Natural and Formal Infinities[†]

David Tall

Mathematics Education Research Centre
Institute of Education
University of Warwick
Coventry CV4 7AL
e-mail: David.Tall@warwick.ac.uk

*Concepts of infinity usually arise by reflecting on finite experiences and imagining them extended to the infinite. This paper will refer to such personal conceptions as **natural infinities**. Research has shown that individuals' natural conceptions of infinity are 'labile and self-contradictory' (Fischbein et al., 1979, p. 31). The formal approach to mathematics in the twentieth century attempted to rationalize these inconsistencies by selecting a finite list of specific properties (or axioms) from which the conception of a **formal infinity** is built by formal deduction. By beginning with different properties of finite numbers, such as **counting**, **ordering** or **arithmetic**, different formal systems may be developed. Counting and ordering lead to **cardinal** and **ordinal** number theory and the properties of arithmetic lead to ordered fields that may contain **infinite** and **infinitesimal** quantities. Cardinal and ordinal numbers can be added and multiplied but not divided or subtracted. The operations of cardinals are commutative, but the operations of ordinals are not. Meanwhile an ordered field has a full system of arithmetic in which the reciprocals of infinite elements are infinitesimals. Thus, while natural concepts of infinity may contain built-in contradictions, there are several different kinds of formal infinity, each with its own coherent properties, yet each system having properties that differ from the others.*

*The construction of both natural and formal infinities are products of human thought and so may be considered in terms of 'embodied cognition' (Lakoff and Nunez, 2000). The viewpoint forwarded here, however, is that formal deduction focuses as far as possible on formal logic in preference to perceptual imagery, developing a network of formal properties that do not depend on specific embodiments. Indeed, I shall show that formal theory can lead to **structure theorems**, whose formal properties may then be re-interpreted as a more subtle form of embodied imagery. Not only can natural embodied theory inspire theorems to be proved formally, but formal theory can also feed back into human embodiment, now subtly enhanced by the underlying network of formal relationships.*

Introduction

After many centuries conceptualising various notions of infinity, and encountering puzzling conflicts, such as the Zeno paradoxes and logical difficulties of infinitesimals, matters came to a head at the end of the nineteenth century. Cantor (1895) developed a theory of infinite cardinals in which division was not possible, and he denied the theoretical existence of infinitesimals. Cantor's theory, however, fell foul of the wrath of Kronecker who accused him of claiming that certain mathematical objects (such as cardinal infinity) existed without showing how they could be constructed. Kronecker's view blossomed into the school of intuitionism led by Brouwer in the early twentieth century.

Two other major schools of thought grew at around the same time, each of which began approaches that encompassed formal definitions of infinite concepts. Hilbert (1900), in his celebrated address at the International Congress of Mathematicians, announced his list of the major problems for twentieth century mathematics. He also endorsed the *formal* approach to mathematics using the axiomatic method. This involves specifying a finite list of axioms to act as a foundation for a particular theory. He envisaged that such a strategy could lead to a total axiomatisation of mathematics, free from logical blemish. What mattered to him was not what concepts such as infinite cardinals *are*, but the consistency of the axiomatic theory that supports them. The third school of thought—the logicist approach led by Bertrand Russell, Gottlieb Frege and others—desired to define mathematical concepts purely in terms of logic[1].

All these theories encountered difficulties. The intuitionist school purposefully refused to use proof by contradiction, and this denied them a major tool used widely in twentieth century mathematics. Bertrand Russell found inconsistencies in aspects of Cantor's set theory and suggested technical devices to overcome them. His pristine approach to logic was much admired, but pragmatic mathematicians carried on posing and solving research problems in ways that suited their problem context.

Gödel (1931) dealt what was apparently a mortal blow to both the logicist and formalist viewpoints. He proved that any axiomatic system that includes the natural numbers must have properties whose truth or falsehood cannot be deduced from the axioms of the system[2].

Hilbert's formalist programme, therefore, cannot be completed in its entirety. In particular, any theory of infinite sets will include theorems that may be true but cannot be proved using only the techniques available in that system. Of course, one may add more axioms to the given theory (for instance, adding the axiom of choice to the limitations imposed by intuitionists). Such a tactic may prove a desired theorem, but the new axiom(s) allow the formulation of even more theorems that cannot be proved in the extended system. This is the Achilles' heel of the formalist approach. Essentially an axiomatic system that attempts to talk about infinite structures has 'too many' properties to be able to prove them all by finite proofs using the available techniques.

This weakness did not destroy the desire to have an accepted axiomatic foundation for the many theories of modern mathematics. Mathematicians found that they can still prove a vast number of useful theorems in their axiomatic systems, even though some intractable problems may remain.

To explain ideas of infinity in a manner appropriate for a wide range of readers, the language used here will initially involve everyday colloquial meanings for a concept such as 'set', and whether that set is 'finite' or 'infinite'. In everyday terms, a set is finite if we can – in principle – count it in the usual way, otherwise it is infinite. When mathematics is introduced to undergraduates, a *formal concept* – such as a 'function', 'group', 'topological space', 'cardinal number', 'infinitesimal', or whatever – is given in terms of a definition (or a list of axioms) in words and symbols, from which the properties of the given concept are deduced by mathematical proof.

Our exposition begins by considering the various ways in which individuals construct formal systems. Our quest for formal notions of infinity starts with the Peano postulates as a formal definition for the natural numbers, and outlines a formal development of number concepts. This extends to the formal concepts of infinite cardinals and ordinals on the one hand and the very different construction of infinitesimals and their infinite reciprocals on the other. We then show how formal ideas of infinitesimals can be re-interpreted in a visual manner as points on a geometric number line. This reveals how an embodied conception may

be founded on a formal construction, complementing the more usual direction in which natural embodied conceptions are used to suggest axioms and proofs in a formal theory.

The cognitive construction of formal systems

The human mind is finite. By thinking about the possibility of performing a process again and again, we can easily reach out towards the potentially infinite. The counting sequence, 1, 2, 3, … is a potentially infinite collection. Given any whole number, there is a next one. The decimal notation allows us to represent whole numbers of arbitrarily large size. As a consequence, we humans can easily imagine a *potential infinity* as a process that goes on without end. But how do we think of the whole system *in total*, including all the numbers, all at one time? The answer is, of course, we don't.

The human brain is not only finite, it copes with complexity by using a small conscious focus of attention, alighting on important information to assess a situation and take appropriate action. We therefore do not consciously focus on *all* the information at a given time, but move between different aspects, connected together mentally in a variety of ways.

For a given concept, we develop a whole *concept image* in the brain consisting of 'the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes,' (Tall & Vinner, 1981, p.152). This concept image grows and changes with experience and reflection. Such a biological entity cannot be totally coherent, for the various parts of the concept image develop at different times and in different ways; connections used on one occasion may be different from those evoked on another. Our concept images are full of partial experiences, focusing on a few aspects of a situation, linked together by various associations. What we attempt to do in mathematics is to rationalize the various experiences to build up as coherent an image as possible.

My colleague, Eddie Gray, observed James, aged six, in a group of children asked to type a sum with answer 8 into a graphic calculator. Whilst his friends typed in problems like 7+1, 4+4, 10–2, James typed:

1000000 – 999992.

How did he *know* that the answer to this problem is 8? He certainly did not do it by counting. James has sensed that there are patterns for the 'next number'. After 9 comes 10, 11, 12, … after 19 comes 20, 21, 22, … after 99 comes 100, 101, 102, … after 999 comes 1000, …, and after 999,999 comes 1,000,000. He knows that 1,000 is 'a thousand' and 1,000,000 is 'a million' and he can type 'a million' into the calculator as 1000000. He has never *counted* a million. But his ideas of place value allow him to give it a meaning, and know that 999992+8 is 'a million' just as 2+8 is 10 and 992+8 is 1000. His brain does not contain 'all the numbers', but it does contain many sophisticated connections that give him strategies for giving meaning to any numbers that he cares to consider. It is from such a complex concept image that individuals attempt to build a shared system to be communicated in gestures, pictures, words and symbols, in ways that are mutually agreed to be appropriate.

Towards formal theories

For the young child, the construction of knowledge involves building on perceptions of the world using the growing sophistication of the concept image. The more sophisticated thinker notices properties of structures and relationships between them. Formal thinking begins when selected properties are isolated and used as concept definitions from which other properties may be deduced by mathematical proof. Pinto (1998), building on the natural/alien distinction of Duffin and Simpson (1993), identified two fundamentally different approaches to constructing axiomatic systems. A *natural* approach builds on concept imagery to give a personal meaning to the formal definition. This means constructing examples of the definition that are sufficiently generative to be used as a basis for thought experiments to imagine possible theorems and possible strategies for proving them. A *formal* approach, on the other hand, focuses essentially on the definitions, using formal deductions to build theorems in a manner that attempts to avoid any appeal to intuition. (Figure 1.)

Most mathematicians use both natural and formal thinking processes, whichever is appropriate in a given context. For instance, it is natural to use a variety of conceptual ideas to *build* ideas for formal theories, and then to use a *formal* approach to prove theorems. Some

mathematicians tend more towards a natural approach (particularly in subjects such as geometry and topology where mental pictures suggest theorems to be proved); others prefer an overall formal approach (for instance, in abstract symbolic theories). Nevertheless, it is a common goal to produce written formal proofs of theorems, whether constructed formally or underpinned by natural imagery.
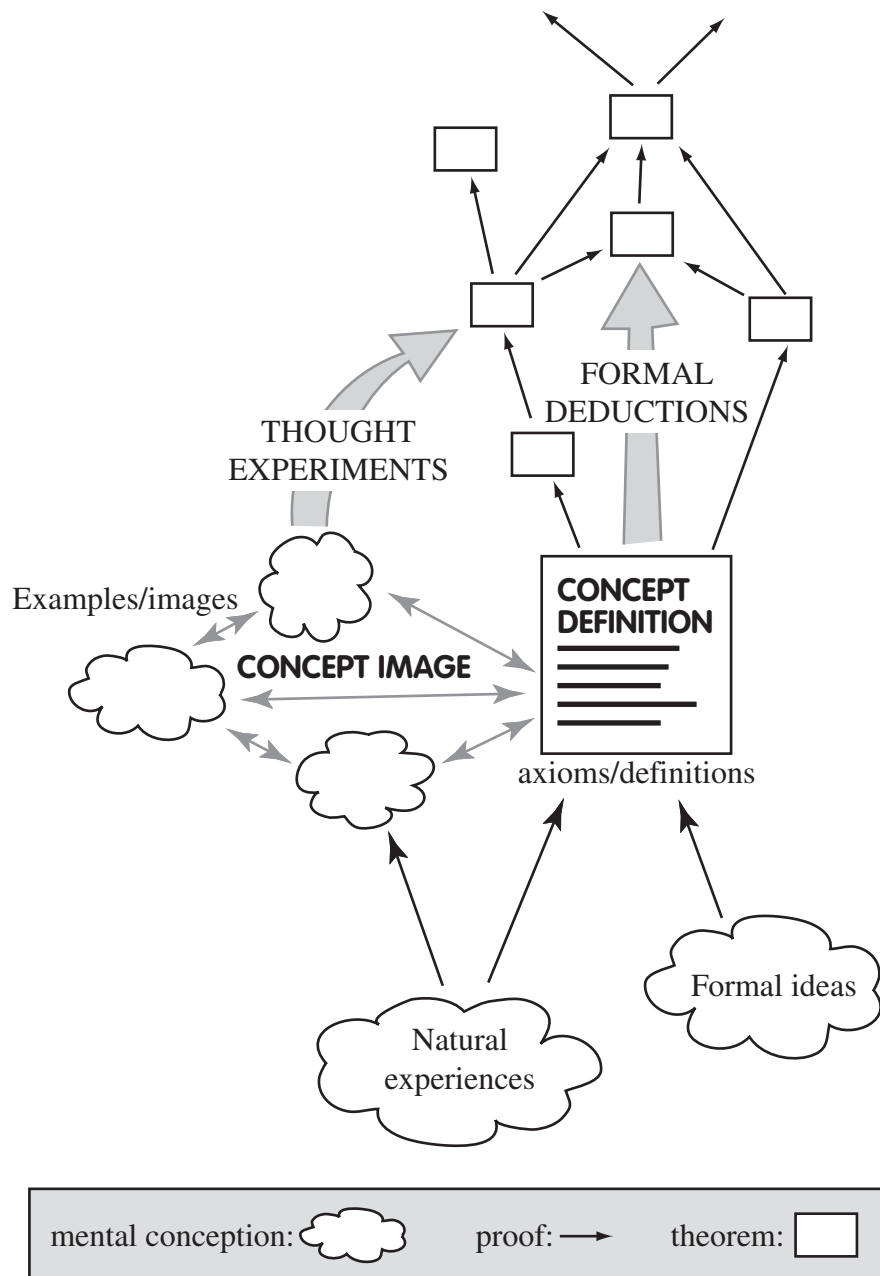
Figure 1: Some constituents in constructing a formal theory

Pinto found that the students she studied often preferred one approach or the other and that both approaches were capable of producing success in developing formal written theories

(Pinto, 1998; Pinto & Tall, 1999, 2001). However, many students failed to use the formal concept definition as a basis of deduction, instead preferring to build on their own personal version of the definition, sometimes inadequate, sometimes distorted, with the result that there was a broad spectrum of success and failure.

