Building Long-term Meaning in Mathematical Thinking: Aha! and Uh-Huh!

David Tall
University of Warwick, UK
david.tall@warwick.ac.uk

The sudden flashes of inspiration that arise when two or more previously unconnected planes of reference are brought together in the mind of an individual is termed an 'Aha!' Moment. These planes may take on various forms, including visual, verbal, symbolic, aesthetic truth, beauty. Sometimes they lead to major advances in insight, but they may also be faulty, leading to the need to review the situation and revise ideas to produce a more reliable solution. On other occasions, differing individuals or differing communities of practice, unfamiliar with one or more planes of reference, may reject the insight, producing a negative 'Uh-Huh!' response. The purpose of this chapter is to offer a framework for meaningful long-term development of mathematical theories that balances these aspects, providing insights that can be widely observed by teachers, learners and experts at all levels.

1. Introduction

This chapter focuses on how we humans make sense of mathematics in the long-term, which involves not only grasping the positive side that gives us insight into understanding and creating more sophisticated mathematical ideas, but also the negative side that impedes our progress.

In *The Act of Creation*, Koestler speaks of the creativity of the Aha! Moment as:

> "the spontaneous leap of insight … which connects previously unconnected matrices of experience and makes us experience reality on several planes at once."
>
> (Koestler, 1964, p.45.)

The connection of previously unlinked frameworks in mathematics operates not only in the creation of original ideas but also in teaching and learning mathematics, where learners are faced with the need to make new connections in their own minds. This involves not only the linking of different mathematical ideas, but also a range of attitudes and emotions personal to each individual. Positive attitudes include developing personal confidence in addressing new problems, building on previous success. Positive emotions include the pleasure of making connections and a sense of aesthetic beauty in the ways in which the mathematical ideas fit together.

The chapter seeks to promote the positive development of mathematical thinking, but this cannot be done while ignoring negative aspects that arise from personal difficulties with mathematics and emotions such as fear and anxiety. For an individual who has a personal sense of confidence based on previous success, it may be possible in the absence of an Aha! insight to write down the word 'Stuck!' and think positively about alternative possibilities (Mason et al., 1981). However, if this does not lead to further progress, the individual may sense a negative emotion that I term an 'Uh-Huh!' experience. This may simply be a barrier that impedes progress at the time. It may also have longer-term consequences not only in the individual, but also shared by others, impeding the progress of the whole community. This is a phenomenon that is particularly significant today as new technological tools offer new insights at a pace that moves too fast to implement in the wider society.

This chapter begins with a study of the structure and operation of the human brain sufficient to understand the relationship between mathematical thinking and personal aspects of emotional reactions and attitudes towards mathematics. Then we consider the cultural development of mathematical thinking to gain insight into how communities can hold very

Prepared for *The Creativity of Aha! Moments and Mathematics Education*, Ed. Bronislaw Czarnocha
This pre-publication version links to files on the internet, including items on my personal website.

different views of the nature of mathematics. This is related to a long-term framework of development of mathematical thinking in the individual to address how the learner faces significant changes in meaning as new mathematical contexts are encountered.

Problem-solving will be seen in conjunction with Skemp's (1979) affective theory of goals (that the individual wishes to achieve) and anti-goals (to be avoided), to integrate cognitive and affective aspects. This will be related to the way in which Japanese Lesson Study seeks to provoke the Aha! integration of different planes of thought, how this may be integrated in the meaningful mathematics framework, and how the desire to transfer different practices to other communities is affected by cultural differences.

I will then consider several of my own personal 'Aha!s' that have offered me insight into the development of mathematical thinking – such as 'concept image' (Tall & Vinner, 1981), the notion of 'local straightness' to give embodied meaning to the derivative (Tall, 1985), mathematical symbols interpreted flexibly as process and concept (procept) (Gray & Tall, 1994), three worlds of mathematics (Tall, 2004), structure theorems to give embodied and symbolic meaning to formal structures (Tall, 2013), and making sense of mathematical expressions through spoken articulation (Tall, Tall & Tall, 2017). I will consider how in certain curricula, some have been partially adopted in modified forms while others have not, and relate these to Aha! or Uh-Huh! experiences. I conclude on a positive note by considering broader ideas of 'Aha!s' in dreams and in the aesthetic qualities of mathematical beauty.

2. Mathematical thinking in the biological brain

Mathematical thinking occurs in the human brain, and is supported by communication with others, building on the accumulated knowledge of succeeding generations. To relate mathematics to the structure and operation of the brain requires a link between mathematics and a completely different plane of reference. This may result in an Aha! experience for the reader, one that gives insight, or an Uh-Huh! experience that fails to make the link. To illustrate the nature of the difficulty, I will offer a brief outline of some aspects of brain structure and operation and then seek to translate the essential features related to mathematics and emotion into more familiar language.

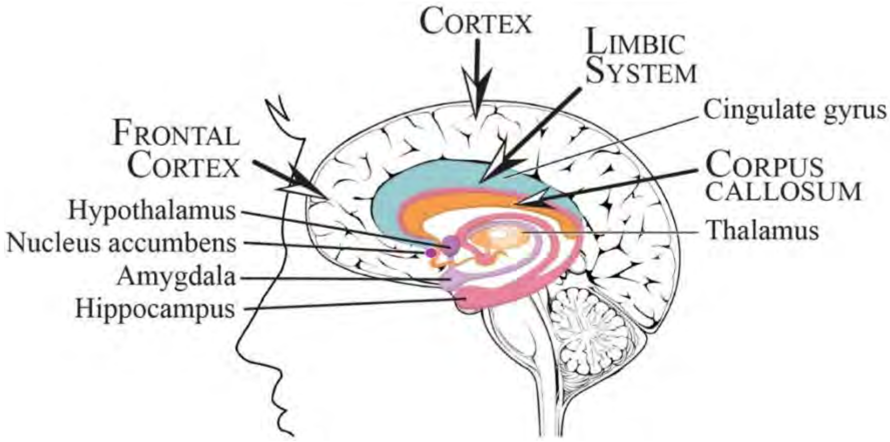

Figure 1: A cross-sectional view of the brain from the left side

Figure 1 shows an outline representation of the left-side of the brain with some details of its structure. The outer layer of the brain surface – the *cortex* – deals with cognitive issues of receiving sensory information and processing it to take actions, with the *frontal cortex* making conscious decisions. In the middle of the brain is a collection of structures – the *limbic system* – that operates subconsciously, fulfilling a diverse range of functions, from moderating the automatic controls of the body, sensing pleasure and pain, reacting immediately in terms of 'fight or flight', and other major activities such as storing and retrieving long-term memory.

The brain has two essentially symmetric parts on the left and the right which serve different functions while cooperating through a structure consisting of around two hundred million nerve fibres called the *corpus callosum*.

Seen from above, the brain's left and right parts have areas in the cortex that receive sensory input and output action across the middle top surface of the brain. Paradoxically, in humans, but not in many other species, the left side of the brain deals with the right side of the body and the right side deals with the left but, in other ways, the two halves of the brain have different functions, cooperating together by connections through the corpus callosum. Most right-handers and 90% of left-handers interpret hearing (from both ears) and output speech in the left brain, which also specializes in sequential activities such as counting, while the right brain specializes in more global aspects such as estimating size (Figure 2). Visual information from the eyes passes along connections to the visual cortex at the back of the brain, again in subtle ways to bring together information from the left and right to give stereoscopic vision.

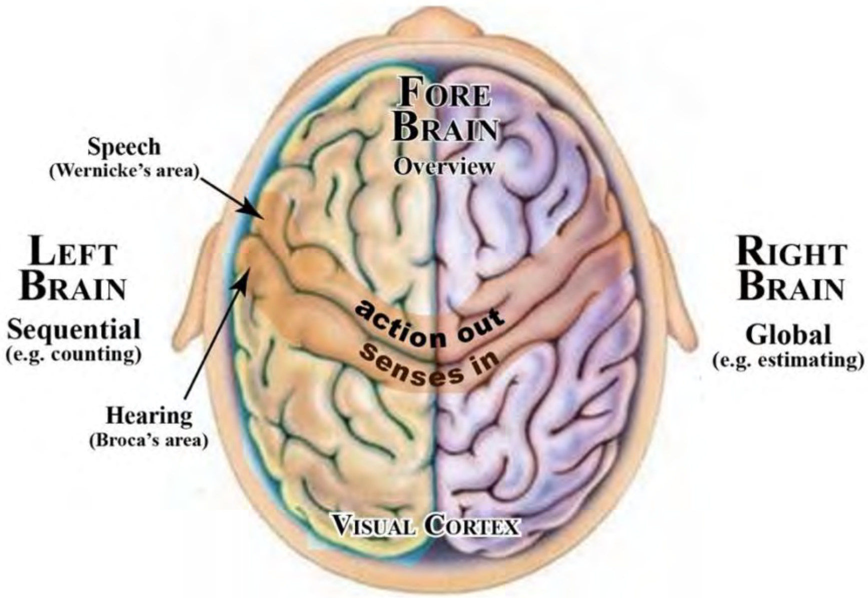

Figure 2: The brain from above

The limbic system is duplicated on each side of the brain, each with its own connections to other parts of the brain, collaborating through the corpus callosum. There is evidence that the right amygdala senses negative emotions such as fear and sadness while the left amygdala plays a role in the brain's reward system. The right amygdala also plays a role associating memories of time and place with emotional properties.[1] Another limbic structure, the *nucleus accumbens*, acts as a kind of 'pleasure centre', responding to a range of reward and reinforcement.

