JOANNA MAMONA-DOWNS

LETTING THE INTUITIVE BEAR ON THE FORMAL; A DIDACTICAL APPROACH FOR THE UNDERSTANDING OF THE LIMIT OF A SEQUENCE

ABSTRACT. This theoretical paper provides: (1) a presentation of some tasks that may be regarded as typical sources for forming students' intuitions and first understandings about limiting processes of real sequences, (2) an analysis of the formal definition of limit via identifying roles for each symbol that occurs to achieve a mental image firmly consonant with the definition, and (3) a description of how this mental imagery may be used to re-examine the validity of some intuitive beliefs. In particular a persistent issue found in (1) is that the sources encourage an intuitive image of a sequence as having an ultimate term associated with the limit; it is this belief that is mostly discussed in (3).

KEY WORDS: dynamic graph image for limit, extracting meaning from formal definition, formative sources for intuition, limit, sequences

1. PROLOGUE

Research in Mathematics Education is very consistent in saying that most students have extreme difficulties to obtain a reliable and robust image of the limit concept, but that it is very hard to find solutions to this situation, see especially Vinner (1991).

The problem lies largely in the surprisingly rich intuitive base that many students seem naturally to be endowed with in the theme. We say surprisingly because infinity is something never directly experienced by the senses in the physical world. However, there appears to be a common part of the human psyche that directs the human brain to contemplate infinity, in particular the unbounded universe and the infinitely small. This carries through to the notion of limit. Historically, the early developers of Calculus were able to refine their intuitive ideas about limiting processes into methodology of great efficiency and accuracy. This suggests that it may be useful to provide a forum that would strengthen (in terms of mathematical validity) students' naïve ideas about sequences and limits. But we cannot expect the students to fully appreciate the motivations that drove the mathematicians in history, and hence their intuitions will remain too vague, self-contradictory and dispersed to constitute any reliable image.

Educational Studies in Mathematics **48:** 259–288, 2001.

Hence we must turn to the formal definition to extract such an image. But a mental image deals with the cognitive. In general, formal definitions, although they may have aspects that impose on intuition, are also quite 'democratic' in their minimalistic style of expression. In a way, they maximize the ways in which their content may be understood. The problem is, for the student, twofold. One is the minimalistic style of the expression, which offers cognitive breadth but at the same time requires quite a mature reflection on structure, see Mamona-Downs and Downs (2000). The second is the content of the definition, which may well have meta-cognitive (meaning here concerns like logical tightness, brevity, ease of application, etc.) as well as cognitive aspects in its design. Both elements of this problem are very much evident in the definition of limit.

A further problem is that even if we do manage to obtain a useful image of limit from the definition, would this necessarily dominate, or could be induced to dominate, naïve ideas? To achieve this we first need compatibility between the two.

The aim of this paper is to present a didactical sequence that may solve, or at least alleviate, the problems mentioned above for the concept of limit of sequence. The basic idea is to allow students to mature their intuitions to a degree in a class-discussion environment and let these intuitions help in understanding the formal definition. We consider the cognitive and meta-cognitive load, required of the students in order to create their own images from the formal statement, too demanding. Because of this we introduce didactical measures to help them to accommodate the concept definition in their mind even at the risk of some loss of the freedom that the formal expression avails. Our first strategy is to identify a central core of the definition, and then build up by stages from there. How this is done already constitutes a restriction, but not a severe one, on how to think about the concept. Beyond this, though, we suggest also to didactically introduce a standard representation that imposes more or less a single, unified image (the well-known graph model of limiting properties of sequences). Bringing in this representation does not deny others, but it is true that we wish it to be regarded as a privileged one. Because of the way that all the content of the definition is embedded into the model, and yet the context of the model is both cognitively appealing and potentially may accommodate all sequences, the image extracted is both robust and can be readily compared against any informal argumentation.

Hence from what we have said above, we are suggesting for the introductory teaching of the concept of limit the following three didactical steps:

(i) Initiating and developing intuition through raising issues in a classroom discussion environment.
(ii) Introducing the formal definition and to analyze it in tandem with the issues in (i). Introducing a particular representation.
(iii) To endorse or revoke opinions made in step (i) by comparison with the formal definition, especially via the representation in (ii).

The argument of the paper is not strictly didactically driven, though it is always relevant to the aims of the didactical framework. We treat all stages (i), (ii), (iii) above in separate sections, but in each section we stress expected student behavior rather than describing particular pedagogical strategies that might be used. Thus our arguments would be directed either to inform teachers what to expect, or what to include in their class plan as an aim and why.

The arguments used represent the author's personal opinions; they are usually consistent with results found in the literature, but quite often the perspectives, within which they are discussed differ. (For example, the tasks used to raise an issue in stage (i) might be the same tasks that researchers could use to examine supposed pre-existing beliefs.) Because of this we employ relatively few references in our main discourse; however some of the relevant literature is briefly reviewed in the next section.

Finally, another motive to adopt the perspective that has been taken by this paper is that we believe it to be a very suitable platform to explain in the most explicit way the problem of the clash between the 'dynamic approaching' intuitive image and the 'static' image evinced by the definition of limit. This is an important theme in the article. In particular, our didactical approach here should go far to address the issue put by Khinchin (1968): "This outwardly static character of the concept of limit, as it appears in modern mathematics, often gives rise to the objection that, having frozen the movement out of the idea of limit, it thereby creates a tendency to dissociate the mathematical concept of limit from the living reality whose reflection and abstraction it is supposed to represent. This criticism is *essentially* invalid, because the modern definition at no point contradicts the earlier one, but merely refines it, and therefore cannot have a different content. More important for us, however, is that from the pedagogical point of view, the objection raised in this criticism deserves every attention."

2. Synopsis of the Research

As some research papers on limits and limiting processes not only deal with real sequences but also real functions and continuity, this overview will cover both themes.

2.1. *Prerequisite notions*

The limit concept has various prerequisite notions. One of the most important is that of the real number. Non-standard models exist, which allow for intuitive ideas like infinitesimals and infinite numbers to be expressed theoretically. As typically students have not been given formal expositions about the reals previously, their intuitive beliefs are very wide on this topic. See for example Fischbein, Tirosh and Hess (1979). Some authors (Tall, 1980a, 1980b; Sullivan, 1976), have discussed the educational implications if non-standard models are introduced. On the other hand, textbooks like (Gardiner, 1982) are available to help students to obtain a firmer image of the reals before starting conventional Real Analysis courses. The belief that infinitesimals exist would seem very influential in how limits are thought of, but one recent paper (Szydlik, 2000) claims to have some evidence against this.

Another prerequisite notion is that of function. Sequences are rarely considered by students as functions on $\mathbb{N}$ (Mamona, 1990), and hence tend to be regarded as processes rather than mathematical objects. Also continuity of real functions in a graphical context tends to be seen in terms of geometric curves and the connection between the x- and the y-coordinates may be obscured (Cottrill et al., 1996).

The statements of the formal definitions of limit also assume some previous skills. One is to cope with logical quantifiers, which trouble most students (Dubinsky et al., 1988). Also the application and/or understanding of absolute values and inequalities cause problems to quite a few students (see again Cottrill et al., 1996).

2.2. *Research on students' beliefs and behavior with respect to limits*

2.2.1. *Limit of a sequence or series*

In Davis and Vinner (1986), a test revealed a typical list of misconceptions that we summarize.

1. A sequence 'must not reach its limit'
2. Implicit monotonicity for a_n
3. Confusing limit with bound
4. Assuming that the sequence has a 'last term'

5. Assuming that you somehow can 'go through infinitely many terms' of the sequence
6. Confusing $f(x_0)$ with

   $\lim_{x \to x_0} f(x)$

   (of significance to sequences in the common practice of exploring the limiting behavior of a function by evaluating function values only for a discrete subset of its domain).
7. Assuming that sequences must have some obvious, consistent pattern.
8. Neglect of the important role of the order in which ε and N appear as you read the definition from the left.
9. (Hypothesized) confusion between the fact that n does not reach infinity and the question of whether a_n may possibly 'reach' the limit L.

