

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): David Tall and Shlomo Vinner

Article Title: Concept image and concept definition in mathematics with particular reference to limits and continuity

Year of publication: 1981

Link to published version: <http://dx.doi.org/10.1007/BF00305619>

Publisher statement: The original publication is available at www.springerlink.com

Concept Image and Concept Definition in Mathematics with particular reference to Limits and Continuity

David Tall

Mathematics Education Research Centre
University of Warwick

Shlomo Vinner

Science Teaching Center
Hebrew University

ABSTRACT The concept image consists of all the cognitive structure in the individual's mind that is associated with a given concept. This may not be globally coherent and may have aspects which are quite different from the formal concept definition. The development of limits and continuity, as taught in secondary school and university, are considered. Various investigations are reported which demonstrate individual concept images differing from the formal theory and containing factors which cause cognitive conflict.

Compared with other fields of human endeavour, mathematics is usually regarded as a subject of great precision in which concepts can be defined accurately to provide a firm foundation for the mathematical theory. The psychological realities are somewhat different. Many concepts we meet in mathematics have been encountered in some form or other before they are formally defined and a complex cognitive structure exists in the mind of every individual, yielding a variety of personal mental images when a concept is evoked. In this paper we formulate a number of general ideas intended to be helpful in analysing these phenomena and apply them to the specific concepts of continuity and limits.

1. Concept image and concept definition

The human brain is not a purely logical entity. The complex manner in which it functions is often at variance with the logic of mathematics. It is not always pure logic which gives us insight, nor is it chance that causes us to make mistakes. To understand how these processes occur, both successfully and erroneously, we must formulate a distinction between the mathematical concepts as formally defined and the cognitive processes by which they are conceived.

Many concepts which we use happily are not formally defined at all, we learn to recognise them by experience and usage in appropriate contexts. Later these concepts may be refined in their meaning and interpreted with increasing subtlety with or without the luxury of a precise definition. Usually in this process the concept is given a symbol or name which enables it to be communicated and aids in its mental manipulation. But the total cognitive structure which colours the meaning of the concept is far greater than the evocation of a single symbol. It is more than any mental picture, be it pictorial, symbolic or otherwise. During the mental processes of recalling and

manipulating a concept, many associated processes are brought into play, consciously and unconsciously affecting the meaning and usage.

We shall use the term *concept image* to describe the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes. It is built up over the years through experiences of all kinds, changing as the individual meets new stimuli and matures.

For instance the concept of subtraction is usually first met as a process involving positive whole numbers. At this stage children may observe that subtraction of a number always *reduces* the answer. For such a child this observation is part of his concept image and may cause problems later on should subtraction of negative numbers be encountered. For this reason all mental attributes associated with a concept, whether they be conscious or unconscious, should be included in the concept image; they may contain the seeds of future conflict.

As the concept image develops it need not be coherent at all times. The brain does not work that way. Sensory input excites certain neuronal pathways and inhibits others. In this way different stimuli can activate different parts of the concept image, developing them in a way which need not make a coherent whole.

We shall call the portion of the concept image which is activated at a particular time the *evoked concept image*. At different times, seemingly conflicting images may be evoked. Only when conflicting aspects are evoked *simultaneously* need there be any actual sense of conflict or confusion. Children doing mathematics often use different processes according to the context, making different errors depending on the specific problem under consideration. For instance adding $\frac{1}{2} + \frac{1}{4}$ may be performed correctly but when confronted by $\frac{1}{2} + \frac{1}{3}$ an erroneous method may be used. Such a child need see no conflict in the different methods, he simply utilises the method he considers appropriate on each occasion.

The definition of a concept (if it has one) is quite a different matter. We shall regard the *concept definition* to be a form of words used to specify that concept. It may be learnt by an individual in a rote fashion or more meaningfully learnt and related to a greater or lesser degree to the concept as a whole. It may also be a personal reconstruction by the student of a definition. It is then the form of words that the student uses for his own explanation of his (evoked) concept image. Whether the concept definition is given to him or constructed by himself, he may vary it from time to time. In this way a *personal* concept definition can differ from a *formal* concept definition, the latter being a concept definition which is accepted by the mathematical community at large.

