



This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Title

Classification and Concept Consistency

Abstract

This paper investigates the extent to which undergraduates consistently use a single mechanism as a basis for classifying mathematical objects. We argue that the concept image/concept definition distinction focuses on whether students use an accepted definition, but does not necessarily capture the more basic notion that there should be a fixed basis for classification. We examine students' classification of real sequences before and after exposure to definitions of 'increasing' and 'decreasing'; we develop an abductive 'plausible explanations' method to estimate the consistency within the participants' responses and suggest that this provides evidence that many students may lack what we call 'concept consistency'.

Introduction

The idea of classifying mathematical objects into well-defined categories or sets is central to mathematics, and has been so at least since Cantor. It is not always easy to implement: as mathematics develops, there are debates about which definitions should be adopted; what is, or is not, classified as a ‘function’, a ‘polyhedron’ or a ‘set’ changes as the definitions are refined by successive generations of mathematicians (Lakatos, 1976; R  thing, 1984). Indeed, set theory has encountered problems (such as the ‘set of all sets’) that have required ever more intricate foundational elaboration to address (Ferreir  s, 2007). However, the idea that there are agreed definitions against which membership is judged is not contentious to most practicing mathematicians, and pure mathematics as experienced by undergraduate students is usually founded on the so-called ‘classical model’ in which that a category is specified by a fixed definition. That is, category membership is completely determined by a set of necessary and sufficient conditions.

Categories are fundamental to everyday cognition too: the act of classifying objects into categories is a natural human process, forming a vital way in which we make sense of the world around us (Markman, 1989). Categories simplify and accelerate cognition, allowing us to predict and control regularities in the world and to make inferences about properties that are not immediately apparent (Gelman & Coley, 1991; Kruschke, 2005). However, in everyday contexts, the classical model is a poor approximation: categories are not usually well-defined.

This means that habits of everyday cognition need to be modified in order to work effectively in mathematics at the undergraduate level. This paper is about the extent to

which the use of the classical model, as described above, is evidenced in students' thinking during the transition to advanced mathematics.

Theoretical background

The classical model, as well as having a foundational role in mathematics, was the commonly held view of general cognition (e.g. Boole, 1854/1951) until research began to uncover effects for which it could not account. Natural thought seems to involve categories with radial structures and ill-defined boundaries: people behave as though categories can have 'branches' that do not share common properties, or as though some objects are more or less 'central' within a category. For example, Smith, Shoben and Rips (1974) found that people were quicker to classify a robin as a bird than to classify a penguin as a bird. To account for such effects, cognitive models for natural human category learning commonly build on prototype theory (Rosch, 1973), in which classification involves testing similarity against an abstracted summary of experience with category members, or on exemplar theory (Medin & Schaffer, 1978), in which classification involves testing against a number of specific stored exemplars. More recent research suggests that natural categorisation may be better modelled with a combination of these mechanisms, incorporating selective attention so that some features of exemplars are weighted more highly for the purposes of classification in certain contexts (see Kruschke, 2005). All such models, however, aim to account for human categorisation in everyday contexts, and not for learning involving categories based on the (naïve) formal mathematical idea that an element belongs to a set if and only if it satisfies some fixed properties.

Lakoff and Nuñez have argued that even the seemingly rigorous foundations of mathematics based on sets and formal logic (the ‘formal foundational principle’) are at odds with ‘mathematics...viewed as a human conceptual system’ (Lakoff & Nuñez, 1998, p.93); so much so, they argue, that ‘sets and logic’ and ‘human mathematics’ are simply different forms of mathematics that ‘yield different results...each of which is perfectly valid in its own sphere’ (*ibid*, p.86). This view is challenged by some professional mathematicians, who claim that it conflates ‘mathematical facts’ with ‘human knowledge of mathematical facts’ (Gold, 2001). Whatever the case, the mathematics as it is encountered traditionally in undergraduate courses is presented according to the classical model. Thus we should expect conflict arising from the difference between natural forms of classification and those required for the formal mathematics encountered by undergraduates.

In mathematics education, Tall and Vinner highlighted the cause of such conflict when they introduced the key notions of concept image and concept definition. They argued that a student might sometimes reason using their concept image, where this term means ‘the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes’ (Tall & Vinner, 1981, p.152). Concept images can be built over years of encounters with mathematical concepts in the form of examples, different representations, tasks etc. Researchers have argued that reasoning with reference to concept images alone may be productive in many cases, but that a student may misclassify if their concept image is not fully developed or not coincident with the extension of the definition (Davis & Vinner, 1986). Such misclassifications and related misinterpretations are prevalent in the cases involving

functions and graphs (Vinner, 1991), limits (Davis & Vinner, *ibid.*), derivatives (Zandieh, 2000), tangents (Biza, Christou & Zachariades, 2008) etc. One could argue that classifying according to a concept image amounts to using natural classification in a mathematical context; concept images might be similar in structure to natural categories while mathematical judgements need to be based on (or at least accord with) formal, agreed definitions. So judgements based on a concept image can sometimes leads to conclusions that conflict with the formal theory.

