

# The Chasm between Thought Experiment and Formal Proof

David Tall

Mathematics Education Research Centre  
Institute of Education  
University of Warwick  
Coventry CV4 7AL  
e-mail: David.Tall@warwick.ac.uk

*This presentation will address the conceptual demands placed on students attempting to deal with formal proof for the first time and present empirical evidence that reveals the subtlety of this transition. It transpires that there is more than one route to move from informal experience of proof to formal proof. Informal proof often occurs in the style of a thought experiment, using a variety of imagery to infer that, when a certain situation occurs, then another must also occur as a consequence of the first. Formal proof, on the other hand, is based on verbal/symbolic definitions and focuses only on those results that can be deduced logically from the definitions. The presentation will show that there are (at least) two cognitively different routes from informal to formal. One builds on imagery and constantly reconstructs it to fit new formalisms. Another starts from the definitions and develops only those properties that can be built by formal deduction. Empirical evidence will be given, collected in longitudinal studies from students in their first year of university mathematics, to demonstrate how both of these routes can lead to success, but that each involves a different array of cognitive difficulties that can lead to failure. For instance, the imagery of thought experiments may include subtle elements at variance with the formalism that causes serious blockages of understanding. On the other hand, formal proof may also lead to structure theorems which have their own mental images that can then be used in informal thought experiments to predict new directions for the formal theory. The results suggest that different students may benefit from different kinds of teaching strategies and what may help one may be of a hindrance to another.*

## Proof in the Mathematical Community

The meaning of formal proof seems clear to the mathematical community as the final “precising” of a mathematical argument, even though the theorem may be constructed in a variety of ways:

In the fall of 1982, Riyadh, Saudi Arabia ... we all mounted to the roof ... to sit at ease in the starlight. Atiyah and MacLane fell into a discussion, as suited the occasion, about how mathematical research is done. For MacLane it meant getting and understanding the needed definitions, working with them to see what could be calculated and what might be true, to finally come up with new “structure” theorems. For Atiyah, it meant thinking hard about a somewhat vague and uncertain situation, trying to guess what might be found out, and only then finally reaching definitions and the definitive theorems and proofs. This story indicates the ways of doing mathematics can vary sharply, as in this case between the fields of algebra and geometry, while at the end there was full agreement on the final goal: theorems with proofs. Thus differently oriented mathematicians have sharply different ways of thought, but also common standards as to the result. (MacLane, 1994, p. 190–191.)

---

This paper was given as a plenary lecture at the 8. *Internationales Symposium zur Didaktik der Mathematik*, Universität Klagenfurt, Austria, 28<sup>th</sup> Sept – 2<sup>nd</sup> October 1998. It brings together under one theme a number of studies on definition and proof performed at the Mathematics Education Research Centre, University of Warwick, UK.

Saunders MacLane, the formal symbol-manipulator plays with ideas in ways that led to such things as the more abstract idea of category theory through abstracting the essential properties of objects and maps in other theories. Michael Atiyah, on the other hand, uses a wide array of techniques from topology, algebra and geometry, to play around with ideas that he considers important and gets a sense of them before he even proposes definitions and theorems. As his research student I can remember him telling me to seek a problem that had some value and interest for other mathematicians rather than simply writing down an arbitrary set of rules to see what came out of them. For him the system should be pregnant with interesting ideas. I well remember a time when he remarked to me that “a vector bundle is a topological variation of a vector space” just as “a module is an algebraic variation of a vector space”, so one should expect there would be some useful parallels between the theories. This “vague and uncertain situation”, linking two distant ideas—topological vector bundles and algebraic modules—produced new theorems in a new subject called algebraic K-theory.

The notion of “concept image” is relevant here, consisting of “the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes. It is built up over the years through experiences of all kinds, changing as the individual meets new stimuli and matures.” (Tall & Vinner, 1981). The creation of a new mathematical proof requires an appropriate concept image of the situation being considered. It can operate in different ways, for instance the concept image of proof from the algebraist MacLane is different from the concept image of proof from the algebraic geometer Atiyah. One builds more from operating on secure existing systems to build new ones, the other operates in more general problem-solving situations and brings in a variety of tools, creating new definitions where appropriate.

In the opening chapter of *Advanced Mathematical Thinking* (Tall, 1991), I quoted from the perceptive writings of Poincaré who also remarks on different ways of thinking in mathematics:

It is impossible to study the works of the great mathematicians, or even those of the lesser, without noticing and distinguishing two opposite tendencies, or rather two entirely different kinds of minds. The one sort are above all preoccupied with logic; to read their works, one is tempted to believe they have advanced only step by step, after the manner of a Vauban<sup>1</sup> who pushes on his trenches against the place besieged, leaving nothing to chance. The other sort are guided by intuition and at the first stroke make quick but sometimes precarious conquests, like bold cavalymen of the advanced guard.

[Poincaré, 1913 page 210]

Poincaré further supported his arguments by contrasting the work of Weierstrass and Riemann:

---

<sup>1</sup> Sebastien de Vauban (1633-1707) was a French military engineer who revolutionized the art of siege craft and defensive fortifications.

Weierstrass leads everything back to the consideration of series and their analytic transformations; to express it better, he reduces analysis to a sort of prolongation of arithmetic; you may turn through all his books without finding a figure. Riemann, on the contrary, at once calls geometry to his aid; each of his conceptions is an image that no one can forget, once he has caught its meaning. [ibid, page 212]

However, on considering the contrast between analytic reasoning and more general visual imagery, Poincaré considered that logical thinkers too were using their own kind of intuition:

... When one talked to M. Hermite, he never evoked a sensuous image, and yet you soon perceived that the most abstract entities were for him like living beings. He did not see them, but he perceived that they are not an artificial assemblage and that they have some principle of internal unity. [ibid page 220]

The conclusion is inescapable. The intuition used by mathematicians to prove new theorems is the product of the concept images of the individual. The more educated the individual in logical thinking, the more likely his concept imagery will resonate with a logical response:

We then have many kinds of intuition; first, the appeal to the senses and the imagination; next, generalization by induction, copied, so to speak, from the procedures of the experimental sciences; finally we have the intuition of pure number... [ibid., page 215.]

### **Mathematical proof for students**

In introducing students to mathematical proof, it is therefore of interest to see if their individual concept images cause them to approach proof in different ways. Poincaré commented on this also:

... Among our students we notice the same differences; some prefer to treat their problems 'by analysis' others 'by geometry.' The first are incapable of 'seeing in space', the others are quickly tired of long calculations and become perplexed. [Poincaré, 1913, page 212.]