For each individual a concept definition generates its own concept image (which might, in a flight of fancy be called the “concept definition image”). This is, of course, part of the concept image. In some individuals it may be

empty, or virtually non-existent. In others it may, or may not, be coherently related to other parts of the concept image. For instance the concept definition of a mathematical function might be taken to be “a relation between two sets A and B in which each element of A is related to precisely one element in B .” But individuals who have studied functions may or may not remember the concept definition and the concept image may include many other aspects, such as the idea that a function is given by a rule or a formula, or perhaps that several different formulae may be used on different parts of the domain A . There may be other notions, for instance the function may be thought of as an *action* which maps a in A to $f(a)$ in B , or as a *graph*, or a *table of values*. All or none of these aspects may be in an individual’s concept image. But a teacher may give the formal definition and work with the general notion for a short while before spending long periods in which all examples are given by formulae. In such a case the concept image may develop into a more restricted notion, only involving formulae, whilst the concept definition is largely inactive in the cognitive structure. Initially the student in this position can operate quite happily with his restricted notion adequate in its restricted context. He may even have been taught to respond with the correct formal definition whilst having an inappropriate concept image. Later, when he meets functions defined in a broader context he may be unable to cope. Yet the teaching programme itself has been responsible for this unhappy situation.

As we shall see shortly the concept images of limit and continuity are quite likely to contain factors which conflict with the formal concept definition. Some of these are subtle and may not even be consciously noted by the individual but they can cause confusion in dealing with the formal theory. The latter is concerned only with that part of the concept definition image which is generally mutually acknowledged by mathematicians at large. For instance, the verbal definition of a limit “ $s_n \rightarrow s$ ” which says “we can make s_n as close to s as we please, provided that we take n sufficiently large” induces in many individuals the notion that s_n cannot be equal to s (see Schwarzenberger and Tall, 1978). In such an individual this notion is part of his concept definition image, but not acknowledged by mathematicians as part of the formal theory.

We shall call a part of the concept image or concept definition which may conflict with another part of the concept image or concept definition, a *potential conflict factor*. Such factors need never be evoked in circumstances which cause actual cognitive conflict but if they are so evoked the factors concerned will then be called *cognitive conflict factors*. For instance the definition of a complex number $x+iy$ as an ordered pair of real numbers (x, y) and the identification of $x+i0 = (x, 0)$ as the real number x is a potential conflict factor in the concept of complex number. This is because it includes a potential conflict with the set-theoretic notion that the element x is distinct from the ordered pair $(x, 0)$. Students in a questionnaire (Tall, 1977b) often regarded a real number such as $\sqrt{2}$ as not being a complex number and yet several of these defined real

numbers as “complex numbers with imaginary part zero.” Thus $\sqrt{2}$ was regarded as real and $\sqrt{2}+i0$ as complex. These were conveniently considered as being distinct entities or the same, depending on the circumstances, without causing any cognitive conflict. They only become *cognitive* conflict factors when evoked simultaneously.

In certain circumstances cognitive conflict factors may be evoked subconsciously with the conflict only manifesting itself by a vague sense of unease. We suggest that this is the underlying cause for such feelings in problem solving or research when the individual senses something wrong somewhere; it may be a considerable time later (if at all) that the reason for the conflict is consciously understood.

A more serious type of potential conflict factor is one in the concept image which is at variance not with another part of the concept image but with the formal concept definition itself. Such factors can seriously impede the learning of a formal theory, for they cannot become actual cognitive conflict factors unless the formal concept definition develops a concept image which can then yield a cognitive conflict. Students having such a potential conflict factor in their concept image may be secure in their own interpretations of the notions concerned and simply regard the formal theory as inoperative and superfluous.

The notions so far described are all clearly manifested in the various concepts of limit and continuity. In the remainder of the article we describe a few of the problems caused by a concept image which does not coherently relate to the concept definition and the resulting potential conflicts.

2. Practical curriculum problems

There are several practical problems imposed in the teaching of the concepts of limits and continuity. If we confine ourselves to the three notions set out:

- (i) limit of a sequence $\lim_{n \rightarrow \infty} s_n$,
- (ii) limit of a function $\lim_{x \rightarrow a} f(x)$,
- (iii) continuity of a function $f: D \rightarrow \mathbf{R}$,

then we find that in English schools these are rarely taught in a logical order. Practical considerations and the need to learn the calculus at the earliest opportunity lead to the notion of a limit of a function first being discussed in differentiation as

$$\lim_{\delta x \rightarrow 0} \frac{f(x + \delta x) - f(x)}{\delta x}$$

or in some other notation. Very soon the standard formulae for derivatives are derived and the general notation of limit recedes into the background. When it is discussed, initially it is usually in a dynamic form:

“ $f(x) \rightarrow c$ as $x \rightarrow a$,”

or

“ $f(x)$ approaches c as x approaches a .”

Later the limit may be given in some quantified form as

“we can make $f(x)$ as close as we like to c provided that we take x sufficiently close to a ,”

or

“for all $\epsilon > 0$, there exists $\delta > 0$ such that
 $0 < |x-a| < \delta$ implies $|f(x)-c| < \epsilon$.”

