Shifts in the Meanings of Literal Symbols

Liz Bills
Mathematics Education Research Centre
University of Warwick, UK
e-mail: Liz.Bills@warwick.ac.uk

The difference in roles played by literal symbols has been used as an organising principle for various aspects of the learning of algebra. In this paper I draw attention to the shifts in the meanings attached to literal symbols which may take place during certain problem solving procedures. I argue that students' appreciation of these shifts can contribute to their meaningful use of these procedures.

Introduction

Much of the literature on children's understanding of algebra uses as an organising principle the notion of different uses of literal symbols.

Küchemann's (1981) early classification of children's use of letters consisted of six types of response to test items: *letter evaluated, letter not used, letter used as an object, letter used as a specific unknown, letter used as a generalised number* and *letter used as a variable*. The last three of these six were thought of as demonstrating some kind of algebraic competence.

Usiskin (1988) describes four meanings of variable linked to different purposes of algebra. These are *generalisations*, if algebra is seen as generalised arithmetic; *unknowns* if algebra is seen as a procedure for solving certain kinds of problems; *parameters* or *arguments* if algebra is seen as the study of relationships between or among quantities; finally, *arbitrary objects which are members of an abstract system.*

In contrast Sfard and Linchevski (1994) trace the historical development of algebra through the following stages: algebra as generalized arithmetic (operational and then structural phases), algebra of a fixed value (of an unknown), functional algebra (of a variable), and finally abstract algebra (algebra of formal operations and abstract structures).

Ursini and Trigueros (1997) consider the algebraic skills necessary for undertaking undergraduate studies under three headings: *variable as unknown*, *variable as general number* and *variables in a functional relationship*. They assert that "College students should be able to cope with all of them, moreover in order to handle the variable as a mathematical object they should be able to integrate its different uses in one concept and shift between them depending on the requirement of the task" (p 256). However in this paper and in Trigueros and Ursini (1999) they focus on tasks which demonstrate understandings of one of these roles rather than of shifts between roles.

Each of these studies has a slightly different purpose for the notion of different uses of literal symbols. Küchemann uses it to classify children's responses; Usiskin links it to different understandings of the purpose of algebra; Sfard and Linchevski use it to identify stages in the historical and epistemological development of algebra and Ursini and Trigueros use it to identify and classify algebraic skills. This paper will consider, not the categorisation of algebraic activity by different uses of variables, but the kind of thinking required to move flexibly between different uses.

The study which gave rise to this paper was undertaken with sixteen and seventeen year old pupils. A key feature of the mathematics curriculum which they were studying was the introduction of problems which involved differing roles of literal symbols within the same problem, for example:

- using "standard" forms such as $y = mx + c$, $ax^2 + bx + c = 0$, where the roles of the x and y variables are familiar, but a, b, c and m are replacements for what have, up till now, been numbers
- considering functions such as $k(k-1)x^2 + 2(k+3)x + 2$ in which the roles of the two variables could be seen as equivalent but are more likely to be seen as very different because of the familiarity of one (x) and the relative unfamiliarity of the other
- using the notion of a variable point, (x, y) or (a, b) rather than a single variable, or of a variable line or curve.

Problems of this kind are often described simply as those involving parameters but this description seems to miss some important distinctions between roles of literal symbols which will arise in the later discussion. Moreover, as my second example above illustrates, the role of a literal symbol within an equation or problem situation need not be inherent to that situation but may be determined by the perspective and actions of the individual. It would be mathematically correct to describe the expression $k(k-1)x^2 + 2(k+3)x + 2$ as a quadratic in k, but much more likely that it would be seen as a quadratic expression in x with k as a parameter.

Amongst recent studies of students' work on problems involving 'parameters', Furinghetti and Paolo (1994) undertook a larger scale study which involved setting questionnaires to 199 sixteen and seventeen year olds. Amongst the points which arise from the researchers' analysis of the replies are the following:

- letters in apparently symmetrical roles, e.g. $kx > 0$, cause difficulties for students
- some letters elicit a stereotyped expectation of role

Their main conclusion, however, is that the majority of the students they surveyed had difficulty in expressing the difference between unknown, variable and parameter.

My study led me to the conclusion that very many of the problems that my students were tackling involved a subtle shift in the role played by the literal symbols; moreover it was this shift which in each case provided the power that made the solutions to these problems 'standard methods'. I will argue that it is these shifts of meaning that allow students to perform these standard methods flexibly and not by rote.

