MAKING MATHEMATICAL NOTATION MORE MEANINGFUL

By GARY PERLMAN
University of California, San Diego
La Jolla, CA 92093

Consider the formula for the arithmetic average,

$$(1) \qquad \frac{\sum_{i=1}^{i=n} X_i}{n},$$

which states that the average of a set with n members, $X_1, X_2, X_3, \ldots, X_n$, is the sum of all its members divided by n. If we assume students can add and divide, then the ability of students to learn how to compute an average depends largely on their ability to decipher the meanings of all the symbols in (1). I believe there are some rules that can simplify the task of communicating mathematical ideas with symbols. For example, to simplify (1), two appropriate rules are the following: *symbols should be familiar* (out with sigma), and *symbols should be mnemonics* (in with SUM). Another notational rule is the following: *a notation should promote generalization.* This suggests the following scheme appropriate for the computer age.

$$(2) \quad \text{AVERAGE}(X) = \text{SUM}(X)/\text{SIZE}(X)$$

Although (2) could be shorter and is not as powerful as general summation notation, its operations are defined in a common functional notation, and their meaning can easily be remembered because of their names.

Introduction to Notation

A notation is a set of symbols and rules for combining symbols to represent mathematical ideas. A notation is often a shorthand for expressing mathematical **ideas,** but it can also be instrumental in facilitating students' understanding and use of mathematics. A notation should not serve merely as an aid to compaction; it should convey meaning by itself and, in doing so, be an aid to solving problems. From Alfred North Whitehead in *An Introduction to Mathematics:*

> By the aid of symbolism, we can make transitions in reasoning almost mechanically by the eye, which otherwise would call upon higher faculties of the brain. By relieving the brain of all unnecessary work, a good notation sets it free to concentrate on more advanced problems. [1911, chap. 5]

In fact, many advances in mathematics have been due to suggestive notations. For example, expressing the rule "the product of the mth power of x with the nth power of x is equal to the $(m + n)$th power of x" with abstract symbols

$$x^m x^n = x^{m+n}$$

led to the discovery that the square root of x is equal to the 1/2 power of x,

$$x^{1/2} x^{1/2} = x^1.$$

Previously, no one had thought of using nonintegral exponents.

Notation affects both the learning and the creation of new mathematics. Despite the apparent importance of notation, there has been little research on what makes a good notation. In the following sections I propose some rules for denoting things, properties of things, and relations between things. Satisfying some rules violates others, a fact that suggests there are no perfect notational schemes, though some are better than others.

A notation should be concise

A notation is generally used to express ideas with few symbols and in a small space, for fast and easy study. Without this rule, we could satisfy all other rules by writing out in long wordy English all the ideas we wanted to convey. A problem with verbose descriptions is that a complicated idea generally requires many words to describe. Because we have limited short-term memories, we have difficulty understanding ideas presented in this way. By packing a lot of information into a small space, more information is available at a glance, facilitating the observation of relations between ideas.

For example, it is difficult to grasp the meaning of the following familiar theorem when it is stated in English.

> *In all triangles with a right angle, the product of the length of the side opposite the right angle with itself is equal to the sum of the products of the lengths of the other two sides, each with itself.*

This verbose version of the Pythagorean theorem does not make use of some English conventions about naming. Although there is a cost associated with the introduction of any new notation, commonly denoted things should have their own names. A triangle with a right angle is called a right triangle; the product of anything with itself is called its square; and the side opposite the right angle in a right triangle is called the hypotenuse. With this bit of notation, the theorem becomes the following:

> *In a right triangle, the square of the length of the hypotenuse is equal to the sum of the squares of the lengths of the other two sides.*

By using some sloppy conventions, we can assume that right triangles are implied anytime anyone talks about hypotenuse, and we can ignore the distinction between the length of a side and the side itself. The following short form is then obtained.

> *The square of the hypotenuse is equal to the sum of the squares of the other two sides.*

To further compact the theorem, abstract symbols can be introduced (e.g., "=" for "is equal to"). One might think it is a good idea to introduce as many symbols as there are concepts, but the wisdom of introducing loo many symbols has been questioned (Kline 1973). It requires effort to learn the meaning of new symbols. If a concept is used frequently, then a notation specifically for it is appropriate, but symbolism just for the sake of using symbols can be an unnecessary burden' on memory.

A notation should be precise.

By identifying the length of a side with its name, the final short form is technically incorrect. 1 do not see this as a lack of precision because people understand the convention. Precision is achieved when people understand the message, not when pedantic mathematicians cannot find fault.

