DYNAMICAL ASPECTS OF MATHEMATICAL PROOF

Tony Barnard

Kings College London

Abstract: *In regarding mathematical thinking as proceeding via operations involving a small number of 'items' at any one time, an important feature is the phenomenon in which a section of mathematical structure may be mentally held as a single unit, possessing an interiority that can be subsequently expanded without loss of detail and trigger connections with other parts of cognitive structure. This article discusses the role of this phenomenon of 'compression' and 'expansion' in the manipulation of statements involved in certain types of mathematical proof. Implications for teaching linked to an awareness of these aspects is also discussed.*

Cognitive units

Much of mathematics has the form of a hierarchy or network of concepts spanning several layers of abstraction and with various interconnections. Even something as elementary as the equality '$\sin 60^\circ = \sqrt{3}/2$' can relate to things such as 'the angles in an equilateral triangle are all equal', 'the angles in a triangle add up to 180°', 'an angle in an equilateral triangle is 60°', 'the line joining the vertex to the midpoint of the base (of an isosceles triangle) meets it at right angles', 'if the side is two units, half a side is 1 unit', 'Pythagoras' Theorem', '$a^2 + b^2 = c^2$', '$b^2 = c^2 - a^2$', 'the square of $\sqrt{3}$ is 3', '$1^2 + (\sqrt{3})^2 = 2^2$', '$\sin^2 + \cos^2 = 1$', 'the sine of an angle is opposite over hypotenuse', etc.

However, the amount that can be held in the mind and attended to at any one time is very limited and, in order to minimise the consequent constraints on thinking, various strategies are adopted for reducing the mental load of data to be considered. Underlying many of these strategies is a phenomenon in which a collection of linked items is mentally held as a single unit, as if it had been compressed in order to fit into the short-term focus of attention. Examples of such collections of items, linked in a section of mathematical structure, are the steps of an algorithmic process, the components of a complex statement, different representations of a number, the steps of a logical deduction, alternative formulations of an algebraic relationship, a set of equivalent statements. There are also instances where mathematical items are linked in ways that are not necessarily part of a recognisable mathematical structure. These could involve, for example, telephone numbers, memorable dates or even mannerisms of a lecturer! The point is that the single entity can be easily manipulated using a minimum of thinking space and subsequently unpacked whenever needed. Such a piece of cognitive structure that can be held in the short-term focus of attention all at one time has been called a ***cognitive unit*** (Barnard & Tall, 1997, Barnard, 1999). Not only does this operate as a form of shorthand for a collection of items, it also carries with it, just beneath the surface, the structure of the collection, and is operative in the sense that the live connections within the structure are able to guide its manipulation. For example, when various relationships associated with the equality '$\sin 60^\circ = \sqrt{3}/2$' are conceived as a cognitive unit, they are not just replaced by a passive mental label. As well as leaving space in the mind for bringing in other ideas, the cognitive unit actively influences this thinking by pushing to the fore elements of its internal structure such as 'if two sides of a triangle are equal then the base angles are equal', 'Pythagoras Theorem', '$1^2 + (\sqrt{3})^2 = 2^2$'.

Furthermore, in much the same way as something that initially seems abstract may, with greater familiarity, become to seem concrete, a collection of linked cognitive units may itself become a (new) cognitive unit. In this way an individual is able to operate with a network of nested mental structures spanning several layers of thought.

These ideas are closely related to Skemp's *varifocal theory* of cognitive concepts (Skemp, 1979), where a concept may be conceived either as a global whole, or viewed under closer scrutiny to reveal various levels of detail. The flexibility of thought accompanying operation with rich cognitive units is related to the *curtailed structures* of Krutetskii (1976), which are evident in the way capable students compress their solutions to problems in a succinct and insightful manner.

> … mathematical abilities are abilities to use mathematical material to form generalized, curtailed, flexible and reversible associations and systems of them. These abilities are expressed in varying degrees in capable, average and incapable pupils. In some conditions these associations are performed "on the spot" by capable pupils, with a minimal number of exercises. In incapable pupils, however, they are formed with extreme difficulty. For average pupils, a necessary condition for the gradual formation of these associations is a system of specially organized exercises and training.

