4 Numbering the Continuum

There are two senses in which the continuum can be said to have been numbered. (1) The (linear) continuum has been replaced by, or is even in some circumstances identified with, a set of numbers – the real numbers. (2) The continuum has been assigned a cardinal number $2^{\aleph_0}$, i.e. sense has been given to the question 'How many points are there in a line?' and a partial answer given. In the light of what was said in chapter 3 it is clear that (2) could not have come about without (1) or something like it. The continuum had to be represented by a set of points with a determinate membership before it could be assigned a number.

How could this transition have come about? How were the paradoxes of the infinite overcome? So far we have seen that the classical finitist can stand his ground, admitting only the notion of the potentially infinite, without thereby being committed to the existence of any actual infinite provided that (a) he distinguishes between continuous and discrete wholes, between wholes given prior to their parts (where the identity of the parts depends crucially on that of the whole of which they are part) and wholes given after their parts (where the identity of the whole is determined by that of its parts); (b) that he insists on the distinction between essential and accidental generalization, or at least between extensional and non-extensional readings of 'all As are B'; and (c) that he insists on the distinction between the indefinite and the potentially infinite.

Moreover, even if one admits that every potential infinite presupposes an actual infinite, this still does not overcome the apparent contradiction involved in thinking of a continuum as made up of points. The actual infinite, even if it is metaphysically inevitable, does not thereby become a possible object of knowledge, or a contradiction-free and therefore usable mathematical concept. Pascal indeed used the actual infinite as a foil, as a means of proving the existence of a being, knowledge and understanding of which transcends all human rational capacities. We can know of its existence but cannot comprehend it. This is not without significance for, as we shall see, Cantor too was obliged to admit a notion of the absolutely infinite, which he also associated with God and which had to be placed outside the range of mathematical computation and comprehension.

In this situation the classical finitist has a strong case. He is at least in a good position to engage in metaphysical and epistemological arguments with his opponent and has the upper hand epistemologically, where it would seem that the actual infinite can play no significant role. But his position will be changed radically if (a) it can be shown that a coherent conception of the continuum as an infinite collection of points is after all possible, and (b) that the actual infinite plays a significant role in the mathematics which is used in and is necessary to the natural sciences, and physics in particular.

1 The Algebraization of Geometry

The pressures which brought down the edifice of classical, Aristotelian finitism did indeed come from within mathematics and physics. Ultimately they derive from the demand for a numerical, practically applicable handling of continuous magnitudes and in particular of continuous change (including, of course, motion). With hindsight it can be seen that the crucial moves had already been made by Descartes in his *La Géométrie*, where he argues that Euclid-style definitions of geometric figures should be replaced by definitions given in the form of algebraic equations. From this the notion of a function rapidly followed in the work of Leibniz and Newton, and it is the subsequent development of this concept (which is all-important to the mathematical physicist) which apparently dictates the punctualization of the continuum. But at the same time it introduces a new, specifically mathematical conception of totality (set or class) – a whole given neither before nor after its parts, whose membership is to be regarded as determinate, generalization over which must be treated as extensional but non-accidental. In other words, there arise mathematical conceptions of

totality which cross-cut prior philosophical distinctions and which primarily inform Cantorian set theory and subsequent axiomatizations.

The logicist programme for providing a rigorous foundation for analysis superimposed the logical notion of class on these mathematical conceptions in a way which conceals their distinctive character (for indeed the logicist claims that there is here no distinction). But, in the face of Zeno's paradoxes, a condition of the possibility of treating the continuum as a totality of points, without absurdity, is the emergence of new ways of thinking about totalities, new ways of conceptualizing and reasoning about continuous wholes which synthesize the traditionally distinct notions of continuous and discrete wholes. This opposition had to be transcended in the production of any such synthesis.

It will, therefore be necessary to sketch the course of this synthesis and the emergence of new ways of conceptualizing totalities. This can be no more than an impressionistic sketch, for the history here is complex and technical (any mathematically and philosophically rigorous treatment would require many volumes). Those wanting more rigour and/or more detail are referred to the suggestions for further reading at the end of the book.

It should come as no surprise to find that mathematical, rather than philosophical, considerations are those which pose the real challenge to the classical finitist. It is important to locate this challenge more precisely than is frequently the case, for it is only in this way that we can come to see exactly what sort of sense is made of the actual infinite within mathematics and so to assess the wider implications of its mathematical use. It is customary to treat the invention of infinitesimal calculus as marking the occasion of the really significant intrusion of the actual infinite into mathematics. While it is true that the calculus was introduced (and perhaps could only have been introduced) in a philosophical climate of metaphysical acceptance of the infinite (a climate of rational theology), it is not the mere introduction of methods of differentiation and integration which dictates the move either to a point continuum or to an actual infinity.

The original introduction of the operations of differentiation and integration was geometrical. As geometrically grounded they can (as the later work of Weierstrass and others showed) make do with

traditional geometrical concepts of continuity, the potential infinity of points of division, potentially infinite sequences and the notion of a limit, even if when loosely described they appear to involve infinitesimal magnitudes and actually infinite division. Differentiation can be pictured as giving the gradient of the tangent to a curve $y = f(x)$ at a given point (x, y) by treating it as the gradient of the line from (x, y) to $(x + \delta, f(x + \delta))$ when δ is infinitely small (figure 4.1). (A condition for this to be defined is that one should get the same result by approaching from the right, i.e. by considering lines from $(x - \delta, f(x - \delta))$ to (x, y).)