Before meeting axiomatic theories, the individual will already have some kind of *informal* (concept) *image*. Research shows that such concept images often persist long after formal ideas are introduced (Fischbein *et al*., 1979). Personal informal images may involve essentially contradictory features. For instance, informal experiences invariably suggest that 'the whole is greater than the part', so that the whole set N of all natural numbers $\{1, 2, 3, …\}$ is clearly greater than the part consisting of the set E of all even numbers $\{2, 4, 6, …\}$. Yet in the formal theory of cardinal numbers, two sets are defined to have 'the same cardinal number' if there is a one-to-one correspondence between them. The relation $n \leftrightarrow 2n$ gives a one-to-one correspondence between N and E and so, in this case, the whole and the part have the *same* cardinal number.

Some natural learners find this experience alien and reject the theory of cardinal numbers, as did Kronecker in the early history of the subject. Some keep the informal ideas in a separate compartment, a strategy common to those formal learners who focus only on the formal theory without any desire to link it to informal knowledge. However, a powerful way forward for both natural and formal thinkers is to attempt to reconstruct their knowledge to resolve the conflict between old experiences and new evidence. For instance, the old experience that 'the whole is greater than the part' is still true for finite sets. The new evidence that this is false for the natural numbers may be turned to advantage; in a formal theory, we can *define* a set to be 'infinite' if and only if it can be put in one-to-one correspondence with a proper subset.

Given a list of axioms to specify a formal concept, we define the *formal image* to consist of that part of the concept image that is formally deduced from the axioms. Since both formal and informal image are contained in the same biological brain, interconnections are bound to occur and confusions are likely to arise. However, the links may also be used to positive advantage by relating the formalism to appropriate imagery:

> Who does not always use along with the double inequality $a > b > c$ the picture of three points following one another on a straight line as the geometrical picture of the idea 'between'? (Hilbert, 1902, p.442)

In this sense, embodied meanings are at the root of mathematics. However, a formal system may have more than one particular embodiment. In a formal system, only a small number of explicit properties are selected as the basis for deduction. Any deductions that are made, depending only on these initial axioms, must also apply to any other situation where the axioms happen to be satisfied. If a system satisfies the axioms of a vector space, it does not matter what the elements of the system are. They could be vectors in a plane, in n-dimensional space, solutions to a system of differential equations, a particular collection of functions. All that matters is that the structure involved satisfies the axioms for a vector space, with the immediate implication that it also satisfies all the properties deduced from the definition of a vector space.

Fully formal expositions of a mathematical theory have been proposed by various mathematicians and logicians, (eg Russell & Whitehead, 1910–1913). In this sense, it would be possible to speak philosophically of a 'formal concept' that arises purely within such a formal theory. However, such a formal concept, as conceived by a human being, must arise as an essential part of that individual's biological and cognitive development. In practice, it is therefore difficult—even impossible—to base one's conceptualisation *solely* on formal deduction—as Hilbert, the founder of the Formalist school was fully aware. In my use of the term *formal concept*, I therefore also acknowledge the existence of an underlying cognitive structure supporting that concept.

Formal structure theorems and embodied representations

Even though a formal theory is designed to prove results by formal deduction, the resulting theorems may then provide the basis for new visual images. Some theorems may be classified as *structure theorems* in that they prove that the defined concept must have a certain explicit structure. Structure theorems are essential building blocks for the development of a theory, not only because they stand as important staging posts in the formal development, but also

because they may allow the individual once more to give the formal concept some kind of underlying embodied representation. (Figure 2.)
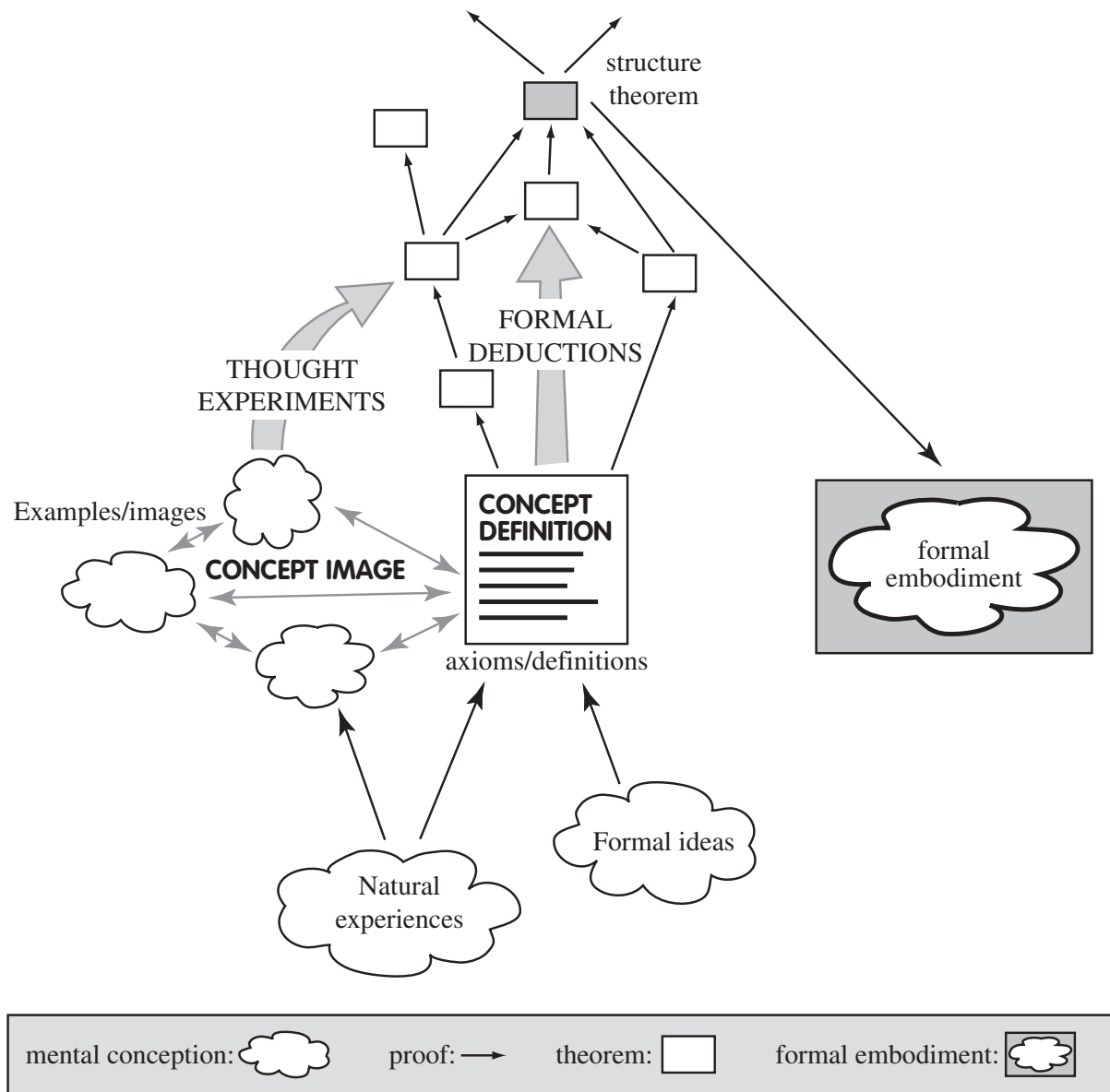

Figure 2: Building new formal embodiments from a formal theory

For instance, the *Peano postulates* are three carefully crafted properties used to formulate the definition of the system of the natural numbers. A powerful structure theorem shows that any system satisfying the Peano postulates is essentially the same as (isomorphic to) the familiar natural numbers 1, 2, 3, … Hence the formal theory of the Peano postulates gives a structure theorem that allows us to visualize the axiomatic system once more as the numbers on a number line.

These ideas resonate with the recent development of 'embodied cognition' in the form described by Lakoff and his colleagues (Lakoff and Johnson, 1999; Lakoff and Nunez, 2000). In particular, Lakoff and Nunez (2000) give a compelling and insightful account of 'where mathematics comes from', in terms of bodily experience and physical perception, supported by the devices of language. In doing so, however, they are closer to the natural part of the spectrum described here than the formal approach distilled through the abstract thoughts of axiomatic thinking. It is my contention that their theoretical position is better at describing 'where mathematics comes from' rather than 'where mathematics goes to'.

Natural learning explicitly grows from the individual's perceptions, actions and reflections, and grows in rigour by transmuting embodied thoughts into formally presented mathematics, often linked intimately to its natural source. Formal learning also occurs within a human brain. However, formalists and logicists struggle to divest themselves of misleading concept imagery, to focus as closely as is humanly possible on the precise interpretations of concept definitions and logical deductions. Although vast tracts of the brain are engaged in perception and construction of imagery, there are also huge areas of the cortex that are plastic and useable for a variety of activities including the many processes involved in thinking mathematically. All mathematical thought of necessity occurs in the human brain, but this does not mean that all thought is related to embodied perception.

As an example, Lakoff (1987) suggests that the 'container metaphor' (that A contained in B and B contained in C automatically gives A is contained in C) underlies the use of logical implication in mathematics. Formal thinkers are aware of naïve uses of implication, but they explicitly seek to manipulate symbols according to specified rules. An axiom such as '$a < b$ and $b < c$ implies $a < c$' can readily be interpreted as relationships between three points in order on a line, as Hilbert observed. However, the identical form '$a \sim b$ and $b \sim c$ implies $a \sim c$' is also used as part of the definition of an equivalence relation, which has an entirely different underlying concept image.

There is physical evidence that different brain areas are used for natural and formal thought. Houdé *et al.* (2000) obtained brain scans of individuals who were asked to solve

logic problems that required them to inhibit perceptual concepts to activate a logical reasoning process. The pictures of the changes in brain areas activated are impressive.

> The main finding is a striking shift in the cortical anatomy of reasoning from the posterior part of the brain (the ventral and dorsal pathways) to a left-pre-frontal network including the middle frontal gyrus, Broca's area, the anterior insula, and the pre-SMA. This result indicates that such brain shifting is an essential element for human access to logical thinking.
>
> (Houdé *et al*. 2000, p. 721)

Thus, there is a genuine physical shift in thought processes to use areas of the cortex that are not so directly related to embodied perception. Not only does formal thought attempt to divest itself of informal imagery, the formal thinker is aware of the danger of endowing a given list of axioms with a particular embodied image, specifically because this may limit thought processes to that particular context. However, having once proved a theorem by formal deduction, it may then be opportune to interpret the results of the theorem in a visual embodied manner which provides more sophisticated images available for thought experiments. It may have subtle implicit visual elements that deceive, but it is also supported by the explicit network of formal properties that enhance thought experiments and suggest useful new directions for formal deduction. In this way, not only can embodied thought be used for thought experiences to support formal thinking, but formal thinking can also produce logically based structure theorems that lead back to a more sophisticated kind of embodied thought.

The potential and actual infinity of counting numbers

> One, two, three four five,
> Once I caught a fish alive,
> Six, seven, eight, nine, ten,
> Then I let him go again.

In this section, we begin with the human activity of counting and consider how natural experiences and formal ideas can build first a definition (the Peano postulates) and then a formal structure for the theory in the sense of the diagram given in figure 2.

