in π is the k^{th} digit after the decimal point, $\rho = \frac{10^k - 1}{3 \cdot 10^k}$.

Now suppose that ρ were not rational; then $\rho = \frac{10^k - 1}{3 \cdot 10^k}$ would be impossible and no sequence 0123456789 could appear in π, so that $\rho = \frac{1}{3}$, which is also impossible. Thus the assumption that ρ is not rational leads to a contradiction; yet we are not warranted to assert that r is rational, for this would mean that we could calculate integers m and n for which $\rho = \frac{m}{n}$. But this evidently requires that we can produce a sequence 0123456789 in π or demonstrate that no such sequence can appear, and at present we can do neither.

Order on the constructive real line. The *order relation* on the reals is given constructively by stipulating that $r < s$ is to mean that we have an explicit lower bound on the distance between r and s. That is,

$$r < s \Leftrightarrow \textit{ n and k can be found so that } s_{n+p} - r_{n+p} > 1/k \textit{ for all } p.$$

It can readily be shown that, for any real numbers x, y such that $x < y$, there is a rational number α such that $x < \alpha < y$.

We observe that $r \# s \Leftrightarrow r < s \vee s < r$. The implication from right to left is clear. Conversely, suppose that $r \# s$. Find n and k so that $|r_{n+p} - s_{n+p}| > \frac{1}{k}$ for every p, and determine $m > n$ so that $|r_m - r_{m+p}| < \frac{1}{4k}$ and $|s_m - s_{m+p}| < \frac{1}{4k}$ for every p. Either $r_m - s_m > \frac{1}{k}$ or $s_m - r_m > \frac{1}{k}$; in the first case $r_{m+p} - s_{m+p} > \frac{1}{2k}$ for every p, whence $s < r$; similarly, in the second case, we obtain $r < s$.

We define $r \leq s$ to mean that $s < r$ is false. Notice that $r \leq s$ is not the same as $r < s$ or $r = s$: in the case of the real number ρ defined above, for instance, clearly $\rho \leq \frac{1}{3}$; but we do not know whether $\rho < \frac{1}{3}$ or $\rho = \frac{1}{3}$. Still, it is true that $r \leq s \wedge s \leq r \Rightarrow r = s$. For the premise is the negation of $r < s \vee s < r$, which, by the above, is equivalent to $\neg r \# s$. But we have already seen that this last implies $r = s$.

There are several common properties of the order relation on real numbers which hold classically but which cannot be established constructively. Consider, for example, the trichotomy law $x < y \vee x = y \vee y < x$. Suppose we had a method enabling us to decide which of the three alternatives holds. Applying it to the case $y = 0$, $x = \pi^{\sqrt{2}} - r$ for rational r would yield an algorithm for determining whether $\pi^{\sqrt{2}} = r$ or not, which we have already observed is an open problem. One can also demonstrate the failure of the trichotomy law (as well as other classical laws) by the use of "fugitive sequences". Here one picks an unsolved problem of the form $\forall n P(n)$, where P is a decidable property of integers—for example, Goldbach's conjecture that every even number ≥ 4 is the sum of two odd primes. Now one defines a sequence—a "fugitive" sequence—of integers (n_k) by $n_k = 0$ if $2k$ is the sum of two primes and $n_k = 1$ otherwise. Let r be the real number defined by $r_k = 0$ if $n_k = 0$ for all $j \leq k$, and $r_k = 1/m$ otherwise, where m is the least positive integer such that $n_m = 1$. It is then easy to check that $r \geq 0$ and $r = 0$ if and only if Goldbach's conjecture holds. Accordingly the correctness of the trichotomy law would imply that we could resolve Goldbach's conjecture. Of course, Goldbach's conjecture might be resolved in the future, in which case we would merely choose another unsolved problem of a similar form to define our fugitive sequence.

A similar argument shows that the law $r \leq s \vee s \leq r$ also fails constructively: define the real number s by $s_k = 0$ if $n_k = 0$ for all $j \leq k$; $s_k = 1/m$ if m is the least positive integer such that $n_m = 1$, and m is even; $s_k = -1/m$ if m is the least positive integer such that $n_m = 1$, and m is odd. Then $s \leq 0$ (resp. $0 \leq s$) would mean that there is no number of the form $2 \cdot 2k$

(resp. $2 \cdot (2k + 1)$) which is not the sum of two primes. Since neither claim is at present known to be correct, we cannot assert the disjunction $s \leq 0 \vee 0 \leq s$.

In constructive analysis there is a convenient substitute for trichotomy known as the *comparison principle.* This is the assertion

$$r < t \Rightarrow r < s \vee s < t.$$

Its validity can be established in a manner similar to the foregoing.

Albebraic operations on the constructive reals. The fundamental operations $+$, $-$, $\cdot$, $^{-1}$ and $|\cdot|$ are defined for real numbers as one would expect, viz.

- $r + s$ is the sequence $(r_n + s_n)$
- $r - s$ is the sequence $(r_n - s_n)$
- $r \cdot s$ or rs is the sequence $(r_n s_n)$
- if $r \# 0$, r^{-1} is the sequence (t_n), where $t_n = r_n^{-1}$ if $t_n \neq 0$ and $t_n = 0$ if $r_n = 0$
- $|r|$ is the sequence $(|r_n|)$

It is then easily shown that $rs \# 0 \Leftrightarrow r \# 0 \wedge s \# 0$. For if $r \# 0 \wedge s \# 0$, we can find k and n such that $|r_{n+p}| > \frac{1}{k}$ and $|s_{n+p}| > \frac{1}{k}$ for every p, so that $|r_{n+p}s_{n+p}| > \frac{1}{k^2}$ for every p, and rs $\# 0$. Conversely, if $rs \# 0$, then we can find k and n so that

$$|r_{n+p}s_{n+p}| > \frac{1}{k}, \quad |r_{n+p} - r_n| < 1, \; |s_{n+p} - s_n| < 1$$

for every p. It follows that

$$|r_{n+p}| > \frac{1}{k}(|s_n|+1) \text{ and } |s_{n+p}| > \frac{1}{k}(|r_n|+1)$$

for every p, whence $r \# 0 \wedge s \# 0$.

But it is not constructively true that, if $rs = 0$, then $r = 0$ or $s = 0$! To see this, use the following prescription to define two real numbers r and s. If in the first n decimals of π no sequence 0123456789 occurs, put $r_n = s_n = 2^{-n}$; if a sequence of this kind does occur in the first n decimals, suppose the 9 in the first such sequence is the k^{th} digit. If k is odd, put $r_n = 2^{-k}$, $s_n = 2^{-n}$; if k is even, put $r_n = 2^{-n}$, $s_n = 2^{-k}$. Then we are unable to decide whether $r = 0$ or $s = 0$. But $rs = 0$. For in the first case above $r_n s_n = 2^{-2n}$; in the second $r_n s_n = 2^{-k-n}$. In either case $|r_n s_n| < \frac{1}{m}$ for $n > m$, so that $rs = 0$.

Convergence of sequences and completeness of the constructive reals. As usual, a sequence (a_n) of real numbers is said to *converge* to a real number b, or to have *limit s* if, given any natural number k, a natural number n can be found so that for every natural number p,

$$|b - a_{n+p}| < 2^{-k}.$$

As in classical analysis, a constructive necessary and condition that a sequence (a_n) of real numbers be convergent is that it be a *Cauchy* sequence, that is, if, given any given any natural number k, a natural number n can be found so that for every natural number p,

$$|a_{n+p} - a_n| < 2^{-k}.$$

However, some classical theorems concerning convergent sequences are no longer valid constructively. For example, a bounded momotone sequence need no longer be convergent. A simple counterexample is provided by the sequence (a_n) defined as follows: $a_n = 1 - 2^{-n}$ if among the first n digits in the decimal expansion of π no sequence 0123456789 occurs, while $a_n = 2 - 2^{-n}$ if among these n digits such a sequence does occur. Since it is not known whether the limit of this sequence, if it exists, is 1 or 2, we cannot claim that that this limit exists as a well defined real number.

In classical analysis ℝ is *complete* in the sense that every nonempty set of real numbers that is bounded above has a supremum. As it stands, this assertion is constructively incorrect. For consider the set A of members $\{x_1, x_2, \ldots\}$ of any fugitive sequence of 0s and 1s. Clearly A is bounded above, and its supremum would be either 0 or 1. If we knew which, we would also know whether $x_n = 0$ for all n, and the sequence would no longer be fugitive.

Nevertheless, the completeness of ℝ can be salvaged by defining suprema and infima somewhat more delicately than is customary in classical mathematics. A nonempty set A of real numbers is *bounded above* if there exists a real number b, called an *upper bound* for A, such that $x \leq b$ for all $x \in A$. A real number b is called a *supremum*, or *least upper bound*, of A if it is an upper bound for A and if for each $\varepsilon > 0$ there exists $x \in A$ with $x > b - \varepsilon$. We say that A is *bounded below* if there exists a real number b, called a *lower bound* for A, such that $b \leq x$ for all $x \in A$. A real number b is called an *infimum,* or *greatest lower bound*, of A if it is a lower bound for A and if for each $\varepsilon > 0$ there exists $x \in A$ with $x < b + \varepsilon$. The supremum (respectively, infimum) of A, is unique if it exists and is written sup A (respectively, inf A).

Let us prove the *constructive least upper bound principle.*

Theorem. Let A be a nonempty set of real numbers that is bounded above. Then sup A exists if and only if for all $x, y \in \mathbb{R}$ with $x < y$, either y is an upper bound for A or there exists $a \in A$ with $x < a$.

Proof. If sup A exists and $x < y$, then either sup $A < y$ or $x <$ sup A; in the latter case we can find $a \in A$ with sup A – (sup $A - x$) $< a$, and hence $x < a$. Thus the stated condition is necessary.

Conversely, suppose the stated condition holds. Let a_1 be an element of A and choose an upper bound b_1 for A with $b_1 > a_1$. We construct recursively a sequence (a_n) in A and (b_n) of upper bounds for A such that, for each $n \geq 0$,

(i) $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$

and

(ii) $b_{n+1} - a_{n+1} \leq \frac{3}{4}(b_n - a_n)$.

Having found $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, if $a_n + \frac{3}{4}(b_n - a_n)$ is an upper bound for A, put $b_{n+1} = a_n + \frac{3}{4}(b_n - a_n)$ and $a_{n+1} = a_n$; while if there exists $a \in A$ with $a > a_n + \frac{3}{4}(b_n - a_n)$, we set $a_{n+1} = a$ and $b_{n+1} = b_n$. This completes the recursive construction.

From (i) and (ii) we have

$$0 \leq b_n - a_n \leq (\tfrac{3}{4})^{n-1}(b_1 - a_1).$$

It follows that the sequences (a_n) and (b_n) converge to a common limit l with $a_n \leq l \leq b_n$ for $n \geq 1$. Since each b_n is an upper bound for A, so is l. On the other hand, given $\varepsilon > 0$, we can choose n so that $l \geq a_n > l - \varepsilon$, where $a_n \in A$. Hence l = sup A.

An analogous result for infima can be stated and proved in a similar way.

Functions on the constructive reals. Considered constructively, a *function* from ℝ to ℝ is a rule F which enables us, when given a real number x, to compute another real number $F(x)$ in such a way that, if $x = y$, then $F(x) = F(y)$. It is easy to check that every polynomial is a function in this sense, and that various power series and integrals, for example those defining tan x and e^x, also determine functions. Viewed constructively, some classically defined "functions" on ℝ can no longer be considered to be defined on the whole of ℝ. Consider, for example, the "blip" function B defined by $B(x) = 0$ if $x \neq 0$ and $B(0) = 0$. Here the domain of the function is $\{x \in \mathbb{R}: x = 0 \vee x \neq 0\}$. But we have seen that we cannot assert dom(B) = ℝ. Accordingly, the blip function is not well defined as a function from ℝ to ℝ. Of course, classically, B is the simplest *discontinuous* function defined on ℝ. The fact that the simplest possible discontinuous function fails to be defined on the whole of ℝ gives grounds for the suspicion that *no* function defined on ℝ can be discontinuous; in other words, that, constructively speaking, *all functions defined on* ℝ *are continuous* (as remarked in the previous secrion, this was a central tenet of intuitionism's founder, Brouwer). The claim is plausible. For if a function F is well-defined on all reals x, it must be possible to compute the value for all rules x determining real numbers, that is, determining their sequences of rational approximations $x_1, x_2, \ldots$. Now $F(x)$ must be computed to accuracy ε in a finite number of steps—the number of steps depending on ε. This means that only finitely many approximations can be used, i.e., $F(x)$ can be computed to within ε only when x is known within δ for some δ. Thus F should indeed be continuous. In fact all known examples of constructive functions defined on ℝ are continuous. Constructively, a real-valued function f is *continuous* if for each $\varepsilon > 0$ we can find a specific $\delta > 0$ such that $|f(x) - f(y)| \leq \varepsilon$ whenever $|x - y| < \delta$. If all functions on ℝ are continuous, then a subset A of ℝ may fail to be genuinely *complemented*: that is, there may be no subset B of ℝ disjoint from A such that $\mathbb{R} = A \cup B$.To see this, we first suppose that A, B are disjoint subsets of $\mathbb{R}$ and that there is a point $a \in A$ which can be approached arbitrarily closely by points of B (or vice-versa). Then, assuming all functions on $\mathbb{R}$ are continuous, it cannot be the case that $\mathbb{R} = A \cup B$. For if so, we may define the function f on $\mathbb{R}$ by $f(x) = 0$ if $x \in A$, $f(x) = 1$ if $x \in B$. Then for all $\delta > 0$ there is $b \in B$ for which $|b - a| < \delta$, but $|f(b) - f(a)| = 1$. Thus f fails to be continuous at a, and we conclude that $\mathbb{R} \neq A \cup B$. Now if we take A to be any finite set of real numbers, any union of open or closed intervals, or the set $\mathbb{Q}$ of rational numbers, then in each case the set B of points "outside" A satisfies the above condition. Accordingly, for each such subset A, $\mathbb{R}$ is not "decomposable" into A and the set of points "outside" A, in the sense that these two sets of points together do not exhaust $\mathbb{R}$. This fact indicates that the constructive continuum is a great deal more "cohesive" than its classical counterpart. For classically, the continuum is merely *connected* in the sense that it is not (nontrivially) decomposable into two open (or closed) subsets. Constructively, however, $\mathbb{R}$ is indecomposable into subsets which are neither open nor closed. In this sense the constructive real line can be brought close to the ideal of a true continuum.

Certain well-known theorems of classical analysis concerning continuous functions fail in constructive analysis. One such is the *theorem of the maximum*: a uniformly continuous function on a closed interval assumes its maximum at some point. For consider, as Figure 1 below, a function $f : [0,1] \to \mathbb{R}$ with two relative maxima, one at $x = \frac{1}{3}$ and the other at $x = \frac{2}{3}$ and of approximately the same value. Now arrange things so that $f\left(\frac{1}{3}\right) = 1$ and $f\left(\frac{2}{3}\right) = 1 + t$, where t is some small parameter. If we could tell where f assumes its absolute maximum, clearly we could also determine whether $t \leq 0$ or $t \geq 0$, which, as we have seen, is not, in general, possible. Nevertheless, it can be shown that from f we can in fact ccalculate the maximum value itself, so that at least one can assert the existence of that maximum, even if one can't tell exactly where it is assumed.