Some of these points may be influenced by not regarding a sequence as a function. For example the sequence (a_n) given by $a_n = 0$ if n is odd and $a_n = 1/2n$ if n is even, is often seen as being two separate sequences (Tall and Vinner, 1981). Some behavior that shows restricted understanding, such as assuming monotonicity, may be refined to obtain logically equivalent definitions to the standard ones for limits, see for example (Burenkov and Terarykova, 1996). In the same spirit, Sierpinska's acts of understanding for the limit concept (Sierpinska, 1990), suggest that some beliefs are best attributed to a position within a schema, which not only contains the limit concept but also various other related notions (like Cauchy sequences). Another factor is how students retain the formal definition in their minds; a study by Pinto (1996), suggests that some students use memory, whilst others reconstruct it from a graph image. (This same image will be described and discussed in Sections 4 and 5.)

2.2.2. *Limits of a function, and continuity*

The following characterization, quoted from Williams (1991), usefully summarizes the different intuitive approaches or descriptions given by students on limits of functions:

Approach type	*'Typical' Informal Statement*
Dynamic-Theoretical	A limit describes how a function moves as x moves to a certain point.
Boundary	A limit is a number or point beyond which a function cannot go.
Formal	A limit is a number, to which the y-values of a function can be made arbitrarily close by restricting x-values.
Unreachable	A limit is a number or point the function gets close to but never reaches.
Approximation	A limit is an approximation that can be made as accurate as you wish.
Dynamic-Practical	A limit is determined by plugging in numbers closer and closer to a given number until the limit is reached.

Although many papers refer to the plurality of the approaches that individual students are prepared to accept, most papers suggest that cognitively the strongest images are the dynamic ones. For example, in Tall and Vinner (1981) dynamic informal arguments were set against those based on the formal definition. Such considerations may (for example) "lead to the possibility that the assumption that $f(x) \neq c$ (the limit of f at x) is a potential conflict factor." In Cottrill et al. (1996), fieldwork was analyzed from the APOS perspective; here, naïve intuitions based on movement towards the point where the limit is to be taken play an important part. In particular this situation often seems to cognitively sever the relationship represented by the function, so that one imagines two separate processes, one along the x-axis, another on the graph curve, which then have to be coordinated (again).

Passing on to continuity, there is a danger that the students' intuitions will entangle in a seriously circuitous way. A graphic image of continuity over an interval is used to understand limits at a point. Then continuity at a point is used to define continuity over an interval. Indeed, in Tall and Vinner (1981) it is stated that the topic of continuous functions "is truly the bête noire of analysis." The study in Vinner (1987) shows that many students can confuse continuity with other notions such as domain of a function, graph curve and differentiability. Finally Mamona (1987), investigates the extent for which students' concept images of the notion of continuity depend on text exposition, teacher presentation or linguistic considerations. A more thorough review may be found in Cornu (1991). Most relevant literature predates this work.

3. Tasks to Bring up Issues on the Intuitive Level

In this section we present some tasks that we use to illustrate possible mental responses towards processes involving limits set in more or less familiar (or easily conceived) situations. The rationale is that these tasks will allow us to (conjecturally) argue about students' intuitive thoughts on this theme from a particularly formative perspective. It should be stressed that the ensuing commentaries do not relate explicitly to any actual fieldwork. Everything should be read hypothetically, in this kind of way: if a student was confronted with this, then *possibly* he / she would react thus. . .

We wish to simulate with some degree of plausibility the way students would argue, especially in a social context, about sequences and their limits as an issue for which there is not necessarily any preconception of such objects. The focus of the discussion will be on the given tasks, which will have concrete contexts (either physically or geometrically based) in which explicit questions are posed. The role of the questions is to inspire debate on the issue, rather than inviting certain answers. (Should the class become too secure on a particular answer, the teacher could intervene.) The selection or design of the tasks then becomes important. This is not because we feel that the tasks used can be chosen to obtain an optimal response in terms of making the students' intuitions more aligned with the definition. The selection is important to encourage the thinking about the issue to be as broad and deep as possible. To achieve this we have to respect some fundamental cognitive strands, otherwise a momentum will not be generated. In particular, to some extent we have to anticipate a certain principle, which in the past has been looked at from slightly different perspectives but has been recently discussed in a linguistic framework by Lakoff and Nunez (2000) as the Basic Metaphor of Infinity. This principle means each task should imply a process that evolves in an ordered string of stages that never ends, but at the same time some sort of conceivable completion of this process is suggested. It is the status of this completion, and how to characterize this status, which will sustain the issue. For an example, if you consider a ball bouncing on a hard floor under gravity, the completion could be considered as the state when the ball is at rest on the floor. Various simple infinite processes can be considered in this system. This indeed will form the context that will concern most of the discussion of this section.

Task I. A ping-pong ball is dropped from a height h onto a level, hard floor. Each time it bounces, the highest height the ball attains is half that it attained for the previous bounce. (Height of the ping-pong ball is always measured relative to the 'lowest' point of the ball from the floor.)

First question: How many times does the ball bounce? If the word 'infinity' occurs in the answer, we ask a *supplementary question* what do we mean by infinity here?

Hypothetical responses: One response might be: the ball bounces a certain finite number of times (which however cannot be determined). There are two distinct possible trains of thought here. The first is based merely on personal experience that by observation the ball always *seems* to come to rest after a finite number of bounces. The issues related to this belief (that could be generated in coming to or in reaction from this position) may be how accurately can you sense motion; how can you extrapolate your thoughts about what you can see to what you cannot; what kind of physical properties may operate at the extreme small scale. In these issues, sensory and physical considerations precede the purely mathematical rule that is intended to dominate the system. This situation really involves the general 'didactic contract', a term coined by Brousseau (1986, 1989). The contract is failing if the students do not see in the question's context an idealized situation to be treated as a completely mathematical framework, which is somehow 'friendly' because it resembles some physical system. Appealing to physical evidence, should it happen, is not helpful and the intervention of the teacher is needed. However the second way of thinking that also leads to the belief that the ball bounces a finite (but large) number of times is more significant, as it does take account of the mathematical rule implied in the question. The argument is centered about the following issue; is it possible to halve a length an indefinite number of times? There are various reasons why a student might regard it as impossible. We suggest a few. First a student might have a literal atomic view, learnt from physics, which (s)he wants to import into his / her model of the real numbers. Here an interval of some explicit (but short) length cannot be divided into two. Second, a student might trace in his / her mind the process of how an interval becomes successively shorter each time it is divided into two, and hence its length eventually becomes less than any familiar positive number. It may be plausible, then, for the student to believe that for a very great number of divisions (not infinite), the length of the resulting interval will be smaller than all 'normal' positive numbers. This interval has width a positive number but not with any particular name. It meets with descriptions that students often use like ‘arbitrarily small number’, ‘infinitely small number’, ‘number as small as we may choose’ or ‘infinitesimal’ (used informally, of course). All these names are somewhat suggestive of the (self-contradictory) notion of smallest positive number. If indeed this notion is cognitively associated with the interval, further halving would bring an inconsistency in meaning; the most natural way to overcome this problem would be to suppose that further division must force you to zero. Third, a counter argument might be attempted to try to persuade that there is a problem with indefinite division. For instance, a student might explain to his / her fellows that as the ball falls from height h to the ground (height = 0) it must pass all heights $h / 2^n$ ($n \in \mathbb{N}$), and each section of the fall between $h / 2^{n-1}$ and $h / 2^n$ must occupy some time, which would then accumulate without bound over all the sections if n was allowed to vary right through $\mathbb{N}$. This fallacious argument will be closely mirrored by one presented later in more detail. (All the arguments above crucially concern images of the real number system; the latter will be dealt with in other papers in the volume. The overlap of content is unavoidable, as the concept of limit of real sequences depends fundamentally on the concept of distance in the reals. Historically the ‘debate’ between indivisible atoms and infinite divisibility is fairly well captured by Zeno’s paradoxes, two of which are aimed at showing the implausibility of atomism, the other two the implausibility of indefinite division, see for example Boyer (1939). There are formal infinitesimals in the pseudonumbers that are less than all real positive numbers yet are not zero; however these are not reals, so there is no contradiction in dividing these into two (say), see Davis and Hersh (1981). However surely the exact status of formal infinitesimals is unlikely to be anticipated by students without didactical interposition.)