(Krutetskii, 1976, p. 352.)

Another related idea is that of process-object duality. Davis discussed the idea of a procedure achieving *noun status:*

> When a procedure is first being learned, one experiences it almost one step at a time; the overall patterns and continuity and

> flow of the entire activity are not perceived. But as the procedure is practised, the procedure itself becomes an entity - it becomes a *thing*. It, itself, is an input or object of scrutiny. All of the full range of perception, analysis, pattern recognition and other information processing capabilities that can be used on any input data can be brought to bear on this particular procedure. Its similarities to some other procedure can be noted, and also its key points of difference. The procedure, formerly only a thing to be done - a verb - has now become an object of scrutiny and analysis; it is now, in this sense, a noun.
> (Davis, 1984, p.36)

Other 'process-object' theories include the ideas of the *encapsulation* of process into object (Dubinsky, 1991) and the *reification* of process into object (Sfard, 1991). Unlike Skemp's theory which sees the schematic structure consisting of object-like concepts which are linked by properties and processes, these encapsulation theories describe how sequences of activities can become routinized into thinkable processes which are then in turn conceived as mental objects. This is described as follows:

> According to APOS theory, an action is a transformation of mathematical objects that is performed by an individual according to some explicit algorithm and hence is seen by the subject as externally driven. When the individual reflects on the action and constructs an internal operation that performs the same transformation then we say that the action has been interiorized to a process. When it becomes necessary to perform actions on a process, the subject must encapsulate it to become a total entity, or an object. In many mathematical operations, it is

> necessary to de-encapsulate an object and work with the process from which it came. A schema is a coherent collection of processes, objects and previously constructed schemas, that is invoked to deal with a mathematical problem situation.
>
> (Asiala et al, 1997, p. 400.)

A different view of the relationship between process and object is that in which the role of the symbol is seen as being pivotal in the thinking process. Gray & Tall (1984) formulated the notion of *procept* as a combination of process and concept evoked by a single symbol. They saw a symbol such as '3+4' as acting as a pivot between a process (of addition) and the concept (of sum). This power - which is characteristic of symbolism in arithmetic, algebra and calculus - allows the thinker to switch between using the symbol as a concept to think about or as a process to calculate or manipulate to solve a problem.

Process-object theories provide a good model regarding computations and procedures, such as adding together two three-digit numbers or solving a linear equation, in which the objects being manipulated are numbers or symbols. However, in addition to sequential procedures, in which each action cues the next, mathematical thinking can also involve steps in which a variety of cues and pieces of knowledge need to be synthesised. Here several items may be contemplated at the same time, reorganised and a conclusion drawn from the collection as a whole. For example, the deduction that an integer is even if its square is even, requires a synthesis of statements such as "an integer is either even or odd" and "the square of an odd integer is odd" . Establishing relationships in elementary Euclidean geometry and proving theorems in undergraduate mathematics consist almost entirely of manipulating statements in this kind of way.

Now the important point about manipulating these things, be they numbers, symbols or statements, is that they *do* have to be 'things'. That is to say, they have to be mentally held as items 'small enough' to fit into the short term focus of attention and leave 'space around them' in which to operate. In the same way that a process might be conceived as a single entity, a 'thing' could also be a chain of logical deduction, a complex statement, a collection of alternative formulations of an algebraic relationship, a set of equivalent statements, etc. Thus the notion of 'cognitive unit' gives a model of aspects of mathematical thinking appropriate for a wider range of mathematical activity.

Constructing mathematical proofs

We shall take the term, 'mathematical proof', to mean a hierarchy of links between givens and a concluding statement, where a 'given' is something that is assumed (explicitly or implicitly) without relating it to anything more primitive. A basic ingredient in the building of such links is the manipulation of mathematical statements. For example, in proving the truth of a statement by contradiction, we assume that the statement is false, consider consequences of this assumption and identify a contradiction among these consequences. Let us consider in more detail what is involved in this process.