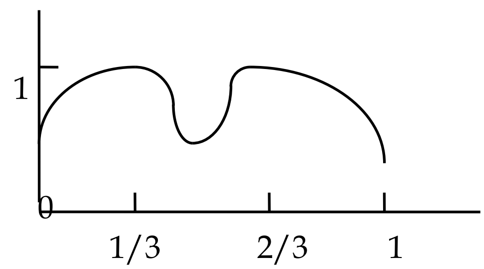


Figure 1

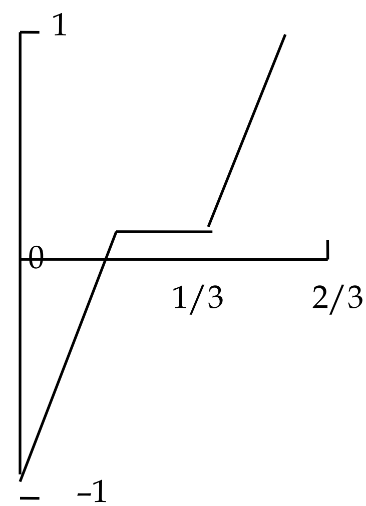


Figure 2

Another classical result that fails to hold constructively in its usual form is the well-known *intermediate value theorem.* This is the assertion that, for any continuous function f from the unit interval [0, 1] to $\mathbb{R}$, such that $f(0) = -1$ and $f(1) = 1$, there exists a real number $a \in [0,1]$ for which $f(a) = 0$. To see that this fails constructively, consider the function f depicted in Figure 2 above: here f is piecewise linear, taking the value t (a small parameter) between $x = \frac{1}{3}$ and $x = \frac{2}{3}$. If the intermediate value theorem held, we could determine a for which $f(a) = 0$. Then either $a < \frac{2}{3}$ or $a > \frac{1}{3}$; in the former case $t \geq 0$; in the latter $t \leq 0$. Thus we would be able to decide whether $t \geq 0$ or $t \leq 0$; but we have seen that this is not constructively possible in general.

Nevertheless, it can be shown that, constructively, the intermediate value theorem is "almost" true in the sense that

$$\forall f\, \forall \varepsilon > 0\; \exists a\; (|f(a)| < \varepsilon)$$

and also in the sense that, if we write $P(f)$ for

$$\forall b\; \forall a{<}b\; \exists c\; (a < c < b \wedge f(c) \neq\neq 0),$$

then

$$\forall f\, [P(f) \rightarrow \exists x\; (f(x) = 0)].$$

This example illustrates how a single classical theorem "refracts" into several constructive theorems.

Axiomatizing the constructive reals. The constructive reals can be furnished with an axiomatic description (Bridges 1999). We begin by assuming the existence of a set $\boldsymbol{R}$ with

- a binary relation > (*greater than*)
- a corresponding *apartness relation* # defined by $x \# y \;\Leftrightarrow\; x > y$ or $y > x$
- a unary operation $x \mapsto -x$
- binary operations $(x, y) \mapsto x + y$ (*addition*) and $(x, y) \mapsto xy$ (*multiplication*)
- distinguished elements 0 (*zero*) and 1 (*one*) with $0 \neq 1$
- a unary operation $x \mapsto x^{-1}$ on the set of elements $\neq 0$.

The elements of $\boldsymbol{R}$ are called *real numbers.* A real number x is *positive* if $x > 0$ and *negative* if $-x > 0$. The relation $\geq$ (*greater than or equal to*) is defined by

$$x \geq y \Leftrightarrow \forall z(y > z \Rightarrow x > z).$$

The relations < and ≤ are defined in the usual way; x is *nonnegative* if $0 \leq x$.

The sets $\boldsymbol{N}$ of natural numbers, $\boldsymbol{N}^+$ of positive integers, $\boldsymbol{Z}$ of integers and $\boldsymbol{Q}$ of rational numbers are identified with the usual subsets of $\boldsymbol{R}$; for instance $\boldsymbol{N}^+$ is identified with the set of elements of R of the form $1 + 1 + \cdots + 1$.

These relations and operations are subject to the following three groups of axioms, which, taken together, form the system **CA** of axioms for *constructive analysis,* or the *constructive real numbers.*

Field Axioms

$$x + y = y + x \quad (x + y) + z = x + y + z) \quad 0 + x = x \quad x + (-x) = 0 \quad xy = yx$$

$$(xy)z = x(yz) \quad 1x = x \quad x \# 0 \rightarrow xx^{-1} = 1 \quad x(y + z) = xy + xz$$

Order Axioms

$$(x \geq y \wedge y \geq x) \leftrightarrow x = y.$$

$$\neg (x > y \wedge y > x) \quad x > y \rightarrow \forall z(x > z \vee z > y) \quad \neg(x \# y) \rightarrow x = y$$

$$x > y \rightarrow \forall z(x + z > y + z) \quad (x > 0 \wedge y > 0) \rightarrow xy > 0.$$

Notice that from the fourth of the order axioms it follows that $\neg(x \neq y) \rightarrow x = y$, that is, the equality relation is *balanced.* (Here and in the sequel $x \neq y$ is an abbreviation for $\neg(x= y)$).

The final two axioms introduce special properties of > and ≥. In the second of these the notions *bounded above, bounded below, and bounded* are defined as in classical mathematics, and the *least upper bound,* if it exists, of a nonempty set S of real numbers is the unique real number b such that

- b is an upper bound for S, and
- for each $c < b$ there exists $s \in S$ with $s > c$.

Here a"nonempty" has the stronger constructive meaning of being inhabited, to wit, that an element of the set in question can actually be constructed.

Special Properties of >.

Archimedean axiom. For each $x \in \mathbf{R}$ such that $x \geq 0$ there exists $n \in \mathbf{N}$ such that $x < n$.

The least upper bound principle. Let S be a nonempty subset of R that is bounded above , such that for all real numbers a, b with $a < b$, either b is an upper bound for S or else there exists $s \in S$ with $s > a$. Then S has a least upper bound.

The following basic properties of > and ≥ can then be established.

$$\neg(x > x) \quad x \geq x \quad (x > y \wedge y > z) \rightarrow x > z \quad \neg(x > y \wedge y \geq x) \quad (x > y \geq z) \rightarrow x > z$$
$$\neg(x > y) \leftrightarrow y \geq x \quad \neg(x \geq y) \leftrightarrow \neg\neg(y > x) \quad (x \geq y \geq z) \rightarrow x \geq z$$

$\neg(x > y \wedge x = y)$ $x \geq 0 \rightarrow (x = 0 \leftrightarrow \forall\varepsilon{>}0\ (x < \varepsilon))$ $x + y > 0 \rightarrow (x > 0 \vee y > 0)$ $x > 0 \rightarrow -x < 0$

$(x > y \wedge z < 0) \rightarrow yz > xz$ $x \# 0 \leftrightarrow x^2 > 0$ $1 > 0$ $0 < x < 1 \rightarrow x > x^2$

$n \in N^+ \rightarrow n^{-1} > 0$

if $x > 0$ *and* $y \geq 0$, *then there exists* $n \in \mathbf{Z}$ *such that* $nx > y$

$x > 0 \rightarrow x^{-1} > 0$ $xy > 0 \rightarrow (x \neq 0 \vee y \neq 0)$

if $a < b$, *then there exists* $r \in \mathbf{Q}$ *such that* $a < r < b$

The constructive real line ℝ as introduced above is a model of **CA.** Are there any other models, that is, models not isomorphic to ℝ? If classical logic is assumed, **CA** is a categorical theory and so the answer is no. But this is not the case within intuitionistic logic, for it can be shown that, in intuitionistic set theory, both the Dedekind and Cantor reals are models of **CA**, while these may fail to be isomorphic.

The intuitionistic continuum. In constructive analysis, a real number is an infinite (convergent) sequence of rational numbers generated by an effective rule, so that the constructive real line is essentially just a restriction of its classical counterpart. Brouwerian intuitionism takes a more liberal view of the matter, resulting in a considerable enrichment of the arithmetical continuum over the version offered by strict constructivism. As conceived by intutionism, the arithmetical continuum admits as real numbers "not only infinite sequences determined in advance by an effective rule for computing their terms, but also ones in whose generation free selection plays a part." (Dummett 1977, p. 62).The latter are called *(free) choice sequences.* Without loss of generality we may and shall assume that the entries in choice sequences are natural numbers.

Hermann Weyl describes Brouwer's conception of choice sequences in the following terms:

In Brouwer's analysis, the individual place in the continuum*, the real number, is to be defined not by a set but by a* sequence *of natural numbers, namely, by a law which correlates with every natural number n a natural number φ(n)... How then do assertions arise which concern... all real numbers, i.e., all values of a real variable? Brouwer shows that frequently statements of this form in traditional analysis, when correctly interpreted, simply concern the totality of natural numbers. In cases where they do not, the notion of sequence changes its meaning: it no longer signifies a sequence determined by some law or other, but rather one that is created* step by step by free acts of choice, *and thus necessarily remains* in statu nascendi. *This "becoming"* selective sequence (werdende Wahlfolge) *represents the continuum, or the variable, while the sequence determined* ad infinitum *by a law represents the individual real number in the continuum. The continuum no longer appears, to use Leibniz's language, as an aggregate of fixed elements but as a medium of free "becoming". Of a selective sequence* in statu nascendi*, naturally only those properties can be meaningfully asserted which already admit of a yes-or-no decision (as to whether or not the property applies to the sequence) when the sequence has been carried to a certain point; while the continuation of the sequence beyond this point, no matter how it turns out, is incapable of overthrowing that decision. (*Weyl 1949, p. 52).

While constructive analysis does not formally contradict classical analysis and may in fact be regarded as a subtheory of the latter, a number of intuitionistically plausible principles have been proposed for the theory of choice sequences which render intuitionistic analysis divergent from its classical counterpart. (Intuitionistic analysis, nevertheless, an extension of **CA.)**

One such principle is *Brouwer's Continuity Principle*. This asserts that, given a relation $R(\alpha, n)$ between choice sequences α and numbers *n,* if for each α a number *n* may be determined for which $R(\alpha, n)$ holds, then n can already be determined on the basis of the knowledge of a finite number of terms of α. This may be seen to be plausible if one considers that the according to Brouwer the construction of a choice sequence is incompletable; at any given moment we can know nothing about it outside the identities of a finite number of its entries. Brouwer's Continuity Principle amounts to the assertion that every function from $\mathbb{N}^{\mathbb{N}}$ to $\mathbb{N}$ is continuous.

From this one can prove a weak version of the Continuity Theorem, namely, that every function from $\mathbb{R}$ to $\mathbb{R}$ is continuous (Bridges and Richman 1987, p. 109). Another such principle is *Bar Induction*, a certain form of induction for well-founded sets of finite sequences (Dummett 1977, Bridges and Richman 1987). Brouwer used Bar Induction and the Continuity Principle in proving his Continuity Theorem that every real-valued function defined on a closed interval is uniformly continuous, from which it follows that the intuitionistic continuum, and all of its closed interval, are cohesive.

Brouwer gave the intuitionistic conception of mathematics an explicitly subjective twist by introducing the *creative subject*. The creative subject was conceived as a kind of idealized mathematician for whom time is divided into discrete sequential stages, during each of which he may test various propositions, attempt to construct proofs, and so on. In particular, it can always be determined whether or not at stage n the creative subject has a proof of a particular mathematical proposition p. While the theory of the creative subject remains controversial, its purely mathematical consequences can be obtained by a simple postulate which is entirely free of subjective and temporal elements. The creative subject allows us to define, for a given proposition p, a binary sequence $\langle a_n \rangle$ by $a_n = 1$ if the creative subject has a proof of p at stage n;, $a_n = 0$ otherwise.

Now if the construction of these sequences is the only use made of the creative subject, then references to the latter may be avoided by postulating the principle known as *Kripke's Scheme:*

> *For each proposition p there exists an increasing binary sequence* $<a_n>$ *such that p holds if and only if* $a_n = 1$ *for some n.*

Taken together, these principles have been shown to have remarkable consequences for the cohesiveness of subsets of the continuum. Not only is the intuitionistic continumm cohesive, but, assuming Brouwer's Continuity Principle and Kripke's Scheme, it remains cohesive even if one pricks it with a pin "The [intuitionistic] continuum has, as it were, a syrupy nature, one cannot simply take away one point." If in addition Bar Induction is assumed, then, even more surprisingly, cohesiveness is maintained even when all the rational points are removed from the continuum. Thus it is appropriate to describe the classical continuum as the "frozen intuitionistic continuum".

5. Nonstandard Analysis and the Hyrerreal Line

Once the continuum had been provided with a set-theoretic foundation, the use of the infinitesimal in mathematical analysis was largely abandoned. And so the situation remained for a number of years. The first signs of a revival of the infinitesimal approach to analysis surfaced in the 1950s with a paper by A. H. Laugwitz and C. Schmieden (Laugwitz and Schmieden 1958). But the major breakthrough came in 1960 when it occurred to the mathematical logician **Abraham Robinson** (1918–1974) that "the concepts and methods of contemporary Mathematical Logic are capable of providing a suitable framework for the development of the Differential and Integral Calculus by means of infinitely small and infinitely large numbers." (Robinson 1996, p. xiii). This insight led to the creation of *Nonstandard Analysis* (**NSA**), which Robinson regarded as realizing Leibniz's conception of infinitesimals and infinities as ideal numbers possessing the same properties as ordinary real numbers. In the introduction to his book on the subject he writes:

> *It is shown in this book that Leibniz's ideas can be fully vindicated and that they lead to a novel and fruitful approach to classical Analysis and to many other branches of mathematics. The key to our method is provided by the detailed analysis of the relation between mathematical languages and mathematical structures which lies at the bottom of contemporary model theory.*

After Robinson's initial insight, a number of ways of presenting Nonstandard Analysis were developed. Here is a sketch of one of them (Bell and Machover 1977, Keisler 1994).