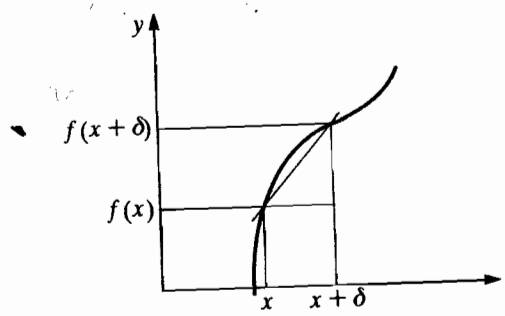

Figure 4.1

This would treat the gradient as the ratio of two infinitely small quantities $f(x + \delta) - f(x)$, and δ, but raises the awkward question of how an infinitely small quantity differs from 0, and of how one can divide by such a quantity and distinguish the result from division by any other infinitely small quantity. Paradoxical conclusions quickly follow as Berkeley pointed out in his criticism of Newton's use of calculus (Berkeley, 1734). These difficulties are avoided by treating the gradient of the tangent as the limit of an infinite sequence of ever closer approximations each of which is a ratio between finite lengths, i.e. as

$$\lim_{\delta \to 0} \frac{f(x + \delta) - f(x)}{\delta}$$

Similarly the definite integral can be pictured as giving the area under a given section of a curve $y = f(x)$. This area can be

approximated by chopping it up into rectangles and adding their areas together. The narrower these rectangles are, the better the approximation is. The areas might thus be thought of as the sum of infinitely many, infinitesimally thin rectangles which still somehow manage to have a non-zero area. But the area can also be defined as a limit approached by summing over successively thinner, but still finite, rectangles (figure 4.2).

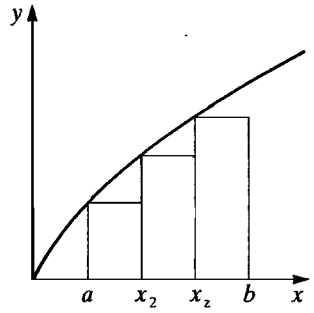

Figure 4.2

$$\int_a^b f(x) = \lim_{n\to\infty} \Sigma((x_1 - a)f(a) + (x_2 - x_1)f(x_1) + \ldots + (b - x_n)f(x_n))$$

The idea of a limit, the limit of an infinite sequence of ever closer approximations to a given quantity, thus supplants the infinitesimal and it would initially seem that such sequences need only be regarded as potentially infinite. For the idea of an infinite sequence of closer approximations to a given limit is already present in Zeno and the classical finitist could get away with potentially infinite sequences there. He can continue to do so in this case provided that he sticks to differentiation and integration as geometrically picturable and interpretable operations. For here the limits to which an approximation is sought are already geometrically defined independently of any sequences of approximation to them. The gradient of the tangent to the curve is geometrically given by the ratio AB/XB, i.e. this ratio exists, the only problem is to put a number on it. Similarly, any closed plane figure is presumed to have an area, even though its measurement

may be problematic. The infinite sequences involved are then just the familiar potentially infinite sequences associated with continued, ever finer division.

Restricted to this almost purely geometric form, differentiation and integration are merely modifications of existing geometrical techniques (Archimedes methods of exhaustion and of indivisibles, and various methods for constructing tangents to curves). What then, was the essential novelty? For there is no doubt that the methods of infinitesimal calculus were new and powerful, and moreover that they were perceived, both at the time of their introduction and since, as involving the infinite in mathematics in a way which did not occur with the geometer's recognition of the infinite divisibility of a continuous magnitude.

What is new is the fact that they form part of a calculus. It is the context of introduction which makes the difference. These things are not merely conceptualized as limits of infinite sequences of approximations, but are associated with methods of computing values of limits of such sequences. As such they are part of, and indeed central to, the motivation of the algebraization of geometry. It is the algebraic representation of the (potential) infinite already inherent in standard geometrical practice that gives it a new and problematic, because number-like, status. The algebraic representation gives us at least the appearance of being able to calculate with the infinite and with infinitesimals. If we write

$$\text{'lim'}_{n\to\infty}$$

for example, it is tempting to read '$n \to \infty$' in the same way as '$n \to 1000$' and thus as if ∞ were some value that n might actually attain, even though the geometrically guided use of the whole limit expression neither warrants nor requires this. Moreover, although the algebraic notation and its associated operations were initially introduced as linked with an intended geometrical interpretation, they soon take on a life of their own, going beyond what is geometrically representable or picturable. The question of what sort of sense is to be made of operations and expressions which have their origin solely in the algebraic, symbolic representation then becomes urgent and there is a pressure to give arithmetical, numerical interpretations primacy over geometrical interpretations and hence

for a rigorously and independently founded arithmetical–numerical representation of the continuum.

It was Descartes who first systematically introduced algebraic methods into geometry, insisting that the objects of geometry, geometrical figures and curves, should be defined not in the manner of Euclid, but by algebraic (polynomial) equations. He thought of such an equation as giving the law according to which a point would have to move in order to generate a curve. Thus for example a circle, centre (a, b) and radius r, is given by the equation

$$(x-a)^2+(y-b)^2=r^2$$

Its circumference is traced out by a point (x, y) which moves in such a way as always to satisfy this equation (figure 4.3). There are here several important moves away from the classical tradition of Euclid and/or Aristotle. They were not all initiated by Descartes but were first brought together by him.

