CHAPTER 6

THE ROLE OF CONCEPTUAL ENTITIES AND THEIR SYMBOLS IN BUILDING ADVANCED MATHEMATICAL CONCEPTS

GUERSHON HAREL & JAMES KAPUT

Mathematical thinking is carried out using mental objects. For example, suppose one asks if a vector space V and its double dual V^{**} are isomorphic. At one level, one is asking about the "objects" V and V^{**} and, to begin describing an isomorphism, one may go on to describe a correspondence between respective vectors in the two spaces, which again, are treated mentally as objects, although they might be n-tuples or matrices, for example. Similarly, one may need to define a mapping between two function spaces, where the elements of the domain and range of the mapping must be treated cognitively as objects, as opposed to the mapping itself, which may be treated as a process, with inputs and outputs. In yet another instance, one may need to reinterpret a universal construction in the sense of MacLane (1971) as an adjoint functor pair, where the existence of a unique mapping with a certain property in fact defines a natural transformation between functors – so the mapping must play the role of an object on which the natural transformation acts. Such experiences are quite common in mathematics at all levels, but they feature widely throughout advanced mathematical thinking. The aim of this chapter is to begin to discuss them and their roles in helping us to build ever more complex mathematical concepts.

The idea of conceptual entities formation was suggested by Piaget (1977) in his distinction between form and content. Recently, several researchers have recognized its value in the learning of mathematics. It has been called *encapsulation* (Ayers, Davis, Dubinsky & Lewin, 1988), *reification* (Sfard, 1989), *integration operation* (Steffe & Cobb, 1988), for example, this process is an instance of reflective abstraction (Beth & Piaget, 1966), in which "a physical or mental action is reconstructed and reorganized on a higher plane of thought and so comes to be understood be the knower" (p. 247). Greeno (1983) defines a conceptual entity as a cognitive object for which the mental system has procedures that can take that object as an argument, as an input. He distinguishes cognitive objects from attributes, operations and relations, which attach to or act on objects. Further, he suggests that to qualify as objects, they must be permanently available in the individual's mental representation (p. 277).

The construction of function as a conceptual entity is an example of the entification process (Thompson, 1985a; Harel, 1985; Ayers *et al*, 1988). One level of understanding the concept of function is to think of a function as a *process* associating elements in a domain with elements in a range. This level of understanding may be sufficient to deal with certain situations, such as interpreting graphs of functions point-wise or solving for x in an equation of the form $f(x)=b$, but it would *not* be sufficient to deal meaningfully with situations which involve certain operators on functions, such as the integral and differential operators, as we will see later in this chapter. For the latter situations, the three components of function – the rule, the domain, and the range – must be encapsulated into a single conceptual entity so that these operators can be considered as procedures that take functions as arguments. Incidentally, a formal definition of a function as a single *set* of ordered pairs, a mathematical entity, does not appear to play a role in these situations – when would one conceive of a function as a set of ordered pairs in the context of applying a differential operator to that function? In this way the concept image evoked in a given context may be different from the formal definition, and may even at times be in conflict with that definition, as discussed in the previous chapter.

The construction of conceptual entities embodies the "vertical" growth of mathematical knowledge (in the sense of Kaput, 1987). For example, at lower levels, the act of counting leads to (whole) numbers as objects, taking part-of leads to fraction numbers, functions as rules for transforming objects become themselves objects that can then be further operated upon, for instance they may be differentiated or integrated. This complements the kind of "horizontal" growth associated with the translation of mathematical ideas across representation systems and between non-mathematical situations and their mathematical models.

In the next section of this chapter we lay out some of the circumstances under which conceptual entities are created and used and what their cognitive function might be, often by pointing to consequences in students' reasoning processes where they have not yet been mentally constructed. In the following part we will shift attention to the complex roles of notation systems in building and using conceptual entities. We regard this chapter as a foray into relatively unexplored territory, and do not make claims of completeness or of empirical substantiation for the framework being suggested.

1. THREE ROLES OF CONCEPTUAL ENTITIES

We will discuss the concepts of function, operator, vector space, and limit in terms of the role that conceptual entities have for:

1. *Alleviating working memory or processing load* when concepts involve multiple constituent elements.
2. Facilitating *comprehension* of complex concepts: the cases of "uniform" operators, "point-wise" operators, and "object-valued" operators.
3. Assisting with the *focus of attention* on appropriate structure in problem solving.