In a similar manner the limit of a sequence is often dealt with briefly then greater stress is laid on the limit of a series, for instance in the form of a geometric series

$$1 + x + \dots + x^{n-1} = \frac{1 - x^n}{1 - x} \rightarrow \frac{1}{1 - x} \text{ for } |x| < 1,$$

or a Taylor series

$$f(x + h) = f(x) + \frac{hf'(x)}{1!} + \frac{h^2 f''(x)}{2!} + \dots$$

The notion of continuity is rarely alluded to as a formal definition, but a concept image is built up from informal usage of the term. For instance in The School Mathematics Project Advanced Level texts (hereinafter referred to as “SMP”) the concept images of limits and continuity are carefully built up over the two years of the course with fairly formal concept definitions only being given at the very end. In this way the concept image is intended to lead naturally to the concept definition, but in practice certain potential conflicts occur which can cause cognitive conflict for those who later study formal analysis.

These comments are not intended as a criticism of the curriculum as it now is. They are intended to demonstrate the realities of the situation where concept images are built up practically and informally, then an initial concept definition is given which is an approximation to the formal definition met by the small percentage of students who go on to study mathematics at university. Our intention is to consider the three given concepts in turn and discuss some of the potential conflict factors which may occur.¹

¹ The authors are grateful to Cambridge University Press and The School Mathematics Project for granting permission to quote from *Advanced Mathematics* (Metric) Books 1-4.

3. Limits of sequences

Sequences are usually introduced without formal definition. For instance in SMP, Book 1, chapter 2 (page 31) it says:

Suppose that you wanted to work out the value of 2^7 how would you do it?
Probably you would recite to yourself the sequence of numbers

2, 4, 8, 16, 32, 64, 128;

and perhaps you would also count off

1, 2, 3, 4, 5, 6, 7

on your fingers at the same time, which would help you to keep your place in the sequence. Thus when you had counted off n fingers you would have recited the number 2^n .

It then goes on to explain how the process can be carried out on a computer, leading through inductive definitions, proof by induction and several intermediate topics before defining the notion of function in chapter 4. After a brief set-theoretic discussion recalling earlier work on relations the notion of function is defined (SMP, Book 1, chapter 4, p. 83):

It will be noticed that the relation ‘is the husband of’ has a special property, at least in Christian countries: that each married man has just one wife. This means that each member of the domain M is associated with one and only one member of the codomain. Relations having this property occur very frequently, and play an important part in mathematics. They are known as ‘functions’.

A function is a relation between elements of two sets – the domain and the codomain – such that each member of the domain is related to one and only one member of the codomain.

A dozen pages later the notion of ‘sequence’ is recalled and reinterpreted as part of the function concept (SMP, Book 1, chapter 4, p. 95):

We have sometimes looked at strings of numbers such as

$0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots$

and used the word ‘sequence’ to describe them. The separate numbers are called the ‘terms’ of the sequence: thus the first term is 0, the second term is $\frac{1}{2}$ and so on. We see at once that with this sequence is associated a mapping, described by the table:

Position of term	Value of term
1	$\rightarrow 0$
2	$\rightarrow \frac{1}{2}$
3	$\rightarrow \frac{2}{3}$
4	$\rightarrow \frac{3}{4}$
...	...

It is therefore possible to define a sequence as a function having for domain the set N of natural numbers $\{1, 2, 3, 4, \dots\}$.[†] The image $f(n)$ of the number n is described as the n th term of the sequence.

†The word sequence is also used for functions having for domain the set of non-negative integers $\{0, 1, 2, 3, \dots\}$. The number $f(n)$ is then properly described as the $(n+1)$ th term.

Apart from giving arithmetic and geometric progressions ($n \rightarrow a+bn$, $n \rightarrow ar^n$ respectively) as examples of sequences, the text then returns to further discussion of the function concept.

Later on various iterative processes are used to approximate certain numbers, for instance the Newton-Raphson technique for a root of an equation or step-by-step methods to solve differential equations. It is only in the very last chapter of the fourth and last book of the course that the notion of convergence of a sequence is discussed in any formal sense. First there is a discussion of the “challenge and response” procedure in the particular case (SMP, Book 4, chapter 43, p. 1309):

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots, \left(\frac{1}{2}\right)^i, \dots$$

