The De Morgan Journal 1 no. 1 (2011), 19–27 www.de-morgan.org

PERCEPTIONS, OPERATIONS AND PROOF IN UNDERGRADUATE MATHEMATICS

DAVID TALL

Teaching and learning undergraduate mathematics involves the introduction of ways of thinking that at the same time are intended to be more precise and logical, yet which operate in ways that are unlike students' previous experience.

When we think of a vector, in school it is a quantity with magnitude and direction that may be visualized as an arrow, or a symbol with coordinates that can be acted upon by matrices. In university mathematics it is an element in an axiomatic vector space.

As I reflected on this situation I realised that these three entirely different ways of thinking apply in general throughout the whole of mathematics [**7**, **8**]. The two ways encountered in school depend on the one hand on our physical perception and action and dynamic thought experiments as we think about relationships, on the other they depend on operations that we learn to perform such as counting and sharing which in turn are symbolised as mathematical concepts such as number and fraction.

At university, all this is turned on its head and reformulated in terms of axiomatic systems and formal deduction. Our previous experiences are now to be refined and properties are only valid if they can be proved from the axioms and definitions using mathematical proof. The formal approach gives a huge bonus. No longer do proofs depend on a particular situation: they will hold good in any future situation we may meet provided only that the new context satisfies the specific axioms and definitions. However, the new experience is also accompanied by mental confusion as links, previously connected in perception and action, now require reorganisation as formal deductions, and subtle implicit links from experience may be at variance with the new formal setting.

Further analysis of the development of mathematical thinking reveals three quite different forms of thinking and development that I term *conceptual embodiment*, *operational symbolism* and *axiomatic formalism*. These operate in such different ways—not only at a given point in time, but also

---

2000 *Mathematics Subject Classification* 97C50 (primary), 97D20 (secondary).

University of Warwick. Email: *david.tall@warwick.ac.uk*

in their long term development—that I called them three mental *worlds* of mathematics.

Conceptual embodiment and operational symbolism develop in complementary ways in school mathematics in which physical operations relate to algebraic symbolism (Thomas [**10**]). The world of conceptual embodiment is based on our operation as biological creatures, with gestures that convey meaning, perceptions of objects that recognise properties and patterns, thought experiments that imagine possibilities, and verbal descriptions and definitions that formulate relationships and deductions as found in Euclidean geometry and other forms of figures and diagrams. The world of operational symbolism involves practising sequences of actions until we can perform them accurately with little conscious effort. It develops beyond the learning of procedures to carry out a given process (such as counting) to the concept created by that process (such as number). Gray and Tall [**3**] formulated this flexibility by speaking of such symbols as 'procepts' that act dually as *process* and *concept*. The operational world of symbolism develops in a spectrum of ways from limited procedural learning to flexible proceptual thinking.

The third world of axiomatic formalism builds from lists of axioms expressed formally through sequences of theorems proved deductively with the intention of building a coherent formal knowledge structure. Its major criterion is that relationships must in principle be deducible by formal proof. However, students and mathematicians interpret formalism in a variety of ways, depending on the links with embodiment and symbolism. Some build *naturally* on their previous experience to *give meaning* to definitions. For instance, the idea of a sequence (s_n) tending to a limit may be seen by plotting the successive points (n, s_n) and seeing that, the sequence tends to a limit L if, given a required error $\epsilon > 0$, then from some value N onwards, (for $n \geqslant N$) the terms s_n lie between two horizontal lines $L \pm \epsilon$. Others build *formally* by *extracting meaning* from the definition by learning to reproduce it and practising formal proofs until it becomes a familiar mode of operation. Both approaches are possible and can lead to successful formal thinking, although both can fail, either because the new formal ideas conflict with beliefs built from earlier experience or because the multi-quantified definitions are just too difficult to handle (Pinto and Tall [**4**, **5**]).

The question arises as to how this framework of three worlds of mathematics can help us as mathematicians to encourage our students to think in successful mathematical ways. The framework is general. Although embodiment starts earlier than operational symbolism, and formalism occurs much later still, when all three possibilities are available at university level, the framework says nothing about the sequence in which teaching should occur. Indeed, in the learning of mathematical analysis some students clearly follow a natural approach based on their thought experiments and concept imagery while others are more comfortable working in a purely formal context. Not only is it possible to use embodied examples to give meaning to a formal theory, it is also possible to use a formal theory to highlight the essential properties in an embodied example.