Greeno (1983) suggested a number of functions of representational knowledge involving conceptual entities: forming analogies between domains, reasoning with general methods, providing computational efficiency, and facilitating planning. He offered empirical findings that are consistent with his suggestions; these findings deal with elementary mathematics – geometry proofs and multi-digit subtraction – as well as physics, puzzle problems, and binomial probability. He also suggests that instructional activities with concrete manipulatives can lead to an acquisition of representational knowledge that includes conceptual entities. Other researchers suggest different types of instructional activities for the construction of conceptual entities. For example, Ayers *et al*, (1988) demonstrate how computer activities in learning mathematical induction and composition of functions can facilitate the construction of these concepts as entities (see the next chapter).

1.1 WORKING-MEMORY LOAD

One psychological justification for forming conceptual entities lies in their role in consolidating or chunking knowledge to compensate for the mind's limited processing capacity, especially with respect to working memory. To avoid loss of information during working memory processes, large units of information must be chunked into single units, or conceptual entities. Thus, thinking of a function as a process would require more working-memory space than if it is encoded as a single object. As a result, complex concepts that involve two or more functions would be more difficult to retrieve, process, or store if the concept of function is viewed as a process. This is true for many concepts in advanced mathematics. Imagine, for example, the working-memory strain in dealing with the concept of the double dual space of a space of *nxn* matrices if none or only a few of the concepts, matrix, vector space, functional, and field are conceived as consolidated entities.

1.2a COMPREHENSION: THE CASE OF "UNIFORM" AND "POINT-WISE" OPERATORS

Despite the heavy working-memory load involved in understanding the dual space of an *nxn* matrix space without most of its subconcepts being entities, it is still possible to make sense out of it, at least momentarily. In some situations, however, the justification for the formation of conceptual entities is more than just a matter of cognitive strain that results from a memory load. In such situations comprehension *requires* that certain concepts act mentally as objects due to an intrinsic characteristic of the construct involved. Examples of such situations include those which involve the integral or differential operators. These types of "uniform" operators *cannot* be understood unless the concept of function is conceived as a total entity. We distinguish these from other types of operators on functions which could be termed "point-wise" operators, and for which there is no need to conceive functions as objects, but only as processes acting on individual elements of their domains. For example, sum and composition can be treated as point-wise operators; this position is different from Ayers *et al*'s (1988) position who argue that composition of functions requires the encapsulation of function as an entity. Further research is needed to examine the two arguments. The cognitive process of understanding these operators involves the conception of a function as a process acting on individual elements of the domain. In constructing the composition of two functions f and g, say f∘g, one must first perform the process g on an arbitrary element x of the domain, generating a result g(x), and then perform the process of f on that result to obtain f(g(x)), all conceivable as acting on individual elements of the domain. These two separate operations are coordinated to produced a new process. Similarly, in constructing f+g, for every input x, the outputs, f(x) and g(x), are produced to construct the sum, f(x)+g(x). This sum can even be illustrated graphically by using a sample set of directed line segments for the distances between the horizontal axis and the graphs of f and g, respectively. Then the graph of f+g is the graph whose distance from the horizontal axis is given by the vector sum of the directed line segments. Clearly, the sum f+g can be illustrated point-wise.

The limit of a one variable function is another case which may be regarded as a point-wise operator. To understand this complex concept, many clusters of knowledge about different concepts in mathematics are required whose rich conceptual content is reflected in the complexity of its historical development. We will not attempt to analyze this knowledge here; however, the process-conception of function is sufficient (and necessary) to understand the limit concept. This is so because $\lim_{x \to a} f(x) = L$ may be viewed in terms of the *point-wise dependency* between the behavior of the numbers "near" a, x's, or inputs of f, and the behavior of their outputs, f(x)'s, "near" L.

By contrast, "uniform" operators arise when the point-by-point process is *inapplicable*. For example, to understand the meaning of:

$$I(t) = \int^{t} f(x)\, dx$$

as a function of t, it is necessary to think of I(t) as an operator that acts on the process $x \rightarrow f(x)$ *as a whole* to produce a new process:

$$t \rightarrow \int^{t} f(x)\, dx$$

It is the awareness of acting on a process as a whole, as a totality – not point-by-point – that constitutes the conception of that process as an object.