Starting with the classical real line $\mathbb{R}$, a set-theoretic universe—the *standard universe*—is first constructed over it: here by such a universe is meant a set U containing $\mathbb{R}$ which is closed under the usual set-theoretic operations of union, power set, Cartesian products and subsets. Now write **U** for the structure (U, $\in$), where $\in$ is the usual membership relation on U: associated with this is the extension $\mathcal{L}(U)$ of the first-order language of set theory to include a name ***u*** for each element u of U. Now, using the well-known compactness theorem for first-order logic, U is extended to a new structure ***U** = ($^*U$, $^*\in$), called a *nonstandard universe*, satisfying the following key principle:

> ***Saturation Principle.*** *Let* **Φ** *be a collection of* $\mathcal{L}(U)$*-formulas with exactly one free variable. If* **Φ** *is* finitely satisfiable *in* **U**, *that is, if for any finite subset* **Φ′** *of* **Φ** *there is an element of U which satisfies all the formulas of* **Φ′** in **U**, *then there is an element of *U which satisfies all the formulas of* **Φ** *in* ***U.**

The saturation property expresses the intuitive idea that the nonstandard universe is very rich in comparison to the standard one. Indeed, while there may exist, for each finite subcollection F of a given collection P of properties, an element of U satisfying the members of F in U, there may not necessarily be an element of U satisfying *all* the members of P. The saturation of ***U** guarantees the existence of an element of $^*U$ which satisfies, in ***U,** all the members of P. For example, suppose the set ℕ of natural numbers is a member of U; for each $n \in \mathbb{N}$ let $P_n(x)$ be the property $x \in \mathbb{N}$ & $n < x$. Then clearly, while each finite subcollection of the collection P = $\{P_n: n \in \mathbb{N}\}$ is satisfiable in U, the whole collection is not. An element of $^*U$ satisfying all the members of P in ***U** will then be an "natural number" greater than every member of ℕ, that is, an *infinite* number.

From the saturation property it follows that ***U** satisfies the important

> ***Transfer Principle.*** *If* σ *is any sentence of* $\mathcal{L}(U)$, *then* σ *holds in* **U** *if and only if it holds in* ***U.**

The transfer principle may be seen as a version of Leibniz's continuity principle: it asserts that all first-order properties are preserved in the passage to or "transfer" from the standard to the nonstandard universe.

The members of U are called *standard sets,* or *standard objects*; those in $^*U - U$ *nonstandard sets* or *nonstandard objects*: $^*U$ thus consists of both standard and nonstandard objects. The members of $^*U$ will also be referred to as **-sets* or **-objects* Since $U \subseteq {}^*U$, under this convention every set (object) is also a *-set (object). The **-members* of a *-set A are the *-objects x for which $x \; {}^*\!\in A$

If A is a standard set, we may consider the collection $A^\star$ —the *inflate* of A – consisting of all the *-members of A: this is not necessarily a standard set nor even a *-set. The inflate of a standard set may be regarded as the same set viewed from a nonstandard vantage point. While clearly $A \subseteq A^\star$, $A^\star$ may contain "nonstandard" elements not in A. It can in fact be shown that *infinite* standard sets always get "inflated" in this way. Using the transfer principle, any function f between standard sets automatically extends to a function—also written f – between their inflates.

Each mathematical structure $\mathbf{A} = (A, R, ...)$ has an inflate $\mathbf{A}^\star = (A^\star, R^\star)$. From the transfer principle it follows that $\mathbf{A}$ and $\mathbf{A}^\star$ have precisely the same first-order properties.

Now suppose that the set $\mathbb{N}$ of natural numbers is a member of U. Then so is the set $\mathbb{R}$ of real numbers, since each real number may be identified with a set of natural numbers. $\mathbb{R}$ may be regarded as an ordered field, and the same is therefore true of its inflate $\mathbb{R}^\star$.This, the *hyperreal line,* has precisely the same first-order properties as $\mathbb{R}$. The members of $\mathbb{R}^\star$ are called *hyperreals.* A standard hyperreal is then just a real, to which we shall refer for emphasis as a *standard real.* Since $\mathbb{R}$ is infinite, nonstandard hyperreals must exist. The saturation principle implies that there must be an *infinite* (nonstandard) hyperreal, that is, a hyperreal *a* such that $a > n$ for every $n \in \mathbb{N}$. In that case its reciprocal $\frac{1}{a}$ is infinitesimal in the sense of exceeding 0 and yet being smaller than $\frac{1}{n+1}$ for every $n \in \mathbb{N}$. In general. we call a hyperreal *a infinitesimal* if its absolute value $|a|$ is < $\frac{1}{n+1}$ for every $n \in \mathbb{N}$. In that case the set *I* of infinitesimals contains not just 0 but a substantial number (in fact, infinitely many) of other elements. Clearly *I* is an additive subgroup of ℝ, that is, if $a, b \in I$, then $a - b \in I$.

Note that ℝ★ is thus a nonarchimedean ordered field. One might question whether this is compatible with the facts that ℝ★ and ℝ share the same first-order properties, but the latter is archimedean. These data are consistent because the archimedean property is not first-order. However, while ℝ★is nonarchimedean, it is **-archimedean* in the sense that, for any $a \in$ ℝ★there is $n \in$ ℝ★ for which $a < n$.

The **members** of the inflate ℕ★ of ℕ are called *hypernatural numbers.* As for the hyperreals, it can be shown that ℕ★ also contains nonstandard elements which must exceed every member of ℕ: these are called *infinite* hypernatural numbers.

For hyperreals *a, b* we define $a \approx b$ and say that *a* and *b* are *infinitesimally close* if $a - b \in I$. This is an equivalence relation on the hyperreal line: for each hyperrreal *a* we write $\mu(a)$ for the equivalence class of *a* under this relation and call it the *monad* of *a.* The monad of a hyperreal *a* thus consists of all the hyperreals that are infinitesimally close to *a*: it may thought of as a small cloud centred at *a.* Note also that $\mu(0) = I$.

A hyperreal *a* is *finite* if it is not infinite; this means that $|a| < n$ for some $n \in \mathbb{N}$. It is not difficult to show that finiteness is equivalent to the condition of *near-standardness*: here a hyperreal *a* is *near-standard* if $a \approx r$ for some standard real *r.*

Much of the usefulness of Nonstandard Analysis stems from the fact that statements of classical analysis involving limits or the (ε, δ) criterion admit succinct, intuitive translations into statements involving infinitesimals or infinite numbers, in turn enabling comparatively straightforward proofs to be given of classical theorems. Here are some examples of such translations:

- *Let* $<s_n>$ be *a standard infinite sequence of real numbers and let s be a standard real number. Then s is the limit of* $<s_n>$ *within* $\mathbb{R}$, $\lim_{n\to\infty} s_n = s$ *in the classical sense, if and only if* $s_n \approx s$ *for all* infinite *subscripts n.*

- *A standard sequence* $<s_n>$ *converges if and only if* $s_n \approx s_m$ *for all infinite n and m.* (Cauchy's criterion for convergence.)

Now suppose that *f* is a real-valued function defined on some open interval (*a, b*). We have remarked above that *f* automatically extends to a function—also written *f*— on (a, b) .

- *In order that the standard real number c be the limit of f(x) as x approaches* x_0, $\lim_{x\to x_0} f(x) = c$ *, with* x_0 *a standard real number in (a, b), it is necessary and sufficient that* $f(x) \approx f(x_0)$ *for all* $x \approx x_0$.

- *The function f is continuous at a standard real number* x_0 *in (a, b) if and only if* $f(x) \approx f(x_0)$ *for all* $x \approx x_0$. (This is *equivalent to saying that f maps the monad of* x_0 *into the monad of* $f(x_0)$.

- *In order that the standard number c be the derivative of f ay* x_0 *it is necessary and sufficient that*

$$\frac{f(x) - f(x_0)}{x - x_0} \approx c$$

for all $x \neq x_0$ *in the monad of* x_0*.*

Many other branches of mathematics admit elegant and fruitful nonstandard formulations.

Finally, it should be pointed out that while the usual models of Nonstandard Analysis are obtained using highly nonconstructive tools, other methods have been developed for producing such models which are constructively acceptable. (Moerdijk 1995 and Palmgren 1998, 2001).

6.The Continuum in Smooth Infinitesimal Analysis

Smooth Infinitesimal Analysis. In differential geometry it is customary to confine attention to functions between spaces which are *smooth,* that is, arbitrarily many times differentiable. *Smooth Infinitesimal Analysis* (**SIA**) provides a framework for mathematical analysis which embodies the practice of differential geometry in a remarkably direct and powerful way, namely by stipulating that *all* functions – however defined - between spaces be smooth. In particular al functions defined on the real line **R** to itself are required to be smooth. In **SIA** this requirement is satisfied through the following considerations. Starting with a differentiable function f: **R** → **R**, draw the graph of f, so generating a curve in the Cartesian plane. Now pick any point on the curve and consider a short length L of the curve around the point. If L is taken to be sufficiently short, it will appear to b *approximately* a straight line. In **SIA** it is required that, if L is taken to be of *infinitesimal* length, then it will be a straight line *exactly.* This idea in its turn is realized by postulating the existence in **R** of a set Δ of infinitesimals with the property that all functions on Δ are linear.

Thus in **SIA** we are given:

- an object **R** called the *real* line (or *smooth line*), on which the usual operations of addition and multiplication are defined and satisfy the usual algebraic laws governing real numbers (in particular 0 and 1 are present and are neutral elements under + and $\times$ respectively;
- a subset **Δ** of the set **R** of real numbers called the *domain of infinitesimals.* **Δ** is assumed to satisfy the following conditions:

  (1) $0 \in$**Δ**

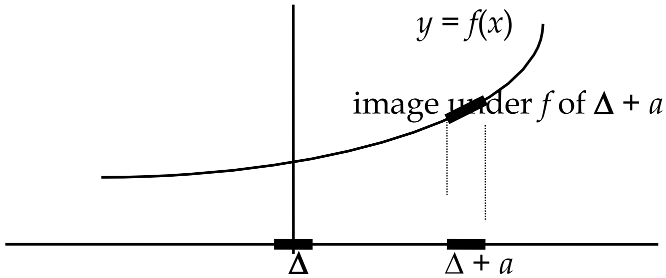
(2) *Any multiple of an infinitesimal is an infinitesimal,* that is, for any $a \in \mathbf{R}$, $x \in \Delta \Rightarrow ax \in \Delta$. It follows immediately from this that Δ is *symmetric around* 0, that is, $\forall x\ (x \in \Delta \Leftrightarrow -x \in \Delta)$.

(3) The ***Infinitesimal Linearity Principle*** or the ***Kock- Lawvere axiom.*** This asserts that all maps $\Delta \to \mathbf{R}$ are *linear* (strictly speaking, *affine*, but no matter) in the following strong sense:

for any $f : \Delta \to \mathbf{R}$ there is a *unique* $a \in \mathbf{R}$ such that, for all $\varepsilon \in \Delta$, $f(\varepsilon) = f(0) + \varepsilon a$. (Here and in the sequel we shall use symbols ε, η to denote arbitrary elements ofΔ.) This real number a is called the *slope* of f, and is written **slp**(f). Thus, for any $f : \Delta \to \mathbf{R}$ and all $\varepsilon \in \Delta$, $f(\varepsilon) = f(0) + \varepsilon$ **slp**(f).

Given $f : \Delta \to \mathbf{R}$, let $f^* : \Delta \to \mathbf{R} \times \mathbf{R}$ defined by $f^*(\varepsilon) = (\varepsilon, f(\varepsilon))$. The map f^* may be thought of as the graph of the map f n the plane. The Infinitesimal Linearity Principle can be seen as asserting that this graph is a straight line with slope **slp**(f) passing through $(0, f(0))$. Thus the effect of any map on Δ is to translate and rotate it; in effect Δ behaves like a short "rigid rod", just long enough to have a slope, but too short to bend. It is, as it were, a geometric object possessing location and direction but lacking extension.

The Infinitesimal Linearity Principle can also be presented in the following way. Think of a real function f: $\mathbf{R} \to \mathbf{R}$ as defining a curve $\mathscr{C}$ in the plane. Then the Infinitesimal Linearity Principle amounts to the assertion that, for any such f and for any for any point $a \in \mathbf{R}$, the image under f of the "infinitesimal interval" $\Delta + a$ obtained by translating Δ to a is a straight line and coincides with the tangent to $\mathscr{C}$ at $x = a$.



Thus each real function has the effect of "bringing" **Δ** into coincidence" with the tangent vector to the curve associated with the function at any point on it. In this sense, then, **Δ** plays the role of a *generic tangent vector.* Also, since the image of **Δ** under a function is necessarily a straight line and a part of the associated curve, it follows that each point on a curve is contained in a nondegenerate infinitesimal straight segment of the curve. Accordingly in **SIA** *curves are infinitesimally straight.* This is called the **Infinitesimal Straightness Principle.**

The Infinitesimal Linearity Principle has two immediate formal consequences:

- **The nondegeneracy of Δ :** $\Delta \neq \{0\}$. For suppose $\Delta = \{0$, and let a, b be any unequal real numbers. Then the two maps $\varepsilon \mapsto \varepsilon a$ and $\varepsilon \mapsto \varepsilon b$ would both be identically 0, contradicting the Linearity Principle.
- **The Cancellation Principle:** for any $a, b \in \mathbf{R}$, if $\varepsilon a = \varepsilon b$ for all $\varepsilon \in \Delta$, then $a = b$. This follows immediately from the uniqueness condition in the Linearity Principle.

We can now show that Δ consists of *nilpotent,* in fact *nilsquare* quantities, that is,

$$\forall x\ (x \in \Delta \Rightarrow x^2 = 0).$$

To prove this, write $s: \Delta \to \mathbf{R}$ for the map $\varepsilon \mapsto \varepsilon^2$. Then we have, for $\varepsilon \in \Delta$,

$$\varepsilon^2 = s(\varepsilon) = s(0) + \varepsilon\, \mathbf{slp}(s) = \varepsilon\, \mathbf{slp}(s)$$

and, since $-\varepsilon \in \Delta$,

$$\varepsilon^2 = (-\varepsilon)^2 = s(-\varepsilon) = s(0) - \varepsilon\, \mathbf{slp}(s) = -\varepsilon\, \mathbf{slp}(s).$$

Hence $\varepsilon^2 = -\varepsilon^2$, so that $\varepsilon^2 = 0$.

Notice that we do not claim that Δ comprises *all* nilsquare quantities. This, although usually assumed in presentations of **SIA**, is not needed.

Here is a further fact of interest. Without assuming the symmetry of Δ around 0, the following conditions are equivalent:

(i) $\forall x\, (x \in \Delta \Rightarrow x^2 = 0)$.

(ii) $\mathbf{slp}(s) = 0$.

(i) $\Rightarrow$ (ii). Assuming **(i),** we have, for all $\varepsilon \in \Delta$, $0 = \varepsilon^2 = s(\varepsilon) = s(0) + \varepsilon\, \mathbf{slp}(s) = \varepsilon.\, \mathbf{slp}(s)$. The Cancellation Principle yields $\mathbf{slp}(s) = 0$.

(ii) $\Rightarrow$ (i) Assuming **(ii),** we have, for all $\varepsilon \in \Delta$, $\varepsilon^2 = s(\varepsilon) = s(0) + \varepsilon\, \mathbf{slp}(s) = 0 + 0 = 0$. Hence **(i).**

The Differential Calculus in SIA. Let us call a *real function* any real-valued function defined on a closed interval in to **R.** The derivative of an arbitrary real function can be formulated as follows. Given a closed interval **I** in **R** and a function $f : \mathbf{I} \to \mathbf{R}$, for each $x \in \mathbf{I}$ define the function $f_x \colon \Delta \to \mathbf{R}$ by $f_x(\varepsilon) = f(x + \varepsilon)$. (Intervals are defined by providing **R**

with a suitable order relation. We assume that closed intervals are stable under the addition of infinitesimals: see below.)