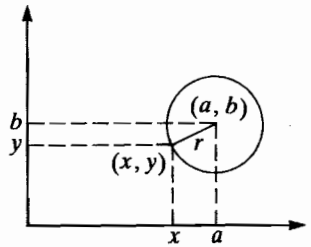

Figure 4.3

In the first place, Descartes is using variables, symbols such as 'x' and 'y' to stand for quantities which change. Their immediate interpretation is a geometric one; they stand for distances or lengths along a given axis from the point where they intersect (the origin). But Descartes also quite explicitly treats these lengths as themselves representing any other continuous magnitude which one might happen to be interested in. Thus change of temperature with time may be pictured geometrically by taking lengths on one axis as representing temperatures and on the other axis as representing times. Moreover the same goes for areas and volumes, these too

may be represented, as continuous magnitudes, by line lengths. This is a departure from earlier practice where it was presumed that if a and b are line lengths then $a \times b$ stands for an area. Because Descartes is prepared to allow that any continuous magnitude can be represented by a line length he can allow that there must be a line length to represent $a \times b$. In this way he makes line lengths closely resemble numbers in that multiplication and division are now operations defined over line lengths, whereas previously only addition and subtraction had seemed to make sense. This allows Descartes to make geometric sense of expressions such as x^5 which would otherwise have to have been thought of as the 'volume' of a five-dimensional cube. It is this step which is crucial to being able to regard a polynomial equation, such as

$$y=a_1x^5+a_2x^4+a_3x^3+a_4x^2+a_5x+a_6$$

as defining a curve in a two-dimensional space. Descartes thus assumes that all continuous magnitudes and all ratios between them (whether they are commensurable or not) can be represented by lengths. In this way the theory of ratios and proportions between continuous magnitudes was swiftly turned into an 'arithmetic' in which ratios are treated as numbers of a new kind. These ratios would include not only those between commensurable magnitudes, but also those between incommensurable magnitudes, i.e. ratios known not to be expressible as ratios between whole numbers, such as $\sqrt{2}$ and π.

Secondly, the focus of geometric attention is turned away from closed figures to continuous paths, whether forming closed curves or not and on their algebraic characterization – the characterization of a 'motion' which will generate the curve. This means that one is no longer dealing with continuous wholes as wholes which are bounded and limited and in this way given before their parts. Instead they are treated as generated wholes which may be potentially infinite but which are given by the algebraically expressed law constraining and determining their generation.

It is important that the early development of algebraic, analytic geometry was closely, and indeed almost inseparably, bound to the development of mathematically expressed theories of mechanics. It is this which means that the most common way of thinking of the

relation between a curve and the algebraic expression which defines it is by thinking of the expression as a law of a generating motion. Calculus itself was developed with a view to providing a quantitative treatment of change, and rates of change. It is thus in terms of motion that limits are interpreted and understood. Thus Newton wrote:

> Perhaps it may be objected that there is no ultimate proportion of evanescent quantities: because the proportion before the quantities have vanished, is not the ultimate: and when they are vanished, is none. But by the same argument, it may be alleged that a body arriving at a certain place, and there stopping, has no ultimate velocity; because the velocity before the body came to that place is not its ultimate velocity: when it has arrived it is none. But the answer is easy; for by the ultimate velocity is meant that with which the body is moved, neither before it arrives at its last place and the motion ceases nor after, but at the very instant it arrives: that is, that velocity with which the body arrives at its last place, and with which the motion ceases. And in like manner, by the ultimate ratio of evanescent quantities is to be understood that ratio not before they vanish, nor afterwards, but with which they vanish. ... There is a limit which the velocity at the end of the motion may obtain, but not exceed. This is the ultimate velocity. And there is the like limit in all quantities and proportions that begin and cease to be. And since such limits are certain and definite, to determine the same is a problem strictly geometrical. (Newton, 1934, pp. 38–9)

In this context differentiation is (a) defined as an algebraic operation, and (b) interpreted as giving the rate of change of one quantity (represented on the y-axis) with respect to another (represented on the x-axis) at a point. Thus if

$$y = ax^2 + bx + c, \qquad \frac{dy}{dx} = 2ax + b$$

and the 'rate' of change of y with respect to x does not have to be worked out for each point separately; it is given by a new equation.

Integration is also defined as an algebraic operation and as the inverse of differentiation; so

$$\int 2ax + b \, dx = ax^2 + bx + k$$

Moreover, their definition as algebraic operations allows for repeated applications (even though these may not always have any natural physical interpretation). If dy/dx gives the rate of change of y (position) with respect to x (time), i.e. velocity, then d^2y/dx^2 gives rate of change of velocity with respect to time, i.e. acceleration. This makes it possible to talk of and put values on instantaneous positions, velocities and accelerations whilst also having equations which characterize the ways in which they are changing. The algebraic characterization of both motions and geometric curves thus marks an enormous increase in descriptive power. An equation is a source of information about any point one chooses (and in this sense is an infinite description of all points) which is also a characterization of the whole which is not built up from information about points.

So besides the potentially infinite, discretely generated sequence of the natural numbers, there is now, in addition, the conception of a line which is continuously generated, in accordance with a law, and which does not involve constructing a point from those which preceded it, but merely ensuring that a constant, complex ratio, expressed algebraically, is preserved. Regarded in this way we can imagine the construction of a graph in the following way: a point can be construed as carrying out motions away from the origin in the direction of both the x- and y-axes simultaneously. Suppose that

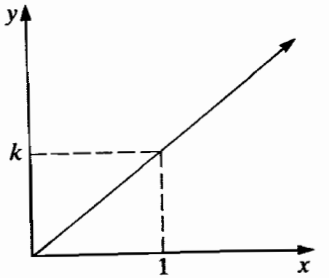

Figure 4.4

motion along the x-axis is uniform, continuous and has unit velocity. If motion in the direction of the y-axis is also uniform, continuous and has velocity k then the resulting graph will be a straight line whose gradient is k and the equation of the motion will be $y = kx$ (both motions being referred to the same equably flowing time). This is shown in figure 4.4.