Mathematically unsophisticated students attempt to interpret "uniform" operators as "point-wise" operators apparently because they cannot conceive of a function as an object. Consider the derivative operator. Our experience in the classroom suggests that many students understand that f′(x) means: for the input x there is the output f(x), and for that output we get the derivative f′(x). Faced with the question,

find the derivative of the function $f(x) = \begin{cases} \sin x & \text{if } x \neq 0, \\ 1 & \text{if } x = 0, \end{cases}$

a common response is:

$$f'(x) = \begin{cases} \cos x & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

The student is no longer treating differentiation as a limit process, but as an algorithm to be applied to the formula at each point (or to the two separate formulas in the expression). To be able to handle this problem, the student needs to be able to consider the values of the function near x and renegotiate the limit process. In Greeno's terms, the function f must act as an argument for the (cognitive) differentiation operator, which it cannot do unless the function is conceived as a conceptual entity.

1.2b COMPREHENSION: THE CASE OF OBJECT-VALUED OPERATORS

As the notion of function develops, it can have different objects as inputs and outputs, in particular, it can output another function. For instance, the real-valued function $f(x,y)$ is usually thought of as a process mapping points on the plane, (x, y), into points on the real line, $f(x,y)$; thus, students who possess the process-conception of function would likely have no difficulty dealing with this interpretation. A more subtle interpretation can view $f(x,y)$ as a process which associates points on the real line, x, with functions, $f_x(y)$ where the latter assigns the value $f(x,y)$ to y. In this interpretation f is regarded as a function with input x and output the function f_x. We believe that, cognitively, thinking of a function as an output is *not* different from thinking of it as an input, in the sense that in both cases a function must be treated as a variable, as a conceptual entity. In this respect, this interpretation of $f(x,y)$, like the "uniform" operator, demands that the concept of function will be treated as an object. However, the cognitive demands of such a viewpoint are often great.

This analysis, which has yet to be empirically substantiated, is supported by our informal observations while teaching undergraduate mathematics classes the concepts of double limit, $\lim_{(x,y)\to(a,b)} f(x,y)$, and the iterated limit, $\lim_{x\to a}\lim_{y\to b} f(x,y)$. As some textbook authors have indicated (e.g., Munroe, 1965, p. 108), we observed that while computationally the iterated limit is easier than the double limit, conceptually the iterated limit involves a more sophisticated idea, which causes difficulty for students in particular circumstances. In stating and proving certain theorems on iterated limits (e.g., theorems concerning conditions on equality between this limit and the double limit), one needs to regard $\lim_{x\to a}\lim_{y\to b} f(x,y)$ as a composition of the following three mappings (see figure 12):

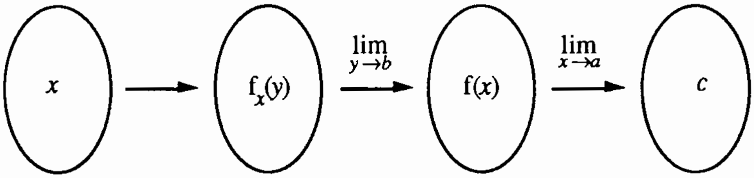

Figure 12 : $\lim_{x\to a}\lim_{y\to b} f(x,y)$ as a composition of three mappings

1. $M: x\to f_x(y)$, whose domain is a set of real numbers and whose range is a set of functions;

2. $\lim_{x\to a}: f_x(y) \to f(x)$, whose domain and range are sets of functions;

3. $\lim_{x\to a}: f(x)\to c$, whose domain is a space of functions and range is a set of numbers.

Students responses and questions indicate difficulty in dealing with aspects concerning the operator M, which, as indicated earlier, requires the object-conception of function. While the operator M must be understood as an object-valued operator, the other two operators, $\lim_{y\to b}$ and $\lim_{x\to a}$ can be viewed in two ways, which determine different levels of understanding the concept of iterated limit. In one way $\lim_{y\to b}$ and $\lim_{x\to a}$ are uniform operators acting on objects which happen to be functions. This level of understanding, although desirable, is not achieved by the average student, who usually views these limits, and the concept of limit in general, in a less sophisticated way as point-wise operators.