When needed, precision can be gained at little cost. For example, parentheses can often be used to remove ambiguities when there are no conventions to follow. If we did not have the convention that arithmetic expressions are evaluated from left to right with some operations having precedence over others, we could clarify the order of operations of

$$a + b \times c/d$$

with parentheses to obtain

$$(a + ((b \times c)/d)).$$

A notation should promote generalization.

Similar ideas should have similar notations. How one idea is represented should be consistent with how related ideas are represented, and it should provide clues about how they are represented. In analytic geometry, (x, y) is used to denote a point in the real plane. To denote a point in space, the notation (x, y, z) generalizes from (x, y) in two ways. One is that the third dimension is added on at the end of x, y so the two-dimensional representation is embedded in the three-dimensional; and the other is that z follows xy in the alphabet as well as in the notation.

In computer programming languages, where functions may have many parameters, an example of bad notational form would be to have two related functions with the same parameters but where the order of the parameters does not agree. A matrix inversion program might have two parameters, a matrix and its size,

invert (matrix, size),

whereas, to compute the determinant another function might take the same parameters but in the opposite order,

determinant (size, matrix).

Either form can be justified, but their combination cannot. Knowing the form of one function does not help but hinders the learning of the other.

Subscripts, although they add some complexity by making the symbols harder to learn, do promote generalization as well as minimize the total number of symbols introduced. If a sequence is represented by

$$X = (X_1, X_2, X_3, \ldots, X_i, \ldots, X_n),$$

we have a way of talking about a sequence, its first, last, and even ith elements in a general manner.

Symbols should be mnemonics.

Arbitrarily chosen symbols will not serve as mnemonics (aids to memory) and so will be harder to learn than ones well chosen. A symbol should have something in common with the object it represents. If a symbol is an aid for remembering what it stands for, it will be easier to learn and use. It may be an abbreviation of an existing name as in representing the number of dollars with d, or it may even **look** like the concept it denotes as in the case of using $\|$ for parallel.

Abbreviations help us remember a new symbol in terms of an existing symbol for an object, a symbol that often has nothing to do with that object. Using d for dollars is a good choice because *dollars* begins with d; however, there is no natural connection between the word *dollars* and the concept of dollars. So whether a mnemonic has anything to do with the concept it denotes is not as important as having a relation to some existing attribute such **as** its name. The strongest connection occurs with symbols for geometric objects where symbols look like the things they denote. For example, **A** can be used to represent triangle.

To help students learn notation, we should supply a rationale whenever new notation is introduced. Otherwise, a symbol's mnemonic value may not be perceived. The rationale might be the truth about the matter, such as the historical reason for the choice, or it can be a simplification of the truth to facilitate instruction. For example, a rationale like the following might be given to aid the learning of the relational operators *greater than* and *less than*, $>$ and $<$, respectively.

> Relational operators are put *between* their operands because they represent relations *between* operands. The smaller quantity is **put** on the smaller side of the symbol.

Whether this is the real reason for the choice of symbols is not really important. The rationale has given a unified explanation that serves as an aid to memory, so it is all that is needed.

An example of innovation in choosing mnemonic symbol names is in computer programs where names rather than single symbols are used for functions and variables. Earlier, I used the notation

invert (matrix, size)

to denote a matrix program. The choice of names can be arbitrary, but *invert* for the program, *matrix* for its matrix argument, and *size* for the size of the matrix are all useful mnemonics. These symbolic names are not as short as possible but are a great aid to understanding the meaning of statements and will be used with similar success in all mathematics as computer science and mathematics education become more integrated.

Symbols should be familiar.

Symbols themselves should be easy to remember so they should have already been encountered. They should be pronounceable, writable, and visually distinct. If they are not, they may be confused or forgotten. Obscure symbols should be avoided. If they must be introduced, their English names, along with their pronunciation, should accompany them. Otherwise, the use of special symbols may confuse students, possibly leading to what is informally called ***math anxiety.*** A common source of frustration is the use of the Greek alphabet. These symbols are both hard to distinguish and hard to pronounce, and learning concepts expressed with them becomes a memory test rather than a learning experience.

Symbols should be unique.

Symbols and concepts should be in one-to-one correspondence within a given mathematical topic. The same concept should not have two representations and the same symbol should not be used to denote more than one object. But because it is desirable that similar concepts be similarly denoted and because there are a limited number of similar representations, it is common for one symbol to have more than one meaning. Multiple meanings for symbols usually do not present a problem of precision if the student knows there can be more than one meaning for symbols and is able to determine which meanings are meant by the contexts in which they appear. Skemp (1971) suggests using unique meanings of symbols *inside* a particular domain, but the same symbol can have different meanings in *different* domains if it is clear to the student to which domain the teacher is referring.