Again we see a suggestion of different tendencies of students colouring the way that they approach mathematics. There are research findings too. Krutetskii (1976, p.178) studied 192 children selected by their teachers as 'very capable' (or 'mathematically gifted'), 'capable', 'average' and 'incapable'. He found a spectrum of performance in which the incapable remembered only incidental irrelevant detail, with long, often erroneous, inflexible solution procedures, whilst the very capable remembered general strategies rather than detail, focussed on essential elements and were able to provide alternative solutions. But amongst the gifted he found a spectrum of performance in which six of the thirty four children studied were classified as "analytic", five as "geometric" and 23 as "harmonic". In other studies, the use of visual methods was found in very few students, most preferring symbolic methods. This is not necessarily inconsistent with Krutetskii, for there is a tendency as students become more proficient symbolically that symbolic methods are preferred. So with university

students we may expect those who are “harmonic” to tend to lean towards symbolic methods, giving a larger group of students using symbols than using visualisation alone.

We therefore have various evidence of different ways in which students approach mathematics. There is a spectrum of success and failure, but there is also another spectrum of cognitive preferences including analytic, geometric and harmonic thinking with perhaps a tendency amongst more advanced students to prefer analytic symbolic thinking as the ideas become more sophisticated.

### Student concept images and previous experience of mathematics

How do these different approaches to mathematics in general relate to approach to proof in particular? It is first helpful to consider the students’ previous experience. When students encounter formal proof they have already developed their own ways of testing whether things are true. In everyday life proof means different things in different contexts. To a judge and jury it means something established by evidence ‘beyond a reasonable doubt’. To a statistician it means something occurring with better than a calculated probability. To a scientist it means something that can be tested by experiment. More generally, simple practical experience gives a belief that “facts” are true. For instance, we “know” that  $2+2$  is 4 by counting, we know that  $2^{10} > 10^3$  because our experience of calculation with decimal numbers allows us to calculate that  $1024 > 1000$ .

This belief in the arithmetical properties of numbers is part of a long apprenticeship in various kinds of mathematics which are encountered before meeting formal proof. Mathematical learning begins with the young child interacting with the external world, perceiving things and acting on them.

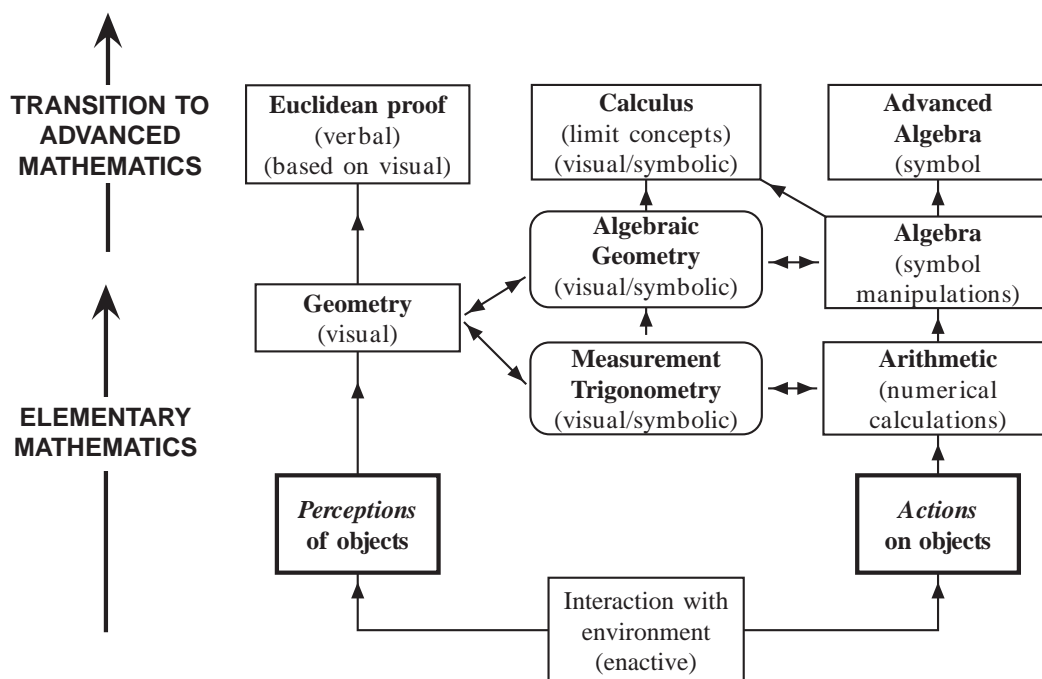


Figure 1: Development of major mathematical themes

Building from perception of objects is a strand studying shape and space which introduces pictorial representations and verbal descriptions in geometry and builds to the verbal form of proof found in Euclidean geometry. Building from actions on objects (counting, sorting, ordering etc) builds the use of symbols used for calculation in arithmetic and manipulation in algebra. Straddling these two are linking topics such as measurement, trigonometry and algebraic geometry.

As I have suggested elsewhere (Tall, in press, a), there are severe discontinuities in what appears to be a coherent and consistent development in the use of symbols. The symbols used in arithmetic, algebra, calculus and formal theory have quite different characteristics (figure 2).

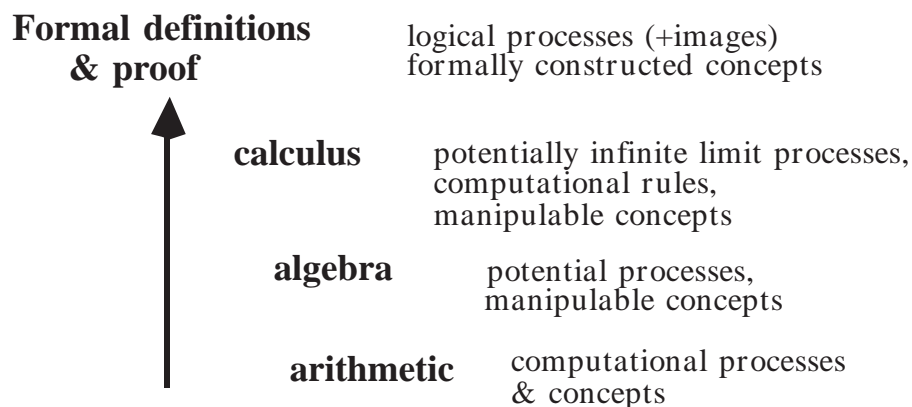


Figure 2: different kinds of characteristics of processes and concepts in selected topics

In arithmetic of whole numbers the operations are always designed to produce a result. This leads to the *computational* use of symbols. In algebra, symbolic expressions such as  $2+3x$  can be evaluated only when arithmetic values are substituted for algebraic variables. Thus algebraic expressions can be considered as *potential processes* and *manipulable concepts*. In the calculus the limit concept causes even more difficulty because it expresses the dynamic idea of potentially getting as close as is required to a limit value: the process of tending to a limit is a *potentially infinite process*. Faced with such a fundamentally new concept, it is no wonder that students prefer the finite operations of the rules of differentiation and the finite tests of convergence to the theoretical consideration of the infinite notion of limit. These will cause further difficulties in formal proof where the concepts are formally constructed by deduction from definitions using logical processes.