In Tall and Vinner's terms, then, students need to learn to work with concept definitions in formal mathematical contexts. This is not easy. Some difficulties arise from the specifics of definitions encountered: it can be challenging to work with statements involving highly compressed mathematical symbols and complex logical structures (Dubinsky, Elterman & Gong, 1988), especially if the evocative nature of the words they define ('converges', 'tends to' etc.) makes it easy to fall back on natural strategies (Tall & Vinner, 1981; Monaghan, 1991).

Such difficulties may be exacerbated by messages implicit in textbooks. Raman (2004), for instance, highlighted a range of relationships between the presentation and use of definitions in texts related to calculus. She found that in a precalculus textbook, definitions of continuity were presented informally and classifications were based on graphical imagery; in a calculus text, formal definitions were given but tasks did not require their use, and in an analysis text, formal definitions were both given and regularly used. The first two approaches may not lead naturally to the third, though such pedagogical sequences may be unavoidable if we wish to introduce a concept before the formal theory is accessible (Barbé et. al., 2005). However, they may result in a lack of

clarity about the role of definitions, which may be further compounded by the introduction of different forms of the same definition: Bergé (2008) observed the definition of completeness transforming through three distinct (but equivalent) variants as students move through three successive courses. Of course, there may be sound pedagogical reasons for using a particular variant of a definition at a certain stage, but it might not be clear to the students that such definitions are supposed to be logically equivalent and thus determine identical extensions.

Students' understanding and use of definitions has been the subject of a considerable research literature, including studies on the nature of definitions (e.g. Van Dormolen & Zaslavsky, 2003), students' beliefs about them (Zaslavsky & Shir, 2005), and students' guided collaborative development of definitions in contexts new to them (Larsen & Zandieh, 2008; Zandieh & Rasmussen, 2010). Studies have suggested that students who spontaneously generate examples in response to new definitions later perform better on tasks involving the defined concepts (Dahlberg & Housman, 1997), and there has been much recent discussion of the way in which example generation tasks might help students to modify their example spaces so that these more closely coincide with the extensions of the defined concepts (Mason, 2002; Watson & Mason, 2005; Zazkis, Liljedahl & Chernoff, 2008; Zazkis & Leikin, 2008). Definitions are also discussed in the extensive literature on proof, which focuses on the role of definitions as a deductive base in arguments about whole classes of objects. This (desirable) use of definitions is often contrasted with students' attempts to support general arguments by using examples in more or less sophisticated ways (Authors, 2004; Harel & Sowder, 2007; Healy & Hoyles, 2000). Inappropriate example-based or concept image-based

reasoning has been shown to persist even when the students know the appropriate definitions (Vinner, 1991) and even when one would think there was little in their previous mathematical experience that would support non-deductive strategies (Edwards & Ward, 2004).

The most basic form of definition use, however, is in classifying objects. Ross and Makin (1999) point out that while ‘classification is rarely your ultimate goal when you are faced with an unfamiliar object ...[it] is a first step in allowing you to know how to deal with the entity’ (p.228). Such knowledge is possible because, in both natural and classical situations, classification and definitions have a two-way relationship.

In the natural world, you may decide that something is a cat because of a level of furriness, four-legged-ness etc., but equally if you are told something is a cat, you can make a reasonable guess about its having fur and possessing four legs (as well as much other information about what it would like to eat, whether it is safe to stroke, how high it can jump etc.). This, of course, involves some measure of personal judgment and uncertainty, as captured in the psychological models of natural category learning.

In the classical world, however, inference is simpler because it does not involve personal judgement or uncertainty; you can always decide that a number is composite by finding two proper factors and can always conclude from being told that a number is composite that it must have (at least) two such factors. In this sense, mathematical classification and argumentation are simpler in logical structure than their everyday equivalents, a fact that mathematicians recognise and sometimes attempt to point out to their students (Author, 2010).

Of course, this does not mean that mathematical classification and reasoning are psychologically simpler, partly because the classical model demands *consistency* in a way that everyday classification does not. As noted earlier, people treat everyday categories as though they have internal structure (with ‘better’ and ‘worse’ examples), sometimes to the extent that the cueing context may evoke conflicting judgements about category membership: in some contexts the Pope may be classified as a bachelor, in he may be classified as not-a-bachelor (Rips, 1989; Machery, 2009). Classification in natural category learning can thus lead to situations which would be impossible in classical theory and unacceptable in mathematics.