Besides the iterated limit, the undergraduate mathematics curriculum is replete with situations involving object-valued operators, for example those which concern parametric functions, such as $f(x)=ax+b$, $f(x)=\sin(ax)$, $f(x)=\log_a x$, etc., or parametric equations involving such functions. In these situations the correspondence between the parameters and the function, or the equation, constitutes an object-valued operator. The difficulties involved in understanding object-valued operators was investigated by Harel (1985) in the context of linear algebra (taught to advanced high-school students in Israel). It was found that students usually had difficulty dealing with such a correspondence, unless they were able to tag the outputs of the correspondence with familiar geometric figures, such as lines or planes (e.g., $t \to (a, b)+t(c, d)$ or $(t_1, t_2) \to (a,b) + t_1(c, d) + t_2(e,f)$). These geometric figures, which were manipulable objects for the students, apparently helped the students to construct such a correspondence as an object-valued operator.

Another common example involves the construction in abstract algebra of the quotient object associated with a "normal" sub-object, e.g., in the case of groups. The cosets must be conceived as objects if they are to participate as elements of a group. However, the existence of a "representative element" for a coset, where the operation defined on cosets can be given in terms of an operation on their representatives, makes it possible to deal successfully with many aspects of the quotient group on a symbol manipulation level without treating the subsets of the group as objects, or even as subsets. Students' inadequate conceptions are revealed when one asks them to attempt to create a group using a non-normal subgroup's cosets – they often cannot understand why the subsets "fall apart" when they attempt to multiply them together as sets, or by using representatives.

Finally, data reported by Kaput (in press) can further support the cognitive distinctions among the different types of operators made above. Secondary level students were asked to determine an algebraic rule that fits a student-controllable set of numerical domain-data (they pick the x's and the computer provides the $f(x)$'s). Examination of their behavior revealed a clear and stable decomposition of the group of students (in a sample of over 40 high school students) into two sets, one of whom consistently used a point-by-point pattern-matching process, mediated by natural language formulations of their proposed "rules," while the other searched for and applied a parametrically mediated formulation of their proposed rules. The latter, for example, would look for constant change in the dependent variable, identify this as the "m" in $y=mx+b$, and proceed from there. For them the process was a search for parameters that indexed functions as objects. In effect, they were dealing with a space of functions (albeit a limited one), whereas the other group of students conceptualized the task as a point-wise attempt to build a function whose point-wise behavior matched the rule that they had formulated using natural language.

1.3 CONCEPTUAL ENTITIES AS AIDS TO FOCUS

The third role of conceptual entities we have identified involves facilitating focus on those aspects of a problem representation that are most relevant to the solution of a problem. In a one-on-one interview with an experimental group of Israeli high-school students regarding the concept of vector space (after several instructional sessions in which this concept was gradually abstracted from two and three dimensional representations; see Harel, 1989a, 1989b), the first author asked the following question:

> Let V be a subspace of a vector space U, and let β be a vector in U but not in V. Is the set $V+\beta=\{v+\beta \mid v \text{ is a vector in } V\}$ a vector space?

There were clearly two groups of students: those who answered this question by checking the whole list of the vector-space axioms, and those whose answer was something like, "you moved the whole thing, it doesn't have the zero vector any more", or "the new thing, $V+\beta$, is not closed under addition". Clearly, the latter group of students viewed V as a total entity, a "thing," and thus they were able to view $+\beta$ as a shift operator which takes V as an argument, an input. This enabled them to focus on those vector space properties that are most relevant to the solution of the given problem, namely, the zero property or one of the closure properties. The other group of students, on the other hand, relied on the formal definition of vector space by checking whether the individual axioms apply. That $V+\beta$ is a subset of the vector space U, which guarantees the existence of most of the axioms, was not visible to these students. Moreover, many of these students failed to check some of the axioms, including those essential to the solution of the problem (e.g., the existence of zero).

2. ROLES OF MATHEMATICAL NOTATIONS

The power of mathematics associated with the roles of conceptual entities is closely related to the roles of mathematical symbolism. Using mathematical notations, complex ideas or mental processes can be chunked and thus represented by physical notations which, in turn, can be reflected on or manipulated to generate new ideas. In this section we will discuss three aspects of the interaction between formation of conceptual entities and mathematical notation:

1. The role of mathematical notation in forming conceptual entities.
2. Different types of mathematical notations, *elaborated* and *tacit* notations, and the manner in which they represent conceptual structure.
3. Notations as substitutes for concepts.

