Success and Failure in Arithmetic and Algebra

David Tall

Mathematics Education Research Centre
University of Warwick
COVENTRY CV4 7AL

Introduction

Arithmetic and algebra are central parts of the National Curriculum throughout the British Isles. In both of these there is a known and disconcerting level of failure, in algebra even more than in arithmetic. Why does this failure occur? Or, to look at the more positive side of the coin, why are some fortunate souls able to do these parts of mathematics almost without effort. By the conservation of energy, if there is no effort, there can be no work done. So, by implication, there must be some other source of energy within the more able that drives their success. Research performed recently at Warwick (Thomas 1988, Gray 1991a, Gray 1991b, Tall & Thomas 1991, Gray & Tall, 1991) reveals that there is indeed a qualitative difference between the thinking processes of those who succeed and those who fail, a difference that makes the mathematics easier for the more able and harder for the less able, exacerbating the chasm between them. The more able develop a way of thinking that fires an inner engine with a feed-back loop creating new knowledge from old, the less able seek solace in being able to carry out procedures that may be successful in the short term but are likely to lead to long-term failure.

Examples of success and failure in arithmetic

Let us begin by looking at how children succeed or fail in simple arithmetic. The examples are taken from Gray (1991a).

Stuart (aged 10) responded to the problem 8+6 by saying "*I know 8 and 2 is 10, but I have a lot of trouble taking 2 from 6. Now 8 is 4 and 4; 6 and 4 makes 10; 10 and another 4 makes 14*".

Stuart is successful, but knows few number bonds, and has to search through his small repertoire to try to solve the problem. He is extremely creative in the mathematics that he is doing, but his methods are arduous and likely to come under considerable strain when he tries more complicated tasks.

Michelle (aged 10), faced with "18–7", said "*ten from eighteen leaves eight, seven from ten leaves three, eight and three makes eleven*". Michelle, like Stuart, seeks to find familiar number bonds to solve the problem. She sees 18 as 8 and 10, but takes the 7 from the 10 rather than from the 8.

Michael (aged 9) Michael chose to write 18–9 as

$$\begin{array}{r} 18\ - \\ \underline{9} \end{array}$$

and, as is usual in the decomposition process, put a 'little one' by the eight. "*This is the easy way of working it out. I can't take nine from eight but if I put a little one it makes it easier because now its nine from eighteen*".

He didn't seem to realise that this was just the same problem all over again. After some considerable time he resorted to his more usual procedure for subtraction from teens. Eighteen marks were placed from left to right on his paper and then starting from the left hand side he crossed out nine marks, counting from one to nine as he crossed out. Recounting from the left those original marks not crossed out, he was able to provide the correct solution.

All these three children were considered "less able" by their teacher, yet were successful at carrying out the arithmetic tasks in their own way. Two progressed by deriving facts from known facts, the third reverted to counting. Amongst less able children, use of known facts to derive facts is rather rare. A more likely tactic is to count. But, as children grow older, counting on fingers becomes *de rigeur* so they must invent new methods to extend their earlier counting procedures.

Jay (aged 10) rejected standard concrete materials, "I'm too old for counters", but neither did he like using his fingers, "my class don't use counters or fingers". For numbers up to twenty he casually splayed his ten fingers on the edge of his desk and imagined another ten fingers to extend his counting techniques.

Gavin (aged 9) "liked counting with his fingers – that is what they are made for", but for problems up to twenty he assigned numbers in the teens to various parts of his body in a clockwise fashion from left shoulder, to waist, to thigh, to calf and ankle, then up his right side. "I've only got ten fingers; I count as if I had a never-ending load".

Philip (aged 8) solved his physical counting another way, using toes to supplement his fingers, though this proved problematic when attempting to move his middle toes.

It can be seen that these creative methods of counting, which extend the physical counting processes of early childhood to larger numbers, are fraught with difficulties and may be leading down a cul-de-sac of failure. To find the source of these difficulties it is useful to go back and consider the ways in which the number concepts arise in the child's development.