As we consider the major themes which arise in elementary mathematics, we may consider where proof occurs (See Tall, in press, b, for an extended discussion). Proof arises quite naturally in geometry. Here there are intuitive ideas of space and shape that are defined in terms of verbal definitions. It is a natural consequence to see if the definitions not only describe the mental geometric objects but can also be used to show that one definition implies another. For instance, that a figure described as a triangle with two equal sides must also be a triangle with two equal angles.

Proof in geometry is a verbal exercise based on underlying visual images. It is interesting to note that such an exercise is *generic*, meaning that the picture of the situation, whether drawn or imagined, represents not just a single case but stands as a representative for the class of all possible figures described by the given verbal description. To generate such a proof involves a *thought experiment*, in which one imagines the conditions of the theorem holding and attempts to “see” if the conclusion follows. It is turned into a Euclidean proof by certain conventions concerning the congruence of triangles which are used to link what is known to what is required. Because it is possible to build up a systematic theory of propositions each deduced either from the definitions or from previous propositions, Euclidean proof supplies part of the experience for formal proof in advanced mathematical thinking. But there are other areas for which it is less appropriate. It gives little experience of dealing formally with set-theoretic defined axioms and the use of propositional logic.

The other major theme of symbolic computation and manipulation uses proof in generalised arithmetic (algebraic manipulation) to describe generalities in arithmetic. For instance one may show that the square of any whole number is one more than the product of the number above and the number below through the algebraic manipulation

$$(n-1)(n+1)+1 = n^2-1^2+1 = n^2.$$

The link between symbols and visualisations also gives an opportunity for demonstrations of proof. For instance, the fact that  $m \times n = n \times m$  for two whole number may be imagined by looking at an array of objects either as  $m$  rows of  $n$  or  $n$  columns of  $m$ . Once again this is a *generic proof*, in that the physical picture actually has a specific number of rows and columns but can be imagined as having “any number” of rows and columns:

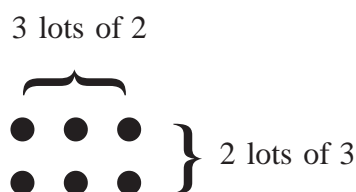


Figure 3: A graphic proof that  $3 \times 2$  is the same as  $2 \times 3$  which works generically for all whole numbers

The student beginning to study formal proof therefore has concept imagery developed in a wide variety of ways, including:

- “knowing” that arithmetic facts are “true” from experience,
- various everyday notions of proof,
- proof by generic example,
- euclidean verbal proof (with underlying generic visual examples),
- algebraic computations to represent general calculations.

All of these involve either symbolic activity in arithmetic or algebra, or some kind of “thought experiment” where one draws or imagines a typical situation satisfying the assumed properties to see if the required conclusion holds.

When formal proof is introduced, it is not exactly like any of these. It is supposed to be performed without using a picture for anything other than organisational support. It starts with explicit definitions and moves through logical deductions to a desired conclusion. To cope with the details, these sophisticated requirements are lessened. Every detail need not be included, all that is required is that it is evident to the reader that the details could be filled in as required. The sense of what should or should not be included in a proof is a cause for great concern by students (Nardi, 199?)

### Generic proof and formal proof: The irrationality of $\sqrt{2}$

One of the first serious proofs presented to students is the proof that  $\sqrt{2}$  is irrational. The standard proof is by contradiction, represented by the plan:

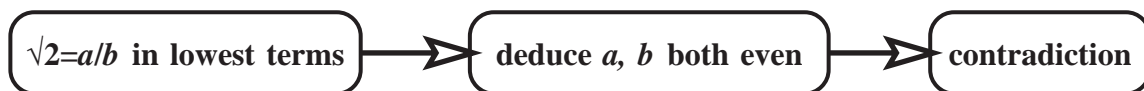


Figure 4: Outline proof of the irrationality of  $\sqrt{2}$

The detail that needs to be included is rather more subtle than may be obvious at a first consideration. Figure 5 shows some of the steps that may, or may not be considered in thinking through the proof, deduced from interviews with university mathematics students (Barnard & Tall, 1997). Solid arrows show straightforward decisions, curly arrows are less obvious, and those marked in grey may be less explicit.

It transpires that students can often remember the general plan of the proof, but are less secure on detail. Student S (who had achieved an A-grade in mathematics in the advanced level school examinations) said:

I'd take the case where I assumed it was a rational and fiddle around with the numbers, squaring, and try to show that ... if it was rational then you'd get the two ratios  $a$  and  $b$  both being even so they could be subdivided further, which we'd assumed earlier on couldn't be true so our assumption it was rational can't be true. (Barnard & Tall, 1997)

Yet, when he tried to explain the steps, he remarked:

... [the lecturer] did some fancy algebra which I couldn't actually reproduce.

When asked to reproduce it, he wrote:

$$\left(\frac{a}{b}\right)^2 = 2,$$

$$a^2 = 4b^2,$$

saying,

I think that's what he did, but he did it in one step whereas normally I would've taken two. (Student S)

