Using Technology to Support an Embodied Approach to Learning Concepts in Mathematics

First ***Coloquio de Historia e Tecnologia no Ensino de Matemática***
at Universidade do Estado do Rio De Janiero,
February 21-3, 2002.

David Tall

Mathematics Education Research Centre
University of Warwick
CV4 7AL, UK
e-mail: david.tall@warwick.ac.uk

In this paper I will explain what I mean by an 'embodied approach' to mathematics. I shall contrast and compare it with two other modes: the 'proceptual' (manipulating symbols as process and concept) and the 'axiomatic' based on formal definitions and formal proof. Each of these has its own standard of 'truth'. I argue that the embodied mode, though it lacks mathematical proof when used alone, can provide a fundamental human basis for meaning in mathematics. I shall give examples of an embodied approach in mathematics, particularly in the calculus, using technology that makes explicit use of a visual and enactive interface.

PHYSICAL AND MENTAL TOOLS FOR THINKING

In his work over thirty years ago, long before the development of modern computers, Bruner (1966) focused on *homo sapiens* as a tool-using species.

> Man's use of mind is dependent upon his ability to develop and use "tools" or "instruments" or "technologies" that make it possible to express and amplify his powers. His very evolution as a species speaks to this point. It was consequent upon the development of bipedalism and the use of spontaneous pebble tools that man's brain and particularly his cortex developed. It was not a large-brained hominid that developed the technical-social life of the human; rather it was the tool-using, cooperative pattern that gradually changed man's morphology by favoring the survival of those who could link themselves with tool systems and disfavoring those who tried to do it on big jaws, heavy dentition, or superior weight. What evolved as a human nervous system was something, then, that required outside devices for expressing its potential.
> (Bruner, *Education as Social Invention*, 1966, p. 25.)

In his essay "Patterns of Growth", Bruner (1966) distinguished three modes of mental representation – the *sensori-motor*, the *iconic* and the *symbolic*.

> What does it mean to translate experience into a model of the world. Let me suggest there are probably three ways in which human beings accomplish this feat. The first is through action. […] There is a second system of representation that depends upon visual or other sensory organization and upon the use of summarizing images. […] We have come to talk about the first form of representation as **enactive**, the second is **iconic**. […] Finally, there is a representation in words or language. Its hallmark is that it is **symbolic** in nature. Bruner, 1966, pp. 10–11.

Tall, D (2003). Using Technology to Support an Embodied Approach to Learning Concepts in Mathematics. In L.M. Carvalho and L.C. Guimarães *História e Tecnologia no Ensino da Matemática*, vol. 1, pp. 1-28, Rio de Janeiro, Brasil.This paper consists of two parts: the first is a developing theory of three modes of thinking: embodied/proceptual/formal. The second part builds on long-established material from Tall (2000).

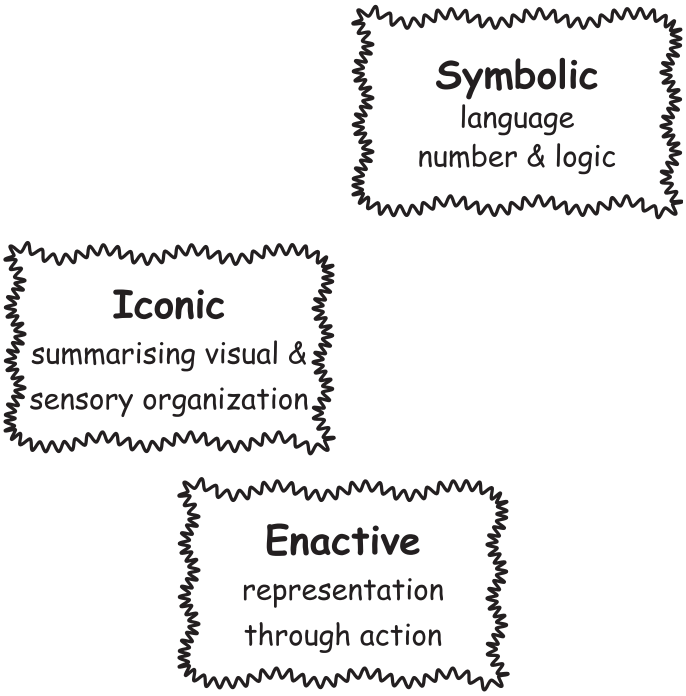

Figure 1: Bruner's three modes of representation

Bruner considered that these representations grow in sequence in the cognitive growth of the individual, first enactive, then iconic and finally the capacity for symbolic representation. He hypothesised that "any idea or problem or body of knowledge can be presented in a form simple enough so that any particular learner can understand it in recognizable form" (ibid. p. 44).

The development of modern computer interfaces show something of Bruner's philosophy in the underlying use of:

- Enactive interface,
- Icons as summarizing images to represent selectable options,
- Symbolism through keyboard input and internal processing.

When representations in mathematics are considered, clearly the single category of 'symbolism'—including both language and mathematical symbols—requires subdivision. Bruner (1966, pp. 18, 19) hinted at such a formulation in saying that that symbolism includes both "language in its natural form" and the two "artificial languages of number and logic." To these categories we must add not just number, but algebraic and other functional symbolism (e.g. trigonometric, exponential, logarithmic) and the wider symbolism of axiomatic mathematics.

The Reform movement in the calculus—for example the Harvard Calculus—focused initially on three representations: *graphic*, *numeric* and *symbolic* (or *analytic*):

> One of the guiding principles is the 'Rule of Three,' which says that wherever possible topics should be taught graphically and numerically, as well as analytically. The aim is to produce a course where the three points of view are balanced, and where students see each major idea from several angles.
>
> (Hughes Hallett 1991, p. 121)

The ‘Rule of Three’ later became the ‘Rule of Four’, extending the representations to include the *verbal*, giving four basic modes:

- verbal,
- graphic,
- numeric,
- symbolic (or analytic).

Several points are interesting here:

i) The enactive mode is completely omitted,

ii) The “verbal” mode was not seen initially as being important,

iii) Formal-axiomatic formulation using logical deduction is absent.

Each of these aspects is significant. The omission of the enactive mode is presumably because it does not seem to be a central focus in the graphs and symbols of the calculus. As I found with my earlier work on Graphic Calculus (1985a), this omission is a serious one because the embodied aspects of the calculus help to give fundamental human meaning. The initial omission of the verbal category is interesting. My interpretation is that this is because the verbal category is a fundamental ingredient underpinning all the other modes of operation. It can be highlighted, as it is in the Harvard Calculus, but it is essentially ever-present. Finally the down-playing of formal considerations is an implicit admission that there is something essentially difficult about this mode of operation that is not part of the calculus and more appropriately postponed for a formal course in analysis.

Taking these observations into account, I decided to categorise the modes of representation into three fundamentally distinct ways of operation:

- **Embodied**: based on human perceptions and actions in a real-world context including but not limited to enactive and visual aspects.
- **Symbolic-proceptual**: combining the role of symbols in arithmetic, algebra and symbolic calculus, based on the theory of these symbols acting dually as both process and concept (procept). (See Tall et al, 2001).
- **Formal-axiomatic**: a formal approach starting from selected axioms and making logical deductions to prove theorems. (Figure 2.)

My solution involves making choices, but I hope to show that the choices made can be justified by the fact that each category operates in a distinct manner, each with its own world of meaning and distinct methods of justification.

The *embodied* world is the fundamental human mode of operation based on perception and action. The *symbolic-proceptual* world is a world of mathematical symbol-processing, and the *formal-axiomatic* world involves the further shift into formalism that proves so difficult for many of our students. Language operates throughout all three modes, enabling increasingly rich and sophisticated conceptions to be developed in each of them.