The assignment between the bounces and the heights is very significant as it forms the basis to leave the physical context and just to consider a sequence of numbers. (The recognition of this assignment is not to be at all taken for granted). Just as it has allowed the issue of indefinite divisibility of intervals to be brought in, it equally admits the following argument, which runs counter to those found in the previous paragraph. Here the sequence is taken as an *ordered list* of numbers for which one number may be obtained from the previous number. To generate the whole sequence, we have to repeat an algebraic operation. We might imagine ourselves having to do the operation ‘manually’. Whenever there seems to be a sense of a process never to end, there is an expectancy that infinity is involved. In our case, each operation corresponds to a bounce, and hence the expectancy carries through that the number of bounces is infinite. But in what meaning do we take infinity here?

It is quite plausible that a student might think on the following lines. Suppose that hypothetically you count the bounces of the ball as far as you are willing to. (We say hypothetically, because physically there would be bounds to where you can reach by counting, as the bounces are progressively getting smaller and more frequent). You understand that there are an infinite number of bounces, but it is natural to feel that your activity is bringing you closer to an ultimate ending somehow doing with this infinity. This motivates the following belief; infinity is the number that you get if you count forever. Note that infinity, thus, is thought of as a *number;* the greatest natural integer. This could lead you in the end to consider an infinite sequence (a_n) to possess a final term a_∞. A value for a_∞ may be designated if there seems to be a sensible and consistent tendency in the values of a_n to become closer and closer as n increases to a particular number. So, in our case, where it is intuitively clear that the heights of the bounces become arbitrarily small, we have $a_\infty = 0$. (The notion of this last term of a sequence surely will be closely associated with the idea of the limit when the latter notion is conceived of, so the limit tends to be considered as an integral part of the sequence. This theme will be met again in this paper. Also, the above argument is consonant with the theory of embodied mathematics as espoused by Lakoff and Nunez (2000). Here it is claimed that cognitively speaking the only way that the human mind can comprehend infinity in its various manifestations in mathematics is through an analogy to the finite, in particular to a finite succession of repeated actions. The understanding that a finite succession of actions will have a unique, final action implies that an infinite succession should share these properties; this is called the 'basic metaphor of infinity' in the theory. Here we could note a similarity to what is known as the 'Leibniz's principle', which we shall refer to again later. The supposed last term of a sequence constitutes a particular case of the 'basic metaphor of infinity': this means that the last term of a sequence does have a legitimate status cognitively, if not strictly mathematically.)

It is a common observation that students do not naturally think in terms of sets and functions (especially when these are not explicitly presented as such). Hence it is not usually intuitive to regard (countable) infinity simply as a *property that a set might have*, rather than some meta-number. (Informal terms often used such as 'an infinite number' do not help.) However if the set of natural integers is not well understood, certainly it follows that the concept of an infinite sequence is also not understood. In the fieldwork of the relevant literature, it is often very clear when students are treating sequences and series as on-going processes that will never end, as opposed to when students regard sequences and series as entities via sets and functions. The two groups of students will evidently have marked differences in their opinions and beliefs. We mention just one obvious difference here; the students working on sequences as on-going processes are much more dependent on a rule that generates the terms.

Second question. How far will the ball travel in total?

We suppose for this question that students accept that the ball bounces infinitely many times. Then it seems most reasonable to assume that the journey must be infinite both in distance and time. However, when we realize that although we have an 'infinite number' of bounces these bounces are becoming successively smaller (as are the time intervals between consecutive bounces), can we be sure? Especially as the physical evidence is very convincing that the ball *does* stop. These issues surely would raise a lively debate in any pre-Calculus class!

Here we shall explicitly write down a simulated class – generated debate, in outline. The camp C_1 advocating the proposition that the ball travels an infinite distance might claim that all the bounces reach some (positive) height and hence there is some very small positive height that all the bounces will surpass. Challenged about the identity of this very small lower bound, it may well be rephrased as an (informal) infinitesimal, which will simply raise the issue what is the sum of an infinite set of numbers all greater than or equal to an infinitesimal, an issue with no firm basis to argue.

The opposite camp C_2 might argue by experiment with the given law, by calculating the distance the ball has traveled in total after a few bounces, and to notice how the increment decreases for each time and the totals seem to be heading for a particular finite number. This does not convince the other camp; you still have to add on an infinite number of terms, which will add up to infinity. The master-stroke, though, is done by C_2 by producing a general term b_n that describes (for whatever natural integer n you might take) the distance that the ball has covered before the n^{th} bounce. C_1 is at first not impressed, and counter as before; it does not matter which n you take, you still have to add an infinite number of terms to b_n, so you have shown nothing. C_2 explains that, on the contrary, it has clinched the whole issue. Why? Because we can easily demonstrate to you that for *all* values of n there is a finite number, B say, for which b_n is less than B. This means that it is impossible for the ball to travel more than the distance B.

The preceding exchanges in our simulated school-class debate are interesting in various ways, particularly as they suggest that some of the intricate issues for which the formal definitions are designed to cover can be realized quite plausibly in casual discourse. (Here, of course, we are not claiming that students, even when prompted carefully, will ever be able to formulate the definitions of limits for themselves. This is quite another thing!) We see that C_2 take the partial sums to convert the first few terms of the (implicit) series involved in the second question into the first few terms of an (implicit) sequence. They discover that they can find any term in the sequence via a general term, which allows a switch from a process to objectification of the sequence. Now they can treat *all* terms together, and when they discover a global upper bound, their claim is justified. Here we see, then, half-formed ideas of the roles of quantifiers, variables and inequalities in finding limits, as well as the vague notion that infinity can only be investigated in terms of the finite.

The second question is not settled yet; we have only decided that the distance traveled will be finite, i.e. must be represented by some real number D. What is this number? The upper bound B that C_2 constructs very likely would also be the least upper bound of the partial sums, therefore B would automatically be considered as the natural candidate. C_2 might adopt a similar argument as before to promote the position that D = B; first to show experimental evidence and finally to explicitly show that you can always find some partial sum such that its distance from B is less than any given (small) positive real number. However, now the issue revolves around accepting or not an explicit number for an answer. C_1 can now avail itself with the idea of an infinitesimal as a number less than all positive real numbers, yet strictly greater than zero. With this C_1 can defend an argument that D is an 'iota' less than B rather than D=B, motivated by an image inspired by potential infinity rather than actual infinity. As Robinson's Non-Standard Analysis (1966) shows, both stances are mathematically acceptable. However perhaps the position of an infinitesimal difference is the more vulnerable to informal linguistic argument. For example, the seeking for a specific number of D (we contend) would favor a potential infinity image, whereas D representing a particular contextual meaning (i.e. total distance) would favor a role of D parallel to B. If a student were to marry the two roles in his mind, we would obtain, according to the argument above, that D is an 'iota' less than itself.

Our simulated debate above corresponds fairly well with the actual classroom community activities reported by Sierpinska (1987). More specifically Sierpinska here states four different cognitive 'models' supporting the belief that all sequences are finite and another four supporting the opposite, i.e. sequences are infinite, yet each encouraging epistemological obstacles related to limits. Two of the latter models fit in with the behavior shown by camp C_1 above; another deals with the belief of a final term of an infinite sequence that we also address; the remaining one concerns bounds, with some relevance to our presentation.