We argue that in mathematics, even before we consider whether a student is using a definition to make a classification judgement, we should consider whether they possess this more fundamental idea of ‘concept consistency’: that is, whether they understand that there should be a single mechanism for judging whether a mathematical object is a member of a set (whether or not that mechanism is a formal definition). We claim that students’ knowledge of this underlying principle is not usually addressed in mathematics education research. Studies have typically made a distinction between classifications or arguments based on accepted definitions and classifications or arguments based on concept images or examples. That approach addresses the issue of whether students know the accepted definitions and classify according to these, but not whether they behave as though there *ought to be* consistent classification even in cases in which they have not yet learned the standard definition.

Zandieh and Rasmussen (2010) and Zaslavsky and Shir (2005) have shown that college students and high-achieving upper high school students can engage in discussions

about desirable properties of definitions and do seem to understand that the goal is to decide on a definition that covers all possible cases. However, in both cases this work was conducted in a collaborative learning situation, so it does not speak to the behaviour of individuals when asked to classify objects without discussion, nor to the issue of whether classification is consistent in the absence of a definition.

We suggest that the principle of concept consistency would be more clearly manifest if students were asked individually to classify objects as they saw fit, prior to an encounter with the accepted definition. In this paper, we take such an approach and present evidence that at least some beginning university students behave as though they do not possess concept consistency – that is, they do not appear to use a single mechanism for making classification judgements. Specifically, we ask:

1. How does classification differ before and after exposure to a relevant definition?
2. To what extent are students' responses consistent either with a reasonable interpretation of these definitions or, more fundamentally, consistent with each other?

Methods

Participants and Context

Our study involved 187 students in the first year of study for a mathematics or mathematics-related degree at a highly ranked UK university. The students were all high achieving, having attained the highest grade in their pre-university mathematics examinations in order to enter their degree course. The study took place in week four of a real analysis course as part of a standard lecture period where the lecturer allotted time for the researchers to conduct the task. The intention was to expose students to a concept

which they had not yet formally encountered, but for which they could be expected to have spontaneous conceptions associated with everyday language and earlier mathematical experience. In addition, we wanted a concept for which some examples could be considered ‘central’ to the categories while others could be considered to be ‘boundary’ cases, so that natural classification might reasonably lead to a range of different responses. We thus used the key concepts of ‘increasing’ and ‘decreasing’ as applied to sequences; while the students already had four weeks experience of formal mathematics (including work with sequences and the formal definition of convergence), definitions for ‘increasing’ and ‘decreasing’ in the context of sequences had not been introduced. These definitions also had the advantage that they are logically very simple, each having only one (universal) quantifier and one condition.

Task

The task had three phases: a pre-definition classification phase; a definition presentation and a post-definition classification phase. In both the pre-definition and post-definition classification phases, the participants were given a sheet with each of the sequences shown in Figure 1 listed on it. Beside each sequence were four boxes marked ‘I’, ‘D’, ‘B’ and ‘N’, and the participants were asked to tick exactly one of these to indicate whether they would classify the sequence as ‘increasing’, ‘decreasing’, ‘both increasing and decreasing’ or ‘neither increasing nor decreasing’. During each classification phase, students were able to work through the questions in any order and were able to alter choices as they wished. However, at the end of the pre-definition phase, participants were asked to turn to the post-definition page and not turn back. There was no evidence that this request was ignored.

INSERT FIGURE 1 ABOUT HERE

Between the classification phases, the participants were given a short presentation in which one of the researchers displayed the formal definitions of ‘increasing’ and ‘decreasing’ for real sequences using symbolism consistent with that used by the lecturer (Figure 2), checked that students were familiar with all of the symbols, and read the definitions aloud. No examples were given, nor was any other support for the classification task (such as natural language meanings) given. The definitions remained visible on the main lecture theatre screen while the students completed the post-definition classification, in which they were explicitly asked to classify according to these definitions.

INSERT FIGURE 2 ABOUT HERE

In the pre-definition phase, students were also given the option to select ‘not sure’, but only a minority of students chose this for any sequence. To allow comparisons across phases, this data was discarded along with any spoiled or incomplete scripts, leaving 146 complete paired sets of data. Data was collated to give both cross-participant response rates and individualised patterns of responses as analysed below.

Results

The Effect of Presenting the Definitions

Table 1 shows the number of participants who classified each sequence in each category in the pre-definition phase. It is clear that, for some sequences, almost every student classified in a way that is in keeping with the definition:

B: 1, 4, 9, 16, 25, 36, 49, 64, ..., C: $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \dots$, I: $10\frac{1}{2}, 10\frac{3}{4}, 10\frac{7}{8}, 10\frac{15}{16}, 10\frac{31}{32}, \dots$ and J:

$-2, -4, -6, -8, -10, \dots$ are all central category members in the sense that everyday uses of ‘increasing’ and ‘decreasing’ coincide naturally with the formal definition. This is somewhat less true for G: $6, 6, 7, 7, 8, 8, 9, 9, \dots$, although most still classified it as increasing. Sequence D: $1, -1, 2, -2, 3, -3, 4, -4, \dots$ provoked substantial agreement amongst the students, who tended to classify it as ‘both increasing and decreasing’, a response that differs from the definition-based classification. For sequences F: $1, 3, 2, 4, 3, 5, 4, 6, \dots$ and H: $0, 1, 0, 2, 0, 3, 0, 4, \dots$ there was a split between the classifications ‘both’ and ‘increasing’, with only a minority selecting the normatively correct classification ‘neither’.