In order to progress from the assumption that the initial statement is false, we need to construct a statement that is implied by the falsity of the initial statement. It is usually safer to begin by constructing the negation of the initial statement (something which is actually equivalent to the falsity of the initial statement) since this does not lose any information, but either way we are involved in constructing a statement that is logically related to the initial statement. The next step is concerned with taking this newly constructed statement and generating statements which are consequences of it. However this step is more subtle than simply listing all the consequences we can find. We need to select the consequential statements with an eye on the goal of arriving at a contradiction, and this involves having a feel for a range of cognitive units related to each statement we consider. Next we have to pick out from all of this two statements which are in contradiction to each other. Finally we have the more routine, but no less complex, step of deducing that the assumption of falsity must itself be false (because it leads to a contradiction) and then interpreting this to mean that the initial statement is true. At all stages in this process we are operating on statements externally in much the same way as we might operate on variables or unknowns. We are modifying statements, recognising logical relationships between statements and thinking about statements to see if they satisfy various given requirements.

Thus manipulating statements in mathematical proof involves a range of operations from constructing negations and converses to recognising implications and contradictions. In order to do this successfully not only do the components of a statement have to be meaningful, but the statement itself has to be a cognitive unit. It has to be a ‘thing’ which can be operated on from the outside, and with attributes such as having the potential to be recognised as a missing piece and to be contemplated in thoughts about whether or not to use it. A major cause of students’ difficulties in constructing mathematical proofs (and to a lesser extent in following them) is not that they necessarily do not understand the statements involved, but that they are unable to get outside of these statements and work with them in the above manner.

One area which permeates mathematical discourse at university, and in which the above aspect comes into play, is that of negating statements involving quantifiers. Consider a statement *S* of the form, "There exists $n \in \mathbf{N}$ such that $A(n)$ is true", where $A(n)$ is some statement involving a positive integer *n*. $A(n)$ could be, for example, a simple statement like "$1+61n^2$ is a perfect square", or a compound statement like "$|a_m|<1$ for all integers $m \geq n$". In order to construct the negation of *S*, a student has to have meaning for its component statements "There exists $n \in \mathbf{N}$" and "$A(n)$", and also be able to conceive them as manipulable entities. The first stage of the manipulation lies in the transformation to the statement, "For all $n \in \mathbf{N}$, $A(n)$ is not true". The next stage lies in unpacking the new component "$A(n)$ is not true". In the simple example above this just becomes, "$1+61n^2$ is not a perfect square". However in the compound example where $A(n)$ is "$|a_m|<1$ for all integers $m \geq n$", the unpacking involves a further manipulation of cognitive units in transforming this to "There exists an integer $m \geq n$ such that $|a_m| \geq 1$". Finally the student has to fit all the pieces together, perhaps making further slight modifications in order to satisfy constraints of language or notation. Thus the negation of the statement, "There exists $n \in \mathbf{N}$ such that $|a_m|<1$ for all integers $m \geq n$", could read, "Given $n \in \mathbf{N}$, there exists an integer $m \geq n$ such that $|a_m| \geq 1$". At all stages of this process, the primary objects of thought are the *statements*, "There exists $n \in \mathbf{N}$", "$|a_m|<1$", "*m* is an integer greater than or equal to *n*", and not the numbers or symbols $1, n, m, a_m$. If a student is wholly preoccupied with the micro objects inside the statements, thinking things like, "I have an element a_m and the modulus of this element is less than 1", then it will not be possible to construct the negation.

Another area in which these limitations are often a major source of difficulty is in proofs using the Principle of Induction. Let us consider, for example, what is involved in proving by induction the statement, "$3^n > 2n-1$ for all positive integers *n*". This statement is implied by the following two statements.

A. The statement, "$3^n > 2n-1$", is true when $n = 1$.

B. For $k \geq 1$, the truth of the statement "$3^n > 2n-1$" for $n = k$, implies the truth of this statement for $n = k+1$.