The *derivative* $f' : \mathbf{I} \to \mathbf{R}$ of f is defined by $f'(x) = \mathbf{slp}(f_x)$. It follows easily that

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

This is the *Fundamental Equation of the Differential Calculus* in **SIA**. The quantity $f'(x)$ is the slope at x of the curve determined by f and the infinitesimal

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

is the infinitesimal change or *increment* in the value of f on passing from x to $x + \varepsilon$.

The fundamental equation (*) can be thought of as a rigorous formulation of the engineers' "practical" version of the Taylor series of a function. For (*) is what we get if in the Taylor series

$$f(x + \varepsilon) = f(x) + \varepsilon f'(x) + 1/2!\, \varepsilon^2 f''(x) + \ldots$$

we follow the practice of engineers to treat as "vanishingly small" the square and higher powers of the "small" quantity ε .

Derivatives of elementary functions are easily calculated in **SIA** using the Cancellation Principle. For example, here is the calculation of the derivative of the function x^n:

$$\varepsilon(x^n)' = (x + \varepsilon)^n - x^n = \varepsilon n\, x^{n-1} + \text{terms in } \varepsilon^{2\text{-}} \text{ and higher powers} = \varepsilon n\, x^{n-1}$$

Hence, by the Cancellation Principle,

$$(x^n)' = n\, x^{n-1}.$$

And here is the calculation of the derivative of the function $1/x$ (for $x > 0$):

$$\begin{aligned}\varepsilon(1/x)' = 1/x+\varepsilon - 1/x = -\varepsilon/x(x+\varepsilon) &= -\varepsilon(x-\varepsilon)/x(x+\varepsilon)(x-\varepsilon)\\ &= -\varepsilon x+\varepsilon^2/x(x^2-\varepsilon^2)\\ &= -\varepsilon x/x^3\\ &= -\varepsilon/x^2.\end{aligned}$$

Cancelling ε on both sides of the equation gives

$$(1/x)' = -1/x^2.$$

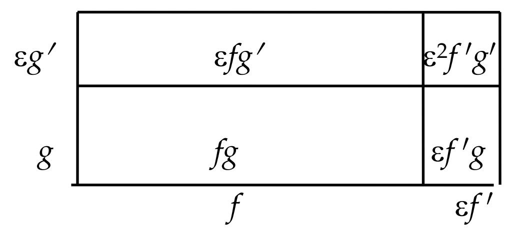
Let us derive in **SIA** a basic law of the differential calculus, the *product rule:*

$$(fg)' = f'g + fg'.$$

To do this we compute

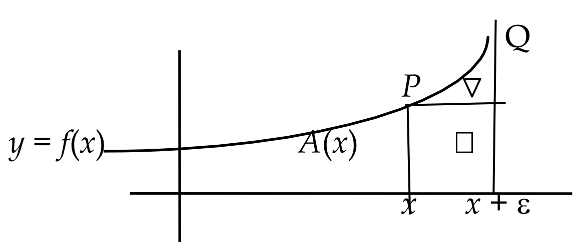
$$\begin{aligned}(fg)(x+\varepsilon) &= (fg)(x) + (fg)'(x) = f(x)g(x) + (fg)'(x),\\ (fg)(x+\varepsilon) &= f(x+\varepsilon)g(x+\varepsilon) = [f(x) + f'(x)].[g(x) + g'(x)]\\ &= f(x)g(x) + \varepsilon(f'g + fg') + \varepsilon^2 f'g'\\ &= f(x)g(x) + \varepsilon(f'g + fg'),\end{aligned}$$

since $\varepsilon^2 = 0$. Therefore $\varepsilon(fg)' = \varepsilon(f'g + fg')$, and the result follows by the Cancellation Principle. This calculation is depicted in the diagram below.



Infinitesimals in **SIA** have two fundamental aspects, *algebraic* and *geometric.* Algebraically, they are real numbers whose squares vanish; geometrically, they are straight segments of curves. Both of these aspects are used in applications.

The interplay of these aspects is well illustrated by the derivation in **SIA** of the *Fundamental Theorem of the Calculus.* To this end, let **I** be a closed interval $\{x: a \leq x \leq b\}$ in **R** — or **R** itself—and let f: **I** → **R**; also let $A(x)$ be the area under the curve $y = f(x)$ as indicated in the figure:



Then

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

Now, by the Infinitesimal Straightness Principle, the arc PQ is a straight line; accordingly $\triangledown$ is a triangle of area $\frac{1}{2}\,\varepsilon.\varepsilon f'(x) = 0$. It follows that $\varepsilon A'(x) = \varepsilon f(x)$, and the Cancellation Principle gives

$$A'(x) = f(x).$$

Thus, if we regard $A(x)$ as the integral of $f(x)$, the above equation asserts that differentiation is the inverse of integration – the Fundamental Theorem of the Calculus.

Smoothness of Functions in SIA. From the Fundamental Equation the *Principle of Continuity* can be deduced, namely, that in **SIA** *all real functions are continuous*, in the sense of *sending neighbouring points to neighbouring points.* (Here two points x, y on **R** are said to be *neighbours* if $x - y$ is in Δ, that is, if x and y differ by an infinitesimal.) To see this, given a real function f and neighbouring points x, y, note that $y = x + \varepsilon$ with ε in Δ, so that

$$f(y) - f(x) = f(x + \varepsilon) - f(x) = \varepsilon f'(x).$$

Since $\varepsilon f'(x)$ is infinitesimal, the result follows.

From the fact that in **SIA** every real function has a derivative, which is itself differentiable, it follows that any real function is arbitrarily many times differentiable. This fact justifies the use of the term "smooth".

In **SIA** there is a sense in which *everything is generated by the domain of infinitesimals.* For consider the set Δ^Δ of all maps $\Delta \to \Delta$. It follows from the Linearity Principle that **R** can be identified as the subset of Δ^Δ consisting of all maps vanishing at 0. In this sense **R** is "generated" by Δ.

Explicitly, Δ^Δ is a monoid under composition which may be regarded as acting on Δ by composition: for $f \in \Delta^\Delta$, $f \cdot \varepsilon = f(\varepsilon)$. The subset V consisting of all maps vanishing at 0 is a submonoid naturally identified as the set of *ratios of infinitesimals.* The identification of **R** and V made possible by the principle of infinitesimal linearity thus leads to the characterization of **R** itself as the set of ratios of infinitesimals. This was essentially the view of Euler, who regarded infinitesimals as formal zeros and real numbers as representing the possible values of 0/0. For this reason Lawvere has suggested that **R** in **SIA** should be called the space of *Euler reals.*

Once one has **R**, Euclidean spaces of all dimensions may be obtained as powers of **R**, and arbitrary Riemannian manifolds may be obtained by patching together subspaces of these.

The "Internal Logic" of SIA is Intuitionistic. We observe that the postulates of **SIA** are *incompatible with the Law of Excluded Middle of classical logic* (LEM)– the assertion that, for any proposition *p,* either *p* holds or not *p* holds. This incompatibility can be demonstrated in two ways, one informal and the other rigorous. First the informal argument. Consider the function f defined for real numbers x by $f(x) = 1$ if $x = 0$ and $f(x) = 0$ whenever $x \neq 0$. If LEM held, each real number would then be either equal or unequal to 0, so that the function f would be defined on the whole of **R.** But, considered as a function with domain **R,** f is clearly discontinuous. Since, as we know, in **SIA** every function on **R** is continuous, f cannot have domain **R** there. (The domain of f is in fact (**R** – {0}) ∪ {0}, which, because of the failure of the law of excluded middle in **SIA,** is provably unequal to **R.**)

So LEM fails in **SIA**. To put it succinctly, *universal continuity implies the failure of the Law of Excluded Middle.*

Here now is the rigorous argument. We derive the failure of LEM from the Cancellation Principle. To begin with, if $x \neq 0$, then $x^2 \neq 0$, so that, if $x^2 = 0$, then necessarily not $x \neq 0$. This means that

$$\textit{for all infinitesimal } \varepsilon, \textit{ not } \varepsilon \neq 0. \qquad (*)$$

Now suppose that LEM were to hold. Then we would have, for any ε, either ε = 0 or ε ≠ 0. But (*) allows us to eliminate the second alternative, and we infer that, for all ε, ε = 0. This may be written

$$\textit{for all } \varepsilon, \;\; \varepsilon.1 = \varepsilon.0,$$

from which we derive by Cancellation Principle the falsehood 1 = 0. So again LEM must fail.

The "internal" logic of **SIA** is accordingly not full classical logic. It is, instead, *intuitionistic* logic, that is, the logic derived from the constructive interpretation of mathematical assertions. In practice when working in **SIA** one does not notice this "change of logic" because, like much of elementary mathematics, the topics discussed there are naturally treated by constructive means such as direct computation.

It is worth noting that the refutability of the Law of Excluded Middle in **SIA** leads to the refutability of an important principle of set theory, the *Axiom of Choice.* This is the assertion

(AC) for any family **A** of inhabited sets, there is a *choice function on* **A**, that is, a function f: $\mathbf{A} \to \bigcup\mathbf{A}$ for which $f(X) \in X$ whenever $X \in \mathbf{A}$. (A set t is said to be *inhabited* if it can be constructively shown to have a member. In intuitionistic logic this is a stronger condition than the assertion that the set be nonempty.)

Now the Law of Excluded Middle can be derived merely from the very special case of the Axiom of Choice which asserts merely that any *doubleton* $\{U, V\}$ has a choice function. For let p be any proposition, define

$$U = \{x \in 2 : x = 0 \vee p\} \quad V = \{x \in 2 : x = 1 \vee p\}$$

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

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

It follows that

$$[a = 0 \vee b = 1] \vee p\ ,$$

whence

(*) $$a \neq b \vee p.$$

Now clearly

$$p \rightarrow U = V = 2 \rightarrow a = b,$$

whence

$$a \neq b \rightarrow \neg p,$$

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

Thus LEM is derivable from this special case of the Axiom of Choice. Since LEM is refutable in **SIA**, so equally, then, is **AC**.

The refutability of the Axiom of Choice in **SIA**, and hence its incompatibility with the Principle of Continuity which prevails there, is not surprising in view of the Axiom's well-known "paradoxical" consequences. One of these is the famous *Banach-Tarski paradox* (Wagon 1985) which asserts that any solid sphere can be decomposed into finitely many pieces which can themselves be reassembled to form two solid spheres each of the same size as the original, or into one solid sphere of any preassigned size. Paradoxical decompositions such as these become possible only when continuous geometric objects are, recalling Dedekind's words, "dissolved to atoms ... [through a] frightful, dizzying discontinuity" into discrete sets of points which the axiom of choice then allows to be rearranged in an arbitrary (discontinuous) manner. Such procedures violate the Principle of Continuity.

The Link between the Infinitesimal and the Real in SIA; the Constancy Principle. Suppose that we are investigating the behavior of some variable quantity represented by a function F. The approach taken in **SIA**, as (implicitly) in the differential calculus, is to begin the investigation by confining it initially to the infinitesimal world. Life in the infinitesimal world is beautifully simple: curves are just straight lines, and the squares of incremental changes vanish. This makes the determination of infinitesimal increments equally simple, enabling the increment $\varepsilon F'(x)$ in $F(x)$ to be presented in the form $\varepsilon k(x)$, where $k(x)$ is some explicit function whose form has been obtained by "infinitesimal" analysis. Thus we obtain an "infinitesimal" equation of the form $\varepsilon F'(x) = \varepsilon k(x)$. Applying the Cancellation Principle in turn yields the "differential" equation

$$(*) \qquad F'(x) = k(x)$$

which holds in the world "in the large".

The Cancellation Principle thus provides a formal, astonishingly simple link between the infinitesimal world and the real world, the world "in the large". The idea of a linkage between these two worlds was the animating principle behind applications of the calculus throughout the 17th and 18th centuries.

In practice, of course, the equation (*), while of fundamental importance, is only the first step in determining the explicit form of the function F. For this, it is necessary to "integrate" k , that is, to provide k with an *antiderivative,* an explicit function G such that $G' = k$. It will then follow that $F' = G'$, from which we will be able to conclude that $F = G$. (Strictly speaking, F and G may differ by a constant function but we shall ignore this here.)

To carry out this procedure in **SIA** we need to introduce an additional postulate. Let **I** be a closed interval. We define a *stationary point* of a function f: $\mathbf{I} \to \mathbf{R}$ to be a point $a \in \mathbf{I}$ in whose vicinity "infinitesimal changes in the value of the argument fail to change the value of f", that is, for which $f(a + \varepsilon) = f(a)$ for all ε. This means that $f(a) + \varepsilon f'(a) = f(a)$, so that $\varepsilon f'(a) = 0$ for all ε, from which it follows by the Cancellation Principle that $f'(a) = 0$. Thus a stationary point of a function is precisely a point at which the derivative of the function vanishes.

In classical analysis, if the derivative of a function is identically zero, the function is constant. This fact is the source of the following postulate concerning stationary points adopted in **SIA**:

Constancy Principle. If every point in a closed interval **I** is a stationary point of f: $\mathbf{I} \to \mathbf{R}$ (that is, if f' is identically 0), then f is constant.

It follows from the Constancy Principle that two functions with identical derivatives differ by at most a constant.

Put succinctly, the Constancy Principle asserts that "universal infinitesimal (or "local") constancy implies global constancy", or "infinitesimal behaviour determines global behaviour" The Constancy Principle brings into sharp focus the difference in **SIA** between points and infinitesimals. For if in the Constancy Principle one replaces "infinitesimal constancy" by "constancy at a point" the resulting "Principle" is false because *any function whatsoever* is constant at every point. But since in **SIA** all functions on **R** are smooth, the Constancy Principle embodies the idea that for such functions local constancy is sufficient for global constancy, that a nonconstant smooth function must be somewhere nonconstant over arbitrarily small intervals.

The Constancy Principle provides another bridge between the infinitesimal world and the world "in the large". Hermann Weyl could not see such a direct linkage between the two worlds, and inferred that this absence doomed the idea of infinitesimal, leading to its inevitable replacement by the limit concept. In his *Philosophy of Mathematics and Natural Science* he says:

> [In its struggle with the infinitely small] *the limiting process was victorious. For the* limit *is an indispensable concept, whose importance is not affected by the acceptance or rejection of the infinitely small. But once the limit concept has been grasped, it is seen to render the infinitely small superfluous. Infinitesimal analysis proposes to draw conclusions by integration from the behavior in the infinitely small, which is governed by elementary laws, to the behavior in the large; for instance, from the universal law of attraction for two material "volume elements" to the magnitude of attraction between two arbitrarily shaped bodies with homogeneous or non-homogeneous mass distribution. If the infinitely small is not interpreted 'potentially' here, in the sense of the limiting process, then the one has nothing to do with the other, the process in infinitesimal and finite dimensions become independent of each other, the tie which binds them together is cut.*

In **SIA** the Constancy Principle reconnects the infinitesimal and the extended. Behaviour "in the large" is completely determined by behaviour "in the infinitely small".