INSERT TABLE 1 ABOUT HERE

It is important to consider the responses to sequence E: $3, 3, 3, 3, 3, 3, 3, \dots$. There are two sensible formal mathematical definitions of the concept of ‘increasing’, depending on whether each term is required to be strictly greater than its predecessor or whether the weaker ‘greater than or equal to’ relationship is used. Item E would be classified as ‘neither’ with the strict interpretation and as ‘both’ with the non-strict alternative. In the pre-definition classification task, every student classified it as ‘neither’. This was counted as correct for analysis of the pre-definition responses only, because the definitions presented were formal definitions of ‘increasing’ and ‘decreasing’ in the non-strict sense.

In the pre-definition phase, students would have been highly unlikely to make classification judgements based on a formal definition, since this had not been presented to them. They clearly made some judgements which matched the extension of the formal definition and some which did not. This data thus fits Davis and Vinner’s (1986) notion of concept image-based classification: for sequences that display simple, familiar

behaviours, most responses matched the extension of the definitions, and for those that do not, responses were considerably more mixed.

In the post-definition phase, students had access to the formal definitions and we might therefore expect their responses to be in line with these. However, they also still had access to whatever decision mechanisms they had previously used, and we can see the impact of this in their post-definition responses, as shown in Table 2. Again, sequences B, C, I and J were classified almost perfectly consistently with the definitions. For all other sequences it was possible to see clear changes in the patterns of responses from the pre-definition to the post-definition phase. We conducted Fisher Exact Tests for each item to examine these changes, and found that for items B, C, I and J there was no significant difference between the response patterns ($p > 0.999$ in each case), but for every other item the difference was significant ($p < 0.001$ in each case).

INSERT TABLE 2 ABOUT HERE

Note that the average number of correct classifications (that is, those which matched the extension of the formal definition given) increased from 6.65 ($\sigma = 1.19$) to 7.82 ($\sigma = 2.04$), which is a significant ($t = 6.1077$, $p < 0.001$) and moderate-to-large difference ($d = 0.70$). The data thus show an improvement in classification relative to the definitions after these were presented. However, it is surprising that the effect is not larger. Given the experience of the students with definitions and the relatively simple concept involved in the task, one could have expected a near perfect post-definition classification. In fact, in the next section, we shall see that only 29% of the participants' classifications were fully in line with the given definitions.

Pre-Definition Classification Consistency

Our principal concern in this paper is not the particular classifications made, but the inferences we can make about the extent to which students appeared to use a single mechanism for judging whether the sequences were increasing/decreasing; that is, the extent to which they exhibited concept consistency. For this purpose, it is clear that some items did not discriminate between students, notably those sequences that were classified almost perfectly in line with the definitions in the pre-definition phase. As a result, our subsequent analysis omits those sequences and also omits the small number of participants who classified any of these apparently straightforward items incorrectly at either stage (leaving $N = 136$). For the remaining items A, D, E, F, G and H we produced a response profile for each participant. The normatively correct profile is shown in Figure 3.

INSERT FIGURE 3 ABOUT HERE

It was not expected that the pre-definition responses would be identical across the group. However, if students possess concept consistency we would expect to see only a small number of response profiles, each linked to a possible classification mechanism. Instead, across 136 students, we saw 24 different response profiles, none of which was the profile consistent with either mathematically acceptable definition (NNBNIN or NNNNIN). Table 3 shows the six most commonly occurring profiles (all 24 are given in Appendix A).

INSERT TABLE 3 ABOUT HERE

To consider the issue of concept consistency in detail, we developed a method of ‘plausible explanations’: for each profile, we tried to identify fixed sets of properties or

key features (in the sense of Ross and Makin, 1999) which could account for all the responses. The method of ‘plausible explanations’ can be seen as a form of abductive reasoning; constructing hypotheses which would account for the responses *if the student possessed concept consistency* (Magnani, 2001). In cases where no plausible explanation could be constructed for the given response, we argue that this constitutes abductive evidence that the student did not possess concept consistency.

One could argue that some students might possess concept consistency – they may be trying to use a single mechanism for making their classification judgements – but make mistakes in applying the mechanism they have in mind. While we accept this possibility, we think it would be unlikely to have a large impact here because the task was a straightforward one involving familiar concepts, the students were all high achieving, and the given definitions are logically simpler than most they had already encountered on their course.

