The role of proof in mathematics teaching and the Plateau Principle

David Easdown, School of Mathematics and Statistics, University of Sydney, de@maths.usyd.edu.au

Abstract: *One of the most difficult learning thresholds for students of mathematics is the concept of proof. The difficulty manifests itself in several ways: (1) appreciating why proofs are important; (2) the tension between verification and understanding; (3) proof construction. The first case study descibes a spurious but 'convincing' proof and a correct but 'unconvincing' proof of a deep result in linear algebra. A brief discussion introduces the Plateau Principle, an unspoken credo for successful research in mathematics, which says simply: 'look for and be prepared to use a variety of plateaus as starting points for a mathematical investigation.' The second case study illustrates an underlying 'proof template' that assists in the development of proof technique, in much the same way as a sense of perspective is essential for the ability to draw well.*

Introduction

Students entering university are often very adept at performing algorithms and finding their way through the maze of sophisticated calculations. However they tend to have very little experience with mathematical proofs even though these are central to verifying mathematical facts and building a corpus of reliable knowledge. It is common for new students to say that they 'like mathematics' but 'hate proofs'. For many, proof technique is a difficult hurdle to overcome and has all of the hallmarks of a *threshold concept*, in the sense of Meyer and Land (2003, 2005). The ability to understand and construct proofs is *transformative*, both in perceiving old ideas and making new and exciting mathematical discoveries. It is *irreversible* and often accompanied by a 'road to Damascus' effect, not unlike a religious conversion or drug addiction. The most inspiring mathematical proofs are *integrative* and almost always expose some hidden *counter-intuitive* interrelations. And of course they are *troublesome*: it can take a long time, even years, for students to learn to appreciate proofs and to develop sufficient technique to write their own proofs with confidence. Wrestling with and finally grasping the concept of proof 'becomes a rite of passage' (Bradbeer 2005), as a student seeks full membership of the mathematical community.

It is common in pure mathematics courses for proofs to take centre stage, presented sequentially and without much discrimination as to quality or quantity of detail or information being processed. It is as though the physical, willy-nilly act of justification is indispensible and inexorably leads to the learner's acquisition of mathematical knowledge. However, it is important to distinguish between knowledge of mathematical truth and the understanding of mathematics (Manin 1990). Indeed the whole premise of the SOLO taxonomy, for example, is that deep learning takes place when one passes from factual knowledge (isolated pieces of information or techniques in the unistructural or multistructural phases) to conceptual knowledge (in the relational and extended abstract phases) (Biggs 2003).

The mathematics teacher therefore has to be very careful about the selection of proofs to include when introducing topics, and filtering out certain details which can obscure important ideas. Indeed the word 'proof' is often equated with 'obfuscation'. A poorly presented proof, even if meticulously prepared, can be frustrating and wasteful in terms of time and effort in concentration. It is extremely common for students to get lost, and think 'why bother?' or 'what's the point?'. 'We believe you if you tell us something is true. There is no need to confuse us or put us to sleep.'

This article is in two parts. The first is a case study involving a deep theorem in linear algebra, with two specimen proofs, the first of which is fallacious but convincing. This introduces briefly the idea of *syntactic reasoning.* The second specimen is correct but the routine verification does not in itself develop the student's conceptual understanding. In cases like that it is the author's view that it is better to apply *The Plateau Principle* and engage the student in higher level consequences and applications.

The second case study introduces a *Proof Template* which is a useful stepping stone in developing facility in proof construction. It exploits the spectrum of possibilities between the compact syntax of a mathematical expression and broader underlying semantics.

Case study: the Cayley-Hamilton Theorem

The material described in this section is derived from the author's teaching of linear algebra, for advanced second year students at the University of Sydney. The Cayley-Hamilton Theorem is a seminal, classical result about matrices. No advanced undergraduate course in linear algebra should be without it! It states that every square matrix of ordinary numbers satisfies its own characteristic equation, that is,

> Cayley-Hamilton Theorem: If M is a matrix and $p(\lambda) = \det(M - \lambda I)$ is its characteristic polynomial, then $p(M)$ is the zero matrix.