In order to comprehend and operate with the assertion that *A* and *B* together imply the desired conclusion, it is necessary to temporarily ignore the details inside these two statements and to think of each of *A* and *B* as a single entity. Then when proceeding to prove the statements *A* and *B*, it is necessary to unpack their contents and work with the interior items. Statement *A*, a substitution and numerical verification, does not present much difficulty. Statement *B*, however, requires a dual literal substitution of n by k and by $k+1$, and subsequent work with the statements

B_1. $3^k > 2k-1$,

B_2. $3^{k+1} > 2(k+1)-1$.

If a student is only able to focus at the micro level of the items $3, k, >, +, 2, -1$ inside these statements, it will be difficult to handle the idea of n being both k and $k+1$ and to think about B_1 and B_2 at the same time. At this level of focus it would have to be one or the other. Yet in order to even contemplate the possibility of B_1 implying B_2, it is necessary to temporarily think of them simultaneously as single units within the compound statement, "B_1 implies B_2".

The next thing is to look inside each of B_1 and B_2 and try to think of a way of proving that B_1 implies B_2. This might involve noting that, as $3^{k+1} = 3.3^k$, we have that B_1 implies $3^{k+1} > 3(2k-1) = 6k-3$, and so it would be enough to show that $6k-3 \geq 2k+1$. Identifying this inequality as an intermediate objective and noting that it is true as a consequence of the fact that $k \geq 1$, involves a technical process of contemplating a variety of statements and expressions (modifications of B_1, B_2 or expressions inside them), following up consequences and matching with the desired objective. In order to do this successfully, and not lose sight of what is being assumed and what is being deduced, students have to compress and expand cognitive units, moving flexibly between compound statements, simple statements, expressions and symbols as their primary objects of thought. Without such flexible shifts of focus, it will not be possible to construct a proof in anything other than a shallow instrumental manner.

Implications for teaching

Students often say that they can follow proofs when the lecturer goes through them in class, but they are unable to construct proofs for themselves when required to do so for homework. One explanation of this phenomenon has to do with the shifting of focus through the different layers of detail in the 'things' to be manipulated: statements, statements within statements, expressions within statements, symbols within expressions, *etc.* When a proof is covered in a lecture, the lecturer's guidance along the path between givens and concluding statement, implicitly includes specification of the level of items that are to be the primary objects of thought at any stage. For example, in the induction proof of the last section, the lead as to when statement B_1 is to be thought of as a compressed item within the statement, "B_1 implies B_2", or when it is to be unpacked for a finer grained manipulation, is provided by the lecturer. It is this focus shift of compression and expansion that often lies at the heart of the difficulty when students try to construct proofs for themselves. It is a bit like knowing when and how to change gear while driving. When students ask the seemingly bizarre question, "How do you do proofs?", they may simply be reacting to a predicament similar to that of trying to drive without awareness of the existence of gears.

A lot has been written about problem-solving and many useful general strategies and heuristics have been suggested. It can be argued, however, that the most powerful problem-solving strategies in mathematics are those which are domain specific. For example in showing that *a* divides *b*, where *a* and *b* are two integers specified only by certain properties (e.g. *a* might be the greatest common divisor of positive integers *c* and *d*, and *b* any common divisor of *c* and *d*), a good strategy is to write down an equation of the form $b = aq + r$ and then try to use the properties to show that $r = 0$. Whereas use of the division algorithm in an abstract setting may initially be seen by a student as a disparate collection of steps of algebra, when used again in a range of different settings, this may become a 'thing' in the mind of the student, a cognitive unit that would then be available for the student's own use. A broader example is the local-global principle in commutative algebra or, in elementary number theory, the theme that a problem involving positive integers can often be reduced to one about primes because every positive integer is a product of primes.

