A Comparative Study of Cognitive Units in Mathematical Thinking

Tony Barnard

Department of Mathematics
Kings College
Strand, London, WC2R 2LS
e-mail: tony.barnard@kcl.ac.uk

David Tall

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

In this paper we consider the theory of 'cognitive units' introduced in 'Cognitive Units, Connections and Mathematical Proof' (Barnard and Tall, 1997) and compare it with other theories of cognitive construction. We find that it includes a theory of 'cognitive compression' to reduce the cognitive strain involved in thinking about more complex concepts that is related to the 'varifocal theory' of Skemp. It generalises the notion of process-object and schema-object encapsulation to give a broader theory of rich cognitive units, with connecting links that minimise cognitive strain and maximise thinking power.

Cognitive units, connections and compression

The notion of 'cognitive unit' that we consider here is 'a piece of cognitive structure that can be held in the focus of attention all at one time' (Barnard and Tall, 1997). We are particularly interested in cognitive units with rich internal links that allow them to be thought of as a single entity. For instance, the equations

$$P = QR, \quad \frac{P}{Q} = R, \quad \frac{P}{R} = Q$$

may initially be conceived as three separate items, together with links between them carried out by performing the same operation to both sides. Later this perception may become compressed into a single relationship between the quantities involved, seen as different ways of saying the same thing (Barnard, 1999).

Such pieces of cognitive structure do not exist in a vacuum. Intimately connected with them are other ideas in the mind that may readily be called to the focus of attention in turn. Our theory is that rich, compact cognitive units allow the thinker to manipulate these ideas in efficient, insightful ways, whereas students with diffuse structures will not find it so easy to make connections between concepts that are themselves diffuse and vague. There is already considerable evidence to support this thesis (Barnard & Tall, 1997; Chin & Tall, 2000; Crowley & Tall, 1999; McGowen, 1998). Not only do such units operate as a form of short-hand, linking many aspects of a complex structure, they also carry along with them connections that are able to guide their manipulation. Such activities may set up new links which may in turn become increasingly strong so that new cognitive units may be formed, building a network of nested mental structures that may span several layers of thought. This offers a manageable level of complexity in which the thought processes can concentrate on a small number of cognitive units at a time, yet link them or unpack them in supportive ways when necessary.

Completed January 9th 2001; submitted to PME25.

These ideas are closely related to Skemp's (1979) "varifocal theory" of cognitive concepts, where a concept may be conceived either as a global whole, or viewed under closer scrutiny to reveal various levels of detail. He referred to this conceptual detail as the interiority of the concept. Powerful cognitive units have a rich interiority appropriate for the task in hand.

The building of cognitive links in such a way that one item in the focus of attention can refer at will to a variety of closely connected pieces of knowledge will be termed ***compression***. In some cases a physical compression is known to occur in brain activity:

> As a task to be learned is practiced, its performance becomes more and more automatic; as this occurs, it fades from consciousness, the number of brain regions involved in the task becomes smaller. (Edelman & Tononi, 2000, p.51)

More generally, however, a process of *long-term potentiation* occurs in which connections between neurons become more easily activated and therefore resonate together as a single structure (Carter, 1998, p. 159). As mathematicians we see an analogy in which the brain has not only 'substructures' that work as a unit, but also 'quotient-structures' caused by the identification of separate units. A possible example is the cortical system for vision which has twenty or so separate regions each of which perform a specific task (eg recognising colours, changes in shade, edges, orientation of edges, movement of edges, identifying objects, and so on). We suggest the manner in which these disjoint areas are connected together to give a unified visual perception may be considered as such a quotient structure. As mathematics educators involved in thinking we have no physical evidence for this, so we pass the metaphor on to neurophysiologists to evaluate its usefulness. Meanwhile it is our role to collect evidence that intimates how mathematical structures may be held as manipulable cognitive units with an interiority that is able to both guide manipulation of the unit and also be subsequently expanded without loss of detail.

Process-object duality

Process-object duality is at the heart of several theories of mathematical development, for instance the encapsulation of process into object (Dubinsky, 1991) or 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 by Asiala et al (1997) 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.)