The framework can be better understood by reflecting on specific cases. Consider, for example the notion of continuity. Embodiment clearly gives powerful insights that can be used to motivate symbolic relationships and formal definitions. For instance, the dynamic idea of natural continuity arises from the physical drawing of a graph with a 'continuous' stroke of the pencil remaining on the paper and leaving a continuous trace. While this is often considered to be an 'intuitive' notion of continuity that lacks a formal definition, it is also possible to envisage the graph as a stroke of a pencil which covers the theoretical graph with a stripe of height $\pm\epsilon$. If a small portion of the graph is stretched horizontally, while maintaining the vertical height, the graph will 'pull flat' in the sense that, for some $\delta > 0$, then for any x between $x_0 - \delta$ and $x_0 + \delta$, the value of $f(x)$ will lie between $f(x_0) - \epsilon$ and $f(x_0) + \epsilon$ (Tall [**9**]). In this way it is possible to have a natural transition from embodied continuity to the formal definition in mathematical analysis, which may help a natural learner but may be unnecessary for a formal learner.

Elementary calculus is highly amenable to a natural approach that links together visual insight and symbolic manipulation without introducing formal epsilon-delta definitions. Using computer technology to magnify graphs reveals the property that many continuous graphs visibly approximate to a straight line under high magnification. Such a graph is said to be 'locally straight'. The slope of a locally straight graph can be *seen* by highly magnifying a portion of a graph to visualize it as essentially straight and to measure its slope. This gives a natural distinction between continuity of a graph drawn with a pencil or with pixels on a graphic display (which will 'pull flat') and differentiability (which involve graphs that are 'locally straight'). It enables students to visualize non-differentiability (with 'corners' having different left and right derivatives, or even functions that are so wrinkled that they do not look straight no matter how much they are magnified) and to realise that most continuous functions are *not* differentiable (Tall, 2009). Such an approach, although based on visual and symbolic techniques only, gives far greater insight into the *meaning* of the notions of continuity and differentiability.

Furthermore, for a locally straight function, the Leibnitz notation dy/dx

may be interpreted as a quotient of the components of the tangent vector, as originally conceived by Leibnitz himself. In such an interpretation, dx and dy can be called differentials, representing the components of the tangent vector up to a scalar multiple. Now a first-order differential equation is just that: it formulates the direction of the tangent in which the differentials are the components dx and dy.

Software can be programmed to build up the numerical slope of a graph dynamically by shifting along and computing

$$\frac{f(x+h)-f(x)}{h}$$

for variable x and fixed h. This can be drawn as a practical slope function that stabilizes on a visible graph on screen for small values of h, revealing the stabilized graph as the derivative. The embodied action of looking along a graph, imagining its changing slope operates on a visual *object*, (the graph of f) and gives a new *object* (the stabilized graph Df). For instance, if $f(x) = \sin x$, then looking at the changing slope along the graph gives $Df(x) = \cos x$. The symbol D is here an embodied operator that means 'look along the graph and see its slope function Df'.

Focusing on a specific point x gives the equation

$$Df(x) = \frac{dy}{dx}$$

where $Df(x)$ is the value of the function produced by the operation D calculated at x and dx and dy are differentials (components of the tangent). This leads to the natural idea of blending of the two meanings by writing

$$\frac{dy}{dx} = \frac{d(f(x))}{dx} = \frac{d}{dx}f(x)$$

and allowing the symbol d/dx to be interchanged with the operation D.

This approach is a quite different from that suggested by the APOS theory of Dubinsky (e.g. Asiala et al., 1996), which speaks of focusing on a *process*, here the limit process

$$\lim_{h\to 0}(f(x+h)-f(x))/h,$$

and encapsulating it as an *object*. Fundamentally, operating on an *object* to construct a visible object is far more elementary than encapsulating a *process* to give an as yet unknown *object*. Research results speak for themselves: the visual approach is highly successful (Tall, 1986) whereas the APOS view, programming functions symbolically to compute a practical derivative that is to be encapsulated as a symbolic object proves to be far more elusive (Cottrill et al. [**2**]).

There is a clear distinction between a natural approach to elementary calculus and a formal approach to mathematical analysis. Elementary calculus blends together experiences in embodiment and symbolism without entering the complicated formal world of mathematical analysis that is characterised by the multi-quantified epsilon-delta definition of limit.

Notice that I am not saying that one approach should be privileged over another. It is not a question of whether one should teach the formal definition of limit or not, it is a question of the objective of the particular course and its appropriateness for the current development of the learner.

