COGNITIVE CONSTRAINTS ON COMMUNICATION

Representations and Processes

Edited by

LUCIA VAINA

Center for Cognitive Science, M.I.T., and Computer Science Department, Boston University

and

JAAKKO HINTIKKA

Department of Philosophy, Florida State University

D. REIDEL PUBLISHING COMPANY

A MEMBER OF THE KLUWER ACADEMIC PUBLISHERS GROUP

DORDRECHT / BOSTON / LANCASTER

LUCIA M. VAINA

TOWARDS A COMPUTATIONAL THEORY OF SEMANTIC MEMORY

INTRODUCTION

Memory is one of the most important functions of the human brain, yet its understanding – why and how it does what it does – has so far eluded us. Research in memory has been a frustrating task not least because of the intimate familiarity with what we are trying to understand, and partly also because the human cognitive system has developed as an interactive whole; it is difficult to isolate its component modules – a necessary prerequisite for their thorough elucidation.

Before one embarks on a new research area its general demands and underlying theoretical hypotheses should be stated.

FUNDAMENTAL DEMANDS OF RESEARCH IN MEMORY

The first demand to the student of memory, and of cognitive abilities in general, is to attempt to separate those topics that appear capable of explanation by available approaches from those for which no ready explanation even in outline seems available.

The second demand concerns the properties and attributes of the human brain that the researcher must aim to account for. Some of these are:

(1) the redundancy and self-restorative nature of the brain mechanisms connected with learning and memory;

(2) the capacity to learn;

(3) the modification and refinement of the information already stored;

(4) the local character of computations (a local change shouldn't require large modifications);

(5) the existence of sensory-specific modules that independently convey information to the language module.

The third demand concerns to the evaluation of the research done and the choice of those methods that seem to lead to the most relevant results. It has become clear that the most reliable approach to the study of the brain

L. Vaina and J. Hintikka (eds.), Cognitive Constraints on Communication, 97–113.

activity is to regard it as a large and complex information processing system. Central to this approach is the belief that human cognitive capacity can fruitfully be viewed as some kind of symbolic system. Thus, much is to be learned by developing computational theories for aspects of human information processing and comparing the results of these theories and their implementations with human performance on the same tasks. Behavioral phenomena may suggest or constrain possible information processing tasks whose properties can be studied computationally, and thus might lead to the search for previously unrecognized behavioral consequences that they imply.

In an information processing approach, such as the one to which I subscribe, there is a distinction (pointed out by Marr and Poggio [6]) between the various levels at which our information processing device may be understood: at one level, there is the theory of computation, which is what is computed and why, and at the next level is the particular algorithm, or way in which the computation is carried out. My present goal is to elaborate the computational theories of some of cognitive abilities of the human brain. The particular implementation, although eventually important, plays only a secondary role at the moment.

THEORETICAL HYPOTHESIS OF RESEARCH

(1) To understand various disabilities resulting from lesions to the brain helps us to understand the normal function of the brain.

(2) Data are important for the process of developing the theory; ideas and hypotheses that are at the variance with the data have to be rejected.

(3) We should bear in mind that the facts that we deal with are soft and the working domain ill understood, so our intellectual resources would be misplaced if at the present they are spent on the construction of elaborate mathematical structures.

LEVELS OF RESEARCH INTO MEMORY

Memory may be studied at several levels. At the most physiological end one would study the neural basis of memory, how the hardware implements the storage process. Examples of theories at this level are the cerebellum [4], [8], the mathematical theory of associative memory devices [7]. Although such research can provide us with illuminating insights into the functioning of the brain, and provides an essential component of an eventual understanding of memory, it is clearly not the whole story. Although studies at this level address the details of how to implement a certain kind of memory in a particular hardware, they unfortunately shed no light on what information should be stored in the memory or how it should be represented there. The underlying reason is that studies at this level contain no analysis of the uses of memory in the broader context of day to day information processing tasks.

Memory cannot be studied in isolation, since it is essentially only an adjunct to the proper execution of our ordinary information processing tasks. In order to try to formulate specifically some of the basic requirements of memory we must therefore examine the structure of the processing tasks for which it is used. A first division in central processing, although a rough one, would be between *modality specific* and *modality unspecific* analysis. Examples of modality specific analysis include the tasks of visual analysis, tactile analysis, auditory analysis, etc. It is clear that these different types of analysis must be taken at least some way before cross-modal interactions of any complexity could be useful, and in fact clinical evidence from neurology suggests that these analyses can proceed a substantial way before their combination. Thus in vision, for example, a sophisticated representation of the shape and disposition of a viewed object can be derived by patients whose realization of the shape's use or purpose is severly impaired.