Cohesiveness of the Continuum in SIA. In classical analysis the continuum and its closed intervals are connected in the sense that they cannot be split into two nonempty subsets neither of which contains a limit point of the other. In **SIA** the Constancy Principle ensures that these have the vastly stronger property of *cohesiveness* (or indecomposability) which, we recall, is the property of not being splittable into two disjoint nonempty parts *in any way whatsoever.* This is clearly equivalent to the condition that any map defined on **R** or one of its closed intervals to {0, 1} takes a constant value.

To show that this condition holds, let A be **R** or any closed interval, and suppose $f\colon A \to \{0, 1\}$. We claim that f is constant. For we have, for any $x \in A$,

$$(f(x) = 0 \text{ or } f(x) = 1) \quad \& \quad (f(x + \varepsilon) = 0 \text{ or } f(x + \varepsilon) = 1).$$

This gives four possibilities:

(i) $f(x) = 0 \;\;\&\;\; f(x + \varepsilon) = 0$

(ii) $f(x) = 0 \;\;\&\;\; f(x + \varepsilon) = 1$

(iii) $f(x) = 1 \;\;\&\;\; f(x + \varepsilon) = 0$

(iv) $f(x) = 1 \;\;\&\;\; f(x + \varepsilon) = 1$

Possibilities (ii) and (iii) may be ruled out because f is continuous. This leaves (i) and (iv), in either of which $f(x) = f(x + \varepsilon)$. So f is locally, and hence globally, constant, that is, constantly 1 or 0.

From the cohesiveness of closed intervals it can be inferred (Bell 2001) that in **SIA** *all intervals in* **R** *are cohesive.*

Smooth infinitesimal analysis as an axiomatic theory; consequences for the continuum. **SIA** can be *axiomatized* as a theory formulated within (higher-order) intuitionistic logic. Here are the basic axioms of the theory (Moerdijk and Reyes 1991).

Axioms for the continuum, or smooth real line R. These include the usual axioms for a commutative ring with unit expressed in terms of two operations + and · , (we usually write xy for $x \cdot y$) and two distinguished elements $0 \neq 1$. In addition we stipulate that **R** is an *intuitionistic field,* i.e., satisfies the following axiom:

$$x \neq 0 \text{ implies } \exists y(xy = 1).$$

Axioms for the strict order relation < on R. These are:

O1. $a < b$ and $b < c$ implies $a < c$.

O2. $\neg(a < a)$

O3. $a < b$ implies $a + c < b + c$ for any c.

O4. $a < b$ and $0 < c$ implies $ac < bc$

O5. either $0 < a$ or $a < 1$.

O6. $a \neq b$ implies $a < b$ or $b < a$.

O7. $0 < x$ implies $\exists y\, (x = y^2)$.

Arithmetical Axioms. These govern the set **N** of *Archimedean* (or smooth) *natural numbers*, and read as follows:

1. **N** is a cofinal or Archimedean subset of **R,** i.e. $\mathbf{N} \subseteq \mathbf{R}$ and $\forall x \in \mathbf{R}\, \exists n \in \mathbf{N}\, x < n$.
2. *Peano axioms*:

$$0 \in \mathbf{N}$$

$$\forall x \in \mathbf{R}(x \in \mathbf{N} \rightarrow x + 1 \in \mathbf{N})$$

$$\forall x \in \mathbf{R}(x \in \mathbf{N} \rightarrow x + 1 \neq 0)$$

3. *Restricted Induction scheme*. For every formula $\varphi(x)$ involving just $=$, $\wedge$, $\vee$, $\top$ ("true"), $\bot$ ("false"), $\exists$

$$\varphi(0) \wedge \forall x \in \mathbf{N}[\varphi(x) \rightarrow \varphi(x + 1)] \rightarrow \forall x \in \mathbf{N}\varphi(x).$$

Using restricted induction it follows that

- **N** has *decidable equality*, i.e. $\forall x \in \mathbf{N} \forall y \in \mathbf{N}\ (x = y \vee x \neq y)$
- **N** is *linearly ordered*, i.e. $\forall x \in \mathbf{N} \forall y \in \mathbf{N}\ (x < y \vee x = y \vee y < x)$.
- **N** satisfies *decidable induction*: for any formula $\varphi(x)$,

$$\forall x \in \mathbf{N}(\varphi(x) \vee \neg\varphi(x)) \rightarrow [[\varphi(0) \wedge \forall x \in \mathbf{N}(\varphi(x) \rightarrow \varphi(x + 1)] \rightarrow \forall x \varphi(x)].$$

The relation $\leq$ on **R** is defined by $a \leq b \Leftrightarrow \neg(b < a)$. The open interval (a, b) and closed interval $[a, b]$ are defined as usual, viz. $(a, b) = \{x: a < x < b\}$ and $[a, b] = \{x: a \leq x \leq b\}$; similarly for half-open, half-closed, and unbounded intervals. It can be shown from the axioms introduced so far that closed intervals are stable under the addition of infinitesimals.

We have written Δ for the subset $\{x: x^2 = 0\}$ of **R** consisting of (nilsquare) *infinitesimals.* As before, we use the letter ε as a variable ranging over Δ.

The two final axioms are:

Infinitesimal Linearity Axiom. *For any map* $g: \Delta \rightarrow \mathbf{R}$ *there exist unique* $a, b \in \mathbf{R}$ *such that, for all* ε, *we have*

$$g(\varepsilon) = a + b\varepsilon.$$

Constancy Axiom. *If* $A \subseteq \mathbf{R}$ *is any closed interval on* **R,** *or* **R** *itself, and* $f: A \to \mathbf{R}$ *satisfies* $f(a + \varepsilon) = f(a)$ *for all* $a \in A$ *and* $\varepsilon \in \Delta$*, then* f *is constant.*

It follows easily from the Infinitesimal Linearity Axiom that Δ is *nondegenerate,* i.e. $\Delta \neq \{0\}$.[3] For if $\Delta = \{0\}$, then the identity map $i: \Delta \to \Delta$ can be represented as $i(\varepsilon) = b\varepsilon$ for any b, in violation of the uniqueness condition on b. It should be noted that, while Δ does not reduce to $\{0\}$, nevertheless 0 is the only *explicitly nameable* element of Δ. For it is easily seen to be inconsistent to assert that Δ actually contains an element $\neq 0$.

From the nondegeneracy of Δ we can also (again) refute the Law of Excluded Middle in **SIA,** more particularly, we can prove

(*) $$\neg\forall\varepsilon(\varepsilon = 0 \vee \varepsilon \neq 0).$$

For we have, for $\varepsilon \in \Delta$, $\varepsilon^2 = 0$, whence $\neg(\varepsilon \neq 0)$, and (*) would give $\varepsilon = 0$. So Δ would be degenerate, contrary to what we have already shown. It follows from (*) that, using x and y as variables ranging over **R,**

$$\neg\forall x\forall y(x = y \vee x \neq y).$$

In a word, the identity relation is *undecidable* on **R.**

Call a binary relation S on **R** *balanced* if it satisfies

[3] It should be noted that, while Δ does not reduce to $\{0\}$, nevertheless 0 is the only *explicitly nameable* element of Δ. For it is easily seen to be inconsistent to assert that Δ actually contains an element $\neq 0$.

$$\forall x \forall y\ (\neg\neg xRy \rightarrow xRy).$$

Then the nondegeneracy of **Δ** implies that, in **SIA** the equality relation is *unbalanced.* For suppose that = were balanced. Then, for any ε, it would be the case that $\neg\varepsilon \neq 0 \rightarrow \varepsilon = 0$. But we have already shown above that $\neg(\varepsilon \neq 0)$, so it would follow that ε =0. This being the case for any ε, **Δ** would be degenerate.

On the other hand, in **SIA** the order relation < is balanced. For suppose $\neg\neg a < b$. Then certainly $a \neq b$, since $a = b \rightarrow \neg a < b$ by irreflexivity. Therefore, by axiom O6, $a < b$ or $b < a$. The second disjunct together with $\neg\neg a < b$ and transitivity gives $\neg\neg a < a$, which contradicts $\neg a < a$. Accordingly we are left with $a < b$. Hence < is balanced.

Even allowing for the presence of intuitionistic logic, we note that the *algebraic structure* on **R** in **SIA** differs little from that of the usual classical system of real numbers. In **SIA**, **R** is equipped with the usual addition and multiplication operations under which it is a field. In particular, **R** satisfies the condition that each $x \neq 0$ has a multiplicative inverse. Notice, however, that since in **SIA** no infinitesimal (apart from 0 itself) is provably ≠ 0, they are not required to have multiplicative inverses (a requirement which would lead to inconsistency). From a strictly algebraic standpoint, **R** in **SIA** differs from its classical counterpart only in being required to satisfy the Cancellation Principle.

The situation is otherwise, however, as regards the order structure of **R** in **SIA**. Since infinitesimals do not have multiplicative inverses, and **R** is an intuitionistic field, it must be the case that $\forall\varepsilon\neg(\varepsilon \neq 0)$, whence

$$\forall\varepsilon\neg(\varepsilon < 0\ \wedge \neg\ \varepsilon > 0),$$

or equivalently

$$\forall\varepsilon(\varepsilon \leq 0 \wedge \varepsilon \geq 0).$$

It follows easily from this and the nondegeneracy of Δ that

$$\neg\forall x\forall y\ (x < y \vee y < x \vee x = y).$$

In other words the order relation < on **R** in **SIA** *fails to satisfy the trichotomy law*; it is a *partial,* rather than a *total* ordering.

The axioms of **SIA** entail that **R** differs in certain key respects from its counterpart in *constructive analysis* **CA** (section 4). For example, in **CA** the equality relation is balanced, while we have shown above that in **SIA** it is unbalanced. Also in **CA** the ordering relation < satisfies

(*) $$\neg(x < y \vee y < x) \rightarrow x = y;$$

and this is incompatible with the axioms of **SIA**. For (*) implies

(**) $$\forall x\neg(\ x < 0 \vee 0 < x) \rightarrow x = 0.$$

But in **SIA** it is easy to derive

$$\forall x\in\mathbf{\Delta}\neg(\ x < 0 \vee 0 < x),$$

and this, together with (**), would give **Δ** = {0}, contradicting the nondegeneracy of **Δ.**

In **CA** the object **Δ** is degenerate while the nondegeneracy of Δ in **SIA** is one of its characteristic features.

In **SIA** cohesive subsets of **R** correspond, *grosso modo,* to connected subsets of ℝ in classical analysis, that is, to intervals. This is borne out by the fact that any puncturing of **R** is *decomposable*, for it follows immediately from Axiom O6 that

$$\mathbf{R} - \{a\} = \{x: x > a\} \cup \{x: x < a\}.$$

The set **Q** of (smooth) *rational numbers* is defined as usual to be the set of all fractions of the form *m/n* with $m, n \in \mathbf{N}$, $n \neq 0$. The fact that **N** is cofinal in **R** ensures that **Q** is dense in **R**.

The set **R – Q** of *irrational* numbers is decomposable as

$$\mathbf{R} - \mathbf{Q} = [\{x: x > 0\} - \mathbf{Q}] \cup [\{x: x < 0\} - \mathbf{Q}].$$

This is in sharp contrast with the situation in *intuitionistic analysis* that is, **CA** augmented by Kripke's scheme, Brouwer's Continuity Principle, and bar induction. For we have observed (section 4) that in intuitionistic analysis not only is any puncturing of **R** cohesive, but that this is even the case for the irrational numbers. This would seem to indicate that in some sense the continuum in **SIA** is considerably less "syrupy" than its counterpart in **SIA.**

Comparing the smooth and Dedekind real lines in SIA. In **SIA** The usual set ℝ of real numbers can be constructed in **SIA** as Dedekind cuts. A *Dedekind real* is a pair $(U, V) \in \mathscr{P}\mathbf{Q} \times \mathscr{P}\mathbf{Q}$ (here $\mathscr{P}\mathbf{Q}$ is the power set of **Q,** the set of rational numbers) satisfying the conditions:

$$\exists x\, \exists y\, (x \in U \wedge y \in V)$$

$$U \cap V = \varnothing$$

$$\forall x\, (x \in U \leftrightarrow \exists y{\in}U.\ x < y)$$

$$\forall x\, (x \in V \leftrightarrow \exists y{\in}V.\ y < x)$$

$$\forall x \forall y (x < y \rightarrow x \in U \vee y \in V).$$

The set ℝ of Dedekind reals as thus defined can be turned into an ordered ring (Johnstone 1977)· This ring is always *constructively complete*, that is, satisfies the condition: Let A be an inhabited subset of ℝ that is bounded above. Then sup A exists if and only if for all x, $y \in$ ℝ with $x < y$, either y is an upper bound for A or there exists $a \in A$ with $x < a$. (A real number b is called a *supremum*, or *least upper bound*, of A if it is an upper bound for A and if for each $\varepsilon > 0$ there exists $x \in A$ with $x > b - \varepsilon$.)

Although ℝ is constructively complete, it is not conditionally complete in the classical sense. (This is because of the failure in **SIA** of the logical law $\neg p \vee \neg\neg p$. The failure of the latter in **SIA** follows immediately from the cohesiveness of **R** by considering the predicate $x \neq 0$.)

But ℝshares some features of the constructive reals not possessed by **R**, e.g.

$$\neg\neg x = y \rightarrow x = y$$

$$x \le y \wedge y \le x \rightarrow x = y$$

$$x^n = 0 \rightarrow x = 0.$$

There is a natural order preserving homomorphism φ: **R** → ℝ given by

$$\varphi(r) = (\{q{\in}\mathbf{Q}: q < r\}, \{q{\in}\mathbf{Q}: q > r\})$$

This is injective on **Q**, and embeds **Q** as the rational numbers in $\mathbb{R}$**.** Moreover, the kernel of φ coincides with the ideal **I** of strict infinitesimals in **R**, so φ induces an embedding of the quotient ring **R**/**I** into $\mathbb{R}$**. R**/**I** is **R** *shorn of its nilpotent infinitesimals*: it is both an intuitionistic field and an integral domain, that is, satisfies

$$\forall x(x \neq 0 \rightarrow x \textit{ is invertible}) \qquad \forall x \forall y[xy = 0 \rightarrow x = 0 \vee y = 0].$$

It can be shown that φ is *surjective*—so that $\mathbf{R}/\mathbf{I} \cong \mathbb{R}$—precisely when **R** is constructively complete in the sense above. In that event $\mathbb{R}$ is both an intuitionistic field and an integral domain.

