Mathematics Education Research Journal 2008, Vol. 20, No. 2, 5-24

The Transition to Formal Thinking in Mathematics

David Tall
University of Warwick, UK

This paper focuses on the changes in thinking involved in the transition from school mathematics to formal proof in pure mathematics at university. School mathematics is seen as a combination of visual representations, including geometry and graphs, together with symbolic calculations and manipulations. Pure mathematics in university shifts towards a formal framework of axiomatic systems and mathematical proof. In this paper, the transition in thinking is formulated within a framework of 'three worlds of mathematics'– the 'conceptual-embodied' world based on perception, action and thought experiment, the 'proceptual-symbolic' world of calculation and algebraic manipulation compressing processes such as counting into concepts such as number, and the 'axiomatic-formal' world of set-theoretic concept definitions and mathematical proof. Each 'world' has its own sequence of development and its own forms of proof that may be blended together to give a rich variety of ways of thinking mathematically. This reveals mathematical thinking as a blend of differing knowledge structures; for instance, the real numbers blend together the embodied number line, symbolic decimal arithmetic and the formal theory of a complete ordered field. Theoretical constructs are introduced to describe how genetic structures set before birth enable the development of mathematical thinking, and how experiences that the individual has met before affect their personal growth. These constructs are used to consider how students negotiate the transition from school to university mathematics as embodiment and symbolism are blended with formalism. At a higher level, structure theorems proved in axiomatic theories link back to more sophisticated forms of embodiment and symbolism, revealing the intimate relationship between the three worlds.

Introduction

The ideas in this paper are situated in an overall view of long-term human learning, building from genetic structures that we all share and developing more sophisticated individual knowledge based on personal experiences. In particular I propose that there are three fundamental human attributes set before our birth in our genes that are essential to mathematical thinking and that personal growth depends on the individual's interpretations of new situations based on experiences they have met before.

Set-befores

I use the term 'set-before' to refer to a mental structure that we are born with, which may take a little time to mature as our brains make connections in early life. For instance, the visual structure of the brain has built-in systems to identify colours and shades, to see changes in shade, identify edges, coordinate the edges to see objects and track their movement. Thus the child is born with a biological system to recognise small numbers of objects (one, two, or perhaps three) that gives a 'set-before' for the concept of 'twoness' before the child learns to count. Other set-befores include conceptions such as 'up' and 'down' related to the pull of gravity and our upright posture, and the related concept of the horizontal. Another is the sense of weight that we encounter through the pull on our muscles as we lift objects. Other set-befores include the social ability to interact with others using gestures such as pointing to draw attention to things.

However, there are three fundamental set-befores that shape our long-term learning and cause us to think mathematically in specific ways. They are:

- *recognition* of patterns, similarities and differences;
- *repetition* of sequences of actions until they become automatic.
- *language* to describe and refine the way we think about things;

While recognition and repetition to practice routines are found in other species, it is the power of language, and the related use of symbols, that enables us to focus on important ideas, to name them and talk about them to refine their meaning. Recognition of patterns is an essential facility for mathematics, including patterns in shape and number.

Repetition that becomes automatic is essential for learning procedures. However, there is a more sophisticated level that involves not only the ability to *perform* the procedure, but also to *think about* it as sophisticated entities in their own right, where symbols operate dually as process and concept (procept) to allow us to think flexibly (Gray & Tall, 1994).

Mathematical development depends profoundly on these three set-befores. By being able to routinise a sequence of actions so that we can do it without effort, we can think about it and do it again, and again. Each counting number is followed by another, and another, leading to potential infinity. By categorising the collection of numbers and giving it a name, or the symbol $\mathbb{N}$, we can conceive of an actual infinity of numbers as a single entity. Thus repetition and categorisation can together lead to the notion of actual infinity.

Met-befores

Personal development builds on experiences that the individual has met before. Previous experiences form connections in the brain that affect how we make sense of new situations. I define a *met-before* to be 'a current mental facility based on specific prior experiences of the individual.'

A met-before is sometimes consistent with the new situation and sometimes inconsistent. For instance, the met-before '2+2 makes 4' is experienced first in whole number arithmetic and continues to be consistent with the arithmetic of fractions, positive and negative integers, rational, real and complex numbers. But the met-before 'taking away gives less' remains consistent with (positive) fractions, but is inconsistent with negatives where taking away –2 gives more. The same met-before works consistently with finite sets, where taking away a subset leaves fewer elements, but is inconsistent in the context of infinite sets, where removing the even numbers from the counting numbers still leaves the odd numbers with the same cardinality. In this way, met-befores can operate covertly, affecting the way that individuals interpret new mathematics, sometimes to advantage, but sometimes causing internal confusion that impedes learning.

Most long-term curricula focus only on broadening experiences based on positive met-befores, failing to address met-befores that cause many learners profound difficulties. For example, mathematicians will have the limit concept as a met-before in their own minds, which, for them, forms the logical basis of calculus and analysis; but it is not a met-before for students beginning calculus and causes profound difficulties. The brain changes in its ability to think over time, reorganising information to create new structures that are often more sophisticated and better at coping with new situations. It is not simply a repository of earlier experiences adding new information to old; it re-formulates old information in new ways, changing how we think as we grow more mature. Experts may have

forgotten how they thought when they were young and are likely to need to reflect on how different students' met-befores affect their ways of learning.

Three Worlds of Mathematics

The development of the individual from a young child to a sophisticated adult builds on the three fundamental set-befores of recognition, repetition and language to construct three interrelated sequences of development that blend together to build a full range of mathematical thinking (Tall, 2004, 2006). This is not to say that there is a one-to-one correspondence between set-befores and sequences of development. However, recognition and categorisation of figures and shapes underpins thought experiments with geometry and graphs, while the repetition of sequences of actions symbolised as thinkable concepts leads to arithmetic and algebra. Each of these constructional processes develop further through the use of language to describe, define and deduce relationships, until, at the highest level, set-theoretic language is used as a basis for formal mathematical theory.

While it may be argued that these developments are simply different modes of thinking that grow in sophistication, I have come to describe them as 'three worlds of mathematics' that develop in sophistication in quite different ways.

- the *conceptual-embodied world*, based on perception of and reflection on properties of objects, initially seen and sensed in the real world but then imagined in the mind;
- the *proceptual-symbolic world* that grows out of the embodied world through action (such as counting) and is symbolised as thinkable concepts (such as number) that function both as processes to do and concepts to think about (procepts);
- the *axiomatic-formal world* (based on formal definitions and proof), which reverses the sequence of construction of meaning from definitions based on known objects to formal concepts based on set-theoretic definitions.

Terms such as 'embodied', 'symbolic', 'formal' have all been used in a range of different ways. Here I use a technique that arose from my friend and supervisor, the late Richard Skemp, in putting two familiar words together in a new way to signal the need to establish a new meaning (such as 'instrumental understanding' and 'relational understanding' or 'concept image' and 'concept definition').