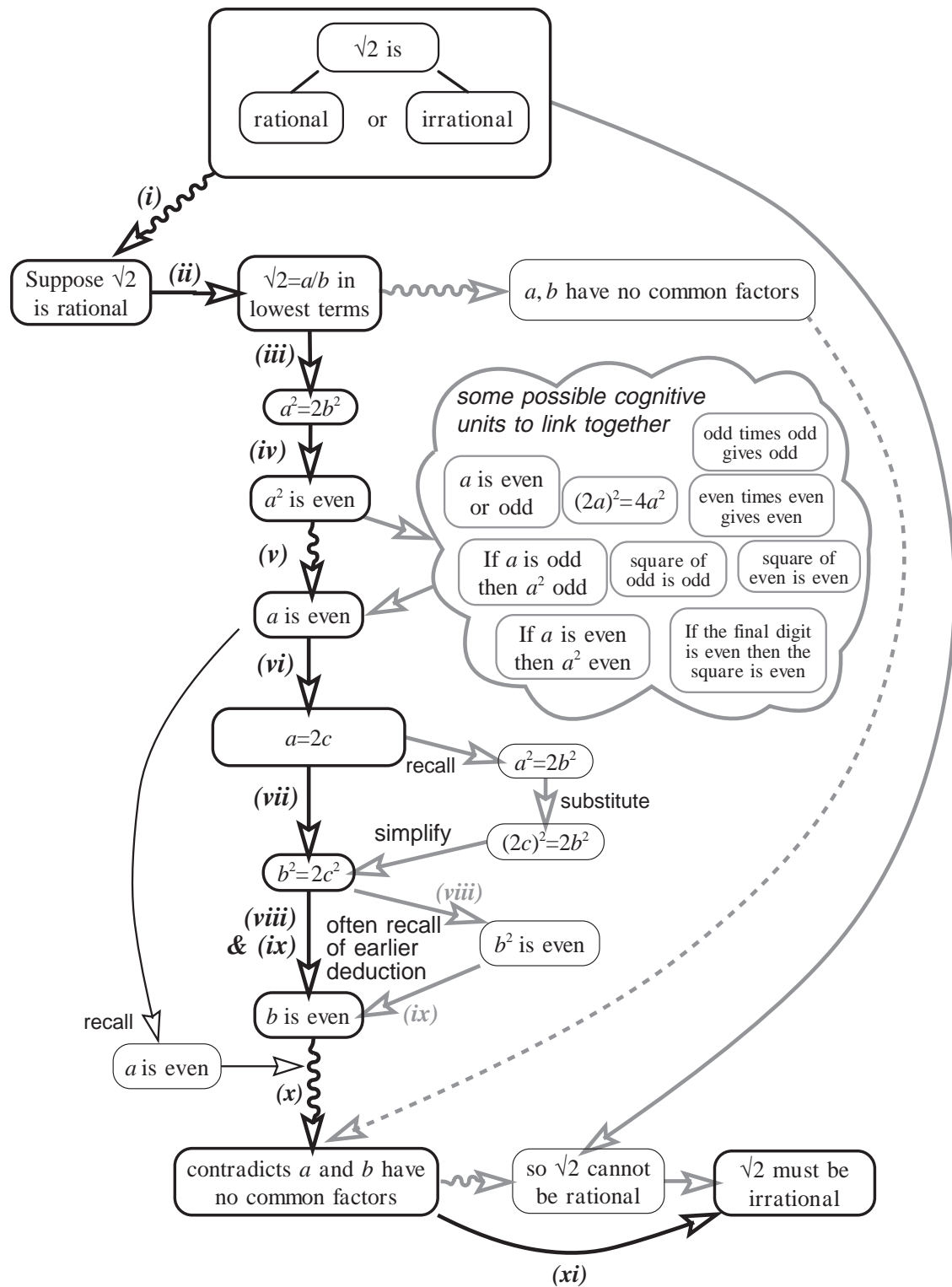


Figure 5 : Cognitive steps which may be used in the proof that  $\sqrt{2}$  is irrational.

When asked to fill in the details, he eventually obtained the correct result  $a^2=2b^2$ . Such difficulties occurred with several students interviewed, however, the majority completed the manipulation routinely in a single step and were able to give further details on request.

More serious difficulties arise with the step:



$a^2$  is even implies  $a$  is even,

which essentially involves a proof by contradiction nested within the overall contradiction proof. Again student S could remember the step but not the argument, citing the authority of the lecturer that  $a$  is even because “the root of an even number is even, he just assumed it.” When asked to give a proof, he simply responded with specific cases such as the square root of 4 and the square root of 16 both being even. He “knew” that a number and its square were both even *at the same time*, but it did not seem apparent that one fact was a *consequence* of the other.

A serious problem with this contradiction proof is that it conflicts with previous experience. Instead of carrying out a sequence of calculations to get a specific answer, it is necessary to hold two mutually exclusive possibilities in mind at once and show that if one were false then a contradiction ensues. The idea that one assumes something to be false when one “knows” it is true is made more difficult by having one contradiction argument nested within another. Once these difficulties are encountered, many students continue to consider the proof, committing the strategy, and perhaps the details, to memory.

Tall (1979) found that students were much happier with a generic thought experiment, which involves imagining given fraction, factorising numerator and denominator and squaring to find that each prime factor occurs an even number of times. For instance,

$$\frac{9}{40} = \frac{3^2}{2^3 \times 5}$$

and the squaring of this number *doubles* the number of each prime factor to give

$$\left(\frac{9}{40}\right)^2 = \frac{3^2}{2^3 \times 5} \times \frac{3^2}{2^3 \times 5} = \frac{3^4}{2^6 \times 5^2}$$

so the primes occurring in the factorization of numerator and denominator of a square number all occur an even number of times. Hence the square of any fraction cannot equal 2 which factorizes as  $2/1$  and has an *odd* number of 2's in the numerator.

Students in the first year of university expressed a strong preference for the generic proof over the standard proof by contradiction. Although those who were already familiar with the standard proof might prefer it, the following day several of these changed their preference when they saw the explanatory power of the generic proof. None of those who preferred the generic proof first changed to prefer the standard proof. The generic proof was preferred even more when it came to proving the generalisation that  $\sqrt[3]{(5/8)}$  is irrational. This follows using the original generic idea because  $5/8=5/2^3$  fails to have an even number of each factor, whereas the standard proof uses the terms “even” and “odd” which many students fail to generalise appropriately.

Given the power of generic proof, it may be advantageous for a part of the university population to use it in preference to a formal proof of this result. Indeed, with students whose long-term intentions are practical or applicable rather than theoretical, this may be a sensible tactic. But it does not mean that the generic proof is preferable in the long term for those who are going to study more formal mathematics where proof by contradiction late proves to be an indispensable tool.

Students meeting formal mathematics are often faced with a difficult transition from the computation and thought experiment of elementary mathematics to the definition and deduction of formal mathematics. How do they cope with the difficulties?

### **The transition from thought experiment to formal proof**

As we have seen, students arrive at the point of studying formal proof when their previous experience tells them that certain truths are “known”, that generic arguments give insight into what is happening, and that much of their mathematics involves specific calculations to obtain answers.

Formal proof is quite different. It requires logical deduction from definitions. One might think that this is obtained by a sequence of activities such as:

1. become acquainted with the definition
2. use the definition to deduce results
3. use the results in further theorems to build up systematic theories

which may be summarised under the successive headings:

- 1 DEFINITIONS,
- 2 DEDUCTIONS,
- 3 SYSTEMATIC THEORY.

However, the cognitive processes involved are more intimately interconnected. For instance, to truly understand the nature of a definition requires the use of deductions to see the implications of that definition. Before the notion of a systematic theory can be fully realised, there is an important interplay of the form

DEFINITIONS  $\longleftrightarrow$  DEDUCTIONS

To take account of this observation, Bills & Tall (1998) defined:

A (mathematical) definition or theorem is said to be *formally operable* for a given individual if that individual is able to use it in creating or (meaningfully) reproducing a formal argument.

Following students through a twenty week course and interviewing selected individuals at regular intervals every three weeks shows that many students do not make their definitions operable in any formal sense. Even students who succeeded might work with a vague impression of the definition for a time

before crystallising it into an operable definition. For instance, the class were given the formal definition of “least upper bound” as follows:

An *upper bound* for a subset  $A \subset \mathbb{R}$  is a number  $K \in \mathbb{R}$  such that  $a \leq K \forall a \in A$ .

A number  $L \in \mathbb{R}$  is a *least upper bound* if  $L$  is an upper bound and each upper bound  $K$  satisfies  $L \leq K$ .