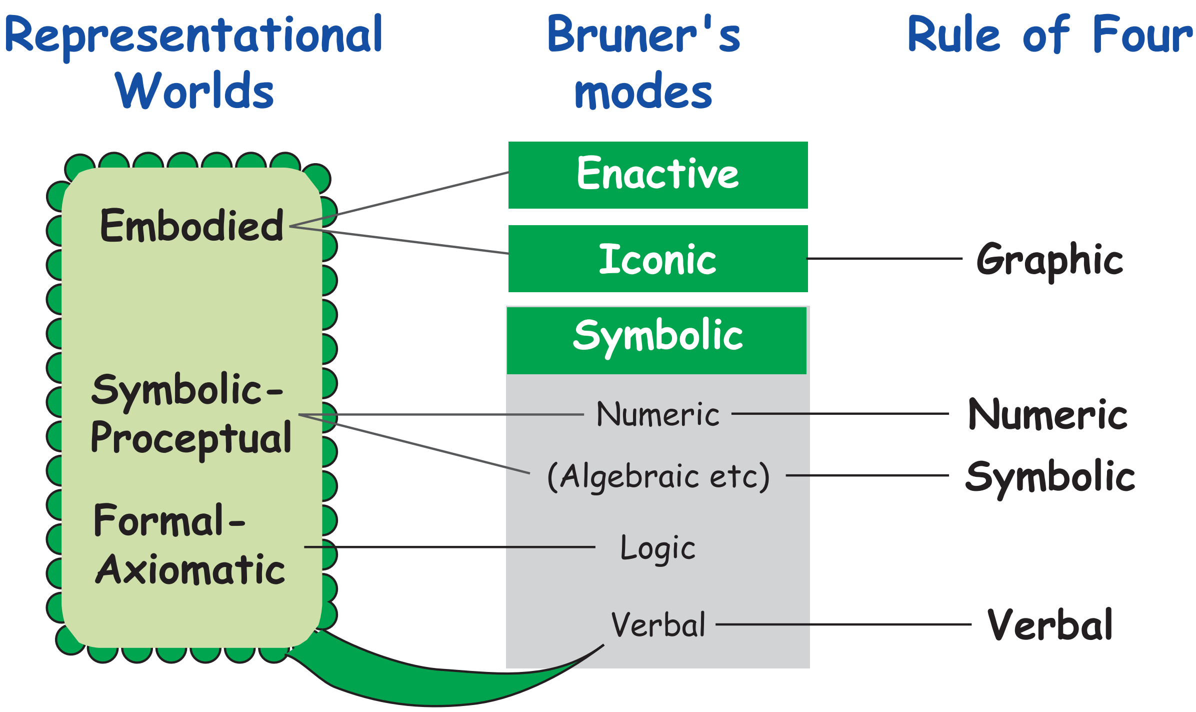

Figure 2: Three representational worlds and their links with other viewpoints

In using the term 'embodied', I am highly aware of the growing theories of 'embodied cognition' in cognitive science in the last two decades. In mathematics, a major contribution has been made by Lakoff and his colleagues. (Lakoff and Johnson, 1999, Lakoff and Nunez, 2000, Nunez et al 1999.) Embodied cognition focuses on the bodily/biological mechanisms underlying cognition and my work lies squarely in this broad scheme of ideas. However, Lakoff uses terms in a different way by asserting that *all* mathematics is embodied, meaning that it depends on constructions in human minds and shared meanings in mathematical cultures. I agree with this position. However, it reduces the power of the word 'embodied' because it refers to *all* mathematical thinking. I prefer to use the term 'embodied' to refer to thought built fundamentally on sensory perception as opposed to symbolic operation and logical deduction. This gives the term 'embodied' a more focused meaning in mathematical thinking.

Each world of operation incorporates a range of different aspects. The embodied mode includes enactive and iconic, and encompasses an increasingly subtle use of visual and spatial imagery. The proceptual mode contains several distinct stages (see Tall *et al* 2001). These include arithmetic calculations, algebraic manipulations and the potentially infinite notion of the limit concept with significant cognitive reconstructions necessary to cope with each successively sophisticated topic. The formal mode begins with an initial deductive stage based on embodied experience (for instance in Euclidean geometry) prior to building a full-blown systematic axiomatic theory.

Relationships with other theories

The subdivision into these three worlds of operation has links with a number of other theories. Piaget's stage theory incorporating

sensori-motor / preconceptual / concrete operational / formal

has a similar structure, though his theory is primarily developmental in origin.

The SOLO taxonomy (Structure of Observed Learning Outcomes) of Biggs and Collis (1982) formulated a subtly different view incorporating Bruner's ideas in terms of the following successive modes of operation:

sensori-motor / ikonic / concrete operational / formal / post-formal.

The SOLO taxonomy differs from that of Piaget, in that it is intended in to provide a template for *assessment*. Within each mode, the development of a specific concept is assessed as to whether the student's response is:

- *pre-structural* (lacking knowledge of the assessed component)
- *unistructural* (focusing on a single aspect)
- *multi-structural* (focussing on several separate aspects)
- *relational* (relating different aspects together)
- *extended abstract* (seeing the concept from an overall viewpoint).

Each mode, therefore, is not a single level of cognitive operation. It grows within the individual and, as each successive mode comes on stream sequentially in cognitive development, earlier modes continue to be available.

At a time when a student is learning mathematics, the sensori-motor and ikonic modes will already be available together and I have essentially combined them to give the *embodied mode*. Formal aspects of thinking in mathematics I have combined into the *formal-axiomatic mode*. This begins with local deduction (meaning 'if I know something … then I can deduce something else') and develops into global systems of axioms and formal proof.

Why *Three* Worlds of Operation?

The highly complex thinking processes in mathematics can be categorised in many ways. My choice of three categories puts together those aspects which have a natural relationship between them whilst allowing sufficient distinction to be of value. The embodied mode, for example, lies at the base of mathematical thinking. It does not stay at a low level of sensori-motor operation in the sense of the first stage of Piagetian development. It becomes more sophisticated as the individual becomes more experienced, while remaining linked, even distantly, to the perception and action typical in human mental processing. A 'straight line', for instance, is sensed initially in an embodied manner through perception and conception of a straight line given by a physical drawing. However, an embodied conception of a straight line may become more subtly sophisticated to cover the idea that a line has length but no breadth, which is a fundamental concept in Euclidean geometry. What matters here is that the conception of a ‘straight line’ remains linked to a perceptual idea even though experience endows it with more sophisticated verbal undertones.

The *proceptual* mode (beginning with Piaget’s concrete operational or SOLO’s concrete symbolic) is based on symbolic manipulation found in arithmetic, algebra and symbolic calculus. It could easily be subdivided, and often is. Research by my colleagues and myself suggest that there are a range of transitional difficulties that occur in moving through proceptual thinking to new kinds of proceptual symbolism (figure 3).

The symbols in arithmetic are *operational*, that is there is an algorithm for calculating the desired process. For instance, the symbol 3+2 is asking for the process of adding 3 and 2, which can be carried out, for example, by starting at 3 and counting on two more. There are several subtle difficulties that occur in handling broader number systems, such as integers (where adding a negative number will now make the result *smaller*, contrary to all experience with counting numbers) or fractions, with the difficulties of equivalence and the arithmetic of fractions.

Algebra has a new kind of symbol, such as $3+2x$, which is no longer *operational*; the desired sum cannot be carried out until x is known, and so the process of evaluation is only *potential*. Yet students are asked to manipulate expressions that have processes that they cannot carry out. No wonder students find the initial contact with algebra so confusing!

Other subtle difficulties occur in later algebraic developments. For instance,

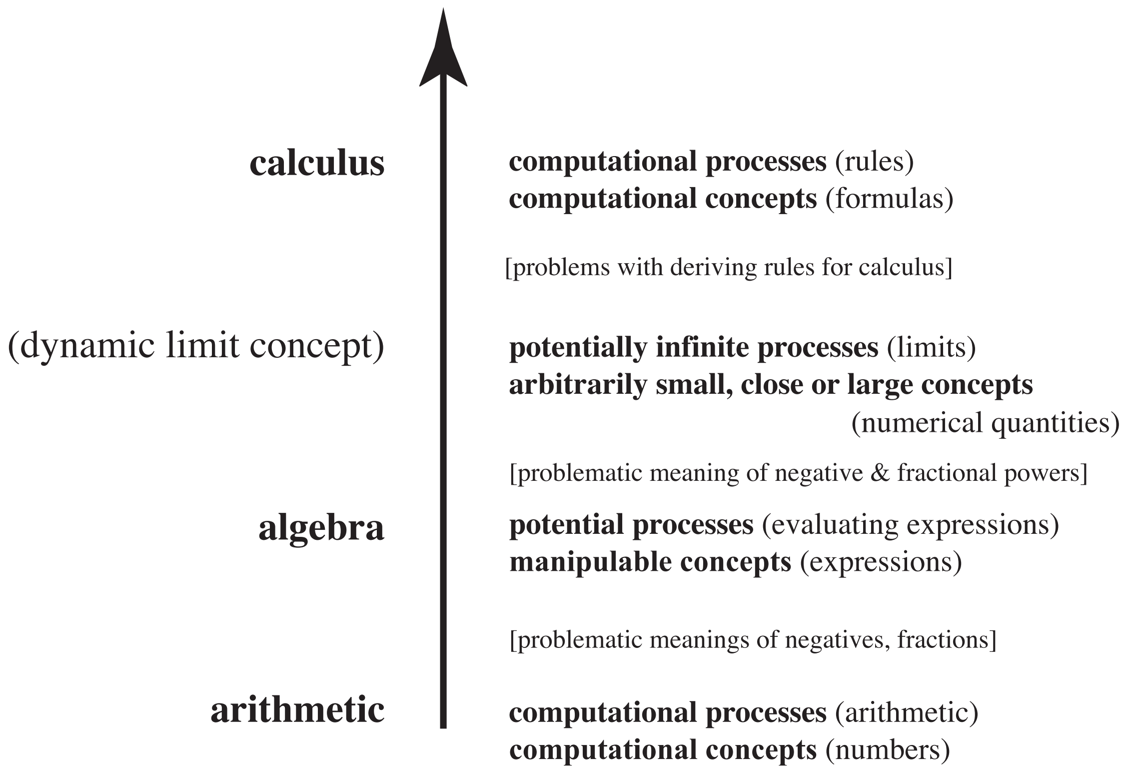

Figure 3: Some different types of procept in mathematics (Tall et al, 2001)

the meaning of x^2 as two lots of x multiplied together is evident, but $x^{1/2}$ has no corresponding meaning. (How can we have 'half a lot of x multiplied together?)