In tasks employed by the literature, a fairly common feature that occurs is that as well as the expected implicit sequence or series there is also an extra focus available in the task environment (independent of the infinite process) which provides a natural candidate for the limit. Examples may be found in Binmore (1977, pp. 26–27), Nunez (1994), Williams (1991) [session 2-problem 2], and Fischbein, Tirosh, and Hess (1979) [problem 5]. The well-studied problem, 'is 0.999... = 1?', may be regarded in the same spirit. The task found in Binmore (1977) is a particularly vivid case: here two bulldozers approach each other so that it is known exactly when and where they will collide, whilst the infinite process is given by a fly toing and froing between them. However if the speed of the fly is known relative to the speed of approach of the two bulldozers, then the distance that the fly travels before its demise may be deduced without reference to the infinite process. Such tasks would give an extra 'handle' in any class debate and in particular would give persuasive naturalistic evidence that an infinite series can indeed have a finite limit. But a danger lies in this too, in that the situation may neglect the consideration of the finite sums; instead a sum of infinitely many numbers may be endorsed as a meaningful structure in itself, because of the apparent availability of a numerical value for it.

Another type of task employed by the literature are those that can be characterized by a context involving an infinite family of objects (e.g. geometric figures) from which a real sequence may be extracted. Examples of the type may be found in Mamona (1987), Artigue (1991), Nunez (1993) and Fischbein, Tirosh, and Hess (1979). We shall consider a task adapted from Orton (1980), in order to illustrate how the presence of the family of objects may influence the students' understanding of the limit of real sequences. However for reasons that will become evident in due course, we consider this particular task unhelpful for a program aimed at prompting students to generate issues relating to limits.

Task II: Imagine a stairway with just two steps, with rise and tread both 1 meter. From the original stairway we construct another with twice the number of steps, by halving the rise and tread. Following the same process inductively, we may construct a whole family of staircases. What can we say about the perimeter of the staircases? What is the final result of the inductive process?

Here we have two infinite processes operating, one may be modeled as a sequence of geometric objects (zigzag curves) and the other is a sequence of reals obtained through a 'measure' of these objects. It should be noted straight away that at this formative level it might be just as natural for students to develop intuitions for a limiting process on a sequence of objects as for a sequence in the real numbers.

Let us start with the real sequence given by the perimeters of the staircases. The situation is rather similar to the bouncing ball in Task I in that the inductive process for the staircases means that the value of each term of the real sequence is understood in terms of the previous state for the representing zigzag curves. This may induce an on-going process psychology, that might lead to beliefs such as the sequence has many terms but is finite, or the sequence is infinite but some of it is so removed that it can never be reached, perhaps with a belief in an ultimate term at an 'infinite position'.

Another aspect here is that the inductive process on the objects produces a constant sequence for the perimeter. A constant sequence is perhaps the easiest kind of sequence to encapsulate as a mathematical entity, but students may treat such a sequence with suspicion. What is the point in having an infinite sequence that simply repeats a value endlessly? Also for this case what is the point in thinking about limits? Students may see an issue in trying to describe a behavior when the terms are perceived to be settling down overall to a particular value, but this becomes a non-issue if all the terms are the same, and do not meet the usual image of having to 'move towards' the limit. At least the present example goes some way to justify the need for the constant sequence, because its terms are formed in a 1:1 correspondence with the elements of a (non-constant) sequence of objects.

Let us now briefly talk about intuitions concerning the limiting objects, but we shall mostly restrict ourselves only to aspects affecting also the intuition on the corresponding real sequence. We contend that if a student does conceive of a limit object, it is likely to be one out of two forms; the first a staircase with an 'infinite number' of steps of 'infinitesimal size', the second a ramp of constant slope. (Which one is selected certainly would be highly dependent on the wording used in the question. For example, as Orton did in his original version, we have used the phrase 'final result', which certainly would bias the issue towards a potential infinity image and the staircase with 'infinitesimal steps', as previously observed in Mamona (1987), Cornu (1991) and Tall (1992).) However, we shall not discuss the arguments by which these results would be obtained or justified by the student. We only wish to note the following two points. One is that the staircases may have, for the student, a much more concrete identity than numbers, and this might affect the way that the student feels about the sequence of the objects. Even if the phrase ‘final result’ (and similar) may be avoided in the question, this may mean that the identification of a limit object will be associated with a notion of last ‘item’ in the sequence. This belief could be transferred to the real sequence given by the perimeters via the 1:1 correspondence. Our second note concerns more the form of the perceived final or limit object. The infinitesimal staircase is a construct that is not supportable in standard analysis, but its cognitive pull is well explained by the Leibniz principle which reflects the bias that the limiting object should have the same properties as the objects in the sequence. According to standard analysis, the ‘correct answer’ is the ramp of slope 1. But here we reach a potential conflict factor. The real sequence formed from the staircases turns out to be a constant one with constant term 4, whereas the perimeter of the ramp is $2+\sqrt{2}$. Hence the natural expectation that the limit object should yield the limiting value for the related real sequence is not realized in this case. Such behavior is likely to seem paradoxical to many students, and it may shake confidence in naïve arguments made in paradigms like the bulldozer task previously discussed.

It is likely that the seemingly contradictory argumentation above will not be conducive to a fruitful classroom debate, but would only bring confusion. Though generally the author agrees with the position that conflicts should be met rather than avoided, she does not support applying it in the particular environment where the students are developing their first ideas about concepts as yet lacking definite form. Hence the above task, and other tasks containing similar apparent paradoxes, are probably best left out at this preliminary stage. However there are plenty of other tasks that employ infinite sequence of objects, and which do not display this type of ‘idiosyncrasy’. One obvious example is how the area of a circle is calculated from regular polygons. You might try to start an issue about whether the argument is justified or not. Perhaps the discussion might lead to the Ancient Greek strategy of the Method of Exhaustion, see Boyer (1939).

The stairway question shows that the choice of tasks will influence the total cognitive intake. However it also acts as a warning that the issues may become very complicated. In a free discussion environment there are no absolute boundaries, so even an apparent innocuous task may eventually lead to a non-useful impasse.

Conclusion of the section

Many authors refer to the strong influence past physical and social experiences seem to have particularly on the concept of limit. For example Cornu

(1981, 1983) talks about 'spontaneous conceptions'; in Schwarzenberger and Tall (1978) the terms 'tend to' and 'limit' have been found to have connotations for the student before teaching the concept that can persist after instruction. Further, in Davis and Vinner (1986), it is suggested that there exists a rich substratum of naïve ideas that compete with the official ones presented in class; during a course studied in the latter paper the official forms were followed and comprehended, yet at an unannounced test asking for informal descriptions, some time after the course was completed, the naïve ideas tended to dominate again. All of these suggest that the spontaneous conceptions constitute an uncontrollable force that cannot be completely side-stepped.

Instead of trying to deny the naïve ideas of the students, what we have tried to do in this section is to advocate their airing and discussion to a degree. We acknowledge fully that the full gamut of misconceptions, alternative conceptions, inconsistencies and vague images that have been described in the literature will remain. Some of these negative factors have been referred to in the section; others have not, such as restricted images encouraged by certain special properties held by the particular sequences used (e.g. the limit forms a bound; monotone behavior). A point not so well acknowledged in the literature is the following: because typically a task used will have a feature in its context that can be identified as the natural 'completion' of the infinite process involved, it may be difficult to separate the two concepts 'sequence' and 'limit'. This issue will be resumed in the final section.

What we feel justifies the pedagogical strategy described in this section (beyond the standard arguments of the social-constructivism tradition in Mathematics Education) is that within a class debate climate it is more than likely that the students will be exposed to the essential ideas that ultimately motivate the definition of limit. More specifically, we contend that the following two competing broad arguments would naturally form the heart of the discussion of the class (as suggested by our simulated class dispute above). Either the attention is put on the infinite process and the student ponders how this behaves towards the completion, or the attention is put on the completion where some expectations relating to the completion are formulated. This may be considered as the genesis of the opposing views of the limit, the dynamic and the static, a theme much discussed in the literature. The exposure to both sides of this debate, we feel, will be more than helpful in order to obtain a cognitive appreciation of the definition when it is presented. Having a sound cognitive base for the definition then may endow it with a mental image more permanent than the one in place for the fieldwork in Davis and Vinner (1986), say.