Sfard (1991) proposed a corresponding sequence, affirming that operational mathematics (use of processes) almost invariably precedes structural mathematics (use of objects). Dubinsky and his colleagues followed the process-object sequence for several years before the strict sequence was loosened:

> … although something like a procession can be discerned, it often appears more like a dialectic in which not only is there a partial development at one level, passage to the next level, returning to the previous and going back and forth, but also the development of each level influences both developments at higher and lower levels. (Czarnocha et al, 1999, p. 98.)

Gray & Tall (1994) took a different view of the relationship between process and object. They saw the role of the *symbol* as being pivotal in the thinking process in a very special way. A symbol such as "3+4" could act as a pivot between a process (of addition) and the concept (of sum). This immense 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. They formulated the notion of *procept* as a combination of process and concept evoked by a single symbol. This theory saw the notion of procept becoming richer (in interiority, to use Skemp's terminology) as different symbols and processes represented the same object, for instance, 6 as $5+1$ or $2+4$. From this range of associations it is possible to compute, say $8+6$, because 6 is $2+4$ and $8+2$ gives 10, and 10 and 4 gives 14. The notion of 'procept' was extended to include all the triples of process-object-symbol that have the same object in a given cognitive context. For instance, 6 is a procept which embraces $3+3$, $5+1$, $2+4$, and so on. Later in the development of the individual, it might also come to embrace $12/2, \sqrt{36}$, $3 \cdot 5+2 \cdot 5$. In our terminology, a procept is therefore a special case of a cognitive unit that grows with interiority as the cognitive structure of the individual gets more sophisticated. This compression of mental schemas or schemes into a cognitive unit features in a range of theories of cognitive development.

Schema-concept duality

Mathematical thinking involves two different kinds of mental activity which are both referred to as schemas or schemes. One is a *sequential* action scheme that occurs in time and is stabilized by long-term potentiation, strengthening and coordinating cognitive links such as the "see-grasp-suck" scheme in the young child. Another refers more specifically to the physical structure of the brain which offers a *multi-connected* schema in which many possible links are available at any given time. The process-object theories seem to focus more on the first of these, theorizing that sequential schemes are encapsulated or reified as mental objects. The second type of schema offers a more subtle way of building up mental concepts that can operate flexibly as cognitive units.

Crowley and Tall (1999) consider how the "linear equation schema", for formulating and solving linear equations may represent the same idea in different forms:

- the equation $y = 3x + 5$,
- the equation $3x - y = -5$,
- the equation $y - 8 = 3(x - 1)$,
- the graph of $y = 3x + 5$ as a line,
- the line through (0,5) with slope 3,
- the line through the points (1,8), (0,5).

For some college algebra students these may all be compressed into a single cognitive unit, with the various representations just alternate ways of expressing the same thing. But it is also clear that there are students who see the structure as consisting of distinct ideas with procedures (that they may not be able to carry out) required to get from one thing to another. A student with such a diffuse view of linear equations may therefore have a partial schema for relating the various representations but not a global schema that easily sees them all as essentially the same cognitive unit.

Dubinsky and his colleagues extended their APOS theory so that, in addition to a process-object construction, there was also a schema-object construction:

> As with encapsulated process, an object is created when a schema is thematized to become another kind of object which can also be de-thematized to obtain the original contents of the schema.
> (Asiala et al, 1997, p.400)

In this way, highly connected mental structures are built up at different levels of detail, connected in various ways.

The distinction between procedural thinking that allows limited success in familiar contexts and conceptual thinking that is more adaptive in new problems has long been a subject of study. Hiebert and Carpenter (1992) proposed two alternative metaphors for cognitive structures—as vertical hierarchies or webs:

> When networks are structured like hierarchies, some representations subsume other representations, representations fit as details underneath or within more general representations. Generalisations are examples of overarching or umbrella representations, whereas special cases are examples of details. In the second metaphor a network may be structured like a spider's web. The junctures, or nodes, can be thought of as the pieces or represented information, and the threads between them as the connections or relationships. (Hiebert & Carpenter, 1992, p. 67.)