In **SIA** the usual *open interval topology* can be defined on $\mathbb{R}$. It can be shown (Stout 1976) that with this topology $\mathbb{R}$ is always *connected* in the sense that it cannot be partitioned into two disjoint inhabited *open* subsets. In **SIA** $\mathbb{R}$ actually inherits a stronger cohesiveness property from **R.** To see this, call a subset X of a set A *detachable* if there is a subset Y of A such that $X \cap Y = \varnothing$, $X \cup Y = A$. Now we can show that, if X is a detachable subset of , then $\varphi[\mathbf{R}] \subseteq X$ or $X \cap \varphi[\mathbf{R}] = \varnothing$. For suppose $X \subseteq \mathbb{R}$ detachable and define $f: \mathbb{R} \rightarrow \{0, 1\}$ by $f(x) = 1$ if $x \in X$, $f(x) = 0$ if $x \notin X$. Then $f \circ \varphi: \mathbf{R} \rightarrow \{0, 1\}$ must be constant since **R** is cohesive. If $f \circ \varphi$ is constantly 1, then $\varphi[\mathbf{R}] \subseteq X$; if constantly 0, then $A \cap \varphi[\mathbf{R}] = \varnothing$. It follows easily that if φ is surjective then $\mathbb{R}$is itself cohesive.

***Nonstandard Analysis in* SIA.** In certain formulations of **SIA** the system of *natural numbers* possesses some intriguing features which make it possible to introduce another type of infinitesimal—the so-called *invertible* infinitesimals—resembling those of Nonstandard Analysis.

We recall that the set **N** of smooth natural numbers is required to satisfy not the full principle of mathematical induction for arbitrary properties but only the weaker restricted induction scheme. This raises the possibility that **N** may not coincide with the set ℕ of *standard natural numbers,* which is defined to be the *smallest* subset of **R** containing 0 and closed under the operation of adding 1. Now, models of **SIA** have been constructed (Moerdijk and Reyes 1991) in which ℕ is a *proper subset* of **N**; accordingly the members of **N** – ℕ may be considered *nonstandard integers.* Multiplicative inverses of nonstandard integers are infinitesimals, but, being themselves invertible, they are of a different type from the (necessarily noninvertible) nilpotent infinitesimals which are basic to **SIA.**

Proceeding formally, we define the set ℕ of *standard* natural numbers to be the intersection of all *inductive* subsets of **N**, i.e.,

$$\mathbb{N} = \{n \in \mathbf{N} : \forall X \subseteq \mathbf{N}[0 \in X \wedge \forall m(m \in X \to m+1 \in X) \to n \in X].$$

ℕ evidently satisfies *full induction*:

$$\forall X \subseteq \mathbf{N}[0 \in X \wedge \forall m(m \in X \to m+1 \in X) \to X = \mathbf{N}].$$

The space of *arithmetical infinitesimals* is the set

$$\mathbf{AIN} = \{x \in \mathbf{R}: \forall n \in \mathbb{N}(-1/(n+1) < x < 1/(n+1)\}.$$

This may be considered the largest infinitesimal neighbourhood of zero in **SIA**: it contains the space **Δ** of nilsquare infinitesimals as well as the space of *invertible* or *Robinsonian infinitesimals*

$$\mathbf{IN} = \{x \in \mathrm{AIN}: x \textit{ is invertible}\}.$$

As inverses of "infinitely large" reals (i.e. reals r satisfying $\forall n \in \mathbb{N}.\ n < r \ \vee \forall n \in \mathbb{N}.\ r < -n$) invertible infinitesimals are the counterparts in **SIA** of the infinitesimals of Nonstandard Analysis.

To assert the existence of invertible infinitesimals is to assert that **I**N be inhabited: this is equivalent to asserting that the set **N** – $\mathbb{N}$ of *nonstandard integers* be inhabited, or equivalently, that the following holds:

$$\exists n \in \mathbf{N} \forall m \in \mathbb{N}\ m < n.$$

When this condition is satisfied, as it is in certain models of **SIA,** we shall say that nonstandard integers, or invertible infinitesimals, are *actually present.* It is consistent to assert the actual presence of invertible infinitesimals, i.e., that **IN** be inhabited,.

One may also postulate the condition

$$\forall n \in \mathbf{N}[\forall x \in \mathbf{N} - \mathbb{N}\ (x > n) \rightarrow n \in \mathbb{N}],$$

i.e. "a natural number which is smaller than all nonstandard natural numbers must be standard". This is in fact equivalent to the condition that $\mathbb{N}$ be a stable subset of **N,** i.e. $\mathbf{N} - (\mathbf{N} - \mathbb{N}) = \mathbb{N}$. Assuming that nonstandard integers are actually present, this latter may be understood as asserting that as many as possible of these are actually present.

In the presence of invertible infinitesimals $\mathbb{R}$is a nonstandard model of the reals lacking nilsquare infinitesimals. The passage via the function φ from **R** to $\mathbb{R}$ eliminates the nilsquare infinitesimals but preserves the invertible infinitesimals. When φ is onto, $\mathbb{R}$ is a cohesive *nonstandard model of the reals.*

Consistency and Models of SIA. Smooth Infinitesimal Analysis is a fascinating theory with many attractive, even exotic features, but one may well ask whether is it *consistent.* Now the consistency of an axiomatic theory is usually established by producing a *model* for it, that is, providing an interpretation of the basic constituents of the theory under which all the axioms of the theory can be shown to be true. The consistency of **SIA** can be established by these means. *Models* of **SIA** are *categories* of a certain kind known as *toposes* (or *topoi*). These originated through the work of Alexander Grothendieck in algebraic geometry in the 1950s and 60s as a generalization of the idea of a category of sheaves on a topological space (hence the term *topos*). In the late 1960s F William Lawvere and Myles Tierney formulated a simple but powerful set of first-order axioms characterizing the idea underlying Grothendieck's concept of topos. A category satisfying the Lawvere-Tierney axioms was initially known as an *elementary topos,* and later just a *topos.* Although the origin of the topos concept lay in topology and algebraic geometry, elementary toposes have deep connections with logic and set theory. In particular, a topos can be thought of as a "generalized model of set theory".. The remarkable thing is that the body of laws satisfied by the logical operations in a topos—its "internal logic"—does not, in general, correspond to classical logic, but rather to the *intuitionistic* or *constructive* logic of Brouwer and Heyting in which the law of excluded middle is not affirmed.

Within a topos mathematical concepts can be formulated, arguments carried out and constructions performed much as one does in "ordinary" set theory. only observing the rules of intuitionistic logic.

In fact, any topos may be regarded as a model of *intuitionistic set theory* **IST**. (By intuitionistic set theory **IST** is meant the theory in intuitionistic first-order logic whose axioms are the "usual" axioms of Zermelo set theory (without the Axiom of Choice), namely: Extensionality, Pairing, Union, Power set, Infinity and Separation. For an exposition of **IST**, see Bell 2014). This means that, within a topos, mathematical constructions can be carried out as if within **IST.** In particular, the rational numbers can be defined as usual and the real numbers constructed by employing either Dedekind's procedure of making cuts in the rationals or Cantor's procedure employing equivalence classes of Cauchy sequences of rationals. While classically these two constructions lead to isomorphic results, this is *not true* in **IST**: indeed, a number of toposes have been constructed in which the ordered rings of Dedekind and Cantor reals fail to be isomorphic (Johnstone 2002).

Accordingly, in a topos-theoretic or constructive universe there is more than one candidate for the role of the mathematical continuum. The classical view that the linear continuum is a uniquely determined entity gives way to a pluralistic conception under which the continuum has a number of embodiments with essentially different properties. Even when one decides to choose the Dedekind reals as one's continuum, its properties may fall far short of those possessed by its classical counterpart. For example, in intuitionistic logic it cannot be proved that the Dedekind reals satisfy the least upper bound property. It can be shown (Johnstone 2002) that, in a topos**,** the Dedekind reals possess this property exactly when the logical law $\neg(\alpha \wedge \beta) \rightarrow (\neg\alpha \vee \neg\beta)$ holds there. This is an arresting instance of the connection between logic and the properties of the mathematical continuum made visible by the shift from classical to constructive logic.

The practice of topos theory quickly spawned an associated philosophy—jocularly known as "toposophy"— whose chief tenet is the idea that, like a model of set theory, any topos may be taken as a taken as an autonomous universe of discourse or "world" in which mathematical concepts can be interpreted and constructions performed (Bell 1986, 1988). These "worlds" opened up by topos theory can have startlingly different properties from the world of classical mathematics. For example, recall that in 1924 Brouwer proved from his intuitionistic principles that every real-valued function on a closed interval of the intuitionistic continuum of real numbers is uniformly continuous, in short, that the intuitionistic continuum has the *Brouwer Property.* This is of course inconsistent with the classical account of the continuum. But a number of toposes have been constructed in which the continuum of Dedekind reals has the Brouwer property (Mac Lane and Moerdijk 1992). In these all closed intervals in the Dedekind continuum are cohesive.

Toposes which are models of **SIA** have also been constructed. The basic ideas here go back to Lawvere. In the 1960s he conceived the ideas of developing the concept of smoothness in category-theoretic terms and of employing nilpotent infinitesimals in the calculus and differential geometry. The framework he formulated is known as *synthetic differential geometry* (**SDG**). The key principle of **SIA**, what we have termed the Infinitesimal Linearity Principle, was given its explicit form by Lawvere and Anders Kock, and, as remarked above, is often called the *Kock-Lawvere axiom.* The explicit construction of topos models of **SIA** in which differential geometry can be fully developed – so-called *smooth toposes* (Moerdijk and Reyes 1991) was first achieved by Dubuc in 1979. This, no easy task, established fully the consistency of **SIA.**

Contrasting Nonstandard Analysis with SIA. **SIA** shares with Nonstandard Analysis (**NSA)** the feature that continuity is represented by the idea of "preservation of infinitesimal closeness". Nevertheless, there are a number of differences between the two approaches:

- In models of **SIA**, only smooth maps between objects are present. In models of **NSA,** all set-theoretically definable maps (including, in particular, discontinuous ones) appear.
- The logic of **SIA** is intuitionistic, while the logic of **SIA** is classical. (It should be pointed out, however, that constructive versions of **NSA** have been developed. See Palmgren 1998).
- In **SIA**, all curves are infinitesimally straight. Nothing resembling this is present in **NSA**.

- The nilpotency of the infinitesimals of **SIA** reduces the differential calculus to simple algebra. In **NSA** the use of infinitesimals is a disguised form of the classical limit method.
- The hyperreal line in **NSA** is obtained by augmenting the classical real line with infinitesimals (and infinite numbers), while the smooth real line **R** comes already equipped with infinitesimals.
- In any model of **NSA**, the hyperreal line $\mathbb{R}^\star$ has exactly the same set-theoretically expressible properties as does the classical real line: in particular $\mathbb{R}^\star$ is an archimedean field in the sense of that model. This means that the infinitesimals (and infinite numbers) of **NSA** are not intrinsically so in the sense of the model in which they "live", but only relative to the "standard" model with which the construction began. That is, speaking figuratively, an inhabitant of a model of **NSA** would be unable to detect the presence of infinitesimals or infinite numbers in $\mathbb{R}^\star$. This contrasts with **SIA** in two respects. First, in models of **SIA** containing invertible infinitesimals, the real line is nonarchimedean with respect to the set of standard natural numbers, which is itself an object of the model. In other words, the presence of (invertible) infinitesimals and infinite numbers would be perfectly detectable by an inhabitant of the model. And secondly, the characteristic property of nilpotency possessed by the microquantities of a model of **SIA** is an intrinsic property, perfectly identifiable within the model. In **NSA** the hyperreals have precisely the same algebraic properties as do the classical real numbers, but the smooth reals in **SIA** do not.

The differences between **NSA** and **SIA** arise because the former is essentially a theory of infinitesimal numbers designed to provide a succinct formulation of the limit concept, while the latter is, by contrast, a theory of infinitesimal geometric objects, designed to provide an intrinsic formulation of the concept of differentiability.

SIA and Physics. In the past physicists showed no hesitation in employing infinitesimal methods, In this connection we recall the words of Hermann Weyl:

> *The principle of gaining knowledge of the external world from the behaviour of its infinitesimal parts is the mainspring of the theory of knowledge in infinitesimal physics as in Riemann's geometry and, indeed, the mainspring of all the eminent work of Riemann* (Weyl 1922, p. 92).

The use of infinitesimals relied on the implicit assumption that the (physical) world is smooth, or at least that the maps encountered there are differentiable as many times as needed. For this reason **SIA** provides an ideal framework for the rigorous derivation, using infinitesimals, of results in classical physics (Bell 1998). We present two of these here.

First, we derive the *equation of continuity for fluids*, first derived in 1757 by Euler.by original derivation by Euler The derivation in **SIA** will follow Euler's very closely, but the use of nilsquare infinitesimals and the Cancellation Principle will render the argument entirely rigorous.

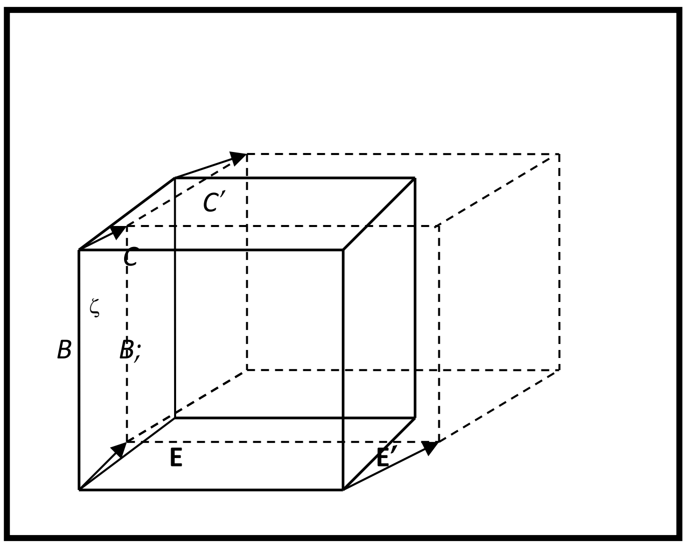
Before we begin we require a few observations on *partial derivatives* in **SIA**. Given a function $f\colon \mathbf{R}^n \to \mathbf{R}$ of n variables $x_1, ..., x_n$, the partial derivative $\frac{\partial f}{\partial x_i}$ is defined as usual to be the derivative of the function $f(a_1, \ldots, x_i, \ldots, a_n)$ obtained by fixing the values of all the variables apart from x_i. In that case, for an arbitrary infinitesimal ε, we have

(1) $$f(x_1, \ldots, x_i + \varepsilon, \ldots, x_n) = f(x_1, \ldots, x_n) + \varepsilon \frac{\partial f}{\partial x_i}(x_1, \ldots, x_n).$$

Using the fact that $\varepsilon^2 = 0$, it is then easily shown that

(2) $$f(x_1 + a_1\varepsilon, \ldots, x_n + a_n\varepsilon) = f(x_1, \ldots, x_n) + \varepsilon \sum_{i=1}^{n} a_i \frac{\partial f}{\partial x_i}(x_1, \ldots, x_n).$$