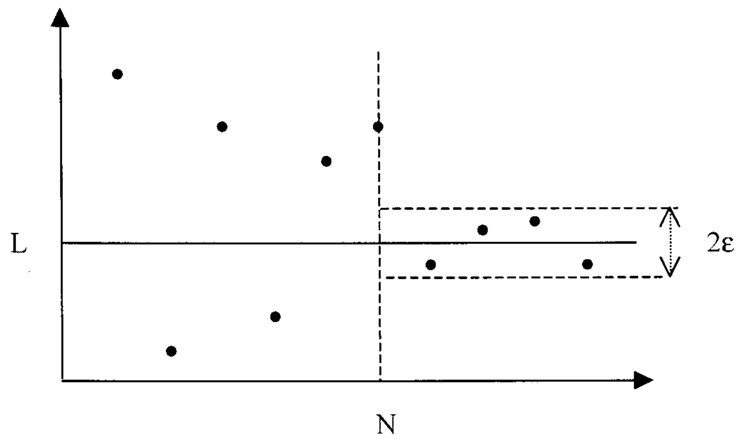
4. On the definition of limit of real sequence

4.1. *Introduction*

In this section we shall examine the (standard) definition concerning the limit of a real sequence. The aim is to reflect directly on the statement of the definition to obtain a mental image that is robustly coherent with the definition. Two factors are involved here. One is to identify a central core of the definition, and then build up by stages from there. The second is to bear existing intuition on the issue (perhaps collected in the context of the previous section), in order to obtain a 'reading' of the definition that is cognitively accessible. Much of the content of this section reflects the author's own philosophical deliberations, rather than representing any collective perspective drawn from the literature.

4.2. *The statement of the definition of the limit of a real sequence with an illustrating graph*

Let (a_n) be an infinite sequence of real numbers (indexed by $n \in \mathbb{N}$). The sequence (a_n) has the limit $L \in \mathbb{R}$ if and only if $\forall \varepsilon > 0\ \exists\ N \in \mathbb{N}$ s.t. $\forall n > N$, $|a_n - L| < \varepsilon$.



4.3. *A discussion on the immediate impact of the definition and its graphical illustration*

A first course in Calculus, as often presented in the last years in secondary education in many countries, can be usefully taught on the informal level, but these courses usually depend heavily on procedural methods while playing down the importance of sharpening the concepts and the theory underlying the methods. However, if we wish the students to be more flexible in their problem-solving, they must have a clear and consistent idea about the concepts involved. But we have seen in section 3 that a wide range of biases and misconceptions may potentially be present in students' intuitive thinking about limits. Probably for the first time in their mathematical experience, there is a need to work with completely formal definitions to clear up the fog created by uncertainties and intuitive beliefs. The definitions are not, however, necessarily designed to replace the original intuition; rather, they give a way to assess and indeed to enhance intuition. Therefore the definitions of limits have to be understood, but how?

When the student sees for the first few times the definition of a limit, he / she is simultaneously trying to make sense of the symbolism and how this meaning fits in with any preconceived informal images which he / she might have. The symbolism is almost bound to seem strange and daunting; although some very few students seem to have a natural facility to be able to digest such statements as given in the definition in a remarkably short time, most students need a considerable time to feel comfortable with the definition and quite a few are completely intimidated and never quite grasp it. If at last the definition is understood it is usually readily accepted as being an appropriate way of describing a limiting process. This is not to say, of course, that we could expect the students to be able to construct the definition (in more casual terms) on their own. Not at all! There is an aspect of the definition which needs a subtle shift of focus from the underlying discrete function $n \rightarrow a_n$ to the intervals $[L-\varepsilon, L+\varepsilon]$ centered about the proposed limit L. More on this will be said later. Once the shift is made, the form that the definition should take is clear; there just remains the problem to express it tightly in mathematical terminology. But although the shift may seem quite small once taken, making that little shift for the first time requires a measure of genius.

Let us now cast a closer look at the statement of the definition of the limit of a sequence, and to try to identify potential difficulties that a student might encounter.

4.3.1.

We identify the focal component of the right hand side of the statement of the definition as $|a_n - L| < \varepsilon$. All of the other information simply elaborates this inequality.

Hence there is a question about whether one should read the expression from the left or from the right. Usual reading practice (in Europe at least) would encourage the left to right order of interiorisation of the expression. Likely, mathematicians retain this practice of reading left to right but they mentally suspend information until they reach the core of the definition. This common feature in formal definitions where in a sense something is qualified before you know what that something is (or signifies) might be difficult for a student.

4.3.2. *How to understand the inequality* $|a_n - L| < \varepsilon$?

Here we are simply relating three real numbers given by the three symbols a_n, L and ε (supposing for now that n is fixed). Even on this level we cannot assume that all students would be comfortable with an expression which is both an inequality and involves absolute values. This, though, might be regarded as a failure of teaching at earlier years. It is a lack perhaps of both number and symbol sense if the students at the stage we are concerned with cannot immediately interpret the statement into something more intuitive like 'the distance of a_n from L is less than ε'. In fact it may be more palatable to them to replace the phrase $|a_n - L| < \varepsilon$ by $a_n \in (L-\varepsilon, L+\varepsilon)$, even though the latter is more set theoretical. (On the other hand, the inequality form has an advantage over the interval form in that it reflects better the notion of distance which is important in extending the theory to metric spaces.)

It is now important for the student to somewhat anticipate the character of the symbols a_n, L and ε. What potential of variability does each have? Well, L stands for the proposed value of the limit, so surely it has the character of a *constant*. The symbol a_n (now letting n to vary) stands for a term of the sequence we are investigating; we wish to check that a_n approaches L as n becomes large, so it is natural to want to regard a_n as a *variable*. Finally ε is primarily a device to measure the closeness of terms of the sequence (a_n) to L; clearly within the structure of the inequality ε has the role of a constant, but it must also carry with it a *potential* to vary otherwise we cannot utilize it to examine the behavior of the sequence as it gets arbitrarily close to its limit. Hence we may say that ε has a character similar to a *parameter*.

Also it is important to understand the role of the inequality in the definition. It is not a statement of truth, but it is the basis of a decision process. This issue is more to do with a general understanding of definitions than with our particular theme, but the complexity of the right hand side of the statement of our definition may obscure how the phrase 'if and only if' fits in. (There is a parallel here to the common phenomenon of students assuming a statement whilst proving that *same* statement.)

Already we see the significance of the past experience of the student on the intuitive level. It is needed for the student to respond to the teacher's explanations of the roles of the symbols in the inequality. Remember also that we have conjectured in the previous section that on the formative level it is difficult to separate the concept of sequence from the concept of limit, which may militate cognitively *against* the idea that the definition acts as a decision process.

4.3.3.

The remaining part of the right hand side of the expression of the definition simply qualifies which terms of (a_n) and which values of ε we should consider when focusing on the inequality. This makes exact the decision process based on the inequality.

Now a_n (as the variable) will vary whilst ε (the parameter) is set constant. Hence we must expect to consider the conditions on a_n first, and ε second. This is reflected by the qualifying term for a_n immediately preceding the inequality. This term introduces a new symbol N, which always denotes some positive integer. The appearance of the symbol N is probably confusing for many students, and makes the definition seem more difficult to grasp than it really is. There are two reasons for the students to feel uneasy about N. The first is that as N does not explicitly enter in the focal feature, i.e. the inequality, we are causing an independent secondary focus (within the description of the qualifications on the first focus) to explain N's role. The student must coordinate the two foci. The second reason why students might be disturbed by N is that they might be looking for an explicit indication of how the value of N is to be found. The student's previous experience in mathematics is likely to be of a largely algorithmic nature. It is difficult then for them at this stage to distinguish between a rule for a process and a definition.