Suppose now that motion in the y direction is initially v, but is uniformly decelerated (acceleration $-a$), then $x = t$ and $y = \frac{1}{2}[v + (v - at)]t$, i.e. $y = vx - \frac{1}{2}ax^2$, and what we get is the path of a projectile (a parabola) as shown in figure 4.5. The composition of the non-uniform motion with the uniform motion has the effect of 'bending up' the straight line which forms the x-axis.

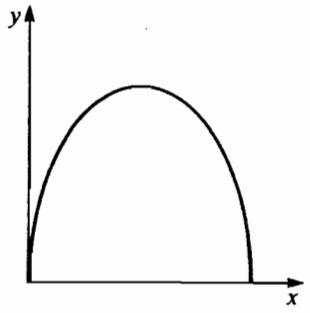

Figure 4.5

But just how non-uniform might the motion in the y-direction get? If $y = \sin x$ we have a wave function. And wave functions can themselves be superimposed on one another to give quite complex patterns – as when the ripples sent out by two or more pebbles meet. So we might get a motion which is 'oscillatory' in a complicated way. Daniel Bernoulli (1700–1782), when approaching the problem of how to write down an equation for the motion of a vibrating string said:

> My conclusion is that all sonorous bodies include an infinity of sounds with a corresponding infinity of regular vibrations.... Each kind multiplies an infinite number of times to accord to each interval between two nodes an infinite number of curves, such that each point starts and achieves at the same instant, these vibrations, while following the theory of Mr Taylor, each interval should assume the form of the companion of the cycloid extremely elongated. (Manheim, 1964, p. 41)

The equation given to reflect the superposition of this infinity of curves is

$$y = \alpha \sin\frac{\pi x}{a} + \beta \sin\frac{2\pi x}{a} + \gamma \sin\frac{3\pi x}{a} + \delta \sin\frac{4\pi x}{a} + \cdots$$

The idea that one could use such a superposition of 'wave' functions to represent, algebraically, a given curve over a given interval proved to be crucial both for the development of the concept of a function and of set theory.

At the time at which Bernoulli was writing, it was presumed that two functions which coincide over an interval will coincide everywhere and that any algebraic equation in which y is given as a function of x is geometrically representable, whilst not all geometrically drawable curves are algebraically representable. The initial problem was precisely that of finding analytic, algebraic expressions (laws) to characterize given curves or 'motions'. And there was disagreement between D'Alembert who equated a function with its algebraic expression, and Euler who identified a function with its graph. The efforts to generalize Bernoulli's results and to find the conditions under which an infinite trigonometric series actually represents a given function led eventually to the theory of point sets and provided the stimulus for Cantor's introduction of ordinal numbers.

Fourier gave a precise statement of the generalized problem: given an arbitrary function $f(x)$, find the coefficients a_n and b_n such that the equation

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

shall be an identity over a prescribed interval of the x-axis. The very statement of this problem marks a shift firmly in the direction of (a)

equating a function with its graph, for Fourier's arbitrary function means an arbitrarily drawn function, (b) treating algebraic expressions as capable of representing (piecewise if necessary, i.e. using different representing functions for different intervals) every drawable geometric curve.

In the course of investigation of this general problem posed by Fourier, a number of 'pathological' functions were discovered, functions which were algebraically expressed in terms of infinite sums but whose 'graph' is unpicturable. To see how these can arise we have to note that, given the indefinite divisibility of the continuum, there is no limit to the number of oscillations that can be packed into a given interval (their frequency), and given its unbounded nature there are no upper limits which can be placed on the amplitude of such oscillations. Functions which are expressed as complicated wave functions will, however, always be continuous (see figure 4.6). A discontinuous function is one which 'jumps' at one or more points (see figure 4.7).

If there is no limit to the frequency with which motion in the y direction might oscillate, then might it oscillate with an infinite frequency? That is, might it be the case that no matter how small an

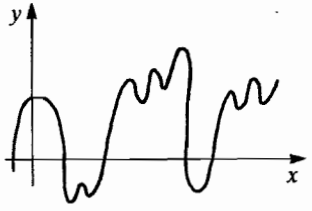

Figure 4.6

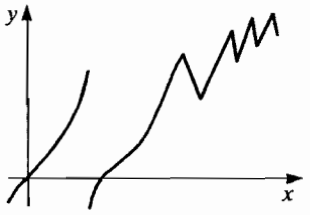

Figure 4.7

interval we take there will be an oscillation contained within it (i.e. the graph will have changed direction) in this interval? This is what would appear to be the case with Weierstrass's everywhere continuous but nowhere differentiable function

$$f(x) = \sum_{n=0}^{\infty} b^n \cos(a^n x)$$

where a is an odd integer greater than 1, b is a positive constant less than 1 and ab is greater than 1. It can be shown that for any point x_0 the difference quotients (giving 'gradients') approaching from the right and from the left have a different sign no matter how close one gets to x_0. So the function, though continuous, is not differentiable at any point. Riemann provided an example of a function which has infinitely many discontinuities between any two limits but which is none the less integrable.

$$f(x) = \frac{(x)}{1} + \frac{(2x)}{4} + \frac{(3x)}{9} + \ldots = \sum_{n=1}^{\infty} \frac{(nx)}{n^2}$$

where

(x) = the excess of x over the nearest integer
$(x) = 0$ if x is midway between two integers

so

$$-\tfrac{1}{2} < x < \tfrac{1}{2}$$

$f(x)$ is convergent for all values of x, but is discontinuous for all x of the form $p/2n$, where p and n are relatively prime. Thus $f(x)$ is discontinuous an infinite number of times in every arbitrarily small interval. But $f(x)$ is not too wild; the number of jumps larger than a given s is always finite. So it is possible to chop the regions into small enough bits so that within each bit the jumps are smaller than s, and so to get successive approximations of the 'area under' $f(x)$.