'Conceptual embodiment' refers not only to the broader claims of Lakoff (1987) that all thinking is embodied, but more specifically to perceptual representations of concepts. We conceptually embody a geometric figure, such as a triangle consisting of three straight line-segments; we *imagine* a triangle as such a figure and allow a specific triangle to act as a *prototype* to represent the whole class of triangles. We 'see' an image of a specific graph as representing a specific or generic function. Conceptual embodiment grows steadily more sophisticated as the individual matures in a manner described by Van Hiele (1986), building from perception of objects, through description, construction and definition, leading to deduction and Euclidean geometry. Other embodied geometries follow, such as projective geometry, spherical geometry, and various non-euclidean geometries, all of which may be given a physical embodiment. It is only when the systems are axiomatised and the properties deduced solely from the axioms using set-theoretic formal proof that the cognitive development of geometry shifts fully to a formal-axiomatic approach (See Figure 1).

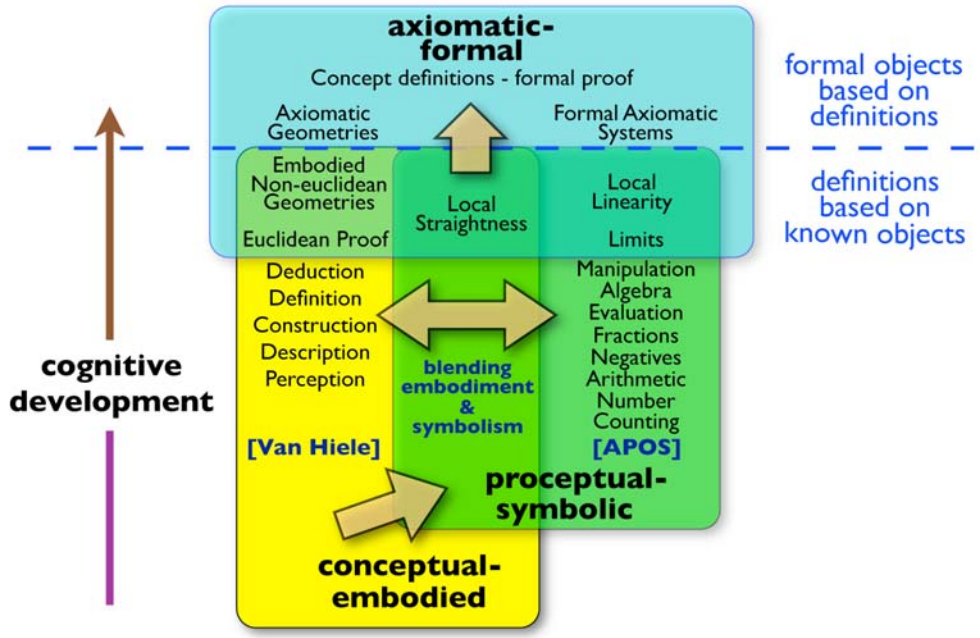

Figure 1. The Three Worlds of Mathematics illustrated by selected aspects.

Proceptual symbolism' refers to the use of symbols that arise from performing an action schema, such as counting, that become thinkable concepts, such as number (Gray & Tall, 1994). A symbol such as 3+2 or $\sqrt{b^2-4ac}$ represents both a process to be carried out or the thinkable concept produced by that process. Such a combination of symbol, process, and concept constructed from the process is called an *elementary procept*; a collection of elementary procepts with the same output concept is called a *procept*.

Process-object encapsulation was first described succinctly by Dubinsky in his APOS theory (e.g. Cottrill et al., 1996) based on the theories of Piaget and was used mainly in programming mathematical constructs in a symbolic development. Later in this paper we will return to APOS theory to show how a blending of embodiment and symbolism gives a more complete way of developing sophistication in mathematical thinking.

'Axiomatic formalism' refers to the formalism of Hilbert that takes us beyond the formal operations of Piaget. Its major distinction from the elementary mathematics of embodiment and symbolism is that in elementary mathematics, the definitions arise from experience with objects whose properties are described and used as definitions; in formal mathematics, as written in mathematical publications, formal presentations *start* with set-theoretic definitions and deduce other properties using formal proof.

Formal mathematics does not arise in isolation. In his famous lecture announcing the twenty-three problems that dominated the twentieth century, Hilbert remarked:

> To new concepts correspond, necessarily, new signs. These we choose in such a way that they remind us of the phenomena which were the occasion for the formation of the new concepts. So the geometrical figures are signs or mnemonic symbols of space intuition and are used as such by all mathematicians. 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, 1900 ICME lecture

It is important to discuss the interrelationship of worlds working together. Putting together two names, such as 'conceptually embodied axiomatic formalism' is clearly inappropriate and compression is required. For this purpose, we now refer to the three worlds simply as 'embodied', 'symbolic' and 'formal', using the meanings for the terms established above, which enables us to combine them to give names such as 'embodied formalism' when formal thinking is underpinned by embodiment.

The overall structure of Figure 1 can now be seen in outline as a combination

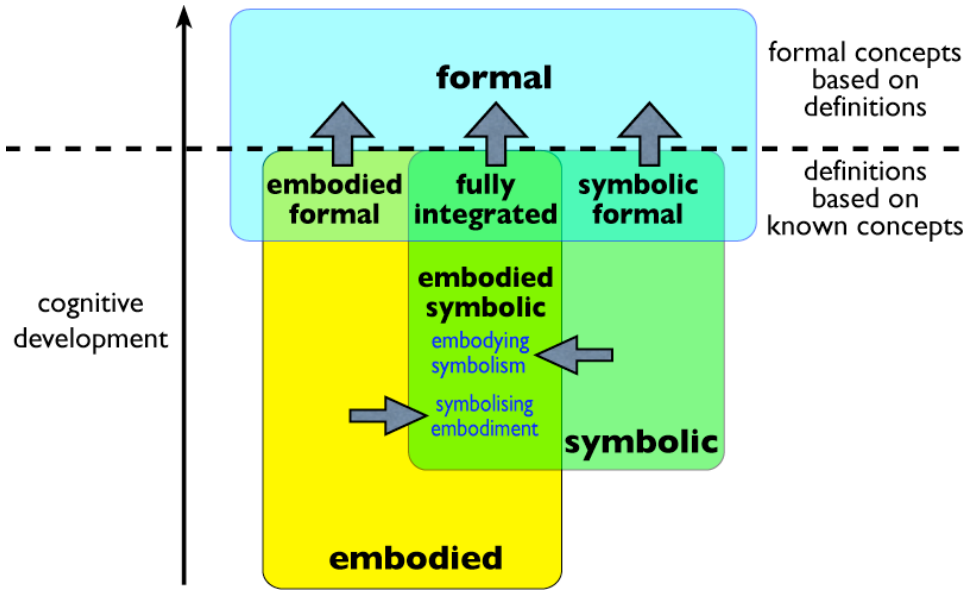

Figure 2. Cognitive development through three worlds of mathematics.

of interacting worlds of mathematics in Figure 2.

School mathematics builds from embodiment of physical conceptions and actions: playing with shapes; putting them in collections; pointing and counting; sharing; measuring. Once these operations are practiced and become routine, they can be symbolised as numbers and used dually as operations or as mental entities on which the operations can be performed. As the focus of attention switched from embodiment to the manipulation of symbols, mathematical thinking switches from the embodied to the (proceptual) symbolic world. Throughout school mathematics, embodiment gives specific meanings in varied contexts while symbolism in

arithmetic and algebra offers a mental world of computational power.

The later transition to the formal axiomatic world builds on these experiences of embodiment and symbolism to formulate formal definitions and to prove theorems using mathematical proof. The written formal proof is the final stage of mathematical thinking; it builds on experiences of what theorems might be worth proving and how the proof might be carried out, often building implicitly on embodied and symbolic experience.

