Cognitive Difficulties in Learning Analysis

David Tall

Mathematics Education Research Centre
Warwick University
COVENTRY CV4 7AL, UK
e-mail: D.O.Tall@csv.warwick.ac.uk

Introduction

Although mathematicians have hunches as to *why* students cannot understand analysis, it is not part of their task as professional researchers to study the cognitive development of how the subject is learned. Instead the general method of trying to improve student success is to reorganise the manner in which the subject is presented. Sometimes this gives a perceived measure of success, but overall there is as yet no agreed consensus as to what must be done to improve the situation. One of the serious reasons for this is that the search for clues is being performed in the wrong area. To understand why students can't do analysis clearly requires a study of the *students* as well as the subject matter. In this article I attempt briefly to summarise some of the research into student conceptions of mathematics which may be of assistance. Much of this has been done in the last years of school and the transition to university and is not always directed at the type of student who will be faced with mathematical analysis at its most abstract. Students across a spectrum of ability and experience differ greatly in the vision they bring to the subject and the problems that they face. Nevertheless, the results are clear enough to focus on difficulties which students experience in learning analysis. Hopefully this will lead to further serious empirical studies of student difficulties in analysis. A pilot study carried out recently at Warwick suggests that there is much of value to be learned.

The student conceptions on which the theory of analysis is built

We all have experiences which build up our own personal imagery of mathematical ideas. Conceptions of such things as *real number*, *limit*, *function*, *continuity*, *completeness*, and so on in every individual are built on previous experiences which are present in the mind at the same time as the formal definitions. It is important therefore to be aware of these other conceptions that coerce our thinking and often produce confusion and conflict in our students (and also in ourselves). The term *concept image*[1] has been used by mathematics educators in a technical sense to describe the total cognitive structure, including all mental pictures and imagery related to the concept, in the mind of an individual. The concept image is likely to include factors which conflict with the formal definition and cause serious cognitive conflict for the learner. Some student concept images are discussed in the book *Advanced Mathematical Thinking*[2] (a compilation written by sixteen international maths educators) and in *The Transition to*

Advanced Mathematical Thinking chapter of the recent *Handbook on the Teaching and Learning of Mathematics*[3] published in the USA by the National Council of Teachers in Mathematics.

The *real number system* itself is by no means clearly formulated in most students' minds. Young children learn about the accurate arithmetic of whole numbers and later of negative numbers and fractions. Their arithmetic is precise and gives *exact* answers. Later, students add other numbers to their repertoire, such as $\sqrt{2}$ and π. As soon as computations are performed with these, they become *approximate*, accurate to a specified number of decimal places. Although these numbers may be visualised as points on a number line, this does not mean that the number line itself is understood in a coherent way. In particular, 'infinite decimals" are often regarded by sixth-formers as being *improper* in the sense that they "go on forever" and do not *exactly* specify the limit value[4]. Instead the notation $\sqrt{2}=1.414\ldots$ says "the square root of two can be computed to any desired number of decimal places and is 1.414 to three places". It does *not* mean that the infinite decimal is the limit of the sequence of decimal approximations. Students' whole experience of numbers in upper secondary school is in terms of adequately accurate *approximations* rather than accurate precision. The conflict between the abstract desire for precision and the practicalities of arithmetic lead to subtle conflicts in the mind.

The notion of *limit* is not clearly conceived by most students arriving at university. It is considered as a *process* of *getting closer* rather than the *concept* of limiting *value*[5]. If it is thought of as an actually *quantity*, the process of *getting smaller* is usually encapsulated as an *arbitrarily small* ***variable*** *quantity*, in a manner which fits more with Cauchy's idea of the continuum than our own[6]. Students have intuitive conceptions of infinitesimals. That is not to say that they are ready for non-standard analysis, for their conceptions conflict with this theory also. For instance "nought point nine recurring" may be the "nearest number to 1 without being 1", whilst the non-standard number

$$0\cdot\underbrace{999\ldots9}_{n\ \ places}=1-{}^{1}\!/_{10^n}\ \text{(for infinite } n\text{)}$$

can be made closer to one by making *n* larger. Beliefs that may surprise mathematicians are possible, including the notion found amongst a sizeable minority of undergraduates at an English university that there is no "smallest positive real number" yet there can be a "first positive real" (in the form "nought point infinite noughts, one").