Each of these modality-specific analyses poses its own self-contained memory problem, whose primary purpose will be to aid the recovery of a structural description (in the case of vision, of the shape of the viewed object from images of it). One might call such memories *modality specific recognition memories* (MSRM). Thus in vision, for example, the MSRM which Marr and Nishihara [5] used to organize and store their 3-D models, (see Figure 1), is deployed during the construction of a specific, arbitrarily detailed, object-centered description of the shape of the viewed object. According to current thinking, visual processes preceding this step do not usually involve the deployment of a learned catalog of shapes; they consist almost exclusively of memory-free perceptual processes, like stereopsis and structure from motion, and can usefully be thought as pure perceptual processes.

If this were generally true, one could view the different recognition modules as roughly consisting of two parts: the first, which one might perhaps call pure perception, consists of essentially memory-free analysis of

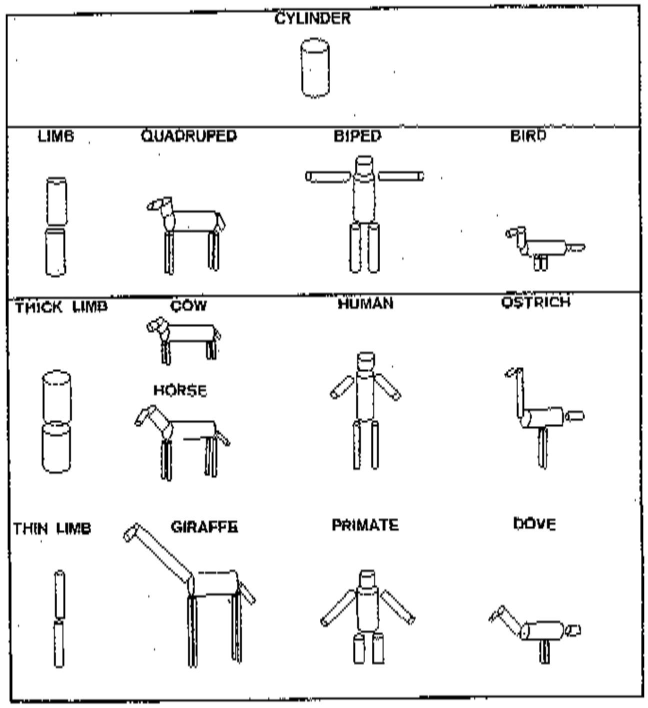


Fig. 1.

the incoming sensory information; and the second involves the use of memory of stored descriptions during the construction of a representation of the incoming information. Thus in vision, for example, a patient lacking his visual recognition memory but retaining his perceptual apparatus should still be able to perform simple visual tasks, like discriminate two lines at different orientations, or two points at different depths, even though unable to describe the shape of the viewed object. The descriptions supplied by the different MSRM, are potentially complex, since they are capable of representing exhaustively all the information that can be acquired via that particular sense. For example the description from vision of even a fairly simple shape can include 3-D models for aspects of the shape at several different scales, as illustrated in Figure 2.

Yet rich as these individual modality specific descriptions can be, we know

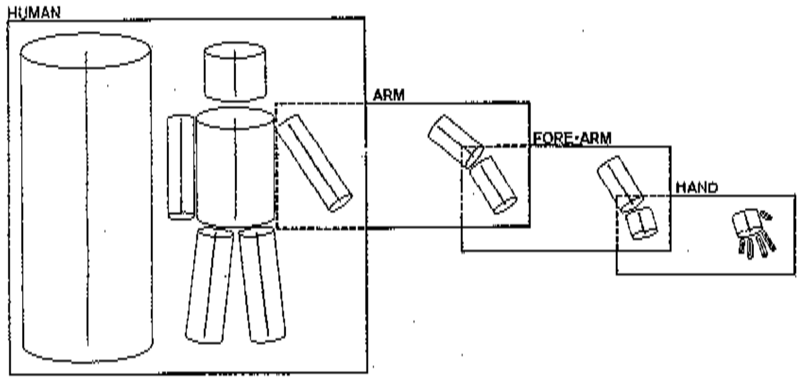


Fig. 2.

from our own experience that the comprehension of what we see, touch, hear, involves more than each one, and more even than their combination. For example, our comprehension of an object includes a knowledge of its use and purpose, to which there are often several aspects, and its name. The organization and representation of this information involves a different category of memories that are not modality specific. To distinguish them from the MSRM, let us call them Cognitive Memories. These memories are our research interest.