Formal theories based on axioms often lead to *structure theorems*, which reveal that an axiomatic system (such as a vector space) has a more sophisticated embodiment and related symbolism—for instance a finite dimensional vector space is an n-dimensional coordinate system. In this way the theoretical framework turns full circle, building from embodiment and symbolism to formalism, returning once more to a more sophisticated form of embodiment and symbolism that, in turn, gives new ways of conceiving even more sophisticated mathematics.

This gives a natural parsimony to the framework of three worlds: as human embodiment leads to the mathematical operations of symbolism and on to the formalism of pure mathematics and back again at higher levels to more embodiment and symbolism. Meanwhile those who *use* mathematics in physics, applied mathematics, economics and so on, formulate mathematical models and symbolism to process the mathematics in the models—an approach justified by the accompanying formal framework that interlinks embodiment, symbolism and formalism.

Compression, Connection and Thinkable Concepts

The study of the development of mathematical thinking is aided by several theoretical concepts to support our analysis. The human brain is highly sophisticated, but it is also surprisingly limited, being able to deal with only a small number of pieces of information at a time. In his famous paper, Miller (1956) suggested the number is around 7 ± 2, based on a review of many articles published at the time. Personally I feel that it is much smaller than this; perhaps I could cope with more when I was younger – but I can't remember.

The human brain copes with this by connecting ideas together into 'thinkable concepts'. (Although all concepts are clearly thinkable, I use the two words together to focus on how the concept is held in the mind as a single entity at a single time.)

Compression into thinkable concepts occurs in several different ways. One, discussed by Lakoff (1987) in his book *Women Fire and Dangerous Things*, is categorisation, where concepts are connected in various ways in a category that itself becomes a thinkable concept. Sometimes the category may be represented by a specific case operating in a generic capacity such the equality $3+4=4+3$ representing commutativity of addition.

Another mode of compression, described by Dubinsky and his colleagues (Cottrill et al., 1996), occurs in APOS theory where an ACTION is internalised as a PROCESS and is encapsulated into an OBJECT, connected to other knowledge within a SCHEMA; they also note that a SCHEMA may also be encapsulated as an OBJECT.

Following Davis (1983), who used the term 'procedure' to mean a specific sequence of steps and a process as the overall input-output relationship that may be implemented by different procedures, Gray, Pitta, Pinto and Tall (1999) represented the successive compression from procedure through multi-procedure, process and procept, expanded in Figure 3 to correspond to the SOLO taxonomy

sequence: unistructural, multi-structural, relational, extended abstract (Pegg & Tall, 2005).

This models the way in which a procedure—as a sequence of steps performed in time—is steadily enriched by developing alternative procedures to allow an efficient choice. The focus switches from the individual steps to the overall process, and may then be compressed as a procept to think about and to manipulate mentally in a flexible way.

Some students who have difficulty may become entrenched in a procedural approach, perhaps reaching a multi-procedural stage that can lead to procedural efficiency. Other students develop greater flexibility by seeing processes as a whole and compressing operations into thinkable concepts. This can lead to a spectrum of outcomes within a single group of learners between those who perform procedurally and those who develop greater flexibility. In arithmetic, Gray and Tall (1994) called this the *proceptual divide*.

The earlier work of Dubinsky and his colleagues (e.g. Cottrill et al., 1996) focused initially on a symbolic approach by programming a procedure as a function and then using the function as the input to another function. The data shows that, while the process level was often attained, encapsulation from process to object was more problematic.

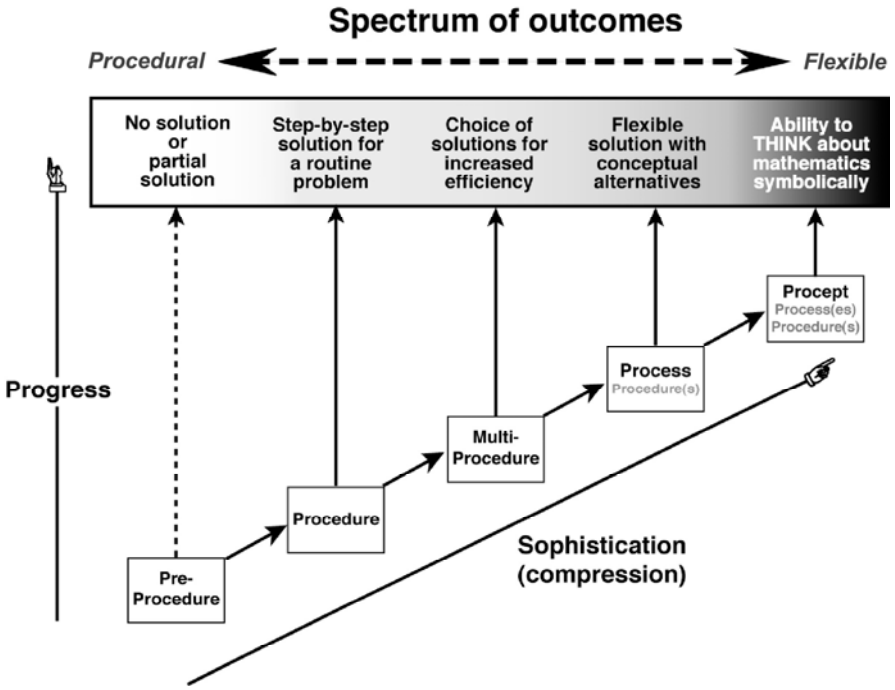

Figure 3. Spectrum of outcomes from increasing compression of symbolism.

A curriculum that focuses on symbolism and not on related embodiments may limit the vision of the learner who may learn to perform a procedure, even conceive of it as an overall process, but fail to be able to imagine or 'encapsulate' the process as an 'object'.

Widening the perspective to link symbolism to embodiment reveals that symbolic compression from procedure to process to object has an embodied counterpart. This happens when the actions involved operate on visible objects. The actions have an *effect* on the objects, for instance, when sharing them into equal shares, permuting them into a new arrangement, or translating an object on a plane. The 'effect' is the change from the initial state to the final state. The compression from procedure to process can be seen by shifting the focus of attention from the *steps* of a procedure to the *effect* of the procedure.