If the objective is to give insight into the calculus as an operational system in applications in which the Leibnitz notation plays its part, then a locally straight approach gives both human meaning and operational symbolism. If the objective is to develop logical mathematical analysis (preferably as a course that follows elementary calculus), then the handling of multi-quantified definitions is part of the toolkit required for rigorous mathematical thinking. The most important aspect is to decide upon the aims of the course and not to inflict formal subtleties on students who are better served by a meaningful blend of embodiment and symbolism.

The three worlds of mathematics each offer their own distinct advantages:

– embodiment gives a basis of human meaning that can be translated into flexible symbolism,
– symbolism offers a powerful tool for suitably accurate computation and precise symbolic solutions,
– formalism offers precise logical deduction that will operate in any context where the axioms and definitions are satisfied.

Consider, for example, the manipulation of multi-quantified statements. Embodiment will allow thought experiments to think about how to negate such a statement, to allow one to realise that to prove that a universal statement is not true, one only needs a single counter-example and that to prove an existence statement is not true requires a universal statement of its falsehood. Symbolism translates these statements into $\neg\forall \equiv \exists\neg$ and $\neg\exists \equiv \forall\neg$. In this way the definition of continuity of a function f at a point x on a domain D can be written as

$$\forall\epsilon > 0\,\exists\delta > 0\,\forall y \in D\,(|x-y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon)$$

and its negation can be found by placing the negation symbol in front and passing it successively over each quantifier, swapping one to the other to get

$$\exists\epsilon > 0\,\forall\delta > 0\,\exists y \in D\,(|x-y| < \delta \text{ and } \Rightarrow |f(x) - f(y)| \geqslant \epsilon).$$

This symbolic manipulation is easier to handle than thinking through the full embodiment of the meaning all at once. It enables a more compressed form of thinking that is supportive in building formal proofs.

What is essential in learning, is to build on the previous experience of the students to enable them to make personal sense of the new constructs. In the case of vectors, a vector also has three different meanings: as a geometric quantity with magnitude and direction, as an algebraic entity written as a column vector and as an element in a formal vector space. As a geometric quantity, it can be represented as a physical action, say as a translation of an object such as a triangle on the surface of a flat table. A given point A on the object will be shifted by a translation to a point B and represented as the shift $\vec{AB}$ from a starting point to a finishing point in which any two such arrows will all have the same magnitude and direction. The translation can therefore be represented by a *single* arrow of given magnitude and direction that can be placed anywhere to represent the start and end of the shift of a particular point. This gives an embodied arrow of given magnitude and direction that represents the translation. Again we start with an *object* on the table and a process of translating it to represent the translation as an embodied object, the *free vector*. Representing the composite of two translations, one after another, the result is represented by the unique free vector that has the same effect. This conception of a free vector then has a meaning that translates naturally to the triangle law or the parallelogram law.

A scalar multiple of a translation can be imagined as retaining the direction but multiplying the magnitude by the scalar (or reversing the direction if the scalar is negative). This applies to free vectors by multiplying the magnitude of the vector by the scalar in the same way.

The symbolic representation of a vector arises naturally through the solution of a system of linear equations in n variables. For $n = 1, 2, 3$, such equations can be represented in 1, 2, or 3 dimensional space. The symbolic techniques naturally extend to n variables and, even if the ideas are no longer easily visualised in higher dimensions, they can be represented by coordinate vectors with n components with transformations represented by matrices.

The formal representation of a vector is quite different. A vector space is specified as an additive abelian group V with the action of a field of scalars F which satisfies appropriate axioms. Such vectors now no longer have magnitude or direction, but by introducing the notion of linear independence and spanning set, a structure theorem may be proved to show that any finite dimensional vector space over F is isomorphic to a space F^n represented as n-dimensional coordinates. In the case of $n = 2$ or 3 and $F = \mathbb{R}$ gives an embodiment of the vector space almost like $\mathbb{R}^2$ or $\mathbb{R}^3$. I say ‘almost’ because the vectors in the vector space do not yet have a conception of magnitude or direction. To do this, one needs to add an inner product to enable one to specify lengths and angles.

The problem for the teacher and the student is to be aware what assumptions are being made. Are vector spaces being studied *formally* based only on deductions from axioms or *naturally*, based on experiences of perceptions and actions in two and three-dimensional real space? The choice is up to the teacher, but it needs to be explicit.