Procedures and concepts

Initial success in mathematics comes about through being able to *do* things. One of the first of these is *counting*. The young child is given experiences that lead to the routine of counting "one, two, three, ..." and the more subtle idea that when these words are spoken while pointing successively at each object in a collection, then the last word spoken is the number of items in the collection. The concept of a number such as "five" is therefore associated with an underlying procedure. Yet the symbolism "five" or "5" takes on a life of its own because it can be spoken, it can be written, it can be seen, it can be heard. It takes on a concrete existence which embodies within it both the procedure of counting and the concept of number.

Addition is initially an extension of counting. The sum "3+2" is first attacked by counting three objects, then two more, then putting the two collections into a single collection and counting them all to get "1, 2, 3, 4, *5*". This first manifestation of counting is called *count-all*. It is a succession of three counting procedures.

With experience the child comes to realise that a sum such as "3+2" does not require the first three objects to be counted a second time. Indeed, it is only necessary to *count-on* two more from 3 as "4, *5*". This, however, involves a double-counting process. As the next two numbers in the continuing number sequence "4,5" are spoken, it is also necessary to maintain a count of how many of these are counted. Often this is done using physical objects – fingers, unifix cubes, numbers on a ruler – so that one of these counting processes is done mentally, whilst the concrete objects are used to keep track of the other. Calculating 3+2 on a number line is done by pointing at 3, then counting on 2 more, which ends up pointing at the result, 5.

Notice here that "count-on" treats the first number 3 as a mental or physical entity, then uses the second number to evoke a counting procedure. Numbers are used once more on the one hand as concept and the other as procedure.

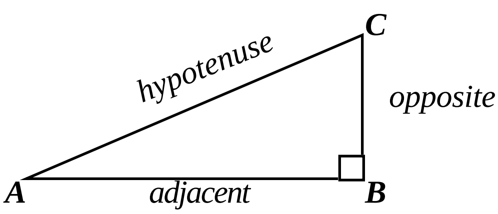
Experience usually leads on to the encapsulation of the addition 3+2 as a "known fact": "three plus two is five". Such known facts can be learned in two distinct ways: by rote, or in a meaningful way. Our earlier examples show that less able children may be hampered by knowing fewer facts, so learning number facts would seem to be of great value. However, it is also clear (as in the case of Michelle) that, even when they have the facts available to them, they may lack the flexibility to use them in the most economical and productive way. Thus, although rote-learning of facts may increase the foundation on which to build, the meaningful learning of facts is essential for flexible thinking.

Flexibility of mathematical notation

The more able work in a much more flexible way. The fact “3+2 is 5” is seen to be the same as a whole cluster of related facts “2+3 is 5”, “5–3 is 2”, “5–2 is 3”. Such facts can lead easily to new facts, for instance “I know 4 and 4 is 8, so 4 and 5 is 9”. This flexibility is well-known. What is less known is that it depends on a dual use of the symbolism which is *ambiguous*. The symbol 3+2 stands both for a *procedure*, the procedure of addition through counting, and also for the *result* of that procedure. The symbolism evokes both process and product. This dual usage of symbolism for both procedure and concept is found throughout mathematics. Yet, because of the mathematician’s desire for precision and rejection of ambiguity, we have failed to fully understand this duality and ambiguity of symbolism, which gives it such flexibility.

Here are a few instances:

- 3+2 represents both the procedure of addition and the concept of sum,
- 3x2 represents both the procedure of multiplication (through repeated addition) and the concept of product,
- +2 represents both the procedure of “add two” (or shift two units to the right on the number line) and also the concept of a positive signed number,
- –2 represents both the procedure of “subtract two” (or shift two units to the left) and also the concept of negative number,
- $\frac{3}{4}$ represents both the procedure of division and the concept of fraction,
- $\sin A = \frac{\text{opposite}}{\text{hypotenuse}}$ represents both the procedure of calculating the trigonometric ratio and also the concept of sine,



- $\pi = 3.14159\ldots$ represents the procedure of calculating π as a succession of more accurate decimals and also the value of π, indeed, the left-hand side of this equation seems to be concept and the right hand side a procedure of approximation,
- In the calculus the notation $\lim_{x\to 1} \frac{x^2-1}{x-1}$, represents both the procedure of tending to a limit and the value of that limit,
- so does the notation $\lim_{n\to\infty} \frac{1-x^n}{1-x}$
- and $\sum_{n=1}^{\infty} a_n$
- and $\lim_{\delta x\to 0} \sum_{x=a}^{b} f(x)\,\delta x$.