For example, a translation of an object on a plane is an action in which each point of the object is moved in the same direction by the same magnitude. At the multi-structural level all the arrows from a start point to finish point can be seen to be *equivalent*, providing a *set* of equivalent translations. However, any *one* of these arrows can be used as a representative of *all* the equivalent arrows. A more subtle interpretation shifts us from the process level (equivalent arrows) to an object level by representing the *effect* of the action as a *single* free vector, as an arrow of given magnitude and direction that may be moved to any point to show how that point moves. This free vector is a conceptual embodiment of the vector translation as a mental embodied object. Adding free vectors is performed by placing them nose to

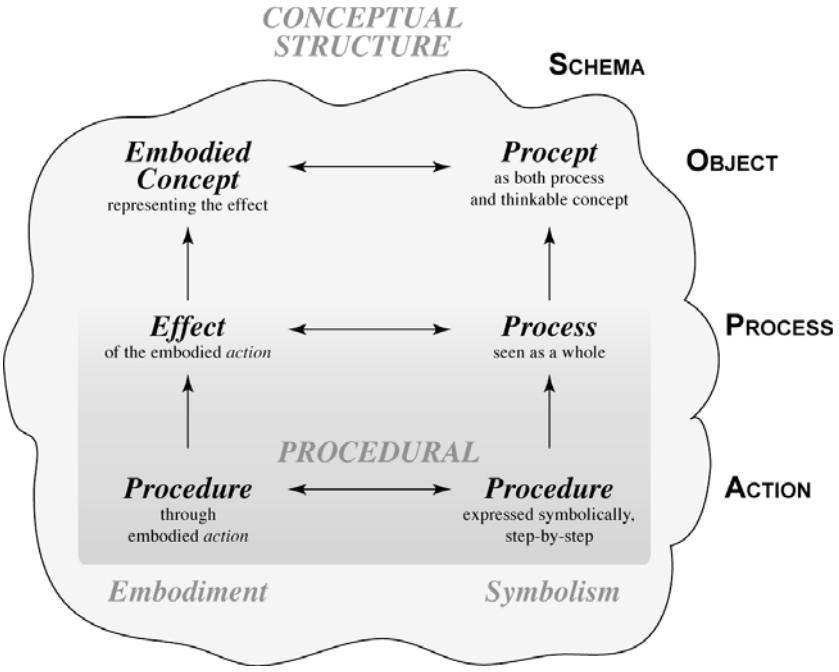

Figure 4. Procedural knowledge as part of conceptual knowledge (from Tall, 2006).

tail to give the unique free vector that has the same effect as the two in succession. In the embodied world, there is therefore a meaningful parallel to symbolic compression in APOS theory by shifting one's attention from the *steps* of an action to the *effect* of that action and imagining the *effect* as an embodied thinkable concept (See Figure 4).

This combination of embodiment and symbolism can give an *embodied meaning* to the desired encapsulated object, changing the learning required from a search for an as yet un-encapsulated symbolic object in APOS theory to the state of having an embodiment of the required object and searching for a numeric or symbolic way to compute it.

As different individuals follow through a mathematics curriculum that introduces ideas in increasing levels of sophistication, they cope with it in different ways. Piaget hypothesised that all individuals pass through the same sequence of stages at different rates, but Gray and Tall (1994) observed the proceptual divide in which children develop in different ways, some clinging to the security of known step-by-step procedures, while others compress their knowledge into the flexible use of symbols as process and concept (procepts). Procedures occur in time and work in limited cases but may not be sufficiently compressed into thinkable concepts to be used flexibly for more sophisticated thinking. Procedural learning may have a short-term advantage to pass an imminent test, but it needs the additional compression into thinkable concepts to enable the long-term development of increasingly sophisticated mathematical thinking.

Knowledge Frameworks and Conceptual Blending

Recent developments in cognitive science suggest an overall picture of long-term growth that is of great value in mathematical thinking. Fauconnier and Turner (2001) present a view of the development of human thinking that focuses on *compression* and *conceptual blending*. Compression is seen as a general cognitive process that compresses situations in time and space into events that can be comprehended in a single structure by the human brain. For instance, the statement 'If Mrs Thatcher stood for President, then she would not get elected because the unions would oppose her' is a compression blending together the American and British democratic systems. The blend links similar ideas, such as the election of a leader in a democratic system subject to the support or opposition of pressure groups and ignores differences such as the fact that the American President is elected by all the people while the British Prime Minister is the elected leader of the party that wins the election. Blending also encourages new creative thinking, such as a higher-level analysis of the ways in which different democracies work.

In general, when we encounter a new situation we interpret it by blending together our met-befores, which may come from different experiences having some aspects in common and others in conflict. Those in common may give pleasurable insight; those in conflict may cause confusion that can act as a challenge for those who feel confident but lead to anxiety for those who do not.

The development of the number concept is a typical case of successive blends. While the number systems $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$ may be seen by a mathematician as successive number systems represented on the number line which lies in the complex plane $\mathbb{C}$, each extension involves a sophisticated blending process for the learner. The number line itself is a blend of counting and measuring where each

whole number has a 'next' in the counting operation, but in measurement there is no 'next' fraction. Operating with whole numbers gives the sense that 'addition and multiplication give a bigger result' and 'take-away gives less' which conflicts with the behaviour of integers, where taking away a negative gives more, and with fractions where multiplication can produce a smaller result.

Later expansions of the number system blend an original knowledge structure within a wider structure with properties that conflict with previous experience. The hypotenuse of a right-angled triangle with rational sides may not itself be rational, the shift from fractions to decimals introduce infinite decimals that never end. The embodied number line includes numbers such as π, e and $\sqrt{2}$ that cannot be expressed as fractions or recurring decimals. Every non-zero number on the number-line has a square which is positive but the complex numbers have a 'number' $\mathbf{i}$ whose square is negative.

Blends can occur within one of the worlds of mathematics or between different worlds. For instance, multiplication is a blend of different embodiments such as the area produced by multiplying two lengths or the number of elements in a rectangular array of objects. On the other hand, algebraic symbolism may be blended with corresponding embodied graphs. The shift from school mathematics to the logical demands of university mathematics involves a major shift in knowledge blending.

Blending Embodiment, Symbolism and Formalism in the Concept of Real Number

The concept of real number is a blend of *embodiment* as a number line, *symbolism* as (infinite) decimals and *formalism* as a complete ordered field. Each has its own properties, some of which are in conflict. For instance, the number line develops in the embodied world from a physical line drawn with pencil and ruler to a 'perfect' platonic construction that has length but no thickness. This is a natural process of compression in which the focus of attention concentrates on the straightness of the line and the position of the lines and points. In Greek geometry, points and lines are different kinds of entity in which a point has position but no size and a point may by 'on' a line or not. The line is an entity in itself; it is not 'made up of points'.

Physically the number line can be traced with a finger and, as the finger passes from 1 to 2, it feels as if it goes through all the points in between. But when this is represented as decimals, each decimal expansion is a different point (except for the difficult case of recurring nines) and so it does not seem possible to imagine running through *all* the points between 1 and 2 in a finite time. There is also the counterfactual dilemma that, if the points have no size, how can even an infinite number of them make up the unit interval? In the embodied world we may imagine a point as a very tiny mark made with a fine pencil, so practical points have an indeterminate small size even if theoretical points do not. Furthermore, if a point had no size and a line no thickness, then we would not be able to see them. Prior to the introduction of the formal definition of real numbers, we live, perhaps somewhat uneasily, with the blend of a practical number line that we draw and imagine and a symbolic number system that can be represented by infinite decimals.

Formally, the real numbers $\mathbb{R}$ is an ordered field satisfying the completeness axiom. This involves entering a completely different world where addition is no longer defined by the algorithms of counting or decimal addition, instead it is

simply asserted that for each pair of real numbers *a*, *b*, there is a third real number call the sum of *a* and *b* and denoted by *a+b*. Formally, it is possible to prove that there is, up to isomorphism, precisely *one* complete ordered field and that this can be represented by infinite decimals which are unique (except for the case where one decimal ends in an infinite sequence of nines and the other increases the previous place by one and ends in an infinite sequence of zeros). Thus it is possible for the human brain to recycle its former experiences and use the arithmetic of experience to blend the symbolic world with the formal world.

