A CONTINUUM OF DIAGRAMMATIC DATA STRUCTURES IN HUMAN COGNITION

by

John A. Barnden
Computer Science Department
Indiana University
Bloomington, IN 47405

TECHNICAL REPORT NO. 131

A CONTINUUM OF DIAGRAMMATIC DATA STRUCTURES
IN HUMAN COGNITION

JOHN A. BARNDEN

OCTOBER, 1982

ABSTRACT

It is proposed that human cognition consists of the manipulation of data structures which are "diagrammatic images". Images are not assumed to be the objects of conscious experience. They are patterns in media which are abstractly in the form of two-dimensional arrays. These patterns can be pictorial, propositional or of a hybrid nature, and lie on a continuum analogous to that of ordinary diagrams. The patterns are interpreted and manipulated by a production system. By virtue of a gross characterization of the physiological implementation of the image-holding media, the model provides in outline a simultaneous answer to three important questions: (1) how are the brain's data structures implemented? (2) where do the lower-level data structures (feature maps) which arise in vision fit in the whole range of data structures in the brain? and (3) how can both pictorial imagery and propositional computation be elegantly included in a cognitive model?

ACKNOWLEDGMENTS

I am grateful to B. Chandrasekaran, Gray Clossman, Douglas Hofstadter, Margaret Intons-Peterson, John O'Donnell and Don Robinson for helpful discussions. The material is based on work supported in part by the National Science Foundation under Grant MCS-8102291.

SECTION 1: INTRODUCTION

"The brain computes by manipulating diagrammatic images." This is the core of an imagery-based cognitive model described in this report. There has been much debate* about whether human cognition makes use of "pictorial images", under various interpretations of that term. Let us take pictorial images to be patterns, in some "graphic medium" in the brain, which resemble the world in much the way ordinary pictures do. The cognitive model postulates that the brain makes use not just of pictorial patterns in such media, but also of patterns from a much wider class. This class of patterns is analogous to the full range of ordinary diagrams (patterns of marks on external graphic media). At one extreme, a pattern can be pictorial. At another extreme, a pattern can be purely ***"propositional"*** – that is, serve the purposes that network structures, logical formulae and other abstract data structures serve in cognitive theories. In particular, the propositional patterns in the brain are typically similar, in a ***gross*** way, to semantic network diagrams to be found in the literature (see e.g. Findler (1979)).

The purpose of this paper is to present in clearer and fuller form the claims just sketched. Detailed discussion and comparison with other work is, for the sake of brevity, postponed to later papers. The model is very briefly outlined in Barnden (1982). Two points should be clearly understood at the outset. First, the issue of conscious imagery is not addressed by the model – the images conjectured to exist are data structures which are not assumed to correspond to or generate conscious experiences. Second, what will be proposed is really a model schema rather than a model – many parameters will have to be specified for a model to appear which is testable by psychological experiment or computer simulation. For convenience, however, we shall continue to use the term "model".

The model provides a rough but unified answer to three important open questions about the nature of human cognition:-

(Q1) How is the brain's data structure manipulation implemented in terms of physiological mechanisms?

(Q2) How does the manipulation of data structures derived from perceptual input fit in with the brain's data structure manipulation generally?

(Q3) What is the relationship between the brain's spatial information about physical domains and its non-spatial information (about physical or non-physical domains)?

Underlying each of these questions there is a certain strong assumption. The assumption for the first question is the following.

---

* See, for example: Anderson (1978, 1979), Baylor (1971), Block (1981), Bower (1972), Chase and Clark (1972), Cooper and Shepard (1973), Fodor (1975, pp.174ff), Hayes (1973), Hayes-Roth (1979), Hebb (1977), Hinton (1979), Kieras (1978), Kolers and Smythe (1979), Kosslyn (1981), Kosslyn and Schwartz (1977), Kosslyn et al. (1979), Neisser (1976), Paivio (1971, 1977, 1980), Palmer (1975, 1978), Peterson, Peterson and Ward-Hull (1977), Peterson, Thomas and Johnson (1977), Posner (1973), Pylyshyn (1973, 1978a, 1978b, 1979, 1981), Richardson (1969), Schwartz (1972), Simon (1972, 1978), Sloman (1971, 1975).