An immediate problem in attempting to come to terms with the structure and operation of the human brain for someone interested mainly in mathematics is that the names given to the parts relate to Latin and Greek descriptions of the position and shape as seen when the brain is dissected and give no clue as to their function. For example, the amygdala has an almond shape, the thalamus is egg-shaped, and the hippocampus looks like a seahorse from 'hippo' meaning 'horse' and 'campo' meaning 'monster'. The 'hypothalamus' is situated 'above' (hypo) the

[1] https://en.wikipedia.org/wiki/Limbic_system

thalamus and the ‘nucleus accumbens’ is a shortened form of ‘nucleus accumbens septi’ which is Latin for ‘nucleus adjacent to the septum’.

For our purposes it is more important to gain a broad sense of how the limbic structures operate. They are intimately interconnected and play a variety of roles. For example, the thalamus receives and analyses sensory input and passes it on to relevant areas of the brain. This includes sensing threatening situations and reacting in a ‘fight or flight’ response that floods the whole brain with biochemical neurotransmitters that enhance or suppress connections between neurons. The ‘fight’ mechanism enhances connections to think about the problem, while the ‘flight’ response suppresses connections to avoid it.

A major role of parts of the limbic system, such as the hypothalamus and cingulate gyrus, is to regulate subconscious autonomic functions such as body temperature and heart rate. This enables emotional responses to be translated into physical responses such as damp sweaty palms or increased pulse rate. More generally the limbic system is part of an overall relationship between the operation of the brain and our feelings, preferences and prejudices.

Conscious thinking requires information to be passed within the brain, such as visual information being taken in by the eyes, passed back to the visual cortex, then moving forward to other parts of the brain before taking conscious decisions in the frontal cortex. Meanwhile, incoming sensory data also passes directly to the limbic system and initiates automatic action before the frontal cortex takes conscious decisions (Kahneman, 2011). This underpins mathematical thinking with personal attitudes and emotions before we take considered conscious decisions.

Neurotransmitters that flood the brain to suppress connections not only affect our feelings, they affect our attitudes and ability to think about mathematics. On the positive side, successful experiences in mathematics – especially Aha! experiences linking ideas together – enhance connections, putting the brain on alert to improve mathematical thinking. On the other hand, negative experiences failing to make sense of mathematics can give rise to negative feelings, such as mathematical anxiety, that suppress connections and make mathematical thinking difficult, or even impossible.

The operation of the limbic system is a built-in part of our human nature that links our cognitive actions in mathematical thinking to our emotional feelings and attitudes. An Aha! reaction that proves to be successful can give insight to an individual and can lead a whole community into new ways of thinking. More subtly, if a community of mathematicians fail to make a particular connection, this can lead to an Uh-Huh! experience that causes the whole community to reject a particular way of thinking, with long-term consequences that may last for many generations, even for thousands of years. (We will shortly see an example when we consider the evolution of the calculus.)

3. The cultural evolution of mathematical thinking

In the book *Evolution of Mathematical Concepts*, the mathematician Raymond Wilder (1968) interpreted the historical development of number and geometry using the anthropological term ‘culture’, which he described as:

> A collection of customs rituals, beliefs, tools, mores, and so on, called *cultural elements*, possessed by a group of people who are related by some associative factor (or factors) such as common membership in a primitive tribe, geographical contiguity, or common occupation. (Wilder, 1968, p. 18.)

Cultural elements, such as counting and number, arose from needs within the culture that proved to be of benefit. Wilder speaks of *cultural stress* that occurs when there is a need in the culture to be satisfied, such as understanding the seasons to know when to plant crops, and developing counting and number systems to plan civilised activities such as barter and trade.

Cultures benefit from shared elements that are stable and useful. These may *diffuse* from one culture to another, but this is likely to take time to do so, called *cultural lag*, and may even encounter *cultural resistance* if the new element challenges current elements that are considered to operate successfully. To balance elements that resist change, I would add the notion of *cultural stability*, which seeks to maintain familiar elements that allow the culture to continue to operate in a shared coherent manner.

In recent times, many changes have occurred in our cultural approach to life in general and mathematics in particular, aided by advances such as those in digital technology. As new possibilities arise, mathematicians and mathematics educators have sought to implement them in their approaches to mathematics and its teaching. But cultural lag and cultural resistance affect progress, as Wilder cautioned:

> Attempts to change the direction of mathematical research by individuals who deem the tendencies prevailing at the given time to be "wrong," seem to be of little avail. Only strong environmental and internal pressures [...] appear to be effective in changing the course of mathematical development. (ibid. p. xi.)

3.1 The example of calculus

In the 1980s, the invention of dynamic computer graphics allowed a new approach to differentiation in which the student could use dynamic graphics to zoom in on familiar graphs such as polynomials, trigonometric and exponential functions to reveal that under higher magnification they 'looked straight'. This offered a completely new frame of reference to interpret the notion of differentiation: knowing that a graph is locally straight enables the learner to look along it to see its changing slope to give a meaning for the changing slope function (Tall, 1985). For a function such as $f(x) = x^2$, it is possible to draw a 'practical slope function' $(f(x + c) - f(x))/c$ for small values of c, in this case, $2x + c$, to see it visibly stabilize on $2x$ for small c.

This enables the student to link several different planes of reference for the graph and for its change in slope in terms of:

- the visual picture of the graph,
- the symbolic representation of the function,
- embodied gestures tracing the graph and its visual change in slope.

In my very first study using my Graphic Calculus software in a mathematics class of sixteen-year-olds (Tall, 1986), some of the students who were also taking physics had been told in their physics lessons that the derivative of $\sin(x)$ is $\cos(x)$. Student Graham wanted to know why. He used the software to draw the practical slope function and there was a delighted Aha! to see that the changing slope of the graph of $\sin(x)$ looked like $\cos(x)$. The students were then asked what they thought the slope function of $\cos(x)$ might be and guessed it might be $\sin(x)$. On drawing the slope function (figure 3) there was an even bigger Aha! when the slope function of $\cos(x)$ looked like the graph of $\sin(x)$ *upside down*, which gives a meaning why the derivative of $\cos(x)$ is minus sin(x) (ibid., p.192).

At this point there was an Uh-Huh! response from student Brian who had transferred from another school where he had been well-drilled in examination techniques. He did not see why he should be asked to guess the formula. Why couldn't the teacher say what the derivatives were so that he could learn them to pass the test?

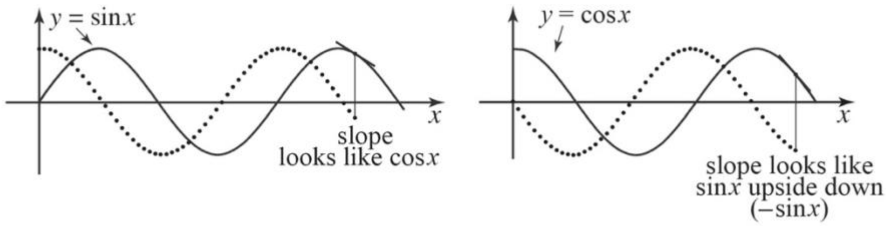

Figure 3: plotting the slope functions for sin$x$ and cos$x$

Other Aha! Moments occurred as the teacher and students used embodied gestures to give meaning to other calculus concepts. For example, in discussing the idea of maxima and minima, the teacher sketched a maximum on the board and traced along it with his finger. He asked the students how they might tell if there was a maximum or minimum and student Andrew gestured with his hand, suggesting that the slope would be positive before and negative after. There was general assent from the rest of the class (ibid., pp. 195-6).

Over time, as the students could see dynamic graphical ideas on the computer, they were shown conceptual examples that enhanced the meaning of the calculus so that it became easier to imagine subtle properties without the need for the computer. Even Brian steadily realized that it was of value to make sense of the ideas.

3.2 Cultural stress in the development of calculus

The graphic approach to the calculus was quickly taken up by projects in the UK and used as an introduction to differentiation. But cultural pressures required the students to pass tests that were mainly based on symbolic techniques for calculating symbolic derivatives. Cultural resistance came into play and the curriculum to this day continues to focus on aspects that can be tested. This can be interpreted not only in terms of a desire for cultural stability but also an Uh-Huh! reaction to reject an alternative approach that builds on natural intuition rather than on the formal definition inherited from the precision of mathematical analysis.

In the USA, the MAA report (Bressoud, Mesa, Rasmussen, 2015) on college calculus complains that the students taking a first course in calculus, either in school or in college, focus on a limited ability to perform routine symbolic processes. The curriculum does not attempt to mention meaning in terms of local straightness. In high school calculus, the forces of cultural stability operate to maintain traditional elements that can be tested in a multi-choice examination system (College Board, 2018).