How is one to discover the computations performed by the cognitive memories? One possibility would be to consider what tasks are accomplished by the cognitive memories in order to investigate the computations they perform. In this case we have to define the goals of cognitive memories, and once we know them we can define the problems involved in attaining these goals. Naturally, it is very important during the formulation of the goals to rely on the right kind of data. Once the computational problems have been formulated, and we have a computational theory it is useful to develop a program. A strong theoretical motivation for having an implementation of a theory is that it helps one to appreciate problems that otherwise risk being passed unnoticed.

The memory whose task it is to process, store and retrieve upon request information about the meaning of words, concepts, facts, etc., was called Semantic Memory by Tulving [9], and my present concern is exactly this. I example, demonstrated a relatively intact understanding of lexical structures, and have a pretty normal ability to integrate perceptual and functional information. They also seem to be quite good at dealing with the fuzziness of conceptual boundaries. Anomics on the other hand, show evidence of lexical distortion, and manifest an inability to integrate perceptual with functional information. In addition to this, they seem to be insensitive to category boundaries.

THE RELEVANCE OF APHASIA TO THREAD MEMORY

The phenomena described in the last section together with my own experience with aphasic patients, have led to the following preliminary formulations of some principles of semantic memory.

(1) The existence of the different agnosias suggests the presence of several structurally distinct modality-specific representation systems. That is, a particular object is represented internally in several different memories.

(2) From the work done on visual-recognition memories, it appears that the description of an object is structured rather than unitary, and the units of description are organized hierarchically from the general to the particular.

(3) There are several ways of representing knowledge about an object: by its modality-specific descriptions, its functions or uses, and the categories to which it belongs. Each of these modes can be independently impaired. This suggests that semantic memory contains at least three modules, one for each of these modes of representation.

(4) Perceptual classifications and semantic classifications are separate, hierarchically organized systems and can be differentially impaired.

(5) Most concepts have both superordinate and subordinate categories. But at their most detailed level the representations of two different objects will be different.

(6) There is rarely an all-or-none response to an object; on the contrary, some semantic meaning is often preserved, and this is invariably of a general rather than of a specific nature. For example, an agnosic patient in Warrington's experiments would respond faster to "Is a duck an animal?" than to "Is a duck a bird?".

(7) From the comparative study of different types of patients (for example anterior vs posterior lesions) we see that the capacity for semantic categorization may be good, the recognition of functionality of objects may be good, and their perceptual description may also be good, yet some types

of objects may be easier to name than others. Some aphasics do much better with concrete than with abstract objects. This might be due to impaired access mechanisms rather than to the degradation of stored information. Thus we can begin to differentiate between processes that access information and processes that store and organize it, and there is some evidence that they can be independently impaired.

(8) We have seen that semantic memory as opposed to episodic memory relies on inferences. So, we can further differentiate the processes that store and organize the information, into processes of inference and processes of storage.

(9) From a comparison of posterior with anterior aphasics, it seems that the anterior lesions (Broca's) produce an impairment of the access processes (probably damaging control processes that deal with contextual differences) whereas posterior lesions may impair mechanisms that store and organize the information. Anomic patients can apparently retrieve a superordinate category of the object they are asked to name together with a correct description of its use. They are, however, apparently often unable to access a more particular representation. (e.g. for a rose, they may get "flower" but not "rose"). It seems then that the damage here is to the inference processes and not necessarily to the storage processes. Wernicke's aphasics, on the other hand, seem to have a general impairment of the memory itself.

(10) Evidence for the difference between impaired access mechanisms and damage to the memory itself has recently come, for example, from patients at the VAH-Boston. One of them initially had a severe inability to manipulate symbolic expressions, and showed no use of semantic memory. On re-learning the ability to manipulate symbols, some of his use of semantic memory returned. This suggests that his primary damage lay in his access mechanisms.

(11) Frequency and familiarity plays an important role in the case of impairment of semantic memory. Frequent or familiar terms are retrieved correctly more often than less frequent and unfamiliar terms.