My use of the word 'shift' follows, but is not the same as that made by Mason and Davies (1988). The kind of 'shift' I am pointing to is a 'shift of attention' which occurs within the mind of the individual. By connecting them to certain classes of problems I am claiming that such shifts are associated with success with these problems and flexibility in the application of the methods of solution.

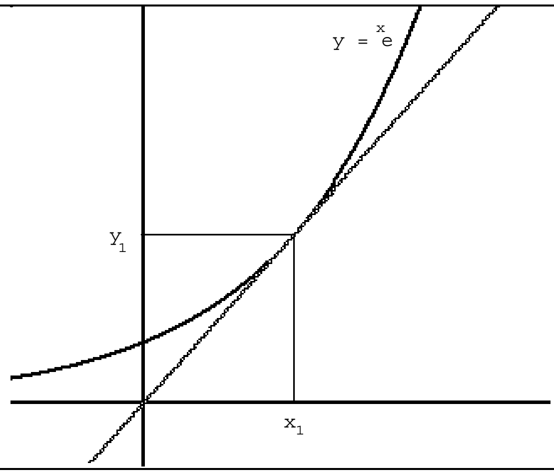
Shifts

'Variable' to 'unknown-to-be-found'

By considering the range of types of problem that the students were working on I identified four kinds of shift. The first shift is in the role of x or y from 'variable' to 'unknown-to-be-found'. By 'variable' I mean a quantity whose importance is entirely in its relationship with another quantity, rather than in its value or values. The term 'unknown-to-be-found' means a literal symbol which has a particular value which is to be discovered. For example, in a solution to the question 'Find the co-ordinates of the point where the line $x + 2y - 4 = 0$ meets the line $y = 2x - 2a + b$ ' x and y are first seen as variables whose importance is in their relationship to each other. Each can take any real value. However as soon as the learner begins to solve these as a pair of simultaneous equations, x and y take on the roles of unknowns, whose numerical values are to be found. Other examples of problems involving this shift are questions which require the co-ordinates of points of intersection with the axes or of turning points. Although it is a significant issue in earlier years at school, I found no evidence that simple problems involving this shift were still causing difficulties for students at age sixteen.

However, when this shift occurred within a more complex problem one group of students produced a very interesting response. Below is a reproduction of the work of one group on the problem "Find the point on the curve $y = e^x$ at which the tangent goes through the origin"

$\boldsymbol{y = e^x}$



gradient of tangent $= e^{x_1}$

equation of tangent $\quad y = e^x x$

$y = e^x.$

$\therefore \quad e^x x = e^x.$

$\therefore \quad x = 1 \qquad y = e.$

On a first reading I was struck by the fact of the various meanings of x in their workings. They drew a sketch of the curve and in their diagram they recognised the *particular* nature of the point of tangency by labelling it (x_1, y_1) and the gradient of the tangent e^{x_1}. However, in their working to find the x coordinate of the point of tangency, they dropped the subscripts and wrote $y = e^x x$ as 'equation of tangent'. According to one analysis, the first 'x' in this equation refers to a particular value of x which occurs at the point where the tangent touches the curve. The second 'x' is a variable in a relationship between x and y which can be represented by a straight line.

At this stage, <u>in their notation</u> they have lost the distinction between x as a variable in an expression of a relationship represented by a straight line and x as an unknown coordinate of the point of tangency (as expressed by x_1). Both these meanings of x exist within the same equation.

In the next equation, $e^x x = e^x$, they equated the y-coordinate of a point on the tangent with the y-coordinate of a point on the curve in order to find the value of x at the point of contact. Now each x in the equation refers to the unknown value previously called x_1 and x is unambiguously an unknown-to-be-found. The students did not see the need to use two different letters to make the earlier distinction. However the elision between the two meanings of x did not prevent them from going on to find the particular value of x required.

This ambiguous use of x occurs because of a shift from variable in the equation of a curve (in the equation dy/dx = e^x, x stands for the first coordinate of any point on the curve $y = e^x$) to unknown particular value of x (in the statement 'gradient of tangent = e^x' x stands for the first coordinate of the point where the tangent touches the curve).