The introduction of the limit concept brings a procept whose calculation is now *potentially infinite*, so that most students believe that such a phenomenon 'goes on forever', without every quite reaching the limiting value. Again this provokes universal difficulties for students. It is with some relief that they find the 'rules of calculus' being *operational* again, albeit with symbolic rather than numerical input and output.

The development of number, algebra and limit are littered with cognitive transitions that require considerable effort for learners and often act as considerable barriers to progress. It is my belief therefore that it is more natural to put all of these numeric and symbolic manipulations together and to subdivide them into smaller categories wherever this is appropriate, taking into account natural transitions that occur in sense-making in developing meaning for symbolism in these branches of mathematics.

The final *axiomatic* category also includes a range of approaches. The earlier modes of thought already have their own proof structures. The embodied mode already supports thought experiments where one imagines a situation occurring and thinks through the consequences. The proceptual mode allows a simple form of proof by checking calculations, or using algebraic symbolism to generalise ideas in arithmetic. In the axiomatic world, *formal* proof comes into play, first in terms of local deductions of the form 'if I know this, then I know that.' What distinguishes the formal mode is the use of formal definitions for concepts from which deductions are made. The formal world again grows in sophistication from local deductions based on definitions into the formulation and construction of axiomatic systems such as those in group theory, analysis, topology, etc. Even here there is a range of ways in which students can come to terms with the formalism. Pinto (1998) distinguishes between 'natural' thinking where the formalism is built by continual refinement of the concept image and 'formal' thinking which builds logically from definitions and formal deductions. In Tall (2002), I take these ideas further to show how 'natural' thought experiments based on imagery may suggest possible theorems which may then be deduced by 'formal' means. In the other direction, formal proof can produce significant *structure theorems* that state that a given formal system has certain structural properties; these can yield their own imagery which allows natural thinking to speculate once more in terms of thought experiment.

DIFFERING MODES OF OPERATION AND BELIEF

The embodied, proceptual and formal modes have differing ways of justifying and proving which reveal them to operate as quite different worlds of experience. Let us consider this in a simple example:

Example: The sum of the first n whole numbers is $\frac{1}{2}n(n+1)$.

Proof 1: (embodied). Lay out rows of stones. Put 1 in the first row, 2 in the second row, 3 in the third, and so on. The picture is shown in the left hand part of figure 3. Now take an equal layout of pebbles, turn it round and fit the two together as in the right-hand picture. It can be seen that the two together make a rectangle size n by n+1, so there are $n(n+1)$ stones altogether, making $\frac{1}{2}n(n+1)$ in the original shape. The validity of this proof is in the visual picture.

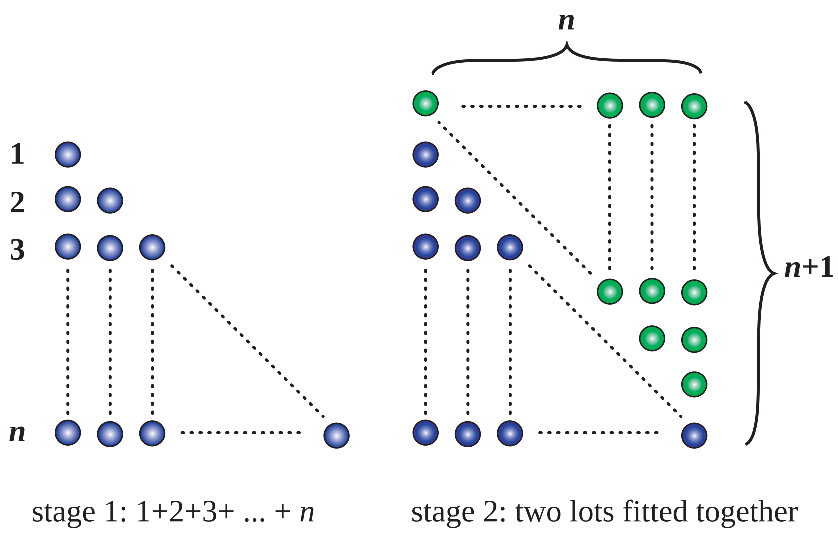

Figure 4: The embodied proof that the sum of the first n whole numbers is $\frac{1}{2}n(n+1)$

Proof 2: (proceptual). Write out the sum

$1+2+3+\ldots n$

backwards as

$n+\ldots +3+2+1$

and add the two together in order, pair by pair, to get

$(1+n) + (2+n-1) + \ldots + (n+1)$

to get n lots of n+1, ie. $n(n+1)$, so, again, the original sum is half this, namely $\frac{1}{2}n(n+1)$.

Proof 3: (*axiomatic*) By induction.

The embodied and proceptual proofs have clear human meaning, the first translating naturally into the second. The induction proof, on the other hand, often proves opaque to students, underlining the gap that occurs between the first two worlds and the formal world.

THE THREE DIFFERENT WORLDS OF THE CALCULUS

The term ‘calculus’ has its origins in the Latin word ‘calculus’ for ‘stone’, which was used as a physical tool during a process of calculation. In this context the theory of numbers has an underlying physical origin, for example in the notion of ‘triangular numbers’ or ‘square numbers’ which can be represented by triangular and square arrays of pebbles respectively. In the same way, calculus has a real world context in which the objects of study involve the rate of change of variable quantities (differentiation), the accumulation of growth (integration) and the relationship between them (the Fundamental Theorem).

The work of Newton and Leibniz moved us into a new realm by providing a mechanical method for calculating rates of change and cumulative growth. Thus calculus can be seen as operating in both a physical real-world sense, the source of our perceptions of change and growth, and in a symbolic sense, through problem-solving employing the algorithms of differentiation and integration. It is therefore a combination of the *embodied* and *proceptual* modes of operation.

In the last century, with the growth of mathematical analysis the formal-axiomatic mode of operation was developed in which numbers are no longer represented only as points on a number line, but are elements of an axiomatic structure, a complete ordered field. This gives a new *formal-axiomatic* framework for the calculus as part of the broader theory of mathematical analysis.

Traditional calculus teaching has focused on the graphical ideas of rate of change and cumulative growth, and the symbolic manipulation of the rules of calculus in differentiation and integration. The initial stages usually begin with informal ideas of the limit concept in geometric, numeric and symbolic form. It thus inhabits a combination of embodied and proceptual worlds, although the embodied world aspects are largely represented by static pictures rather than dynamic movement.

The arrival of the computer gives new possibilities: it has a graphical interface which allows the user to interact in a physical way by pointing, selecting and dragging objects onscreen to extend the embodied context of real-world calculus. Symbol manipulators such as *Mathematica*, *Maple* and *Derive* have the capacity to carry out the algorithms of the calculus on behalf of the user. However, these applications have a largely symbolic interface, producing graphic *output* on the screen, but with little embodied *input*. I contend that to give the calculus a physical human meaning, we should take advantage of an enactive interface which is now possible, and rethink the calculus to expand the standard graphic and symbolic modes of thought to take advantage of the full embodied mode on the one hand and to consider how this can lay the basis of a formal mode of thought for those students who will benefit from further study. Each mode brings its own viewpoint and its own mechanism for establishing truth and we continue our quest by considering each of these in turn.

DIFFERENT WARRANTS FOR TRUTH IN EACH WORLD

The three worlds of meaning have quite different ways of establishing truth. The **embodied world** is a world of *sensory* meaning. Its warrant for truth is that things behave predictably in an expected way.

The **proceptual world** is the familiar traditional world of calculus where calculations can be made (both arithmetic and algebraic). A graph has a slope (derivative) or an area (integral) because you can *calculate* it.

The **axiomatic world** is a world where explicit axioms are assumed to hold and definitions are given formally in terms of quantified set-theoretic statements. A function has derivative or integral because you can *prove* it.