THREAD MEMORY

The requirements formulated in the previous paragraphs have led us to formulate a new type of semantic memory, called "thread memory". A preliminary version of thread memory was introduced by Vaina & Greenblatt ([10] 1979), and the ideas behind it have been evolving over the past year.

I will now describe briefly the structure of the memory and some of the processes associated with it.

Structure of the Thread Memory

The basic element in the representation of objects in memory is called a thread, and it consists of a set of symbols (nodes), ordered according to precise rules. The relation between symbols might have different meanings, each of these meanings being associated with a module of representation. Three modules may be distinguished:

(1) a category module, which is a hierarchical organization of the symbols in memory, from the more general category to the more particular category. For example, mallard → thing → living-thing → animal → bird → duck → species-of-duck → mallard.

(2) a functional module, which contains information about the function, or the uses of the objects. In this module, actions and objects are represented in relation. Thus for example (S LEADS-TO) SEE → BOY → RUN.

(3) a descriptional module, which contains information about the component parts of an object, about its appearance. Each element of this module results in the creation of a thread. This module establishes relations between various threads that can be associated with a symbol. For example (BIRD FEATHERS) relates the thread keyed on FEATHERS to the symbol BIRD.

The first element in the thread, called the key, is the stimulus by which the thread is accessed. The threads end in the same symbol as the key. The difference is that when we reach the access symbol at the end of the thread it is loaded by then with all the meaning that is represented on the thread. The symbols in the thread are not unique to the thread, they might appear in several threads. This is a very useful property because it allows a leveled partitioning of information in the memory. So for example, most of the objects from the world could have as their most general representation a symbol like "thing". But that it would not tell us anything, we need to be more specific. The extent to which the memory gets specific can vary, but a semantic memory has to be able in principle to give a unique description to every object in the world.

Objects are represented by a set of threads that give a multiple description of the object and its functionality. The set of all threads associated with an object, or in other words, having the same key, constitute a *general thread* (Figure 3). There are many symbols in common among the threads in a general thread. Some of them are explicit and, some of them as we will see in

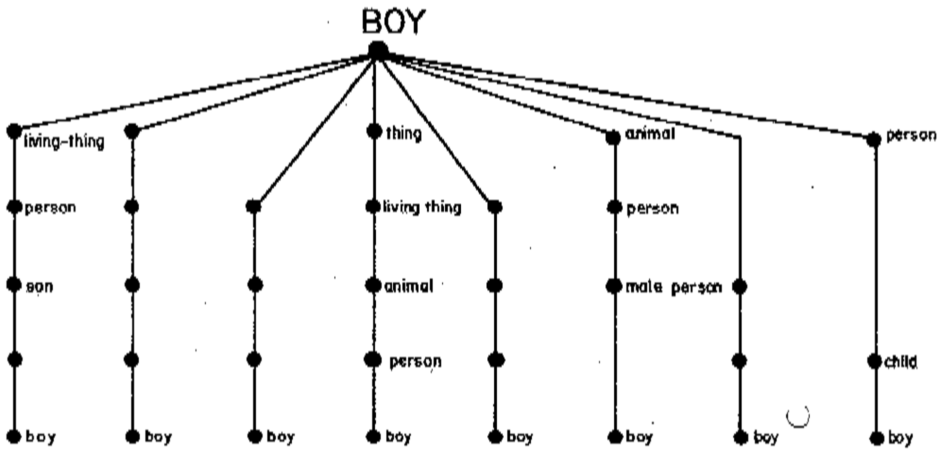


Fig. 3. Example of a general-thread.

the next section can be deduced by the application of the simple thread operation of deduction. A complete semantic representation of an object is given in all the modules of the semantic memory.

We call the more specific common node of two (or more) threads in the same module a *fork-point.* The fork-point of all threads in a general thread constitutes the most common knowledge that one has about an object and which is not dependent on its particular occurrence. In each general thread there are different fork-points, depending on the threads that one is looking at. At the fork point one has to decide which representations or descriptions might contain the needed specific information and choose that thread. The fork point of threads that belong to different general threads reveals what the compared representation have in common in a particular module of representation. For example, what is common between a chair and a table in the category module? The usual response is that they are both pieces of furniture. Fork-points obtained in the descriptional module might be misleading for interpretation in another module, like the categorial module for example. Thus a WHALE in the descriptional module is represented as LIVES-IN-WATER and its shape description is like the shape of a fish. From this, one might believe that a whale is a kind of fish. To avoid confusions, simple threads are formed to tell us that a whale is not a fish for example. Or specifying that IS-NOT-A PLANT ANIMAL enables us to distinguish all plants from animals by means of a single thread operation.

