Issues with Notation at the Interface between Technologies

Chris Barling and Peter Jones
Swinburne University of Technology
cbarling@swin.edu.au
pjones@swin.edu.au

Abstract
Mathematics is underpinned by the notation in which we write it. Traditional notation is not free of ambiguity, but on the whole the conventions we use to write mathematics on paper are flexible and visually clear. It is important, though, to view this as the product of a particular technology. When we wish to transfer our mathematics from one technology to another, the interface between the systems of notation becomes paramount. The problem is highlighted particularly when students are introduced to computer algebra systems (CAS).

In this paper we highlight the differences between pencil-and-paper and electronic (linear) notation, some well-recognized and others not, particularly noting certain points at which students often encounter difficulties. The solution to these problems is not intuitive or obvious, and students encountering the interface between traditional and CAS technologies need specific intervention to assist them in overcoming the many difficulties they encounter.

Mathematics through notation
We can do simple calculations in our head, and follow a simple mathematical argument presented verbally, but most of the time when we do mathematics we have to store some or all of our work. This is most commonly by writing or recording what we are doing on paper, or a blackboard, or a calculator screen or in a computer, or perhaps in some other calculating device such as an abacus. This paper is concerned primarily with the dominant mode of writing in its various forms, whether pen or pencil on paper, on a black or white board or a computer screen, on a slate or clay tablet or in sand: the whole spectrum that for brevity we will call working “on paper”. Let it be said right at the beginning that this in itself constitutes a technology, without which very little mathematics is possible.

Notation “on paper”
If we write the number 237 we are using characters with a long history, and a convention of denoting value by position, the placing of the three characters adjacently itself creating a meaning.

If we add two numbers and write $237 + 125 = 362$ we are using further characters with equally interesting histories, one to denote an operation and one a relationship of equality.

If we set out the addition in the usual way:

\[
\begin{array}{c}
237 \\
125 \\
362 
\end{array}
\]

we are doing something else, now creating meaning through the relative positions of the three numbers on the page, and the use of the dividing bar to separate the two numbers being added from their sum.

If we decide to square one of the numbers and write $237^2 = 56169$ the raised 2 has yet further meaning, denoted by its alignment vertically above the other digits, $237^2$ being quite different from $237^2$.

All of this illustrates an essential feature of the “on paper” technology and its notation, that meaning is created by position. Placing characters in alignment creates meaning, whether in the horizontal line of the single number or the vertical alignment in the columns of the addition. Meaning is also given by displacement, as in the raised power, or the various meanings of subscripts. Even simple and familiar concepts such as fractions are notated by a sophisticated combination of positionings such as in the equation

\[
\frac{238}{2} + \frac{126}{6} = 140
\]
Whether we write this on a blackboard with chalk or on paper with a pen, or indeed trace it in sand, the essential feature of this group of technologies is that the hand is free to move in interacting with a two-dimensional medium, so that movements left to right or up and down are all equally possible and easy to achieve.

So far we have restricted examples to simple arithmetic, but once we move on into algebra and higher mathematics the same observations are equally important. We routinely create complex formulas such as

\[ T = 2\sqrt{l/g} \]

or

\[ x = \frac{b \pm \sqrt{b^2 - 4ac}}{2a} \]

and elaborate objects such as

\[ \int \frac{x^2 + 1}{\sqrt{x^2 + 1}} \, dx \]

or

\[ \lim_{n \to \infty} (1^n)^n u_n \]

or

\[ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \]

in which both alignment and displacement are critical to creating and understanding the meaning.

Problems with traditional “paper” notation

The notation that most of us think of as "normal" is actually the product of long evolution. The standard text on this history is a beautiful and amazing book by Florian Cajori (1928/9), which gives the whole history of modern European notation up to about 1900, and illustrates how recent and how localized is the acceptance of what we think of as standard. Whether from historical study or from our own experience, it is clear that there a number of issues that potentially create difficulties even within this apparently homogenous and well-accepted “system”.

First, there is no uniformity in the actual agreed character set to be used: both within one society and even more internationally one is aware of a great deal of variation between the actual shapes of the digits 1 to 9, for instance, and in non-western scripts such as Arabic or Chinese, of course, the notation is quite different. Even more, there is enormous variation in even the most basic conventions such as the notation of the four basic operations of arithmetic or the use of commas and points in long numbers. In the USA and Britain, for instance, 12,345 denotes twelve thousand, three hundred and forty-five (the comma marking place), while 12.345 denotes twelve and 345 thousandths (the point indicating the beginning of the fractional part). In many parts of Europe the convention is exactly reversed.