Experts in formal proof often distrust intuition, especially in analysis because, without a proof, so many intuitions can be proved to be false by a carefully designed counter-example. This distrust for intuition can lead to widely shared beliefs that impede insight for alternative theories. For instance, there is a common belief that the completeness axiom fills out the whole number line and there is no room for infinitesimals, leading to an 'Uh-Huh!' reaction from pure mathematical analysts who deprecate the use of infinitesimals. Meanwhile, applied mathematicians often think of 'infinitesimal calculus' in terms of variable quantities that can be 'arbitrarily small'. We will return to this difference of meaning in §7.4.

4. Transitions in mathematical meaning over the long term

In *How Humans Learn to Think Mathematically* (Tall, 2013), I formulated a long-term framework for the evolution of ideas in the individual and in communities over time, from early beginnings to highly sophisticated theories. This reveals transitions in meaning as new contexts are encountered where established methods of approach require modification to proceed.

Mathematical thinking takes place in the human brain where the left side usually deals with language and sequential thought and the right side deals with more global aspects, with the two sides cooperating as a whole. The fore-brain takes conscious decisions through more in-depth mathematical thinking while the limbic system in the mid-brain links cognitive actions to emotional feelings and attitudes.

This leads to a cognitive framework for mathematical thinking which integrates human perception, action and internal imagery into a long-term development that I termed *conceptual embodiment*, together with the long-term development of *operational symbolism* and the more sophisticated set-theoretic development of *axiomatic formalism*. I referred to these as three 'worlds' of mathematics. Broadly speaking, they may be related to the global activity of the right brain, the sequential activity of the left brain and the reasoning of the fore-brain, with connections between these cognitive activities working together and linking to emotional attitudes and feelings through the limbic system. I will consider this in more detail in §7.3.

As we encounter new contexts, we need to link together ideas in new ways. Sometimes this involves an Aha! Moment linking together previously unconnected planes of thought, but sometimes the link may not occur, preventing the transition from being made. This happens with individuals as they encounter more sophisticated mathematical contexts in learning mathematics and also in different communities of practice as cultural forces come into play.

4.1 Transition in context

As an individual learns to think mathematically over the long term, different kinds of transition occur within and between different worlds of mathematics. In conceptual embodiment, practical activities involve recognizing and describing properties of objects, later developing into theoretical definition and deduction. Van Hiele (1986) analysed the growth of ideas in geometry in terms of successive levels where the language changes subtly in meaning. This applies not only in Euclidean geometry, but in the transition to calculus where the definition of a tangent in Euclidean geometry as a 'line which touches the curve in one point only' defines a tangent to a circle but is problematic in defining a tangent for more general curves in calculus.

Operational symbolism has more complicated transitions. Not only do operations, such as addition, become mental objects, such as sum, but there is also a succession of new number systems with new properties, from whole numbers, to fractions, signed numbers, rationals and irrationals which make up the real number system, real and imaginary numbers in the complex number system, and various extensions such as vectors, matrices and so on. Algebra grows out of the patterns of arithmetic and develops into a theoretical framework of symbol manipulation to formulate and solve problems. Theoretical proof arises from formulating definitions and making deductions based on practical experience and mental imagination. There is a major change from the theoretical use of definition and proof to the axiomatic formal mathematical world in which mathematical structures are formulated solely in terms of set-theoretic definitions and all other properties must be deduced by formal proof.

4.2 Cultural transition between communities

As mathematics evolves over the generations in different cultural settings, it goes through various transitions that may lead to very different kinds of mathematics. The cultural differences may be of such a nature that it builds a boundary between them that is an impediment to linking them together. These cultural differences may arise between different civilisations, different religions, or between differing subcultures in the same society, such as theoretical engineering and formal pure mathematics. Using religious terminology, if an individual in community A changes beliefs to those in community B, this may be considered as a *transgression* by community A, and an *enlightenment* by community B. If their differences are fundamental, they may seem irreconcilable. However, it may also be possible to build bridges between the two in a *multi-community overview* in which the two sides recognise aspects on which they agree and use them as common ground to address their differences (figure 4).

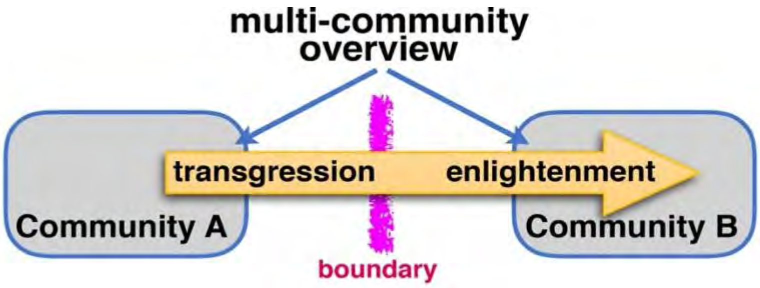

Figure 4: Cultural transition between different communities (Tall, 2019a)

In summary, the possibilities may be characterised as:

Impediment: inability to leave the current community to cross over a boundary
Transgression: crossing out of the current community over a boundary
Enlightenment: crossing into a new community over a boundary
Overview: encouraging communication between communities.

Examples include differences between communities of pure and applied mathematicians, between mathematicians and educators, between politicians who prescribe the curriculum, curriculum designers, teachers and assessors, between different levels of teaching in early learning, primary, secondary, university, and different forms of expertise in mathematics.

7b.4.3 Personal transitions

A similar analysis may be performed for an individual learner attempting to make sense of increasingly sophisticated mathematical ideas over the long-term. The individual may be familiar with one context, say context A, and is attempting to make sense of a second context B which may not be understood because the change in meaning places a mental boundary between the two. For example, the individual may be familiar with whole number arithmetic, but is unable to contemplate taking away a larger number from a smaller one, because it is not possible to take away more objects from a collection that has less.

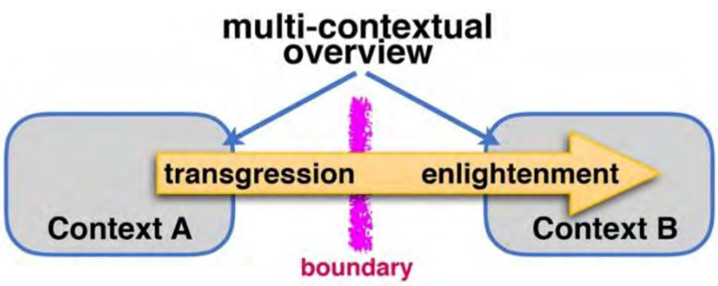

Figure 5: Personal transition between different mathematical contexts (Tall, 2019a)

In the case of an individual seeking to make a change in context within a single community, the possibilities include:

Impediment: inability to change context
Transgression: unwillingness to change context
Enlightenment: ability to change context
Overview: ability to switch between contexts.

Examples include generalising number systems from counting numbers to fractions, to signed numbers, to rational numbers, reals, complex numbers, from arithmetic to algebra, from practical drawing to Euclidean proof, through changes in meaning in geometry (van Hiele, 1986), from school mathematics to university, and so on.

The learner may also be faced with the possibility that he is able to invent new ideas that are not familiar to the teacher. My son Nic at the age of 4 years and 5 months was watching the television weather forecast and asked what 'minus two centigrade' meant. I took him into the garden and showed him a wall thermometer and talked about temperatures above and below zero, which is when water freezes to form ice. We had the first of a number of conversations, some of which I recorded, that allowed me to later write up a number of instances of an Aha! Moment, including the idea that starting at a temperature of 2° and going down 3°, ends up at –1° (Tall, 2001). At school, he was invited to write down a sum of his own choosing and invented his own symbolism using a downward pointing arrow to write this as 2 ↓ 3 = –1. His teacher marked it wrong (Uh-Huh!). I asked Nic if he was upset by this, but he simply said, 'No, he [the teacher] didn't understand.'

5. Problem solving and Skemp's goals and anti-goals

To encourage undergraduates to make sense of mathematics for themselves and not simply learn facts by rote, I designed a course based on *Thinking Mathematically* (Mason et al., 1982) for mathematics undergraduates. It was a ten-week course with one two-hour large group problem-solving session with around 40 students, and one-hour small group sessions to reflect on the problems more deeply. The two-hour session began with a plenary of about 20 minutes to introduce the focus for the day, then I left the room for an hour and, on my return, always found the class buzzing with activity ready for a concluding discussion of around 20 to 30 minutes. As a policy I did not work on the problems myself beforehand, so that the solving activity was genuinely student-led, though, over several years I did see patterns emerging in the solutions which I did not reveal to the students until we had reflected on their ideas. The students were encouraged to keep notes on their developing work, including writing 'Aha!' when they had an insight and 'Stuck!' when they came to a difficulty they could not resolve. The assessment included a written analysis of their problem-solving development, so it was actually sometimes an advantage to be stuck.

To encourage the students to reflect on their emotional reactions, the course was accompanied with a study of Richard Skemp's (1979) psychological theory of goals that are desired objectives and anti-goals that are to be avoided. He theorized that very different emotions are sensed as one moves towards or away from a goal or antigoal, and other feelings were caused by the belief that a goal could be achieved or an anti-goal avoided.

A goal that one believes is achievable is accompanied by a feeling of confidence, which may change to frustration if it proves subsequently to be difficult to achieve. Frustration sensed by a confident person is likely to act as a positive encouragement to redouble the effort to achieve the goal. Moving towards a goal gives pleasure and moving away from it gives unpleasure—a term used in Freudian analysis to denote the opposite of pleasure.