Indeed, we would argue that the plausible explanation method may overestimate consistency: the existence of a plausible explanation does not guarantee that any given student used a single mechanism for making their classification judgements. Some students may have used different mechanisms for different items (i.e. lacked concept consistency) but nonetheless given a response profile for which it happens that there is a plausible explanation. In any case, we argue that the method does provide some indication of the number of students who did or did not use a consistent classification scheme.

The response profiles for which we could generate a plausible explanation are:

BBNBIB: This profile appears to be consistent with a ‘local’ interpretation. That is, if some terms are less than their immediate predecessors and others are greater then the sequence is classified as ‘both’, and if all the terms are equal then it is classified as ‘neither’.

NBNIII: There were two plausible explanations for this profile. In one, the participant might treat the sequence as two separate sequences and classify them separately (so, for instance, F: 1,3,2,4,3,5,4,6,... is increasing because it is composed of two interleaved increasing sequences). Alternatively, the participant may associate the informal notion of ‘increasing’ with what would formally be called ‘unbounded above’.

NNNIII: This profile appears to be consistent with a ‘holistic’ interpretation, seeking trends across the sequence as a whole rather than being influenced by local behaviour. For instance, F: 1,3,2,4,3,5,4,6,... could be seen as increasing ‘overall’ with small steps backward not treated as an important feature. One could view this as an implicit consideration of the ‘limiting bounds’: the sequence of infima and suprema of the sequence of tail sequences.

BBNBIB: This profile is very similar to the modal profile and fits the ‘local’ interpretation, except that G: 6,6,7,7,8,8,9,9,... is classified as ‘neither’, perhaps because of an interpretation in which between some consecutive pairs there is an increase and between others there is not.

We note that these classification mechanisms are entirely sensible; in the absence of any specified criteria, each represents a reasonable interpretation of the meanings of

increasing and decreasing in this context, and the way these terms combine. In particular, our plausible explanations suggest that students' concept images might cause them to focus locally (on pairs or small finite sets of consecutive terms) or holistically (the limiting bounds of a sequence or its long-term behaviour) rather than universally (across all pairs of consecutive terms). Indeed, lecturers teaching this subject might find it useful to know that large numbers of students seem to naturally interpret 'increasing and decreasing' in these local or holistic senses, and that this potential for conflict will need to be addressed if their interpretation is to be brought into line with the definitions.

Here, however, we are interested in the fact that these four categories account for only 72% of the responses. For the other 28% (18 distinct response profiles), we were unable to generate a plausible explanation, and we argue that this indicates that these students may lack an underlying idea of concept consistency. To illustrate this phenomenon, we consider two such profiles.

First, consider BBNIII. One can develop plausible explanations that account for parts of this profile: for example, A: 0,1,0,1,0,1,0,1,... and D: 1,-1,2,-2,3,-3,4,-4,... might be classified as 'both' because they increase and decrease 'locally'. F: 1,3,2,4,3,5,4,6,... and G: 6,6,7,7,8,8,9,9,... might be classified as 'increasing' because they are increasing 'holistically'. However, neither explanation accounts for the whole response profile and we thus conclude that these students are unlikely to have judged consistently using a single judgement mechanism.

Similarly, in profile NBNBIB, sequences D: 1,-1,2,-2,3,-3,4,-4,..., F: 1,3,2,4,3,5,4,6,... and H: 0,1,0,2,0,3,0,4,... can be classified as 'both' with the 'local' interpretation, but this cannot account for the 'neither' classification given for A:

0,1,0,1,0,1,0,1,.... For both profiles, it appears that the students evoked different parts of their concept image or focused on different key features when responding to the different sequences; their classifications did not manifest concept consistency.

Overall, then, this ‘plausible explanation’ method suggests that around a quarter of the participants appeared not to classify consistently.

Post-Definition Classification Consistency

After exposure to the definition, participants had a new basis on which to classify. They were specifically instructed to classify according to the given definitions, which remained on display throughout the time the students worked on the task. One might therefore expect fewer distinct response profiles, but in fact there were more: 28 in total, all of which appear in Appendix B. Table 4 lists the six most commonly occurring response profiles for the post-definition classification task.

INSERT TABLE 4 ABOUT HERE

When looking at the pre-definition task, we were concerned with the extent to which students responded consistently according to some set of (internally held) criteria. The presence of given definitions in this task, however, changes our focus. For this task, if a student possessed concept consistency, we would expect either a response profile given by one of the plausible explanations above or, if the student understood that they needed to classify with the definition, a normatively correct response profile.

The profiles in Table 4 show that 40 students gave the normatively correct profile NNBNIN, and a further 10 gave one which is consistent with a plausible misreading of the definitions – reading the inequalities as strict – giving the NNNNNN profile. In

addition, nine people gave responses that were consistent with one of the profiles generated by the plausible explanations method outlined above (BBNBNB or NBNIII). The remaining 77 students (57%) gave responses which were neither consistent with the definitions, nor with the 'strict' misreading, nor with a plausible set of internal criteria. In all it appears that the availability of the definitions did not immediately result in normatively correct classifications and actually led to a drop in apparent concept consistency across the sample.

