Cognitive aspects of proof, with special reference to the irrationality of √2

David Tall

University of
Warwick England

The criteria by which a proof is judged in mathematics seem, on the face of it to be quite different when considered by sophisticated mathematicians as compared with learners. The sophisticated mathematician may concern himself with logical structure, mathematical style, the degree of generality, the aesthetic quality of the proof, and so on. The learner may lack the sophistication to appreciate these criteria fully and may concern himself more with the manner in which the proof explains the result and demonstrates why it must be true, based on his current state of development.

A basis for the cognitive development of proof is already available in the psychology of learning in terms of the meaningful learning of Ausubel (1978) or the relational understanding of Skemp (1976). At any stage in development the ideas presented need to be potentially meaningful (in Ausubel's terminology), which may mean, in the short term, presenting proofs in a radically different form from that ultimately desired. In this paper we will see an example of this phenomenon. The initial proof may be more cumbersome, less aesthetically pleasing, yet prove more meaningful to the learner at the particular stage under consideration. Even so, the long-term desire for full sophistication must be kept in mind, yielding two complementary but, at times, conflicting, principles:

1. To present the material in a potentially meaningful manner for the learner,

2. To aid the learner in achieving long-term sophistication.

These two principles in fact underlie the whole of education. They must be carefully balanced; lack of balance in either direction leads to its own particular problems. One may conjecture that psychologists tend to err by over-stressing the first principle (for example in interpreting Piaget's theory of stages) whilst mathematicians pay more attention to the second. Over-indulgence in the first can lead to stunted mathematical growth whilst the second can lead to less able students having no growth at all.

When these principles are applied to the understanding of proof, the first suggests that proofs should be prepared in a manner which is intended to be meaningful at the time whilst the second suggests that the teacher should remain mindful of the need to develop more subtle and powerful proof styles, choosing the appropriate time to stretch the student's capacity.

The principles also govern the apparently conflicting criteria of the sophisticated mathematician and the learner. The learner may need more attention paid to the first principle in the initial stages, meanwhile the sophisticated mathematician will have progressed so that the first principle implies that he already has enough experience for subtle proof structure to be meaningful. However, it is not out of the question that he may meet a completely new style of proof (for instance the use of first order logical principals in non-standard analysis) where he may find himself in the same position as the learner.

The irrationality of $\sqrt{2}$

In English school sixth forms a standard introduction to proof by contradiction is the following:

> *Proof C* We show that the assumption that a rational p/q exists such that $p^2/q^2 = 2$ leads to a contradiction. Suppose that $p^2/q^2 = 2$ where p and q are integers with no factor in common, then
>
> $$p^2 = 2q^2$$
>
> Thus p^2 is even. But if p were odd, then p^2 would be odd, so this means that p must be even, $p = 2r$ where r is an integer. Substituting in $p^2 = 2q^2$, we obtain $4r^2 = 2q^2$, so $2r^2 = q^2$. The same argument shows that q must also be even. So p and q have common factor 2, contradicting the fact that they have no common factor.

This is an aesthetically pleasing proof by contradiction. However, the contradiction does not arise by contradicting the given statement:

"$\sqrt{2}$ is not a rational number,"

but a different one:

"$\sqrt{2}$ is not a rational number in lowest terms."

Learners often feel a sense of emptiness and lack of explanation as to why $\sqrt{2}$ is not irrational. There are also linguistic factors which render this proof very special, namely that the English language contains the words "even" and "odd" to describe the property of being divisible by 2 or not divisible by 2, but no other integer has corresponding words for such a divisibility property.

An alternative approach to the problem may be to describe precisely the nature of the square of a rational number and to show that 2 cannot be such a square:

Proof D We will show that if we start with any rational p/q and square it, then the resulting square cannot be 2. On squaring an integer n, the number of times any prime factor appears in the factorisation of n is *doubled* in the prime factorisation of n^2, so that each prime factor occurs an *even* number of times in n^2. In the fraction p^2/q^2 we factorise the numerator p^2 and denominator q^2, cancelling common factors where possible. Then each factor either cancels exactly, or we are left with an even number of appearances of that factor in the numerator or the denominator. The fraction p^2/q^2 cannot be simplified to give 2/1 because the latter has an *odd* number of 2s in the numerator. So the square of a rational p/q is never equal to 2.