Natural experiences of number involve learning the number names in sequence, initially to say the first few numbers one after the other, and then building up the succession of whole

numbers in decimal notation. To select appropriate axioms for a formal version of these 'natural' numbers proves (in retrospect!) to be subtly simple. We need:

(a) to start somewhere (with the number 1),
(b) for each number, we need to have a 'next number',
(c) these numbers need to be all different,
(d) they are *all* the numbers we get.

To translate these into formal mathematical axioms, we begin with a set N considered as a formal concept in an axiomatic theory, and a function $s:N \to N$ which gives the set-theoretic idea of the 'next number' (or 'successor)' $s(x)$ for each $x \in N$. This takes care of (b). Properties (a), (c), (d) can then be formulated, starting from an element $\mathbf{1}$ (written in bold face to show it is a specific element in the set N, rather than the number 'one' from childhood). Suitable axioms are as follows:

(i) there is at least one element (denoted by $\mathbf{1} \in N$) which is not a successor, i.e. no $k \in N$ satisfies $s(k) = \mathbf{1}$.

(ii) s is one-to-one, i.e. $a \neq b$ implies $s(a) \neq s(b)$,

(iii) For any subset S of N satisfying
$$\mathbf{1} \in S, \text{ and, whenever } k \in S \text{ so also is } s(k) \in S,$$
then $S = N$.

These three axioms are based on the *Peano Postulates*, formulated by the Italian mathematician Guiseppe Peano in around 1895 and published in his magnum opus (Peano, 1908; Kennedy, 1973). The last of them (axiom (iii)) is called the 'induction axiom'. It is formulated to say that the succession of elements starting from $\mathbf{1}$ gives the full set N, in that if a subset S contains the starting element $\mathbf{1}$ and, whenever it contains an element k it also contains its successor $s(k)$, then S is the whole of N.

Here is a brief informal explanation of how these axioms may be used to build a formal theory of whole number arithmetic. During this discussion, remember that N is *not* the familiar system of natural numbers, but a purely theoretical concept that satisfies the three axioms.

Axiom (i) states 'there is *at least* one element which is not a successor'. We may use the other axioms to show that there is *precisely* one element, namely $\mathbf{1}$. For if not, there would be another element $c \neq \mathbf{1}$ in N where $c \neq s(k)$ for any $k \in N$. Consider the set

$$S = \{\, n \in N \mid n \neq c \,\}$$

of all elements in N except c. Then $\mathbf{1} \in S$ (because $\mathbf{1} \neq c$). If $k \in S$ then, because $s(k) \neq c$, it follows that $s(k) \in S$. The induction axiom (iii) then gives $S = N$, contradicting the assumptions $c \in N$ but $c \notin S$.

Thus, there is only *one* 'first' element in N, namely $\mathbf{1}$. We denote its successor $s(\mathbf{1})$ by $\mathbf{2}$, $s(\mathbf{2})$ by $\mathbf{3}$, $s(\mathbf{3})$ by $\mathbf{4}$, and may continue in this way to build the usual decimal notation for elements of N (although this is easier to do in practice than to write out formally) .

To formulate the arithmetic in N, we begin by 'adding one' to define $m+\mathbf{1}$ to be $s(m)$ for every $m \in N$. By definition, we get $\mathbf{1}+\mathbf{1} = s(\mathbf{1}) = \mathbf{2}$, $\mathbf{2}+\mathbf{1} = s(\mathbf{2}) = \mathbf{3}$, $\mathbf{3}+\mathbf{1} = s(\mathbf{3}) = \mathbf{4}$, and so on. To define $m+n$ in general we use induction, as follows.

> Consider the set S of elements n for which it is possible to define $m+n$ (for all $m \in N$). First, note that $\mathbf{1} \in S$ (having just defined $m+\mathbf{1} = s(m)$ for all $m \in N$). Suppose now that $k \in S$. Then $m+k$ will already have been defined for all $m \in N$. Then we define $m+s(k)$ to be the successor of $m+k$, namely $m+s(k) = s(m+k)$. This in turn shows that $s(k) \in S$. Using axiom (iii), we deduce that $S = N$. So, $m+n$ can be defined for all $m, n \in N$.

As examples of this, we obtain:

$$\begin{aligned}
\mathbf{1}+\mathbf{2} \;\; &= \mathbf{1}+s(\mathbf{1}) && \text{(because } \mathbf{2} = s(\mathbf{1})\text{, by definition of } \mathbf{2}\text{)} \\
&= s(\mathbf{1}+\mathbf{1}) && \text{(using the definition } m+s(n)=s(m+n) \text{ for } m = n = \mathbf{1}\text{)} \\
&= s(\mathbf{2}) && \text{(because } \mathbf{1}+\mathbf{1} \text{ is defined to be } s(\mathbf{1}) \; = \mathbf{2}\text{)} \\
&= \mathbf{3} && \text{(because } \mathbf{3} \text{ is defined to be } s(\mathbf{2})\text{)}.
\end{aligned}$$

Similarly:

$$\mathbf{2}+\mathbf{2} = \mathbf{2}+s(\mathbf{1}) \; = s(\mathbf{2}+\mathbf{1}) \; = s(\mathbf{3}) \; = \mathbf{4}.$$

In this way, we have *proved* that

$$\mathbf{2}+\mathbf{2} = \mathbf{4}$$

(where $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}$ are the initial elements of the set N rather than the usual numbers of childhood).

By these techniques it is possible to build up a formal system N which has the familiar properties of the 'natural numbers' we have known since childhood.

Is it worth it? For most people, probably not. After many years building up a concept image of whole number arithmetic with all its familiar properties, it seems hardly worthwhile repeating all the detail for all the other arithmetic facts, such as **4+4 = 8** or **123+9 = 132**, using the formal numbers of N. It is not something that I have done in full, nor something I wish to do in future. Once it is seen to be possible, mathematicians usually revert to the familiar notation for numbers and simply use the familiar rules of arithmetic, such as 2+2 = 4, in the belief that any of these results *could* be proved from the Peano axioms, wherever it is really necessary. In essence, one accepts that, once the generative nature of the Peano axioms have been realized, the informal image can be used for thinking about whole number arithmetic.

However, in some ways it *is* of enormous importance. For instance, the so-called 'law of commutativity' is something that is encountered by experience in everyday arithmetic. We 'know' that $m+n = n+m$ for all natural numbers m and n, but may sense some unease because we cannot 'prove' it true using our informal experience alone. However, it *can* be proved from the Peano axioms, although the proof is somewhat tedious. (First prove that $m+1 = 1+m$ by induction on m, then extend it to $m+n = n+m$ by induction on n.)

A more powerful idea is to realise that proof by induction need no longer be considered as a potentially infinite process (start at 1, and having established the general step from k to $k+1$, use this to get from 1 to 2, from 2 to 3, *and so on*).

Given a property P(n) that is true or false for every $n \in N$, define:

   $S = \{n \in N \mid P(n) \text{ is true}\}$,

Axiom (iii) above then becomes:

> **The induction axiom**: If P(1) is true and the truth of P(k) implies the truth of P($k+1$) (for all $k \in N$) then P(n) is true for all natural numbers n,

To prove a statement P(n) by induction, it is only necessary to do *three* steps:

(a) Prove that P(1) is true.

(b) Prove that the truth of P(k) implies the truth of P(k+1).

(c) Quote the induction axiom to give P(n) true for all $n \in N$.

A proof about the *potential infinity* of the natural numbers can be formulated as a *finite* three-stage proof based on the Peano axioms. The difficulty in shifting from potential to actual infinity is therefore avoided by the technical device of focusing on the finite sequence of steps in a proof by induction.

This may cause warning bells to ring in the back of the mind! How can we suddenly step from potential to actual infinity in this way? The answer relates to the manner in which the human brain operates by switching focus from one viewpoint to another. The three axioms (i), (ii), (iii) contain within them the implication that the set N must have an infinite succession of different elements. However, instead of focusing on the infinite qualities of the set N, we focus on the *finite* nature of the three axioms. This is a considerable conceptual gain. It does not make *all* proofs about the natural numbers into finite proofs, but it does allow *induction* proofs to be considered finite.

So where does Gödel's theorem fit in this scheme of things? This may arise for a statement P(n), known to be true or false for every natural number n, but for which the induction step from P(k) to P(k+1) is not available. A possible candidate is Goldbach's conjecture where P(n) states that 'every even number $2n$ is the sum of two primes' (where, in this case, 1 is considered as a prime). For example,

$$2 = 1+1$$
$$4 = 1+3 = 2+2$$
$$6 = 1+5 = 3+3$$
$$8 = 1+7 = 3+5$$
$$10 = 3+7 = 5+5$$

and so on. Every case P(k) that has ever been investigated has turned out to be true. But no-one has ever found a general way to make the deduction from P(k) to P(k+1). As I wrote this article, I thought about the problem, and got excited to find that sometimes the move from P(k) to P(k+1) can be accomplished by adding 2 to one of the primes concerned. For instance,

$$6 = 1+5 \text{ leads to } 6+2 = 1+5+2 \text{ so } 8 = 1+7.$$

This method obviously depends on $2n$ being the sum of odd primes $m+n$ where at least one of $m+2$ or $n+2$ is also prime. It fails for $14 = 7+7$ where the next odd number 9 is not prime. Perhaps one can salvage the situation by using a different decomposition, say,

$14 = 3+11$, leading to $16 = 5+11$ or $16 = 3+13$.

Regrettably, although this works in this particular case, it doesn't seem to lead to a general method that works for all cases. After a while, I gave up trying.

After this exercise, I still cannot say if Goldbach's conjecture is true or false. I don't know. Nor (at the moment) does anyone else. Just because no-one has found a proof (yet), however, does not mean that there is *not* a proof of some kind. Were someone eventually to find one, perhaps by some clever application of induction, it would only tell us for sure that Goldbach's conjecture is not a member of the list of arithmetic truths with no finite proof using the familiar axioms of arithmetic. Gödel's theorem confirms that this list is non-empty.

This discussion highlights a flaw in hoping that induction proofs can always be used to prove statements about whole numbers. If there is no proof to link the general case P(k) to P(k+1), then induction is not an available option. This should not be a total disappointment. As finite beings, we cannot cope with an infinite number of totally different cases at the same time. At best, we can use finite proofs that have a general pattern applying to an infinite number of similar cases. Goldbach's conjecture illustrates the possibility of formulating a general theorem in such a way that it applies to an infinite number of cases, but may lack a finite proof that applies to all of them.

Despite this set-back, there are still plenty of properties that *can* be proved by induction (including all the usual rules of arithmetic, such as $m+n=n+m$, $mn=nm$, $m(n+r)=mn+mr$, and so on.) For this reason, the axiomatic method continues to be used. It cannot prove *all* theorems for the natural numbers, but it can prove a wide range of useful ones. In particular, it leads to a formulation of axioms for the natural numbers in which the general laws of arithmetic are not just *natural* observations of what happens in the real world, but now recast as *formal* deductions from the specified axioms.

Cardinal numbers

The shift of attention, from the natural numbers as a potentially infinite collection going on and on, to a single entity—the *set* N, given axiomatically—leads to considering the relationships with other infinite sets. The theory of comparison of infinite sets developed by Cantor is generated by starting with the definition that:

> Two sets A, B are said to *have the same cardinal number* if there is a bijection $f : A \rightarrow B$.

Here the term 'bijection' means a function f which is one-one, (ie if $a_1 \neq a_2$ then $f(a_1) \neq f(a_2)$) and onto (ie, for every $y \in B$ there is an $x \in A$ such that $f(x) = y$).

Note that the definition, as given here, does not say what a cardinal number *is*, it only defines when two sets have the same number. For finite sets, the cardinal number may be taken as the counting number of elements in the set. For infinite sets we must invent new names for cardinal numbers, for instance $\aleph_0$ (aleph zero) is the name given to the cardinal number of the set N. The question arises as to which other sets have cardinal number $\aleph_0$. Certainly, the odd numbers O and the even numbers E can be put in one-to-one correspondence with N, as seen in figure 3.

$$
\begin{array}{llllll}
N : & 1 & 2 & 3 & \dots & n & \dots \\
 & \updownarrow & \updownarrow & \updownarrow & & \updownarrow \\
E : & 2 & 4 & 6 & \dots & 2n & \dots
\end{array}
$$

Figure 3: A one-to-one correspondence between the natural numbers and the even numbers

To extend the notion of addition to infinite cardinals, all that is necessary is to formalise the natural 'count-all' technique of childhood: count one set, count another and put them together and count all the elements in both sets. For cardinals α and β (finite or infinite), the sum $\alpha+\beta$ is found by selecting a set A with cardinal number α, a set B with cardinal number β (where A, B have no common elements) and defining $a+\beta$ to be the cardinal number of $A \cup B$. Because $A \cup B = B \cup A$, then $a+\beta = \beta+\alpha$, so that addition of cardinals is commutative.