No term of this sequence is zero, but if we programmed a computer to evaluate them, the terms would soon be printed as zeros, since they would be smaller than the smallest number that the machine could handle. Furthermore, this would be true however many decimal places the output could print. Challenged to produce a number N which will enable us to say ‘all terms from the N th onwards will be zero to the first 50 decimal places’, we can certainly find such a number; and we can do so whatever number the challenger puts in the place of 50. Notice two things: we are not committed to produce the smallest such number N , but only one suitable N ; secondly, we cannot produce a single number N for all time, which will dispose of the challenge no matter what number is put in place of 50. Our response depends on the challenge with which we are faced. Formally, we can express the situation like this: Given any number p , there can be found a number N such that

$$\left|\frac{1}{2^i}\right| < 10^{1-p} \text{ for all } i > N.$$

When this is true, we say that the limit of $1/2^i$ is zero, or that $1/2^i$ converges to zero, as i increases without limit. This is written

$$\lim_{i \rightarrow \infty} \frac{1}{2^i} = 0.$$

There is, of course, no particular virtue in decimal representation; from the mathematical point of view all that is necessary is to say that

For every positive number k , there is a number N such that $\left|\frac{1}{2^i}\right| < k$ for all $i > N$.

Further examples are given (in which the functions are all given neatly by formulae) of both convergent and divergent sequences until after nearly five pages of discussion we reach (SMP, Book 4, chapter 43, p. 1312):

Definition. If for every positive number k there is a number N such that

$$i > N \Rightarrow |f(i) - c| < k,$$

the sequence $\{f(i)\}$ has the *limit* c , and we write

$$f(i) \rightarrow c \text{ as } i \rightarrow \infty, \text{ or } \lim_{i \rightarrow \infty} f(i) = c.$$

In particular, therefore, we find a spiral technique in use, introducing the concept contextually in specific examples, then returning to it at various times

in increasing stages of subtlety until a formal definition is reached and then briefly used at the end of the course. We may legitimately ask what concept image is produced by these tactics and whether the formal concept definition is thoroughly integrated into this image. On this matter to date we only have data collected from students arriving at an English university having taken a variety of different ‘advanced level’ courses in their last two years at school. In a questionnaire for mathematics students arriving at University, students were asked if they had met the concept of the limit of a sequence $s_1, s_2, \dots, s_n, \dots$ at school, either with a precise definition, informally or not at all. They were then all asked to find the limits of certain sequences, including

$$\lim_{n \rightarrow \infty} \left(1 + \frac{9}{10} + \frac{9}{100} + \dots + \frac{9}{10^n} \right).$$

After being asked to write down the definition of a limit (if they knew one), they were asked to say whether $0.\dot{9}$ (nought point nine recurring) was equal to 1, or just less, giving the reason for their answer. The details of their responses are discussed in Tall (1977a). A number contained potential conflicts. A mathematician might consider that the computation of the limit in (1) above was a similar problem to the question of $0.\dot{9}$. On the contrary, fourteen out of 36 students claimed

$$\lim_{n \rightarrow \infty} \left(1 + \frac{9}{10} + \frac{9}{100} + \dots + \frac{9}{10^n} \right) = 2 \text{ but } 0.\dot{9} < 1.$$

The responses to the last part were often couched in ‘infinitesimal’ terms:

“It is just less than one, because the difference between it and one is infinitely small.”

“Just less than one, because even at infinity the number though close to one is still not technically one.”

Clearly the two questions evoked different parts of the concept image of the limiting process. In a subsequent test the same students were asked to write the following decimals in fractions:

$$(2) \quad .25, .05, .3, .\dot{3} = .333\dots, .\dot{9} = .999\dots,$$

Thirteen of the fourteen who previously said $0.\dot{9} < 1$ now said $0.\dot{9} = 1$. Several now experienced actual cognitive conflict, illustrated in the following responses to the last two decimals in list (2).

A $0.\dot{3} = \frac{1}{3}$

$0.\dot{9} = 3 \times 0.\dot{3} = 3 \times \frac{1}{3} = \text{rubbish}$

B $0.333\dots = \frac{1}{3}$ no fraction [this answer was crossed out]

C $0.\dot{3} = \frac{1}{3}$

$0.\dot{9} = 1$ or (none exists)

- D $0.\dot{3} = \frac{1}{3}$
 $0.\dot{9} = 0.999$
 E $0.\dot{3} = \frac{1}{3}$
 $0.\dot{9} \approx 1$

Other potential conflict factors which occurred are discussed (though not using this terminology) in Schwarzenberger & Tall (1978) and Tall (1977b). One of great significance is that students often form a concept image of “ $s_n \rightarrow s$ ” to imply s_n approaches s , but never actually reaches there, as one student put it:

“ $s_n \rightarrow s$ means s_n gets close to s as n gets large, but does not actually reach s until infinity.”