If a computer is used, then the software can be programmed in a manner that supports these various modes. For instance, the *Visual Calculus* software programmed by Teresinha Kawasaki enables a graph to be magnified and moved under enactive control so that the user may zoom in to *see* the graph is *locally straight* and move the window along the graph to *feel* the changing slope of the graph (figure 5).

The embodied mode does not *prove* things are true in a mathematical sense, but it has the potential of building meaning far beyond traditional symbolic calculus. It is my belief that calculus software should be programmed to enable the user to explore not just nice smooth graphs, but graphs with corners, or more wrinkled graphs where the slope varies wildly. In this way an embodied approach can give a meaningful foundation for the most subtle of ideas of analysis. These ideas are currently omitted from most current calculus reform courses, which focuses mainly on the workings of regular functions. I ask how students can be expected to see the need for *proof* in a formal sense when they

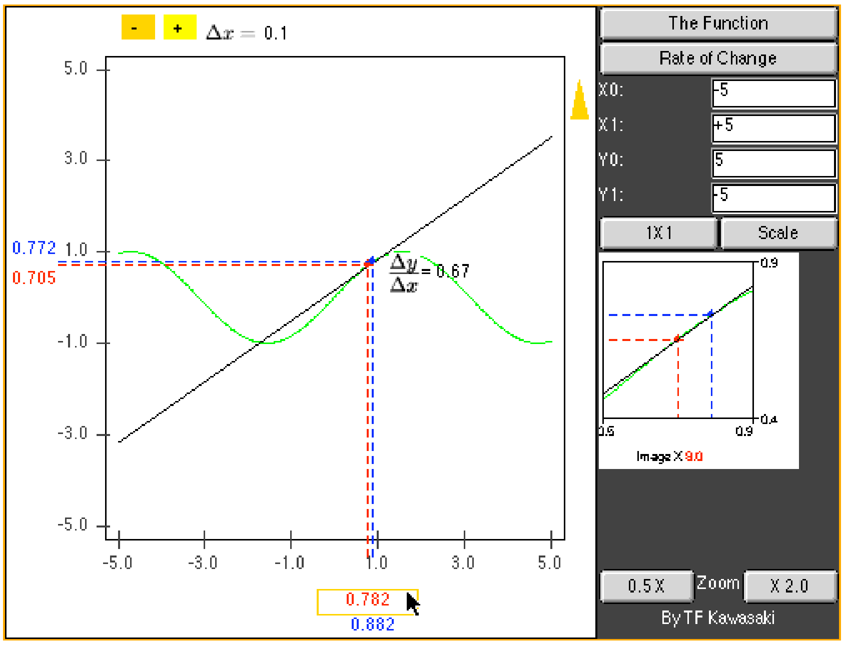

Figure 5: Dragging the view-point along the graph to see the changing slope of a locally straight graph

have no experience of what can go wrong in a meaningful context. My method allows students to *see* functions that are clearly *nowhere* locally straight, when classically trained students have no mental image as to what it means for a function to be non-differentiable.

AN EMBODIED APPROACH TO THE CALCULUS

An embodied approach to the calculus focuses on fundamental perceptual ideas before introducing any symbolism. It is *not* an approach that begins with formal ideas of limits, but with embodied ideas of graphical representations of functions. Nor is it an approach based *solely* on real-world applications although these are natural components of the total picture. It must encompass sensory ideas of the *mathematics* as well as sensory ideas of the applications. Each real world application involving say length, area, velocity, acceleration, density, weight etc has specific sensory perceptions that are *in addition* to the ideas of the calculus and these may cloud the underlying mathematics. For instance, if we build on the idea that the slope of a time-distance graph is a velocity, and the slope of a time-velocity graph is an acceleration, then we focus on the embodied senses of distance, velocity and acceleration rather than on the simpler underlying mathematics that each is obtained from the previous one as the slope of its graph.

The central idea of an embodied approach to the calculus builds on interaction with the physical picture of the graph of a function. It is important to emphasise that these functions involve variables that are *numbers*. The slope of such a graph is a variable *number*, the area under the graph is a *number* and slopes and areas themselves have graphs that are numerical quantities. Thus we may plot the graphs of the derivatives and ant-derivatives *on the same axes* if we so desire. In this way we may, for example, study the graphs of 2^x and 3^x to see how they have the same shape as their slope functions and seek a number e such that the slope of e^x is again e^x.

An embodied approach to the calculus is at its best when it links into the related world of symbolism, with its numeric calculations and algebraic manipulations to give the symbolic procedures of differentiation to calculate the slope of a graph. The derivative of a function is again a function, and (if this derivative is also locally straight) it can be differentiated again and again.

Where appropriate, I seek to motivate ideas in ways that can later be turned into axiomatic proofs. However, I see the theory of calculus fundamentally living in the two worlds of embodiment and proceptual symbolism.

In the remainder of this paper, I draw heavily on examples from my plenary lecture to the *Fifth Asian Technology Conference in Mathematics* (Tall, 2000).

COMPUTER ENVIRONMENTS FOR COGNITIVE DEVELOPMENT

Two concepts are useful in building an embodied approach to mathematics:

- a *generic organiser* is an environment (or microworld) which enables the learner to manipulate *examples* and (if possible) *non-examples* of a specific mathematical concept or a related system of concepts. (Tall, 1989.)
- a *cognitive root* (Tall,1989) is a concept which is (potentially) meaningful to the student at the time, yet contain the seeds of cognitive expansion to formal definitions and later theoretical development.

A cognitive root is usually an embodied concept. For instance, the notion of local straightness is a cognitive root for differentiation. This was first demonstrated in the program *Magnify* shown in figure 6 (Tall, 1985b), an early pre-cursor of the much more enactive *Visual Calculus* software of Kawasaki.

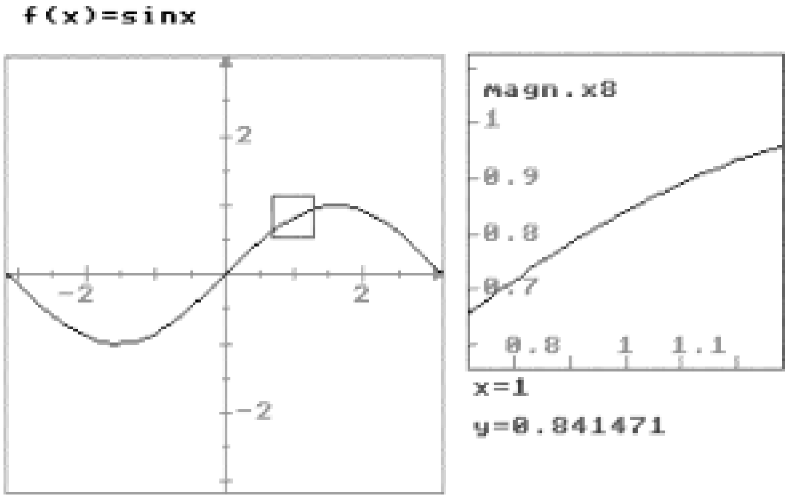

Figure 6: magnifying a graph that looks less curved when magnified
and, highly magnified looks 'locally straight'

The program *Magnify* is a generic organizer for the notion of local straightness. However, unlike almost all approaches to the calculus, which deal only with formulae for functions that are differentiable, this includes functions such as the *blancmange function* (figure 7), which is *nowhere* locally straight, thus fulfilling the need for a generic organizer to focus on *non-examples* as well as examples.

The blancmange function $bl(x)$ is the sum of saw-teeth:

$s(x) = \min(d(x), 1 - d(x))$, where $d(x) = x - \mathrm{INT}x$ is the decimal part of x,

$s_n(x) = s(2^{n-1}x)/2^{n-1}$,

and the function itself is:

$bl(x) = s_1(x) + s_2(x) + s_3(x) + \ldots$

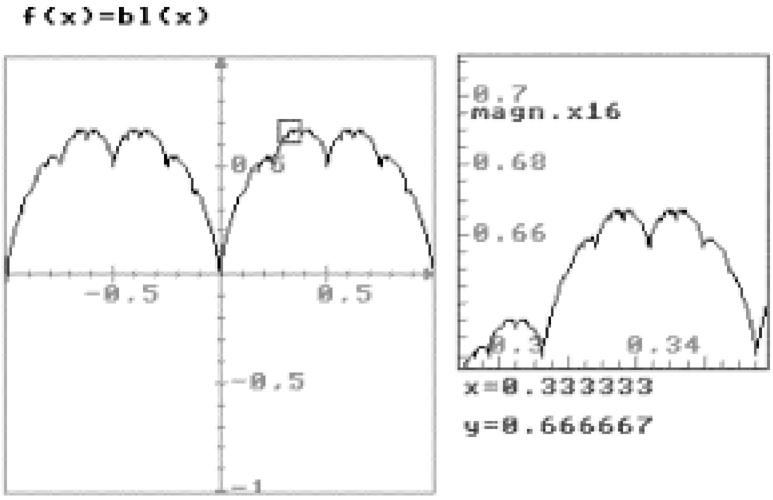

Figure 7: a graph which nowhere looks straight

If we consider the ‘nasty’ function:

$n(x)=bl(1000x)/1000$

then it is a tiny blancmange which is everywhere smaller than $\frac{1}{1000}$. The two graphs $g(x) = \sin x$ and $f(x) = \sin x + n(x)$ differ by less than $\frac{1}{1000}$ and yet one is locally straight everywhere and one is locally straight nowhere! (Figure 8.)