The author has given the following well-known single line 'proof' to many different student audiences, in a traditional lecture setting, and has *never* fielded an objection. It may be that there are students who see the fallacy straight away and prefer to say nothing, but the overwhelming impression is one of universal, uncritical acceptance. The 'proof' is this:

> Specimen 1: $p(M) \;=\; \det(M - MI) \;=\; \det(M - M) \;=\; \det(0) \;=\; 0\,.$

The brevity of Specimen 1 is appealing and the chain of equalities is convincing. Indeed the last three equalities are absolutely correct, and the first equality looks so plausible that it seems churlish to question it. But of course the proof cannot be correct, because $p(M)$ is a matrix and 0 at the end is just a number. (The 0 inside $\det(0)$ is in fact a matrix, so there is the added abuse of notation, but that is harmless and not a fault in the argument.)

The fallacy is in the substitution at the first step. The matrix $p(M)$ is formed by taking the characteristic polynomial $p(\lambda)$ and replacing λ by M throughout and the constant term $p(0)$ by $p(0)I$ and then evaluating to the zero matrix using matrix arithmetic. The expression $\det(M - MI)$ at the first step is trying to circumvent this process quickly by performing a matrix substitution before taking a determinant!

In the language of algebra, the two *operations*, of matrix substitution and of taking determinants, *do not commute.* Operations A and B in mathematics often *do* commute, expressed by the simple formula $AB = BA$. This expresses a common pattern and following it superficially without regard to meaning is an example of *syntactic reasoning*, so-called because all that is important is the syntax of the formula. This mirrors 'formulaic' behaviour, which one tends to identify with 'mindlessness'. A famous example of this is *The Freshman's Dream*, where the naive student conveniently, but unfortunately incorrectly, assumes operations of addition and taking powers commute:

> The Freshman's Dream: $(x + y)^n \;=\; x^n + y^n\,.$

By contrast, here is a correct but 'opaque' proof of the Cayley-Hamilton Theorem, just for 3×3 matrices, which the author asks students to work through as an exercise or further reading, and adjust slightly for the general case of an $n \times n$ matrix:

Specimen 2: Write the characteristic polynomial of M as

$$p(\lambda) = b_0 + b_1\lambda + b_2\lambda^2 + \lambda^3$$

and put $B = \text{adj}(\lambda I - M) = B_0 + \lambda B_1 + \lambda^2 B_2$, the adjugate matrix, where B_0, B_1, B_2 are matrices whose entries do not contain any expressions involving λ. By properties of the adjugate, we have

$$(\lambda I - M)(B_0 + \lambda B_1 + \lambda^2 B_2) = (b_0 + b_1\lambda + b_2\lambda^2 + \lambda^3)I \ .$$

Equating coefficients yields

$$-MB_0 = b_0 I \ , \quad B_0 - MB_1 = b_1 I \ , \quad B_1 - MB_2 = b_2 I \ , \quad B_2 = I \ .$$

But $p(M) = b_0 I + b_1 M + b_2 M^2 + M^3$, so that, from the above equations,

$$p(M) = -MB_0 + M(B_0 - MB_1) + M^2(B_1 - MB_2) + M^3 B_2 = 0 \ ,$$

the zero matrix, and the proof of the Cayley-Hamilton Theorem is complete.

Specimen 2 is straightforward for a diligent student to check, but immensely difficult to think of in the first place. Indeed, it relies on a property of the adjugate matrix, which in turn relies on a technical and nontrivial theory about matrix determinants. *Yet it is quite inappropriate for a lecture or class presentation.* The majority of the class will be none the wiser for having seen it, and have little more insight into why the Cayley-Hamilton Theorem is true than if they saw Specimen 1, which is not a proof at all (except for 1×1 matrices), but a joke! Practicing mathematicians use the Cayley-Hamilton Theorem extensively (at least implicitly), but do not think for a moment about why it is true. They are applying the following:

The Plateau Principle: Look for and be prepared to use a variety of plateaus as starting points for a mathematical investigation.

It is not necessary to carefully and painstakingly prove everything in a mathematics course. The greatest enemy of understanding is coverage (Gardner 1993, quoted in Biggs 2003). It is more worthwhile to treat the Cayley-Hamilton Theorem as a plateau, and explore the consequences (leaving a proof like Specimen 2 as a reading exercise). For example, the Cayley Hamilton Theorem is one stepping stone towards developing the Jordan Canonical Form, which is another 'higher' plateau. Indeed, granted the Jordan Canonical Form, then one can climb down and provide a conceptual derivation of the Cayley-Hamilton Theorem, involving such diverse ideas as factorisation of polynomials and nilpotent matrices. Technically speaking, such a 'proof' then would be circular, because one plateau is derived from the other and vice versa. But it does not matter: the connections are important and lead to a robust and deep understanding of the theory. Rather than dwelling on unistructural or multistructural detail checking, such as is required to follow Specimen 2, the time and energy is better spent leading the student into the relational and extended abstract phases of the SOLO taxonomy (Biggs and Collis 1982).