'Placeholder-in-a-form 'to 'Unknown-to-be-found'

The second shift I have identified involves the role of placeholder-within-a-form. By this I mean literal symbols which carry a special meaning by virtue of being frequently encountered in a specific context e.g. m and c in $y = mx + c$. The special status established by these literal symbols is discussed in Bills (1997). The shift in question takes place from the role of placeholder-in-a-form to that of unknown-to-be-found. For example in answering the question 'What is the equation of a straight line with gradient 3 which passes through the point (2, 8)?' a student might substitute $x = 2$ and $y = 8$ into the form $y = 3x + c$ to find a value for c. As a result of this substitution, c changes from being a placeholder for 'the y intercept' within a standard form, to being an unknown-to-be-found. This shift was also identified by Bloedy-Vinner (1994) as follows:

> 'Moreover, the meaning of a letter as a parameter or as an unknown or variable, might change throughout the process of solving a problem ..... Solving this problem ("Find an equation for the line through (2, 5) with slope 3") starts with writing an equation $y = ax + b$, where common knowledge determines that x and y are variables whereas a and b are parameters. The process continues by substituting the constant 3 for a, and solving an equation with unknown b, where constants are substituted for x and y. The process terminates by substituting the constants found for a and b, and by letting x and y be variables in $y = 3x + 1$' (p89-90)

Again, students at this level find it relatively straightforward to learn and use procedures which involve this shift. However a slightly different picture emerges when the problem offered to students is non-standard in some way. For example, in the extract below there is insufficient information given to find the equation of the line.

Paul and Trevor had asked me to give them a revision session on aspects of coordinate geometry because both had missed some of the lessons on this topic. I began by speaking to them about the general form for the equation of a straight line, $y = mx + c$, and how they would use it to find the equations of particular lines. I asked them to explain what they could tell about the equation of a straight line if they knew the line went through the point (1, 2), and, after their answer, I continued by asking them what they could say about the equation of a line which passes through (0, 4).

Trevor: c has to equal 4.

Paul: c equals 4, because 4 equals $0 + c$

Trevor: So the gradient is 0.

Liz: It tells us that c is 4, which is, you could have done that by a slightly different sort of reasoning because, c, you said to me was the point on the y-axis where it cuts.

Trevor: Yeah.

Liz: And this point (0, 4) is on the y-axis. It goes through (0, 4) then c is 4. What does it tell us about m?

Paul: That it's 0 because ....

Trevor: I don't know if it would be 0, cause you are just saying that x is 0. It still could be at an angle

Paul: We know, we know that $y = 4$, in this particular case and we know that 4 is c, so we know that mx has got to equal 0.
Liz: Yeah.
Paul: And the only way mx could equal 0 is if m is 0.
Trevor: No but
Paul: No, No, No - because x is 0.
Trevor: x is zero, so m could be anything
Paul: Yeah that's it so m could be anything

Trevor and Paul's attention throughout this extract and most of the rest of the conversation was on substituting values for x and y into an equation for a straight line in order to find the values of the placeholders, m and c, now treated as unknowns. They began by treating the point (0, 4) in the same way. My agenda was different. I wanted them to see that (0, 4) can be treated differently because it is on the y-axis, but my intervention failed to shift their attention away from the substitution they had made. In response to my question about the value of m they returned to their equation $y = 0 + 4$ and Paul deduced that m must be equal to zero. One interpretation of his line of argument is that , knowing that he was seeking information about m, he chose to treat x as indeterminate, a varying quantity which must be given freedom to vary, rather than treating it as a known value, 0. The situation was compounded by the fact that there was insufficient information to calculate a value for m. The students expected to be able to find the value of m and taking x as indeterminate rather than given enabled them to do so. However their focus on the role of m as unknown eventually allowed them to see a solution.

This incident is also an example of what Furinghetti and Paolo (1994) speak of as students' difficulties with letters in apparently symmetrical roles i.e. $mx = 0$. The confusion might be understood as one between the two different roles of the letters in an equation which gives no clues as to which letter is playing the role of unknown and which the role of given.

'Unknown-to-be-taken-as-given' to 'Unknown-to-be-found'

The third shift can occur when an *analytic* solution method is used. I use the term analytic in the sense of Klein (1968). He describes "analysis of the first kind" as a method of solving problems algebraically which assumes the unknown as known and then transforms the equation to identify the unknown. For example, in solving analytically the problem 'Find the point of contact of the tangent to the curve $y = x^2 + 1$ which passes through the origin', the first step is to name the unknown by choosing a letter to stand for the x coordinate of the point of contact. The chosen letter, a say, is then treated as given and used to form equations which express relationships between a and other quantities. Finally those equations are solved for a . The shift that takes place then is from unknown-to-be-taken-as-given to unknown-to-be-found.