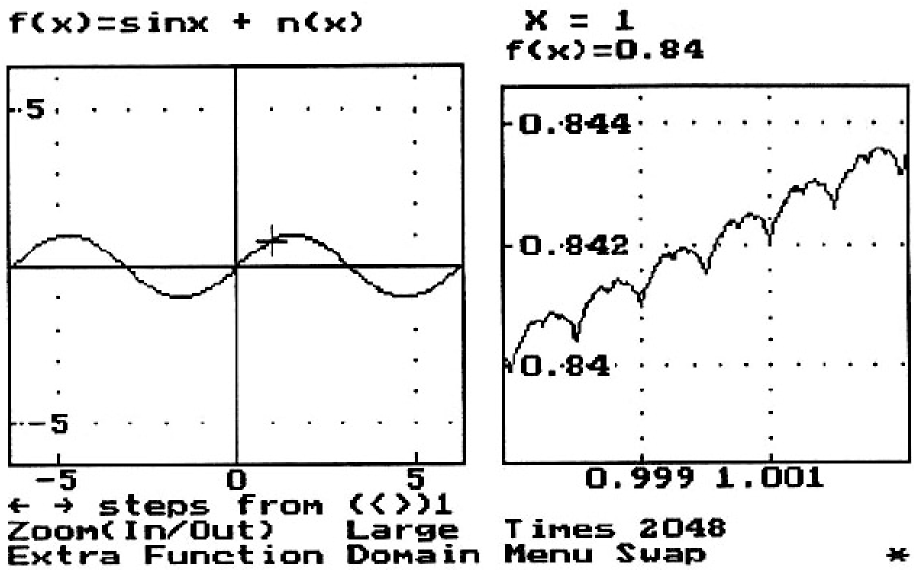

Figure 8. A ‘smooth-looking curve’ that magnifies ‘rough’.

This simple picture has an amazing consequence. If $f(x)$ is any differentiable function, then $f(x)+n(x)$ is *nowhere* differentiable, but looks exactly the same at one scale. The distinction only appears under higher magnification. Thus the software reveals its own limitations. A function may *look* straight in a given picture. What matters is that it must look straight at *all* magnifications. In this sense the generic organiser *Magnify* contains within it the visible evidence of its own limitations. It therefore has the potential to focus on the need for a more sophisticated mathematical theory.

EMBODIED LOCAL STRAIGHTNESS

Having stated categorically that I think that the formal limit notion is an entirely wrong place to begin the calculus (although it is precisely the right place to begin an axiomatic development in analysis), it is necessary to explain in what way a 'locally straight; approach to the calculus should begin.

Using appropriate software, a range of experiences can be arranged which lead to an embodied insight into calculus concepts. These include:

a) zoom in under enactive control to sense the lessening curvature and establish local straightness by sensing it 'happen'.

b) drag a magnification window along a locally straight graph to see its changing slope.

c) explore 'corners' (with different left and right slopes) and more general 'wrinkled' curves to sense that not all graphs are locally straight.

d) use software to draw the slope function to establish visual relationship between a locally straight function and its slope expressed symbolically.

e) Consider visual slope functions of sin*x*, cos*x*, and 'explore' the minus sign that arises in the derivative of cosx, which is –sin*x* and is visibly the graph of sin*x* reflected in the *x*-axis.

f) Explore 2^x, 3^x and vary the parameter *k* in k^x to find a value of *k* such that the slope of k^x is again k^x.

EMBODIED LOCAL STRAIGHTNESS AND MATHEMATICAL LOCAL LINEARITY

There are great cognitive and mathematical differences between local straightness and local linearity. 'Local straightness' is a primitive human perception of the visual aspects of a graph. It has global implications as the individual looks *along* the graph and sees the changes in gradient, so that the gradient of the whole graph is seen *as a global entity*.

Local linearity is a *symbolic linear approximation* to the slope *at a single point* on the graph, having a linear *function* approximating the graph at that point. It is a *mathematical* formulation of slope, taken first as a limit at a point *x*, and only then varying *x* to give the formal derivative as a function. Local straightness *remains at an embodied level* and links readily to visualising the slope of a given graph. Local linearity focuses on the 'best' local linear approximation expressed symbolically.

For instance, the derivative of cos*x* is seen to be equal to –sin*x* in the embodied mode because 'it is the graph of sin*x* upside down' (figure 9). This does not mean that this is a *proof* in a formal sense. However, the symbolic proof of the derivative by finding the limit of $(\cos(x+h)-\cos x)/h$ as *h* tends to zero is rarely convincing to students in my experience. In practice, it is based on the use of trigonometric formulae which are not 'proved' symbolically at this stage and on an *ad hoc* argument (usually presented visually) that $\sin x/x$ tends to

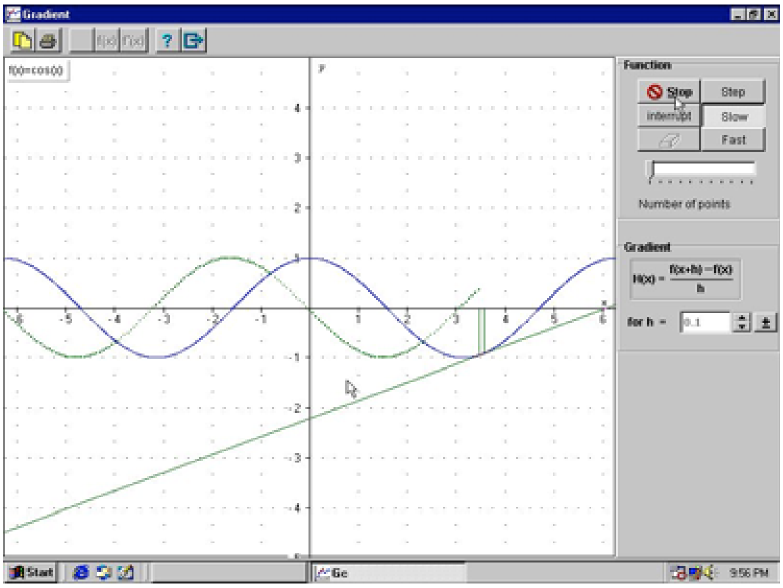

Figure 9: The gradient of cos*x* (drawn with Blokland et al, 2000).

1 as *x* tends to zero. I would contend that an embodied experience with meaning is more appropriate at this point. The time for the more manipulative and formal aspects can come at a later stage when they have more chance of making sense.

My own belief is that a locally straight approach is an option that is appropriate for the widest spectrum of students. It is:

- an 'embodied approach' which can be supported by enactive software to give it a human meaning.
- it can be linked directly to the usual numeric and graphic derivatives.
- it fits exactly with the notion of local straightness, which can be linked to local linearity (for those for whom it is important).
- it involves visual and symbolic ideas which can later be linked to formal analysis in either standard, or non-standard, form.

LOCAL LINEARITY AND THE SOLUTION OF DIFFERENTIAL EQUATIONS

As an example of the distinction between the embodied notion of local linearity and the symbolic-proceptual notion of local straightness, consider the inverse problem to the finding of a derivative. Mathematicians for many generations have used the fundamental theorem of the calculus to declare that the inverse of differentiation is integration. This is a conceptual blunder. The inverse of finding the slope of a function is to be given the slope of a function and to be asked how to find a function with this slope. The inverse of differentiation is *anti-differentiation*: given the derivative, find the function. In traditional calculus this is given in terms of linear differential equations in the form

$$\frac{dy}{dx} = F(x,y).$$

In traditional symbolic calculus this is attacked by a rag-bag of specific techniques suitable for a small number of types of differential equation. The meaning is (usually) lost. But the *embodied* meaning is plain. It is this:

> If I point my finger at any point (x,y) in the plane, then I can calculate the slope of the solution curve at that point as $m = F(x,y)$ and draw a short line segment of gradient m through the point (x,y).