Secondly, the one symbol or combination of symbols can take on many different meanings in various contexts. This is not always problematic, but even at the school level it can be confusing, for example, as to whether the symbol

\( (1, 2) \)

represents a point in two-dimensional space or an open interval on the number-line — and in higher mathematics it can denote so many other things (ordered pair, greatest common divisor etc). Another ambiguity, even more worrying, is the use in some countries of the point to denote both the decimal marker (12.345) and multiplication (4! = 4.3.2.1): without some context it can be very difficult to know what an expression like 2.3 might mean. At different times there have been attempts to differentiate the two meanings by use of a raised dot for one and a dot level with the foot of the numerals for the other, but as Cajori (1928, paras 287, 288) makes clear, the rule as to which dot signifies which operation has changed many times over the decades and between countries; and the need for rapid typing on standard keyboards seems to have led to the complete demise of the raised dot in modern times.

Thirdly, the one concept can often be notated in more than one way. Sometimes this is simply a matter of handwriting conventions, like
\[ \frac{3}{4} \text{ versus } \frac{3}{4} \text{ versus } \frac{3}{4} \]

or

\[ \sqrt{2} \text{ versus } \sqrt{2} \]

and sometimes it is a matter of different but equivalent concepts, like

\[ \sqrt{2} \text{ versus } 2^{\frac{1}{2}} \text{ versus } 2^{\frac{1}{2}} \]

**Traditional notation in printing and typing**

All of this notation is essentially designed for handwriting on a two-dimensional surface, where the freedom to move up, down and sideways is fully exploited. It soon became apparent, though, that in other technologies this movement was a problem. The first signal of this came with printing, when it became apparent that setting mathematics into type was much more difficult and costly than standard lettering. Fractions, for instance, with their three layers of text and the need to centre the numerator and denominator, create serious difficulties, and more elaborate structures like integrals with terminals created serious problems for specialist typesetters. A committee of the British Association looking into this in 1875 reported that

The cost of ‘composing’ mathematical matter may in general be estimated at three times that of ordinary or plain matter …’

(British Association Report for 1875, p. 337, cited in Cajori (1929 p. 335)).

Cajori (1929, pp 180ff) reports that the first mathematician to appreciate this was Leibniz, who devoted a great deal of thought to alternative notations and actively promoted the use of those which do not require any variation to continuous linear typesetting, for instance preferring to write the fraction 3/4 as 3 : 4 . The British Association report just cited includes a number of recommendations to avoid notations that require “justification” (in the sense of complex typesetting procedures), such as

replace \[ \frac{3}{4} \] with \[ 3 + 4 \] or \[ 3 : 4 \]

or replace \[ \sqrt{2} \] with \[ \sqrt{2} \]

The Council of the London Mathematical Society made a number of similar recommendations in 1915, reported in Anon (1915) and briefly discussed in Bryan (1916).

In more recent times many of us faced the same difficulties closer to home when attempting to have mathematical work typed on conventional typewriters: even with the sophistication of “golf balls” and the like this was still a difficult area requiring special expertise (and often special social skills in negotiating with the typist).

**Linear notation**

The one-dimensional notation promoted by Leibniz and his successors has turned out to have more importance than just a reduction in labour and costs, because it has now become fundamental to the calculators and computers that we all depend on. The instructions that drive all electronic devices are essentially linear: things happen in a structured, one-dimensional way, one thing after the other. And from the first the input devices (at least in English-speaking countries) have been based on the QWERTY keyboard, with little if any possibility for additional symbols, or for sophistications such as subscripts and superscripts or unusual fonts or characters.

If we look at the earlier programming languages they are full of conventions that replace all of these things with less attractive but clearer and simpler notations. Addition and subtraction use their standard signs, which are in the standard keyboard; but multiplication is denoted by the asterisk * (possibly to avoid confusion with the letter x, which must be reserved for other uses in any case), and division with the slash /. Powers are generally denoted by the “hat” sign, replacing \[ 4^2 \] with \[ 4^{\hat{2}} \].