Sometimes we need to particularize an object. Thus, dog is a generic term that refers to all dogs in general, Spot a particular one. Animal is a generic term too, and Spot is an example of an animal as well as an example of a dog.

The functional module ought, for example, to contain things like a knife is an example of "a sharp thing used to cut". Or, if we cannot find a table, we could use whatever object that "has a flat surface, big enough to eat on", so a rock could serve this purpose.

Computations within Thread Memory

Three main types of computations can be distinguished in thread memory: access processes, inference processes and storage processes. The access processes refer to the interface of the semantic memory with other memories. The inference and storage processes are computations of the semantic memory per se. In talking about each of these processes we see that we can differentiate them further. Thus the access processes are of two types: the access processes themselves and processes that control them. By the general access processes a general thread is brought in the temporary buffer, and then the simple threads are examined one by one (by an operation called "advancing the thread" (Figure 4)). If one had to examine exhaustively every single thread of a general thread, the performance would be very slow and uneconomical. The access control processes and some inference process optimize this. The access control operations can "turn-on" or "turn-off" a group of facts. The context which is "on" at the time of accessing a thread is always specified; thus in advancing the thread we know that a specific-context, with its restrictions and requirement is needed. The specific context affects only the extraction or the choice of threads. Once a thread was chosen and activated in the temporary buffer, the specific context doesn't interfere anymore. We could suppose probably that the specific context is inherited or moved over from the episodic memory.

Another access control mechanism is the general context which allows the threads to be related. The information handled by the general context mechanisms is the information represented in the semantic memory. We can say then that the control access mechanisms control the interface between the information in episodic memory, the information in the semantic memory, and the temporarily active buffer.

INFERENCE PROCESSES

In the section about the structure of thread memory the fork-points were discussed as being special nodes on the threads. Finding the fork-point is a

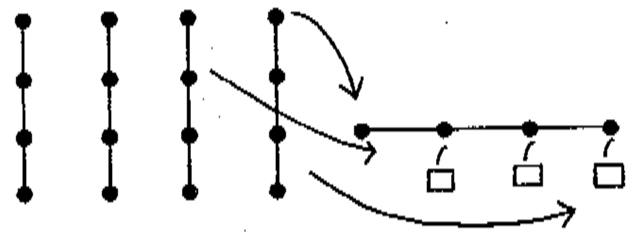


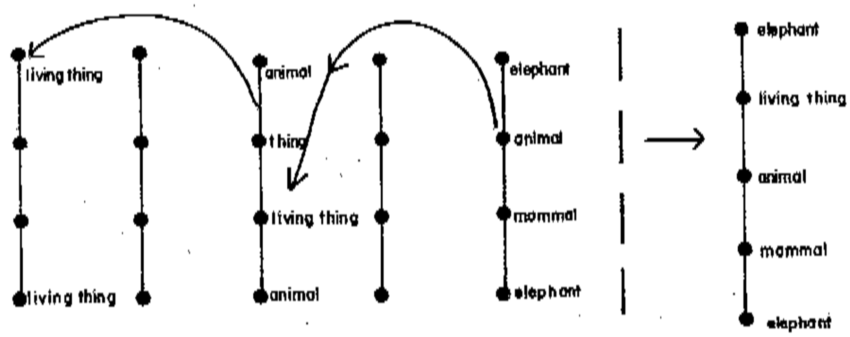
Fig. 4. Access processes: advancing the thread.

special inference process that deals with compare and contrast questions. Another inference process is the process of deduction. For example from the two following threads

ELEPHANT → ANIMAL → ELEPHANT
ANIMAL → LIVING-THING → ANIMAL

we can deduce the answer to this question: Is an elephant a living thing? The necessary one step deduction is performed by searching each thread whose semantic node is on the elephant thread, for the node living-thing. After the node ANIMAL is found, a storage operation is performed (the process of assimilation) to store the deduction. An interesting feature of deduction in thread memory is that the depth of deduction is not the distance between two nodes, as in most proposed semantic memories, but the number of jumps from one thread to another (Figure 5a). Thus a very modest search depth can find solutions to non-trivial questions.

There are several ways of optimizing the inference processes. For example,



Through one step of deduction we obtain the answer, and at the same time the general thread of living-thing (the facts about living-thing).