This is a perfect opportunity to design a generic organiser on the computer. Simply write a piece of software so that when the mouse points at a point in the plane, a short line segment of the appropriate gradient is drawn, and as the mouse moves, the line segment moves, changing its gradient as it goes. As the solution curve is locally straight—it has a slope given by the equation—this line segment is *part* of the solution (at least, it *approximates* to part of a solution). The software allows the segment to be left in position by clicking the mouse. Hence by pointing and clicking, then moving the line segment until it fits with the end of the curve drawn so far, an approximate solution curve can be constructed by sight and hand-movement—an embodied link between a first order differential equation and its solution (figure 10.)

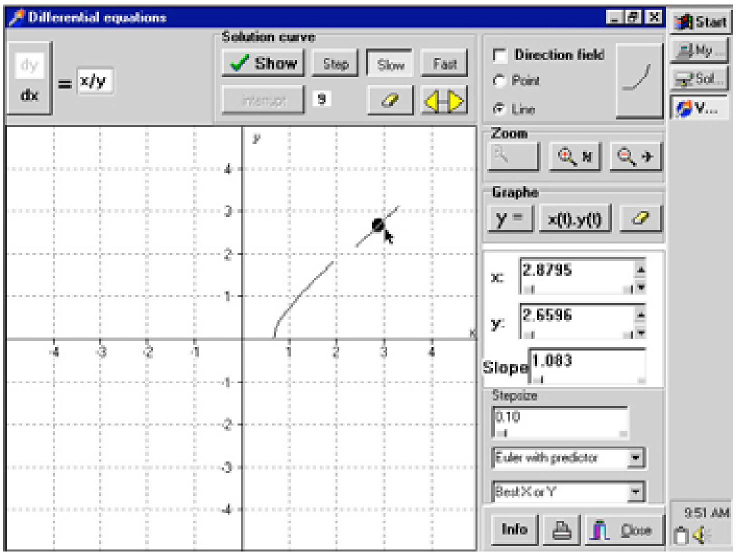

Figure 10: A generic organiser to build a solution of a first order differential equation by hand, (Blokland *et al*, 2000).

CONTINUITY

Tall (1985a) showed how the notion of continuity can be illustrated for a real function. All that is required is to stretch the graph much more horizontally than vertically. In figure 11 we see the blancmange function with a rectangle that is tall and thin. This is stretched to give the picture in figure 12. It can be seen that the graph 'pulls flat' and that further stretching will flatten it horizontally. The translation from this embodied notion of continuity to the formal definition is not very far. Imagine the graph is drawn in a window with $(x_0, f(x_0))$ in the centre of the picture, in the centre of a pixel height 2ε. Suppose it 'pulls flat". Then the graph lies in a horizontal row of pixels and if the window is now of width 2δ, we have:

$$|x - x_0| < \delta \text{ implies } |f(x) - f(x_0)| < \varepsilon \text{ [QED].}$$

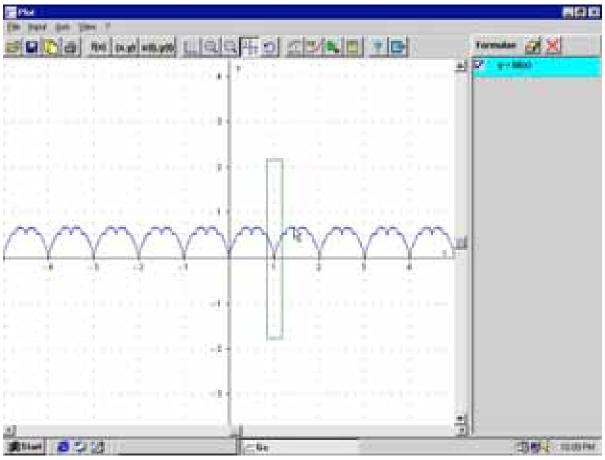

Figure 11: The blancmange graph and a rectangle to be stretched to fill the screen

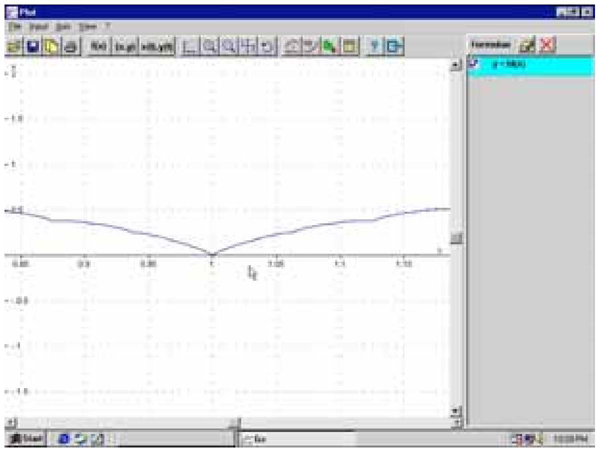

Figure 12: The blancmange function being stretched horizontally.

THE EMBODIED NOTION OF AREA

In the embodied world, the area under a continuous curve can be *seen* and calculated approximately by covering the area with squares and counting them. (Figure 13.)

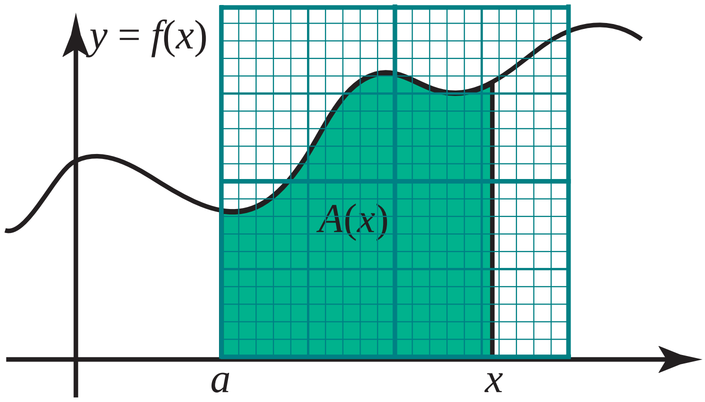

Figure 13: Measuring the area under a graph with a grid

The area from a to x under the graph is a function $A(x)$ called the 'area-so-far' function. In the practical embodied world of physical measurement, by using small enough squares, a numerical value of the area can be found to a degree of accuracy limited only by the accuracy of drawing and measuring.

Just as the cognitive root of 'local straightness' can be used to lead to more sophisticated theory, so the embodied notions of 'area' and 'area-so-far' can support Riemann and even Lebesgue integration. The use of technology to draw strips under graphs and calculate the numerical area is widely used. With a little imagination, and well-planned software, it can be used to give insight into such things as the sign of the area (taking positive and negative steps as well as positive and negative ordinates) and to consider ideas such as how the notion of continuity relates to the notion of integration. For instance, figure 14 shows the area under $\sin x$ from 1 to 1.001 with the graph stretched horizontally. It shows that the increase in area is approximately $f(x)$ times the change in x. Thus the

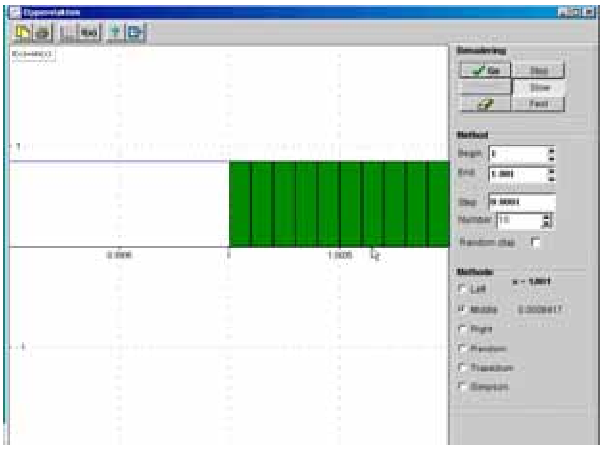

Figure 14: Area under sin*x* from 1 to 1.001 stretched horizontally

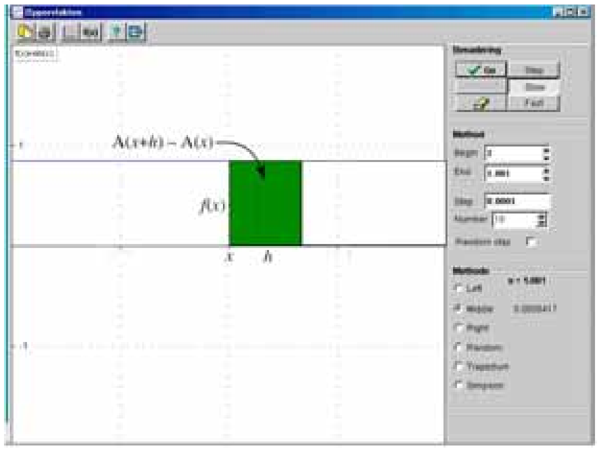

Figure 15: Towards the Fundamental Theorem of Calculus

ratio of change in area to change in *x* approximates to *f*(*x*), which gives insight into the fundamental theorem of calculus that the rate of change of area is the original function.