Sean (the same student as Student S mentioned previously) explained this in his own terms:

*Sean (interview 1):* The supremum of a set is the highest number in the set.

*Interview 2:* [you get the supremum by] looking at all the elements of the set to find out which is the greatest and choosing that number. I always have trouble remembering whether the supremum has to be in the set.

*Interview 4:* It’s the greatest number of ... it’s a number that’s bigger than all the numbers in the set.

*Interview 5:* The set  $\{1, 2, 3\}$  has upper bound 3. [Is 7 an upper bound?] No, it’s not in the set. (Bills & Tall, 1998)

When in the second interview I tried to make him reflect on his ideas by asking him for the least upper bound of the set  $S$  of real numbers  $x$  where  $x < 1$ , he suggested:

[the least upper bound is] a very small number subtracted from one ... nought point nine, nine, nine recurring.

He felt vindicated because he “knew” that “nought point nine recurring” is *less than one*. Thus a personal concept image which suggests there are variable quantities which are essentially “infinitesimally close” (Cornu, 1991) caused a roll-on difficulty with the notion of least upper bound. He was quite articulate about his struggle:

... when we have theorems in analysis lectures, stuff like supremums are just the basic workings; since I can only just understand these individually, one of these basic foundations, I can’t look at all of them together and understand the theorem. (Bills & Tall 1998)

Even successful students did not necessarily memorise the definition at first. In her second interview, Lucy was able to verbalise the definition of least upper bound in a manner close to the symbolic form:

*Interviewer:* If I asked you what was a least upper bound what would you say ...?

*Lucy:* Well for a start it has to be an upper bound.

*Interviewer:* Right so what does that mean?

*Lucy:* An upper bound for a set  $S$ , if you take any element of  $S$  to be  $a$ , say, and for all  $a$  you can find, say the upper bound was  $k$ , for all  $a$ ,  $k$  will be greater than or equal to  $a$  for any number in that set.

*Interviewer:* So that’s the definition for  $k$  being an upper bound.

*Lucy:* ... and the least upper bound is also an upper bound but it’s the least of all the upper bounds so  $l$  has to be less than or equal to  $k$  for all  $k$  greater than or equal to  $a$ .

In the fourth interview she is very confident expressing the definition verbally:

Well it's got to be an upper bound itself and it's got to be the least of all the upper bounds.

But even in the fifth interview, when asked to write down the definition of the least upper bound of a non-empty set  $S \subset \mathbb{R}$ , she wrote:

$\forall s \in S, s \leq \mu$  [saying “ $\mu$  is an upper bound”]

$\forall k \in \mathbb{R}$  s. t.  $s \leq k$  and  $\mu \leq k$ .

After a discussion she modified the last part to “ $\forall s \in S, s \leq k \Rightarrow \mu \leq k$ .” We thus see that students can build notions of proof without the definitions being fully committed to memory. It is possible for definitions to begin to be formally operable before they are fully stable concepts for the learner.

### **Concept Image and Deduction of Structure Theorems**

The developing of a formal theory is paradoxically built on informal experience. Under the veneer of logic there needs to be some kind of mental structure that guides the formal proof. This, I suggest is the concept image. What the individual must do is to use their concept image of a given situation to suggest what might be true and then attempt to convert this intuition into formal deductions. The concepts that are built up formally are termed the “formal (concept) image. The formal image is part of the whole concept image. However, one must distinguish this carefully from other intuitions which may or may not be relevant to the formal image.

In the opening quotation, we saw Maclane talking about “structure theorems” which are used as building blocks in a formal theory. These have not only a formal role, but also a cognitive one. Once a structure theorem has been satisfactorily proved, we have seen that it can take its place in the concept image, now reinforced in the knowledge that it is part of a growing formal theory.

There are many examples in mathematics. For instance, having had experience of numbers on the number line, one may formulate the system as a complete ordered field. A structure theorem says that (up to isomorphism) there is precisely *one* ordered field. This means that such a number system can be represented as a number line with decimal representations as usual. In linear algebra, experience of vectors in two and three dimensional space can be extended by writing down the definitions of a vector space over the real numbers. Here the powerful structure theorem is that any *finite-dimensional vector space* is essentially isomorphic to  $\mathbf{R}^n$ . In group theory, the computational example of permutations of a finite set may be extended to the definition of a group. The structure theorem here is that any theoretical finite group is a subgroup of the familiar group of permutations.

In this way we see that a structure theorem constructs a formal image for the defined concept which can then be taken as a firm cognitive and logical basis for further formal concept building.

There is a subtle problem here. During the building of a formal image (and after), the individual has two kinds of mental image in the cognitive structure: the broad experiential concept image and the subset consisting of the formal image which relates only to the defined concept. How does the individual manage to distinguish between concept image and formal image?

A short answer is: most mathematicians don't! For instance, if the natural numbers are defined formally through the Peano postulates, it is possible to build a huge array of results from the definitions alone. For instance, we might define addition using the Peano postulates and show that  $2+2=4$ . But in making subsequent calculations, do we always prove things from first principles? Life is too short. We know  $10^3 < 2^{10}$  by using the arithmetic we learned in school. Once we know that there is essentially only *one* set of natural numbers satisfying the Peano postulates, then we fall back on our everyday arithmetic, knowing it will be the same as the (isomorphic) theoretical concept. In this way, though we wish technically to prove everything in an appropriate way, the need for suitable economy of style means that we do not prove *everything*. Once we have confidence in a given formal situation, we are concerned only that a result *could* be proved in detail. A proof establishes this to the satisfaction of other mathematicians. But how do students cope with the idea of "proving what satisfies others" when they initially lack the sophistication to share in the culture of the mathematical community?

### **Students giving and extracting meaning**

To gain more insight into the process of student understanding of formal proof, Marcia Pinto (1998) followed the process of studying the development of students notion of proof by interviewing selected students at regular intervals during their first analysis course. However, rather than simply following the classifications that have occurred in other research, she decided to follow the method of "grounded theory" formulated by Strauss, 1987 (see also Strauss & Corbin, 1990). Here data is collected under headings that arise from an ongoing analysis which are modified according to the data obtained. She selected students to represent a wide class of ability (including some of the most able mathematics students in the University of Warwick), to give a spectrum of success and failure, by giving a pre-test and selecting individuals who performed exceptionally well, average, and below-average. She also took into account students who seemed flexible (or harmonic in Krutetskii's formulation) and others who seemed more procedural.

To begin her research she laid down the initial headings mentioned earlier:

- DEFINITIONS,
- DEDUCTIONS,

- SYSTEMATIC THEORY.