Discussion

Our study explored the concepts of 'increasing' and 'decreasing' in the context of real sequences, with the aim of examining how students' classifications differ in the absence and presence of a definition and, in particular, whether these classifications were consistent with a single judgement mechanism.

In the development of the abductive 'plausible explanations' method, we believe we may also have shed some light on students' concept images of 'increasing' and 'decreasing' when they are used in a new context. In particular, the results suggest that students may focus on local behaviour (on pairs or small finite sets of consecutive terms) or on holistic behaviour (the limiting bounds of a sequence or its long-term behaviour) rather than on the universal behaviour captured by the formal definition.

Our main aim, however, was to investigate the issue of concept consistency. Tall and Vinner's concept image/concept definition distinction, along with the considerable body of research based on it, has illuminated one vital aspect of what students need to learn in order to make sense of advanced mathematics. However, we have argued that

investigation of this issue has usually focused on whether or not students operate according to the accepted definitions and, in so doing, has overlooked the issue of whether or not they possess the more fundamental principle of concept consistency. In this paper we have explored this issue by considering students' responses to pre- and post-definition classification tasks.

We found, first, that exposure to the formal definition did improve classification performance, in the sense that more students classified more sequences in accordance with the definitions. However, this improvement was not as drastic as might be expected given that the context was familiar and the definitions were simple: in the post-definition classification task, only 29% classified in a way which fully matched the extension of the given definitions.

More notably, in both the pre- and post-definition tasks, a large number of students did not appear to use a single mechanism for classification at all – that is, their responses lacked concept consistency. While it has previously been observed that students evoke different, sometimes contradictory concept images in response to different cues or problems (Dahlberg & Housman, 1997), we found the extent to which this occurred in this case somewhat surprising: within each phase the classifications were completed on the same page, in the space of a few minutes, with all the participant's responses visible at once and with no restriction on going back to change a previous answer.

The fact that the degree of consistency actually dropped from the pre- to the post-definition task suggests that it may be difficult even for a student who is trying to classify consistently to maintain this criterion while also incorporating new information. Such

difficulties might occur in three ways: first, such a student might read the definitions insufficiently carefully and thus misapply them throughout or in particular cases; second, they might understand the definitions correctly but be swayed by earlier interpretations in some or all cases; third, they might not even consider changing an answer that they think is obvious (which seems particularly likely in the case of the constant sequence).

Whatever the relative frequencies of these occurrences, the result raises questions for those designing instruction at this level. It might seem intuitively obvious that students should be exposed to a variety of examples in order to improve the correspondence between their individual example spaces and the conventional spaces, but our results suggest that even in simple cases, instructors should be prepared for a period of confusion and errors resulting from conflict between concept image- and concept definition-based judgments.

Clearly, of course, this study has looked at only one concept in one topic area during the transition to advanced mathematical thinking. It may be that some of the issues would be different if the students had more (or less) prior experience of related concepts, or if the concepts were described using terms that did not have familiar everyday meanings. However, the issue of concept consistency seems important as a precursor for learning to work effectively with definitions in advanced mathematics: if a student is not aware that classification should be done on a consistent basis, they are less likely to attend to the agreed definitions. It may therefore be that some students would benefit from instruction that focused on this principle.

References

- Author, (2010).
- Authors, (2004).
- Barbé, J., Bosch, M., Espinoza, L. & Gascón, J. (2005). Didactic restrictions on the teacher's practice: The case of *limits of functions in Spanish high schools*. *Educational Studies in Mathematics*, 59, 235-268.
- Bergé, A. (2008). The completeness property of the set of real numbers in the transition from calculus to analysis. *Educational Studies in Mathematics*, 67, 217-235.
- Biza, I. Christou, C., & Zachariades, T. (2008). Student perspectives on the relationship between a curve and its tangent in the transition from Euclidean geometry to analysis. *Research in Mathematics Education*, 10, 53-70.
- Boole, G. (1854/1951). *An investigation of the laws of thought*. New York: Dover.
- Dahlberg, R. P. & Housman, D.L. (1997). Facilitating learning events through example generation. *Educational Studies in Mathematics*, 33, 283-299.
- Davis, R. B. & Vinner, S. (1986). The notion of limit: Some seemingly unavoidable misconception stages. *Journal of Mathematical Behavior*, 5, 281-303.
- Dubinsky, E., Elterman, F. & Gong, C. (1988). The student's construction of quantification. *For the Learning of Mathematics*, 8(2), 44-51.
- Edwards, B. S. & Ward, M. B. (2004). Surprises from mathematics education research: Student (mis)use of mathematical definitions. *American Mathematical Monthly*, 111, 411-424.
- Ferreirós, J. (2007). *Labyrinth of thought: A history of set theory and its role in modern mathematics*. Basel: Birkhäuser Verlag.