Given this widespread phenomenon of the duality and ambiguity of mathematical notation as procedure and concept, it is quite amazing that it has not been named. I suspect this is because we first observe the specific and evident and much later focus on the subtle and generative deeper concepts. But once these deeper concepts are named, it is amazing how much power they give us in terms of explanation and prediction.

Procept

The amalgam of *pro*cedure and con*cept* which is represented by the same notation is defined to be a *procept*. Once the term has been verbalized it assists in explaining what is going on in the learning of mathematics, or rather the learning of mathematical procepts. For instance, *number* is a procept, evoking both the procedure of counting and the concept of number. *Addition* is a procept, which operates on the procept of number. The various levels of the encapsulation of the procedure of counting to the concept of sum can be seen to be successively sophisticated growth of the procept of sum.

From procedure to procept in arithmetic

Count-all consists of three procedures : count one set, count the other, then count the combination. However, it is something that happens *in time*. The numbers to be added are input several seconds before the sum is output, so the child performing the procedure successfully may not develop the linkage between input and output that is crystallised as a "known fact".

Count-on views the first number as concept and the second as procedure, using a double-count procedure to give the output. The count-on procedure is more complex, but it does reduce the number of steps in the procedure to give a greater possibility of linking the two input numbers with the output as a known fact.

If input and output become linked and remembered, then the resulting known fact has a proceptual quality. It is both procedure and concept.

The divergence between procedure and procept

The difference between procedure and procept leads to the qualitative difference between the less successful and the more successful:

> *The more successful see addition as a flexible procept, others see it as a procedure that occurs in time, either as count-on or count-all.*

The more able develop a proceptual system of deriving new facts from old and have an inbuilt feed-back loop that creates new number facts. The less able are locked into a procedural system in which they are faced with harder and harder procedures of counting.

Figures 1 and 2 give empirical support for this hypothesis. Seventy two children were selected by their teachers in two "typical" schools to represent the chronological ages 7+ to 12+, with each school providing three pairs of children in each year to represent the below average, average, and above average attainers (Gray 1991a). These children were interviewed individually for half an hour on at least two separate occasions a week apart, and in each session were asked to solve between eighteen and twenty arithmetic problems at various levels of difficulty. Figure 1 (Gray 1991b) illustrates the different strategies used by children of differing abilities in solving single-digit addition and subtraction problems.

Note the almost complete absence of derived facts in the less able (particularly in addition), whereas the average and above average start with a high proportion of known facts and use derived facts to generate other facts. As the ages of the children increase, the proportion of known facts increase, but to a lesser extent in the less able.

Figure 2 (taken from Gray & Tall, 1991) shows the total range of strategies used by more able and less able children in the ages 8+ to 12+ for specific subtraction problems whose answer is not a known fact.