For instance, if we take A to be a set with a single element a (which is not in N), then there is a bijection between N and $A \cup N$. (Figure 4).

$$N: \quad 1 \quad 2 \quad 3 \ \dots \ n \ \dots$$
$$\quad\quad \updownarrow \quad \updownarrow \quad \updownarrow \quad\quad \updownarrow$$
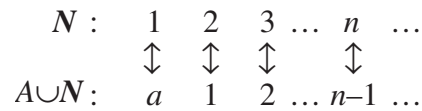$$A \cup N: \quad a \quad 1 \quad 2 \ \dots \ n{-}1 \ \dots$$

Figure 4: A one-to-one correspondence between the natural numbers
and the same set 'plus one'.

This shows that the cardinal numbers of $A \cup N$ and N are the same, so:

$$1 + \aleph_0 = \aleph_0 + 1 = \aleph_0.$$

Replacing A by a set with n elements shows that:

$$n + \aleph_0 = \aleph_0 + n = \aleph_0 \text{ for any natural number } n.$$

Also, because the union of odd and even numbers gives $O \cup E = N$, it follows that:

$$\aleph_0 + \aleph_0 = \aleph_0.$$

The equation

$$\aleph_0 + x = \aleph_0$$

can therefore have many solutions, including 0, 1, 2, 3, … and $\aleph_0$. So, subtraction involving infinite cardinals cannot be uniquely defined.

Similar phenomena occur with attempts at defining multiplication of cardinals. For any two sets A, B, the *cartesian product $A \times B$* is defined to be the set of all ordered pairs (a, b) with $a \in A$, and $b \in B$. The product of two cardinal numbers α, β can be defined by taking sets A and B with cardinals α, β, respectively, and defining $\alpha\beta$ to be the cardinal number of $A \times B$. For instance, if we take a set A with two elements, say $A = \{1,2\}$ and B a set with three elements, $B = \{1,2,3\}$, then the cartesian product has six ordered pairs:

$$(1,1), (1,2), (1,3), (2,1), (2,2), (2,3),$$

corresponding to the fact that 2 times 3 is 6.

Notice that, in this case, the cartesian product $B \times A$ consists of

$$(1,1), (1,2), (2,1), (2,2), (3,1), (3,2),$$

and is *not* the same set as $A \times B$ (for instance, $(2,3) \in A \times B$, but $(2,3) \notin B \times A$). In general, $A \times B$, $B \times A$ are not equal, but the correspondence $(a, b) \leftrightarrow (b, a)$ sets up a bijection between $A \times B$ and $B \times A$, so the two sets have the same cardinal number.

If A is a set with two objects, x and y, then $A \times N$ is the set of ordered pairs of the form (x, n) or (y, n) for any natural number n and, by definition, it has cardinal number $2\aleph_0$. However, this set may be put into one to one correspondence with N as in figure 5:
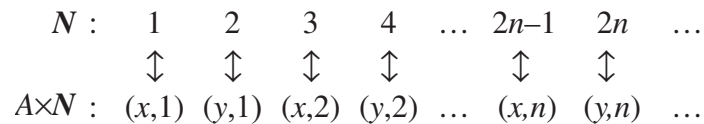
$$
\begin{array}{ccccccccc}
N: & 1 & 2 & 3 & 4 & \dots & 2n{-}1 & 2n & \dots \\
 & \updownarrow & \updownarrow & \updownarrow & \updownarrow & & \updownarrow & \updownarrow & \\
A \times N: & (x,1) & (y,1) & (x,2) & (y,2) & \dots & (x,n) & (y,n) & \dots
\end{array}
$$

Figure 5: A one-to-one correspondence between the natural numbers
and a set with 'twice as many' elements

This proves that $2\aleph_0 = \aleph_0$.

A similar argument shows that $n\aleph_0 = \aleph_0$ for any $n \in N$. A bijection $f : N \to N \times N$ may be defined by counting down the successive 'backward diagonals' as in figure 6 (where $f(1) = (1,1)$, $f(2) = (1,2)$, $f(3) = (2,1)$, $f(4) = (1,3)$, $f(5) = (2,2)$, …). This proves that $\aleph_0 \aleph_0 = \aleph_0$.

$$
\begin{array}{ccccc}
(1,1) & (1,2) & (1,3) & (1,4) & \dots \\
(2,1) & (2,2) & (2,3) & (2,4) & \dots \\
(3,1) & (3,2) & (3,3) & (3,4) & \dots \\
(4,1) & (4,2) & (4,3) & (4,4) & \dots
\end{array}
$$

Figure 6: 'Counting' the set of all ordered pairs (m,n) of natural numbers m, n

The equation $\aleph_0 \times x = \aleph_0$ therefore has solutions which include $x = 1$, $x = 2$, … and $x = \aleph_0$. So, *division by infinite cardinal numbers cannot be uniquely defined*.

At this point all the examples given are of sets with the same cardinal number as the set N of natural numbers. However, for any set S, we can construct another set T whose cardinal number is strictly larger. The set T consisting of all subsets of S is a good candidate. Clearly the set S has cardinal number less than or equal to T (because the map from S to T mapping the element x to the single element subset $\{x\}$ is a bijection onto a subset of T). Now consider *any* map $f : S \to T$. We will show that it cannot map onto the whole of T. Consider the subset $K = \{x \in S \mid x \notin f(x)\}$. If K were in the image of f, then $K = f(x_0)$ for some $x_0 \in S$. and we ask whether $x_0 \in f(x_0)$. If $x_0 \in f(x_0)$ then, by definition of K, we have $x \notin K$, contradicting

$K = f(x_0)$. On the other hand, if $x_0 \notin f(x_0)$ then, again by definition of K, we have $x \in K$, once more contradicting $K = f(x_0)$. Thus, no map from S to T can be a bijection, so T must have a strictly larger cardinal number than S.

By repetition of this construction, we get an unending sequence of cardinal numbers, each one strictly bigger than the one before.

Ordinal numbers

Natural numbers are not only used for *counting*, but also for putting a (finite) set into an order, saying 'this element is the first', this is the second', and so on. Just as the counting property can be extended to infinite sets using the cardinal number concept, there is an infinite extension of ordering numbers to ordinal numbers. This is done using *ordered* sets where a set A is said to be ordered if there is a relation denoted by $a < b$ between some pairs of elements $a, b \in A$, such that:

- Given $a, b \in A$, precisely one of the following hold:
  $a < b$ or $b < a$ or $a = b$,

- If $a < b$ and $b < c$, then $a < c$.

Examples of ordered sets include the set of integers $Z = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ and the set $N = \{1, 2, 3, \ldots\}$ with their usual order.

To define the notion of ordinal number requires the ordering also to satisfy the definition:

- an ordered set A is said to be *well-ordered* if every non-empty subset S of A has a least element, that is there is an element $l \in S$ where $l \leq s$ for all $s \in S$.

Of the two examples given, N in its usual ordering is well-ordered, but Z is not, because the set Z itself fails to have a least element in the given ordering.

A function $f : A \rightarrow B$ is said to be 'order-preserving' if $a < b$ implies $f(a) < f(b)$.

Two well-ordered sets A, B have the same *ordinal number* if there is an order-preserving bijection $f : A \rightarrow B$.

The map in figure 7 between the natural numbers N and the even numbers E clearly preserves order, so the sets N and E have the same ordinal number, denoted by ω (omega). (A different

symbol is used from the cardinal number because it happens to have very different properties, as we shall shortly show.)

$$N : \quad 1 \quad 2 \quad 3 \ldots \quad n \quad \ldots$$
$$\updownarrow \quad \updownarrow \quad \updownarrow \qquad \updownarrow$$
$$E : \quad 2 \quad 4 \quad 6 \ldots \quad 2n \ldots$$

Figure 7: An ordered correspondence between the natural numbers and the even numbers

We can build a mental image of a well-ordered set A, as a thought experiment, by considering successive elements in turn. Either A is empty (with ordinal number 0), or contains a least element a_1. On removing a_1, the remaining subset of elements is either empty, or contains a least element a_2. Continuing in this way, either A is finite with n elements in order $a_1 < a_2 < \ldots < a_n$, or it contains an infinite set F of elements, $a_1 < a_2 < \ldots < a_n < \ldots$. In the latter case there is an order-preserving bijection $f : N \rightarrow F$ where $f(n) = a_n$, showing that the ordinal number of F is ω. Thus every well-ordered set S is either finite, or begins with a set with ordinal number ω. (It can, of course have further elements that come after the elements in F in the given order.)

The addition of two ordinal numbers, corresponding to well-ordered sets A and B, is performed by ordering $A \cup B$ by placing the elements in A *before* the elements in B. For instance, to obtain a set with ordinal number $1+\omega$, we give the set $\{c\} \cup N$ an order in which the element c comes before all the elements of N, so $c < 1 < 2 < \ldots < n$ for all $n \in N$. This can clearly be put in a one-to-one order preserving relationship. (Figure 8.)

$$N : \quad 1 \quad 2 \quad 3 \ldots \quad n \quad \ldots$$
$$\updownarrow \quad \updownarrow \quad \updownarrow \qquad \updownarrow$$
$$\{c\} \cup N : \quad c \quad 1 \quad 2 \ldots n{-}1 \ldots$$

Figure 8: An ordered correspondence between the two given sets

The cardinal number of $\{c\} \cup N$ is thus the same as the cardinal number of N, giving:

$1+\omega = \omega$.

Just as in the cardinal sense, we seem to get the same ordinal once more. However, the situation with ordinals is more subtle. To get an ordered set with ordinal number $\omega+1$, we order $N \cup \{c\}$ so that $1 < 2 < \ldots < n < \ldots < c$ where the element c now satisfies $n < c$ for all

$n \in N$. If we try to put this ordered set in one-one ordered correspondence with N, then c must correspond to some $k \in N$ as in figure 9.

$$N \cup \{c\}: \quad 1 < 2 < ... < c \qquad ?$$
$$\updownarrow \qquad \updownarrow$$
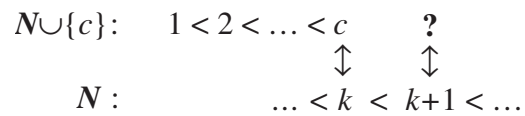$$N: \qquad\qquad ... < k \ < \ k+1 < ...$$

Figure 9: There is no ordered correspondence between these ordered sets

But there is no element in $N \cup \{c\}$ exceeding c in the order, so there is no element in $N \cup \{c\}$ which can correspond to $k+1$ in N. Hence there is no possible order preserving correspondence between $N \cup \{c\}$ and N, so $\omega+1$ is *different* from ω. This gives the unexpected results for ordinals that

$$1+\omega = \omega,$$

but

$$\omega+1 \neq \omega,$$

which in turn gives

$$1+\omega \neq \omega+1.$$

Addition of ordinals is, therefore, *not commutative*.

This discussion shows that cardinal infinity and ordinal infinity have different properties from each other and both differ from finite experiences of number. So strange were these ideas to the mathematical community when first announced that Kronecker prevented the initial publication of Cantor's theory of infinite cardinals and ordinals. Such pressures subsequently caused Cantor to have a nervous breakdown. Now the theory is widely accepted but this does not make it any easier for the learner, who must negotiate a minefield of conflicting intuitions and resolve them by a reconstruction of personal knowledge. Although there are many theorems about cardinal and ordinal numbers, an intuitive grasp of their structure invariably comes *after* developing the formalism and reconstructing personal knowledge, not before.

These types of infinity have very different properties from those that arise in other areas of mathematics, particularly those that arise naturally in the calculus from imagining

'arbitrarily large' and 'arbitrarily small' numbers in the calculus. It is to this different type of infinity we now turn.