Coping with an anti-goal is quite different. According to Skemp, an anti-goal that one believes one can avoid gives a sense of security but, when it cannot be avoided, the emotion turns to anxiety. Moving towards a goal instils a sense of fear, while moving away gives relief.

This theory is represented in figure 6 (as drawn in Tall, 2013, p.120) where arrows represent movement to or away from a goal or anti-goal and smiling or frowning icons represent the belief related to the ability to achieve a goal or avoid an anti-goal.

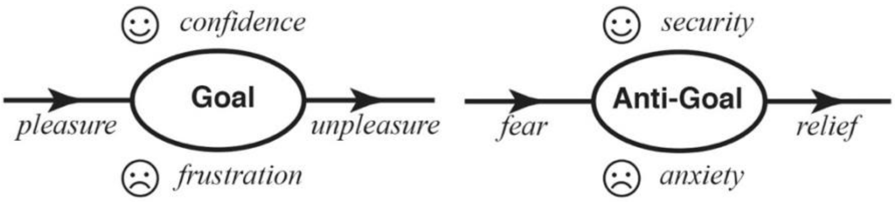

Figure 6: Emotional reactions to goals and anti-goals

This reveals the vast difference between positive emotions of confidence and pleasure relating to goals that are considered achievable and emotions relating to anti-goals which offer, at best, a sense of security and relief and, at worst, a sense of anxiety and fear.

This theory was included to encourage students to realise that when they were feeling negative thoughts, this should not be interpreted as personal failure, but as a sign that the path being taken was not currently productive and it may be helpful to reflect on other possibilities, such as looking at a simpler case to see if progress can be made in another direction.

6. Provoking Aha! in Lesson Study

Japanese Lesson Study is a teaching philosophy that has been developed for over a century, working in large classroom settings. In mathematics, lessons are planned to encourage each learner to think about different approaches to any given problem to make sense of the ideas in their own way (Isoda, Stephens, Miyakawa, 2007). The sequence of lessons is carefully planned and tested to develop an approach that can be used by the wider teaching community. The lessons are organised to begin with simple problems that can be solved in a range of different ways, so that at a later stage, when the main topic is introduced, the learners have a specific array of strategies available that enable the teacher to build up the mathematics using their ideas. In essence, the system is designed to provoke a corporate Aha! reaction where different learners participate using their own ideas.

It was my privilege to work as a consultant with the APEC (Asia Pacific Economic Cooperation) study group on Lesson Study involving 20 communities around the Pacific and to contribute to the theoretical framework. Tall (2015) presents an analysis of four specific lessons covering a range of different kinds of topic: a non-routine problem involving triangular and rectangular shapes (grade 2), multiplication of a double digit number by a single digit number (grade 3), area of a circle (grade 5), thinking systematically (grade 6). What is impressive is the way that children of widely differing levels of development can benefit in personal ways from the same lesson. For example, the lesson on multiplication we saw had been preceded by a lesson solving a problem to find the number of disks in two rows which happened to have twenty disks. In the new lesson the children were faced with three long parallel rows of disks. Figure 7 shows the working of an individual child, and two parts of the notation on the long wallboard filling the whole width of the classroom wall.

Figure 7: calculating 3 times 23, visually and symbolically

On the left side of the wall-board, children have taken it in turns to subdivide the rows into various subsets: 20+3, 10+10+3, 10+3+10 to multiply each part by 3 and add them together. There was an audible Aha! from many in the class who were moved by the beautiful symmetry with two tens on either side of a three in the middle, something that I had not expected as my experience with arithmetic saw 10+3+10 being the same as 10+10+3. On the other hand, I was moved when a little boy stood up and remarked that the solutions were essentially the same, all separating each line of disks into a twenty and a three. Other children suggested alternatives such as 9+5+9 or 11+12 and another possibility in terms of 2 ten-yen coins and 3 one-yen coins. These were written on the centre of the board (not shown) and the picture on the right shows the teacher near the end of the discussion as he asks questions of the children and writes up the responses to show the links between the visual and symbolic representations. At the end of the session, the whole story unfolding the ideas remained on the board from left to right.

There were several different planes of reference, including the embodied layout of the three rows subdivided in various ways, the embodied symbolism of 2 tens and 3 ones, and the standard symbolic layout for long multiplication. Some children were pleased with their clever alternative representations such as 9+5+9, but this was overcomplicated for others. Some experienced an Aha! Moment linking together different frames of reference. Even those who were struggling with the arithmetic had the possibility of seeing that the standard layout was the simplest and most efficient symbolic way of calculating the product.

The children were asked to write up their experiences for homework. What impressed me most was the total enthusiasm of the whole class and the impression that children of different abilities seemed to make progress appropriate for their personal needs.

6.1 Cultural resistance to transfer

After my experience with Lesson Study in Japan, I was invited to act as a consultant for a research project in the Netherlands, the home of 'realistic mathematics', which focuses on children solving realistic problems to make sense of mathematics for themselves. Although this approach found international interest, in its own country there was cultural resistance from university academics who found students were arriving at university lacking in traditional mathematical skills.

The research project planned to teach high school students calculus of polynomials and trigonometric functions using a Lesson Study approach, together with my three-world format blending embodiment and symbolism, supported by Skemp's (1976) theory comparing instrumental and relational understanding. The teachers and researchers in the research team were committed to making a success of their project but were hampered by cultural differences. Each of the three Dutch teachers initially interpreted the project according to their own personal views based on the cultural approach in the current Dutch syllabus, competing to present their best version of the mathematics with little room for student discussion (Verhoef et al., 2014). An awareness of student thinking developed as the project continued in subsequent years, settling on a locally straight approach to differentiation using software developed in *GeoGebra* by one of the team (Verhoef et al., 2015). The project exemplified the cultural effects on diffusion between different traditions, with the desire for cultural stability, the cultural lag in incorporating new ideas and the cultural resistance of subcultures.

Personal Aha!s and Uh-Huh!s

It has been a privilege to live through a time of significant innovation in human society including technological change that was unthinkable when I was young. As Wilder commented:

> When a cultural system grows to a point where a new concept or method is likely to be invented, then one can predict that, not only will it be invented but that more than one of the scientists concerned will independently carry out the invention. […] original discoveries, especially if related to an important and rapidly developing area, will usually be made by more than one researcher. (Wilder, 1981, p. 23.)

In my own case, I was fortunate to put together several different planes of reference, often arising in Aha! insights inspired by sharing ideas with colleagues. Sometimes an insight may not reach the wider community, it may spread to others and it may evoke various Uh-Huh! reactions arising from a desire for cultural stability, or a rejection through cultural resistance. In the following I will consider various examples.

7.1 Concept image

In 1969, I was appointed to a position in the mathematics department at Warwick University as 'Lecturer in Mathematics with Special Interests in Education'. The appointment was to oversee the mathematical content of a new undergraduate degree for mathematics teachers at Coventry College of Education on an adjoining campus, but I was encouraged to research in mathematics and/or in mathematics education. At the time, apart from behaviourist theory and the general theories of Piaget, there was no research tradition in mathematics education and I began to gather data about our undergraduate mathematics students' ideas. This included the observation that most undergraduate students and school teachers thought that 0.999… was 'just less than 1' and a number of subtle meanings for phrases such as 'some rationals are real' which many considered false because 'all rationals are real.' My problem was that I had no theoretical framework to make sense of all these disparate pieces of information.

Then Shlomo Vinner visited Warwick in 1980 and showed me his latest paper which said:

> Let C denote a concept and let P denote a certain person. Then P's mental picture of C is the set of all pictures that have ever been associated with C in P's mind.
>
> Besides the mental picture of a concept there might be a set of properties associated with the concept (in the mind of our person P). […] This set of properties together with the mental picture will be called by us the concept image.
>
> (Vinner & Hershkowitz, 1980, p.177, Vinner, 1983, p.293.)

Aha! In a flash I realised that the term 'concept image' brought all my data together. I drafted a paper to share with him between lunchtime on a Monday and late afternoon the next day in which I freely used his idea to formulate the definition in the following terms:

> We shall use the term concept image to describe the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes. (Tall & Vinner, 1981.)

The two final words 'and processes' did not appear in his original, which related to geometry but were essential in the new paper on the processes of limit and continuity. The original idea of 'image' also included visual imagery of symbolism, but not yet in terms of the relationship between process and concept. It would be another ten years before the notion of concept image of a process was interpreted as the duality of process and concept in the flexible use of 'procept'.

Vinner's original idea referred to concept image and concept definition as 'two cells (not biological) in the cognitive structure' while I referred to them in terms of the operation of the biological human brain. Despite the fact that the two authors have different interpretations for the notion of concept image, it has entered the folk-lore of the subject with over 2,500 citations and was selected for inclusion in the NCTM publication as one of seventeen papers from the 20th century that every mathematics teacher should read.