Dynamic views of limit and consequent misconceptions are commonplace amongst undergraduates. The change from limit *as a variable process* to limit *as a fixed value* is a significant conceptual change which is a worthy focus of attention.

Other concept images cause conflict for students. For instance, earlier experiences suggest that a function is given by a formula[7], so the "funny functions" mathematicians may try to draw as piece-wise connected graphs to 'motivate' analysis theorems are received by some students with incredulity. A *continuous* function[8] is conceived as one given by "drawing without taking the pencil off the paper", and is often further motivated in this way in analysis, but it leads to more grief than insight, as we shall see later in this article.

Other concepts of analysis, such as *sequences* and *series* give cause for confusion. Essentially students see the words as being interchangeable at school and continue to confuse them even when clearly told the definitions. As definitions come thick and fast, from convergence of a sequence, to convergence of a series, to cauchy convergence of sequence and series, and then on to specific tests of convergence, the development may be clear to the lecturer, but not to the student who must first construct the meaning of these concepts and is often not able to cope with them in the given time. Instead a serendipity focus on isolated pieces of information that happen to appeal to the individual occurs during the lectures. Thus a student may remember an individual test of convergence of series that happens to work for him or her, yet may have no recall of the fundamental definition of sequence, series or convergence on which the test is based.

Defining concepts

The idea of *defining* a concept is new to most students. In a platonic sense a concept either exists or it doesn't. Given a concept that is already familiar, one might *describe* it (a square is a polygon with four equal sides and one internal angle is a right angle). One might attempt to be more prescriptive ("if we meet a polygon with four equal sides and one internal angle a right angle *then it is a square*"). But the idea of *defining* a concept and then asserting that *because it is defined it exists* causes severe cognitive problems. The difficulties are greater in mathematical analysis where definitions are often formulated as 'if-then' *processes* – "given an ε we need to find an N such that …" – rather than in terms of an *object* (a triangle is a figure with three straight sides). The processes are so long that they exceed short-term memory processing capacity. Beginners forget the start of the definition before they reach the end. (In an experiment with non-maths majors studying analysis, *none* could remember the definition of convergence of a sequence after two weeks of hearing lectures on the subject.) For all but the most able students (and even these struggle), when definitions are first encountered, they are not compressed into a form that they can readily be manipulated and so cannot be comprehended.

Mathematical proof

It is useful to distinguish three types of proof: *mathematical proof* based on formal definitions, *euclidean proof* based on definitions which are actually descriptions of concrete objects having associated visual mental images and *thought experiments* where the givens are imagined true and then one sees if this leads to the proof of the theorem. Many theorems in analysis are *trivially true* by thought experiment. For instance, if a continuous function is negative somewhere and positive somewhere else, it is clearly zero somewhere in between – just imagine a picture of a graph drawn from negative to positive without taking the pencil off the paper and see! If a_n tends to a and b_n to b, then *clearly* a_n+b_n tends to $a+b$ – how could it be otherwise? Euclidean type proofs of these theorems – where the mathematical objects are pictured and described – have visual or dynamical elements which may lead to oversimplified beliefs characteristic of thought experiments.

Visualisations are important in conceptualising a wide range of mathematical ideas. The psychologist Paivio[9] (1986) proposed a *dual-coding theory* in which visual and verbal/symbolic representations are linked in the brain for mutual support. Pictures prove to give enhanced memories of linked verbal data and vice versa. In my work I have tried to provide students with suitably complex computer generated pictures[10] (e.g. everywhere continuous, nowhere differentiable functions whose area functions are locally straight, or functions with differing values on rationals and irrationals to motivate Lebesgue theory[11]). These enable students to *see* why theorems need to be proved and how things might go wrong but the visual elements rarely have the same sequential/deductive links characteristic of words and symbols. There remains the need to translate the visual insight from a thought experiment into a sequential proof, and this is non-trivial.

The problem of using definitions in mathematical proofs

We have already noted that mathematical definitions are often inordinately long and often formulated in the form of a process rather than a direct description of the constituent parts – for instance, the definition of convergence of a sequence (given ε, there exists N such that …). They are better handled in *written* form where the whole sentence can be scanned and any part focused on at will, than in *spoken* form. Hence lectures on the subject are an unnatural method of communication! If the communication is spoken, the language tends to revert to imagery which is more appropriate for thought experiment than formal proof. Because the thought experiments are “known” to be true by the students, the theorems are seen to be unnecessarily complicated.