Thus “point nine recurring” is not equal to one because the process of getting closer to one *goes on for ever* without ever being completed.

In forming the concept image it does not help when most of the examples of the limiting process consist of sequences given by regular formulae which never actually equal the limit. If a student has a concept image that does not allow s_n to equal s (because it gets ‘closer’ to s as n increases) then we conjecture that he may not absorb a concocted example when it is presented to him. When a small group of four such students were shown the example

$$s_n = \begin{cases} 0 & (n \text{ odd}) \\ \frac{1}{2n} & (n \text{ even}) \end{cases}$$

they insisted that it was not one sequence, but *two*. The even terms *tended* to one and the odd terms were *equal* to one. So this was not a genuine example of a sequence some of whose terms equalled the limit! When a student is given a formal concept definition, the concept definition image that it forms in his cognitive structure may be very weak. We have noted, for example, that students have great initial difficulties with the use of quantifiers “all” and “some” and the standard definitions of limits and continuity all can present problems to the student. His concept image makes it obvious to him that if s_n gets close to s then $1/s_n$ gets close to $1/s$ (provided the latter is not zero). But a weak understanding of the concept definition can make the formal proof of this result very hard for him. This is a typical phenomenon occurring with a strong concept image and a weak concept definition image that permeates the whole university study of analysis, especially when there are potential conflict factors between the two.

4. Limits of functions

Limits of functions of the form $\lim_{x \rightarrow a} f(x) = c$ are also approached in the sixth form in an intuitive manner, with an informal explanation, then at a later stage a more formal definition may be given. Often the limit process is first introduced

when differentiation is being discussed, in which case the concept image of a limit may include a mental picture of a chord tending to a tangent. Other approaches to the derivative may affect the concept image via magnification factor or instantaneous velocity. As with the limit of a sequence, the limit of a function is often considered as a dynamic process, where x approaches a , causing $f(x)$ to get close to c . Once again students may consider $f(x) \neq c$ as part of their concept image. On the other hand the formal concept definition may be given in an imprecise form. In the text of SMP book 4, the example is discussed of

$$\lim_{x \rightarrow 1} \frac{x^2 - 1}{x - 1} = 2$$

where $f(x) = (x^2 - 1)/(x - 1)$ is not defined for $x = 1$. The implication is that one does not consider the case $x = 1$, so it is deduced that (SMP, Book 4, Chapter 43, p. 1325):

For every positive number h , there is a number k , such that

$$|x - 1| < k \Rightarrow |f(x) - 2| < h.$$

From this the formal definition of limit is (SMP, Book 4, Chapter 43, p. 1325):

Definition. If for every positive number h there is a number k , such that

$$|x - a| < k \Rightarrow |f(x) - c| < h,$$

then we write

$$f(x) \rightarrow c \text{ as } x \rightarrow a, \text{ or } \lim_{x \rightarrow a} f(x) = c.$$

Explicitly the definition should at least specify $x \neq a$, so that it should read:

Given $h > 0$, there exists $k > 0$ such that

$$0 < |x - a| < k \Rightarrow |f(x) - c| < h.$$

Implicitly this may be part of the concept image. We would conjecture that the intuitive approach prior to the definition (if the latter is given at all) is often so strong that the feeling of the students is a dynamic one:

as x approaches a , so $f(x)$ approaches c

with a definite feeling of motion.

A questionnaire was given to first year students in which they were asked to explain what is meant by

$$\lim_{x \rightarrow 1} \left(\frac{x^3 - 1}{x - 1} \right) = 3.$$

On the next page of the questionnaire they were asked to write down a definition of $\lim_{x \rightarrow a} f(x) = c$ if they knew one.

Seventy students responded (each with an ‘A’ or ‘B’ grade in A-level mathematics). Those responding to the second question, the concept definition for “ $\lim_{x \rightarrow a} f(x) = c$ ” could be classified as follows:

	correct	incorrect
formal	4	14
dynamic	27	4

Only four gave a correct formal definition, of which one was topological and three were of the form given above.

The fourteen incorrect attempts at a formal definition included seven that were mixed up with other limit notions such as:

$$\text{As } x \rightarrow a, c - \varepsilon \leq f(x) \leq c + \varepsilon \text{ for all } n > n_0$$

or

$$|f(n) - f(n+1)| < \varepsilon \text{ for all } n > \text{given } N_0$$

and seven were incomplete or inaccurate, such as

$$|f(x) - c| < \varepsilon \text{ for all positive values of } \varepsilon \text{ with } x \text{ sufficiently close to } a.$$

Of the incomplete ones, two omitted the condition $x \neq a$.