It is interesting here that a little bit of verbalization may help the situation. N simply denotes a position in the sequence and as such N does not have to be explicitly mentioned in speech. Consider the following informal description: "given ε, there is a certain position beyond which all further terms of the sequence are within the distance ε from L". Such verbalizations should not replace the definition, but they can help tremendously to pin down the relative importance of the various different 'players' involved. In particular they tend to stress the more conceptual content of the definition rather than the practical content.

4.3.4.
Here we will consider the parameter ε. Of all the symbols that enter in the statement of the definition, ε is the only one that has no tangible connection with the sequence. As the limit of the sequence surely depends only on the properties of the sequence itself, what is this 'foreign' parameter doing in the definition? Well, we shall tackle this question later, when we reflect why the definition has the form it has. A second question, though, we shall tackle now: Given that the form of the definition is accepted, how should we interpret the role of ε?

In fact the role that ε plays has already been partly discussed; it provides us a way to analyze which terms of the sequence (a_n) lie within a prescribed distance from the proposed limit L. The variability of ε is needed to allow this analysis to be done for arbitrarily small distances. However there are two points that might cause uneasiness. The first point is that ε seems not to correspond to anything in particular, it is just a number which can be varied sitting self-contained within the statement. This may be difficult to accept. The second point is that the right hand side expression, when read right to left, ends with '*for all* $\varepsilon >$ 0'. This seems abrupt, and does not seem to fulfill our expectations. What we might have expected is that, loosely speaking, the definition to somehow guide ε to *decrease* in value and in such a way that ε becomes arbitrarily small. *This* is the way to check whether eventually all the terms of the sequence will become arbitrarily close to L. However, the definition does not seem to follow this line; why?

Well, the framework of the definition certainly gives us the freedom to choose any certain ε, so we can recapture the more dynamic image by testing the decision criterion for every value of ε for any particular sequence diminishing to zero for ε. However the definition also demands that all other ε values to be tested. This further checking is never needed: for any value, ε_0 say, of ε not lying in the sequence we can always find ε_1 smaller than ε_0 which *does* lie in the sequence; then the N used for ε_1 may equally be used for ε_0.

Hence we see that the formulation of the definition with the term '$\forall\varepsilon>0$' can be taken to be compatible to our intuitive feelings how ε should be used. But this reconciliation needs reflection on the form of the definition. This illustrates how definitions should not be regarded as mere formal statements but bases on which insight may be built. It shows also how what seems to have a very static character (i.e. the definition) can be manipulated mentally to yield a far more dynamic intuitive meaning. One reason not to attempt to introduce this more intuitive meaning directly in the definition is that surely this would make the statement much more clumsy. Another reason, perhaps, is that we should not introduce explicitly in the definition a *secondary* infinite process in ε when the whole point of the definition is to clear up what we mean by a limit of the primary sequence (a_n). There would be temptations to bear intuitive biases on how ε should behave in the limit that in turn might detrimentally influence the students ' understanding of the definition. Finally in applications of the definition it is certainly convenient to be able to directly invoke any particular value of ε in a proof.

The understanding of the phrase 'for all ε>0' certainly brings in meta-cognitive aspects (see Schoenfeld (1992) for an exposition on metacognition), as well as cognitive, because there are other considerations coming here beyond the merely conceptual. Eventually many students learn functional ways of how ε should be manipulated in various types of circumstances, but this seems to be done independently of an appreciation of the role of the phrase. We contend that an attempt ought to be made to explain the rationale of the phrase in a didactical context, as otherwise the definition cannot be made a primary expression of the concept of limit in cognitive terms.

Finally, there is a somewhat similar feature of the definition that might also ultimately bring some confusion. It is natural to believe that as you decrease ε, N should either stay as it is or increase. However there is nothing in the definition to enforce this condition. Cognitively the pull is to take the least possible value for N; meta-cognitively, the definition is designed not to insist on this because otherwise the application of the definition would be severely curtailed. The relationship between ε and N is further examined in §4.3.7.

4.3.5.

Here we consider the illustrating graph. This can be thought of either as an aid in comprehension of the definition or as an aid for memory in the form of a mental image from which the definition may be reconstructed.

Probably any student first meeting the definition of a limit (of a sequence) will want to refer to an illustrating graph like the one we have presented. (Such diagrams seem more or less standard features in textbooks.) Even though the formal expression involved in the definition can be 'thought through' on the lines we have just described, it has to be said that it is very unlikely that a student could carry through these thoughts without some kind of help. The geometric framework of the illustrating graph, although it can only be suggestive rather than prescriptive, provides a far more amenable environment compared with the rather formidable looking formal statement involving inequalities and quantifiers provided by the definition.

More specifically, using a graph strengthens the feel that the sequence (a_n) is a function and its ordering (induced from the ordering of the positive integers) is clearly suggested by the examination of the graph looking from the left to the right. The parameter ε seems to have a more concrete appearance in the form of the two bounding horizontal lines around the line $y = L$. Finally the graph provides a good medium to vary ε as you may think is most suitable, so that the potential difficulty about the term '$\forall\varepsilon>0$' discussed earlier may seem less problematic.

However the graph itself has to be understood in the right way, especially as by its nature it is something static, whereas clearly we want it to suggest something dynamic. Hence it should be understood that the points representing the elements of (a_n) and the line $y = L$ are to be taken as fixed features, whilst the two horizontal boundary lines and the line $x = N$ are variable features. An appropriate way to think about the situation is as follows. If the sequence *does* have the limit L, then we imagine making the boundary strip around the line $y = L$ progressively narrower (by decreasing ε); *in response* we expect the vertical line described by $x = N$ to move to the right (i.e. N becomes greater) in order to 'remove' any points which suddenly fall outside the strip as it narrows. If the sequence *does not* have the limit L then, as we narrow the boundary strip, at some stage we shall get into the situation that, *however* far we move the line $x = N$ to the right, there will still be terms of the sequence lying further to the right (i.e. a_n with $n>N$) *outside* the bounds. The latter expression made in terms of the illustrating graph is a good basis to form the negation of the statement of the definition, which might be more difficult for the students to achieve without this channeling of thought via the graph.

Research (in particular Pinto, 1996) suggests that some students mentally retain the content of the definition through graphs whilst others simply memorize the formal statement. This however does not inform us about the influence of the graphs when the students *start* tackling the meaning of the definition.

The illustrating graph may be certainly considered as a 'faithful representation' in that *all* the content of the definition is completely transferred and imbedded in the geometric context. However cognitively speaking we must not immediately assume that the resultant mental image will always be more effective than ones extracted from the definition more directly. An example may be found in a paper by Pinto and Tall (2001), where a particular student was unable to fit the limiting behavior of a constant sequence within the framework of the illustrating graph, whereas he was able to conform it to the definition. This behavior is understandable as the graph image of the limit cognitively supports a dynamic view of approach to the limit, whilst the constant sequence does not correspond to this image. This means that the arbitrary choice of value available for N whatever value ε takes is more readily accepted from the definition, even though the same information of course could be deduced within the context of the graph.

4.3.6.

We have finished our examination of the statement of the definition on the level of understanding and how it fits in with our conceptual experiences. Here we shall say a little about a concern that a student is likely to have when he / she reflects on the practicality of the definition. What seems strange is that the definition suggests that you have to find the limit (somehow) and the definition is only good for checking that your choice is correct. What would seem much more satisfactory is that the definition more or less guides you explicitly how to decide first if a limit exists and second, if a limit exists, how to find its value.

A student can hardly be blamed at this stage for confusing the character of a definition with that of a procedural method. Only illustration, theory building and experience will reveal the potential of the usage of the definition. On this we will say nothing more. However even the immediate application of the definition may well be underestimated. The trouble lies in that for a conceptual understanding of the definition it is natural to consider L as a constant. This deters treating L as a parameter and hence deters the use of the definition to its full potentiality, where in theory every value of L may be tested, hence eventually yielding the value of the limit if it exists. In practice, even, this can be done by retaining the symbol L (rather than assigning any particular value to it) so we might be able to test all values of L simultaneously; this would involve using the negation of the definition.