In algebra we have to learn to make do without many things, most obviously fractions, so that

\[ \frac{a+b}{c} \]

becomes

\[ (a + b)/c \]
for instance. The treatment of subscripts, the letter $\pi$ and a host of more complex notations (matrices, definite integrals) remain a problem that has to be dealt with from system to system or language to language. The only common thread is that ultimately every mathematical expression is turned into a linear string of standard characters. In the earliest systems, and still in simple devices such as graphics calculators, the string is explicit, with only a limited number of programmed functions (such as the square root) allowed. In many later systems there were numerous system macros that took the place of more explicit spelling-out of functions as text: for example in earlier versions of Microsoft Word for the Macintosh computer there were a number of “formula” commands that allowed the construction of quite complex objects. The expression

$$\sqrt{x^2 + y^2}$$

in this system would have been typed as

```
\sqrt{x^2 + y^2}
```

in which the $\sqrt{}$ constructs the square root and the $\up$ commands form the superscripts (raised by 5 points). The quadratic equation formula

$$b \pm \sqrt{b^2 - 4ac}$$

would have been typed as

```
F(-b \pm \sqrt{b^2 - 4ac}, 2a)
```

Later again the process has often become invisible, as in a modern word-processor that uses a point-and-click process to create visually complex formulas and expressions. This, though, is still not always the case, and there are many contexts on many devices that require strict linear notation, such as creating a function in a graphics calculator or a CAS system. In CAS most devices will then display the output in more visually friendly form, but the input itself must be strictly linear.

Why has linear notation not replaced the more traditional notations, despite advocacy from Leibniz onwards? The answer seems to be that it is simply too clumsy and hard to “read” in all but the simplest cases. Presented with the sequence

$$\sqrt{x^2 + y^2} \quad (x^2 + y^2)^{0.5} \quad ((x^2) + (y^2))^{0.5}$$

the point hardly needs making: as we move left to right the ease of typing increases, but the “user-friendliness” or “readability” clearly decreases, however we might try to define such a vague concept. Fraction notation is another example of a visually evolved notation that is much easier to “read” than its linear equivalent in all but the simplest cases: compare

\[
\frac{1}{2} \div \frac{1}{3} \quad \text{with} \quad \frac{(1/2) - (1/3)}{(1/2) + (1/3)}
\]

or

\[
\frac{x + y}{y} \quad \text{with} \quad \frac{x - (x+y)/y}{y + x/(x-y)}
\]

or try to “spot the difference” in linear notation between

\[
\frac{x}{1 + x + y} = \frac{x}{(x - y)/(1 + x + y)/x}
\]

and

\[
\frac{x}{1 + x + y} = \frac{x}{(x - y)/(1 + (x + y)/x)}
\]

and other similar “variations”.
The transition problem
Most of us and most of our students are still a great deal more comfortable with traditional two-
dimensional notation, and thus when we come to use a device based on linear notation we encounter
what is essentially an *interface* between two technologies: we have to make a *transition* or *translation*
between the two systems of notation.
In the oldest calculators it was possible to perform only one operation at a time, which in some ways
simplified things (and forced clarity about order of operations). To calculate $2\sqrt{3} + 4\sqrt{5}$ on a simple
machine one would first calculate $2\sqrt{3}$ (and store it if possible), then $4\sqrt{5}$, and then add the two
answers. With more recent machines such as graphics calculators it is now possible to enter the whole
expression; the calculator “understands” the correct order of operations and gives the answer in a
single procedure. In this expression the traditional notation and the linear machine notation are
identical.
But if the expression involves any of the layered or displaced notations we have been discussing, it is
necessary for the operator to translate the expression first into linear notation before entering it into the
calculator. The most common situation where this happens in arithmetic is with fractions, where it
must always be emphasized that, in the absence of other indications, both numerator and denominator
must be bracketed (unless a single number):

- $a/b$ translates into $(a + b) / x$ [not $a + b / x$]
- $a/(x + y)$ translates into $a / (x + y)$ [not $a / x + y$]
- $a/xy$ translates into $a / (x * y)$ or $a / x / y$ [not $a / x * y$]

Here the correct order of operations is critical, and misunderstanding of that order a frequent cause of
error.
More complex expressions will create further difficulties, but the resolution of them will depend on the
specific system being used as to how roots, for instance, are created or how standard functions like
trigonometric or exponential functions are defined. Each calculator, or each CAS system, will have its
own syntax that must be learned and followed meticulously, with particular attention to the order of
operations defined by the system.