The number line

The number line is met visually in a natural embodied sense as a straight line with a chosen unit measured from 0 to 1, then repeated at equal intervals to 2, 3, … as on a ruler. (Figure 10.)
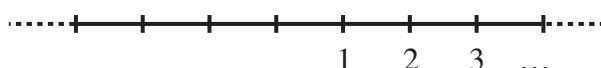

Figure 10: Numbers as points on a line

By marking negative numbers in the other direction and positive and negative fractions in between, this may be further refined to give the 'rational' number line. (Figure 11.)
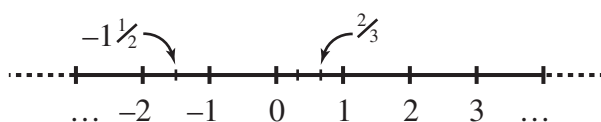
Figure 11: Inserting rationals and negatives

The points on the number line are intimately connected to the line segment measured from the origin. The length of the line segment, taking direction into account, is also used to label the point as well. When students are assumed to be familiar with this idea, teachers reveal that there are points other than rational points on the number line. For, if an isosceles right-angled triangle has side-length 1, its hypotenuse is of length √2, and this length measured from the origin does not end up at a rational number.

At this stage, a sleight of hand occurs. As students, we are told that those points on the number line, which are not rational, are called 'irrational'. The theoretical 'real' line therefore consists of rational and irrational points. In essence, this 'explanation' is no explanation at all! It does not say what an irrational number genuinely *is*, only what it is *not*. (It is *not* rational.)

Students therefore must use their intuitions of the number line to give meaning to the notion of 'irrational'. But individuals have many different intuitions, depending on their previous experience, and one which is commonly held is the belief that quantities exist which

can be 'arbitrarily small', for instance the difference between 0.999…('nought point nine recurring') and 1 is extremely small, but not zero (Cornu, 1991). Most students, many teachers, and some mathematicians (certainly many physicists), do good work by imagining 'infinitesimally small' quantities.

The idea that there is a clearly defined 'number line' shared intuitively by everyone is therefore not acceptable. Despite the efforts of many mathematicians, including Cantor, to outlaw infinitesimals from mathematical analysis, they are alive and well today and continue to be used fruitfully by many experts. What we must do, therefore is to explain in what sense an 'infinitesimal' can be said to 'exist', both in a formal sense, and also as a natural embodied object that we can imagine in our mind's eye.

Infinitesimals

> … to conceive a quantity infinitely small, that is, infinitely less than any sensible or imaginable quantity, or than any the least finite magnitude is, I confess, above my capacity. (Bishop Berkeley, (reprinted 1951) p. 67.)

How can any sensible person conceive of something smaller than any sensible quantity? The answer is that we have great difficulty if we use only the *perceptions* of things in the world we live in. But we do not just perceive, we *reflect* on our perceptions to create new cognitive images within our personal cognitive structure. It turns out that many students who have experiences of quantities 'becoming' small in limiting processes, begin to imagine a 'variable quantity' that can be 'as small as you like' (Cornu, 1981, 1991). The language used in the calculus classroom tends to promote such thoughts, leading to individuals thinking of variables tending to zero as being 'arbitrarily small'. Certainly, the early pioneers in the calculus thought in such terms (see Kleiner, this volume).

In the following sections we begin with an informal image of the number line, select appropriate fundamental properties of the arithmetic and order as axioms for an 'ordered field', then deduce a formal structure theorem which applies to all systems satisfying the axioms. This structure theorem then leads to a visual embodiment of the formal idea of an infinitesimal as a point on a number line.

Ordered fields

We begin our formal journey by recalling the notion of *ordered field* as defined in an undergraduate mathematics course. This is given in terms of a set F with two operations $+, \times$, which, for every pair of elements $x, y \in F$, there are elements $x+y, x\times y \in F$, where $x\times y$ is usually written as xy. The arithmetic axioms are as follows:

- $x+y = y+x$, $xy = yx$, $(x+y)+z = x+(y+z)$, $(xy)z = x(yz)$ (for all $x, y, z \in F$).

- There are (different) elements 0, 1 such that $0+x = 0$, $1x = x$ for all $x \in F$.

- $x(y+z) = xy+xz$, for all $x, y, z \in F$.

- For each $x \in F$ there is $u \in F$ such that $x+u = 0$ and (if $x \neq 0$) there is $v \in F$ such that $xv = 1$. [The elements u and v are usually written as $-x$ and x^{-1}, respectively.]

The 'order axioms' can be formulated in various ways. Here, they will be given in terms of a subset P of F (which may be considered as the 'positive' elements of F) satisfying the following:

- If $x, y \in P$ then $x+y$ and xy are also in P.

- Every $x \in F$ satisfies one and only one of these properties:
  $x \in P$, or $-x \in P$, or $x = 0$.

The order relation $x > y$ (also written as $y < x$) is then defined to hold if and only if $x-y \in P$. An illustration of the axiomatic method is to use these axioms to derive the 'usual' order properties of the relation $>$. For instance, given $a, b \in F$ then the second property implies one and only one of the following hold:

$a-b \in P$, or $-(a-b) = b-a \in P$, or $a-b = 0$.

Hence, $a > b$ or $b > a$ or $a = b$.

Given

$x > y$ and $y > z$,

then

$x-y, y-z \in P$,

so, adding them, we get

$$(x–y)+(y–z) = x–z \in P,$$

which gives

$$x > z.$$

Hence $x > y$ and $y > z$ implies $x > z$.

This latter deduction is highly significant in terms of whether thinking is embodied or not. It establishes this order property not by the mental image of points on a line or by the Lakoff container metaphor, but by substitution of the symbols into specific order axioms prescribed for the set P.

Similar arguments allow other standard properties of order relations to be built up formally from the axioms.

If a field F also satisfies the axioms of order, then the system is said to be an 'ordered field'. Such systems include the rational numbers Q and the real numbers R.

The real numbers also satisfy a further axiom, the axiom of 'completeness':

> An ordered field is said to be *complete* if any non-empty subset S of F with an upper bound $k \in F$ (meaning $x \leq k$ for all $x \in S$) has a *least* upper bound λ (that is an element λ that is both an upper bound of F and smaller than any other upper bound of S).

(For further details, see for instance, Stewart & Tall, 1977).

There is an interesting reversal that occurs in formal theory. A system with *more* axioms usually applies to *fewer* examples. For instance, there are many different ordered fields, including Q and R. However, there is essentially *only one* complete ordered field, the familiar field of real numbers. (This follows from a structure theorem proving that, given two systems satisfying the axioms for a complete ordered field, then there is a bijection between them that is also a direct translation between arithmetic and order. In formal terms, any two complete ordered fields are *isomorphic*.)

Theorems proved about complete ordered fields therefore apply only to the single system of real numbers R. However, any theorems proved without the use of the completeness axiom will automatically apply to *every* ordered field.

There are *many* ordered fields, including some that contain the real numbers R as a subfield. Given that the mental image of the real numbers is essentially to 'fill in the gaps' on the number line (for instance using the theory of Dedekind cuts), it may come as a surprise to find that it is possible to fit in even *more* points than the real numbers on a visual number line.

This is addressed in the next section, by presenting a specific example of such an ordered field expressed formally. This example is given two distinct visual representations. The focus then turns to the axioms for *any* ordered field containing the real numbers R, leading to a structure theorem that enables us to visualize any such ordered field as a visual number line.

An example of an ordered field containing the real numbers

Consider the set $R(x)$ of rational expressions in a variable x, of the form
$$\frac{a_0 + a_1 x + \ldots + a_n x^n}{b_0 + b_1 x + \ldots + b_m x^m}$$
where the coefficients a_r and b_s are all real numbers, with $b_m \neq 0$. These rational expressions have the familiar algebraic properties of addition, subtraction, multiplication and division and satisfy all the axioms for a field. Moreover, given $\alpha \in R$, by taking $a_0 = \alpha$, $b_0 = 1$ and all other coefficients zero, we may consider the real numbers as a *subfield $R \subset R(x)$*, or equivalently, seeing $R(x)$ as a *larger* system containing the real numbers R.

The student may already be familiar with these expressions, simply through manipulating them in school algebra. However, our natural experience of school algebra does not put these elements in an order that turns $R(x)$ into an ordered field. We may do this either by a formal or natural approach. The formal approach simply makes a definition of a subset P of rational functions that satisfies the properties of an ordered field, in the strict sense of Hilbert's programme. The alternative, a natural approach, attempts first to motivate the idea with a practical example.

A natural approach might be to interpret rational functions embodied as *graphs* and try to put some kind of order on them, or, equivalently, to specify a set P of rational functions that

is to be taken as the 'positive' elements. As rational functions are capable of being positive in some places and negative in others, a little cunning is in order.

A rational function is either the zero function, or it is the quotient of two non-zero polynomials, each of which has only a finite number of places where it can be zero. Thus a non-zero rational function has only a finite number of points where it is zero or undefined. Take an interval $0 < x < k$ to the right of the origin (where $k \in \mathbf{R}$) which contains none of these places. Then in the interval $0 < x < k$ the function is neither zero nor infinite. It cannot be negative at one point in the interval and positive somewhere else in the interval, otherwise, being continuous, it would be zero in between. It is therefore either *strictly* positive or *strictly negative* in this interval to the right of the origin.

A rational function is declared to belong to P if there is a real number k such that the function has positive values for all real numbers $0 < x < k$. Figure 12 shows three rational functions that are positive in an open interval to the right of the origin. The rational function $f(x)=x$ is positive for x in an open interval from 0 to k for any positive k, $g(x)=x(1–x)$ is positive in the open interval from 0 to 1, and $h(x)=1/x$ is positive in any positive interval from 0 to k.
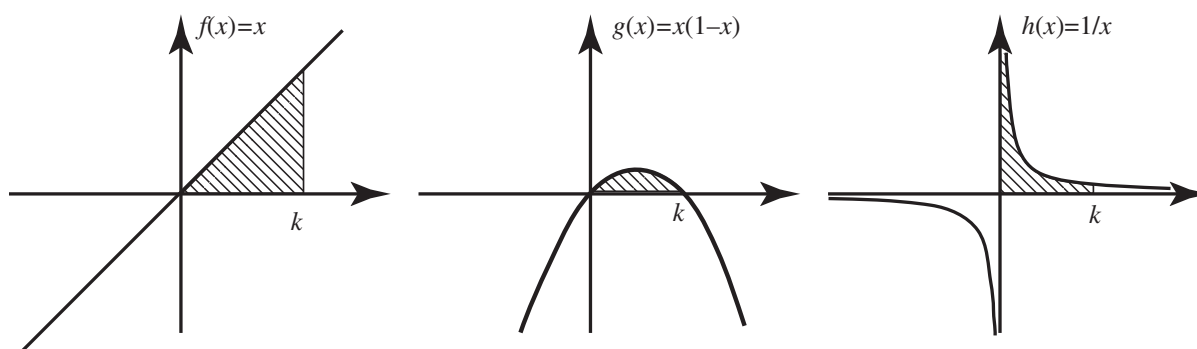

Figure 12: Examples of rational functions in the set *P* of 'positive' elements
(positive in an interval to the right of the origin)

This definition satisfies the axioms for an order given above and, in common with convention, we say that an element in P is 'positive'.

This particular example has an instructive purpose. The rational function is not 'positive' in the everyday meaning of the term, but purely in terms of the formal definition that it is a member of the set P. However, such a definition is *meaningful* in a different embodied sense,

in that it relates the new meaning of the term 'positive' to a specific visual property of a graph. This shows that formal mathematics is *not* a meaningless activity that only works by moving symbols around. During theorem-proving it is essential to focus on the deductive argument. However, in performing thought experiments, or interpreting formal theorems in an embodied sense, the formalism may be given *any* meaningful embodiment that satisfies the formal definition. This is particularly important when a given formal structure can have more than one informal embodiment.

Having defined what it means to be 'positive', the 'order' between rational functions f, g is specified by saying:

$f > g$ if and only if $f - g \in P$.

This does not mean that the graph of f is *everywhere* greater than g, simply that the graph of f is higher than that of g in an interval immediately to the right of the origin.

In this new sense $x > 0$ and $r > x$ for any positive real number r (because $x \in P$ and $r - x \in P$). Thus $0 < x < r$ using the new order defined in the larger ordered field $R(x)$. The element x is therefore 'positive' yet 'smaller' than any positive real number r. It is an *infinitesimal*.