A number of students worked on a similar task at my request. The adapted task was "Find the equation of the tangent to the curve $y = x^2 + 1$ which passes through the origin".

When Paul was given this question to work on he recognised that it could not be solved by the synthetic approaches he had used so far:

Paul: Because you've, you've got no place to start after you've done that. You know that it passes through the origin, it could do that, it could do that (*he indicates lines passing through the origin with different gradients*) whatever, be a tangent to the curve and pass through it. I don't know, you can only do it one. Well you know it passes through the origin but you've got no, no idea where it touches and you need to know where it touches to be able to get the gradient and you need to have the gradient to know where it touches. So you've got a loop which you can't, you can't solve that easily.

From my standpoint I can understand Paul's description as saying that this question requires an analytic rather than synthetic approach. An 'analytic' approach works from the unknown, treating it as known in order to discover its value. A synthetic approach starts from the known and works towards the unknown. Paul wanted to work from the known to the unknown but found that he could not do so.

I tried to encourage him to use an analytic approach. However he understood my suggestion as advising him to guess a particular value rather than to express ignorance by the use of a letter. He eventually solved the problem by a guess and check procedure, beginning by guessing a value for the gradient. Next he found the x coordinate of the point on the curve where the gradient was equal to his chosen value. Then he checked that this point on the curve also lay on the line through the origin with chosen gradient.

Paul went through this guess and check procedure twice, first starting with a gradient of 1, which he found did not fit all the conditions, and secondly with a gradient of 2, which did. His first step in each case was to find the x coordinate of the point on the curve which had the given gradient. In other words, having decided on a trial value for m he used it to find a value of x which fitted certain conditions. A little later in the conversation I encouraged him to develop this approach into an analytic one, where the unknown gradient is named as m and the testing procedure is adapted to set up equations from which the value of m can be calculated.

Liz: .. think back to what you did to start with. You said 'suppose the gradient's 1,'

Paul: oh, yes.

Liz: Now go back to that stage and think 'suppose the gradient's m.'

Paul: If you put .... $y = mx + c$ *(inaudible)*. You know it's zero, you know it's $y = mx$

Liz: yes

Paul: equals, ... ah $m = 2x$ doesn't it, because the gradient's - m is going to equal $2x$, so it's $y = 2x^2$

His earlier reasoning was along the following lines (although I have no evidence that his mental image was in terms of equations and implications)

gradient = 2 $\quad\quad$ $2x = 2$ $\quad\quad$ at point of contact

$x = 1$

With m as gradient however he proceeded in this way

gradient = m $\quad\quad$ $2x = m$

$m = 2x$

tangent is $y = mx$ $\quad\quad$ $y = 2x^2$

If Paul had followed his guess and check procedure, his first stage would have been to say that the point on the curve at which the gradient is m is given by $2x = m$, so that $x = \mathrm{m}/2$ at this point.

However, rather than finding x in terms of m, that is treating m as the known and x as the unknown, he took an expression for m in terms of x as his next stage. Even though it would have led him along a route parallel to that which he had already travelled when using his 'guess and check' approach, he was not able to treat m as known. This approach led him into difficulties because the x he was working with here was the x coordinate of the point of contact rather than the x coordinate of any point on the tangent. Seeing that he had derived $y = 2x^2$ as the equation of the tangent alerted him to the need to rethink.

'Unknown-to-be-taken-as-given' to 'Variable'

The fourth shift takes place when a quantity which was originally conceived of as constant, though unspecified, is allowed to vary, that is it is a shift from unknown-to-be-taken-as-given to variable. This shift frequently occurs in solutions to locus problems, for example 'A point P, co-ordinates (a, b) is equidistant from the x-axis and the point (3, 2). Find a relationship connecting a and b.' In the solution to this problem, a and b are first taken to be fixed but unspecified, so that expressions for the distances from (a, b) to the x-axis and (3, 2) can be formulated in terms of these unknown-to-be-taken-as-givens. Once these expressions have been equated the equation formed can be seen as a relationship between variables and a and b can be allowed to vary in order to map out a parabola. In this question this last stage, which represents the locus aspect of the problem, is not emphasised, because the question asks merely for a relationship between a and b. An emphasis on the locus aspect of the problem is usually accompanied by a change in notation which allows the final relationship to be expressed in terms of x and y. This notational change allows the shift to seeing the letters as variables to take place more easily because the conventional roles of x and y are as variables.