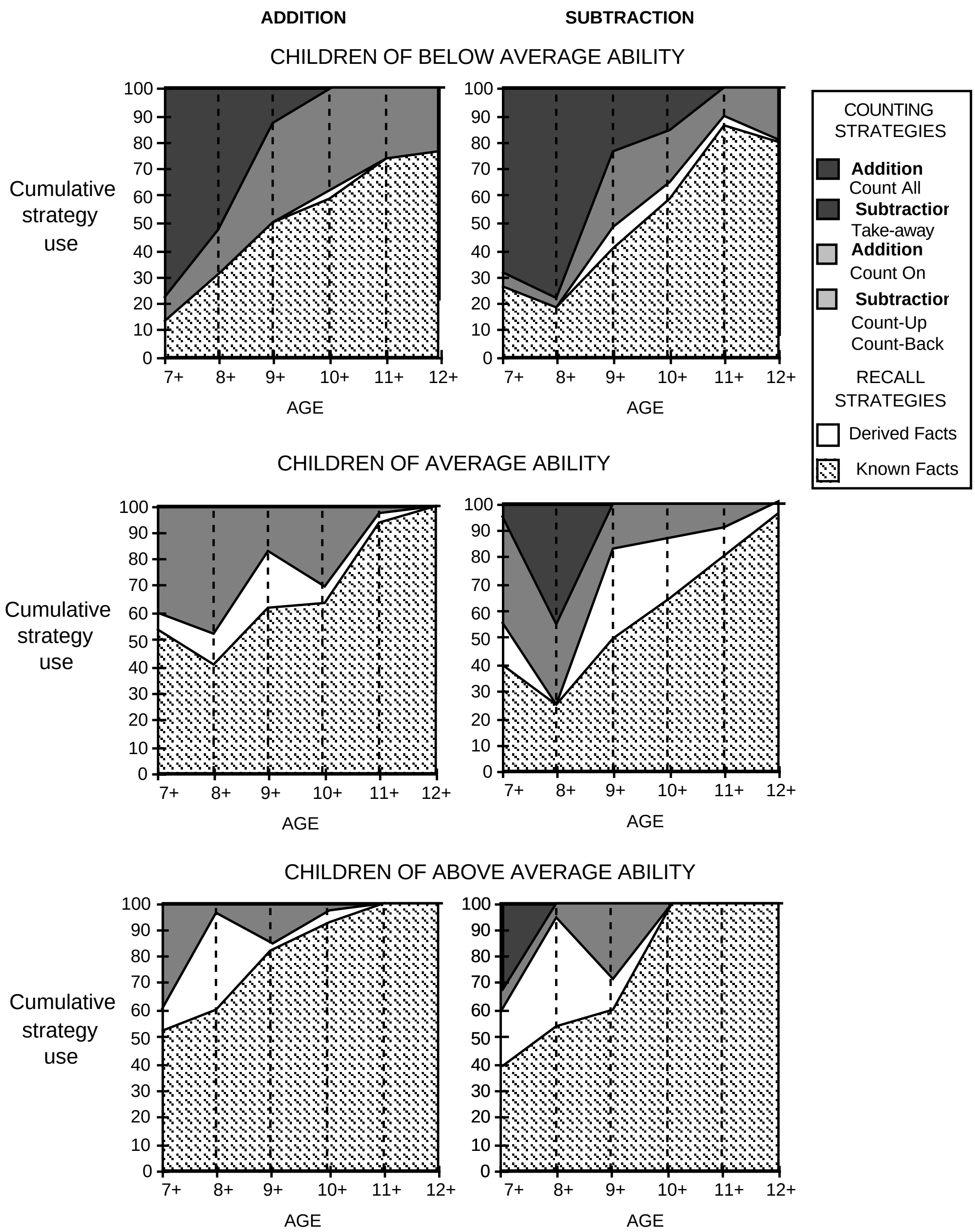


Figure 1 : Strategies for solving addition and subtraction involving numbers up to ten

The left hand side of figure 2 shows the above average children using almost all derived facts and a few examples of counting, whilst the right hand side shows the below average children using few derived facts and a large percentage of counting, take away, and errors. This graphically illustrates the qualitative divergence in thinking processes.