There is often heated debate amongst university mathematics teachers of calculus about the efficacy or otherwise of ϵ-δ definitions and arguments. Students are often the meat in the sandwich. From their perspective at least, it might be helpful and avoid confusion if the Plateau Principle became part of the parlance of calculus teaching, and students were made aware of the plethara of instances when it is being invoked (for example when using limit theorems). Students can be given choices, depending on their level of skill, experience and needs, whether to become embroiled, for example, in the fascinating vagaries of the ϵ-δ theory, or plan their academic program (if suitable) to avoid it altogether. Students need to make choices, and choices are linked to awareness. The Plateau Principle is a useful tool that allows one to knowingly filter out complication or detail at will. Important results like the Chain Rule, the Intermediate Value Theorem and the Mean Value Theorem are best explained or illustrated heuristically. The formidable task of ploughing through rigorous justifications is pointless, and indeed counterproductive, for the bulk of students. A delicate balance needs to be struck between developing a student's technical skill and seeing the mathematics in context and relation to underlying concepts or applications. The ability to recognise and use the Plateau Principle is a skill in itself and, at least implicitly, is the basis for almost all meaningful research in pure mathematics.

Case Study: A Proof Template

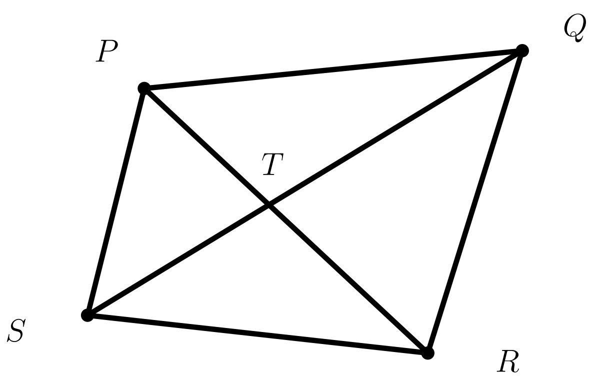
So, which kinds of proof are most appropriate for a lecture or class presentation? The short answer is those which lead as quickly as possible to deep or conceptual understanding. In this section, an example is given from the author's recent teaching of linear algebra to first year university students, who have only a secondary school background and limited or no experience with proof construction. It utilises the following

> Proof Template: expand – apply something – contract

which in turn is a manifestation of the Conjugation Principle, explained later. We will prove the 'if' part of the following

> Proposition: A quadrilateral is a parallelogram if and only if the diagonals bisect each other.

Consider a quadrilateral $PQRS$ and diagonals PR and QS which intersect in a point T. We will suppose that the diagonals bisect each other so that T is the midpoint of both.



We will prove that $PQRS$ is a parallelogram. Our supposition is just a thought experiment, but already something interesting is going on. Evidently our drawing in the page is not a parallelogram, but one can very easily *imagine* that it is! For example if the quadrilateral represented an envelope in space, then it could well be a parallelogram, whose image had been projected onto the page. Students have no difficulty with this. The paradoxical ability to flip backwards and forwards between depictions of a quadrilateral which *both is and is not* a parallelogram is a manifestation of *semantic reasoning* (Easdown 2006). We are already touching on a subject called projective geometry, and the proof has not yet begun! Here is the proof that the hypothesis about T forces $PQRS$ to be a parallelogram:

Specimen 3:

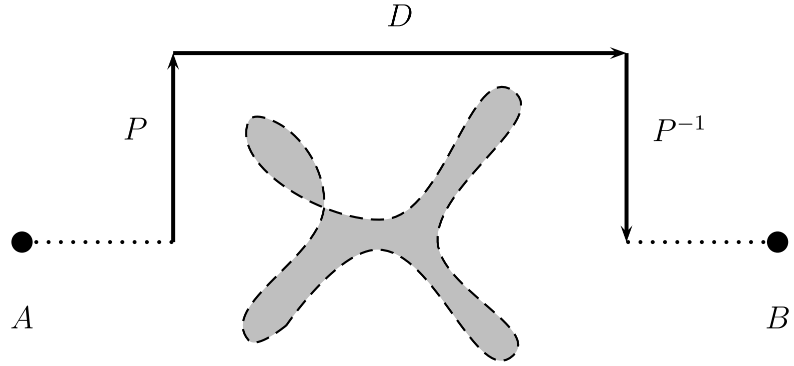
$$\begin{aligned} \overrightarrow{PQ} &= \overrightarrow{PT} + \overrightarrow{TQ} \\ &= \overrightarrow{TR} + \overrightarrow{ST} \\ &= \overrightarrow{ST} + \overrightarrow{TR} \quad = \quad \overrightarrow{SR}\,. \end{aligned}$$