Fig. 5. Deduction: Is an elephant a living thing?

we can define a measure of closeness between symbols in each module of representation called neighborliness. I think this may be important for problem solving or for improving the access process that advances the thread. If the result found is acceptable then the neighborhood is reinforced. The neighborliness is not a simple information processing task; it relies on many other operations. For example, crucial in the measurement of neighborliness is the number of threads in a general thread whose fork point is quite deep (we shall see in the storage processes that these threads are "bundled" together: Figure 6f). In the functional module the process of

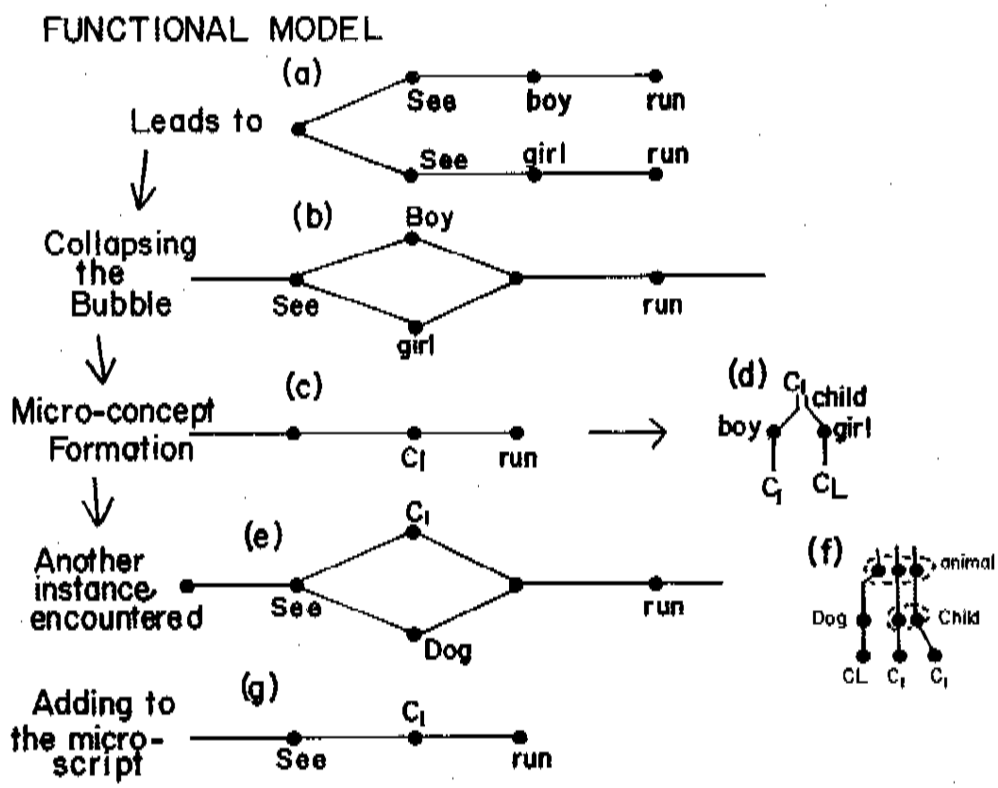


Fig. 6.

neighborliness is used to make generalizations; from particular examples inferences are made to more general categories. This is clearly an optimization process because it actually reduces the enormous number of particular examples of objects that are in the same relation with verbs, to their common fork-point in the category module. For example from

(S LEADS-TO) JOHN CRY
(S LEADS-TO) JANE CRY
(S LEADS-TO) DOG CRY
(S LEADS-TO) BABY CRY
(S LEADS-TO) BOY CRY
(S LEADS-TO) MOTHER CRY

By applying the neighborness process we obtain that (S LEADS-TO) ANIMAL CRY, and by that we would know, for example, that we can say ELEPHANT CRY as well as RABBIT CRY.

An interesting observation can be made, namely that the processes of inference are not simple processes, they presuppose processes of storage. (These remarks relate to a hypothesis made in the study of aphasic patients, that in the case of Wernicke aphasics probably all inference processes as well as the storage processes are impaired and yet in the case of anomic aphasics the impairment is of the inference processes. It is supposed in the literature that Wernicke's aphasics who improve usually recover to become anomic aphasics. This seems very much to support our hypothesis that the processes of inference are based on the processes of storage and the storage processes are first recovered.)