With such pathological functions we see the power of the algebraic symbolism and symbolic operations to outstrip geometric

intuition. These functions are not picturable and thus disrupt the assumption, based on picturable cases, that continuity, integrability and differentiability go together. But more than this, the infinitely dense packing of either continuous or discontinuous oscillations compels recognition of a complex structure in the apparently simple, homogeneous, equably smooth flowing linear continuum. Such functions have the power to introduce divisions in the continuum, not one at a time, but in an unpicturably infinite density all at once, as it were. It should be noted that this is done with functions which are themselves defined using limits of infinite series.

It is here that we have the ground of the undermining of the classical finitist position. Functions had been conceived in inseparable association with their graphs – the 'paths' traced by points moving in accordance with an algebraically expressed law. But when that law dictates a 'motion' which involves infinitely frequent oscillations, or infinitely frequent jumps, it is a path which can no longer be geometrically traced either in the mind's eye or on paper. But if the law can be written and by this means rationally investigated, the graph of the function must be presumed, in some sense, to exist and to be a totality of points over which our only hold is now algebraic. These points are the members of the set of values of a function $f(x)$ for each x considered as a numerical argument. Thus it becomes necessary to think of the original, smoothly continuous line as itself a set of points, each indexed by a number, and which has an unimaginably, because infinitely, complex order structure.

Sets so conceived are actually infinite totalities, given neither before nor after their points. Not before, because the points are no longer points of potential division successively generated and not after because the totality is not defined by reference to characteristics of points (for points are in themselves identical to one another). Indeed, if we are thinking geometrically, a function does not uniquely define a set of points in the plane, for the set of points which constitute its graph is dependent on the unit chosen. To take a very simple example, consider our parabola given by the function $y = ax - bx^2$. Under two different choices of unit it will define two different sets of points (figure 4.8). (The inner curve results from doubling the unit on both axes, i.e. is the graph of $y'' = ax'' - bx''^2$ with $x'' = 2x$, $y'' = 2y$.)

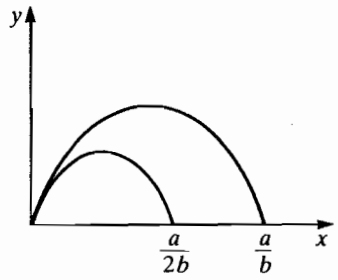

Figure 4.8

Geometrically, then, it is only relative to a unit, or measure, that a function can be correlated with a set of points. What the algebraic expression does is to replace the old geometric definitions of figures in the sense that these are definitions characterizing a certain kind of spatial structure or configuration, which can be realized on all sorts of different scales. When we think algebraically and think in terms of sets of numbers, rather than sets of points, then it is clear that the set of numerical values of a function is unique. The numbers here serve as a means of specifying a structure which, in some cases we have no other means of specifying, a structure which can be instantiated in many different ways depending on the way in which points, lengths or other continuous magnitudes are assigned numbers (measured). In the case of the function neither law nor graph takes precedence, for the graph, the path, is the geometrical/mathematical object of study, but it is given as such only via the function which defines it. Complex structures are the wholes with which the mathematician is concerned and which his algebraic notation also suggests are wholes composed of indivisible points – sets of points indexed by numbers. If the function is well defined (can be shown to have a unique value for every argument) then membership of the set of points constituting its graph (relative to a given measure) must be determinate. But since the path may be indefinitely long, the sets of points may be indeterminate in the sense of being unbounded. Generalizations over such sets of points are grounded in the function of which they are the graph, not in the characteristics of points. It is the structure of points, the relations between them, which are important.

The pathological functions finally showed the power of algebra to outstrip geometrical intuition. Geometry and motion could no longer be relied upon to provide the basis for analysis; it could not function as the background against which its operations were to be interpreted and tested. Analysis would have to become algebraically autonomous. This was the motivation behind arithmetization. There was a need to provide an account of continuity, differentiability and integrability in numerical terms, terms not drawing on geometrical intuition.

2 The Arithmetization of Analysis

But how was a purely arithmetic theory of limits to be constructed? Starting from the positive and negative integers it is possible to define positive and negative rational numbers (or fractions) as ratios between integers, and to give rules for their addition, subtraction, multiplication and division. But the rational numbers, although densely ordered (between any two rational numbers there is always a further rational number), do not form a continuum. There are more points on a line than can, after selection of a unit, be represented by rational numbers. This is the traditional problem of the existence of incommensurable magnitudes. Lines of, for example, length $\sqrt{2}$ can be geometrically constructed and proved not to be representable by any rational number. So geometrically the limit of a sequence of rational approximations to $\sqrt{2}$ is known to exist. But if geometrical intuition is to be dispensed with, then it cannot be presumed that the limit of such a sequence of rational numbers exists; rather such limits have to be introduced by means of definitions which do not presuppose their existence. From the purely arithmetic basis there are as yet only rational numbers and infinite sequences of rational numbers, some of which converge and some of which do not. The aim is to arrive at an arithmetically defined/constructed set of numbers which could adequately represent (by indexing) the points on a line.