FORMALIZING THE EMBODIMENT OF THE FUNDAMENTAL THEOREM OF CALCULUS

The embodied idea of continuity leads naturally to a formal proof of the Fundamental Theorem. Let $A(x)$ be the area under a continuous graph over a closed interval $[a,b]$ from a to a variable point x. In the embodied mode, the area 'exists' because it can be seen and calculated as accurately as required. Continuity means the graph 'may be stretched horizontally to "look flat".' (Figure 15.) This means that:

> Given an $\varepsilon > 0$, and a drawing in which the value $(x_0, f(x_0))$ lies in the centre of a practical line of thickness $f(x_0) \pm \varepsilon$, then a value $\delta > 0$ can be found so that the graph over the interval from $x_0 - \delta$ to $x_0 + \delta$ lies completely within the practical line.

Then (for $-\delta < h < \delta$), the area $A(x+h)$–$A(x)$ lies between $(f(x)-\varepsilon)h$ and $(f(x)+\varepsilon)h$, so (for $h \neq 0$),

$$\frac{A(x+h) - A(x)}{h} \text{ lies between } f(x)-\varepsilon \text{ and } f(x)+\varepsilon.$$

For $|h| < \delta$, we therefore have:

$$\left|\frac{A(x+h) - A(x)}{h}\right| < \varepsilon\,.$$

As ε is arbitrary, this shows that the embodied idea of continuity leads to a corresponding formal definition and to a formal proof of the Fundamental Theorem of Calculus.

FURTHER EMBODIED INSIGHT INTO FORMAL THEORY

I conclude this paper by showing a few visual examples of various sophisticated concepts in mathematical analysis.

The blancmange function is continuous (Tall, 1982), and therefore its area function is differentiable. Figure16 shows the numerical area function for the blancmange and the *gradient of the area function*. This looks like the original graph. Of course it does, because the derivative of the area is the original function again.

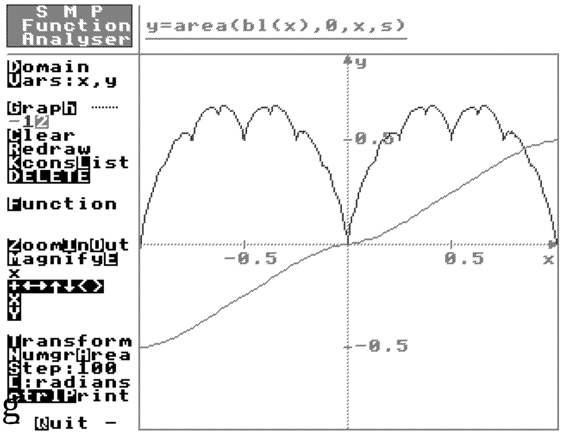

Figure 16: the area function of the blancmange and the derivative of this area (from Tall, 1991b)

Another much more interesting situation is to consider the 'area' under a function which has a number of discontinuities. The function *x*–int(*x*) is discontinuous at each integer and is continuous everywhere else. The area function is continuous everywhere and is also differentiable everywhere that the original function is continuous (figure 17). However, at the integer points, if the graph of the area function is magnified, it can be seen to have a corner at each integer point, because here the area graph has different left and right gradients (figure 18). If you look at the change in the area under the function you may be able to *see* why this happens.

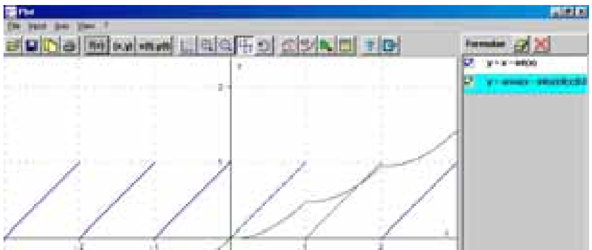

Figure 17: the area function for *x–int*(*x*)

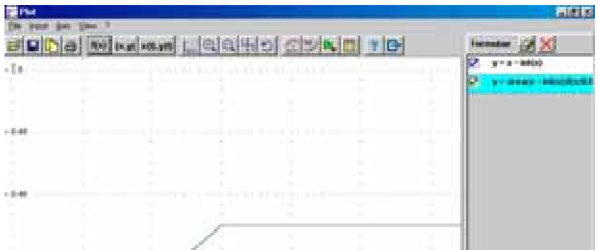

Figure 18: The area function magnified

It was an ambition of mine to draw functions such as *f(x)=x for x rational, f(x)=1–x for x irrational*. The fact that this was impossible for numerical calculations on a computer (which are all rational) did not deter me. In Tall (1993), I found a method that enabled me to make such a model.

The Ancient Greeks used an algorithm to find rational approximations to any (real) number *x*. It begins by finding the integer part *n*, and decimal part *d*:

$$x = n + d \quad (\text{where } 0 \le d < 1).$$

If d=0, then x is a (rational) integer. If not, the subtle part is to note that its reciprocal $1/d$ is greater than 1, so we can take the integer part again and write

$$1/d = n_2 + d_2 \text{ (where } 0 \le d_2 < 1).$$

By continuing this process, the equations can be unravelled to give closer and closer rational approximations to any number x. For instance,

$$\pi = 3 + d \text{ (where } d = 0.14159...)$$

$$1/d = 7.0626...$$

and so a good approximation to π is

$$\pi = 3\tfrac{1}{7} = \tfrac{22}{7}.$$

If the process is applied to a rational number such as $\tfrac{22}{7}$, then the remainder eventually becomes zero:

$$\tfrac{22}{7} = 3 + \tfrac{1}{7}$$

$$1/(\tfrac{1}{7}) = 7+0$$

and the process terminates.

The process gives a sequence of fractions, r_1, r_2, ... which tend to the real number x. If x is rational, the sequence is eventually constant, equalling x expressed in lowest terms. If x is irrational, it is easy to see that the numerators and denominators of r_n must grow without limit. (For if the denominators were all less than an integer N, then the sequence $N!r_n$ would be a sequence of *integers* tending to $N!x$, so the terms must eventually be a fixed integer, implying $N!x$ is an integer, contradicting the fact that x is irrational.)

This gives a method of distinguishing between rationals and irrationals:

> Compute the continued fraction expansion of x. If the rational approximations have denominators that grow without limit, then x is irrational, otherwise it is rational.

Working in the practical world of computers there are technical difficulties. The process involves taking reciprocals; if d is small, then $1/d$ is huge. If d should be zero, but errors make it tiny, then taking the reciprocal causes the method to blow up. The practical way out is to cease when the process gives a decimal part smaller than a specified error e, and check if the size of the denominator of the approximating fraction is bigger than a specified (large) number K.

Using this idea, I formulated the following technical definition to simulate the notions of rational and irrational in a finite computer world:

> **Definition**: A real number x is said to be *(e,K)-rational* if, on computing the continued fraction approximation to x, the first rational approximation within e of x has denominator less than K, otherwise x is said to be *(e,K)-irrational*.

The pseudo-code, returning TRUE for pseudo-rationals and FALSE for pseudo-irrationals, translates easily to most computer languages, is as follows:

```
DEFINITION rational(x,e,K)
r=x : a1=0 : b1=1 : a2=1 : b2=0
REPEAT: n=INTr : d=r–n : a=n*a2+a1 : b=n*b2+b1
        IF d<>0 THEN r=1/d : a1=a2 : b1=b2 :a2=a : b2=b
UNTIL ABS(a/b–x)<e
IF b<K THEN return TRUE ELSE return FALSE
```

Sensible values for e and K are, say, e $=10^{-8}$, K = 10^4 for single precision arithmetic, or e $=10^{-16}$, $N = 10^8$ for double precision.

Such an algorithm allows us to subdivide numbers into two disjoint sets numerically, which I called 'pseudo-rational' and 'pseudo-irrational'. In Tall (1991) I programmed a routine plotting random points, which were mainly 'pseudo irrational' and a second routine that plotted mainly 'pseudo-rationals'.

Figure 19 shows pictures of the function which is x on the rationals and $1-x$ on irrationals together with a graph for the area 'under the graph' from 0 to x. This uses the mid-ordinate rule with a fixed with (rational) step. It encounters mainly (pseudo-) rationals where $f(x)=x$, so the resulting area function approximates to $x^2/2$. When the area is calculated using a random step-length and a random point in the strip to calculate the area, it encounters mainly (pseudo-) irrationals where the function has values $f(x) = 1-x$. The area function drawn in this case reflects the latter formula (figure 20). (Here I have drawn several plots of the area curve. Because of the errors calculating pseudo-rationals and irrationals, there are small discrepancies with the random area that is slightly different each time.)