We have found that even first-year university students, who in theory are already competent at basic
mathematics and familiar with conventional and graphics calculators, are often far from confident in
using their calculators well. Often the problem is one of notation, and specifically of this conversion
process.
In a small class of 33 first-year university students in 2004 a calculus problem was set in which the
answer led to an expression such as $\frac{23}{64\pi}$, which was then to be evaluated. Twenty of the students
obtained a correct value; three obtained a correct expression but did not evaluate it, and four made
various idiosyncratic errors in the process. The remaining six (out of 33) are particularly interesting
because in each case they obtained a value which showed they had calculated not (in this case) $23 / (64 * \pi)$
but $23 / 64 * \pi = (23 / 64) * \pi$. Almost certainly they had simply entered the expression
of the form $\frac{a}{xy}$ as $a / x * y$, failing to note the necessity of bracketing the multiplication. For
comparison, in a similar exercise in 2003 39 students attempted the corresponding problem: 19
obtained a correct answer, 8 obtained a correct expression but did not evaluate it, 8 made other errors,
and 4 obtained the correct expression and then made the error described in evaluating it. The problem,
then, is not uncommon, and persistent. Numerous other examples are encountered in everyday
interaction with such students.
But there seems to be no text or other materials that seriously addresses these issues as a teaching problem at the senior level. Many students would benefit enormously from good explicit materials that would help them work through and practice conversions like the above.

To show how quickly simple things can become daunting for many students, consider the conventional quadratic equation formula: for the equation

$$ax^2 + bx + c = 0$$

the standard expression for the roots in traditional notation is

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Leaving aside the difficulty of the ± sign, and taking the + option for simplicity, how is this translated into linear notation?

On the Texas Instruments graphics calculators there is a square root function available, and in addition an implied multiplication that recognizes a number followed by a letter as multiplication, so in this system we would need to type

$$(-b + \sqrt{b^2 - 4ac})/(2a)$$  (Fig. 1).

This assumes an algebraic exercise in which a, b and c are literal expressions: if they are to be entered as actual numbers we will need to spell out the implied multiplications. This is also true in most CAS syntaxes:

$$(-b + \sqrt{(b^2 - 4ac)})/(2a)$$  (Fig. 2).

If we applied the same formula applied to a more particular equation

$$ax^2 + (4a) x + (a + 3) = 0$$

the expression (taking the + option) would become

$$(-(4a) + \sqrt{(4a)^2 - 4a(a+3)))/(2a)$$  (Fig. 3);
or (with asterisks)
\[-(4*a) + \sqrt{(4*a)^2 - 4*a*(a+3)))/(2*a) \quad (\text{Fig. 4}).\]

Those triple right brackets in particular will force students to think clearly about their syntax! CAS, of course, is far less forgiving of notational inaccuracies than even a modern calculator, some of which are more tolerant of “natural” notation than their older predecessors. Implied multiplications, for instance, are forbidden in CAS, as “ab” will be read as a two-letter variable name rather than the product of two variables “a” and “b”. Asterisks and brackets therefore abound and multiply. And there are many other notation issues peculiar to CAS systems, where one might find what is intended as a simple algebraic product like a(b+c) read as a function and argument like f(x+h), and the endless problems with intended variables reverting to earlier stored values or definitions within a session.

There are further points to think about even in simple systems. What if, instead of x^(1/2), we use a built-in reciprocal function and type \(x^2 - 1\)? What is the order of operations here? Will the machine read \(x^2 - 1\) as \(x^{(2 - 1)}\) or as \((x^2)^{-1}\)? Or, perhaps more clearly, does \(x^2*y\) mean \((x^2)*y\) or \(x^{(2*y)}\)?

Clearly one must have detailed knowledge of any particular machine or package one is using, as orders of operations and recognition of implied operations can vary enormously. Students should be encouraged to experiment with simple numbers when in doubt, and to use brackets wherever necessary to clarify what they mean; they should also try to become comfortable with the order of operations understood by their own calculators or CAS packages.

Order of operations in fact is still a thorny issue at many levels of mathematics education. Historically there have been a number of mnemonics that attempt to clarify the correct order of priority, such as

- **BODMAS** [brackets, of, division, multiplication, addition, subtraction]
- **BIDMAS** [the same, but with I = indices]
- **GEMA** [groupings, exponents, multiplication / division, addition / subtraction]
- **BEMA** [the same but with B = brackets].