Skemp's formulation would allow the nodes in a web to be seen as varifocal hierarchies, thus allowing the two structures to be used together. Likewise a concept could be considered as a web of connected ideas, allowing both hierarchical and web-like structures to coexist in a single structure.

However, the notion of webs and nets are still simplified metaphors for a far more sophisticated mental system. Greater subtlety is essential to be able to reflect on the way we think in mathematics. Consider for example, the statement: $\sin 60^\circ = \sqrt{3}/2$. This may be conceived by an individual as a cognitive unit and linked to a picture such as that in figure 1.

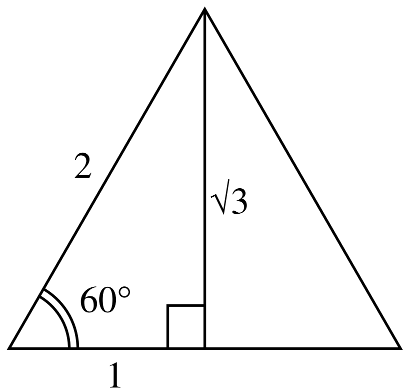

Figure 1: Relationships within an equilateral triangle

This in turn is related to many other ideas 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", "$1^2 + (\sqrt{3})^2 = 2^2$", "the sine of an angle is opposite over hypotenuse", "the opposite is $\sqrt{3}$, the hypotenuse is 2", etc. In this way we see a single cognitive unit linking theorems about triangles, definitions of the trigonometric functions, algebraic representation of a sum of squares, numerical facts about a specific triangle, and so on. It is possible to formulate these partly in terms of hierarchies, for instance, Pythagoras' Theorem has "$1^2 + \left(\sqrt{3}\right)^2 = 2^2$" as a special case, the definition of sine also includes this special case in terms of $\sqrt{3}/2$ as "opposite over hypotenuse". The many links involved also relate to other ideas; the definitions of trigonometric formulae relate to notions of similar triangles having sides in the same ratio; the trigonometric functions have relationships between them such as $\sin^2 + \cos^2 = 1$. Processes of the brain allow these ideas to become intimately connected in such a way that they are easily linked and manipulated.

Professional mathematicians build up highly subtle cognitive units packed with meaning. By focusing on commonly occurring properties which prove useful in making deductions, they build up a range of different theories based on generative concepts translated into chosen systems of axioms. For example

> ... the concept of a group captures the essence of the notion of symmetry and is connected in a precise way to the concept of an equivalence relation, which itself is a precise abstract formulation of the notion of 'sameness' with respect to a given property. Not only are the properties defining a group sufficiently general to be satisfied by a large variety of relatively concrete mathematical objects, but they are also sufficiently special to have lots of powerful consequences at the abstract level. Thus a group is a precisely defined concept which sits at a major junction in the mathematical network of relations. Indeed, one of the most beautiful features of mathematics is the way it allows such precision at even the deepest levels of abstraction. (Barnard, 1996)

Performing cognitive compression

A common method of compression is to use words and symbols as tokens for complex ideas. In particular, words can be used in a way that allows a hierarchical structure to be conceived. For instance, a square is a special case of a rectangle, which is itself a parallelogram, which is a quadrilateral. In the early stages a square and a rectangle may be seen as quite different entities, both having four right angles, but a square has all sides equal, whereas a rectangle has only opposite sides equal. The cognitive unit "rectangle" develops greater sophistication and interiority, growing from meaning just a perceived figure, to a whole collection of rectangular figures (including squares) in any orientation. The realization that squares are special cases of the class of rectangles is an example of cognitive compression where one class of objects is subsumed (for certain purposes) within another.

Language supports the communication and refinement of ideas, both between individuals and within the mind of the individual. It allows verbalized properties of perceptions to be used as a foundation for the development of cognitive units that are sophisticated mental idealizations. These include the notion of a point having position, but no size, or a line having arbitrary length and no thickness. These go on to play their role in proving relationships in elementary Euclidean geometry and perhaps later in more abstract forms of geometry.