7.2 Symbols as process and concept

The 1970s and 80s saw the notion of 'compression of knowledge' from process to object emerge in mathematics education. In 1970 I specified Dienes' *Building Up Mathematics* (1960) as one recommended text in a course I gave, which included the notion of 'the predicate of a sentence becoming the subject in another'. My departmental chairman Christopher Zeeman talked about the grammatical notion of gerund, where the participle 'running' in the sentence 'I am running' becomes the subject in the sentence 'running is good for my health' and applied the construction in mathematics. I was fully aware of Dienes' idea that repeated addition becomes the product and repeated multiplication becomes the power of a number, but I could not see how to take the idea further (Uh-Huh!). When I was invited to Israel in 1986 to demonstrate my Graphic Calculus software, Anna Sfard told me of her PhD study on operational and structural approaches to mathematics, which I interpreted differently as a combination of Dienes' ideas and the French Bourbaki structure (another Uh-Huh!). I visited Ed Dubinsky in early 1989 when he was developing college level mathematics based on Piaget's reflective abstraction, compressing process into object in his APOS theory (Action → Process → Object → Schema).

In the autumn of 1989, my colleague Eddie Gray was completing his PhD (Gray, 1993) under my supervision and explained his experiences talking to children aged 6-11 about their methods of calculating simple addition and subtraction. Anna Sfard was a visitor to the university at the time and the three of us had various discussions about her operational-structural theory which was published shortly afterwards (Sfard, 1991). As Eddie and I considered his data, it became evident that when a child spoke an expression such as 'three plus two', then we were unable to distinguish as to whether the child was thinking about the phrase as a process (an instruction to add) or as a concept (the object as a sum). Then a light-bulb moment occurred. Aha! Perhaps the child was able to switch from one meaning to the other or deal with both at the same time. To be able to talk about this, we needed a new word that could be used to represent either, or both, and the name 'procept' was formed. (Anna Sfard later suggested the alternative word 'project' to represent process or object, but this was not appropriate.) It was immediately evident that the term procept applied throughout mathematical expressions as process or concept, including expressions in arithmetic, algebra and calculus.

The PhD thesis was not received well by its examiners, both internationally known professors, one a philosopher who objected to the notion of procept, the other a mathematics educator who objected to the methodology (Uh-Huh!). To counter this conclusion, Eddie and I wrote up the research (Gray & Tall, 1994) *in its original form*, where it has been widely accepted, with over 2,500 citations in the research literature.

7.3 Three worlds of mathematics

Over the years I had experience of mathematics learning covering a wide age range from young children to university mathematics at undergraduate and graduate level. So I was primed for the next major personal discovery. The insight that led to the three worlds of mathematics occurred in a single Aha! Moment in 2002, reviewing the research data of my doctoral student Anna Poynter.

As an engineer, she taught mathematics using physical modelling with a free vector represented by pushing a triangle across a table ending up pointing in the same direction. Student Joshua explained that the sum of two free vectors was the single free vector that *had the same effect* as the combination of two individual free vectors. This key unlocked the door of the relationship between embodiment and process-object encapsulation. The switch in focus from operation to effect paralleled the compression of a process into an object. This gave me a link between the previously unconnected frames of reference in embodiment and symbolism. Linking this to the formal theory of vector spaces at university level I declared: 'there are three worlds of mathematics: embodiment, symbolism and formalism.'

I had conflicting feelings. The insight made sense to me, but I realised that experts from different specialisms were likely to interpret the framework in different ways and it needed considerable reflection and discussion in seminars before it could be published. The notion of 'embodiment' for me started with the physical embodiment of Dienes and developed through my use of visual dynamic representations and enactive gestures in the calculus. In geometry, embodiment developed in sophistication from practical drawing to theoretical definitions and Euclidean proof, so I named the long-term development '*conceptual* embodiment' to relate to its long-term development from practical activity to mental imagery and theoretical definition and deductive proof.

In the literature, 'symbolism' is used to refer to all kinds of symbols, including words, mathematical symbols and pictures used to represent something. Because I wanted to focus on the flexible (proceptual) use of mathematical expressions, I first called it 'proceptual symbolism', but later changed it to 'operational symbolism' to include the reality that many individuals learnt symbolic algorithms by rote-learning.

The more sophisticated levels of embodiment and of symbolism both include formal definitions and deductions, based on selecting familiar properties as the foundations of a theoretical framework. This includes Euclidean geometry, with its axioms and 'common notions' as a basis for Euclidean proof, and arithmetic and algebra, based on the rules of arithmetic. To describe this use of definition and deduction I used the term 'theoretical mathematics'.

This is very different from formal mathematics based on set theory where structures are defined using set-theoretic axioms and definitions formulated as quantified statements and all other properties of the structures must be deduced from these axioms and definitions using formal proof. I term this type of mathematics 'axiomatic formal mathematics'. To comply with the pure mathematician's view of Euclidean proof as the first stage of formal mathematics, I see formal mathematics consisting of theoretical and axiomatic formal mathematics. Most individuals encounter only practical mathematics in everyday use and theoretical mathematics in applications.

An outline of this cognitive development of the three worlds is given in figure 8.

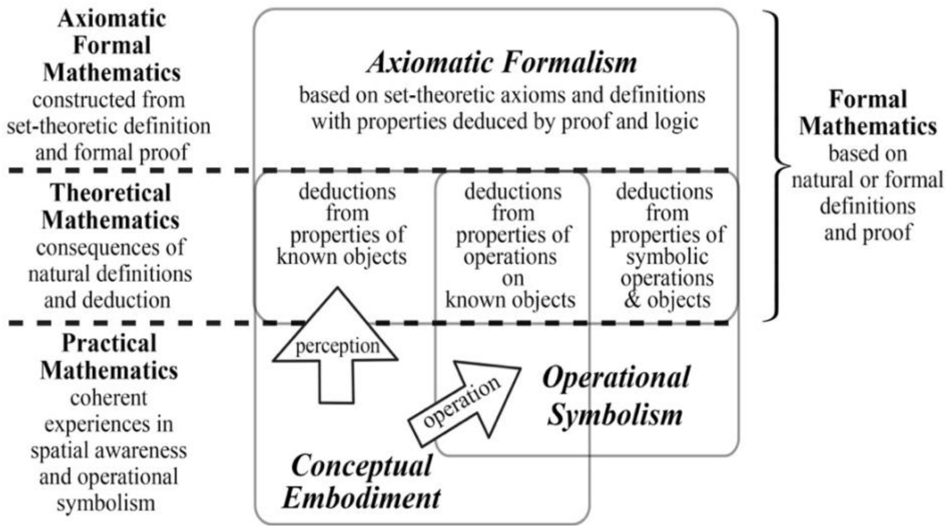

Figure 8. The three-world framework

Given the wide-ranging communities that use mathematics in some form or other, this is likely to be interpreted very differently by individuals who do not share the same frames of reference. Some will only be aware of part of the framework and others will have very different interpretations of various aspects. The framework was published in Tall (2013), seeking to present the ideas in a manner that could make sense to the general reader. Since then, further insights have broadened the framework.

7.4 Structure theorems

The three-world framework as given in figure 8 makes it seem as if axiomatic formalism is the highest manifestation of mathematical thinking. However, this does not mean that we cannot have successively higher levels of embodied and symbolic thinking. An axiomatic structure such as a complete ordered field specified by axioms of arithmetic, order and the completeness axiom (that every non-empty subset that is bounded above has a least upper bound) can be proved to have structures that allow it to be embodied (as points on a number line) and symbolised (as infinite decimals). In this case the familiar line consists only of real numbers.

However, it does not mean that it is not possible to have a number line that contains infinitesimals. They may not lie in the real number system, but they can easily be imagined in a larger system that includes real numbers and infinitesimals.

I will now explain how this can be done using axiomatic formal mathematics. For a reader who does not share this mathematical culture, I will offer an informal interpretation of the principles involved. So do not despair if the following discussion makes you feel uncomfortable (Uh-Huh!). It is an essential part of the big picture to show that experts in different communities may have beliefs that are at variance with each other yet are totally satisfactory in their own context. In particular, I will show how the completeness axiom, which proves that the real numbers cannot contain infinitesimals also proves that *any* ordered field K that contains the real numbers as an ordered subfield *must* contain infinitesimals.

I will refer to elements of the real numbers $\mathbb{R}$ as 'constants' and elements of the larger system K as 'quantities'. A quantity x is said to be *finite* if it lies between two real numbers a, b, so that, in the ordering of K, we have $a < x < b$. A quantity ε is said to be an *infinitesimal* if $\varepsilon \neq 0$ and $-a < \varepsilon < a$ for every positive real number a.

It is then straightforward to prove:

Structure Theorem for any ordered field extension K of the real numbers.
Every finite quantity is either a real number or a real number plus an infinitesimal.

The proof is straightforward. If x is a finite quantity, the set of real numbers $L = \{t \in \mathbb{R} \mid t < x\}$ is non-empty (because it contains a) and is bounded above by b, so it has a unique least upper bound $c \in \mathbb{R}$. Let $\varepsilon = x - c$, then, by a contradiction argument, it can be proved that ε is either zero or infinitesimal. The unique real number c is called the *standard part* of x, written as $c = \mathrm{st}(x)$.

Infinitesimal detail for a quantity t near x can then be seen using the linear map

$$m(t) = (t - c)/\varepsilon$$

This map is called *the ε-microscope pointed at c*. The subset V of quantities such that $(t - c)/\varepsilon$ is finite is called *the field of view* of the microscope. Taking the standard part of $m(t)$ gives the *optical ε-microscope pointed at c* as $\mu : V \to \mathbb{R}$ where

$$\mu(t) = \mathrm{st}(m(t)) = \mathrm{st}(t - c)/\varepsilon)$$

For a real number k, $\mu(c - k\varepsilon) = k$, so the optical microscope maps the field of view onto the whole real line.