2.1 NOTATION AND FORMATION OF COGNITIVE ENTITIES

Greeno (1983) stated two conditions that help distinguish entities from other mental events. One is its continual presence in a mental representation; the other (mentioned earlier) is its ability to act as an argument in another mental procedure or argument. By providing continual *perceptual* experience, material notations help provide the basis for continuing *conceptual* presence. This role is based simply on notations as names – the notation serves to *name* an item in our conceptual world. We might term this the "nominal" role. Note that the parts of the syntax of a notation system associated with identification and discrimination of notational objects plays an important role here. Having an explicit name for a mental event helps objectify it through a kind of transference of object permanence – from the permanence of the physical notational name (which produces perceptual experience on a more or less continuous basis) to a cognitive permanence. Of course, the perceptual item must somehow come to be integrated with the conceptual one. Otherwise, all one might end up with is, say, an easily reproducible mental experience of a mark or character string, with no other mental activity or structure beyond that primitive experience – which is the experience of altogether too many students.

The nominal role of symbols is frequently played out using conventions that help distinguish the status or differing roles of objects in complex situations – convention-based variations in the names of objects help distinguish the classes to which they belong. Suppose a concept involves a process which takes entities of a different order as inputs and outputs, e.g. differentiation operating on functions. Then there is a need to distinguish between the higher level process and its lower level inputs and outputs, a need which is typically satisfied by using systematically different symbols for the items at each level. Then the conceptual activity of keeping the things distinguished is off-loaded onto the notation system. For example, many higher level mathematical activities involve defining functions between sets of functions – as between a vector space and its double dual. Another typical example occurs in topology, when one defines various compactifications, e.g., the Stone-Czech compactification of a regular Hausdorff space based on sets of continuous functions on the unit interval. In all such cases, one finds that, typically, different classes of characters are used to distinguish the different levels of functions - say, one Greek and the other contemporary English-based.

Systematic variation in names also is employed through the use of different classes of symbols to distinguish when an object is being treated in two different ways, where it has essentially two different identities. Consider the conventions used to distinguish the identity of a real number x from its identity as a member of the field of complex numbers, where it may be denoted by $x+0i$. Similar distinctions are made whenever a canonical embedding is being employed, not merely in the case of algebraic closures, because it is a characteristic of "canonicalness" that the substructure is maintained within the larger structure. A related case involves the distinction between a constant function and its value. In all these cases, object identity is identified and maintained notationally.

Relative to Greeno's second condition for cognitive entities acting as arguments in other procedures, the syntax of a notation system specifically structures the place of the material notational objects in a coherently organized physical system. Such a system is designed to support a given type of thinking. For example, the character string notation for functions supports highly sophisticated manipulations, which in turn, are used to facilitate a wide variety of mental operations on the conceptual objects that those character strings denote. Thus, the act of factoring the character-string representation of a polynomial function to help identify its roots may be based on some syntactic rule (e.g., applied to the difference of cubics), which obviates the need to justify all the steps of the process. The strength of a notation system may be measured by whether, and to what degree of fidelity, syntactically guided actions on its objects reflect and/or subsume important mental operations.

We conclude this section with two specific examples to illustrate the variety of ways notations either help encapsulate mathematical concepts as entities or supplant conceptual entities in reasoning processes. Goldin (1982) discusses the impacts of languages or notations on the different stages of the problem solving process, citing his own data as well as the well-known problem-isomorph work by Simon and colleagues. The following discussion can be thought of as somewhat preliminary to the issues discussed by these researchers in the sense that we are dealing with the concept-notation relationship at a more primitive level.

> *Example 1*: Consider the use of graphical notation, the slope of straight lines, to facilitate the order comparison between ratios described as linear functions between sets of objects, measures, or even numbers. To compare two such on the basis of a table of data (a sequence of ordered pairs) or even on the basis of a pair of fractions is not as easy as comparing the slopes of their associated straight lines in a coordinate plane. In this case one need only attend to two things (2 lines) as distinguished by their most salient attribute, their slope. Each single line embodies an infinite set of equivalent pairs of ratio values. This seems to be an instance of a one-for-many substitution of a single notational object for a set of mental objects, although from another perspective it amounts to an integration of detailed features into a single object.
>
> *Example 2*: Recall the study mentioned in §1.2 where students were determining functions from numerical data. There were two types of students: One type of students were essentially "pre-algebraic" in their thinking, and treated every potential rule that they inferred from the numerical data in a table (which they generated) as a natural language-based rule. That is, they thought of $2x+1$ as doubling and adding one, in terms of a natural language interpretation, rather than in terms of parameters m and b in $mx+b$. Thus they did not see growth in the numerical data in the same way as those who were looking for values of these parameters. Basically, the latter were looking for growth rates, which they interpreted as the first parameter's value, etc. For them, a linear function was experienced as a "thing", a conceptual entity, whose identity is determined by the two parameters. The other students were looking for a way to translate from their natural language-based encoding of an unencapsulated *process* to algebra. They quite often succeeded – as long as the parameters involved were positive whole numbers. For negatives, they fell apart, because they were not able to get easy natural language encodings of what for them was a process rather than a thing (Kaput, in press). An open question is what is the relation between the conceptual entity and the parameter notation? Which came first? Or did they co-evolve? In any case, this example seems to offer an instance of the functional power of the nominal use of symbols – as do most systematic uses of parameters.