According to the evidence which appears in text books, the mathematical community at large prefers proof C to proof D.

A questionnaire asking first year university students which of these they understood on first reading, and which caused confusion, showed no significant difference between the two, but this was highly coloured by the fact that many students had seen proof C before. A second questionnaire presented similar proofs, but with 2 replaced by 5/8, also asking the respondents whether they were familiar with either type of proof. About half were familiar with C and none with D.

Proof C* We will show that the assumption that a rational p/q exists such that $p^2/q^2 = 5/8$ leads to a contradiction.
Suppose that $p^2/q^2 = 5/8$ where p and q are integers with no factor in common. Then $8p^2 = 5q^2$. Thus 5 is a factor of $8p^2$. If 5 did not divide p then it could not divide $8p^2$, hence we deduce that $5r = p$ where r is an integer. Substituting in the original equation we get $8(5r)^2 = 5q^2$.
This simplifies to $40r^2 = q^2$. From this we see that 5 must divide q^2. and, by the same argument as before, it divides q, contradicting the fact that p and q have no common factor.

Proof D* We will show that if we start with any rational p/q and square it, then the result p^2/q^2 cannot be 5/8. On squaring any integer n the number of times that any prime factor appears in the factorisation of n is *doubled* in the prime factorisation of n, so each prime factor occurs an *even* number of times in n. (For instance, if $n = 12 = 2^2 \times 3$, then $12^2 = 2^4 \times 3^2$.)
In the fraction p^2/q^2, factorise p^2 and q^2 into primes and cancel common factors where possible. Each factor will either cancel exactly or we are left with an even number of appearances of that factor in the numerator or denominator of the fraction. The fraction p^2/q^2 can never be simplified to give 5/8, for the latter is $5/2^3$, which has an odd number of 5s in the numerator (and an odd number of 2s in the denominator). So the square of a rational p/q can never equal 5/8.

In the second questionnaire, all students were given a page of basic facts about factorisation (including uniqueness of factorisation into primes). They all responded that they were familiar with these facts. They were asked if they understood either proof or were confused by either proof on

the first read through. They were allowed to keep the questionnaires for a few days to see if their attitudes changed and, if so, were asked to record the change and when it happened.

Overall, 33 students responded to the first questionnaire, 37 to the second out of a university mathematics intake where two thirds had at least one grade A in advanced level mathematics. There was no significant difference in the understanding (or confusion) between D and D* in the two samples, but C* proved to be highly significantly more difficult than C.

Amongst the students in the second questionnaire, D* was significantly preferred to C*, both in terms of understanding and (lack of) confusion. Most interesting of all, of those who claimed to have seen neither type of proof before, the preference for D* over C* was very highly significant. (Explicit computations are given later.)

The passage of time improved the understanding of both proofs, but the preference for D* over C* increased still further. At the end of the response only three students out of 37 failed to understand D*. All of these had claimed prior acquaintance with proof C, stated immediately that they preferred C* and did not record any change of attitude. Meanwhile four other students who had seen C before and understood C* initially but not D* ended up by changing their preference to D*. There were no preferential changes from D* to C*. No student out of sixteen who thought about the matter overnight ended by stating a preference for C*; eight said that they preferred (or only understood) D*.

A number of reasons may be put forward to try to explain these results. For instance, proof D* illustrates the reasoning with an example whereas C* merely repeats the ideas of C whilst dropping the even/odd terminology. Proof D* deals with the roles of both prime factors 2 and 5 whereas proof C* concentrates on the prime 5 and neglects 2 completely.

A sophisticated view of proof would deprecate the use of an example, for an individual case proves nothing in general; it would also applaud the concentration on 5 because this is sufficient for the argument. However, the learner sees things differently.