To follow Specimen 3, the student needs only rudimentary knowledge of vector arithmetic. The technical demands do not obscure the underlying structure of the proof. Specimen 3 is symmetric and an application of the Proof Template, beginning with a vector expansion and concluding with a contraction. The step in the middle requires the hypothesis that T is the midpoint of both diagonals. (The equality of vectors can also be interpreted as a manipulation involving congruent triangles, the usual method of proof of the above proposition.) When students understand the underlying pattern of Specimen 3, they are able to develop with ease their own proofs, using vectors, of sophisticated and technically difficult results in geometry (Easdown 2007). The above Proof Template is one of the most common basic ideas in mathematical proof construction and itself is an instance of something more general:

The Conjugation Principle: To do something difficult, change position so that things get easier, and then return.

In Specimen 3, finding the route of equalities from $\overrightarrow{PQ}$ to $\overrightarrow{SR}$ is 'difficult'. The expansion 'changes position' and the final contraction 'returns'. The step in the middle is 'easy' given the hypothesis about T. In other words, the Conjugation Principle is *obstacle avoidance*:

Specimen 4:



Getting from point A to point B in Specimen 4 is difficult if one tries to move in a straight line. By detouring (by first applying the process P), and then moving directly ahead (doing D) where the coast is clear, and then by undoing the detour (undoing P, that is, doing P^{-1}), one returns to the straight unencumbered path to B. In symbols the Conjugation Principle becomes the famous equation:

Specimen 5: $$M = PDP^{-1}$$

where M is the entire 'difficult' process of getting from A to B. This equation is, for example, just the diagonalisation process in matrix theory, where M is a matrix, D is a diagonal matrix of eigenvalues and P is a change of basis matrix consisting of corresponding eigenvectors. If everything is invertible then all of these elements lie in a *group* and M is a *conjugate* of D (which accounts for the terminology *Conjugation Principle*).

Now, each of Specimens 3, 4 and 5 essentially captures the same idea or 'structure'. However, Specimen 5 is a compressed sequence of symbols. Probably only a mathematician would know how to 'unpack' the syntax of PDP^{-1} and recognise its universality. Specimen 3 also uses symbols, in a completely different way, but is already starting to take shape and moving towards some kind of 'semantic' representation. Specimen 4 is almost purely pictorial and is instantly meaningful, even to nonmathematicians. Developing the nexus, between syntax of symbolic representation and the underlying semantics, is perhaps the most important feature of successful mathematics teaching and learning. Resolution of the interplay between syntax and semantics is very interesting and leads to novel teaching principles, such as the Principle of Reflected Blindness and the Principle of Trivial Complexity, which inhibit learning, and the Halmos Principle, which enhances learning (Easdown 2006).

References

Biggs, J. (2003) *Teaching for quality learning at university.* 2nd edition. The Society for Research into Higher Education and Open University Press.

Biggs, J.B. and Collis, K.F. (1982) *Evaluating the Quality of Learning: the SOLO Taxonomy.* New York: Academic Press.

Bradbeer, J. (2005) Threshold concepts and troublesome knowledge in the GEES disciplines. *Planet.* No. 15, December.

Easdown, D. (2006) Teaching mathematics: the gulf between semantics (meaning) and syntax (form). *Proceedings of the 3rd International Conference on the Teaching of Mathematics at the Undergraduate Level.* Istanbul: Turkish Mathematical Society.

Easdown, D. (2007) *A First Course in Linear Algebra.* Pearson Education, Frenchs Forest, N.S.W.

Manin, Yu. (1990) Mathematics as Metaphor. *Proceedings of the International Congress of Mathematicians.* Kyoto.

Meyer, J.H.F. and and Land, R. (2003) Threshold concepts and troublesome knowledge: linkages to ways of thinking and practising within the discipline. In Rust, C. (ed.) *Improving Student Learning: Improving Student Learning Theory and Practice – Ten Years On.* Oxford: Oxford Centre for Staff and Learning Development.

Meyer, J.H.F. and and Land, R. (2005) Threshold concepts and troublesome knowledge (2): epistemological considerations and a conceptual framework for teaching and learning. *Higher Education* **49**, 373-388.