Personally I continue to be concerned that I 'know' things symbolically that I have never proved axiomatically. In the symbolic world, I 'know' that 2^{10} is bigger than 10^3, because the first is 1024 and the second is 1000. But I have never proved this from the axioms for a complete ordered field or from the Peano postulates for the whole numbers. I am happy to *accept* that the familiar arithmetic of decimals is the unique arithmetic of the axiomatic complete ordered field because it fits together so coherently. But 'acceptance' is not mathematical proof.

In the transition from school arithmetic to formal mathematics we need to confront many issues such as this. Is it any wonder that Halmos in his book *I want to be a mathematician* remarked, 'I never understood epsilon-delta analysis, I just got used to it.' As mathematicians we begin to appreciate the purity and logic of the formal approach, but as human beings we should recognise the cognitive journey through embodiment and symbolism that enabled us to reach this viewpoint and helps us sustain it.

Blending Embodiment, Symbolism and Formalism in Calculus and Analysis

Calculus builds in three very different worlds of mathematics. Calculus in school is a blend of the world of embodiment (drawing graphs) and symbolism (manipulating formulae). The geometric notion of slope of a graph is often represented by the action of moving a secant through a point on the graph towards a tangent at the point or, more subtly, through magnifying the graph near the point to see it look like a straight line under high magnification. The latter enables the learner to 'see' the changing slope of a curve and to imagine the slope itself as a changing function. The symbolic aspect allows the slope between two distinct points to be computed numerically or symbolically and a limiting process is required to get the symbolic slope of the tangent as the symbolic derivative. The embodied version has the limit process *implicit* in the process of magnification, while the symbolic version involves computing an *explicit* symbolic representation.

It is interesting to note that the mathematical expert, who already has conceptions of derivative, integral and so on, has the limit concept as a met-before and sees the calculus as building logically from the limit concept, hence designing the curriculum to build on an 'informal' version of the limit concept. However, the novice may feel more comfortable with the embodied approach through magnification to 'see' the slope function *before* being introduced to symbolic techniques for computing it and formal language to define it.

Reform calculus in the USA was built on combining graphic, symbolic and analytic representations of functions using computer software and graphical calculators. However, those of us occupied in research in undergraduate mathematics need to look a little deeper into how the concepts of calculus are constructed. Mathematicians, who live in a world built on the met-before of the

limit concept, have a view of calculus that sees the need to introduce the limit concept explicitly at the beginning of the calculus sequence. My own view is different. For students building on the embodiment and symbolism of school mathematics, I see a more natural route into the calculus combining embodiment and symbolism in a manner that has the full potential to lead either to standard mathematical analysis, non-standard infinitesimal analysis, or practical calculus in applications.

This approach involves using the embodied notion of local straightness that is cognitively different from the symbolic notion of local linearity. Local straightness involves an embodied thought experiment looking closely at graphs to see that, as small portions of certain graphs are highly magnified, they look straight. Some mathematicians have difficulties with such an approach because it seems difficult to formalise at first encounter. But it makes sense to students as they look at a computer screen successively magnifying a graph of a familiar function composed of polynomials, trigonometric functions, exponentials or logarithms. It also makes sense that a function like $|\sin x|$ has a corner at every multiple of π so that on can begin to imagine not only local straightness, but also situations that are not locally straight. It is also relatively simple to give an embodied proof with hand gestures, that the recursively-defined blancmange function is everywhere continuous, but nowhere differentiable (Tall & Giacomo, 2000). Here magnification of the graph shows tiny blancmanges growing everywhere, so the magnification never looks straight (Figure 5).

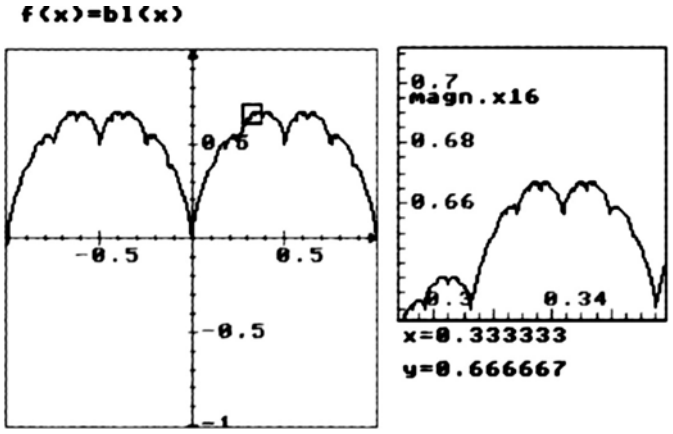

Figure 5. A graph that nowhere looks straight under magnification.

The arguments and pictures are found in several of my papers (see for example, Tall 2003). The embodied ideas can give highly insightful ideas not found in a normal symbolic approach. For example, defining the 'nasty function' $n(x) = bl(1000x)/1000$ then $\sin x$, and $\sin x + n(x)$ look the same when drawn on a computer over a range say –5 to 5, but one is differentiable everywhere and the other is differentiable nowhere! This can be seen just by magnification. It shows that just looking at a static graph is not enough. To be sure of differentiability one

needs to deal directly with the function given symbolically. Hence embodiment reveals subtle meanings that encourage the use of symbolism and formal definitions.

No regular calculus course attempts to give insight into what it means to be nowhere differentiable, yet I do it in my first lesson on calculus to show some functions are locally straight and some are not. If one can imagine, in the mind's eye, that a graph is locally straight, then as the eye follows the curve from left to right, focusing on the slope of the curve, it is possible to *see* the changing slope as a function that can be graphed in its own right. This brings us precisely to the principle enunciated earlier, that the slope can be embodied and visualised giving a slope function that can be *seen* but now needs to be calculated either numerically or symbolically. The need for a limit arises *from* the embodiment to calculate the slope function, not the other way round.

An approach using local linearity, as in College Calculus, on the other hand, involves a *symbolic* concept, seeking the best linear approximation to the curve at a single point. It involves an *explicit* limiting concept from the beginning instead of an implicit limiting concept that occurs when zooming in to see how steep the curve is over a short interval. Non-differentiability is the non-existence of a limit, which lacks the immediacy of the embodied idea of a graph that does not magnify to look locally straight.

The function $a(x) = \int_0^x bl(t)\,dt$ has $bl(x)$ as its derivative, so it is differentiable once everywhere and twice nowhere. When I showed a class of undergraduates the graph of $a(x)$ calculated numerically by a computer program, one of the students (not a mathematics major) said, 'you mean that function is differentiable once but not twice.' (Tall, 1995.) If you know of any other mathematics professor who has had a student imagine a function that is differentiable once and not twice, tell him or her to e-mail me.

Local straightness is particularly apt when dealing with differential equations. A differential equation $dy/dx = F(x,y)$ tells us the slope of a locally straight curve at a point (x,y) is $F(x,y)$, so it is easy to program software to draw a small segment of the appropriate slope when the mouse points to (x,y) and by depositing such segments end to end, the user can build an approximate solution onscreen. This was done in the *Solution Sketcher* (Tall, 1991) and has been implemented in the currently available *Graphic Calculus* software (Blokland & Giessen, 2000, Figure 6).

The Reform Calculus Movement in the USA focuses on the notion of *local linearity*, with the derivative as the best linear approximation to the curve at a single point. It seeks a symbolic representation at a point, using a limiting procedure to calculate the best linear fit perhaps even with a formal epsilon-delta construction. Then the fixed point is varied to give the global derivative function. I cannot imagine a worse approach to present to beginning calculus students.