The intuitive link between the conceptions 'point' and 'real number' is clear in Dedekind's way of defining real numbers. After noting that the straight line L is indefinitely richer in point individuals that the domain R of rational numbers is in number

individuals, because there are infinitely many points in the straight line that correspond to no rational number, he says:

> If now, as is our desire, we try to follow up arithmetically all phenomena in the straight line, the domain of rational numbers is insufficient and it becomes absolutely necessary that the instrument R constructed by the creation of the rational numbers be essentially improved by the generation of new numbers such that the domain of numbers shall gain the same completeness, or as we may say at once, the same *continuity*, as the straight line. (Dedekind, 1963, p. 9)

This led him to ask 'In what does this continuity consist?' His answer is that the essence of continuity lies in the following principle:

> If all the 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, 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, 1963, p. 11)

The 'discontinuity' of the rational number sequence is thus seen to lie in the fact that not all cuts in it are produced by rational numbers, where a cut in the rational numbers R is

> any separation of the system R into two classes A_1, A_2 which possesses only *this* characteristic property that every number a_1 in A_1 in less than every number a_2 in A_2. (Dedekind, 1963, pp. 12–13)

Thus if A_1 were the set of all rational numbers less than $\frac{1}{2}$ and A_2 were the set of all rational numbers greater than or equal to $\frac{1}{2}$ this would be a cut produced by the rational number $\frac{1}{2}$. Whereas if A_1 is the set of rationals less than $\sqrt{2}$ and A_2 is the set of rationals greater than $\sqrt{2}$ this would be a cut (every rational number lies on one side or other of $\sqrt{2}$) but it is not produced by a rational number.

In order to create a continuous numerical domain one needs a number system in which all cuts are produced by, or correspond to,

numbers of that system. This is achieved by saying that whenever a cut (A_1, A_2) is not produced by any rational number, 'we create a new, an *irrational* number α, which we regard as completely defined by 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* always and only when they correspond to essentially different cuts. (Dedekind, 1963, p. 15)

Moreover, the extended domain of numbers created in this way proves to be closed under the operation of forming cuts. i.e. if one considers forming cuts in the sequence of real numbers, there will always be a real number which produces it. Thus, in this sense, the domain of real numbers can be said to be continuous. In addition, in order to justify his claim to have defined new *numbers*, Dedekind had to show that 'cuts' in the sequence of rational numbers can be added, subtracted, multiplied and divided in such a way that when the cut is produced by a rational number these operations reduce to the familiar operations on rational numbers.

The way in which Dedekind 'creates' his real numbers draws on the combinatorial approach to sets considered in chapter 3. His cuts are amongst the possible selections from the set of rational numbers and the totality of cuts is the totality of arbitrary selections which meet the defining condition for being a cut. It is thus a very non-constructive approach in that it treats cuts as pairs of sets without considering how these sets may themselves be specified. He has to presume that the totality of rational numbers is a fixed, actually infinite totality with a determinate membership; moreover it is a densely but linearly ordered totality (between any two rational numbers there exists another rational number and given any two rational numbers, a, b, $a = b$, or $a < b$, or $a > b$).

If the sets of rational numbers constituting a cut were not thought to have a determinate membership, then the cut itself would not be precisely defined. This approach is non-constructive not only in its use of actually infinite totalities, but also in that the cuts are presumed to exist independently of sequences of approximation or of any means of defining the sets which constitute them. They are

introduced by strict analogy with the geometric potential divisibility of the continuum, but once divorced from the geometric analogy, as in all strictness they are supposed to be, they have been cut off from any operation which might 'produce' the cuts and which gives a hold on the notion of division as a potential. The assumption of the existence of limits is replaced by the assumption of the existence of arbitrary selections from a given set.

The alternative route to the creation of real numbers, adopted by Cantor, is in some respects more constructive in its approach. The set of rational numbers R, considered as a representation of the points on a line, is incomplete in the sense that there are lengths, and hence points of division, which can be given a rational approximation of any desired degree of accuracy but can receive no exact numerical representation, i.e. there are infinite convergent sequences of rational numbers which have no rational limit. A sequence $\langle a_n \rangle$ of rational numbers is a *fundamental* sequence if for any positive rational ε there is an integer k such that, $|a_{n+m} - a_n| < \varepsilon$ for any m and all $n > k$. Cantor identified the new (real) numbers with these sequences and showed that one could define an order on them by saying that if $\langle a_n \rangle = b$, $\langle a_n'' \rangle = b''$, and if for any positive rational ε there is an integer k such that for all $n > k$

$$\begin{aligned} |a_n - a_n''| < \varepsilon \quad &\text{then} \quad b = b'' \\ a_n - a_n'' > \varepsilon \quad &\text{then} \quad b < b'' \\ a_n - a_n'' > -\varepsilon \quad &\text{then} \quad b < b'' \end{aligned}$$

So two such numbers are equal if the sequences which define them vary by less than any given ε after a finite distance. It follows that given any rational number a, the constant sequence $\langle a \rangle$ (whose limit is a) is such that either $b = a$, or $b < a$, or $b > a$, for any real number b. Arithmetical operations can then be defined for the new numbers

$$\begin{aligned} b + b' = b'' \quad &\text{means that } \lim_{n \to \infty}(a_n + a_n' - a_n'') = 0 \\ b \cdot b' = b'' \quad &\text{means that } \lim_{n \to \infty}(a_n' a_n' - a_n'') = 0 \end{aligned}$$

So the bs now look and behave like numbers and they incorporate within them a model of the rational numbers. Cantor went on to iterate this process:

C is the set of numbers expressed as fundamental sequences of members of B

.
.
.