Theorems using definitions require a process of translation from words in the theorem, (such as *limit*, *continuous function*, *least upper bound*) and their attendant imagery, to their formal definitions. The game is to translate all the given assumptions in this form, to manipulate the resultant data and to deduce the definitions required to return to the conclusion of the theorem. Regrettably even this does not follow by a simple sequential computations met in school mathematics. Consider the proof of the sum of the limits is the limits of the sum, with its half epsilons and choice of the maximum of the two consequent Ns to give the final N.

The growing difficulty of learning analysis

In a recent research project at Warwick University five selected students were interviewed at three week intervals throughout an analysis course. Certain factors became clear. At the beginning, when presented with the axioms of real numbers, two (Mark with two As in A-level mathematics, Linda with an international baccalaureate) saw the need to specify what was assumed as a vital starting point (definitions) to make deductions, Three others (Nicholas with a B in maths, Anthony with B/E, Brian with A/F in maths/further maths) were confused at being given a list of facts about numbers that they had known since they were children (descriptions!). In the second and third interviews, students were asked, among other things, to define the *least upper bound*. Only Linda could give a fluent *formal* definition of the notion of least upper bound. *She had learned to visualise the page on which it was written and could look along her image of the definition to see its constituent parts.* So she saw the definition as a spatial array of grouped symbols rather than a simple logical sequence of words. But she was currently confused by other concepts, for instance, she did not know the difference between a sequence and a series which had been the topic of discussion for a couple of weeks and when asked to give the definition of convergence of a sequence, offered the cauchy definition of convergence of a series.

Mark remembered *most* of the definition, but here and elsewhere often came to a total stop in the middle of an explanation of a topic when his memory failed, getting back on the rails when he refreshed himself by looking at the notes.

The other three students struggled desperately to remember the definition, often falling back on concept imagery of the type "an upper bound is bigger than any element in the set" with the implication (for them) that the least upper bound was "the largest element in the set". Quite simply, they had not compressed the definition into a form where it could be used fluently and flexibly in their short-term memory store. In the lectures they usually interpreted the theorems in terms of imagery more suited to 'thought experiment' rather than 'formal proof'. They saw theorems as "obvious" (as thought experiments) but "could not understand them" (as formal theorems). This was not solely through lack of memory, as one of the weaker students showed when later committing the definition of least upper bound to memory. It seemed more due to the conflict of their concept imagery and what they believed to be true from their previous experience with what needed to be 'proved' through a method which, for them, failed to be accompanied by any necessity or conviction.

As the lectures continued, the gap between the concepts as meaningfully related to definitions became wider as the students were faced with the increasing overload of handling so much information in their short-term working memory store. By the time the definition of Riemann integral came in the next term, students were floundering with the notion of partition, with each partition having an upper and lower sum using a theorem about continuous functions on each closed subinterval, and then the proposition that all upper sums are greater than all lower sums and hence upper sums have a greatest lower bound and lower sums a least upper bound and that these can be shown to be equal, and that this value is called the integral. The enormity of this excessive sequence of complex statements compared with their earlier algorithmic ideas that the integral of x^n is $\frac{x^{n+1}}{n+1}$ (for $n\neq 1$) needs little comment.

A strange phenomenon occurred. Brian and Nicholas (who were in serious difficulties) could give some kind of verbal description of the ideas but only as far as a rough definition of upper sum and lower sums. The more successful Mark was still attempting to cope with information overload. He knew the definitions of continuity, that theorems were involved to show that a continuous function attains its maximum and minimum in a closed interval, could formulate upper and lower sums, but began to get a stage where the complexity of his description collapsed from its own excessive internal weight. Meanwhile Anthony, one of the weaker students who had missed a lot of lectures, realised that he had twice been asked about the definition of least upper bound in interview and had committed it to memory. He could repeat the precise definition of least upper bound, could describe the whole of the process of upper and lower sums, and their respective greatest lower bound and least upper bound and that when these were equal then this value was the integral. It all sounded incredibly convincing. But a little probing revealed that he had no formal concept of continuous function and still considered that the least upper bound was the biggest element in the set. His description of Riemann sums was based on intuitive concept images, not formal definitions, where the least upper bound in each subinterval was visualised as the highest point on an intuitively 'continuous' graph. Although he might obtain high marks on a question about Riemann sums, his formal conceptualisation was deficient. The joke is that he was only able to operate *because* his imagery was deficient. Had he attempted to deal with the definitions in all their glory he would almost certainly have crashed under the overload in the same manner as some of his more successful colleagues. At the end of the lectures he withdrew before the examination and changed subject.