This construction poses some possible problems. The formal learner may be satisfied with the formal symbolic approach to the definition of the ordered field $R(x)$ and have no need for the picture. The natural learner who has a predisposition for thinking of an infinitesimal as a 'very small quantity' may have difficulty because the embodiment is a *graph*, not a number, or a *point* on a line.

There are several ways of visualising infinitesimals as points on a line. One is to imagine a vertical line near the origin through the point $x = t$. In figure 13, the line meets the constant graph $f(x) = c$ at the point C, the straight line $l(x) = x$ at L, and the parabola $p(x) = x^2$ at P. As the value of t gets close to zero, the line moves near to the y-axis, the point C stays at a fixed height c above the x-axis, but L moves below C and P moves below L. In this sense, one can imagine C as a constant point on the vertical line, with L and P as 'variable points' that remain positive but are ultimately *below* any positive real constant c. This thought

experiment allows the novice to 'see' infinitesimals as 'variable points' that are 'arbitrarily smaller' than any positive real constant c. This is analogous to the notion prevalent at the beginning of the nineteenth century when Cauchy described infinitesimals as 'variable quantities that tend to zero'.
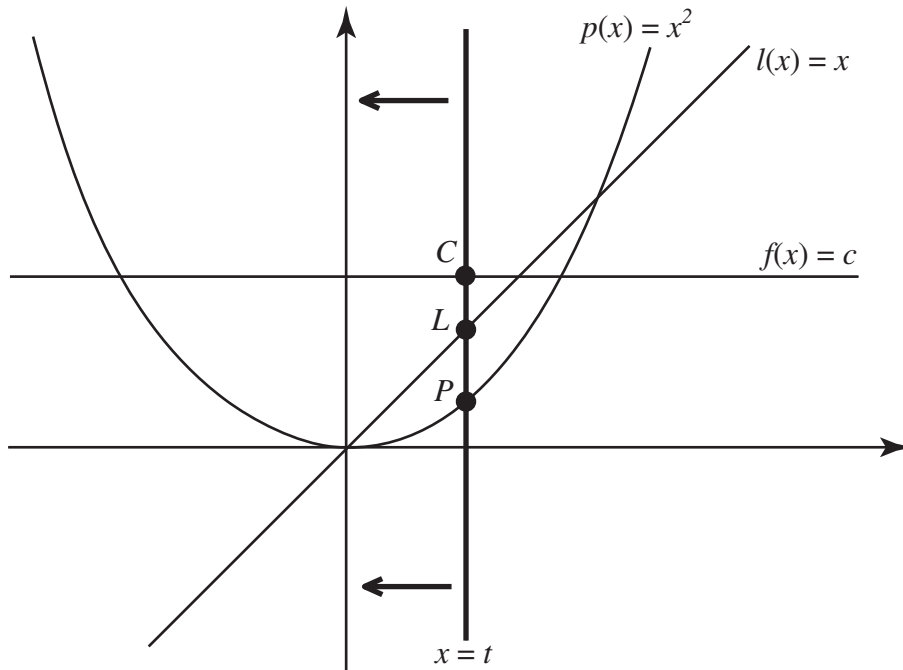

Figure 13: Variable quantities seen as infinitesimals on a moving vertical line

Such an image may be helpful to some natural learners, but it can be very unhelpful to others. It fails to represent an infinitesimal as a genuine *fixed* point on a number line.

A formal image of an ordered field with infinitesimals as fixed points

The individual who has developed a *formal* concept image of an ordered field, involving only deductions from the axiomatic definition, can prove a number of formal properties that provide a context for the meaning of an infinitesimal. First one can prove the structure theorem that F contains a subfield (isomorphic to) the rationals Q. (The proof begins with $1 \in F$. Forming successive sums 1, 1+1, 1+1+1, ... using an induction argument generates N, including negatives and multiplicative inverses generates the subfield Q.)

By thinking of Q as a subset of F, an element $x \in F$ is said to be a *infinitesimal* if $-r < x < r$ for all positive $r \in Q$. This includes the case $x=0$. However, no other positive

rational number $x \in Q$ can be infinitesimal because then $r=x/2 \in Q$ would satisfy $r < x$, contrary to the definition. A similar argument holds for x negative, so there are no non-zero infinitesimals in Q, although there may be infinitesimals in F that are not in Q.

If we try to represent the elements of F as points on a number line, we could begin by marking the elements in the subfield Q as rational points. But where do we fit the other elements of F? Our experience tells us that they might be irrationals as in R, or some other kind of element we need to think more about.

To simplify our discussion, we consider now only an ordered field F, such as $R(x)$, which already contains *all* the real numbers R. An infinitesimal will then satisfy $-r < x < r$ for all positive $r \in R$. By the same argument as for Q just given, we have the following theorem:

> *A non-zero infinitesimal is not a real number; it is an element of a larger ordered field.*

An element $x \in F$ is defined to be *positive infinite* if $x > r$ for every $r \in R$, and *negative infinite* if $x < r$ for every $x \in R$. A *finite* element $x \in F$ satisfies $a < x < b$ for some $a, b \in R$.

With these definitions, we can use the completeness of R to prove a structure theorem that describes the structure of *any* ordered field containing the real numbers:

> **Structure Theorem**: *Every ordered field F with the real numbers R as a proper subfield contains both infinitesimals and infinite elements. Every finite element is of the form $c+\varepsilon$ where $c \in R$ and $\varepsilon \in F$ is infinitesimal. Every infinite number is of the form $1/\varepsilon$ where ε is infinitesimal.*
>
> *Proof*: F has at least one element $x \in F$ not in R. This element x must either be finite or infinite. If x is infinite, a simple calculation shows that $1/x$ is infinitesimal.
>
> If x is finite, then it satisfies $a < x < b$ for $a, b \in R$. Consider the set
>
> $$S = \{t \in R \mid t < x\}.$$
>
> Then S is non-empty (because $a \in S$) and is bounded above (by $b \in R$), so by completeness it has a *least upper bound* $c \in R$. Let $\varepsilon = c-x$, (so that $x = c+\varepsilon$). If $c < x$, and there can be no $k \in R$ in $c < k < x$, as the definition of S gives $k \in S$ where k then exceeds the upper bound c of S. For $x < c$ there can be no $k \in R$ in $x < k < c$ since then, for all $t \in S$, we have $t < x$, so $t < k$, implying k is an upper bound for S that is smaller than the *least* upper bound c. In either case there is no real number between x and c. So $\varepsilon = x-c$ is an infinitesimal.

A visual embodiment of infinitesimals in *any* ordered field containing *R*

The structure theorem we have just proved has been established by deduction from the formal definition. It does not depend on any embodied visual properties. However, it allows us to translate the formal statement into a visual embodiment as a 'number line'. The structure theorem implies every element of *F* is either infinite, which would seem 'too far off' in one direction or the other to see in a finite picture, or it is of the form

$x = c + \varepsilon, (c \in F, \varepsilon$ infinitesimal$)$

which seems 'too close' to *c* to be visually distinguishable from it.
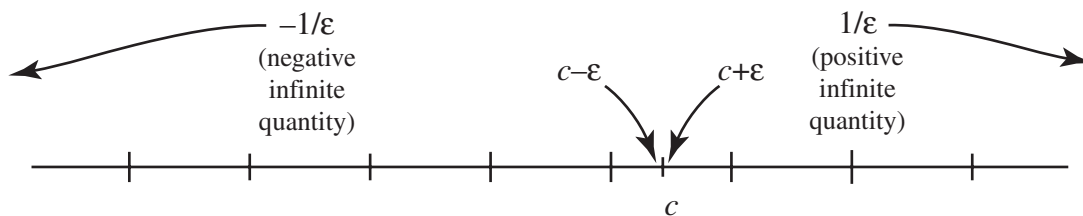

Figure 14: An ordered field *F* with infinitesimals and infinite elements that seem difficult to see!

We are, apparently, in a worse dilemma than in attempting to distinguish visually between rationals and irrationals. However, unlike Bishop Berkeley, who attempted to make sense of infinitesimals purely from his perception and intellect, we have a *formal image* of a complete ordered field with formal structure theorem to give us support.

The problem in 'seeing' infinitesimals is not essentially different from the problem of 'seeing' distinct real numbers on the real number line. To distinguish visually between the irrational number $\sqrt{2}$ and the rational number 1.41421, we use a sufficiently large scale.

The same idea works for visualising the difference between *c* and *c*+ε for c ∈ *R* and ε infinitesimal. We introduce the formal function μ:*F*→*F* given by:

$$\mu(x) = \frac{x - c}{\varepsilon}.$$

Then $\mu(c) = 0, \mu(c+\varepsilon) = 1$, so μ maps *c* and *c*+ε to the (visibly) distinct points 0,1.
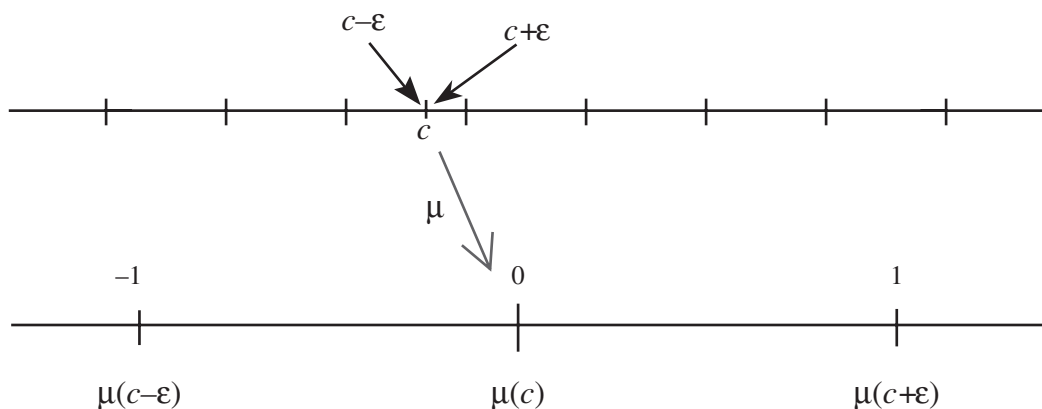
Figure 15: Visualising the function μ as a mapping of a number line with infinitesimals

In map-making the location of a place (eg 'New York') on a map is denoted by its name (New York). Replacing the names of the images by the names of the original points in the same way gives the picture in figure 16.
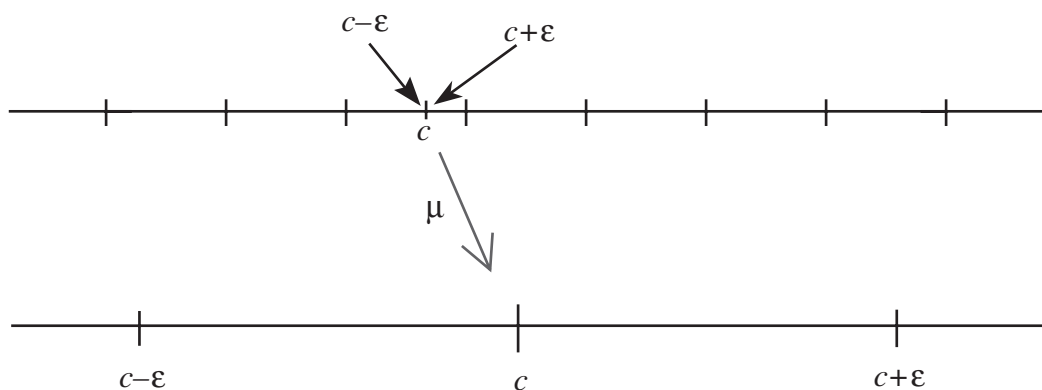

Figure 16: A magnified portion of number line, showing infinitesimals

This technique allows the formal map μ to be visualised in terms of a process that rescales and separates the points c and $c+\varepsilon$.

More generally, for any $\alpha, \delta \in F$, the map $\mu:F \to F$ given by

$$\mu(x) = \frac{x - \alpha}{\delta} \quad (\alpha, \delta \in F, \delta \neq 0)$$

is called the δ-*lens pointing at* α. This operation defined purely in symbolic terms, based on the operations available in a field, has a *visual* interpretation that maps any two points α and α+δ onto visibly distinct points in a finite picture. By taking α or δ to be finite, infinite, or infinitesimal, we obtain a visual method of seeing different parts of the extended number line rescaled to be seen in a finite picture. Remarkably, *this translates a formally defined construction into an embodied visual representation*. It is more than just an *informal*

representation, it is a visual interpretation of the *formal* construction, and, as such, has quite explicit formal properties that can be used in formal deductions.