My students were set a test and one of the questions was as follows:

"A circle has centre (2, 4) and passes through the point (-1, 5). The point (p, q) lies on the tangent which touches the circle at (-1, 5). Find an equation linking p and q. Hence write down the equation of the tangent."

Of the students who made any substantial attempt at the question, all but one worked from the outset with x and y rather than p and q. Some obtained an equation in terms of x and y and then substituted p and q into it. Some did not include p and q in their answer at all.

The method used by the students was to find the gradient of the radius and hence of the tangent and then to obtain the equation of a straight line with this gradient and passing through (–1, 5). The focus was always on the tangent as a line and not on the point (p, q) on the tangent. A method which focused on the point (p, q) might have been based on expressing this fact: that the line joining (p, q) and (-1, 5) and the line joining (2, 4) and (-1, 5) are perpendicular and therefore that the product of their gradients is 1.

The students' choice of method and of letters indicates that they were following a standard procedure for finding the equation of a tangent to a circle, rather than responding meaningfully to the problem which had been set. Their understanding of the standard procedure does not include an awareness that at the outset (x, y) represents a point on the tangent, and hence they ignore the cue in the question to use (p, q) as that point. An interview with one of the students suggested that his choice of (x, y) was unconscious.

Conclusion

I have shown how considering shifts in the meaning of variables can give a new perspective on standard problems and routine procedures in this area of algebra, and illustrated how students' appreciation of these shifts can be related to their meaningful performance of these routines. In my first example students showed by their use of notation that the shift they were making was unconscious. This unconsciousness put them at risk of making an error. In my second example Paul and Trevor were able to solve a non-routine problem because they were confident in the role of m as unknown-to-be-found, even though it had been introduced as a placeholder. In the third, Paul was, at least initially, unable to pursue the analytic approach because he wanted to shift too soon from m as unknown-to-be-taken-as-given to unknown-to-be-found. Finally, students faced with a locus question use an established routine rather than approaching the question as set because their understanding of the routine does not include an appreciation of the shift in meaning of the letters.

If, as the papers I have cited earlier suggest, understanding of the different meanings of variables is a problem for students, it may be that drawing their attention to the shifts in meanings involved in some standard problems and their routine solutions would be an effective way of helping students to improve such understanding.

References

Bills E (1997). Stereotypes of Literal Symbol Use in Senior School Algebra. Twenty first Conference of the International Group for the Psychology of Mathematics Education, Lahti, Finland, Program Committee of PME 21, **2:** 73-80.

Bloedy-Vinner H (1994). The Analgebraic Mode of Thinking - the Case of the Parameter. Eighteenth International Conference for the Psychology of Mathematics Education, Lisbon, Portugal, Program Committee of PME 18, **2:** 88-95.

Furinghetti F, P. D. (1994). Parameters, Unknowns and Variables: a Little Difference? Eighteenth International Conference for the Psychology of Mathematics Education, Lisbon, Portugal, Program Committee of PME 18, **2:** 368-375.

Klein J (1968). Greek Mathematical Thought and the Origins of Algebra. Cambridge, Mass., MIT Press.

Küchemann, D. (1981). Algebra. in Children's Understanding of Mathematics :11-16. Hart K. (Ed.) London, John Murray.

Mason J and Davis J (1988). Cognitive and Metacognitive Shifts. Twelfth International Conference of the International Group for the Psychology of Mathematics Education, Vezprém, Hungary, Program Committee of PME 12, **2:** 487-494.

Sfard A, Linchevski L. (1994). "The Gains and Pitfalls of Reification - the Case of Algebra." Educational Studies in Mathematics **26**: 191-228.

Trigueros, M., Ursini, S, (1999). Does the Understanding of Variable Evolve through Schooling? Twenty third Conference of the International Group for Psychology of Mathematics Education, Haifa, Israel, Program Committee of PME 23, **4:** 273-280.

Ursini, S., Trigueros, M, (1997). Understanding of Different Uses of Variable: A Study with Starting College Students. Twenty First Annual Conference of the International Group for the Psychology of Mathematics Education, Lahti, Finland, Program Committee of PME 21, **4:** 254-261.

Usiskin, Z. (1988). "Conceptions of School Algebra and Uses of Variables" in *Ideas of Algebra, K-12, (1988 Yearbook)*. Coxford, A and Schulte, A P, (Eds.),. Reston, VA., National Council of Teachers of Mathematics: 8-19.