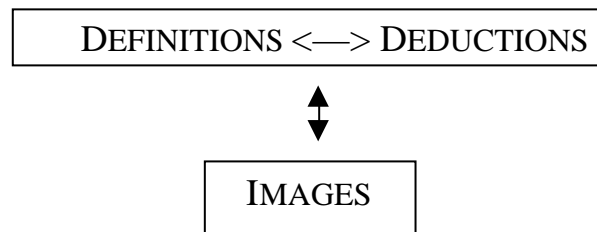
The first group of students were following a course of teacher training for primary children. All had passed their school mathematics Advanced Level exam at a grade D or above. However, mathematics was only part of a broader course which included a wide range of modules on education and practical teaching in mathematics and other subjects. Pinto found that the proposed headings were totally inappropriate for this group. Only two out of twenty were able to give a formal definition of the concept of limit, the remaining eighteen working from an informal concept image. Formal deductions were therefore almost non-existent, and those deductions that occurred tended to be of an informal justification referring to a special case. Pinto therefore revised her headings to focus on

- DEFINITIONS,
- DEDUCTIONS,
- MISCONCEPTIONS.

When it was realised that the third of these had a negative tone to it, she made a third modification and focused on:

- DEFINITIONS,
- DEDUCTIONS,
- IMAGES.

The first two headings were analysed in turn with each being related to underlying concept images as follows:



### **Students building operable definitions and corresponding deductions**

There is not a single way of building a coherent definition, given that this is built upon individual concept images. Pinto (1998) found the data suggesting two recurring possibilities:

- *giving meaning* to the concept definition from concept imagery,
- *extracting meaning* from the concept definition through using it to make formal deductions.

These are not fully analogous to different mathematicians' approaches in creating new theories because the students were given the definitions as a starting point. However, there are certain parallels. The *giving* of meaning

involved using all kinds of personal clues to enrich the definition with examples that could suggest intuitions, in a synthetic way, building up like the visualising geometer. The *extracting* of meaning involved first a routinising of the definition, perhaps saying it over until it became possible to consider it as a workable definition that could be used to give a formal development of the mathematics. Pinto built up her knowledge of the particular students' work by contrasting selected students in pairs.

First, consider two successful students, Ross, who was considered a formal *extractor* of meaning and Chris, a conceptual *giver* of meaning.

Ross wrote down the definition as follows (Pinto, 1996):

$$\begin{aligned} & \text{A sequence } (a_n) \text{ tends to limit } L \text{ if, } \forall \epsilon > 0, \exists N(\epsilon) \in \mathbb{N} \\ & \text{s.t. } \forall n \geq N; \\ & |a_n - L| < \epsilon. \end{aligned}$$

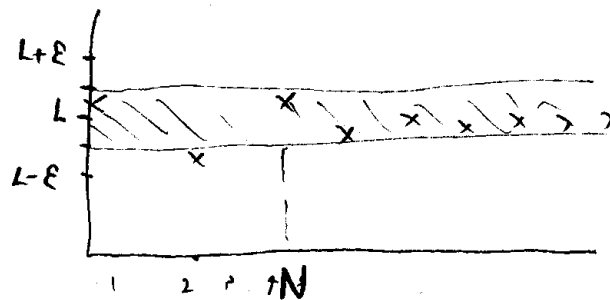
(Ross, first interview)

He explained that he coped by:

“Just memorising it, well it’s mostly that we have written it down quite a few times in lectures and then whenever I do a question I try to write down the definition and just by writing it down over and over again it get imprinted and then I remember it.”

(Ross, first interview)

He also drew pictures and attempted to relate them to his logical ideas, although underlying even the pictures was a sense that there is a changing term handled symbolically.



(Ross, first interview)

He explained:

“Well, before, I mean before I saw anyone draw that, it was just umm ... thinking basically as  $n$  gets larger than  $N$ ,  $a_n$  is going to get closer to  $L$ , so that the difference between them is going to come very small and basically, whatever value you try to make it smaller than, if you go far enough out then the gap between them is going to be smaller. That’s what I thought before seeing the diagrams ... something like that.”

(Ross, first interview)

Three weeks later, when asked to write down what was meant to state that a sequence was *not* convergent to a limit  $L$ , he first wrote down:

$$\forall \epsilon > 0, \exists N(\epsilon) \in \mathbb{N} \text{ st. } \forall n \geq N.$$

$$|a_n - L| < \epsilon.$$

(Ross, second interview)

then negated it by using the formal negation of quantifiers (passing a “not” over a universal or existential quantifier changes one to the other), to give:

$$\forall L, \exists \epsilon > 0 \text{ st. } \forall N(\epsilon) \in \mathbb{N} \exists n \geq N, \text{ st.}$$

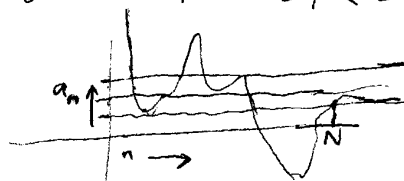
$$|a_n - L| \geq \epsilon$$

(Ross, second interview)

Note how he introduces  $\forall L$  in the second statement, corresponding to the implicit unwritten  $\exists L$  in the first statement. But despite this subtle understanding of the original definition and its formal negation, there is an error in the negation. By writing “ $\forall \epsilon > 0, \exists N(\epsilon)$ ” in the original definition to indicate that  $N$  depends on  $\epsilon$ , he then erroneously wrote the negation as “ $\exists \epsilon > 0, \forall N(\epsilon)$ ” but now  $N$  does *not* depend on  $\epsilon$ . Had he simply written “ $\forall \epsilon > 0, \exists N$ ”, then the portion of the negation “ $\exists \epsilon > 0, \forall N$ ” would have been satisfactory, by default. In other words, by giving *additional* meaning to the definition, he exposed an error in the routine negation that would not have been visible had he written out the definition in its basic form. Such a problem proved easy to remedy by asking him to think it through, which led to him being able to self-correct his error.

Chris used his imagery to support his ideas, writing:

~~If  $a_n \rightarrow L$  then there exists~~  
 For all  $\epsilon > 0$ , there exists  $N \in \mathbb{N}$   
 such that  $|a_n - L| < \epsilon$  for all  $n \geq N$



(Chris, first interview)

As he drew the diagram, he motioned with his hands to suggest the ideas underlying the definition:

“I don’t memorise that [the definition of limit]. I think of this [picture] every time I work it out, and then you just get used to it. I can nearly write that straight down.”

”I think of it graphically ... you got a graph there and the function there, and I think that it’s got the limit there ... and then  $\epsilon$  once like that, and you can draw along and then all the ... points after  $N$  are inside of those bounds. ... When I



first thought of this, it was hard to understand, so I thought of it like that's the  $n$  going across there and that's  $a_n$ . ... Err this shouldn't really be a graph, it should be points.”  
 (Chris, first interview)

He too made an error (drawing a curve instead of points for the values of the sequence), but was able to self-correct using his wider cognitive connections. Throughout the whole course he seemed to be negotiating with the ideas. He believed that a formal proof must include *all* the logical steps to build from the given assumptions, not even allowing the quotation of results previously established, because the “proof” must be complete in itself. It was not until the eighth week of the course that he gave up this ideal when his proofs became interminably long. He also played with other forms of the definition, for instance seeing if increasing the value of  $N$  forced  $\epsilon$  to be small would be a better definition before settling on the formal definition as a matter of preference. It seemed as if he enjoyed the tension of the challenge. He had clearly faced mathematical challenges before and felt the thrill of success. Now he was maintained on an emotional high level which seemed to give him pleasure even when under stress. He always looked for clarity and precise ideas, and surmounted errors by refocusing his attack.