4.3.7.

Here we consider *why* the official definition has the form it has. The issue concerns mostly the parameter ε. Why is this parameter introduced into the definition when it seems to be independent of the given object of study, that is the sequence (a_n)? Well, let us conduct a 'thought experiment'. By building up from our raw material (the sequence) and from our intuition what a limit should involve we attempt to develop a mental image of the limiting process suitable for conversion into formal mathematical language. A shortcoming identified in the mental image suggests a remedial measure that explains the role of ε. This is how it goes:

Take any $N \in \mathbb{N}$. Consider the distance between a_n and L for each $n > N$. A rather technical case might occur; the set of such values may be unbounded. In this case it has to be recognized that L *cannot* be the limit, so we neglect this case. We suppose then that there is a bound to the distance between a_n and L whenever $n>N$. Let us make this bound as small as possible, and call it ε. Note here that ε *does have* a meaning vis-a-vis the sequence (a_n). By increasing N we simply remove some terms of the sequence (a_n) from consideration, hence it is clear that ε either stays as it is or it decreases. At first sight, perhaps, this seems to be enough to ensure the convergence of (a_n) to L; by increasing N enough we can diminish ε to be less than any positive number however small. However this is where the shortcoming comes in; *the previous statement does not necessarily hold.* Even though ε is decreasing as N increases, the values of ε might 'pile – up' at a point not equaling 0. What is worse is that we cannot have any way to decide whether the values of ε do converge to 0 or not, because otherwise we would be in the absurd situation to have to use the definition within the definition!

What is clear is that we require ε to be able to *freely* vary so that we can set ε as small as desired to check that the terms of (a_n) eventually become arbitrarily close to L. Our mental image described above is flawed because ε is constrained by N. This image may be modeled as a decreasing function $f: \mathbb{N} \to \mathbb{R}$ given by $N \to \varepsilon$. Let's denote f(0) as ε_0. Then we construct the following function

$$g:(0, \varepsilon_0] \to \mathbb{N} \text{ given by } g: \varepsilon \to N$$

where N is the unique integer satisfying

$$f(N) \geq \varepsilon > f(N+1).$$

Hence we have rendered ε as the *independent variable* and L is the limit of (a_n) if and only if $g(\varepsilon)$ is defined for all $\varepsilon \in (0, \varepsilon_0]$. Hence by switching the order of dependence between ε and N we can in fact, via g, recover a form of the definition (though one that demands much more precise information than the standard one).

The explicit construction of the function g may well seem to be too formidable for many students, even though the rationale of its structure may be fairly well captured by use of diagrams. However, our use of functions here was only dictated by a wish to be as explicit as possible in treating the issue. The issue may also be thought through in more informal ways (though perhaps at a loss of some clarity). Indeed it is reported in Pinto and Tall (2001) that a student via his own reflections was able to argue why the dependence of ε on N (which might be naively expected) had to be reversed. But the student in question took a considerable time after first meeting the definition to be able to communicate this insight, even though the student was deemed as being strong mathematically. What we conjecture is that the reason why it took so long for the student to tackle the issue is that it took that long for the issue to surface. We feel that if we want to have the definition to have some close rapport with the students' intuitive ideas, the issue, about why the definition has the form it does, must be raised as part of its presentation. Further we contend that the basis on which this explanation would be built on should be fairly well developed by the class-discussion sessions as envisaged in Section 3. Here the crux of the argument already lies on whether to put the focus on the process or on its 'completion' (or limit). The didactical aim would then be to persuade the students of the following: for the definition to express behavior expected by naïve ideas, the attention must be placed on the completion; in colloquial terms, the 'tail' of the sequence must always be contained in any 'neighborhood' of the completion. This program, we feel, is both feasible and useful.

5. Refining the intuition via images evoked from the definition; Concluding remarks

We have noted before in this paper that a sequence is often considered as an infinite process, with the limit acting as a sort of ultimate value for the process. But as well as value, the limit may be regarded to have a special 'ultimate' position within the sequence. The belief that limit specifies both position and value may explain the common remark that students make that a sequence 'must not reach its limit'.

It seems important then to clearly distinguish limit from the sequence itself. But it is in this issue that the typical sources of initializing intuition about sequences (c.f. the paradigms in section 3) seem very influential in a negative way. They introduce the idea of a limiting process more or less simultaneously with that of the sequence.

Although motivating a study on sequences for their own sake may be difficult, it seems important to try to do this. (It is noticeable how often items in lists of misconceptions about limits really refer only to misconceptions about sequences, e.g. items 2 and 7 on p. 6). The limit should not be introduced to appear to be integral with the sequence. One approach (in outline) might be to first induce images consonant to the notion of Cauchy sequences, which may be thought of as a property that the sequence might or might not have. A particular number (to be called the limit) is associated with the property if it holds; the number limit is then to be shown to be the unique one satisfying the familiar conditions (found in the standard definition). However the teaching sequence implicit in this probably would introduce more problems than it would save. Another approach is to let the images developed from the definition as described in the last section to alter the original intuition (as encouraged by paradigms like those in section 3). With this theme, we shall finish the paper.

We shall concentrate on the illustrating graph for the real sequence, pictured in §4.2, with a suitable description of how it is to be interpreted dynamically, as given in §4.3.5. As intimated in the previous section, the graph may be regarded as a system where all of the information entailed in the definition may be faithfully imbedded. The relative concreteness of the graph makes it a powerful depository to 'download' your understandings of the definition, even though the correspondence between definition and graph may have cognitive hitches, such as the special case of constant sequences. (Also, we do not claim that alternative images just as robust do not exist.) Potentially, though, all intuitions, beliefs or behavior about limits of sequences may be compared with the 'official' image: those that fit in with this dynamic graph model may be retained, some may be adjusted cognitively to allow compatibility, whereas some might have to be dismissed from the mind as being basically counter in spirit from the official line that has been established.

However this tactic inherits a problem. Students typically do not have a mature image of the reals, and it is not the role of the definition of limit to elucidate this matter. Thus, for example, a student, could believe that a sequence has a final term in some 'ultimate' position, and because this is to do with an understanding of the reals, we are not in a situation to give the student an argument that will provide an absolute disproof. But we claim that the imagery evoked from the definition strongly encourages a helpful image about the natural numbers at least.

Taking the illustrating graph for a limiting case, the static situation shows some initial terms of the sequence to the left of a threshold position N, where a few further terms representing all subsequent terms all fall within a horizontal strip of a certain breadth. By reducing this breadth, we induce N to 'move to the right' and the representing terms will be further along the sequence. However small we make the thickness of the strip, we never have a sense that we are 'exhausting' the positions (i.e. the positive integers) taken by the terms of the sequence. In particular the notion of a final term of the sequence is at least a redundant one in the model.

Another point is that if the final term is conceived of, surely it would be associated with the limit L, yet the way L is represented in the graph is completely different from the values of the 'finite' elements of the sequence. How L is 'drawn' as a horizontal line suggests an examination of how the values of the finite elements differ from L, and nothing more. In fact the manner in which L is represented is very likely to help students to identify L as a number and a construct related, true, to the sequence but not as an integral part of it.

Hence, plausible arguments made on the dynamic graph may help to persuade students to abandon beliefs such as a final term existing in an infinite sequence, the limit being both the value and position of this final term, and the limit never being reached (in the on-going process involved in the sequence) because the ultimate position of limit is never reached. But we contend that the potential use of this graph imagery is far broader than that suggested by this one issue. For example, it seems an ideal platform to try to describe various types of behavior (e.g. alternating terms above and below the limit) that a sequence may have to widen images restricted to special cases. Hence we may hope, for instance, to dispel the belief that the existence of a limit requires (after some point) a monotonic behavior. In fact all of the misconceptions listed in §2.2.1 may be worth a comparison with the model. But we may go further than this role of censoring or endorsing intuitive ideas. The imagery potentially reflects with precision the content of the definition, and so whenever it is convenient it can be evoked whilst doing proof involving first principles. This topic though will be left as a subject for another paper. (Similar dynamic graphs are available for the concept of differentiation, yet are not usually didactically presented; their use both conceptually and for utility in solving problems has been considered in Downs and Mamona-Downs (2000)).