The dynamic responses included typical replies such as

the value that $f(x)$ approaches as values of x are taken closer to a is c .

or

as x tends towards a , the value of $f(x)$ tends towards c .

Incorrect dynamic responses included

as x gets nearer and nearer to a , the value of the function = c ,

or versions mixed with other limit processes:

$$\lim_{x \rightarrow a} f(x) = \frac{f(x) - f(a)}{x - a} = c.$$

The earlier question requiring explanation of

$$\lim_{x \rightarrow 1} \left(\frac{x^3 - 1}{x - 1} \right) = 3$$

revealed that most of the students could make an attempt at an example even though they could not give a concept definition. Those not giving a concept definition responded to this example as follows:

	correct	incorrect
used a dynamic idea	11	1
other	6	3

Those using a dynamic idea essentially said that $(x^3-1)/(x-1)$ approaches 3 as x approaches 1, some of them dividing $x-1$ into x^3-1 and evaluating the limit as $\lim_{x \rightarrow 1} (x^2 + x + 1) = 3$. The other approaches included continuity ideas (“filling in the missing point in the graph”) and L’Hôpital’s rule,

$$\lim_{x \rightarrow 1} \left(\frac{x^3 - 1}{x - 1} \right) = \lim_{x \rightarrow 1} \left(\frac{3x^2}{1} \right) = 3.$$

Of interest was whether those giving a concept definition actually used this definition in the example. Three out of the four giving a correct formal definition applied the same method in the example, the fourth using a dynamic method. Meanwhile the 14 giving incorrect formal definitions split 10 dynamic (of which 2 were again incorrect) and 4 using the wrong formal definition. All the students giving a dynamic definition used the dynamic method with varying levels of precision. In total 54 out of 70 used a dynamic approach in the example.

Many of these used the words “approaches,” “gets close to,” “tends to” which lead to the possibility that the assumption that $f(x) \neq c$ is a potential conflict factor.

As a follow-up to this questionnaire, two years later when the students were in their final year, twenty-two were asked suddenly in a lecture, whether the following theorem is true or false:

Suppose as $x \rightarrow a$ then $f(x) \rightarrow b$
and as $y \rightarrow b$ then $g(y) \rightarrow c$
then it follows that
as $x \rightarrow a$, then $g(f(x)) \rightarrow c$.

The students were asked to write down ‘true’ or ‘false’ and responded twenty-one ‘true’ and one ‘false’. When pressed, those who said ‘true’ refused to change, asserting that the theorem is true.

It is false. The concept image of $y=f(x)$ getting close to b and $g(f(x)) = g(y)$ getting close to c is very powerful. The theorem would be true if $f(x) \neq b$ for x near a . But in the formal definition, which the students had handled for two years, we have “ $\lim_{x \rightarrow a} f(x) = b$ ” means

for all $\varepsilon > 0$, there exists $\delta > 0$ such that
 $0 < |x-a| < \delta$ implies $|f(x)-b| < \varepsilon$.

Here we may have $f(x) = b$ and then $g(f(x)) = g(b)$ need not be defined, and if it is we may have $g(b) \neq c$.

For instance if

$$f(x) = 0$$

$$g(y) = \begin{cases} 0 & (y \neq 0) \\ 1 & (y = 0) \end{cases}$$

then

$$\text{as } x \rightarrow 0, f(x) \rightarrow 0$$

$$\text{as } y \rightarrow 0, g(y) \rightarrow 0$$

$$\text{as } x \rightarrow 0, g(f(x)) = g(0) \rightarrow 1.$$

The one student who was concerned the theorem might be false was worried that $g(b)$ might not be defined. The remainder relied on their concept image of a limit, not the formal concept definition which they had handled in the course over the previous two years.

5. Continuous functions

This topic is truly the *bête noire* of analysis. The concept image derives initially from a variety of sources, for instance the colloquial usage of the term “continuous” in such phrases as

“the rail is continuously welded”

(meaning the track has no gaps)

or

“it rained continuously all day”

(meaning there were no breaks in the rainfall).

The initial use of the term ‘continuous function’ often implies a similar idea that the graph of the function may be drawn “continuously” so that it has no gaps. An incorrect popular description of topology as ‘rubber-sheet’ geometry, which allows stretching but no tearing, does nothing to dispel this illusion. (In fact topology does allow tearing as long as the tears are glued together again. For instance a square $ABCD$ in which AD and DC are glued together is homeomorphic to a surface made out of a square $WXYZ$ in which a double twist has been performed before glueing WX and ZY . There is no way that one of these surfaces can be transformed to the other in three dimensional space without tearing and reglueing! So the popular idea of continuity has a potential conflict in topology too.)