Students were asked to give the reasons for their responses. Time and again there was a reference to unfamiliarity causing lack of understanding and confusion, whilst familiarity was linked with understanding – a far cry from the strict application of logic being the criterion for a good proof. One would be justified in conjecturing that it was familiarity with proof C that was in large part responsible for the students' claim to understand it, suggesting an instrumental understanding rather than a relational one in some cases.

There are, however, sound reasons why proof D* is preferred to C*,

despite the mathematical community's implied preference for C over D. The main reason is that proof D readily generalises to demonstrate the irrationality of all square roots of non-squares, whereas proof C, though generalizable in spirit, is more specialist in nature because of the linguistic considerations. Thus proof D is generic in the sense that it contains within it a complete spectrum of proofs for all square roots of non-squares. It shows why no square of a rational equals 2 and the same proof readily adapts to 5/8 or any other non-square. Even the example $12^2 = 2^4 \times 3^2$ in D* is also generic in that it readily adapts to any other square. The proof C, on the other hand, achieves its ends by contradicting a subtle variant of the original statement. The idea of a generic proof ties in very naturally with the concept of "generic thought" described in Tall (1978).

General and Generalizable

It has been argued by Mark Steiner (1976) that a proof of type D is more satisfying from a philosophical point of view than C. (I am indebted to Shlomo Vinner for discussion on this point.) Steiner argues that it is not the more *general* proof that is preferable, but the *generalizable* proof. He argues against the proof C and for the proof D on these grounds. He also puts forward the view that a more general proof of the type to be found in Hardy's *Pure Mathematics* does not give a better explanation just because it is more general.
　　Briefly, Hardy's argument is as follows:

> *Proof H* There is no rational number whose square is m/n, where m/n is a positive fraction in lowest terms, unless m and n are perfect squares. For suppose, if possible, that
> $$p^2/q^2 = m/n,$$
> p having no factor in common with q and m no factor in common with n. Then $np^2 = mq^2$. Every factor of q^2 must divide np^2 and, as p and q have no common factor, every factor of q^2 must divide n. Hence $n = rq^2$ where r is an integer. But this involves $m = rp^2$; and as m and n have no common factor, r must be unity. Thus $m = p^2$, $n = q^2$, as was to be proved. In particular, it follows by taking $m = 2$, $n = 1$ that 2 cannot be the square of a rational number.

Proof H (all but the last sentence) comes from the textbook which revolutionized the teaching of mathematics in Britain earlier this century The proof was included on the first questionnaire, along with proofs C, D. It is interesting to note that it produced less understanding and more confusion than either C or D to a level that in the sample concerned was highly significant statistically.
　　It seems that Steiner's thesis on the explanatory value of a proof from a philosophical viewpoint is supported also from a cognitive one. A generalizable proof (such as D or D*) works at the example level (such as

2 or 5/8) in such a way that the examples chosen are typical of the whole class of examples (and hence are generic). A general proof (such as H) works at a more general level (with the rational m/n) and therefore requires a higher level of abstraction.

Long-term sophistication

So far we have considered the suitability of various proofs for beginners. This is not to suggest in any way that a proof suitable at this stage remains the most appropriate at all stages. When a higher level of abstraction is required, a more general proof may very well come into its own. For instance, the technique of considering a fraction in its lowest terms (as in C or H) is a valuable one for more advanced work. As an example, suppose that

$$f(x) = x^n + a_1 x^{n-1} + \ldots + a_n$$

is a monic polynomial with integer coefficients. Then any rational root of f is an integer. A suitable proof is to suppose that the root is p/q in its lowest terms. then

$$(p/q)^n + a_1 (p/q)^{n-1} + \ldots + a_n = 0$$

so

$$-p^n = a_1 p^{n-1} q + \ldots + a_n q^n.$$

The right-hand side is divisible by q, so q divides p^n. Thus q is a factor of p and q and, since they have no non-trivial common factor, $q = \pm 1$ and $p/q = \pm p$ is an integer.