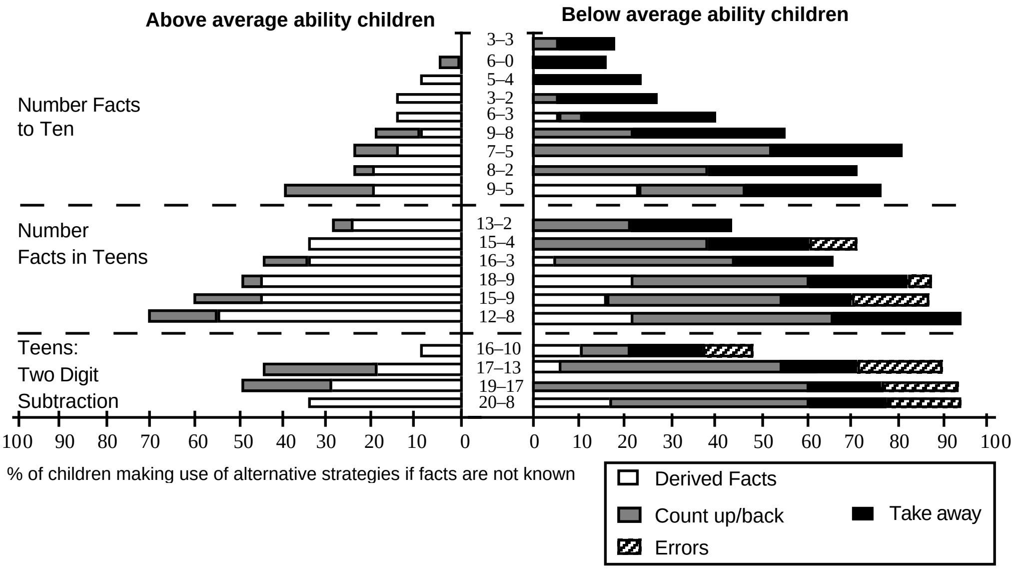


Figure 2: Strategies for solving problems whose answer is not immediately known

Problems in the initial encounter with algebra

Children meeting algebra for the first time often have great problems in understanding the meaning of the notation. They may write *x* and *y* next to each other as *x y* and think of it as "*x* and *y*". But they are told that *xy* is "*x times y*". They may be confused by the meaning of a symbol such as 2+3*x*. If it means the procedure "add 2 to 3 times *x*" then there is a problem that it cannot be calculated until *x* is known. On the other hand, if *x is* known, why use algebra anyway? The algebraic expression is difficult to interpret. Read from left to right in the usual way it says "two plus three times *x*". Because 2+3 is 5, children may think that 2+3*x* is 5*x*. But it isn't. It is all mumbo-jumbo. Algebra soon becomes a meaningless manipulation of meaningless symbols, each week of study bringing a new procedure to carry out "collect together like terms", "do operations inside brackets first", "do multiplication before addition", "do the same thing to both sides", "change sides, change signs", …

There are many problems here. But a major difficulty is the underlying meaning of the notation. The expression 2+3*x* means two different things, it is the *process* of adding together 2 and 3 times *x*, and the *product* of that process. In other words, it is a *procept*. Children who see 2+3*x* only as process find it difficult to understand because it is a process that they cannot carry out until they know the value of *x*. They try desperately to swallow their difficulties and to cope by carrying out the procedures they practice each week in manipulating algebra.

Having seen the difficulties in arithmetic with coordinating different procedures, even greater difficulties occur in algebra. If 2+3*x* is a procedure, how does a child cope with factorising an expression like

$$3(2+3x)+2x(2+3x).$$

A child who can see the expression $2+3x$ as an entity can collect together terms to get $(3+2x)(2+3x)$. A child who cannot do this may seek security by following the rules: multiply out brackets, collect together like terms, look for a factorisation of the resultant quadratic $6+11x+6x^2$. Once more we see the divergence between procept and procedure. The mathematics involving procept, using the notation to represent either procedure or concept, is flexible and powerful. The mathematics involving only procedures is more complicated and lacks insight.

The procedural conspiracy

As educators we want to help our student *do* mathematics. Yet herein lies a dangerous implicit conspiracy between pupils and teachers. When a child cannot cope with mathematics the cry is "show me how to *do* it". When the chips are down, in a large class with pressures all around, it is natural to do just that. And, for a time, everyone is happy. The child has instant gratification, the teacher is pleased that the child can do something, parents and politicians are satisfied that progress is being made. Yet, if we simply show children the *procedures* of mathematics, we may end up by confining them to a cul-de-sac of mathematical short-sightedness which ends up in terminal failure.