In the English sixth form the concept of continuity is often suppressed or dealt with only in passing. The more central topic is differentiation and, since a differentiable function is necessarily continuous, the latter topic is relegated to the background (although it is essential in certain proofs).

Interestingly enough, SMP does stress *discontinuity* in a mathematically non-standard manner in curve sketching. (SMP, Book 1, Chapter 5, p. 133):

Behaviour near a discontinuity. Often a *discontinuity*, or break in a graph, occurs because there is a value of x which would make the denominator of a rational form zero. Thus, in Example 4,

$f: x \rightarrow 1/(x-1)$ is not defined for $x = 1$, and the graph shows a discontinuity there.†

In a footnote this is modified to the standard definition (SMP, Book 1, Chapter 5, p. 133);

† It is customary, though not strictly correct, to say that the function is discontinuous at $x = 1$: in fact it is neither continuous nor discontinuous, because it is undefined. We could define $f(1) = 0$, or any other number, but we cannot choose this number so as to make $f(x)$ continuous.

The net effect on the concept image, however, is, almost certainly, a reinforcement of the intuitive idea that the graph has “no gaps” and may be drawn freely without lifting the pencil from the paper.

Once more the formal definition in SMP comes at the end of the course (SMP, Book 4, Chapter 43, p. 1326):

Redefining continuity. Exercise E gives a number of examples of discontinuities, and should have made it clear that $\lim_{x \rightarrow a} f(x)$ may be entirely different from $f(a)$, and that either may exist independently of the other. This, in fact, is the essence of a discontinuity, and we may provisionally redefine continuity as follows:

Definition. f is said to be continuous at a if $\lim_{x \rightarrow a} f(x) = f(a)$.

This can immediately be rewritten in ‘challenge and response’ form, from our knowledge of limits:

Theorem. f is continuous at a if and only if for every positive number h there is a number k such that

$$|x - a| < k \Rightarrow |f(x) - f(a)| < h.$$

The theorem follows at once from the definition of a limit.

In this case the value $x = a$ need not be excluded, for then $f(x) = f(a)$ and the theorem is true.

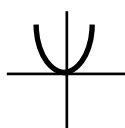
What concept images do students have of continuity when they arrive at university? Once again we only have evidence from students with a variety of backgrounds.

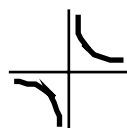
A questionnaire was administered to 41 students with an A or B grade in A-level mathematics. They were asked:

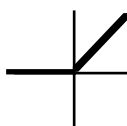
“Which of the following functions are continuous? If possible, give your reason for your answer.”

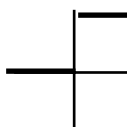
They were given five functions, as follows:

$$f_1(x) = x^2$$



$$f_2(x) = 1/x \quad (x \neq 0)$$


$$f_3(x) = \begin{cases} 0 & (x \leq 0) \\ x & (x \geq 0) \end{cases}$$


$$f_4(x) = \begin{cases} 0 & (x \leq 0) \\ 1 & (x \geq 0) \end{cases}$$


$$f_5(x) = \begin{cases} 0 & (x \text{ rational}) \\ 1 & (x \text{ irrational}) \end{cases}$$

(No picture was drawn in the last case).

The responses were:

	1	2	3	4	5
continuous	41	6	27	1	8
discontinuous	0	35	12	28	26
no answer	0	0	2	2	7

The first of these is to be expected, though the “right” answer may be given for the “wrong” reasons (where “right” and “wrong” are determined by the strict mathematical formulation). For instance it was pronounced continuous “because it was given by only one formula.” The second answer intimates that most students have a concept image which does not allow “gaps” in the picture. This is supported by the comments and, among those who thought f_2 was discontinuous, we have remarks such as

“the graph is not in one piece,”

“the function is not defined at the origin,”

“the function gets infinite at the origin.”

Even amongst those who asserted that f_2 was continuous we find concept images which do not coincide with the accepted formal definition:

“it is continuous because the function is given by a single formula.”

The third function was interesting. Most students who thought it continuous and made a comment made remarks like

“it is all in one piece.”

The twelve who thought it discontinuous gave several different reasons:

“it is not given by a single formula,”

“there is a sudden change in gradient.”

The fourth function was considered discontinuous by most, with several different kinds of reasoning:

“it is not in one piece,”

“there is a jump at the origin,”

“it is not a single formula.”

The last function caused more problems (seven students did not respond). For several it was discontinuous because

“it is impossible to draw,”

for one it was continuous because

“it has a continuous pattern of definition.”