We would also argue that a reason why domain specific strategies are so powerful has to do with the domain dependency of the compressibility of mathematical structure into cognitive units. Although a cognitive unit can arise from a collection of items linked in almost any way, when the links are based on mathematical structure the associations carried with the cognitive unit are likely to be relevant to the mathematical problem at hand. For example consider the following properties of a subgroup *H* of a group *G*. For elements $x, y, z \in G$, (i) $x \in H$ and $y \in H \Rightarrow xy \in H$, and (ii) $x \in H$ and $z \in H \Rightarrow xz^{-1} \in H$. The first of these properties may become associated with a feeling that 'if you are in *H*, multiplying by an element of *H* keeps you in *H*', and the second with a feeling that 'if you are in *H*, cancelling by an element of *H* keeps you in *H*'. Explicitly suggesting to students these informal interpretations of the properties may help them to hold them in their minds as single mental entities. Such a saving of mental space may be important when students come to use one of these properties as part of a mathematical argument. Consider for example the following step occurring in a proof of the fact that every subgroup of a cyclic group is cyclic. *G* is a cyclic group with generator *g*, *H* is a non-trivial subgroup of *G*, *a* and *b* are integers such that $g^a \in H$ and $g^b \in H$, *q* and *r* are integers such that $g^b = (g^a)^q g^r$ and we wish to prove that $g^r \in H$. (The context of this step is that *a* is the *least* positive integer such that $g^a \in H$, and, in line with the cognitive unit of the previous paragraph, $b = aq + r$ with $0 \le r < a$. From the conclusion $g^r \in H$ it therefore follows that $r = 0$ and then that every element of *H* is a power of g^a.) Thinking of going from $(g^a)^q g^r$ (an element of *H*) to g^r as a process of cancelling by (the element of *H*) g^a, the cognitive unit corresponding to property (ii) above can leave space in the short-term focus of attention to enable the conclusion to be seen without strain. On the other hand, working directly with the symbolic representation (in which the implication '$x \in H$ and $z \in H \Rightarrow xz^{-1} \in H$' has to be synchronised with the equation

'$g^b = (g^a)^q g^r$') involves algebraic complexities that would leave some students losing sight of the direction of the argument.

While articulating specific links might help students form particular cognitive units, this might not be enough to help them with the shift of focus through the different layers of detail, in the 'things' to be manipulated at various stages in a mathematical proof. However, increased awareness of a difficulty can often be a first step in overcoming it and accordingly it can be helpful, in the discussion of the mathematics of a proof, if the feature of compression and expansion be made explicit. Indeed having consciously thought about these meta-processes while following proofs in class, certainly helps some students when they come to construct their own. But it does not help all of the students all of the time. The ability to mentally compress a section of mathematical structure into a single manipulable cognitive unit is not something that can be easily taught. Indeed one of the biggest challenges in mathematics teaching is to find ways to help students in this development.

Acknowledgement

Most of the section, 'Consructing mathematical proofs', is taken from my article, 'Why are proofs difficult?', in *The Mathematical Gazette*, Vol. 84, No. 501, 415-422 (November 2000). I am grateful to the Mathematical Association for permission to reproduce this here.

References

Asiala, M., Cottrill, J., Dubinsky, E. and Schwingendorf, K. (1997) The students' understanding of the derivative as slope, *Journal of Mathematical Behavior*, **16** (4), 399-431.

Barnard,T. and Tall,D.O.(1997). Cognitive units, connections and mathematical proof, *Proceedings of the 21st International Conference for the Psychology of Mathematics Education*, Vol.2, 41-48.

Barnard,T.(1999). Compressed units of mathematical thought, *Journal of Mathematical Behaviour*, **17**(4), 401-404.

Davis,R.B.(1984). *Learning Mathematics: The Cognitive Science Approach to Mathematics Education*. Norwood, NJ: Ablex.

Dubinsky,E.(1991). Reflective abstraction in advanced mathematical thinking. In D.O.Tall(Ed.), *Advanced Mathematical Thinking*, (pp.95-123), Dordrecht:Kluwer.

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

Krutetskii,V.A.(1976). *The Psychology of Mathematical Abilities in Schoolchildren* (Trans. J.Teller, Ed. J.Kilpatrick & I.Wirzup). Chicago IL: University of Chicago.

Sfard,A.(1991). On the dual nature of mathematical conceptions: reflections on processes and objects as different sides of the same coin, *Educational Studies in Mathematics*, **22**(1), 1-36.

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