These equations are pivotal in deriving the equation of continuity. Here we are given a inviscid fluid of varying density flowing smoothly in space. At any point $O = (x, y, z)$ in the fluid and at any time t, the fluid's density ρ and the components u, v, w of the fluid's velocity are given as functions of x, y, z, t. Following Euler, we consider the elementary volume element **E**—an infinitesimal parallelepiped—with origin O and edges OA, OB, BC of infinitesimal lengths ε, η, ζ and so of mass εηζρ:



Fluid flow during the infinitesimal time τ transforms the volume element **E** into the infinitesimal parallelepiped **E′** (see figure above) with vertices O' , A', B', C'. We first calculate the length of the side $O'A'$. Now, using (1), the rate at which A is moving away from O in the x-direction is

$$u(x + \varepsilon, y, z, t) - u(x , y, z, t) \; = \varepsilon\frac{\partial u}{\partial x}.$$

The change in length of OA during the infinitesimal time τ is thus $\varepsilon\tau\frac{\partial u}{\partial x}$, so that the length of $O'A'$ is

$$\varepsilon + \varepsilon\tau\frac{\partial u}{\partial x} = \varepsilon\left(1 + \tau\frac{\partial u}{\partial x}\right).$$

Similarly, the lengths of $O'B'$ and $O'C'$ are, respectively,

$$\eta\left(1 + \tau\frac{\partial v}{\partial y}\right), \quad \zeta\left(1 + \tau\frac{\partial w}{\partial z}\right).$$

The volume of **E′** is the product of these three quantities, which, using the fact that $\tau^2 = 0$, comes out as

(3) $$\varepsilon\eta\zeta\left[1 + \tau\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right)\right].$$

Since the coordinates of O' are $(x+u\tau,\ y+v\tau,\ z+w\tau)$, the fluid density ρ' there at time$t + \tau$ is, using (2),

(4) $$\rho + \tau\left(\frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + w\frac{\partial \rho}{\partial z}\right).$$

The mass of **E′** is then the product of (3) and (4), which, again using the fact that that $\tau^2 = 0$, comes out as

(5) $$\varepsilon\eta\zeta\rho + \varepsilon\eta\zeta\tau\left(\frac{\partial \rho}{\partial t} + \rho\frac{\partial u}{\partial x} + \rho\frac{\partial v}{\partial y} + \rho\frac{\partial w}{\partial z} + u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + w\frac{\partial \rho}{\partial z}\right).$$

Now by the principle of conservation of mass, the masses of the fluid in **E** and **E′** are the same, so equating the mass $\varepsilon\eta\zeta\rho$ of **E** to the mass of **E′** given by (5) yields

$$\varepsilon\eta\zeta\tau\left(\frac{\partial \rho}{\partial t} + \rho\frac{\partial u}{\partial x} + \rho\frac{\partial v}{\partial y} + \rho\frac{\partial w}{\partial z} + u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + w\frac{\partial \rho}{\partial z}\right) = 0.$$

The Cancellation Principle gives

$$\frac{\partial \rho}{\partial t} + \rho\frac{\partial u}{\partial x} + \rho\frac{\partial v}{\partial y} + \rho\frac{\partial w}{\partial z} + u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + w\frac{\partial \rho}{\partial z} = 0,$$

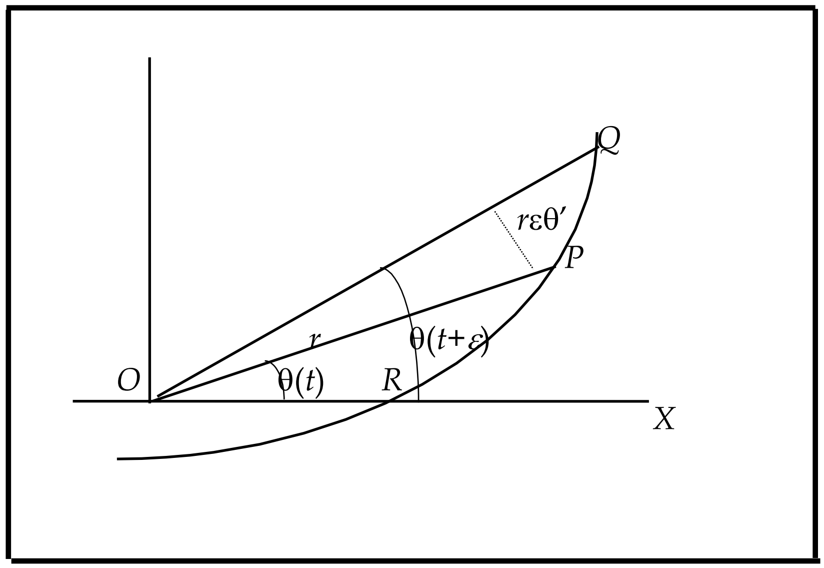
i.e.,

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) + \frac{\partial}{\partial z}(\rho w) = 0\ .$$

This is Euler's equation of continuity.

Next, we derive the *Kepler-Newton areal law of motion under a central force.* We suppose that a particle executes plane motion under the influence of a force directed towards some fixed point O. If P is a point on the particle's trajectory with coordinates x, y, we write r for the length of the line PO and θ for the angle that it makes with the x-axis OX. Let A be the area of the sector ORP, where R is the point of intersection of the trajectory with OX. We regard x, y, r, θ as functions of a time variable t: thus

$$x = x(t),\ y = y(t),\ r = r(t),\ \theta = \theta(t),\ A = A(t).$$



Now let Q be a point on the trajectory at which the time variable has value $t + \varepsilon$, with ε in Δ (figure above). Then by Infinitesimal Straightness the sector OPQ is a triangle of base $r(t + \varepsilon) = r + \varepsilon r'$ and height

$$r \sin[\theta(t + \varepsilon) - \theta(t)] = r \sin \varepsilon\theta' = r\varepsilon\theta'.$$

Here we have used the fact in **SIA** $\sin \varepsilon = \varepsilon$ for infinitesimal ε (Bell 1998).

The area of OPQ is accordingly

$$2 \text{ base} \times \text{height} = 2\,(r + \varepsilon r')r\varepsilon\theta' = 2(r^2\varepsilon\theta' + \varepsilon^2 rr'\,\theta') = 2\,r^2\varepsilon\theta'.$$

Therefore

$$\varepsilon A'(t) = A(t + \varepsilon) - A(t) = \mathit{area}\ OPQ = 2\varepsilon r^2\theta',$$

so that, cancelling ε,

$$A'(t) = 2r^2\theta'. \qquad (*)$$

Now let $H = H(t)$ be the acceleration towards O induced by the force. Resolving the acceleration along and normal to OX, we have

$$x'' = H\cos\theta \qquad y'' = H\sin\theta.$$

Also $x = r\cos\theta$, $y = r\sin\theta$. Hence

$$yx'' = Hy\cos\theta = Hr\sin\theta\cos\theta \qquad xy'' = Hx\sin\theta = Hr\sin\theta\cos\theta,$$

from which we infer that

$$(xy' - yx')' = xy'' - yx'' = 0.$$

Hence

$$xy' - yx' = k, \qquad (**)$$

where k is a constant.

Finally, from $x = r\cos\theta$, $y = r\sin\theta$, it follows in the usual way that

$$xy' - yx' = r^2\theta',$$

and hence, by (**) and (*), that

$$A'(t) = 2k.$$

Assuming $A(0) = 0$, we conclude that

$$A(t) = 2kt.$$

Thus the radius vector joining the body to the point of origin sweeps out equal areas in equal times (Kepler's law).

There is an intriguing use of infinitesimals in Einstein's celebrated 1905 paper *On the Electrodynamics of Moving Bodies* (Einstein et al 1952) in which the special theory of relativity is first formulated. In deriving the Lorentz transformations from the principle of the constancy of the velocity of light Einstein obtains the following equation for the time coordinate $\tau(x', y, z, t)$ of a moving frame:

(i) $$\tfrac{1}{2}\left[\tau(0,0,0,t) + \tau\left(0,0,0,t + \frac{x'}{c-v} + \frac{x'}{c+v}\right)\right] = \tau\left(x',0,0,t + \frac{x'}{c-v}\right).$$

He continues:

Hence, if x′ be chosen infinitesimally small,

(ii) $$\tfrac{1}{2}\left(\frac{1}{c-v} + \frac{1}{c+v}\right)\frac{\partial\tau}{\partial t} = \frac{\partial\tau}{\partial x'} + \frac{1}{c-v}\frac{\partial\tau}{\partial t},$$

or

$$\frac{\partial\tau}{\partial x'} + \frac{v}{c^2 - v^2}\frac{\partial\tau}{\partial t} = 0.$$

Now the derivation of equation (ii) from equation (i) can be simply and rigorously carried out in **SIA** by *choosing x′ to be a nilsquare infinitesimal* ε. For then (i) becomes

$$\tfrac{1}{2}\left[\tau(0,0,0,t) + \tau\left(0,0,0,t + \varepsilon\left(\frac{1}{c-v} + \frac{1}{c+v}\right)\right)\right] = \tau\left(\varepsilon,0,0,t + \frac{\varepsilon}{c-v}\right).$$

From this we get, using equation (1) above,

$$\tau(0,0,0,t) + \tfrac{1}{2}\varepsilon\left(\frac{1}{c-v} + \frac{1}{c+v}\right)\frac{\partial\tau}{\partial t} = \tau(0,0,0,t) + \varepsilon\left(\frac{\partial\tau}{\partial x'} + \frac{1}{c-v}\frac{\partial\tau}{\partial t}\right).$$

So

$$\tfrac{1}{2}\varepsilon\left(\frac{1}{c-v} + \frac{1}{c+v}\right)\frac{\partial\tau}{\partial t} = \varepsilon\left(\frac{\partial\tau}{\partial x'} + \frac{1}{c-v}\frac{\partial\tau}{\partial t}\right),$$

and (ii) follows by the Cancellation Principle.

Spacetime metrics have some intriguing properties in **SIA**. In a spacetime the metric can be written in the form

(*) $$ds^2 = \Sigma g_{\mu\nu}dx_\mu dx_\nu \qquad \mu,\nu = 1, 2, 3, 4.$$

In the classical setting (*) is in fact an abbreviation for an equation involving derivatives and the "differentials" ds and dx_μ are not really quantities at all. What form does this equation take in **SIA**? Notice that the "differentials" cannot be taken as infinitesimals in the sense of **SIA** since all the squared terms would vanish. But the equation does have a very natural form in **SIA**. Here is an informal way of obtaining it.

Think of the dx_μ as being multiples $k_\mu e$ of some small quantity e. Then (*) becomes

$$ds^2 = e^2\Sigma g_{\mu\nu}k_\mu k_\nu,$$

so that

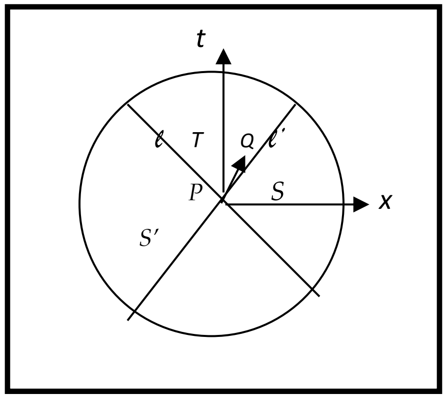
$$ds = e\sqrt{\Sigma g_{\mu\nu}k_\mu k_\nu}\ .$$

Now replace *e* by a nilsquare infinitesimal ε. Then we obtain the *metric relation*: in **SIA**

$$ds = \varepsilon\sqrt{\Sigma g_{\mu\nu}k_\mu k_\nu}\ .$$

This tells us that the "infinitesimal distance" d*s* between a point *P* with coordinates (x_1, x_2, x_3, x_4) and an infinitesimally near point *Q* with coordinate ($x_1 + k_1\varepsilon$, $x_2 + k_2\varepsilon$, $x_3 + k_3\varepsilon$, $x_4 + k_4\varepsilon$) is $\varepsilon\sqrt{\Sigma g_{\mu\nu}k_\mu k_\nu}$. Here a curious situation arises. For when the "infinitesimal interval" d*s* between *P* and *Q* is timelike (or lightlike), the quantity $\Sigma g_{\mu\nu}k_\mu k_\nu$ is nonnegative, so that its square root is a real number. In this case d*s* may be written as ε*d*, where *d* is a real number. On the other hand, if d*s* is spacelike, then $\Sigma g_{\mu\nu}k_\mu k_\nu$ is negative, so that its square root is imaginary. In this case, then, d*s* assumes the form iε*d*, where *d* is a real number (and, of course $i = \sqrt{-1}$). On comparing these we see that, if we take ε as the "infinitesimal unit" for measuring infinitesimal *timelike* distances, then iε serves as the "imaginary infinitesimal unit" for measuring infinitesimal *spacelike* distances.

For purposes of illustration (see figure below), let us restrict the spacetime to two dimensions (*x, t*), and assume that the metric takes the simple form $ds^2 = dt^2 - dx^2$. The infinitesimal light cone at a point *P* divides the infinitesimal neighbourhood at *P* into a timelike region *T* and a spacelike region *S* bounded by the null lines l and l′ respectively (see figure below). If we take *P* as origin of coordinates, a typical point *Q* in this neighbourhood will have coordinates ($a\varepsilon$, $b\varepsilon$) with *a* and *b* real numbers: if $|b| > |a|$, *Q* lies in *T;* if $a = b$, *P* lies on l or l′; if $|a| < |b|$, *P* lies in *S*. If we write $d = \sqrt{|a^2 - b^2|}$, then in the first case, the infinitesimal distance between *P* and *Q* is ε*d,* in the second, it is 0, and in the third it is iε*d.*



At the beginning of the 20th century Poincaré and Minkowski introduced "i*ct*" to replace the "*t*" coordinate so as to make the metric of relativistic spacetime positive definite. This was purely a matter of formal convenience and was later rejected by (general) relativists. In conventional physics one never works with nilpotent quantities so it is always possible to replace formal imaginaries by their (negative) squares. But spacetime theory in **SIA** *forces* one to use imaginary units, since, infinitesimally, one can't "square oneself out of trouble". This being the case, it would seem that, infinitesimally the dictum *Farewell to ict* (Misner, Thorne and Wheeler 1971) needs to be replaced by

Vale "i*ct*", *ave* "iε" !

We conclude this section with a speculation. The infinitesimal domain Δ can be seen to be "tiny" in the order-theoretic sense. For, using ε, η, as variables ranging over Δ, it is not hard to show that

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

whence

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

In particular, the members of **Δ** are all simultaneously ≤ 0 and ≥ 0 but cannot (because of the nondegeneracy of **Δ**) be shown to coincide with zero.

In his book *Just Six Numbers* (Rees 2001) the astrophysicist Martin Rees comments on 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).*

On this account, Planck scales seem very similar in certain respects to **Δ**. In particular, the sentence (*) 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** serve as a suitable model for "Planck scales"? While **Δ** is unquestionably small enough to play the role, it 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 universally continuous. So if Planck scales could indeed be modelled by microneighbourhoods in **SIA**, then one might begin to suspect 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 major victory for the Continuous in its long struggle with the Discrete.