When it came to negating the definition, he first clarified the issue by asking “did you mean *does not tend to a limit L* or does not tend to *any limit*” and then thought through the whole thing meaningfully, writing:

A sequence  $(a_n)$  does not tend to a limit if  
 for any  $L$ , there exists  $\epsilon > 0$  such that  
 $|a_n - L| \geq \epsilon$  whenever  ~~$n \geq N$~~ , for  
 for some  $n \geq N$  for all  $N \in \mathbb{N}$

(Chris, second interview)

This is an exceptional feat of thinking that was not found in many students. It is more demanding than the formal method of negating quantifiers used by Ross. Of 250 students majoring in mathematics asked to describe how they remembered the definition of limit later in the course, *only five mentioned the use of a picture*. Even if pictures are used in the first place (which may very well have happened with more than the five who later recalled it), it seems that the visual approach is largely supplanted by the symbolic definition when the latter becomes operable.

### Less successful students

A large number of students cannot cope with the definition of limit. Robin tried to remember it:

“It's just *memorising* the exact form of it, being the actual idea sort of *understandable* ... which is saying...”

(Robin, first interview)

However, he remembers it inaccurately (Pinto, 1996):

A sequence  $(a_n)$  tends to a limit  $L$  <sup>for</sup>  $\forall \epsilon > 0$  if there exists  $N \in \mathbb{N}$  s.t.  $|a_n - L| < \epsilon$  provided  $n > N$ .  
 (Robin, first interview)

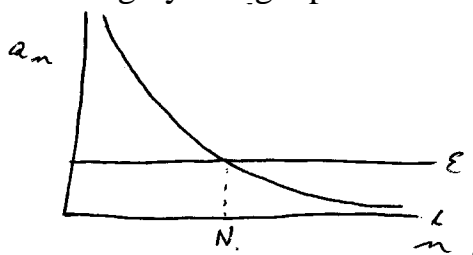
He tries to operate with the definition by giving specific values to the variables, however, instead of starting with a value for  $\epsilon$ , he begins with  $N=5$  and does not speak of any relationship between  $N$  and  $\epsilon$ . He cannot grasp the definition sufficiently to make sense of it (see Barnard & Tall, 1997, for more detailed discussion).

Colin also could not write down the definition successfully:

If  $a_n \rightarrow l$ , then there exists  $\epsilon > 0$ , such that  $|a_n - l| < \epsilon$  for all  $n \geq N$ , where  $N$  is a large positive integer.

(Colin, first interview)

He tried to give the idea meaning by using a picture to support his thinking:



Unfortunately, this involved a restricted concept image of a decreasing function or sequence which proved inadequate to give him a broad enough image to support the full notion of limit. Although he denoted the limit by  $l$ , he wrote  $\epsilon$  instead of  $l+\epsilon$  and considered the limit as a lower bound (a common concept image, see for example, Cornu, 1981, 1991). He explained:

“.. umm, I sort of imagine the curve just coming down like this and dipping below a point which is  $\epsilon$  ... and this would be  $N$ . So as soon as they dip below this point then ... the terms bigger than this [pointing from  $N$  to the right] tend to a certain limit, if you make this small enough [pointing to the value of  $\epsilon$ ].”

### Unsuccessful Negation

Neither of these students could cope with the formal idea of a non-convergent sequence. For instance, Robin wrote:

“A sequence  $a_n$  does not tend to the limit  $L$  if for any  $\epsilon > 0$ , there exists a positive integer  $N$  s.t.  $|a_n - L| > \epsilon$ , whenever  $n \geq N$ .”

(Robin, second interview)

The original quantifiers for the definition of limit here remain unchanged, and all that is changed is the inner inequality  $|a_n - L| < \epsilon$  incorrectly negated to give  $|a_n - L| > \epsilon$ . He is unable to treat the whole definition as a meaningful cognitive unit, and simply focuses on the inner statement as something which he can attempt to handle.

The other student, Colin, said:

“Umm ... I would just say there doesn't exist a positive integer because we can't work it out ... no ... you cannot find an integer  $N$  ...”.

and wrote

There exists a term where  $|a_n - L| \geq \epsilon$  where  $n \geq N$ , where  $N$  is a positive integer.

(Colin, second interview)

Were these less successful students givers or extractors of meaning?

### Students failing to get beyond their concept images

Other students' remembrances of the definition of limit of a sequence (two weeks after encountering it) reveal a collage of isolated ideas (Tall, 1996):

(a) Given  $\epsilon > 0$ ,  $\exists$  a  $\delta$  s.t. ....

(b) given  $\epsilon > 0 \exists l \in \mathbb{R}$  s.t.  $|l - x_n| \leq \epsilon$

(c)  $\left\{ \begin{array}{l} \forall \epsilon > 0 \quad \lim_{n \rightarrow \infty} S_n = l \\ \text{series } \text{---} \text{ tends to a limit} \\ \epsilon + \lambda \quad \epsilon - \lambda \\ S_n \rightarrow \text{limit s.t. } \epsilon + \lambda, \epsilon - \lambda \\ \text{where } \lambda \end{array} \right\}$

(d)  $\lim_{n \rightarrow \infty} S_n = l$

as  $n \rightarrow \infty$  the terms approach and get closer + closer to  $l$   
(or may reach it) but  $l$  is not exceeded

i.e.  $S_n - S_{n+1} \rightarrow 0$  as  $n \rightarrow \infty$

In these cases the students remember isolated facets from the definition and from their concept image. The first remembers something about the limit  $l$ , then changes the symbol to a  $\delta$ , relating to the definition of limit of a real function rather than limit of a sequence. The second remembers  $x_n$  getting within  $\epsilon$  of the limit,  $|l - x_n| < \epsilon$ , but fails to remember the role of  $N$ . The third evokes the term

“series” (which is more recent in their experience) than sequence, remembers the range of values “something  $\pm$  something” but gets it as  $\epsilon \pm \lambda$  rather than  $l \pm \epsilon$ . The fourth remembers dynamic imagery, including a warning that the sequence *can* reach the limit, yet has an image of an increasing sequence which does not exceed the limit, and adds a further fact that  $s_{n+1} - s_n$  tends to zero, which is true but not part of the definition.