In this way the second principle mentioned earlier must come into play and the learner must be helped in achieving long-term sophistication. If this is done by using proof C as an introduction to proofs by contradiction, some degree of confusion is to be expected. This suggests that simpler contradiction proofs should be met first. Examples could include:

m^2 even implies m is even (for an integer m),

$0 < a < b$ implies $0 < \sqrt{a} < \sqrt{b}$ (for real numbers a, b).

Such effort would be worthwhile, for the contradiction proof C is easier to write in symbols than proof D, and with greater sophistication it can assume a more central role.

Responses to questionnaires
The 33 students responding to the first questionnaire understood the proofs, or were confused in the following numbers:

Total 33	D	C	H
understand	19	24	9
confused	10	6	23

Table 1

Comparing the understanding of D (19 understood, 14 no comment) with that of C (24 understood, 9 no comment) and calculating χ^2 (with Yates correction) gives $\chi^2 = 1.1$, which is very insignificant (30% level). A similar conclusion holds for the level of confusion. However, as we shall see in the second questionnaire, a large number of students were likely to have been familiar with C, but hardly any with D, so in table 1 the rating of C is enhanced. Though C and D are not significantly different, comparing H with either gives a different tale. For instance, comparing H with C for understanding in the same way gives $\chi^2 = 11.88$, which is highly significant (0.06%). Similar figures occur in comparing H with D.

The 37 students responding to the second questionnaire replied as follows:

Total 37	D*	C*
understand	25	14
confused	11	21

Table 2

More interesting were the responses of those who claimed to have seen neither type of proof before:

Total 20	D*	C*
understand	15	3
confused	4	17

Table 3: Students who had not seen either type of proof before

Total 17	D*	C*
understand	10	11
confused	7	4

Table 4: Students who had seen proof C before[1]

---

[1]This table was not given explicitly in the original publication.

Comparing the total tables 1 and 2, we find that the level of understanding of D and D* are not significantly different (56%) nor do they cause significantly different confusion (96%). On the other hand, C* is

significantly more difficult to understand than C (1%) and causes more confusion (0.3%). We can see that this is significantly affected by the large number of students (17) who had seen C before and, presumably were able to understand it and had no confusion from the start. In fact, subtracting the entries in table 3 from those in table 2 to obtain the figures for those who have seen C before (but not D), we find 11 of the 14 who claimed to understand C* have seen C before, whilst only 10 of the 25 who understand D* have seen C before. In this way we find that the effect of seeing C before has a highly significant effect on understanding C* (0.6%), whilst, as one might expect, it has no significant effect on the understanding of D* (30%). The figures for confusion are similar, having seen C causes a highly significant reduction in confusion for C* (0.06%) but has no effect on confusion in D* (29%).

The most important figures are found in table 3. For those students who have seen neither type of proof before, the level of understanding D* over C* is very highly significant (0.05%) and reduction in confusion for D* over C* is also very highly significant (0.02%).

Conclusions

We have seen in a group of 20 able students the significant preference in understanding of the proof D* over the proof C* when they have had no experience of either type of proof. The standard contradiction proof C, though mathematically elegant, lacks explanatory power and generalises with difficulty because of linguistic considerations. The Hardy proof does not have greater explanatory power for students, despite its greater generality. Meanwhile the generic family of proofs including D and D*, though more verbal and less easy to write in symbols, has the elusive quality of explanation which enhances understanding. The evidence suggests therefore that we should seek the explanatory power of generic proofs for beginners, rather than the aesthetic elegance or generality of general proofs. The latter can (and should) come at a stage when they are more likely to be appreciated.

References
Ausubel, D. P. *et al*: 1978. *Educational Psychology, a Cognitive View*. Holt, Rinehart & Winston.

Skemp, R. R.: 1976. Relational & Instrumental Understanding, *Mathematics Teaching* 77.

Steiner, M.: 1976. *Mathematical Explanation*, mimeographed notes, Columbia University.

Tall, D. O.: 1978. Mathematical Thinking & the Brain, *IGPME Proceedings*.