Thurston (1994) suggested seven different ways to think of the derivative:

(1) **Infinitesimal**: the ratio of the infinitesimal change in the value of a function to the infinitesimal change in a function.

(2) **Symbolic**: the derivative of x^n is nx^{n-1}, the derivative of sin(x) is cos(x), the derivative of $f \circ g$ is $f' \circ g * g'$, etc.

(3) **Logical**: $f'(x) = d$ if and only if for every ε there is a δ such that when

$0 < |\Delta x| < \delta$, then $\left| \frac{f(x+\Delta x) - f(x)}{\Delta x} - d \right| < \delta.$

(4) **Geometric**: the derivative is the slope of a line tangent to the graph of the function, if the graph has a tangent.
(5) **Rate**: the instantaneous speed of $f(t)$, when t is time.
(6) **Approximation**: The derivative of a function is the best linear approximation to the function near a point.
(7) **Microscopic**: The derivative of a function is the limit of what you get by looking at it under a microscope of higher and higher power. (Thurston, 1994.)

These ideas show a mathematician with great insight blending together a range of possible meanings, including local straightness expressed at a point (item 7). However it omits the *global* concept of local straightness from which all others can grow:

(0) **Embodied**: the (changing) slope of the graph itself.

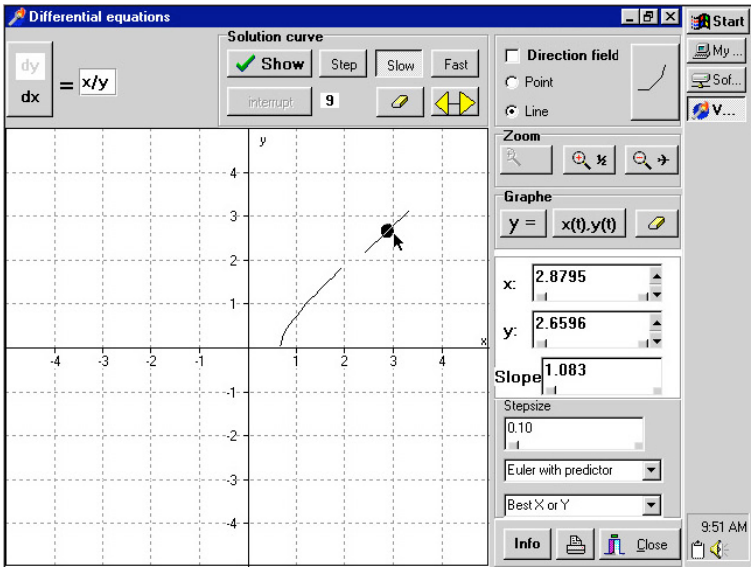

Figure 6. Building the solution of a differential equation by following its given slope (Blokland & Giessen, 2000).

Mathematicians, with their met-befores based on the limit concept have long passed beyond this missing level 0. Learners without experience of the limit concept benefit from such an embodied introduction.

It is my contention (Mejia & Tall, 2004) that the calculus belongs not to the formal world of analysis, 'looking down' on it from above: it belongs in the vision of Newton and Leibniz, looking up from met-befores in embodiment and symbolism used appropriately.

Using a framework of embodiment and symbolism, Hahkiöniëmi (2006) studied his own calculus teaching to find students following different

developments, including an embodied route, a symbolic route and various combinations of the two. He found that 'the embodied world offers powerful thinking tools for students' who 'consider the derivative as an object at an early stage.'

This simple observation is at variance with APOS theory suggesting the building up of the limit concept from (symbolic) ACTION to PROCESS and then to OBJECT. It questions Sfard's (1991) suggestion that operational thinking invariably must precede structural. In our technological age, one can *see* the structure of the derivative globally as a slope function stabilizing onscreen and seek to operationalise it by computing it numerically or symbolically. The formal limit can follow later as a natural way of completing the process already seen as an object in the mind's eye.

To cope with the complexity of the derivative, Hahkiöniëmi proposed a framework in which the teacher is responsible as a mentor for guiding the students through a variety of possible routes by which the students may blend together the various knowledge structures in a way that is personally meaningful (Figure 7).

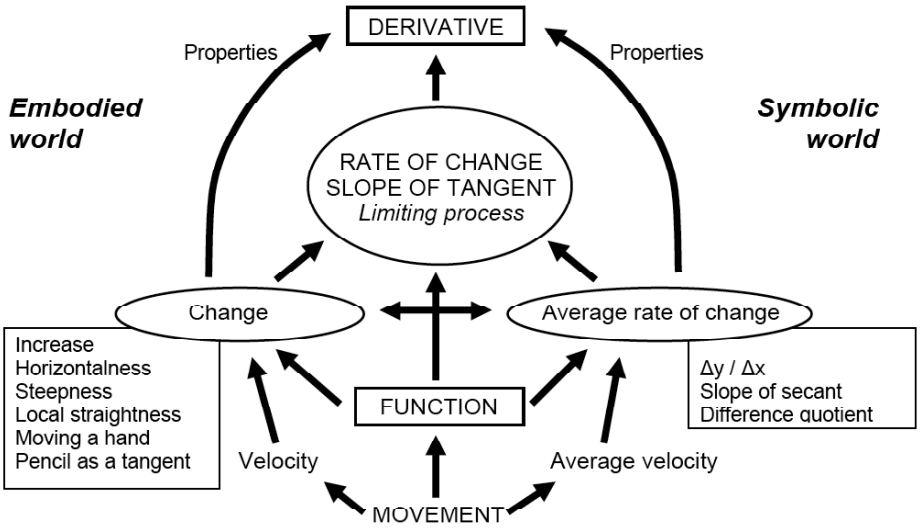

Figure 7. Hypothesised learning framework (Hahkiöniëmi, 2006).

The Cognitive Development of Proof

Proof is handled differently in each of the three worlds (Mejia-Ramos & Tall, 2006). In the embodied world the child may begin with specific experiments represented by specific pictures to confirm that something is true, for instance, a rectangle of items with 3 rows and 2 columns shows that the same array can be seen as 3 lots of 2 or 2 lots of 3, so $3\times 2 = 2\times 3$. The same picture may also be seen as a generic picture demonstrating this property for *any* two whole numbers. Later, as language is used more carefully to make definitions, geometric proofs in

Euclidean geometry become verbalised and build into an organised structure of proof from definitions. Meanwhile, in symbolic development, proof of specific properties may be performed using specific arithmetic calculations, perhaps seen as generic demonstrations, later developing into proof by algebraic manipulation.

The major shift in proof occurs from the embodiment and symbolism of school mathematics to the formalism of advanced mathematical thinking (Tall, 1991). Proof in the embodied and symbolic worlds is based on concepts that are given definitions, so the concepts underpin any sense of proof. Proof in the formal world is ostensibly based only on set-theoretic definitions and mathematical deduction. However, as students come to appreciate formal proof, they build on their previous experience, as do mathematicians who use a variety of approaches, perhaps using embodiment to suggest new hypotheses that are subsequently proved as formal theorems, or counting arguments and other calculations and manipulations that can develop into formal proofs.

My colleague and PhD student, Marcia Pinto (1998) followed students learning concepts in formal mathematical analysis and found there were two distinct routes, one a 'natural' route giving meaning to definitions from the met-befores of the individual's concept image (including both embodiment and symbolism), the other a 'formal' route extracting meaning from the concept definition (Figure 8).