A natural approach would involve beginning from conceptions that are familiar: solving linear equations in one, two and three variables and *generalising* them to n variables, which involves essentially the same symbolic solution technique although no longer visualisable in higher dimensional space. A formal approach would begin by *abstracting* the axioms for a vector space and writing down the list of axioms, and eventually proving a structure theorem from the axioms that vectors in a finite dimensional vector space can be represented by coordinate vectors with n components. Of course, if natural learners are presented with a formal approach, then the initial theorems and proofs may make little sense and the course may only come alive for them when the structure theorem for finite dimensional vector spaces has been proved and they are asked to solve linear equations operationally using symbolic vectors.

The same can be said for other topic areas, for instance, groups studied as embodied operations of actions on figures with symmetry, or symbolic operations as permutations of n elements prior to a formal axiomatic approach.

Formally, the various lecture courses, be they in analysis, vector space theory, group theory, or whatever, often begin with a formal axiomatic structure and formal deductions. Part of the way through the course a structure theorem is proven to give the axiomatic system a structure that can be embodied in a manner now based deductively on the axioms with an operational symbolism that can be used solve problems symbolically.

For instance, in analysis, the axioms for a complete ordered field identify it uniquely up to isomorphism, allowing it to be visualised as a real line and symbolised as infinite decimals. In vector space theory, a finite dimensional vector space over F is isomorphic to F^n, allowing it to be symbolised as ntuples and embodied in $\mathbb{R}^2$ or $\mathbb{R}^3$. In group theory, a finite group is isomorphic to a subgroup of permutations.

The roles played by embodiment, symbolism and formalism are very different and the teacher has to make it explicitly clear what approach is being taken. Is the course to be a formal course that requires formal deduction from axioms? This may be built entirely formally until structure theorems give it forms of embodiment and symbolism based on those axioms. Is it a formal course to be constructed naturally to enable students to give meaning to formal definitions through a range of examples? Or is the course intended to develop the necessary symbolic algorithms to enable the ideas to be used in specific applications, with examples relevant to the area of application?

My own view is that it would help students enormously to gain an insight into the strategy, which many lecturers use implicitly but is rarely made explicit. That is that formal mathematics clarifies issues by specifying explicit axioms that are the 'rules of the game' and formal proofs deduced using these rules are proven once and for all in any situation where the rules are satisfied. The initial deductions from the rules are often quite technical and form a barrier for many students. But once a structure theorem has been proved, the techniques developed are now proven to work in all situations, whether known now or to be encountered in the future. This formal foundation is a gift worth having and it can be acquired by the formal thinker who deduces only from the axioms using formal proof, or by the natural thinker who sees the generalities bringing together many experiences that give meanings to the formalities. An understanding of three different approaches to mathematics would be invaluable, made explicit both to teachers and to students to be aware of the different objectives of mathematical thinking, consisting of:

- ideas based on human perceptions and actions with thought experiments to suggest what might be true,
- operations based on actions that give subtle mathematical processes to express and solve problems symbolically,

and

- formal axioms, definitions and proof that give a coherent framework of mathematics, supporting perception and operation with an underlying formal structure that applies in any situation where the axioms and definitions hold.

References

1. M. Asiala, J. Cottrill, E. Dubinsky, and K. E. Schwingendorf. The development of students' graphical understanding of the derivative. Journal of Mathematical Behavior, 16 (1997), 399–431.

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

3. E. M. Gray and D. O. Tall. Duality, Ambiguity and Flexibility: A Proceptual View of Simple Arithmetic. The Journal for Research in Mathematics Education, 26 no. 2 (1994), 115-141.

4. M. M. F. Pinto. Students' Understanding of Real Analysis. PhD Thesis, Warwick University, 1988.

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

6. D. O. Tall. Using computer graphics as generic organisers for the concept image of differentiation. Proceedings of the Ninth International Conference of the International Group for the Psychology of Mathematics Education, Holland, 1 (1985), 105-110.

7. D. O. Tall. The three worlds of mathematics. For the Learning of Mathematics, 23 no. 3 (2004), 29-33.
8. D. O. Tall. The Transition to Formal Thinking in Mathematics. Mathematics Education Research Journal, 20 no. 2 (2008), 5–24.
9. D. O. Tall. Dynamic mathematics and the blending of knowledge structures in the calculus. ZDM—The International Journal on Mathematics Education, 41 no. 4 (2009), 481-492.
10. M. O. J. Thomas. A Conceptual Approach to the Early Learning of Algebra Using a Computer, Ph.D. thesis, University of Warwick, 1988.