- Gelman, S. A. & Coley, J.D. (1991). Language and categorisation: The acquisition of natural kind terms. In S.A. Gelman & J.P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp.146-196). Cambridge: Cambridge University Press.
- Gold, B. (2001). Review of 'Where mathematics comes from: How the embodied mind brings mathematics into being', www.maa.org/reviews/wheremath.html (accessed 10/09/10).
- Harel, G. & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Education* (pp. 805-842). Greenwich, CT: Information Age Pub Inc.
- Healy, L. & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 31, 396–428.
- Kruschke, J. K. (2005). Category learning. In K. Lamberts & R.L. Goldstone (Eds.), *The handbook of cognition* (pp. 183-201). London: Sage.
- Lakatos, I. (1976). *Proofs and Refutations*. Cambridge: Cambridge University Press.
- Lakoff, G. & Nuñez, R. (1998). What did Weierstrass really define? The cognitive structure of natural and epsilon-delta continuity, *Mathematical Cognition*, 4, 85-101.
- Larsen, S. & Zandieh, M. (2008). Proofs and refutations in the undergraduate mathematics classroom. *Educational Studies in Mathematics*, 67, 185-198.
- Machery, E. (2009) *Doing without concepts*. Oxford University Press.
- Magnani, L. (2001). *Abduction, reason and science. Processes of discovery and explanation*. Dordrecht, The Netherlands: Kluwer.

- Markman, E.M. (1989). *Categorisation and naming in children: Problems of induction*.
Cambridge, MA: MIT Press.
- Mason, J. (2002). *Mathematics teaching practice: A guide for university and college lecturers*. Chichester: Horwood Publishing.
- Medin, D. L. & Schaffer, M.M. (1978). Context model of classification learning.
Psychological Review, 85, 207–238.
- Monaghan, J. (1991). Problems with the language of limits. *For the Learning of Mathematics*, 11, 20–24.
- Raman, M. (2004). Epistemological messages conveyed by three high-school and college mathematics textbooks. *Journal of Mathematical Behavior*, 23, 389-404.
- Rips, L. J. (1989). Similarity, typicality, and categorization. In Vosniadou, S., & Ortony, A. (Eds.), *Similarity and Analogical Reasoning*, (21–59). Cambridge: Cambridge University Press.
- Rosch, E.H. (1973). Natural categories. *Cognitive Psychology*, 4, 328-350.
- Rüthing, D. (1984). Some definitions of the concept of function from Bernoulli to Bourbaki, *Mathematical Intelligencer* 6 (4), 72-77.
- Ross, B. H. & Makin, V.S. (1999). Prototype versus exemplar models in cognition. In R.J. Sternberg (Ed.), *The nature of cognition* (pp. 205–241). Cambridge, MA: MIT Press.
- Smith, E. E., Shoben, E.J. & Rips, L.J. (1974). Structure and process in semantic memory: A featural model of semantic decisions. *Psychological Review*, 81, 214-241.

- Tall, D. O. & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12, 151-169.
- Van Dormolen, J. & Zaslavsky, O. (2003). The many facets of a definition: The case of periodicity. *Journal of Mathematical Behavior*, 22, 91-106.
- Vinner, S. 1991. The role of definitions in teaching and learning. In D.O. Tall (Ed.), *Advanced Mathematical Thinking* (pp. 65-81). Dordrecht: Kluwer.
- Watson, A. & Mason, J. (2005). *Mathematics as a constructive activity: Learners generating examples*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Zandieh, M. (2000). A theoretical framework for analyzing student understanding of the concept of derivative. *Research in Collegiate Mathematics Education IV*, 8, 103–127.
- Zandieh, M. & Rasmussen, C. (2010). Defining as a mathematical activity: A framework for characterizing progress from informal to more formal ways of reasoning. *Journal of Mathematical Behavior*, 29, 57-75.
- Zaslavsky, O. & Shir, K. (2005). Students' conceptions of a mathematical definition. *Journal for Research in Mathematics Education*, 36, 317-346.
- Zazkis, R. & Leikin, R. (2008). Exemplifying definitions: A case of a square. *Educational Studies in Mathematics*, 69, 131-148.
- Zazkis, R., Liljedahl, P. & Chernoff, E.J. (2008). The role of examples in forming and refuting generalizations. *ZDM: The International Journal on Mathematics Education*, 40, 131-141.