The notion of optical microscope was first published in Tall, 1982: more detailed information can be found in Tall, 2009. This can be generalised to multiple dimensions by using an optical microscope on each coordinate, for instance in two dimensions the (ε, δ)-microscope pointed at (c, d) is

$$\mu(s, t) = (\mathrm{st}((s - c)/\varepsilon, \mathrm{st}(t - d)/\delta)$$

It is not appropriate to go into further detail here. I content myself by showing a picture in two dimensions where the infinitesimals ε, δ are taken to be equal, allowing an infinite magnification of a differentiable function $y = f(x)$. Here I have denoted the points in the image with their original names in K^2 and denoted any real number change in x in the image as dx and its corresponding change in y as dy. Then $f'(x) = dy/dx$.

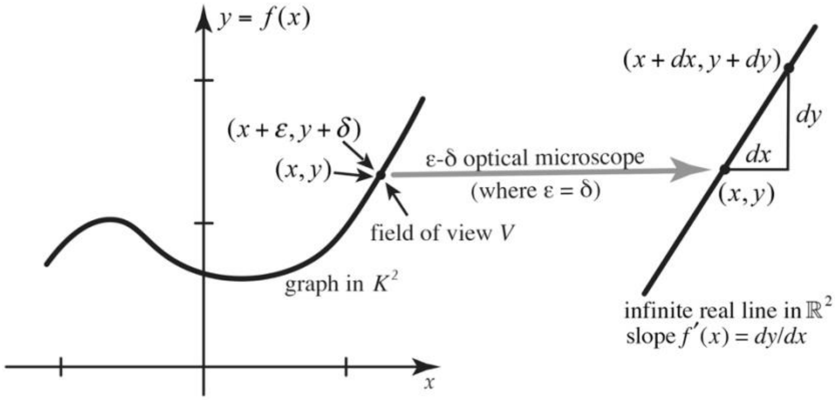

Figure 9: Infinite magnification of a locally straight graph

7.4.1 The full cognitive framework for the three worlds of mathematics

The example of extending previous meanings of mathematical contexts – here developing new embodied and symbolic forms of axiomatic mathematics – is yet another example of a system that works in one context being extended with new meanings in a more sophisticated context. Those who make sense of the context may see the extension as an Aha! insight that increases their power in mathematical thinking. Others may not share all the required planes of reference and may reject the transition as a transgression into unknown territory as an Uh-Huh!, with the power of cultural stability causing them to see their action in a positive light as they remain in a framework that is familiar to them and their peers.

The notion of 'structure theorem' applies widely in axiomatic formal mathematics where many axiomatically-defined structures can involve proving theorems that enable more sophisticated levels of embodiment and symbolism. Our practical perception and operation are limited by the nature of our biological brains and the three-dimensional world we live in, but our understanding of these limitations can allow us to imagine mathematical ideas beyond our physical experience.

Not only can we imagine multi-dimensional ideas in analysis, including such things as the two-dimensional Klein bottle that cannot be represented in three dimensions, we can also prove structure theorems in algebra, topology, and other areas of mathematics that offer new forms of embodiment and symbolism. For example, the formal notion of vector space includes definitions of linear independence and spanning sets that allow us to speak of finite dimensional vector spaces over the real numbers that we can embody in two and three-dimensional space. Meanwhile, using symbolic coordinate systems and matrices it is possible in theory to operate symbolically for any value of n.

The formal notion of finite group with *n* elements can be interpreted with the operations of the group permuting the elements, allowing us, in principle, to visualize ideas as operations on a figure with n vertices and operate symbolically with the elements of the group in terms of the theory of generators and relations.

The cognitive frame for the three worlds of mathematics therefore goes beyond the axiomatic level to new forms of embodiment and symbolism that can be imagined as an upward spiral of successive levels or, more simply as structure theorems folding back from formal mathematics to embodiment and symbolism (figure 10).

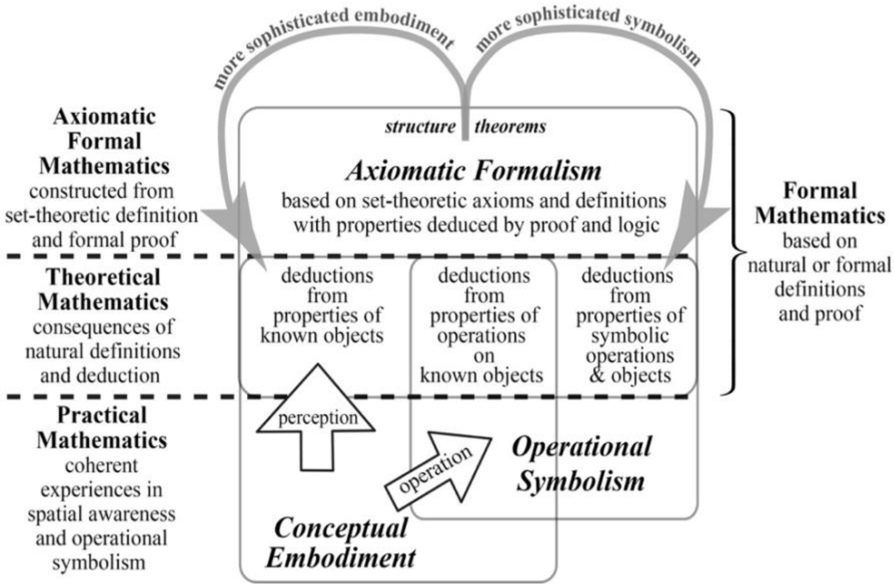

Figure 10: From formalism, folding back to more sophisticated embodiment and symbolism

This, in turn, needs to be seen in the wider aspects of mathematical thinking, including personal attitudinal and emotional development within evolving cultures that encourage or impede progress.

7.5 The Articulation Principle

Since the formulation of the long-term development in *How Humans Learn to Think Mathematically*, I have sought to reflect on fundamental ideas that are widely recognised by teachers and experts and yet also have potential to give long-term insight into mathematical thinking at all levels. These include an amazing Aha! experience shared with my then 11-year old grandson that completely changed my own view of the long-term development of operational symbolism (Tall, Tall & Tall, 2017).

The idea is simple. When we *speak* a mathematical expression, or hear it spoken, then its meaning is affected by the way it is articulated. Generally speaking, a phrase like 'two plus two times two' is processed in the order it is heard. However, we can give different meanings by leaving slight gaps between phrases, here denoted by three dots '…' (an ellipsis). Thus

'Two plus two … times two' is interpreted as 'four … times two', giving 'eight',

while

'Two plus … two times two' is interpreted as 'two plus … four', giving 'six'.

Speaking informally to different individuals at different ages, including young children and adults, I have found that almost all are able to make sense of the distinction, often responding with an audible 'Aha!'. This includes many who admit to mathematics anxiety and also many responsible adults who use mathematics professionally yet do not remember the rote-learnt rule ‘multiplication takes precedence over addition’. For so many, 2+2×2 is calculated in the order it is presented as 8, not the technically correct value 6. *Uh-Huh!*

The reader should consider different articulations of the following expressions:

‘five take away three plus one’

‘minus three squared’

‘the square root of nine times nine’.

Written as symbolic expressions, brackets can be used to distinguish between the two meanings, for example between $(-3)^2$ and $-(3^2)$. The reader should experience this by writing the other two expressions using brackets.

The huge difference between giving meaning through articulation and learning by rote is that making sense using articulation connects ideas together while rote learning may not. This has been elaborated in a number of papers written for differing communities including elementary mathematics teachers (Tall, 2017), cognitive scientists (Tall, 2019b) and mathematicians (Tall, 2019c).[2] I introduce the Articulation Principle as follows:

> **The Articulation Principle**: The meaning of a sequence of operations can be expressed by the manner in which the sequence is articulated.

Note that, even though this applies to mathematics, it is not a mathematical definition. Instead, it links different frames of reference in ways that have the potential to make sense for a wider population. It seems to me that this is something that is widely ‘known’ implicitly by many teachers and experts, but for some strange reason, as far as I know, it has not been formulated as an explicit principle for long-term mathematics learning in teacher preparation or in research.

It is a simple idea that can be introduced at any level, to prepare the young learner for meaningful operations in arithmetic and algebra or at any later stage to offer insight to rote learners who have so far failed to build meaningful connections. It can be coupled with the idea of the flexibility of expressions as processes or objects to offer a coherent development of the meaning of increasingly sophisticated symbolism. This offers a major extension of the original conceptual embodiment of my original framework in Graphic Calculus. It now includes a meaningful long-term development of operational symbolism, the further axiomatic formal approach to mathematical analysis and more sophisticated forms of embodiment and symbolism. It opens up the continuing evolution of a broader theoretical framework for the whole long-term development of mathematical thinking.