A cognitive unit may, or may not, be associated with a natural visual label. For example, in contemplating a mental image of angles written around a circle, in degrees and in radians, there is no need to convert, say, $\pi/4$ radians to 45 degrees (by multiplying by the factor $\frac{\pi}{180}$). The symbols $\pi/4$ and 45° are simply different labels for the same angle. On the other hand, the steps of algebraic manipulation involved in seeing that 'a linear combination of a linear combination of quantities x and y is a linear combination of x and y' is an idea that may be compressed into a cognitive unit without there being a clearly defined label associated with the manipulation process.

Compression can also occur in other ways. For instance, when a collection of ideas or symbols is 'too big' to fit into the focus of attention, it can sometimes be 'chunked' to group into a single unit using some kind of alternative knowledge structure. The four digit number 1914 may be seen not just as a number, but as the year at the beginning of the first world war. This kind of associative link may be used to chunk numbers together into sub-units that can now be held in the limited short-term memory. Most individuals would find the 12-digit number 138234098743, impossible to remember on a single hearing. However, the twelve-digit number 246819141918 can be 'chunked' and remembered as the sequence '2, 4, 6, 8' followed by the dates 1914–1918 of the First World War.

Although it is possible to formulate a range of possible compression strategies, individuals do not use them all to the same extent. This leads to different individuals processing mathematical ideas in ways which may have very different outcomes that may lead to success in some and failure for others. Gray and Tall (1994) noted that children who use the longest counting process—count-all (count one set, count the second, put them together and count them all)—could also remember certain "known facts" such as 1+1 is 2 or 2+2 is 4. But none of these ever put together "known facts" to obtain "derived facts", such as "4+3 is 7" because "4+4 is 8", or "12+2 is 14" because "2+2 is 4". Instead, they always computed arithmetic problems (whose answers were not immediately known facts) by counting. We hypothesize that the sums performed by these children are seen as counting processes and not as meaningful cognitive units with any interiority. The "known facts" for them were isolated and not in a sufficiently rich compressed form which could be mentally manipulated as cognitive units. On the other hand, children who were able to derive new number facts from known related ones were able to perform arithmetic in a far more flexible way which used numbers as compressed cognitive units with powerful interiority.

Another bifurcation occurs in elementary algebra. Here an expression such as "$2+3x$" stands for a potential arithmetic operation such as "add 2 to the product of three times whatever x is". This can cause discomfort for students who feel that a problem "must have an answer" as it does in arithmetic. They are therefore faced with manipulating expressions as mental objects that have only a potential, rather than an actual, internal process of evaluation. This can lead simply to procedural compression in which students learn to carry out a solution process by rote ("collect together like terms", "get the numbers on one side and the variable on the other", "simplify to get the solution", etc). Others are able to conceive the algebraic expressions as entities that can be manipulated. They may go on to conceive of the equation itself as a cognitive unit expressing a given relationship, with a "solution process" as a cognitive unit that can be unpacked to give an efficient route to the solution.

Krutetskii (1976) studied this curtailment of mathematical reasoning, in which capable students would compress their solutions 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.)

Great success in calculation may be developed with a huge range of connected ideas, some meaningful, some rote-learnt, as Nobel Prizewinner, Richard Feynman reports:

> I memorized a few logs and began to notice things. For instance, if somebody says, "What is 28 squared?", you notice that the square root of 2 is 1.4 and 28 is 20 times 1.4, so the square of 28 must be around 400 times 2, or 800. If somebody comes along and wants to divide 1 by 1.73, you can tell them immediately that it's .577 because you notice that 1.73 is nearly the square root of 3, so 1/1.73 must be one-third of the square root of 3. And if it's 1/1.75, that's equal to the inverse of 7/4 and you've memorized the repeating decimals for sevenths: .571428… .
> (R. Feynman, 1985, p. 194.)