Appendices

Appendix A

Full list of pre-definition response profiles:

Response profile	Frequency
B B N B I B	43
N B N I I I	28
N N N I I I	7
B B N B N B	6
B B N I I I	6
N B N B I B	6
N N N N I N	5
B B N B I I	4
N B N I I B	4
N B N I I N	4
N N N I I N	4
B B N I I B	3
N D N I I I	2
N N N N N N	2
B B N B D B	1
B B N I I N	1
B D N I I B	1
B D N N N B	1
B N N I I I	1
N B N B I I	1
N B N N I I	1
N N N B I N	1
N N N I N N	1
N N N N I I	1

Appendix B

Full list of post-definition response profiles:

Response profile	Frequency
N N B N I N	40
N N N N I N	22
B B N B I B	13
B B B B I B	11
N N N N N N	10
B B N B N B	8
B B B B B B	3
N B N B N B	3
B N B N I N	2
N B N B I N	2
N B N N I N	2
N B N N N N	2
N N B I I N	2
N N I N I N	2
B B B I I B	1
B B I B I B	1
B B N B N N	1
B B N I I B	1
B N B N I B	1
B N N B I B	1
N B I N I N	1
N B N B I B	1
N B N I I I	1
N N B N B N	1
N N B N N N	1
N N I B I N	1
N N I I I N	1
N N N I N I	1

Tables

Table 1. Pre-definition responses (asterisks indicate normative definition-based classifications).

	I	D	B	N
A: 0,1,0,1,0,1,0,1,...	0	0	74	72*
B: 1,4,9,16,25,36,49,64,...	146*	0	0	0
C: $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \dots$	0	145*	1	0
D: 1, -1, 2, -2, 3, -3, 4, -4, ...	0	4	119	23*
E: 3, 3, 3, 3, 3, 3, 3, 3, ...	0	0	0*	146*
F: 1, 3, 2, 4, 3, 5, 4, 6, ...	65	0	69	12*
G: 6, 6, 7, 7, 8, 8, 9, 9, ...	133*	1	0	12
H: 0, 1, 0, 2, 0, 3, 0, 4, ...	55	0	72	19*
I: $10\frac{1}{2}, 10\frac{3}{4}, 10\frac{7}{8}, 10\frac{15}{16}, 10\frac{31}{32}, \dots$	140*	4	2	0
J: -2, -4, -6, -8, -10, ...	1	145*	0	0

Table 2: Post-definition responses (asterisks indicate normative definition-based classifications).

	I	D	B	N
A: 0,1,0,1,0,1,0,1,...	0	1	44	101*
B: 1,4,9,16,25,36,49,64,...	146*	0	0	0
C: $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \dots$	1	145*	0	0
D: 1, -1, 2, -2, 3, -3, 4, -4, ...	0	1	53	92*
E: 3,3,3,3,3,3,3,3,...	6	0	69*	71
F: 1,3,2,4,3,5,4,6,...	7	0	48	91*
G: 6,6,7,7,8,8,9,9,...	113*	0	5	28
H: 0,1,0,2,0,3,0,4,...	2	0	47	97*
I: $10\frac{1}{2}, 10\frac{3}{4}, 10\frac{7}{8}, 10\frac{15}{16}, 10\frac{31}{32}, \dots$	141*	3	2	0
J: -2, -4, -6, -8, -10, ...	0	145*	0	1

Table 3: The six most common pre-definition response profiles for items A,D,E,F,G,H.

Response profile	Frequency
B B N B I B	43
N B N I I I	28
N N N I I I	7
B B N B N B	6
B B N I I I	6
N B N B I B	6

Table 4: The six most frequent post-definition responses for items A,D,E,F,G,H.

Post-definition response profile	Frequency
N N B N I N	40
N N N N I N	22
B B N B I B	13
B B B B I B	11
N N N N N N	10
B B N B N B	8

Figures

Figure 1. The sequence items for the pre-definition and post-definition task.

A: 0,1,0,1,0,1,0,1,...

B: 1,4,9,16,25,36,49,64,...

C: $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \dots$

D: 1, -1, 2, -2, 3, -3, 4, -4, ...

E: 3,3,3,3,3,3,3,3,...

F: 1,3,2,4,3,5,4,6,...

G: 6,6,7,7,8,8,9,9,...

H: 0,1,0,2,0,3,0,4,...

I: $10\frac{1}{2}, 10\frac{3}{4}, 10\frac{7}{8}, 10\frac{15}{16}, 10\frac{31}{32}, \dots$

J: -2, -4, -6, -8, -10, ...

Figure 2. The formal definitions given in the presentation phase.

A sequence $\{x_n\}_{n=1}^{\infty}$ is increasing if and only if $\forall n \in \mathbb{N}, x_{n+1} \geq x_n$.

A sequence $\{x_n\}_{n=1}^{\infty}$ is decreasing if and only if $\forall n \in \mathbb{N}, x_{n+1} \leq x_n$.

Figure 3. The normatively correct responses for the reduced list of sequences, giving response profile NNBNIN (note that this would read NNNNIN if the strict interpretation were taken).

A: 0,1,0,1,0,1,0,1,...	N
D: 1,-1,2,-2,3,-3,4,-4,...	N
E: 3,3,3,3,3,3,3,3,...	B
F: 1,3,2,4,3,5,4,6,...	N
G: 6,6,7,7,8,8,9,9,...	I
H: 0,1,0,2,0,3,0,4,...	N