Are there solutions?

We should not assume that all problems have solutions. If a child is more capable of holding several things in the mind at once and more able to compress this knowledge to treat it as a single piece of information to be mentally manipulated, then we should not assume that all other children either have, or can be educated to have, this capacity. The difficulties of Stuart and Michelle mentioned earlier showed that they had the idea of deriving new facts from old, but they didn't do it in the most economical way, so they made rods for their own backs. By helping them become more aware of this problem, and showing them new strategies, we may be giving them too much detailed information which only serves to obscure the conceptual simplicity seen by the more able.

However, there *are* possible ways ahead using the new technology. If a procedure can be formulated mechanically then it can be carried out on a computer. *All* the procepts discussed in this article are of this type. Therefore it becomes an educational possibility to get the computer to carry out the procedures to allow the individual to concentrate on the higher level relationships involving the objects produced by the procedures.

Arithmetic and the calculator

With arithmetic, this means allowing the child to use a calculator, and helping the child focus on meaningful relationships. For instance, using the facility on modern graphic calculators, several successive calculations remain onscreen at a time. Thus it is possible to use a display such as figure 3 to allow the child to be able to focus on the nature of the relationship rather than becoming embroiled in the details of the creative procedures some of them use to carry out the given calculations.

```
8–5
                  3
18–5
                 13
28–5
                 23
```

Figure 3 : representing numerical relationships on a suitable calculator

It is a pity that the calculators usually employed in the classroom only allow a single number to be input at a time, forcing the child to see the arithmetic as a process between distinct numbers in time rather than seeing the relationships represented together in a single display. Even given the inadequacies of current calculators, the CAN project has shown that children who use calculators get a better appreciation of number concepts and fare as well, or better, in knowing standard number facts.

Introducing algebra with a computer

The difficulty of seeing 2+3*x* as a single expression, rather than as a process to be carried out can be greatly assisted using a computer.

It is a simple task to program the computer in BASIC, to type in the command X=3, and then ask a child what happens when we type in the command PRINT X+1. When the computer prints 4, and the phenomenon is repeated with other expressions, it becomes easy to predict what will happen with PRINT X+3 or PRINT 2*X.

We may now investigate what happens with 2+3*X. Here there may be different opinions, for instance, when X is 3 then “2+3 is 5, so the result is 5*X, which is 15”. Testing this on the computer shows a different result, 11. How can this be obtained? A discussion is naturally initiated on how the computer carries out the computation, and it occurs in a sensible context where the computer is using rules that can be articulated, predicted and tested. Algebra in this context makes sense because it is part of the language of communication with a computer.

The Mathematical Association (1989) publishes notes for an approach to algebra using this method. The approach uses a single computer for initial discussion between teacher and class. It needs access to one or two more computers to give the class a chance to explore the ideas themselves. But it also uses “cardboard computers” which are designed to enable the child to play the game of internally storing the numbers in boxes labelled with the names of variables.

The cardboard computer consists of two pieces of card. One represents the screen, on which a sequence of instructions is placed; the other represents the internal storage of the computer with boxes which can be labelled with a letter and a number placed inside (figure 4). For instance, the assignment A=1 is carried out by labelling a box with the letter A and placing 1 inside. B=A+3 requires the operators to find the value of A (which is 1), label a box B and place the value of the sum A+3 (which is 4) inside. PRINT B+2 then requires the operators to find the value of B (which is 4) and print the value of 4+2 (which is 6). Experience shows that the children enjoy the fun of playing with the cardboard computer every bit as much as the real ones!