I used this software to discuss the area under such graphs (Tall, 1993). Students who were not mathematics majors and who would normally not cope very well in an analysis course were able to discuss this example intelligently, noting that 'a random decimal is highly unlikely to repeat, so random decimals are almost certainly irrational'. This led to a highly interesting discussion on the 'area' under 'peculiar' graphs which offered subtle ideas that could move on to Lesbegue integration. It shows how the mathematical mind can gain insights from visuo-spatial ideas in areas where the formal theory would be far too abstruse. But, for some of those who later do go on to the formal theory, visualization can provide a powerful cognitive foundation.

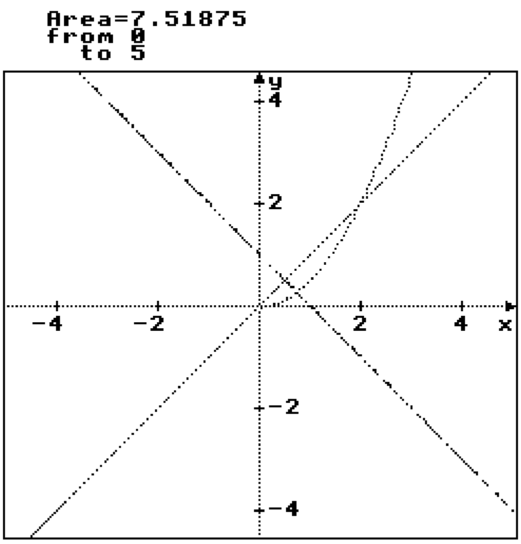

Figure 19: The (pseudo-) rational area

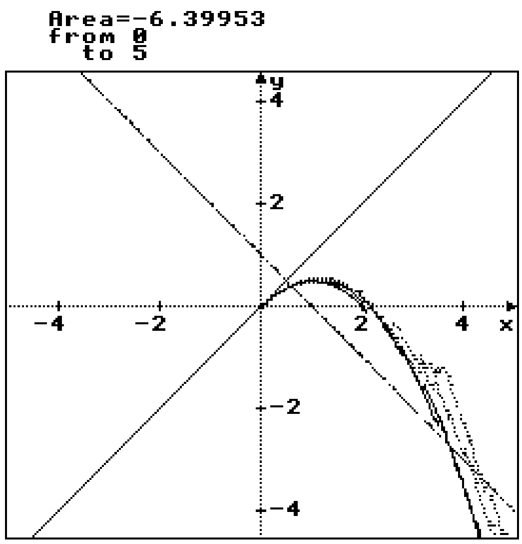

Figure 20 : The (pseudo-) irrational area

REFLECTIONS

In a recent review of the use of technology in calculus, Tall, Smith and Piez (in press) found that many projects (such as the *Harvard Calculus* and *ProjectCalc*) based their ideas on the use of computer representations with various graph programmes and symbol manipulators. However, as we saw earlier, such curricula did not in general consider the underlying embodied ideas of the type presented here. One might wonder why. In my plenary talk to the International Congress of Mathematics Education in 1996 (Tall, 1998), I noted that the sequence in which technology developed caused the pioneers to switch successively to new facilities. The sequence included:

- Programming numerical algorithms (pre-1980),
- Graphics (early 1980s), eg using graph-plotting programs,
- Enactive control (1984), allowing interactive exploration (eg Cabri),
- Computer algebra systems (early 80s, generally available in the late 80s),
- Personal portable tools (1990s) (eg TI-92, PDAs, portables, iBooks with wireless, etc),
- Multi-media (1990s),
- The World Wide Web (1990s).

Constant innovation caused new ideas to be implemented. Mathematicians naturally wanted the latest and “best” tools. Thus new tools took over before the use of earlier tools had been fully worked out. The fledgling use of numeric programming and graphic visualisation was overtaken by the power of computer algebra systems at a time when the power of an enactive interface was still to be fully understood. Now we have had time for reflection I suggest that an embodied approach provides a particularly human foundation for ideas in calculus and analysis that can be a study in itself but can lead on naturally to proceptual-symbolic calculus and formal-axiomatic analysis.

SUMMARY

In this presentation I have described three distinct worlds of human operation, the embodied mode based on human perception and sensation, enhanced by verbal theorising and communication, the proceptual-symbolic world of arithmetic, algebra and functions in calculus, and the formal world of mathematical analysis. I have made the case that a combination of embodied and proceptual operation is appropriate for the calculus and the formal mode can be postponed to a later study of analysis. However, I have also been able to show that extremely deep ideas in mathematical analysis have a cognitive foundation in the embodied mode in a manner which is meaningful to a much broader spectrum of students.

I reported how the use of local straightness and visual ideas of area can be cognitive roots that are foundational in building an embodied understanding of the calculus, but have the potential to grow naturally into the formal theory of analysis.

In particular, I have presented arguments in support of the following hypotheses:

- Local straightness is an embodied foundation (cognitive root) for the calculus.
- The local slope of the graph as rate of change is an embodied foundation (cognitive root) for the slope function (derivative).
- Finding a graph given its slope is an embodied foundation (cognitive root) for differential equations. This is the true inverse operation to differentiation.
- The notion of area under a graph is an embodied notion that can be calculated to suitable accuracy by embodied methods.
- Local flatness (stretching a graph horizontally) is a cognitive foundation for continuity.
- The embodied notion of continuity leads naturally to both an embodied idea and a formal proof of the fundamental theorem of calculus showing that the derivative of the area-so-far function is the original function.

In short, an embodied approach to the notion of change and rate-of-change of quantities represented by graphs has the necessary conceptual power to lead to a potentially meaningful theory of:

☆ *proceptual symbolism in calculus*

and

☆ *axiomatic proof in analysis.*

References

Biggs, J. & Collis, K. (1982). *Evaluating the Quality of Learning: the SOLO Taxonomy.* New York: Academic Press.

Blokland, P., Giessen, C, & Tall, D. O. (2000). *Graphic Calculus for Windows*. (see www.vusoft.nl.)

Bruner, J. S. (1966). *Towards a Theory of Instruction*. Cambridge: Harvard.

Hughes Hallett, D. (1991). 'Visualization and Calculus Reform'. In W. Zimmermann & S. Cunningham (eds.), *Visualization in Teaching and Learning Mathematics*, MAA Notes No. 19, 121-126.

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

Lakoff, G. and Núñez, R. E. (2000). *Where Mathematics Comes From*. New York: Basic Books.

Núñez, R. E., Edwards L. D, & Matos, J. F., (1999). Embodied cognition as grounding for situatedness and context in mathematics education, *Educational Studies in Mathematics* **39**, 45–65, 1999.

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

Tall, D. O., (1982). The blancmange function, continuous everywhere but differentiable nowhere, *Mathematical Gazette,* **66,** 11–22.

Tall, D. O. (1985a). Understanding the calculus, *Mathematics Teaching*, **110**, 49–53.

Tall, D. O. (1985b). *Graphic Calculus* (for the BBC Computer, now superseded by Blokland et al (2000) for windows, above.

Tall, D. O. (1989). Concept Images, Generic Organizers, Computers & Curriculum Change, *For the Learning of Mathematics*, **9**,3, 37–42.

Tall, D. O. (1991). *Function Analyser*. Cambridge University Press. (Software for the Archimedes computer, now superseded by Blokland et al (2000) for PC.)

Tall, D. O. (1993). Real Mathematics, Rational Computers and Complex People. In Lum, L. (ed.), *Proceedings of the Fifth Annual International Conference on Technology in College Mathematics Teaching*, 243–258.

Tall, D. O. (1998). Information Technology and Mathematics Education: Enthusiasms, Possibilities & Realities. In C. Alsina, J. M. Alvarez, M. Niss, A. Perez, L. Rico, A. Sfard (Eds), *Proceedings of the 8th International Congress on Mathematical Education*, Seville: SAEM Thales, 65-82.

Tall, D. O. (2000). Biological Brain, Mathematical Mind & Computational Computers (how the computer can support mathematical thinking and learning). In Wei-Chi Yang, Sung-Chi Chu, Jen-Chung Chuan (Eds), *Proceedings of the Fifth Asian Technology Conference in Mathematics*, Chiang Mai, Thailand (pp. 3–20). ATCM Inc, Blackwood VA..

Tall, D. O. (2002). Natural and Formal Infinities. *Educational Studies in Mathematics* (in press).

Tall, D. O, Gray, E M., Ali, M., Crowley, L., DeMarois, P., McGowen, M., Pitta, D., Pinto, M., Thomas, M., &d Yusof, Y. (2001). Symbols and the Bifurcation between Procedural and Conceptual Thinking, *Canadian Journal of Science, Mathematics and Technology Education* 1, 80–104.