L is the set of numbers expressed by fundamental sequences of members of K

.
.
.

Now the relation between A and B is different from that between B and any subsequent domain reached by this style of definition in that although every a belonging to A is represented in B, there are elements in B which have no counterpart in A. But for C it can be shown that every element of C already has a counterpart in B (and B can be shown to be isomorphic to Dedekind's real numbers). Thus in one sense no new elements are created, however many times the process is reiterated. Yet as sequences the members of C are quite distinct from members of B, they are sequences of sequences of rationals.

3 Toward Infinite Ordinal Numbers

After selection of a unit of measurement and an origin it can be shown that every point on a line can be indexed by a unique element of B, and Cantor postulated that to every element of B there corresponds a unique point on the line which it represents. So in B we would already seem to have an arithmetical model of the continuum. Why, then, did Cantor go on to iterate the procedure and think it important to distinguish between the members of B and subsequently formed 'numbers' even though B already contains the limits of all fundamental sequences of elements of B? This relates to the motives for wanting an arithmetical representation of the continuum in the first place. It is to try to achieve some representation and understanding of the complexity of its point structure as revealed, for example, by the pathological functions, and to produce a definitive and general answer to Fourier's representation problem. Dedekind's method of introducing real numbers reflects only the basic geometrical intuition of arbitrary divisibility. Cantor

is concerned to exhibit the complex fine structure – the order structure – of points on a line. It is to assist with this that he distinguishes between points in terms of the types of series of which they are the limit. He considers a set P of points on the line (considered as indexed by numbers) and gives the following definition.

> By a *limit point of a point set P* I mean a point of the line for which in any neighbourhood of the same, infinitely many points of P are found, whereby it can happen that the (limit) point itself also belongs to the set. By a 'neighbourhood' of a point is understood an interval which contains the point in its interior. Accordingly it is easy to prove that a point set consisting of an infinite number of points always has at least one limit point. (Dauben, 1979, p. 41)

Given any point set P and any other point on the line, it either is or is not a limit point of P. Thus each P has a well-defined set $P^{(1)}$ of limit points, the first derived set of P. If $P = R$ (the set of rationals), $P^{(1)} = B$, the set of all real numbers expressed as limits of fundamental sequences of rationals. But P might be any infinite set, so $P^{(1)}$ might not be B. If $P^{(1)}$ is infinite the operation can be repeated. Either there is some finite n such that $P^{(n)}$ is finite and hence $P^{(n+1)}$ does not exist, or there is not. In the case of R, since iteration always leads to the full set of points of the line, there is clearly no such n. This suggests that if there are sets of points for which the iteration stops at some n, they are not fully continuous although they have a certain non-uniform kind of density. A point p which is expressed as a limit of sequences of sequences of rational numbers is the derived set of a set of points each of which is the limit of a fundamental sequence of rationals.

$$P^{(1)} \quad \{p\} = P^{(2)} \qquad P^{(2)} = \{\lim_{n\to\infty} \langle b_n \rangle\}$$

$P^{(1)}$ is then a set of points which cluster round p. But each of these is itself expressible as a limit of a sequence of rationals. So P is a set of rationals which cluster round p with a particular type of dense ordering. The n, if it exists, for which $P^{(n)}$ becomes finite is

then a sort of measure of the 'density' of clustering of rationals in P round a finite number of points. P is not an evenly distributed dense set, but is locally dense in a finite number of places.

But Cantor did not stop here. For if $P^{(n)}$ is not empty for any finite n, then clearly, even if the operation were iterated an infinite number of times, there would still be a set, and Cantor introduced P^{∞} to indicate this and $P^{\infty+1}$ to indicate the derived set of P^{∞}. ∞ and $\infty + 1$ are new infinite, ordinal numbers and they make their first appearance as indices of the iterations of an operation designed to characterize the structural characteristics of sets within a point continuum and hence the possible intricacies of behaviour of functions and their representation by trigonometric series.

The transfinite ordinal numbers thus first come into being as a way of indexing iterations of the operation of forming the derived set of a set of points. They appear to be required by the attempt to characterize the distribution of points in a continuum. As initially introduced, they do not, and were not intended to put a number on the points in the continuum, although this was the question which later came to preoccupy Cantor. Before seeing how this question arises, however, it is worth considering where the developments described above leave the finitist.

4 Conflict with Classical Finitism

The crucial question for the classical finitist is whether the fundamental sequences of rationals, in terms of which the real numbers are introduced, can be considered as potentially infinite sequences, and indeed whether the rational numbers themselves can be considered as a merely potentially infinite totality. Initially the answer in both cases might be thought to be 'Yes', but a closer inspection of what is required to produce an arithmetic point continuum, which can dispense with geometrical intuition whilst at the same time preserving results founded upon it, suggests that this is not the case. The reason for this is that a class cannot be regarded as potentially infinite unless it is thought of as generated by a non-terminating process. So the possible potentially infinite sequences of rational numbers are those which can be generated in some way, either by a law or by a sequence of choices, whether free or constrained. Any such sequence is only ever given as a finite

fragment plus a generating process, with the consequence that such a sequence is not uniquely given by its (actual) terms, any more than a species, or a subset, of the natural numbers is given by its members.