For instance, Chris followed the natural route building on his imagery to give

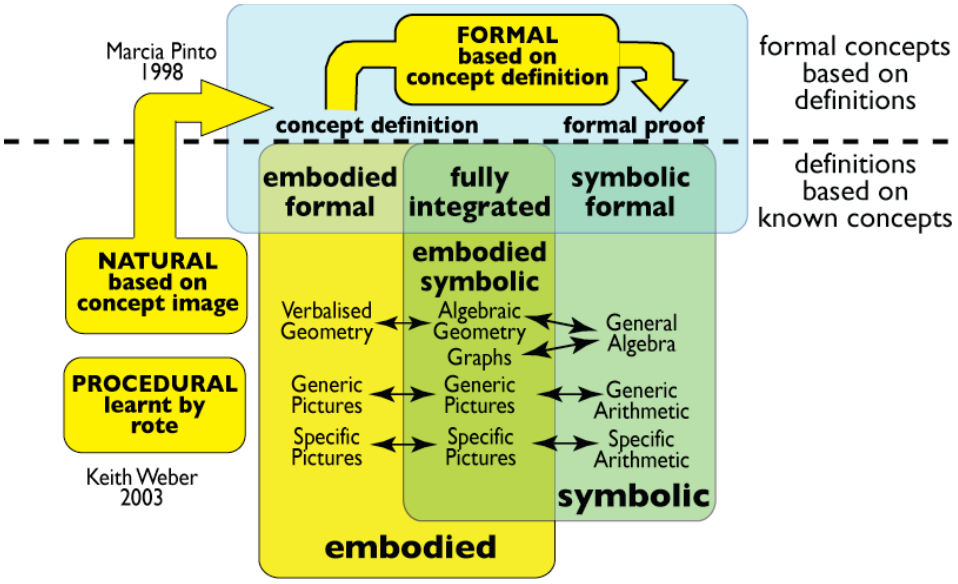

Figure 8. Natural thinking builds on embodiment and symbolism, while formal thinking builds on concept definition.

meaning to the limit concept, 'seeing' the terms (s_n) of the sequence plotted as points (n, s_n) in the plane and imagining that for any $\varepsilon>0$, he could find an N such that the points (n, s_n) for $n \geq N$ lie between horizontal lines $L \pm \varepsilon$. Ross, on the other

hand, followed the formal route by repeating the definition until he could say it in full detail and carefully studying proofs to see how they deduced a theorem from its assumptions.

Cliff also followed the natural route, but his met-befores clashed with the formal definition. He believed that a function on the integers could not be continuous as its graph consisted of disconnected dots, not a 'continuously drawn' pencil line.

Meanwhile, Rolf built on his symbolic experience and could show numerically that if $a_n = 1/n^2$ and $\varepsilon = 10^{-6}$, then he could calculate $N = 10^3$ for which $a_n < \varepsilon$ when $n > N$. However, he could not show that if $a_n \to 1$, then for some N, if $n>N$, then $a_n > \frac{3}{4}$. Not knowing the formula for a_n he could not carry out a *numerical* calculation to find N.

Weber (2004) refined this analysis by a qualitative case study on a particular analysis lecturer and his students. He found that the lecturer began with an initial *logico-structural* teaching style in which he guided the students into constructing a sequence of deductions to prove a theorem. He divided his working space on the board into two columns, with the left column to be filled in with the text of the proof and the right column as 'scratch work'. He wrote the definitions at the top of the left column and the final statement at the bottom, then he used the scratch-work area to translate information across and to think about the possible deductions to lead from the assumptions to the final result. Later, he became more streamlined, presenting proofs in a sequential *procedural* style, writing the proof down in the left column and using the right column to work out detail such as routine manipulation of symbols. Later, he taught topological ideas in what Weber termed a *semantic* style, building on visual diagrams to give meaning, then translating into formal proof.

He analysed student approaches into three types, building on the theory of Pinto:

- a *natural* approach involved giving an intuitive description and using it to lead to formal proof,
- a *formal* approach where students had little initial intuition but could logically justify their proofs,
- a *procedural* approach where students learnt the proofs given them by the professor by rote without being able to given any formal justification.

The term 'natural' corresponds to that of Pinto in giving meaning from intuitive (embodied) knowledge, 'formal' now refers to those who are *successful* in following a formal approach and 'procedural' refers to those who attempt to learn the formal proofs by rote without either embodied or logico-structural meaning. Of the students considered in Pinto's research, Chris was successful in giving embodied meaning to formal theory via a 'natural' route. Ross was successful in a 'formal' approach, extracting meaning from the definitions and the logical structure of theorems. Cliff was unable to make sense of the formal definition because it conflicted with his embodied imagery. Rolf attempted to extract meaning from the definitions based on his symbolic experience. Essentially, both Cliff and Rolf follow Weber's procedural route, but Cliff was unsettled because of a conflict with his embodied ideas, while Rolf was happy to relate the definition to his met-befores in performing calculations to find a numerical N given a numerical ε; Rolf conceived his task as learning procedures by rote to use in solving problems but this was insufficient to cope with more sophisticated ideas and he

left the course halfway through.

Weber's data also shows that students can vary in approach dependent on the context in which they work. Six students interviewed after the course all responded in a natural manner to a topological question (where topology had been taught in a semantic manner building from visual imagery). However, in two other questions about functions and limits, only *one* student responded naturally. The other responses to a question on functions were 4 formal and 1 procedural, and to a question on limits, 2 formal and 3 procedural.

Other research studies reveal how embodiment can operate in subtle ways to affect how students interpret formal definitions. For instance, in a formal lecture course that took the logical route of defining a relation as a set of ordered pairs, and then specialized the definition to specify functions, order relations, and equivalence relations, students gave a variety of meanings to the definitions that affected their interpretation of the mathematics. For example, the transitive law $a \sim b$ and $b \sim c$ implies $a \sim c$ was given subtle embodiments in which *a*, *b*, *c* were implicitly assumed to be all different, which is true for a strong order relation $a < b$, but not for an equivalence relation (Chin & Tall, 2002).

From Formal Proof Back to Embodiment and Symbolism

A major goal in building axiomatic theories is to construct a structure theorem, which essentially reveals aspects of the mathematical structure in embodied and symbolic ways. Typical examples of such structure theorems are:

- An equivalence relation on a set *A* corresponds to a partition of *A*;
- A finite dimensional vector space over a field *F* is isomorphic to F^n;
- Every finite group is isomorphic to a subgroup of a group of permutations;
- Any complete ordered field is isomorphic to the real numbers.

In every case, the structure theorem tells us that the formally defined axiomatic structure can be conceived in an embodied way and in the last three cases there is a corresponding manipulable symbolism.

Thus, not only do embodiment and symbolism act as a foundation for ideas that are formalized in the formal-axiomatic world, structure theorems can also lead back from the formal world to the worlds of embodiment and symbolism. This means that those who use mathematics as a tool can use the embodiment and symbolism to imagine problem situations and model them symbolically. In this way, engineers, economists, physicists, biologists and others often use embodiment and symbolism as a foundation for their work.

The new embodiments depend not just on experience in the world, but on concept definition and formal deduction, leading to new formal insights.