2.2 REFLECTING STRUCTURE IN ELABORATED NOTATIONS

The inventors of mathematical notations created them to express the contents of their own minds, both to themselves, to aid their own thinking, and to others, to aid in the communication of their conceptions. As Leibniz, that great master of notation-invention put it,

> In signs one observes an advantage in discovery which is greatest when they express the exact nature of a thing briefly, and, as it were, picture it; then indeed the labor of thought is wonderfully diminished. (Quoted in Cajori, 1929, p.184)

Extending his remark, we might add that the structure of the conceptions is, in some way, being reflected in the structure of the notations, especially in their syntax. Or, put more constructively, the experience of perceiving the notations shares important features with the experience of the conception apart from any perceptual act. Extending this observation further, we suggest that it is even more important that actions on notational objects in some regular way reflect mental actions on the conceptions. (We again hasten to add, however, that we are not suggesting any kind of simple relationship between notation and conception!)

But mathematical symbols differ in the extent to which they include features that reflect the structure of the mathematical objects, relations or operations that they stand for. Some are more elaborated than others (Harel, 1987). For example, the place-valued symbol 324 expresses a specific structure of the quantity it represents: three hundreds, two tens, and four ones. Of course, this number written in expanded notation is even more elaborated. Similarly, the more abstract symbols, (x, y) for an ordered pair of numbers, $f(x) = 3x^2$ for a specific real-valued function, AB for a line segment whose endpoints are A and B, and

$$\begin{pmatrix} a_{11} & a_{12} \cdots & a_{1n} \\ a_{21} & a_{22} \cdots & a_{2n} \\ & \vdots & \\ a_{m1} & a_{m2} \cdots & a_{mn} \end{pmatrix}$$

for an *mxn* matrix, are all relatively *elaborated* symbols, because they encode the structures or relationships among components of their referents.

On the other hand, for example, the concept "matrix of the linear transformation T relative to the pair of ordered bases μ and ν" can be symbolized by the significantly less elaborated symbol $[T]_{\mu,\nu}$. A more elaborated symbol for this concept could be $[T]_{\mu\rightarrow\nu}$, which indicates that the matrix representation of a transformation T depends on the relationship between the bases in its range and in its domain. An even further elaborated symbol for this concept is:

$[[T(\mu_1)]_V: [T(\mu_2)]_V: \ldots : [T(\mu_n)]_V]$ (used by Anton, 1981),

which encodes many of the variables included and its referent. In contrast, the symbol $[t_{ij}]$ (used by Nering, 1970, to represent the same concept) is far less elaborated, whilst the bare symbol T is a non-elaborated, or *tacit*, symbol. Tacit symbols provide essentially an indexical function – they *name* things, without denoting aspects of the structure of what is named.

One category of tacit symbols consists of those which, during a discussion or proof, are used to represent variables. For example, the statement, "let β be an ordered pair ...", typifies a context in which such a tacit symbol is used; here, the symbol β does not encode the structure of its referent – an ordered sequence of two objects – but it, together with the surrounding phrases, *does* name the set over which the variable varies.