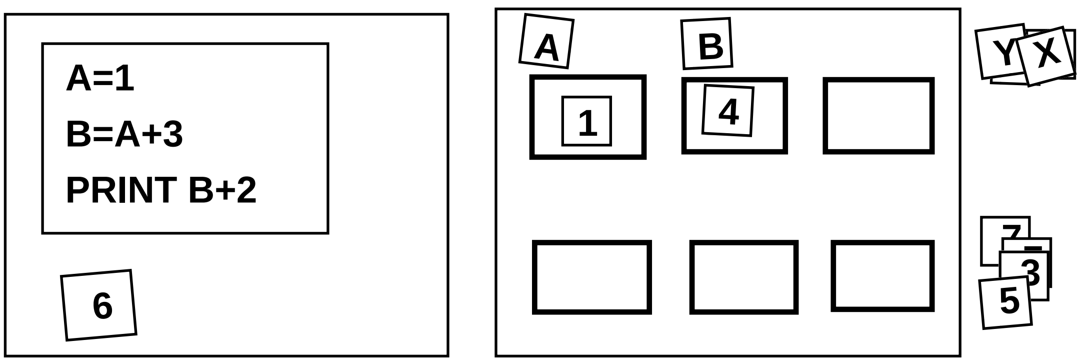


Figure 4 : Using the Cardboard Computer

Notice that the cardboard computer requires the pupils to carry out the specified procedures and actually calculate the value of A+3. On the other hand, the BASIC program carries out the internal procedures of numerical computation and allows the user to consider the outputs of these procedures.

For instance, the program:

```
10 INPUT X
30 PRINT 2+3*X
40 PRINT 5*X
```

requires a number to be input, but then performs the required calculations and prints two numbers. Here the child can see that the two outputs need not be the same and that the value given to 2+3*X is consistent with adding 2 to 3*X , not adding 2 and 3 and then multiplying by X.

But

```
10 INPUT X
20 INPUT Y
30 PRINT 2*(X+Y)
40 PRINT 2*X+2*Y
```

always prints the same value. The expressions 2*(X+Y) and 2*X+2*Y are seen to have the same effect, even though they involve a different sequence of calculations.

Of course it would be useful to print the symbols in normal algebraic notation. The *Algebraic Calculator*, published in the Mathematical Association Pack, allows just this. So it is possible to input a value for x and to print $2+3x$ and $5x$ to see that they are different. The version published by the M.A. is currently for BBC computer only. A more powerful program, the *Function Calculator* is available from Cambridge University Press for BBC, Master, Nimbus and Archimedes computers as part of the *Real Functions and Graphs* package. Figure 5 shows a display of the *Function Calculator* which encourages a discussion of the meaning of the expression 2+3a and the fact that it is not calculated in the same way as 5a.

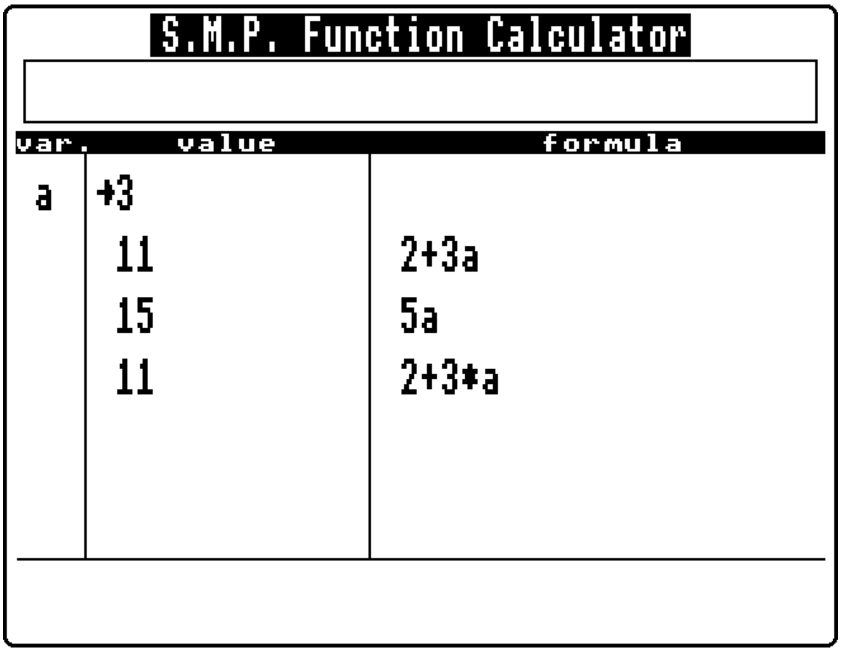


Figure 5 : What does 2+3a mean?