As an example, the completion of the rationals to give the reals using Dedekind cuts was seen by many as 'filling in' the gaps between rational numbers with real numbers so that the line is 'complete', with 'no room' for other numbers such as infinitesimals.

This interpretation is false. Once the formal definition of ordered field has been formulated and its properties determined by mathematical proof, then we can conceive of an ordered field *K* that is a proper ordered extension of the field $\mathbb{R}$. It is then easy to prove that any element in *K* is either greater than, or less than all elements in $\mathbb{R}$, or is of the form $a+\varepsilon$ where $a \in \mathbb{R}$ and ε is an infinitesimal (meaning that $-k < \varepsilon < k$ for all positive real numbers *k*). In a regular picture of the

line, it will be impossible to distinguish between a and $a+\varepsilon$ because they differ by something too small to see. However, the map $\mu : K \to K$ given by $\mu(x) = (x-a)/\varepsilon$ maps a to 0 and $a+\varepsilon$ to 1, which allows them to be 'seen' separately under the magnification μ. Now we can imagine the number line to have not only real numbers, but infinitesimals that we can 'see' under high magnification.

Reflections

The final return of formalism to a more sophisticated form of embodiment and symbolism through structure theorems leads me to see the three worlds of mathematics as a natural structure through which the biological brain builds a mathematical mind. The child builds from the three major set-befores of recognition, repetition and language to recognise and categorise geometric objects, to repeat procedures until they become automatic and perhaps compressed into thinkable procepts, and later to use the more technical language of set theory and logic to construct formal mathematical structures at the highest level.

A wider awareness of the met-befores of embodiment and symbolism and their subtle effects on the students transition to formal mathematical thinking now offers the possibility of explicit discussion between mathematicians and students of the nature of the transition that is occurring in learning formal mathematics.

While university mathematicians differ in their perception of the relevance of embodiment to formal proof—and some may insist that their research is purely formal—all human beings enter this world as children who cannot speak and thus go through a long-term development that builds through embodiment and symbolism to formalism. Axiomatic systems are not designed arbitrarily; they need some form of insight as to what axioms are appropriate, and here met-befores in embodiment and symbolism play subtle roles. Furthermore, formalism itself leads back to structure theorems that have embodied and symbolic meanings, giving a parsimonious framework that returns to its origins.

The proposed theory of conceptual embodiment, proceptual symbolism and axiomatic formalism offers a rich framework in which to interpret mathematical learning and thinking at all levels from the earliest pre-school mathematics through to mathematical research, and, in particular, in the transition from school to undergraduate mathematics.

References

Blokland, P., & Giessen, C. (2000). *Graphic calculus for windows*. http://www.vusoft2.nl.

Chin, E. T., & Tall D. O. (2002). University students' embodiment of quantifier. In A. D. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, 273–280). Norwich, UK: IGPME.

Cottrill, J., Dubinsky, E., Nichols, D., Schwingendorf, K., Thomas, K., & Vidakovic, D. (1996). Understanding the limit concept: Beginning with a coordinated process scheme. *Journal of Mathematical Behavior*, *15*(2), 167-192.

Davis, R. B. (1983). Complex mathematical cognition. In H. P. Ginsburg (Ed.) *The Development of Mathematical Thinking* (pp. 254–290). New York: Academic Press.

Gray, E. M., Pitta, D., Pinto M. M. F., & Tall, D. O. (1999). Knowledge construction and diverging thinking in elementary and advanced mathematics. *Educational Studies in Mathematics*, *38*(1–3), 111–133.

Gray, E., & Tall, D. O. (1994). Duality, ambiguity and flexibility: A proceptual view of simple arithmetic. *Journal for Research in Mathematics Education*, *26*(2), 115–141.

Hahkiöniemi, M. (2006). *Tools for studying the derivative*. Unpublished PhD, Jyväskylä,

Finland.
Hilbert, D. (1900). *Mathematische probleme, göttinger nachrichten*, 253-297, translated into English by Mary Winton Newson, retrieved from http://aleph0.clarku.edu /~djoyce/hilbert/problems.html.
Lakoff, G. (1987). *Women, fire and dangerous things*. Chicago: Chicago University Press.
Mejia-Ramos, J. P., & Tall, D. O. (2004). Reflecting on post-calculus-reform. *Opening plenary Topic Group 12: Calculus, International Congress of Mathematics Education*, Copenhagen, Denmark. http://www.icme-organisers.dk/tsg12/papers/tall-mejia-tsg12.pdf.
Mejia-Ramos, J. P., & Tall, D. O. (2006). The long-term cognitive development of different types of reasoning and proof. Paper presented at the *Conference on explanation and proof in mathematics: Philosophical and educational perspectives*, Essen, Germany. Available on the web at: http://www.warwick.ac.uk/staff/David.Tall/pdfs/dot2006g-mejia-tall.pdf.
Miller, G. A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.
Pinto, M. M. F. (1998). *Students' understanding of real analysis*. Unpublished PhD, Warwick University.
Pegg, J., & Tall, D. O. (2005). The fundamental cycle of concept construction underlying various theoretical frameworks. *International Reviews on Mathematical Education (ZDM)*, *37*(6), 468-475.
Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, *22*, 1–36.
Tall, D. O. (1985). Understanding the calculus. *Mathematics Teaching*, *110*, 49-53.
Tall, D. O. (Editor), (1991). *Advanced mathematical thinking*. Kluwer: Dordrecht (now Springer-Verlag).
Tall D. O. (1991). *Real functions and graphs* (*for the BBC computer, Master, Compact, Nimbus PC & Archimedes computer*). Cambridge: Cambridge University Press.
Tall, D. O. (1995). Visual organizers for formal mathematics. In R. Sutherland & J. Mason (Eds.), *Exploiting mental imagery with computers in mathematics education*, (pp. 52–70). Berlin: Springer-Verlag.
Tall, D. O. (2003). Using technology to support an embodied approach to learning concepts in mathematics. In L. M. Carvalho and L. C. Guimarães (Eds.), *História e tecnologia no ensino da matemática* (Vol. 1, pp. 1–28), Rio de Janeiro, Brasil.
Tall, D. O. (2004), The three worlds of mathematics. *For the Learning of Mathematics*, *23*(3), 29–33.
Tall, D. O. (2006). A theory of mathematical growth through embodiment, symbolism and proof. *Annales de Didactique et de Sciences Cognitives, Irem de Strasbourg*, *11*, 195–215.
Tall, D. O., & Giacomo, S. (2000). Cosa vediamo nei disegni geometrici? (il caso della funzione blancmange), *Progetto alice 1,* (2), 321–336. English version: What do we "see" in geometric pictures?, available at: http://www.warwick.ac.uk/staff/David.Tall/pdfs/dot2000f-blancmange-english.pdf
Thurston, W. P. (1994). On proof and progress in mathematics, *Bulletin of the American Mathematical Society*, *30*(3), 161–177.
Weber, K. (2004). Traditional instruction in advanced mathematics courses: A case study of one professor's lectures and proofs in an introductory real analysis course. *Journal of Mathematical Behavior*, *23*, 115–133.
van Hiele, P. M. (1986). *Structure and insight*. New York: Academic Press.

Author

David O. Tall, Institute of Education, The University of Warwick, Coventry CV4 7AL, UK. Email: <david.tall@warwick.ac.uk>.