7.6 Towards a framework for long-term meaningful mathematical thinking

A theoretical framework does not evolve in a sequential manner. As the previous discussion shows, I first encountered the notion of ‘embodiment’ from the work of Dienes manipulating physical materials. I extended this to manipulating visual imagery by programming software to allow the learner to see the local straightness of familiar graphs and trace along the curve to see the practical slope function that visually settles on the derivative as the ‘theoretical slope function’. Paradoxically, because our vision is less precise than numerical calculation, this allows us to have a sense of the derivative as the slope function, while the numerical approximations, calculated sufficiently accurately, may be sensed as getting ‘as close as is desired, but never quite reaching the limit’. Embodiment offers a meaningful perceptual ‘sense’ of the derivative as a mathematical object produced by the process of getting close. For learners it may be an Aha!, offering fundamental insight, while for pure mathematical experts it may be an Uh-Huh! because it is seen as offering an intuitive version that lacks formal precision.

[2] My papers are available for download in pre-publication form from http://homepages.warwick.ac.uk/staff/David.Tall/downloads.html

However, now that it can be shown that the use of structure theorems can extend formal structures to more sophisticated forms of embodiment and symbolism, pure mathematicians should consider how sophisticated embodiment can support the development of more subtle mathematical theories. We all differ as individuals and it would be sensible to shift to a multi-community overview to share our different perspectives rather than to engage in mathematics wars between different communities.

The continuing development of theory requires an ongoing clarification of the differences that occur between current theoretical positions, particularly between those relating to the working of the human brain as represented in neurophysiology compared to what may be observed by teachers, learners, mathematicians and other users of mathematics.

Neurophysiological evidence is collected using various forms of brain scanning technology limited to changes that occur in a period of a couple of seconds. The limbic system reacts unconsciously before the frontal cortex can make conscious decisions. As a consequence, studies that seek immediate short-term responses may only register unconscious activities that occur spontaneously rather than reveal considered mathematical reasoning operating over a longer period of time.

This relates to the major question concerning the use of language in mathematics. Language is essential to name ideas and be able to build up sophisticated theories. But recent studies in neurophysiology question the role of language in mathematical thinking:

> By scanning professional mathematicians, we show that high-level mathematical reasoning rests on a set of brain areas that do not overlap with the classical left-hemisphere regions involved in language processing or verbal semantics. Instead, all domains of mathematics we tested (algebra, analysis, geometry, and topology) recruit a bilateral network, of prefrontal, parietal, and inferior temporal regions, which is also activated when mathematicians or nonmathematicians recognize and manipulate numbers mentally. Our results suggest that high level mathematical thinking makes minimal use of language areas and instead recruits circuits initially involved in space and number. This result may explain why knowledge of number and space, during early childhood, predicts mathematical achievement. (Amalric & Dehaene, 2016.)

This research is based on fMRI scans where mathematicians and non-mathematicians listened to a spoken statement and responded four seconds later to classify it as true, false, or meaningless. In a further study (Amalric & Dehaene, 2019), mathematicians listened to spoken mathematical and non-mathematical statements and were given 2.5 seconds to press a right-hand button for true or a left-hand button for false. This collects very different data from the long-term meaningful learning theory given in the current paper, where conceptual embodiment involves changes of meaning in language as contexts change. At the same time, operational symbolic learning involves sequences of mental operations in time being compressed into a network of related concepts that offers a rich environment for mathematical thinking.

What they have in common is that rote learning of mathematical methods and concepts without meaning may involve spoken statements in language areas which do not connect to sophisticated networks of mathematical relationships. Such networks can be a rich environment for creating new mathematical theories.

I had the privilege of studying for my doctoral degree in mathematics with Michael Atiyah who was awarded his Fields medal while I was one of his students. He always sought to think geometrically, bringing together ideas from very different planes of reference. For instance, my own thesis arose from a suggestion he made about links between geometry, topology and algebra, saying that 'a vector bundle is a topological generalization of a vector space, while a module is an algebraic generalization of a vector space, so they must have something in common.' While this statement will make little sense to readers unfamiliar with these highly technical areas of mathematics, it is indicative of higher thinking processes in the creation of significant new mathematical theories through linking together previously unconnected frames of reference. When the connection is made, it can be accompanied by an enormous Aha! reaction related to a huge sense of achievement and aesthetic insight. But where does it come from?

8. Dreams and mathematical beauty

Mathematical thinking is the construction of human minds, going beyond what can be encountered in the natural world to an inner mental world that can create amazing new insights. Aha! Moments which connect together previously disparate frames of reference give a sense of pleasure and aesthetic beauty which may be likened to experiences in other areas of artistic achievement.

> People think mathematics begins when you write down a theorem followed by a proof. That's not the beginning, that's the end. For me the creative place in mathematics comes before you start to put things down on paper, before you try to write a formula. You picture various things, you turn them over in your mind. You're trying to create, just as a musician is trying to create music, or a poet. There are no rules laid down. You have to do it your own way. But at the end, just as a composer has to put it down on paper, you have to write things down. But the most important stage is understanding. A proof by itself doesn't give you understanding. You can have a long proof and no idea at the end of why it works. But to understand why it works, you have to have a kind of gut reaction to the thing. You've got to feel it.
>
> (Michael Atiyah, quoted in Roberts, 2016.)

The emotional aesthetic reaction to mathematics has aspects in common with other areas of endeavour. For instance, Atiyah likened the most beautiful equation of all to Shakespeare:

> Ah, the most famous of all, Euler's equation:
>
> $e^{i\pi} + 1 = 0.$
>
> It involves π; the mathematical constant *e* [Euler's number, 2.71828 …]; *i*, the imaginary unit; 1; and 0 — it combines all the most important things in mathematics in one formula, and that formula is really quite deep. So everybody agreed that that was the most beautiful equation. I used to say it was the mathematical equivalent of Hamlet's phrase "To be, or not to be" — very short, very succinct, but at the same time very deep. Euler's equation uses only five symbols, but it also encapsulates beautifully deep ideas, and brevity is an important part of beauty.
>
> (Michael Atiyah, ibid.)

He related his creative activities to dreaming:

> [...] Dreams happen during the daytime, they happen at night. You can call them a vision or intuition. But basically they're a state of mind — without words, pictures, formulas or statements. It's "pre" all that. It's pre-Plato. It's a very primordial feeling. And again, if you try to grasp it, it always dies. So when you wake up in the morning, some vague residue lingers, the ghost of an idea. You try to remember what it was and you only get half of it right, and maybe that's the best you can do.
>
> (Michael Atiyah, ibid.)

This happened to me as I wrote this paper, struggling with some of the links I was seeking to make and then having an amazing breakthrough, until I awoke in the morning to realise how many more steps I would need to take before I could hope to write the complicated detail necessary to make the ideas available to a broader readership.

The problem with an Aha! Moment is that it makes a possible breakthrough based on a personally connected cognitive structure but it may not make sense to others who do not share essential parts of that structure. In the case of the equation, $e^{i\pi} + 1 = 0$, to make sense of its beauty requires a knowledge of so many disparate frames of reference: the calculation of π in geometrical terms, the notion of limit, complex numbers, complex powers, and so on. It is only beautiful to those who have a sense of the sophisticated relationships involved.

Nevertheless, for those who do share sophisticated cognitive structures in mathematics, this does gives a shared sense of aesthetic beauty:

> The logical deductive system of the brain, whatever its details, is inherited and is therefore similar in mathematicians belonging otherwise to different races and cultures. It is in this sense that mathematical beauty has its roots in a biologically inherited logical-deductive system that is similar for all brains.
>
> (Zeki, Chén & Romaya, 2018.)

This relates back to an earlier paper by Zeki, Romaya, Benincasa, and Atiyah, (2014) which Atiyah gleefully said:

> That's the most-read article I've ever written! It's been known for a long time that some part of the brain lights up when you listen to nice music, or read nice poetry, or look at nice pictures—and all of those reactions happen in the same place. And the question was: Is the appreciation of mathematical beauty the same, or is it different? And the conclusion was, it is the same. The same bit of the brain that appreciates beauty in music, art and poetry is also involved in the appreciation of mathematical beauty. And that was a big discovery. (Michael Atiyah, quoted in Roberts, 2016.)

9. Reflections

In this paper I have considered the personal and cultural aspects of an Aha! Moment where two or more previously unconnected planes of reference are suddenly brought together and related it to what I call an 'Uh-Huh!' experience where the connection fails to be made. An Aha! may not be correct. It may need reflection to think through its implications and seek a better solution. On the other hand, an Uh-Huh! experience may be very insidious, impeding the development of a better theory, especially when it occurs not just at an individual level but as a shared belief in a wider community.

An Uh-Huh! may arise from a particular belief that works well at a given point in development. Consider, for example, the Euclidean common notion that 'the whole is greater than the part'. This is totally satisfactory in dealing with finite sets or with parts of a geometric figure. It survived for over two thousand years and led to cultural rejection of Galileo's paradox that there is a one-to-one correspondence between the set of whole numbers 1, 2, 3, … and the set of squares 1, 4, 9, …, even though the set of squares is clearly a subset of the whole numbers. When Cantor realised that the idea that a set can be put in one-to-one correspondence with a subset is the very definition of an infinite set, he was met with hostility and cultural rejection that caused him to have a mental breakdown.

I observe, with some regret, that cultural rejection, even when it is done for laudable reasons of maintaining cultural stability, can result in causing mathematical anxiety for a significant portion of the population. For instance, I believe this is happening with many modern approaches to teaching mathematics, in arithmetic, algebra, geometry and calculus based on rote learning.