Mathematical proof

Mathematical proof involves cognitive units and connections of a more general type than those encountered in elementary mathematics. (Barnard & Tall, 1997). In addition to sequential procedures of calculation or symbol manipulation found in arithmetic and algebra, mathematical proof often requires the synthesis of several distinct cognitive links to derive a new *synthetic* connection. For instance, in the standard proof that $\sqrt{2}$ is irrational, having written $\sqrt{2} = a/b$ as a fraction in its lowest terms, the step from "$\sqrt{2} = a/b$" to "$a^2 = 2b^2$" is an elementary sequence of algebraic operations, but the step from "a^2 is even" to "a is even" requires a subtle synthesis of other cognitive units, such as "a^2 is either even or odd" and "if a were odd, then a^2 would be odd." These synthetic links constitute an essential difference between elementary procedures in arithmetic or algebra and more sophisticated linkages involved in mathematical proof.

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 (Barnard, 2000) has to do with the shifting of focus through the different layers of detail in the cognitive units to be manipulated: statements, statements within statements, expressions within statements, symbols within expressions, etc. In a lecture, the lecturer may implicitly specify the level of items that are to be the primary objects of thought at any stage. For example, in a proof by induction on n of a statement $P(n)$, the distinction needs to be made as to when $P(n)$ is to be thought of as a compressed item within the statement, "$P(n)$ implies $P(n+1)$", or when it is to be unpacked for a finer grained manipulation.

> 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. (Barnard, 2000.)

The wider challenge in mathematics education is how we can help students to construct appropriately linked cognitive units that are flexible and precise to help them build mathematics as a coherent and meaningful structure. These cognitive units arise naturally in human thinking and take on a wide range of roles – general strategies, specific information, routinized sequences of steps, linked together to produce mathematical thinking. Without cognitive units of appropriate manipulable size, thinking becomes diffuse and imprecise and is far less likely to be successful. Even with the development of manipulable cognitive units in individuals that give current success, there will still be challenges requiring intelligent reconstruction to cope with novel situations.

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. (1996), Structure in mathematics and mathematical thinking, *Mathematics Teaching*, No.155, 6-10.

Barnard, T. (1999), Compressed units of mathematical thought, *Journal of Mathematical Behavior*, 17, 4, 1–4.

Barnard, T. (2000), Why are proofs difficult?, *The Mathematical Gazette*, Vol. 84, No. 501, 415-422.

Barnard, T. & Tall, D.O. (1997). Cognitive Units, Connections, and Mathematical Proof. In E. Pehkonen, (Ed.), *Proceedings of the 21st Annual Conference for the Psychology of Mathematics Education*, Vol. 2 (pp. 41-48). Lahti, Finland.

Crowley, L & Tall, D. O. (1999). The Roles of Cognitive Units, Connections and Procedures in achieving Goals in College Algebra. In O. Zaslavsky (Ed.), *Proceedings of the 23rd Conference of PME*, Haifa, Israel, 2, 225–232.

Carter, R. (1998). *Mapping the Mind*, London: Weidenfeld & Nicholson.

Czarnocha, B. Dubinsky, E., Prabhu, V. & Vidakovic, D. (1999). One Theoretical Perspective in Under-graduate Mathematics Education Research. In Zaslavsky, O. (Ed.) *Proceedings of the 23rd Conference for the Psychology of Mathematics Education*, I, 95–110.

Dubinsky, E. (1991), Reflective Abstraction in Advanced Mathematical Thinking. In D. O. Tall (ed.) *Advanced Mathematical Thinking*, Dordrecht: Kluwer, 95-123.

Chin, E-T. & Tall, D. O. (2000). Making, Having and Compressing Formal Mathematical Concepts. In T. Nakahara and M. Koyama (eds.), *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education* 2, 177–184. Hiroshima, Japan.

Feynman, R. (1985). *Surely you're Joking Mr Feynman*, New York: W. W. Norton & Co. Inc, (paperback edition, London: Vintage, 1992).

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.

Hiebert, J. & Carpenter, T. P. (1992). Learning and Teaching with Understanding. In D. Grouws, (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 65–97). New York: MacMillan.

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

McGowen, M. A., (1998). *Cognitive Units, Concept Images, and Cognitive Units: An Examination of the Process of Knowledge Construction*, PhD Thesis, University of Warwick, UK.

Sfard, A. (1991). On the Dual Nature of Mathematical Conceptions: Reflections on processes and objects as different sides of the same coin, *Educational Studies in Mathematics* 22, 1–36.

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