STORAGE PROCESSES

The simplest storage process is the process that creates a thread, a new representation in the memory. Usually new information is based on pre-existing information, or at least on the descriptional information that is given by the single modality recognition memories, or information from other cognitive memories. The module in which the new thread is created depends on the way in which the information is acquired. An important class of storage processes is assimilation processes. These assimilation processes chain together intermediate deduction steps so that the entire chain is available as a single deduction step. The assimilation processes build the thread, and make it more complex by storing the result of deductions (Figure 5b).

There are other types of storage processes, such as processes that optimize the thread. First, threads that share a number of common nodes constitute bundles of threads. These bundles are thicker on the upper part of the thread, where the information is less specific, but they divide into smaller and smaller bundles until they break down to single strands. Bundles have a way of recording their thickness. This recording is very important for the accessing of the information. Unless a special context instruction is given, usually the thickest bundle is looked up first. The thickest bundle in a general thread represents the prototypical representation of an object, and it is accessed first. The thickness of bundles is not fixed, for it changes with every incoming thread. Thus the prototype might change too.

Another optimization process is used when it appears that several threads are sharing a common part. Then it seems useful to make a rearrangement so that all the simple threads share a single pointer to the common part. In other words, the uncommon part constitutes a paradigm in the context determined by the common part. The elements in the paradigm behave as if they were the same. In that context we can replace the paradigm with a new, more general symbol. For example, in the functional module we can have different elements that have the same function: one can throw a rock, a ball, a plate, a pencil, a piece of wood, etc. These objects form a paradigm in the context of throw. We can replace them with a common symbol, such as small physical object. This new type of symbol, called *micro-concept,* will be represented in the descriptional module by the elements that the elements in the paradigm have in common. Thus we see that this new, more abstract concept is derived from particular examples of objects in the real world that can function in certain way (e.g., can be thrown). *Micro-concepts are representational units for the object in the real world* (Figure 6). They serve for their better understanding and manipulation. They also serve for the interface with other cognitive modules, such as the modules involving language.

Massachusetts Institute of Technology Cognitive Science Center and
Boston University Computer Science Dept.

ACKNOWLEDGEMENTS

I am grateful to David Marr, for many stimulating discussions and to Richard Greenblatt, with whom much of the overall framework of thread memory was developed. I thank Dr. Norman Geschwind for carefully reading my manuscript and for his illuminating observations.

Many thanks to the members of the Aphasia Research Center and the speech therapists of the Veterans Administration Hospital Boston, for making my experimental research possible. I especially thank M. Walsh and Dr. Harold Goodglass for their encouraging support.

REFERENCES

[1] Geschwind, N.: 1967, 'The varieties of naming errors', *Cortex* 3, 97–112.

[2] Goodglass, H., Hyde, M. R., and Blumstein, S.: 1969, 'Frequency, picturability and availability of nouns in aphasia' *Cortex* 5, 104–119.
[3] Goodglass, H.: 1978, 'Disorders of naming in aphasia', Fellow's address, Division 6, American Psychological Association, Toronto.
[4] Marr. D.: 1969, 'A theory of cerebellar cortex' *J. Physiol. (Lond.)* 202, 437–470.
[5] Marr, D. and Nishihara, H. K.: 1978, 'Representation and recognition of the spatial organization of three-dimensional shapes', *Proc. R. Soc. Lond.* B 200, 269–294.
[6] Marr, D. and Poggio, T.: 1977, 'From understanding computation to understanding neural circuitry', *Neurosciences Res. Prog. Bull.* 15, 470–488.
[7] Poggio, T.: 1975, 'On optimal nonlinear associative recall', *Biol. Cybernetics* 19, 201–209.
[8] Szentagothai, J.: 1971, 'Memory functions and the structural organization of the brain', *in Adam, G. (ed.), Biology of Memory, Symposia Biologica Hungarica* 10, 21–25.
[9] Tulving, E.: 1972, 'Episodic and semantic Memory', in Tulving, E. (ed.), *Organization and Memory*, Acad. Press, New York.
[10] Vaina, L. M, and Greenblatt, R.: 1970, 'The use of thread memory in anomic aphasia and concept learning', *AI Working Paper* 195.
[11] Warrington, E. K.: 1975, 'The selective impairment of semantic memory', *Quarterly Journal of Experimental Psychology* 27, 635–657.
[12] Whitehouse, P., Caramazza, A, and Zurif, E.: 1978, 'Naming in aphasia: Interacting effects of form and function', *Brain and Language* 6, 63–74.