The first two are still effectively endorsed by the Mathematical Association in Britain (e.g. in Barnard (1999)), but they have many opponents; none of these mnemonics has very wide acceptance, and each has problems even at the level of basic arithmetic: for instance the relative hierarchies of mixed divisions and multiplications or mixed additions and subtractions, and the lack of advice in handling fractions. Moreover none of them begins to address more thorny problems such as the question of \(x^{2^{-1}}\) mentioned earlier, or how to translate things like \(\sin^2x\). There is room for serious effort to try to find a rule or set of rules that will be truly useful and achieve some level of consensus. (See, for instance, MacGregor and Stacey (1995)).
What does the literature have to say about this group of issues? There is little that specifically addresses the notational issues raised in this paper, although many more general papers touch on difficulties with notation and representation in the context of other discussions. Williams (1993), for instance, mentions order of operations as one of three potential areas of difficulty with graphics calculators. Boelkins (1998) writes of the particular difficulty that occurs when a symbol takes on different meanings in different contexts. Eric Schechter in his excellent website (http://atlas.math.vanderbilt.edu/~schectex/commerrs/), in the long article The Most Common Errors in Undergraduate Mathematics, makes many useful points in the section “Confusion about Notation”. With CAS specifically there is a certain amount of literature on the impact on learning of other CAS-specific constraints, such as screen pixilation or the order of entry of solution forms (see for example Lagrange 1999), but on the issue of notational conflict as a source of difficulty the now extensive literature of CAS seems largely silent. None of the papers referred to in the comprehensive survey by Mueller, Forster and Bloom (2002) seems to relate specifically to issues of notation, nor do any of the papers on the University of Melbourne’s CAS-CAT Project website (http://www.edfac.unimelb.edu.au/DSME/CAS-CAT/publications.html); but of course there are many papers that include a certain amount of mention under other headings. Bloom, Forster and Mueller (2001), for instance, report many individual difficulties that seem to involve elements of notation and/or syntax or representation of expressions in the devices they used in their study. Galbraith and Pemberton (2002) identify many problems at the notational interface which they label as “syntax” problems with the CAS package. Heck (2001) touches on some notational issues in a more general discussion on the nature of “variable”.

In teaching with the various CAS systems in particular these problems are paramount. In our experience of helping students deal with Maple, Mathematica and the basic CAS language of the TI–89 we have always found that a great deal of energy has to be put into dealing explicitly with problems of notation and syntax, and specifically with dealing with the mental tricks necessary for changing traditional notation into linear notation and back again. Learning to accommodate the technological constraints of the operation of a technology, such as the need to translate pencil-and-paper notation into a machine-recognizable notation, is critical to the learning process in two ways. Firstly, and most obviously, it is necessary to enable the student to operate the technology effectively. But secondly, and perhaps more importantly, it forces the student to think more deeply about the underlying mathematical ideas and processes as they manifest themselves in the technological environment. As a simple example: in pre-calculator days students were well aware of the concept of $\sqrt{2}$, and most were clear that the table-obtained value of 1.414 was a simple approximation. Then, with the first electronic calculators, the “value” obtained is displayed as 1.414213562 or similar, and many students seem to have lost the understanding that this is an approximate value: they display a belief that “exact value” means one with lots of figures. More recently, such students encountering CAS systems find that the “value” of $\sqrt{2}$ is in fact $\sqrt{2}$, and the numerical value has to be sought as such (if desired): this has forced a more thorough and probably better understanding of the difference between exact and approximate working.

The process works in two ways: the technology shapes the user’s mathematical knowledge and understanding, and at the same time the user makes use of their own mathematical knowledge and understandings to use the technology to perform meaningful mathematical tasks. This two-way process has been termed ‘instrumental genesis’: see, for example, Guin and Trouche (1999).

When senior mathematics was a minority interest and by and large the majority of students studying advanced mathematics had a natural interest and aptitude for it, we could (perhaps) trust students to make adjustments of notation more or less painlessly; but that world has long since gone. In the greatly expanded senior mathematics classes of today’s schools, and in the early tertiary environment, we are increasingly finding large numbers of students for whom these notational inconsistencies and adjustments are themselves problematic. Standardized uniformity will never be possible, even within one culture, let alone internationally. The details may have changed, but the situation described by Cajori nearly a century ago is still worth quoting:

Uniformity of mathematical notations has been a dream of many mathematicians — hitherto an iridescent dream. That an Italian scientist might open an English book on
elasticity and find all formulae expressed in symbols familiar to him, that a Russian actuary might recognize in an English text signs known to him through the study of other works, that a German physicist might open an American book on vector analysis and be spared the necessity of mastering a new language, that a Spanish specialist might relish an English authority on symbolic logic without experiencing the need of preliminary memorizing a new sign vocabulary, that an American traveling in Asiatic Turkey might be able to decipher, without the aid of an interpreter, a bill made out in the numerals current in that country, is indeed a consummation devoutly to be wished. Is the attainment of such a goal a reasonable hope, or is it a utopian idea over which no mathematician should lose precious time? (Cajori, 1929, para 740)

The only remedy for the here and now is to be aware of the difficulties and to address them openly and systematically: in fact to teach notation in the new contexts and to set explicit exercises in making the necessary changes.

Anon (1915) Suggestions for Notation and Printing. Mathematical Gazette VIII n120 p. 172