What becomes apparent when we look at students' difficulties in analysis is that they are genuine, deep and inadequately researched. There is a broad spectrum of ability and insight where each individual has personal conflicts to face. The "able mathematician" amongst them may be able to address some of these problems without explicit instruction and may succeed despite the many difficulties which litter the way. Others in the spectrum are faced with difficulties of meaning which simply cannot be addressed by dedication and hard work. Those who do work hard may learn to recall and repeat sequences of arguments, but the complexity and lack of compression of the knowledge is such that it cannot be held in the mind in a compact enough form to allow it to be mentally manipulated in a manner characteristic of successful mathematicians.

The talented lecturer may receive student approval by giving more inspiring lectures than the norm, but the problem can only be addressed seriously by finding out the nature of students' conceptions and how they grow. This is not something that mathematicians give kudos for. Recently the London Mathematical Society refused to allow research into advanced mathematical thinking as a named area of interest for a mathematician (although 'history of maths' and 'general' are allowed). The reasoning was that Advanced Mathematical Thinking is not an accepted category in *Mathematical Reviews.* The American Mathematical Society later considered the time was not yet ripe to change its list of categories.

I predict that a change of heart will occur in the next five years. It will have to. In the changing society we live in, we cannot go on teaching mathematics to students in ways that the majority patently fail to understand. It is not that they are poor students, but that the mathematical community has not grasped the nettle that the cognitive study of mathematical thinking is as important to mathematics as group theory, topology or functional analysis. If it were as easy as those subjects, surely *somebody* would have got it right! The study of advanced mathematical thinking will one day be regarded as important to the professional mathematician as one distant branch of mathematics research is important to another. The study of finite simple groups may not be important to the chaos theorist, but it *is* important to the community of mathematicians. In like manner, the study of *how* we think in advanced mathematics may not be as important to mathematicians as *what* we think, but if mathematicians do not support those doing research in this area by giving encouragement and the opportunity to earn respect in an intellectually demanding task, then mathematics itself can only be the poorer.

References

1. 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.
2. *Advanced Mathematical Thinking*, 1992, (ed David Tall), published by Kluwer, Dordrecht – a collection of articles by 16 international mathematics educators describing the current situation in cognitive research at this level.
3. *Handbook of Research in the Teaching and Learning of Mathematics*, 1992, (ed D. A. Grouws), New York: Macmillan. (The most recent and authoritative review of mathematics education research from an American perspective with an international view of Advanced Mathematics.)
4. Monaghan, J. D., 1986: *Adolescents' Understanding of Limits and Infinity*. unpublished Ph.D. thesis, Warwick University, U. K.
5. Schwarzenberger, R. L. E. & Tall, D. O., 1976: Conflicts in the learning of real numbers and limits, *Mathematics Teaching*, 82, 44–49.
6. Cornu, B. 1992: Limits. In Tall, D. O. (editor), *Advanced Mathematical Thinking*, Kluwer: Dordrecht
7. Bakar, M. & Tall, D. O., 1992: Students' Mental Prototypes for Functions and Graphs, *Int. J. Math Ed Sci & Techn.*, 23 1, 39–50.
8. Paivio, A., 1986: *Mental representations: A dual coding approach*, Oxford University Press.
9. Tall, D. O., 1987: *Readings in Mathematical Education: Understanding the calculus*, Association of Teachers of Mathematics.
10. Tall, D. O. 1993: Real Mathematics, Rational Computers and Complex People, *Proceedings of the Fifth Annual International Conference on Technology in College Mathematics Teaching*, 243–258, Addison Wesley: New York. Software: *Real Functions and Graphs* (for the Archimedes Computer), Cambridge University Press.