Thus we see programming as an interactive exercise to give meaning to the conventions of algebraic symbolism using computer notation, with the *Function Calculator* focusing on standard algebraic notation. Whilst both of these systems carry out the procedure of calculation and allow the user to focus on the meaning of the expressions, the cardboard calculator focuses on the internal procedures themselves. So the different activities can focus on different aspects at different times and considerably reduce the cognitive strain.

Evidence from Tall & Thomas (1991) shows the benefits of this approach. Pupils using this approach for three weeks may initially not be as good at conventional manipulation as those who have had an equal amount of time devoted to standard practice of algebraic techniques. But long-term they make up the deficiency in technique and show a far higher understanding of the flexible nature of algebraic symbolism.

From the many pieces of evidence quoted in this paper, let me report just one. Pupils in interviews were asked to solve the equation $3p–1=5$. Both experimental pupils (who had used the computer) and control pupils (who had concentrated on techniques) were able to solve this. But the control pupils used procedures "add one to both sides, divide both sides by 2" to get the answer. The experimental pupils were more likely to say "if $3p–1=5$, then $3p$ has got to be 6, and so p must be 2." Faced subsequently with the equation $3s–1=5$, the control pupils may have realised that the equation was similar, but they still needed to go through the procedure to get the result $s=2$. The experimental pupils were more likely to say "it's the same equation".

A little later in the interview the children were asked to solve

$$3(p+1)–1=5.$$

Several experimental students said something like "it's the same equation; $p+1$ is 2, so p is 1". But none of the control students said this. Instead, those who tried to solve the equation "multiplied out brackets, collected together like terms, subtracted 2 from both sides, divided both sides by 3 and found that p is 1".

The difference between the flexible, proceptual approach of the experimental students who saw the equation as being essentially the same each time, and the procedural approach of the control students who attempted to solve the equation is clear.

A way ahead

The experiences using a calculator in arithmetic and a computer environment in algebra suggest a general principle. The computer (and its more primitive relative, the calculator) can carry out routine procedures, allowing the student to focus attention on the objects produced. In this way students, who might otherwise become focused on the procedural aspects, can be refocused on the concepts without the strain of carrying out the procedures (in a possibly idiosycratic way). Sometimes one can concentrate on the procedures, and on others, using the computer, one may concentrate on the concepts produced by the procedures, without needing to carry out the procedure at the same time.

In this way one might be able to lessen the gap that occurs between the procedural thinking that gives short term results and the proceptual thinking that gives the flexible thought processes characteristic of the successful mathematician.

References

Gray, E. M. ,1991a: *Ph. D. Thesis*, University of Warwick.

Gray, E. M., 1991b: 'An analysis of diverging approaches to simple arithmetic: preferences and its consequences", *Educational Studies in Mathematics* (in press).

Mathematical Association, 1989, *Secondary Mathematics with Micros - A Resource Pack*.

Tall, D. O., 1991: *Real Functions and Graphs* for the BBC computer, Master, Compact, Nimbus PC & Archimedes (A3000) computer, Cambridge University Press.

Tall, D. O. & Gray, E. M., 1991: "Duality, Ambiguity & Flexibility in Successful Mathematical Thinking", Proceedings of *PME 15*, **2**, Assisi, Italy, 72–79.

Tall, D. O. & Thomas M. O. J., 1991, "Encouraging Versatile Thinking in Algebra using the Computer", *Educational Studies in Mathematics*, 22 2, 125–147.

Thomas M.O.J. 1988: *A Conceptual Approach to the Early Learning of Algebra Using a Computer*, Ph.D. thesis, University of Warwick.