Precisely what is seen using a δ-lens can be explained using the following definition:

λ ∈ *F* is of *higher order* detail than δ ∈ *F* if λ/δ is infinitesimal, it is the *same order* if λ/δ is finite (but not infinitesimal) and *lower order* if λ/δ is infinite.

A δ-lens pointed at α reveals details which differ from α by the same order as δ; a point which differs from α by higher order difference is 'too close' or 'too small' to see, and lower order differences are 'too far away to be seen'.

The same ideas extend to two or more dimensions using a lens on each coordinate. For instance, a δ-lens pointed at (a, a^2) reveals a nearby point $(a+h, (a+h)^2)$ on the graph of $y = x^2$ as:

$$\mu(a + h,(a + h)^2) = \left(\frac{(a + h) - a}{\delta}, \frac{(a + h)^2 - a^2}{\delta} \right)$$
$$= \left(\frac{h}{\delta}, \frac{2ah}{\delta} + \frac{h^2}{\delta} \right)$$
$$= (\rho,\ 2a\rho + h\rho) \text{ where } \rho = h / \delta.$$

If δ is infinitesimal and h is of the same order, then ρ is finite and $h\rho$ is infinitesimal. The standard part of the image is therefore

$$(\text{st}(\rho),\ \text{st}(2a\rho + h\rho)) = (r, 2ar) \text{ where } r = \text{st}(\rho) \in \textbf{\textit{R}}.$$

Looking at the standard part of an image is the same as looking at real numbers. We define the composite of δ-lens μ followed by taking the standard part to be called an *optical* δ-lens (Tall, 1980b, 1982). The finite part of the image of an optical δ-lens is called the *optical image*. An optical δ-lens maps part of a line with infinitesimal detail to a picture that reveals detail of the same order as δ, as a picture in the real plane.

In the case just mentioned, the optical image is the set of points ($r, 2ar^2$) for $r \in \textbf{\textit{R}}$. This is precisely the real straight line, gradient $2a$, through $(0, 0) = \mu(a, a^2)$. We therefore have the following theorem:

The optical image of a differentiable function under infinite magnification is a straight real line.

As the infinitesimal parts have been suppressed, the optical image is not *nearly* straight, but *really* straight.
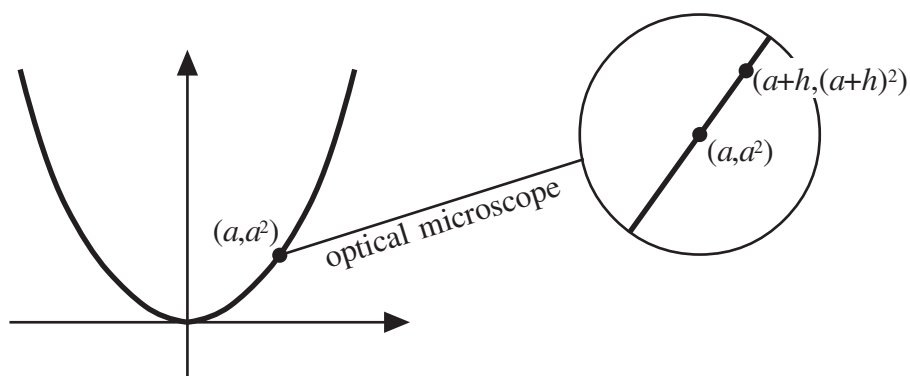

Figure 17: Infinite magnification of infinitesimal portion of a graph as a visual image

One small step for man, one giant leap for mankind

When Neil Armstrong stepped on the moon, he moved our conceptualisation of space a great leap forward by making human imagination of the moon a reality, for himself at least. For the rest of us, of course, it is still imagination, but imagination founded on the actual experience of another.

In mathematics the step that we take in saying, 'let there be a set N satisfying the Peano postulates,' is one which *none* of us can take in actuality, for we cannot hold an infinite number of concepts in our finite mind. Yet, it is a step that many of us take easily in our imagination.

Nevertheless, there is an even greater step taken by some mathematicians, who say:

> Let us suppose that we have a collection of sets (possibly infinite) where none of them is empty, then *simultaneously* it is possible to choose a single element from each set in the collection.

This is the *axiom of choice*. It is a powerful conception that gives great power in mathematical proof. Some sceptical individuals, including those from the school of intuitionism, point to its subtle weakness, that it asserts the existence of a choice without saying how it is to be done. This leads to diverging views of formal mathematics, those who accept the axiom of choice as a legitimate tool in mathematical proof, and those, including

the school of intuitionism, who reject it. In the latter case, all existence proofs need to involve actual constructions given by specified algorithms, as in the approach of Bishop (1967).

The twist in the tale came in the nineteen sixties when Paul Cohen (1966) proved that the axiom of choice is *independent* of the other axioms of set theory. Using it allows the proof of more theorems than would be possible without it, but adding it gives no new contradictions that are not present in set theory without the axiom of choice. From this viewpoint, it would seem that the intuitionist approach is just as liable to contain contradictions as a formal approach with the axiom of choice added. However, the proof of this argument uses the axiom of choice, and so it would be unlikely to be accepted from an intuitionist viewpoint.

The axiom of choice has powerful uses in analysis. For instance, it can be employed to prove the existence of a field $*R$ called a *hyperreal* field, which is an extension of R and is logically appropriate to perform all the usual tasks of mathematical analysis in a field containing infinitesimals (Robinson, 1966). In this theory, each subset S of R has a 'non-standard extension' $*S$ where $S \subseteq *S \subseteq *R$ and every function $f:S \to R$ extends to $*f:*S \to *R$ where $f(x) = *f(x)$ for all $x \in S$.

In simple cases, such as an open interval

$$S = \{x \in R \mid a < x < b\},$$

the extension satisfies the same formulae in the larger system:

$$*S = \{x \in *R \mid a < x < b\}.$$

In the same way, a polynomial function such as

$$f(x) = x^2 \ (x \in R)$$

is extended to

$$*f(x) = x^2 \ (x \in *R).$$

Other more general extensions are guaranteed by a theorem proved using the axiom of choice, essentially 'choosing' the extension to have the appropriate properties (e.g. Lindstrøm, 1988).

Appealing to the axiom of choice asserts the existence of new sets that we can attempt to imagine and talk about, even deriving their properties. For instance, the set of natural

numbers N has a natural extension $*N$ which has infinite elements. The new infinite elements in $*N$ are called (non-standard) infinite natural numbers. If N is an infinite natural number, then so are $N+1$, $N+2$, …; $N–1$, $N–2$, …; $2N$, $2N+1$, … and even bigger numbers such as N^2 or even N^N, and so on. This gives a new theory of infinite natural numbers whose properties are quite different from Cantor's cardinal infinities (Tall, 1980a). The non-standard natural numbers have a full and satisfactory arithmetic. If N is infinite, then $N+1$ is the 'next' natural number and is strictly bigger than N. There is a preceding natural number, $N–1$, and an infinitesimal reciprocal $1/N$. By extending the axioms of ordered fields using the axiom of choice, we therefore get a powerful new extension to the infinite that complements the notions of infinitesimals.

This theory can be used to give powerful formal images of ideas that already appeal to the intuitive senses. For instance, any sequence (a_n) of real numbers is a function from N to R and so can be extended to a function from $*N$ to $*R$. The limit of the convergent sequence (a_n) is then the standard part of a_N for any infinite element $N \in *N$.

This theory gives a new way of speaking about analysis. It does not give any new theorems in everyday analysis, but it can prove them more efficiently. It does, however, give new theorems about non-standard ideas that are not part of the traditional ε-δ approach.

For a function f defined in an open interval containing $x \in R$, then $*f(x+ε)$ can be defined for $x+ε \in *R$ where x is real and ε is infinitesimal. By computing the ratio $(*f(x+ε)–*f(x))/ε$ and taking its standard part, one can calculate the derivative of f at x. In the simple case $f(x)=x^2$, we have

$$\frac{*f(x+\varepsilon) - *f(x)}{\varepsilon} = \frac{(x+\varepsilon)^2 - x^2}{\varepsilon} = 2x + \varepsilon.$$

Taking the standard part suppresses the infinitesimal ε and gives the derivative as $2x$.

Likewise, any sequence (a_n) of real numbers is a function from N to R and so can be extended to a function from $*N$ to $*R$. If (a_n) is convergent, then the limit is defined to be the standard part of a_N for any infinite element $N \in *N$. This gives an alternative approach to the standard theory.

What's in a name?

The notation of non-standard analysis is seen by some as beginning to get messy and over-fussy. For every (standard) set or function we have a non-standard extension which we notate with a preceding '*'. The essential fact is that, restricted to the standard case, the '*' notation becomes redundant. So why introduce a *new* notation? Why not just *extend* the old one?

As ideas are extended, we invariably extend the old language to include the new. For instance, a 'number' may first be encountered meaning a 'natural number', then fractions, negatives, rationals, irrationals, even complex numbers, are all eventually called 'numbers'.

In the same way, the name 'real number' may be extended to encompass the set of non-standard real numbers. Where necessary we can distinguish between the everyday real numbers by calling them *standard* real numbers, and the extended set *non-standard* real numbers. With this linguistic change, the number line can be seen in two different ways. It can be visualised as the (standard) complete ordered field R, or as a full field of non-standard real numbers with infinitesimal and infinite detail. Using optical lenses, we can visualise appropriate detail of the full (non-standard) real line as a standard real picture. In this way, we obtain a visualization of a purely *formal* version of this extended real line in a manner that has echoes of the intuitions of the great mathematicians in history, still found in our students today. Of course, to construct the system formally requires the individual to work through the axiomatic construction of the hyperreals using the axiom of choice (e.g. Lindstrøm, 1988; Robert, 1988). Even then, individuals may not feel that the mental construction makes sense (including those who hold an intuitionist view). Those who do achieve the cognitive reconstruction can communicate to each other about the wonders of non-standard analysis. Others must be content with an informal concept image of these powerful ideas.

The limit concept

This paper began with the potential infinity of a recurring process and the idea of potential infinity and then constructed several distinct infinite extensions of the number system. Many in our current culture have the experience of ε-δ analysis and of Cantor's theory. The latter

has no role for infinitesimals, and this, wrongly, has led to a belief that there is something highly suspect about using infinitesimals. Non-standard analysis provides an alternative, which, whilst not concerning itself with the infinite cardinality of sets, has a full arithmetic of infinite integers and infinitesimal quantities.

It is therefore opportune to close this paper with a discussion of the limit concept in the light of these differing paradigms. The limit notion occurs in a number of different guises. It may involve the behaviour of a sequence of real numbers $a_1, a_2, \ldots, a_n, \ldots$ as n increases, the continuous limit of a function $f(x)$ as x approaches a particular value a, or a more sophisticated limit such as the Riemann sum of areas of rectangles under a curve. All of these have in common a *process* of getting arbitrarily close to a fixed value (the limit). In every case, the same symbolism is used both for the *process* of convergence and also for the *concept* of limit. For instance

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

evokes both the process of calculating the limit of the sum of the series and also for the numerical limiting value. Likewise

$$\lim_{x \to a} \frac{x^3 - a^3}{x - a}$$

also evokes both the process of calculating the limit as x gets close to a and also for the limiting value $3a^2$.

Mathematical concepts where a process and the output of that process are represented by the same symbol are called *procepts* (Gray & Tall, 1994). Limit procepts are very different from other procepts met previously by the student (Tall *et al*., 2001). Arithmetic procepts, such as the process of addition coupled with the concept of sum, or the process of multiplication and the concept of product, usually have a built-in (finite) algorithm of computation which readily give the answer. Algebraic procepts, such as the expression $3+2x$ have only a *potential* process of evaluation, which can only be performed explicitly when x is given a numerical value. Such a view of the expression as an incomplete process gives a sense of 'lack of closure'; an expression sensed as an incomplete process is not easily

conceived as a 'thing' that can be manipulated. The uneasiness felt by individuals often seems to lead them to learn procedural rules without understanding what is going on. However, even procedural rules do at least provide strategies in school algebra to simplify an expression or solve an equation.