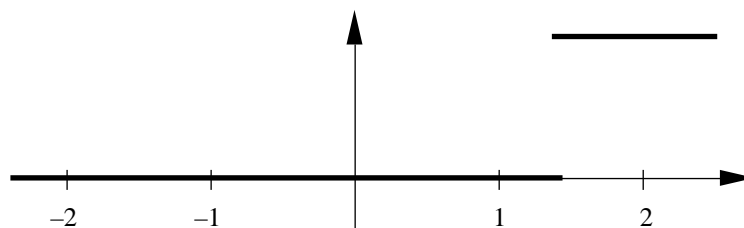
In these responses we see that most students evoked a concept image involving the graph having “no gaps,” being “all in one piece,” for a small number there is an evoked concept of “a single formula” and for a minority there are other images (such as “smoothly changing gradient”). All of these concept images just mentioned have potential conflict factors which conflict with the formal concept definition. We take the latter to be that $f:D \rightarrow \mathbf{R}$ is continuous at $a \in D$ if

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ such that } x \in D, |x-a| < \delta \Rightarrow |f(x)-f(a)| < \varepsilon.$$

Some potential conflicts are concerned with the nature of D . The mental picture most mathematicians have of the real line is “continuous” in the colloquial sense that it has no gaps. It is extremely difficult to form a satisfactory mental picture of the rational line instead of the real line. If we consider functions defined only on the rationals then we may have formally continuous functions which conflict with all the concept images mentioned above. For instance $f:\mathbf{Q} \rightarrow \mathbf{Q}$ (where \mathbf{Q} denotes the field of rationals) given by

$$f(x) = \begin{cases} 0 & (x < 0 \text{ or } x^2 < 2) \\ 1 & (x > 0 \text{ and } x^2 > 2) \end{cases}$$

has a graph like this:



This function is continuous but the graph has a “gap.”

Then again, the concept image is a *global* property, the notion of continuity over an *interval*, not continuity at a point. There are strange animals like

$$f(x) = \begin{cases} x & (x \text{ rational}) \\ 1-x & (x \text{ irrational}) \end{cases}$$

which are formally continuous at a single point (in this case at $\frac{1}{2}$) yet conflict with all three evoked images mentioned above. (It is not in one piece, nor given by a single formula, nor has it a smoothly varying graph.)

If a student is presented with one of these exotic creatures and it causes cognitive conflict with his concept image, then he may be in difficulties. When the teacher is aware of the possible concept images it may be possible to bring incorrect images to the surface and, by discussion, rationalise the problem. However, parts of the concept image may be very strong, especially the idea that the graph is all in one piece over real intervals. This latter mental picture is, in fact, correct. Even so, it can still cause problems in the formal theory.

In a mental picture of the intermediate value theorem and the extreme value theorem, the results are patently obvious, but the proofs of those results using the concept definition are more subtle. Without entering too deeply into another area, we briefly mention that current investigations are revealing the extent to which students have problems with the meaning of “all” and “some” and the manipulation of quantifiers. This means that many students have great difficulty (as is well known) with manipulating the definitions of limits and continuity. They are then in the situation where they may have a strong mental picture yet the concept definition image is weak. They understand the statements of theorems as being obvious, but cannot follow the proofs.

In such a situation, students who do succeed become very suspicious of ideas in analysis. At more advanced levels it becomes far more difficult to visualise the concepts as mental pictures and they can never be sure of the intuitions suggested by their concept image which may be now a mixture of strong concept images having potential conflicts with the concept definition. For instance, how does one visualise a function which is continuous everywhere but differentiable nowhere? Or what mental picture can one have of a function f whose derivatives f' exists yet f' may not be continuous? The mental pictures which served the students well at an earlier stage may now become an impediment. Bruner suggested that iconic processing limited ideas and urged a movement onto the symbolic level. But the student, saddled with his inadequate concept image, may find such a development difficult to achieve. In these and other ways, the difficulty of forming an appropriate concept image, and the coercive effects of an inappropriate one having potential conflicts, can seriously hinder the development of the formal theory in the mind of the individual student.

References

- School Mathematics Project: 1967 (revised 1970). *Advanced Mathematics* (Metric), Books 1–4, Cambridge University Press.
- Schwarzenberger, R. L. E. and Tall, D. O.: 1978. Conflict in the learning of real numbers and limits, *Mathematics Teaching* 82, 44–9.
- Tall, D. O.: 1977a. *Cognitive conflict and the learning of mathematics*, paper presented to the International Group for the Psychology of Mathematics Education, Utrecht, Holland.
- Tall, D. O.: 1977b. Conflicts and catastrophes in the learning of mathematics, *Mathematical Education for Teaching*, 24, 2–18.