The author thanks David Tall and Shlomo Vinner for their insightful comments on an earlier draft of this paper; Cathy Kessel and Alan Schoenfeld for their remarks towards the writing of its final form. She also thanks Martin Downs for helping in the honing down of some of the finer points of the paper.

REFERENCES

Artigue, M.: 1991, 'Analysis', in D. Tall (ed.), *Advanced Mathematical Thinking,* Kluwer Academic Publishers, Dordrecht, pp. 168–198.

Binmore, K.G.: 1977, *Mathematical Analysis, A Straightforward Approach*, Cambridge University Press, Cambridge, pp. 26–27.

Boyer, C.B.: 1939, *The History of the Calculus and its Conceptual Development,* reprinted by Dover, New York, 1959.

Brousseau, G.: 1986, 'Fondements et methodes de la didactique des mathematiques', *Recherches en Didactique des Mathematiques* 7(2), 33–115.

Brousseau, G.: 1989, 'Le contrat didactique: Le milieu', *Recherches en Didactique des Mathematiques* 9(3), 309–336.

Burenkov, V.I. and Terarykova, T.V.: 1996, 'Overcoming difficulties in learning limits' in the *Book of Abstracts of Short Presentations ICME 8*, Sevilla.

Cornu, B.: 1981, 'Apprentissage de la notion de limite: modeles spontanes et modeles propes', *Actes du Cinquieme Colloque du Groupe Internationale PME*, Grenoble, 322–326.

Cornu, B.: 1983, *Apprentissage de la notion de limite: conceptions et obstacles*, These de doctorat de troisieme cycle, L'Universite Scientifique et Medicale de Grenoble.

Cornu, B.: 1991, 'Limits', in D. Tall (ed.), *Advanced Mathematical Thinking*, Kluwer Academic Publishers, Dordrecht, pp. 153–166.

Cottrill, J., Dubinsky, E., Nichols, D., Schwingendorf, K., Thomas, K. and Vidakovic, D.: 1996, 'Understanding the limit concept: Beginning with a coordinated process scheme', *Journal of Mathematical Behavior* 15, 167–192.

Davis, P.J. and Hersh, R.: 1981, *The Mathematical Experience*, Birkhauser, Boston.

Davis, R.B. and Vinner, S.: 1986, 'The notion of limit: Some seemingly unavoidable misconceptional stages', *Journal of Mathematical Behavior* 5(3), 281–303.

Downs, M. and Mamona-Downs, J.: 2000, 'On graphic representation of differentiation of real functions', *THEMES in Education* 1(2), 173–198.

Dubinsky, Ed, Elterman, F. and Gong, K.: 1988, 'The student's construction of quantification', *For the Learning of Mathematics* 8, 44–51.

Fischbein, E., Tirosh, D. and Hess, P.: 1979, 'The intuition of infinity', *Educational Studies in Mathematics* 10, 3–40.

Gardiner, A.: 1982, *Infinite Processes Background to Analysis*, Springer-Verlag, New York.

Khinchin, A.Ya.: 1968, *The Teaching of Mathematics*, (Essays by A. Ya. Khinchin edited by B.V. Gnedenko), The English University Press, pp. 23–24.

Kline, M.: 1972, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York, pp. 35–37.

Lakoff, G. and Nunez, R.E.: 2000, *Where Mathematics Comes From*, Basic Books.

Mamona, J.: 1987, *Students Interpretations of Some Concepts of Mathematical Analysis*, Unpublished Ph.D. Thesis, Faculty of Mathematical Studies, University of Southampton.

Mamona, J.: 1990, 'Sequences and series – Sequences and functions: Students' confusions', *International Journal of Mathematical Education in Science and Technology* 21(2), 333–337.

Mamona-Downs, J. and Downs, M.: 2000, 'Advanced mathematical thinking with a special reference to reflection on mathematical structure', to appear in Lyn English (Chief Ed.), *Handbook of International Research in Mathematics Education,* Lawrence Erlbaum Ass., NJ.

Nunez, R.: 1993, 'Big and small infinities: Psychocognitive aspects', *Proceedings of PME 17*, Tsukuba, Japan, II, 121–128.

Nunez, R.: 1994, 'Subdivision and small infinities: Zeno, paradoxes and cognition', *Proceedings of PME 18*, Lisbon, Portugal, III, 368–375.

Orton, A.: 1980, *A Cross-sectional Study of the Understanding of Elementary Calculus in Adolescents and Young Adults*, Unpublished Ph.D. Thesis, Leeds University.

Pinto, M.M.F.: 1996, 'Students use of quantifiers' Paper presented to the Advanced Mathematical Thinking Group of PME 20, Valencia, Spain.

Pinto, M.M.F. and Tall, D.O.: 2001, 'Following students development in a traditional university classroom', *Proceedings of the 25th Conference of the International Group of Mathematics Education* 4, 57–64. Utrecht, The Netherlands.

Robinson, A.: 1966, *Non-Standard Analysis*, Amsterdam, North-Holland.

Schoenfeld, A.: 1992, 'Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics', in D. Grouws (ed.), *Handbook of research on* *mathematics teaching and learning*, Macmillan and National Council of Teachers of Mathematics, New York, pp. 334–370.

Schwarzenberger, R.L.E. and Tall, D.O.: 1978, 'Conflicts in the learning of real numbers and limits', *Mathematics Teaching* 8(2), 44–49.

Sierpinska, A.: 1987, 'Humanities students and epistemological obstacles related to limits', *Educational Studies in Mathematics* 18, 371–397.

Sierpinska, A.: 1990, 'Some remarks on understanding in mathematics', *For the Learning of Mathematics* 10(3), 24–36.

Sullivan, K.: 1976, 'The teaching of elementary calculus: an approach using infinitesimals', *American Mathematical Monthly* 83(5), 370–375.

Szydlik, J.E.: 2000, 'Mathematical beliefs and conceptual understanding of the limit of a function', *Journal for Research in Mathematics Education* 31(3), 258–276.

Tall, D.O.: 1980a, 'Intuitive infinitesimals in the calculus', *Abstracts of short communications,* ICME IV, Berkeley, page C5.

Tall, D.O.: 1980b, 'The notion of infinite measuring numbers and its relevance in the intuition of infinity', *Educational Studies in Mathematics* 11, 271–284.

Tall, D.: 1992, 'The transition to advanced mathematical thinking: Functions, limits, infinity and proof', in D. Grouws (ed.), *Handbook of Research on Mathematics Teaching and Learning*, Macmillan and National Council of Teachers of Mathematics, New York, pp. 495–511.

Tall, D.O. and Vinner, S.: 1981, 'Concept image and concept definition in mathematics with particular reference to limits and continuity', *Educational Studies in Mathematics* 12(2), 151–169.

Vinner, S.: 1987, 'Continuous functions-images and reasoning in college students', *Proceedings of PME 11*, Montreal, Canada. III, pp. 177–183.

Vinner, S.: 1991, 'The role of definitions in the teaching and learning of mathematics', in D. Tall (ed.), *Advanced Mathematical Thinking*, Kluwer Academic Publishers, Dordrecht, pp. 65–81.

Williams, S.R.: 1991, 'Models of limit held by college calculus students', *Journal for Research in Mathematics Education* 22(3), 237–251.

University of Macedonia,
156 Egnatia str., P.O. Box 1591, 540 06,
Telephone +30 31 891382 (office), +30 31 915064 (home)
Fax +30 31 844536 (for J. Mamona-Downs)
E-mail: mamona@uom.gr