I offer alternative approaches in this chapter which build on ideas that make natural sense for a wider population if we only make ourselves consciously aware of simple ideas that we can use in practice. The Principle of Articulation shows how we can make better sense of the meaning of operations in arithmetic and algebra, and it can be introduced at any stage to improve understanding and to build more coherent relationships as mathematics becomes more sophisticated. Being aware of its application can highlight the patterns that arise in arithmetic and make the transition to algebra more transparent. This can be enhanced by a specific focus on the flexibility of operational symbols as operations or as mental objects.

The transition to the calculus can be supported by the embodiment of local straightness. In fact, this is just one aspect of embodiment that can make the calculus more meaningful (see, for example, Tall, 2009). Taken with the meaningful interpretation of symbolism, this offers a more compelling reason for taking a locally straight embodied approach to the calculus, leading naturally to an embodied basis for mathematical analysis.

The use of structure theorems to reveal the power of embodied ideas even at the highest level shows how different interpretations of mathematics may be appropriate for different communities dependent on the role that they play in society.

But will it happen? Is it possible to seek the Aha! experience by focusing on making meaningful links or will the Uh-Huh! of cultural resistance – even in the form of cultural stability – hold evolution back?

Only time will tell. It is evident that the varied cultures currently immersed in the teaching, learning and using of mathematics will often be subject to cultural lag compared with the speed of change of technology. On the other hand, technology has reached a level where the widespread availability of smart phones with insightful mathematical apps, such as *GeoGebra* and *Desmos*, to manipulate symbols and graphs has the potential to move things on.

As Buckminster Fuller has said:

> You never change things by fighting the existing reality. To change something, build a new model that makes the existing model obsolete. (Buckminster Fuller, 1981.)

References

Amalric, M. & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians, *Proceedings of the National Academy of Science of the USA*. 016 113 (18) 4909-4917. https://www.pnas.org/content/113/18/4909

Amalric, M. & Dehaene, S. (2019). A distinct cortical network for mathematical knowledge in the human brain. *NeuroImage*, 189, 19-31. https://doi.org/10.1016/j.neuroimage.2019.01.001

Bressoud, D., Mesa, V., Rasmussen, C. (Eds). (2015). *Insights and Recommendations from the MAA National Study of College Calculus*. https://www.maa.org/programs/faculty-and-departments/curriculum-development-resources/national-studies-college-calculus.

Directly downloadable from https://www.maa.org/sites/default/files/pdf/cspcc/InsightsandRecommendations.pdf

Buckminster Fuller, R. (1981). *Critical Path*. New York: St Martin's Press.

College Board (2018). *Calculus AB*. https://apcentral.collegeboard.org/courses/ap-calculus-ab/course.

Czarnocha, Bronislaw. (2014). Bisociation – the theory of the Aha! moment. The basis for human mathematical creativity and for computer creativity of data mining. *Mathematics Teaching-Research Journal Online*, Summer 2014 7, 1.

Desmos (no date). See https://en.wikipedia.org/wiki/Desmos_(graphing) and https://www.desmos.com

Dienes Z. P. (1960). *Building Up Mathematics*. London: Hutchinson.

GeoGebra (no date). See https://en.wikipedia.org/wiki/GeoGebra and https://www.geogebra.org

Gray, E. M. (1993). *Qualitatively Different Approaches to Arithmetic*. Ph.D. Thesis, University of Warwick. http://wrap.warwick.ac.uk/2309/1/WRAP_THESIS_Gray_1993.pdf.

Gray, E. M. & Tall, D. O. (1994). Duality, Ambiguity and Flexibility: A Proceptual View of Simple Arithmetic, *Journal for Research in Mathematics Education*, 26 2, 115-141.

Isoda, M.; Stephens, M.; Miyakawa, T. (2007). *Japanese Lesson Study in Mathematics: Its impact, diversity and potential for educational improvement*. Singapore: World Scientific.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Macmillan.

Koestler, A. (1964). *The Act of Creation*. London: Hutchinson.

Mason, J., Burton, L. & Stacey, K. (1982). *Thinking Mathematically*, London: Addison Wesley.

Roberts, S. (2016). Mathematical Beauty: A Q&A with Fields Medalist Michael Atiyah. *Quanta Magazine*. https://www.scientificamerican.com/article/mathematical-beauty-a-q-a-with-fields-medalist-michael-atiyah/

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 Mathematic*s, 22, 1–36.

Skemp, R. R. (1976). Relational Understanding and Instrumental Understanding. *Mathematics Teaching*, 77, 20–26.

Skemp, R. R. (1979). *Intelligence, Learning, and Action*. Wiley: London.

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

Tall, D. O. (1985). Understanding the calculus, *Mathematics Teaching*, 10, 49– 53.

Tall, D. O. (1986). *Building and testing a cognitive approach to the calculus using interactive computer graphics*. PhD thesis, University of Warwick: http://wrap.warwick.ac.uk/2409/

Tall, D. O. (2001). A Child Thinking about Infinity. *Journal of Mathematical Behavior*, 20, 7–19.

Tall, D. O. (2004). Introducing Three Worlds of Mathematics. *For the Learning of Mathematics*, 23 (3). 29–33.

Tall, D. O. (2009). Dynamic mathematics and the blending of knowledge structures in the calculus. *ZDM – The International Journal on Mathematics Education*, 41 (4) 481–492.

Tall, D. O. (2013). *How Humans Learn to Think Mathematically*. New York: Cambridge University Press. ISBN: 9781139565202.

Tall, D. O. (2015). Setting Lesson Study within a Long-Term Framework for Learning. In Maitree Inprasitha, Masami Isoda, Patsy Wang Iverson, Yeap Ban Har (Eds): *Lesson Study: Challenges in Mathematics Education*. pp. 27–50. World Scientific, Singapore.

Tall, D. O. (2017). Making sense of elementary arithmetic and algebra for long-term success. Draft chapter for Japanese Elementary School Teachers. http://homepages.warwick.ac.uk/staff/David.Tall/pdfs/dot2017x-long-term-sense-making.pdf

Tall, D. O. (2019a). Complementing supportive and problematic aspects of mathematics to resolve transgressions in long-term sense making. *Fourth Interdisciplinary Scientific Conference on Mathematical Transgressions*, Krakow, March 2019. Opening Plenary Presentation. http://homepages.warwick.ac.uk/staff/David.Tall/pdfs/dot2019a-transgressions-krakow.pdf

Tall, D. O. (2019b). From Biological Brain to Mathematical Mind: The Long-term Evolution of Mathematical Thinking. In Danesi, M. (Ed.): *Interdisciplinary Perspectives on Math Cognition*. Springer. https://www.springer.com/gp/book/9783030225360

Tall, D. O. (2019c) Long-term principles for meaningful teaching and learning of mathematics. In Sepideh Stewart (ed.): Mathematicians' Reflections on Teaching: *A Symbiosis with Mathematics Education Theories*. (Submitted May 22 2019). http://homepages.warwick.ac.uk/staff/David.Tall/pdfs/dot2019c-long-term-framework.pdf

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

Tall, D. O., Tall, N. D. & Tall, S. J. (2017). Problem posing in the long-term conceptual development of a gifted child. In Martin Stein (Ed.): *A Life's Time for Mathematics Education and Problem Solving. On the Occasion of Andràs Ambrus' 75th Birthday*, pp. 445-457. WTM-Verlag, Münster.

Van Hiele, P. M. (1986). *Structure and Insight*. Orlando: Academic Press.

Verhoef, N. C., Tall, D. O., Coenders, F. G. M., van Smaalen, D. (2014). The complexities of a lesson study in a Dutch situation: Mathematics Teacher Learning. *International Journal of Science and Mathematics*

Education 12: 859–881. https://doi.org/10.1007/s10763-013-9436-6

Verhoef, N. C., Coenders, F. G. M., Pieters, J. M., van Smaalen, D., & Tall, D. O. (2015). Professional development through lesson study: teaching the derivative using GeoGebra. *Professional development in education*, 41(1), 109-126. https://doi.org/10.1080/19415257.2014.886285

Vinner, S. (1983). Concept definition, concept image and the notion of function, *International Journal of Mathematical Education in Science and Technology*, 14:3, 293-305, DOI: 10.1080/0020739830140305

Vinner, S. & Hershkowitz, R. (1980). Concept Images and some common cognitive paths in the development of some simple geometric concepts, *Proceedings of the Fourth International Conference of PME*, Berkeley, 177–184.

Wilder, R. L. (1968). *Evolution of Mathematical Concepts*. New York: Wiley.

Wilder, R. L. (1981). *Mathematics as a Cultural System*. Oxford: Pergamon.

Zeki, S.; Chén, O.; Romaya, J. (2018). The Biological Basis of Mathematical Beauty, *Frontiers in Human Neuroscience* 12, 467. https://www.frontiersin.org/articles/10.3389/fnhum.2018.00467/full

Zeki, S.; Romaya, J.; Benincasa, D.; Atiyah, M. (2014). The experience of mathematical beauty and its neural correlates. *Frontiers in Human Neuroscience*. https://www.frontiersin.org/article/10.3389/fnhum.2014.00068