Thus the sense in which such a sequence may be said to have a limit must be given by the condition for convergence. A sequence $\langle p_1 \rangle$ of rational numbers is convergent if and only if, for any positive rational number δ, there is a positive integer n such that the absolute value of the difference between p_n and any subsequent term of the sequence is less than δ. To say when a sequence converges is to say what it is for a sequence to have a limit. The limit is not something which is independently given and to which the terms of the sequence can be thought to approximate ever more closely. This means that any extensional statement about such a sequence (i.e. one which is true of any other sequence which is extensionally equivalent to the given sequence) has to be true or false on the basis of a finite amount of information about the sequence. For example, two convergent sequences are said to converge to the same limit if and only if there is some point in the sequences after which the difference between corresponding terms becomes arbitrarily small. If there is no law generating the sequences (they are free choice sequences) then it will not be possible to say that they do or do not converge to the same limit, for all we will have to go on will be finite initial segments of the sequences, segments which give no assurance about how they will continue. Since there are indefinitely many sequences with the same finite initial segment this means that any statement we can make about a lawless sequence (on the basis of knowledge of some finite initial segment of it) will not be true of just that sequence but of indefinitely many – all those which share the particular initial segment.

The totality of all potentially infinite sequences of rationals, or even of all convergent or fundamental sequences of rationals, is thus like the set of all subsets of the natural numbers, in that it is a potentially infinite whole containing members which are never fully determinate, for they themselves are potentially infinite, and always actually finite but necessarily incomplete. This does not yield a point continuum in anything like the sense presumed by Cantor, for its members do not have the precise identity required of points.

When this approach to real numbers is interpreted geometrically we get precisely identifiable points of division corresponding to the rational numbers together with the possibility of focusing on an ever smaller region in order to locate a 'point' indexed by a non-rational number, a possibility which is guaranteed by the dense ordering of the rationals (between any two there is another). But since the process of focusing is not, and never can be, complete, the continuum is never resolved into points, only into ever smaller regions.

For example, consider the set of real numbers between 0 and 1 as given by their binary decimal notation. All such sequences can (as in figure 3.3) be represented by the full binary tree which can be thought of both as the making of successive choices about whether to put 0 or 1 in the nth decimal place and as the making of successive divisions of the unit line. All the points corresponding to an infinite sequence with an initial segment .0001 ... lie in the interval $[\frac{1}{16}, \frac{1}{8}]$. Each addition of a level to the tree chops the line up into smaller bits (figure 4.9).

It is, however, possible to develop a theory of 'real numbers' and functions over them on this basis (see for example Bishop, 1967; Troelstra, 1977). The result is intuitionist, rather than classical, analysis and they are by no means equivalent. But what is important from the point of view of our present investigations is the fact that both approaches can be pursued. This means that the classical finitist does not have to back down. On the other hand, his position has lost much of its plausibility. With the development of the classical theory of real numbers on the foundations suggested by Cantor and Dedekind it has become possible, without obvious

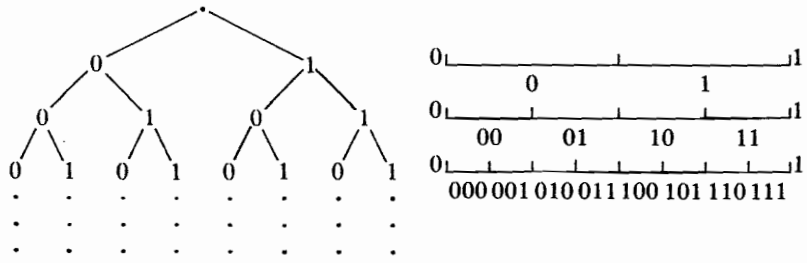

Figure 4.9

incoherency, to think of a continuum as 'made up' of points (strictly as a set of points). Moreover, this view seemed to be required by the way in which functions defined over real numbers are customarily associated with their 'graphs'.

The arguments derived from Zeno's paradoxes against taking a continuum to be made up out of points are circumvented because the 'construction' of the real numbers, which is the initial vehicle for thinking about a point continuum, is not a geometrical construction. It is not a matter of building a continuum by distributing points in space, but of defining the real numbers and showing that these can be ordered so as to be order isomorphic to the points on a line. This means that a one–one correspondence can be established between the real numbers in the interval (0, 1) and any finite line, and between the positive (or positive and negative) real numbers and any infinite line.

It is with the pathological functions that the continuum is actually infinitely divided in a way which is not geometrically picturable. A function which has infinitely many discontinuities between any two limits effects a divorce between the introduction of discontinuities (divisions) and the idea of a successive process. It thus does not support a conception of potentially infinite division, but of actually infinite division. It is the geometrical origin of the notion of a function that suggests that to any well-defined function there corresponds a 'graph' as something which exists as a determinate object of mathematical investigation. It is also the geometric interpretation which is heuristically important in very many of the applications of analysis. But it is the algebraic expression of a function that makes it natural to think of this determinate object as a geometrically determinate, actually infinite set of points.

So the purely arithmetical development of the theory of real numbers (defined as equivalence classes of convergent infinite sequences of rationals) and of functions defined over them does not automatically support a theory of real numbers which can validate the conception of an actually existing infinite point continuum. This is because the infinite sequences of rationals can be treated as either actually or potentially infinite. But when the definition of real numbers as the limits of convergent sequences of rational numbers is given in the context of the intended geometrical interpretation it appears (a) that the actual infinite must be accepted if the real