References

Alexander, H.G., ed. (1956). *The Leibniz-Clarke Correspondence.* University of Manchester Press.

Barnes, J. (1986). *The Presocratic Philosophers.* Routledge.

Baron, M. E. (1987). *The Origins of the Infinitesimal Calculus.* Dover.

Beeson, M.J. (1985). *Foundations of Constructive Mathematics.* Springer-Verlag.

Bell, E. T. (1945). *The Development of Mathematics*, 2nd ed. McGraw-Hill.

——(1965). *Men of Mathematics,* 2 vols. Penguin Books.

Bell, John L. (1986). *From absolute to local mathematics.* Synthese 69, 409-426.

——(1988). *Toposes and Local Set Theories*: *An Introduction.* Oxford Logic Guides 14. Clarendon Press, Oxford. Reprinted by Dove, 2008.

——(1998) *A Primer of Infinitesimal Analysis.* Cambridge University Press. Second edition 2oo8.

——(2001). *The continuum in smooth infinitesimal analysis.* In Schuster, Berger and Osswald (2001), pp. 19-24.

——(2016). *Oppositions and Paradoxes: Philosophical Perplexities in Science and Mathematics.* Broadview Press.

—— (2019). *The Continuous, the Discrete and the Infinitesimal in Philosophy and Mathematics.* Speinger.

——and Machover, M. (1977). *A Course in Mathematical Logic.* North-Holland.

Benacerraf, P., and H. Putnam (1977) *Philosophy of Mathematics: Selected Readings*, 2nd ed. Cambridge University Press.

Bishop, E. (1967). *Foundations of Constructive Analysis.* New York: McGraw-Hill.

—— and D. Bridges (1985). *Constructive Analysis*. Berlin: Springer.

Bolzano, B. (1950). *Paradoxes of the Infinite.* Tr. Prihovsky. Routledge & Kegan Paul.

Boyer, C., (1959). *The History of the Calculus and its Conceptual Development.* Dover.

——(1968). *A History of Mathematics.* Wiley.

——and U. Merzbach (1989). *A History of Mathematics,* 2nd edition. Wiley.

Brentano, F. (1974) *Psychology from an Empirical Standpoint.* Humanities Press.

——(1988). *Philosophical Investigations on Space, Time and the Continuum.* Tr. Smith, B.. Croom Helm.

Bridges, D. (1994). *A constructive look at the real line.* In Ehrlich, ed. (1994), pp. 29-92.

——(1999). *Constructive mathematics: a foundation for computable analysis.* Theoretical Computer Science **219**, pp. 95-109.

——and F. Richman (1987). *Varieties of Constructive Mathematics.* Cambridge University Press.

Brouwer, L.E.J. (1975). *Collected Works, 1.* Edited by A. Heyting. North-Holland.

Cajori, F. (1919). *A Short History of the concepts of Limits and Fluxions.* Chicago: Open Court.

Cantor, G. (1872). *Über die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen*, Mathematische Annalen 5, 123–132.

—— (1878). *Ein Beitrag zur Mannigfaltigkeitslehre,* Journal für die reine und angewandte Mathematik 84, 242–58.

—— (1961). *Contributions to the Founding of the Theory of Transfinite Numbers.* Dover.

Cohen, P.J. (1966). *Set Theory and the Continuum Hypothesis.* W. A. Benjamin Inc.

Cohn, P. M. (1957). *Lie Groups.* Cambridge University Press.

Dauben, J. (1979). *Georg Cantor: His Mathematics and Philosophy of the Infinite.* Harvard University Press.

Dedekind, R. (1963). *Essays on the Theory of Numbers.* Dover.

Dummett, M. (1977). *Elements of Intuitionism.* Clarendon Press, Oxford.

Ehrlich, P., ed. (1994). *Real Numbers, Generalizations of the Reals, and Theories of Continua*. Kluwer.

Ehrlich, P. (1994a). *All numbers great and small.* In Ehrlich, ed. (1994), pp. 239-258.

—— (2006), *The rise of non-Archimedean mathematics and the roots of a misconception. I. The emergence of non-Archimedean systems of magnitudes.* Archive for History of Exact Sciences ***60*** *(1)* pp. 1–121.

Einstein, A. et al. (1952). *The Principle of Relativity : a Collection of Original Memoirs on the Special and General Theory of Relativity.* Tr. Perrett and Jeffrey. Dover. Reprint of 1923 Methuen edition.

Euler, L. (1990). *Introduction to Analysis of the Infinite.* Tr. Blanton. New York: Springer.

Evans, M. (1955). *Aristotle, Newton, and the theory of continuous magnitude,* Journal of the History of Ideas **16,** pp. 548-557.

Ewald, W. (1999). *From Kant to Hilbert, A Source Book in the Foundations of Mathematics.* Vols. I and II. Oxford University Press.

Fine, H. B. (1917) *Ratio, proportion and measurement in the Elements of Euclid.* Annals of Mathematics, Second Series, Vol. 19, No. 1 , pp. 70 -76.

Fisher, G. (1981), *The infinite and infinitesimal quantities of du Bois-Reymond and their reception.* Archive for History of Exact Sciences. **24** (2) pp. 101-163.

Folina, J. (1992). *Poincaré and the Philosophy of Mathematics.* St. Martin's Press.

Fraenkel, A. (1976). *Abstract Set Theory.* Fourth edition of first (1953) edition, revised by A. Levy. North-Holland.

Fraenkel, A., Y. Bar-Hillel and A. Levy (1973). *Foundations of Set Theory,* 2nd edition. Amsterdam: North-Holland.

Frege, G. (1984). *Collected Papers on Mathematics, Logic and Philosophy.* B. McGuinness, ed. Blackwell.

——(1980) *Foundations of Arithmetic.* Tr. Austin. Northwestern University Press.

Gray, J. (1973). *Ideas of Space: Euclidean, Non-Euclidean, and Relativistic.* Clarendon Press, Oxford.

Hallett, M. (1984). *Cantorian Set Theory and Limitation of Size.* Clarendon Press, Oxford.

Heath, T. (1949). *Mathematics in Aristotle.* Oxford University Press.

——(1981). *A History of Greek Mathematics,* 2 vols. Dover.

Heijenoort, J. van (1981). *From Frege to Gödel: A Sourcebook in Mathematical Logic 1879 – 1931.* Harvard University Press.

Hellman, G, and S. Shapiro (2018). *Varieties of Continua: From Regions to Points and Back.* Oxford University Press.

Heyting, A. (1956). *Intuitionism: An Introduction.* North-Holland.

Hobson, E. W. (1957). *The Theory of Functions of a Real Variable.* Dover.

Johnstone, P.T. (1977) *Topos Theory.* Academic Press.

— —(1982) *Stone Spaces.* Cambridge Studies in Advanced Mathematics 3. Cambridge University Press.

— —(2002) *Sketches of an Elephant: A Topos Theory Compendium, Vols. I and II.* Oxford Logic Guides Vols. 43 and 44. Clarendon Press.

Kahn, C. (2001). *Pythagoras and the Pythagoreans: A Brief History.* Hackett.

Kant, I. (1964). *Critique of Pure Reason.* Macmillan.

Keisler, H. (1994). *The hyperreal line.* In Ehrlich (1994), pp. 207-238.

Kirk, G. S., J. E. Raven, and M. Schofield (1983). *The Presocratic Philosophers*, 2nd ed. Cambridge University Press.

Kline, M. (1972). *Mathematical Thought from Ancient to Modern Times.* 3 volumes. Oxford University Press.

Kneebone, G.T. (1963) *Mathematical Logic and the Foundations of Mathematics: An Introductory Survey.* London: Van Nostrand.

Kock, A. (1981). *Synthetic Differential Geometry.* Cambridge: Cambridge University Press.

Kretzmann, N, (ed.) (1982). *Infinity and Continuity in Ancient and Medieval Thought.* Cornell University Press.

Laugwitz, A.H. and Schmieden, C. (1958). *Eine Erweiterung der Infinitesimalrechnung,* Mathematische Zeitschrift **69**, pp. 1–39.

Lawvere, F. W. (1971). *Quantifiers and sheaves.* In Actes du Congrès Intern. Des Math. Nice 1970, tome I. Paris: Gauthier-Villars, pp. 329-34.

— —(2011) *Euler's continuum functorially vindicated.* In Vintage Enthusiasms: Essays in Honour of John L. Bell, D.Devidi, P. Clark and M. Hallett, eds., Springer.

Leibniz, G. (1951). *Selections.* Ed. Wiener. Scribners.

— —(1961). *Philosophical Writings.* Tr. Morris. Everyman Library, Dent Dutton.

——(2001). *The Labyrinth of the Continuum : Writings on the Continuum Problem, 1672-1686* . Tr. Arthur. Yale University Press.

Mac Lane, S. (1986). *Mathematics: Form and Function.* Springer-Verlag.

Mac Lane, S. and I. Moerdijk, I. (1992). *Sheaves in Geometry and Logic: A First Introduction to Topos Theory.* Springer-Verlag.

Mancosu, P. (1996). *Philosophy of Mathematics and Mathematical Practice in the Seventeenth Century.* New York, Oxford: Oxford University Press.

——(1998). *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the1920s*. Clarendon Press.

McShane. E. J. and Botts, T. A. (1959). *Real Analysis.* Van Nostrand.

Misner, C., K. Thorne, and J. Wheeler (1972). *Gravitation.* Freeman.

Moerdijk, I. (1995). *A model for intuitionistic non-standard arithmetic.* Annals of Pure and Applied Logic **73**, pp. 297-325.

——and G. E. Reyes (1991). *Models for Smooth Infinitesimal Analysis.* Springer-Verlag.

Moore, A. W. (1990). *The Infinite.* Routledge.

Palmgren, E. (1998). *Developments in constructive nonstandard analysis.* Bulletin of Symbolic Logic **4**, pp. 233-272.

——(2001). *Unifying constructive and nonstandard analysis.* In Schuster, Berger and Osswald (2001), pp. 167-83.

Peirce, C.S. (1976). *The New Elements of Mathematics,* Vol. III. Edited by Carolyn Eisele. Mouton Publishers and Humanities Press.

——(1992). *Reasoning and the Logic of Things.* Edited by Kenneth Laine Ketner. Harvard University Press.

Poincaré. H. (1946). *Foundations of Science.* Tr. Halstead. Science Press.

——(1963). *Mathematics and Science: Last Essays.* Dover.

Pyle, A. (1997). *Atomism and Its Critics.* Thoemmes Press.

Rescher, N. (1967). *The Philosophy of Leibniz*. Prentice-Hall.

Rees, M. (2001). *Just Six Numbers: The Deep Forces that Shape the Universe.* Basic Books.

Robinson, A. (1996). *Non-Standard Analysis.* Princeton University Press.

Russell, B. (1945). *A History of Western Philosophy*. Simon and Schuster.

——(1958). *A Critical Exposition of the Philosophy of Leibniz,* 2nd ed. Allen & Unwin.

——(1964). *The Principles of Mathematics.* Allen and Unwin.

——(1995) *Introduction to Mathematical Philosophy*. Routledge.

Schuster, P., U. Berger and H. Osswald (2001). *Reuniting the Antipodes – Constructive and Nonstandard Views of the Continuum.* Dordrecht: Kluwer.

Smith, D.E. (1959). *A Source Book in Mathematics.* Dover.

Spivak, M. (1975). *A Comprehensive Introduction to Differential Geometry***.** Publish or Perish Press.

Stout, L. (1976). *Topological properties of the real numbers object in a topos*. Cahiers Topologie Géom. Différentielle **17**, no. 3, (1976), pp. 295-326.

Van Atten, M., D. van Dalen and R. Tieszen (2002). *The phenomenology and mathematics of the intuitive continuum.* Philosophia Mathematica **10**, pp. 203-236.

Van Dalen, D. (1995). *Hermann Weyl's intuitionistic mathematics*. *Bulletin of Symbolic Logic* **1**, no. 2, pp. 145-169.

— — (1997). *How connected is the intuitionistic continuum?* Journal of Symbolic Logic **62**, pp. 1147-50.

— — (1998). *From a Brouwerian point of view.* Philosophia Mathematica **6**, pp. 209-226.

Wagon, S. (1993). *The Banach-Tarski Paradox.* Cambridge University Press.

Weyl, H. (1921*), On the new foundational crisis in mathematics.* (English translation of 'Über der neue Grundslagenkrise der Mathematik,' Mathematische Zeitschrift **10**, 1921, 37-79.) In Mancosu (1998), pp. 86 - 122.

— — (1929). *Consistency in Mathematics*. Rice Institute Pamphlet 16, 245-265. Reprinted in Weyl (1968) II, pp. 150-170.

— — (1932). *The Open World: Three Lectures on the Metaphysical Implications of Science.* Yale University Press.

— — (1940). *The ghost of modality*. Philosophical Essays in Memory of Edmund Husserl, Harvard University Press. Reprinted in Weyl (1968) III, pp. 684-709.

— — (1949). *Philosophy of Mathematics and Natural Science.* Princeton University Press. (An expanded Engish version of *Philosophie der Mathematik und Naturwissenschaft,* Leibniz Verlag, 1927.)

— — (1950). *Space-Time-Matter*. Tr. Brose, H. Dover. (English translation of *Raum, Zeit, Materie,* Springer Verlag, 1918.)

— — (1954). *Address on the unity of knowledge*. Columbia University Bicentennial Celebration. Reprinted in Weyl (1968) IV, pp. 623 - 630.

— — (1968). *Gesammelte Abhandlungen,* I-IV, K. Chandrasehharan. Springer-Verlag.

— — (1969). *Insight and reflection*. (Lecture delivered at the University of Lausanne, Switzerland, May 1954. Translated from German original in *Studia Philosophica,* **15**, 1955.) In T.L. Saaty and F.J. Weyl, eds., *The Spirit and Uses of the Mathematical Sciences,* pp. 281-301. McGraw-Hill, 1969.

——(1985). *Axiomatic versus constructive procedures in mathematics.*, ed. T. Tonietti. *Mathematical Intelligencer* **7**, no. 4, pp. 10-17, 38.

——(1987). *The Continuum: A Critical Examination of the Foundation of Analysis,* tr. S. Pollard and T. Bole. Thomas Jefferson University Press. (English translation of *Das Kontinuum,* Leipzig: Veit, 1918.)

——(1998). *On the New Foundational Crisis in Mathematics*. (English translation of 'Über der neue Grundslagenkrise der Mathematik,' *Mathematische Zeitschrift* **10**, 1921, pp. 37-79.) In Mancosu (1998), pp. 86 - 122.

——(1998a). *On the Current Epistemological Situation in Mathematics*. English translation of 'Die Heutige Erkenntnislage in der Mathematik, *Symposion* **1**, 1925-27, pp. 1-32.) In Mancosu (1998), pp. 123-142.

White, M. J. (1992). *The Continuous and the Discrete: Ancient Physical Theories from a Contemporary Perspective* Clarendon Press, Oxford.