The extent to which a notation is elaborated is determined by the extent to which it ties to prior mathematical knowledge, which is very much a cognitive matter. Indeed, what is elaborated for one person may appear very bare and tacit for another. Nonetheless, the act of connecting a bare notation to an elaborated one is a translation act, which, depending on circumstances, may operate in either direction. The notation's perceived connection with prior knowledge takes the form of perceived features that reflect features of the prior knowledge. For example, two different symbols are usually used to represent the composition of two functions f and g: $f(g(x))$ and $(f\circ g)(x)$. The symbol $f(g(x))$ expresses the process in which the two functions are composed: the input x in the function-machine g produces the output $g(x)$, where $g(x)$ now acts as an input in the function-machine f to produce the output $f(g(x))$. (Note the strong use of temporality here.) Thus the symbol $f(g(x))$ is amenable to the thinking of a function as a process, but depends on the prior knowledge of input-output relations expressed using the standard $f(x)$ notation. The symbol $(f\circ g)(x)$, on the other hand, describes an operation between two functions – f and g – which produces a third one – $(f\circ g)(x)$. This symbol describes f and g as inputs in the [meta] function-machine $\circ$, and thus to understand its meaning functions must be viewed as conceptual entities. In this example, the prior knowledge is that of operating on inputs to functions, and the notation feature is reflected in a parallelism of structure, except that the first function in the composition acts as the input.

The pedagogical importance of this example is that some mathematical symbols cannot be understood via the symbol $f(g(x))$; for example, the "uniform" operator:

$$I(x) = \int^{x} f(t)\,dt.$$

Students have trouble thinking of the integral as a function of x – which is revealed when they are asked to treat it like a function. Our notation $I(x)$ for it is itself intended to help with this – it assists entification by treating it notationally as a function, elaborating it in such a way that the functional dependence on the variable x is highlighted.

The distinction between elaborated symbols and tacit symbols has important consequences for learnability and usability. In Harel (1987) it was hypothesized that an elaborated symbol would be better understood and remembered if it expresses the main and salient variables in its referent. Here, we additionally hypothesize that a tacit symbol can be more meaningfully used when its referent is encapsulated into a conceptual entity. That is, in developing a symbol for a concept one must try to match the degree of elaboration of the symbol with the degree of elaboration of the user's concept, which in turn must match the user's needs for the task at hand. After all, in some cases it is important to suppress detail, and in others the detailed structure plays a role in what one is trying to do. It seems, then, that one's control of the amount of structure explicitly represented in the symbolism is a major factor in mathematical thinking, because one can adjust the "focus of one's mental microscope" by adjusting the notation. This we believe to be an important facilitating factor that notations offer us.

3. SUMMARY

We hope to have introduced some useful ways of thinking about some important aspects of the learning of mathematics that highlight the role of conceptual entities and their relationships with mathematical notations. We regard this chapter as but a beginning into an area of research that others may find productive to pursue in the future.

In §1 we laid out some of the circumstances under which conceptual entities are created and used and what their cognitive function might be, often by pointing to consequences in students' reasoning processes where they have not yet been mentally constructed. We observed three cognitive functions:

- Alleviating working memory or processing load when concepts involve multiple constituent elements, facilitating comprehension of complex concepts,
- the cases of "uniform" operators, "point-wise" operators, and "object-valued operators",
- assisting with the focus of attention on appropriate structure in problem solving.

These functions, undoubtedly, play an important role in mathematical thinking and in fostering the vertical growth of mathematical ideas, at all levels.

In §2 we analyzed the key role that notations play in the entification process by helping substitute names for complex conceptual structures and/or operations. We have discussed three aspects of the interaction between formation of conceptual entities and mathematical notation:

- the role of mathematical notation in forming conceptual entities,
- different types of mathematical notations – *elaborated* and *tacit* notations, and the manner in which they represent conceptual structure,
- notations as substitutes for concepts.

Just as notations can help the formation and application of mental entities, notations can act as *substitutes* for conceptual entities, supplanting the need for them. It is here where both the great power and the great danger in using mathematical notation systems become particularly and unavoidably evident. Accompanying the great power of notations as aids to mathematical thought based on their identity-management role and their structure-substitution role is the great danger that the notations do not refer to any mental content beyond the experienced physical structure of the notations themselves, e.g., as when one deals with an algebraic statement as a character string. This seems to be the case with altogether too many students. While the inventors of notations created them to express and perhaps elaborate their own pre-existing conceptions, in schools we often begin in reverse order, concentrating on manipulation of notations, e.g., the techniques of differentiation and integration in calculus, before providing sufficient experience that would enable the building of mental referents for those notations (Davis, 1986). Students should be given opportunities to build their own notational expressions of their ideas, which can then be guided in the direction of the standard ones. In this way, one builds both notations and conceptions simultaneously, rather than building one or the other first and then attempting to connect the two.