A symbol that has many meanings is the equal sign. In some cases it indicates identity:

$$\text{if } i = 0, \text{ then } \ldots .$$

In others it means assignment:

$$x = y + z.$$

It can also be used to define symbols:

$$t = \text{elapsed time.}$$

The meanings in all cases are highly related and promote generalization by their similar (identical) representations.

The converse of multiple meanings is also common. The appearance of synonymous symbols is usually due to similar theoretical developments on several fronts. The most commonly encountered synonyms are those for multiplication,

$$a \times b \quad a \cdot b \quad ab \quad a * b,$$

and division,

$$a \div b \quad a{:}b \quad a/b \quad \frac{a}{b}.$$

Learning several different symbols for the same objects is usually a waste of time and effort.

Existing notation should be maintained.

> There is no hope of replacing long established notation — the only hope of establishing a good notation is at the outset. [Richard Skemp 1971, chap. 5]

The adoption of *notational conventions* makes it easier for people to communicate without prefacing each communication with definitions. For example, we can't have someone telling us that + will be used for multiplication; the symbol is too deeply entrenched in our educational system. The use of conventions also helps us avoid synonymous symbols that occur when different authors use personal favorite notations.

Format should reflect organization.

I have so far concentrated on symbolic notation, but graphical devices are also useful to convey meaning. Nicely formatted text facilitates understanding because its organization can parallel its content. Rules for good format include placing one idea per line and using indentation and columnation to indicate structure. In the following example, by having the ith term in S under the $(i + 1)$th term of $2S$, the trick of the proof is made transparent.

THEOREM. *The sum of the inverses of the positive powers of* 2 *is 1.*

Proof. Let $S = 1/2 + 1/4 + 1/8 + 1/16 + \cdots$.

Using subtraction, we have the following:

$$\begin{array}{rl} 2S = & 1 + 1/2 + 1/4 + 1/8 + 1/16 + \cdots \\ - \quad S = & 1/2 + 1/4 + 1/8 + 1/16 + \cdots \\ \hline S = & 1 \end{array}$$

Concluding Remarks

> When a code is familiar enough, it ceases appearing like a code; one forgets there is a decoding mechanism. The message is identified with its meaning. [Douglas Hofstadter 1979, p. 267]

We can accept almost any name for a concept. For complicated thoughts, careful choice of form is crucial. Incongruities between the structure of expressions and their meaning not only may make learning more difficult but also may perpetually be the cause of errors.

There is a strong analogy between learning the language of mathematics and one's mother tongue. Vocabulary and grammar must be learned before meaningful communication can be achieved. When students begin learning an area of mathematics, a large proportion of effort should be devoted to teaching notation. By the time students get to more advanced courses in mathematics, the time devoted *to* introducing notation gets shorter, but no less crucial. Even the simplest of ideas cannot be communicated if the teacher and student are not "speaking" the same language.

What implications do these rules have for education? We should have them clear in our minds so that we can make communication of mathematical ideas easier for our students. Whether it would be useful for students to learn about general properties of notational schemes is an open question. Clearly, they must understand the notation system being used. and the only way this can be done, **as** with any language, is through practice.

ACKNOWLEDGMENTS

This paper was inspired by T. C. Hu. Thanks are due to Jean Mandler, Jay McClelland, Don Norman, Dave Rumelhart, and especially Tom Erickson for their helpful comments on earlier drafts. Referee David J. Glatzer was helpful in pointing out places where J violated some of *my* own rules. This paper was written while I was supported by a Natural Science and Engineering Research Council of Canada Postgraduate Scholarship. Partial support was also provided by the Office of Naval Research under contract NOOO14-79-C-0323, NR 157-437, and from the Advanced Research Projects Agency, monitored by ONR under contract NOOO14-79-0515, NR 157-434.

REFERENCES

Hofstadter, Douglas **R.** *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Basic **Books.** 1979.

Kline, M. *Why Johnny Can't Add: The Foilrrre of the New Moth.* New York: St. Martin's Press, 1973.

Skemp, Richard R. *The Psychology of Learning Mathematics.* London: Penguin Books, 1971.

Whitehead, Alfred North. *An Introduction to Mathematics.* London: Oxford University Press. 1911.