Bewilderment shines throughout these responses. In the exam the students pick up marks by using procedural methods for proving convergence of series involving computations which make them feel comfortable, but few have any conception of the role of definition. Some find the world of formal analysis at variance with their real-world experiences. Some training to be teachers daily teach young children by example and see no relevance for esoteric proof in their future profession. In interviews remarks such as the following arise:

I can do examples with numbers and things like that but I can't do things with definitions. I just don't know what they are about.

I mean it's not as if these things are real. You need to swot them up to pass exams but you are never going to use them again.

When I try to prove things in my teaching, I show examples and do it in particular cases. In school I'll never have to teach stuff like this.

I work at the example sheets but after a while I get so mad, all I want to do is throw the papers all over the kitchen.

There is a huge gulf between those who develop formal mathematical techniques in which definitions are used to deduce other properties of the defined concept, and those more rooted in real-world practice, where a definition involves saying enough about a particular concept to enable someone else to be able to identify it. The two ends of the spectrum are illustrated by the reactions of two students on first being given the thirteen axioms for the real numbers. When asked why they thought the lecturer had introduced the axioms, Caroline, a mathematician in the making said:

Well, when you prove things properly you need to say exactly where to start, what it is you are assuming, and that is what the axioms are for.

But Martin, who later gained a good degree in economics was bemused, saying:

I dunno really. I've seen most of it before. I knew most of this stuff when I was about five.

He lives in the real world where his pragmatic grasp of economics will probably earn him a higher salary than a research mathematician and he has no conception of the world of formal definitions distinct from his concept imagery.

## Summary

In this paper we began by noting that mathematicians use different cognitive techniques to generate new theorems. Some work carefully with formal definitions, carefully extracting meaning from them by working with them and gaining a symbolic intuition into theorems which may be true and can be

proved. Some have a wider notion of problem-solving which builds up ideas, developing new concepts that may be true before even before making appropriate definitions to form a basis for the theory. In all cases they have a concept image built up from their previous experience which underpins the proof process.

Students learning mathematics have a different problem. They come from elementary mathematics, deeply ingrained in the computation of arithmetic and the symbol manipulation of algebra which usually involve standard algorithms to solve certain types of problems. The forms of proof at this level (often given other names such as “demonstration” or “justification”) usually either involve algebra to give a symbolic description for a general arithmetic statement, or some kind of thought experiment focussing on a “typical” or “generic” case.

The transition from elementary mathematics to formal proof is a huge chasm for many students. The underlying concept image is unable to sustain the formalism. Many students who fail to appreciate the formalism (including the majority of those in our sample preparing to teach primary mathematics) have informal images which dominate their thinking. The definitions are so complex that some may cope with part of the structure but not all.

Success comes in (at least) two ways, either *giving* meaning by working from the concept image, or *extracting* meaning by working formally with the definition. These two techniques can both be successful and unsuccessful. For the successful student, *giving meaning* involves constantly working on various images, reconstructing ideas so that they support the formal theory. The successful student who extracts meaning from the definition has a different task of building up a formal image based mainly on the proof activities themselves.

Those who fail to cope at all with formal proof who try to give meaning from their concept imagery may be able to imagine thought experiments which give generic proofs with an overall feeling for some of the ideas, though failing to understand the proof at all. Those who try to extract meaning fail to be able to cope with the complexity of the definitions and may be totally confused. A fall-back strategy to attempt to pass exams is to learn proofs by rote.

The teaching and learning of formal proof remains an important component of theory building in advanced mathematical thinking. As such it still needs to be introduced in an appropriate theoretical form for future mathematicians. We note here however, that *giving meaning* from concept images requires ongoing reconstruction of ones ideas to focus on the essential properties of the definition. On the other hand, *extracting meaning* by formal deduction, involves considerable new cognitive stress which happens in an entirely different way. This indicates that it may not always be possible to deal with these different kinds of approach with a single teaching method. The most serious problem is that so many students who intend to teach mathematics, even though they may not be required to teach the ideas of formal proof, are unable to appreciate the full extent of the subject which they pass on to the next generation.

## References

- Barnard, A. D. & Tall, D. O. (1997). Cognitive units, connections and mathematical proof. *Proceedings of the Twenty first International Conference for the Psychology of Mathematics Education*, Lahti, Finland, 2, 41–48.
- Bills, L. & Tall, D. O. (1998). Operable Definitions in Advanced Mathematics: The case of the Least Upper Bound, *Proceedings of PME 22*, Stellenbosch, South Africa, 2, 104–111.
- Cornu, B. (1981). Apprentissage de la notion de limite: modèles spontanés et modèles propres, *Actes du Cinquième Colloque du Groupe Internationale PME*, Grenoble, 322–326.
- Cornu, B. (1991). Limits. In D. O. Tall (Ed.). *Advanced Mathematical Thinking* (pp. 253–166). Dordrecht: Kluwer.
- Krutetskii, V. A. (1976). *The psychology of mathematical abilities in schoolchildren* (J. Teller, Trans.). Kilpatrick, J. & Wirszup, I. (Eds). University of Chicago Press, Chicago.
- MacLane, S. (1994). Responses to Theoretical Mathematics, *Bulletin (new series) of the American Mathematical Society*, 30, 2, 190–191.
- Nardi, E. (1996). Tensions in the novice mathematicians' induction to mathematical abstraction. *Proceedings of the Twentieth International Conference for the Psychology of Mathematics Education*, Valencia, Spain, IV, 51–58.
- Pinto (1998) Unpublished PhD Thesis on Students conceptions of definitions and proof, Warwick University.
- Pinto, M. M. F. (1996). Students' use of quantifiers. Paper presented to the Advanced Mathematical Thinking Working Group at *The Twentieth Conference of the International Group for the Psychology of Mathematics Education*, Valencia, Spain, July 1996, (available from [david.tall@warwick.ac.uk](mailto:david.tall@warwick.ac.uk)).
- Poincaré, H. (1913), *The Foundations of Science* (translated by Halsted G.B.), The Science Press, New York (page references as in University Press of America edition, 1982).
- Strauss, A. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.
- Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. London, UK: SAGE Publications.
- Tall, D. O. (1979). Cognitive aspects of proof, with special reference to the irrationality of  $\sqrt{2}$ . *Proceedings of the Third International Conference for the Psychology of Mathematics Education*, Warwick, England, 206–207.
- Tall, D. O. (1995). Mathematical Growth in Elementary and Advanced Mathematical Thinking, *Proceedings of the Nineteenth International Conference for the Psychology of Mathematics Education*, Recife, Brazil, I, 61–75.
- Tall, D. O. (1996) Understanding the processes of advanced mathematical thinking, *L'Enseignement Mathématiques*, 42, 395–415.
- Tall, D. O. (in press, a) Symbols and the Bifurcation between Procedural and Conceptual Thinking, *International Conference on the Teaching of Mathematics, Samos*. (to appear).
- Tall, D. O. (in press, b) The Cognitive Development of Proof: Is Mathematical Proof For All or For Some? *UCSMP Fourth Conference on Mathematics Education*, to appear.
- 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.