The limit procept is even more subtle. Apart from a few simple cases such as the sum of a geometric series, there is no finite algorithm to compute the value of the limit. Instead, the process of convergence seems to be *potentially infinite*, often giving a sense of 'going on forever', 'tending to the limit', 'approaching the limit', but never seeming exactly to get there.

The literature abounds in describing cases of cognitive obstacles caused by sensing the infinite ongoing process rather than the finite limit value. (See the summaries in Cornu 1991; Tall, 1992). The idea of 'getting closer and closer' carries an embodied sense of 'nearly there, but not quite', giving a sense of an unattainable conclusion to the limit process. The natural idea of a limit therefore has built-in cognitive obstacles.

The formal idea, with its multiple layers of quantification, has other difficulties. For instance, the full definition of continuity of $f : D \to \boldsymbol{R}$ is:

$$\forall a \in D \; \forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in D : |x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon.$$

This is so complex that it is usually broken down into stages, first say that the function is continuous at the point a:

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in D : |x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon,$$

perhaps in a more 'friendly' linguistic form, such as:

given any $\varepsilon > 0$, there is a $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ whenever $|x - a| < \varepsilon$.

We then say that a function is *continuous* (on D) if it is continuous at every $a \in D$.

The value of δ here depends not only on ε, but also on the value of a. If a single value of δ can be found to satisfy every value of $a \in D$, the function is said to be *uniformly continuous* on D. This is formulated in the definition:

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x, a \in D : |x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon$$

The equivalent non-standard definition of continuity is of the form:

$$\forall a, x \in D : x - a \text{ infinitesimal} \Rightarrow f(x) - f(a) \text{ infinitesimal}.$$

If a is restricted to being a standard element of D, this is equivalent to *point-wise* continuity at a. Otherwise if x, a are both allowed to be non-stardard, then it is equivalent to the definition of *uniform* continuity.

The evident simplicity of this idea has led to a number of non-standard approaches to teaching calculus (eg Keisler 1976; Henle & Kleinberg, 1979; Lindstrøm, 1988; Robert, 1988). Various mathematics educators have tested such an approach and shown an apparent improvement in understanding (eg Sullivan, 1976; Artigue, 1991; Frid, 1994). Criticism of the use of the axiom of choice in the non-standard approach however, draws extreme criticism from those such as Bishop (1977) who insisted on explicit construction of concepts in the intuitionist tradition. The vast majority of mathematicians do not take such an extreme view, however, there remains an entrenched suspicion of infinitesimal methods in analysis. The main body of mathematicians seem to remain pragmatic formalists, solving problems in any way that gives a solution, then seeking the clarity and security of formal definitions and mathematical proof to give a logical structure to the relationships.

Summary

In this article, a distinction has been made between natural concepts of infinity, developed from experiences in the finite world, and formal concepts of infinity, built from formal definitions and deductions. Initial natural concepts often involve conflicting ideas, however, by building from different formal definitions, different formal notions of infinity can be constructed, each consistent in itself.

Two different methods of developing the formal theory are identified. 'Natural thinking' builds from concept imagery towards formalism. 'Formal thinking' builds from the concept definition, marginalizing imagery and focusing on logical deduction. Natural thinking is appropriate for thought experiments that suggest possible theorems; formal thinking is appropriate for establishing formal proofs. Individuals may prefer one approach or the other,

or use each method as appropriate. It was shown that, not only can natural thinking be used to imagine possible theorems, but formal thinking may also lead to 'structure theorems', whose properties may be used to develop more subtle visual imagery, now enhanced by a network of formal relationships.

The natural idea of the potential infinity of counting was axiomatized as a formal infinity defined by the Peano postulates. A structure theorem shows that this uniquely defines a single structure formalizing the natural numbers and reformulating potentially infinite repetitions as a finite application of the principle of induction. The focus then turned to the comparison between sets, extended to the formal theory of infinite *cardinals*. These have commutative and associative operations of addition and multiplication, but no corresponding operations of subtraction or division. The correspondence between well-ordered sets leads to the theory of infinite *ordinals*, again with operations of addition and multiplication. Although these are associative, they are no longer commutative.

An alternative focus on the familiar properties of arithmetic and order of number systems leads alternatively to the formal definition of an *ordered field*. Adding the axiom of completeness gives the notion of a complete ordered field. There are many different ordered fields, but (up to isomorphism) only one complete ordered field, namely the field of real numbers. A structure theorem shows that any ordered field contains (a subfield isomorphic to) the rational numbers. A second structure theorem about ordered fields that are bigger than the real numbers, shows that the finite elements are of the form $a+\varepsilon$ where a is real and ε is infinitesimal (possibly zero), and the infinite elements are of the form $1/\varepsilon$ where ε is infinitesimal. Such ordered fields may be visualised as number lines, with formal transformations (δ-lenses) that can be used to reveal infinitesimal or infinite detail.

Using the axiom of choice, it is possible to construct an extension ordered field of **R** that includes not only infinitesimals and infinite elements, but also additional structures such as infinite non-standard integers. The latter satisfy not only the usual rules of addition and multiplication, but, unlike cardinal and ordinal numbers, they can be subtracted, to give a standard or non-standard integers, and divided, yielding infinitesimals.

The moral of this tale is that there is not a single concept of infinity, but several distinct ones, and that it is not appropriate to assign to any one of these concepts the title of 'the' concept of infinity. In particular, it is not appropriate to consider infinite cardinal number theory (with its lack of subtraction and division) as the *only* mathematical infinity. There are other infinite concepts, including the intuitive infinities and infinitesimals in the calculus, that can be given formal theories of their own, consistent in themselves and yet *different* from the notion of infinite cardinal.

Notes

1. Elementary descriptions of intuitionism, formalism, logicism may be found on the World Wide Web, eg, at:

http://www.objectivistcenter.org/pubs/F_phimat.asp.

2. There are a number of elementary pages on the World Wide Web on Gödel's Incompleteness Theorem, which can be found by typing *Goedel, incompleteness* into a search engine. These currently include:

http://www.miskatonic.org/godel.html

and

http://pespmc1.vub.ac.be/ASC/INCOMP_THEOR.html.

References

Artigue, M., 1991: 'Analysis'. In D. O. Tall (Ed.), *Advanced Mathematical Thinking*, (pp.166–198), Dordrecht: Kluwer.

Berkeley, G., 1951: 'The Analyst or A discourse Addressed to an Infidel Mathematician Wherein it is examined whether the object, principles, and inferences of the modern Analysis are more distinctly conceived, or more evidently deduced, than religious Mysteries and points of Faith'. In A. A. Luce (Ed.), *The Works of George Berkeley Vol 4*, London: Thomas Nelson.

Bishop, E., 1967: *Foundations of Constructive Analysis*, McGraw-Hill.

Bishop, E., 1977: Review of 'Elementary Calculus' by H. J. Keisler, *Bulletin of the American Mathematical Society*, 83, 2, 205–208.

Cantor, G., 1895: *Beiträge zur Begründung der transfiniten Mengelehre*, (English trans.: *Contributions to the Founding of the Theory of Transfinite numbers*, 1915).

Cohen, P. J., 1966: *Set Theory and the Continuum Hypothesis*, New York: Benjamin.

Cornu, B., 1981: 'Apprentissage de la notion de limite: modèles spontanés et modèles propres,' *Actes due Cinquième Colloque du Groupe International P.M.E.*, Grenoble 322–326.

Cornu, B., 1991: 'Limits'. In D. O. Tall (Ed.), *Advanced Mathematical Thinking*, (pp.153–166), Dordrecht: Kluwer.

Duffin, J. M. & Simpson, A. P., 1993: 'Natural, conflicting and alien,' *Journal of Mathematical Behaviour*, 12, 4, 313–328.

Fischbein, E., Tirosh, D., & Hess, P. 1979: 'The intuition of infinity', *Educational Studies in Mathematics*, 10, 3–40.

Frid, S., 1994: 'Three approaches to undergraduate calculus instruction: Their nature and potential impact on students' language use and sources of conviction.' In E. Dubinsky, J. Kaput & A. Schoenfeld (eds.), *Research in Collegiate Mathematics Education I,* Providence, RI: AMS.

Gödel, K., 1931: 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme,' *Monatshefte für Mathematik und Physik*, vol. 38.

Gray, E. M., & Tall, D. O., 1994: 'Duality, ambiguity and flexibility: A proceptual view of simple arithmetic,' *Journal of Research in Mathematics Education,* 25 2, 115–141.

Henle, J. & Kleinberg, E., 1979: *Infinitesimal Calculus*, MIT Press, 1979.

Hilbert, D., 1900: *The Problems of Mathematics*. Address to the International Congress of Mathematicians, Paris. Translated into English by M. W. Newson, for the *Bulletin of the American Mathematical Society*, (1902), 8, 437–479.

Houdé, O, Zago, L., Mellet, E., Moutier, S. Pineau, A., Mazoyer, B, Tzourio-Mazoyer, N., 2000: 'Shifting from the perceptual brain to the logical brain: the neural impact of cognitive inhibition training', *Journal of Cognitive Neuroscience*, 1**2** 5, 721–728.

Keisler, H. J., 1976: *Elementary Calculus*, Prindle, Weber & Schmidt.

Kennedy, H. C. (ed.), 1973: *Selected works of Giuseppe Peano* (1858–1932) translated from the Italian. London: Allen and Unwin.

Kleiner, I., in press: 'The Infinitely Small and the Infinitely Large in Calculus'. *Educational Studies in Mathematics*. (This volume.)

Lakoff, G., 1987: *Women, Fire and Dangerous Things*, Chicago: University of Chicago Press.

Lakoff, G. and Johnson, M., 1999: *Philosophy in the Flesh*. New York: Basic Books.

Lakoff, G. & Nunez, R., 2000: *Where Mathematics Comes From*. New York: Basic Books.

Lindstrøm, T., 1988: 'An Invitation to Nonstandard Analysis.' In N. Cutland (ed.), *Nonstandard Analysis and its Applications*, (pp. 1–105). Cambridge: Cambridge University Press.

Peano, G., 1908: *Formulario Mathematico*. Torino: Fratelli Bocca.

Pinto, M. M. F., 1998: *Students' Understanding of Real Analysis*. Unpublished PhD Thesis, Warwick University.

Pinto, M., M. F. & Tall, D. O., 1999: 'Student constructions of formal theory: giving and extracting meaning,' *Proceedings of the Twenty third International Conference for the Psychology of Mathematics Education*, Haifa, Israel, 2, 41–48.

Pinto, M., M. F. & Tall, D. O., 2001: 'Following students' development in a traditional university classroom.' In Marja van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25ᵗʰ Conference of the International Group for the Psychology of Mathematics Education* 4, 57–64. Utrecht, The Netherlands.

Robert, A., 1988: *Nonstandard Analysis*, New York: Wiley.

Robinson, A., 1966: *Non-standard Analysis*. North Holland.

Stewart, I. N. & Tall, D. O., 1977: *Foundations of Mathematics*. Oxford: O.U.P.

Sullivan, K., 1976: 'The teaching of elementary calculus: an approach using infinitesimals,' *American Mathematical Monthly* 83, 370–375.

Tall D.O., 1980a: 'The notion of infinite measuring number and its relevance in the intuition of infinity', *Educational Studies in Mathematics*, 11, 271–284.

Tall, D. O., 1980b: 'Looking at graphs through infinitesimal microscopes, windows and telescopes', *Math. Gazette*, 64, 22–49.

Tall, D. O. & Vinner, S., 1981: 'Concept image and concept definition in mathematics, with special reference to limits and continuity', *Educational Studies in Mathematics*, 12, 151–169.

Tall, D. O., 1982: 'Elementary axioms and pictures for infinitesimal calculus', *Bulletin of the IMA*, 18, 43–48.

Tall, D. O., 1992: 'The Transition to Advanced Mathematical Thinking: Functions, Limits, Infinity, and Proof'. In D. A .Grouws (ed.) *Handbook of Research on Mathematics Teaching and Learning*, (pp. 495–511). New York: Macmillan.

Tall, D. O., Gray, E. M., Ali, M. B., Crowley, L. R. F., DeMarois, P., McGowen, M. A., Pitta, D., Pinto, M. M. F., Thomas, M. O. J. & Yusof, Y. B. M., 2001:. 'Symbols and the Bifurcation between Procedural and Conceptual Thinking', *Canadian Journal of Science, Mathematics and Technology Education* 1, 81–104.