The Data-Structure Hypothesis

> The brain's short-term cognition (natural-language understanding, perception, problem solving, general thought processes) can be explained largely as the manipulation of data structures in approximately as straightforward a sense as the sense in which the operations of a digital computer (running a program) can be explained as the manipulation of data structures.

This hypothesis is our version of the prevalent "symbol-manipulation" view of cognition, espoused by many researchers in cognitive science (notably Newell (1973) and Newell and Simon (1972)). We shall not argue for its validity here (even though it is not universally believed, and Dennett (1978) has argued that there is no cogent reason for accepting it). We content ourselves with taking Fodor's stance (Fodor (1975, p.27), Fodor (1981, p.29) that it is the most fruitful working hypothesis which is currently available. The word "largely" in the hypothesis allows some cognitive activity, such as low-level components of perceptual and motor activity, to escape the necessity to be straightforwardly explained in terms of data structure manipulation. Note that the hypothesis does not say that the data structures in the brain bear any resemblance to those in a computer. Nor does it say that the sense in which the brain manipulates data structures resembles the sense in which computers do.

The assumption underlying the second question is the following, and concerns a special form of data structure in the brain.

The Feature Map Hypothesis for Vision

> Low-level preprocessing converts retinal stimulation into a "feature map" physiologically implemented in the brain. A feature map is a temporary association of values with the elements of a finite 2D array. Each element corresponds to a small region on the retina; neighborhood in the array corresponds to neighborhood of regions; and the values associated with an element constitute information about the presence of certain features (such as small line segments, intensity changes, color, etc.) of the retinal stimulation in the corresponding region.
>
> (Thus the array is a medium analogous to an external graphic medium, and a feature map is a pattern in that medium. Feature maps are special cases of the images postulated by the model.)

Feature maps are similar to the lower-level representations commonly used in artificial-intelligence vision systems (Hanson and Riseman (1978)) and to Marr's raw primal sketches (Marr (1976)). It is sufficient for the purposes of this paper to adopt the simplifying pretence that we are monocular and that our feature maps are rectangular arrays (in an abstract sense, not necessarily a physical one).

It seems probable that some sort of feature map mechanism is physiologically implemented in the brain, because of the nature of the "columns" in primary visual cortex (Hubel and Wiesel (1979)). It is important to note, however, that the feature map supported by these columns need not be the only one in the brain, and that the idea that the main function of the columns is to support a feature map has been challenged (e.g. Pribram (1971, Ch. 6)). Non-visual perception will not be covered in this paper, although the author is studying an expansion of the ideas of the paper to incorporate it (see the brief comments in Section 8).

We may regard a feature map as being in some sense a picture of part of the physical world. We say this because the feature map is similar in structure to a system of adjacent regions in a certain (retinal) projection of the real world, and the information at each feature map element is of a low-level type. In a moment we shall move on to say that, not only is the brain capable of ***interpreting*** feature maps, but is also capable of ***internally generating*** patterns like feature maps, and of manipulating them for problem-solving and other cognitive purposes. These patterns, by virtue of their similarity to feature maps, are deemed to be "pictorial" and are analogous to pictorial drawings. We bring in this analogy to pictures to provide a transition to the next assumption, which was inspired by consideration of pictorial and other sorts of drawing.

Suppose onehas a (perhaps very schematic) pictorial drawing of some furniture in a room seen from above, and that one is using this picture to plan the layout of the room. Then, given any position P on the drawing it is easy to determine whether there is part of a piece of furniture at the room place corresponding to P. Similarly, given a position P on the drawing it is easy to find other positions which correspond to places in the room neighboring the room place associated with P. It is these properties of the drawing which allow one to plan the positions efficiently, of the furniture so that no two pieces of furniture occupy the same space, pieces of furniture are not too close together, certain types of furniture are close to the walls, and so on. In a very similar way, a bird's-eye drawing of a billiard table allows efficient solution of problems concerning the approximate movement of balls on the table. To switch to a more abstract example, consider a geographical map. If one has a finger on a particular place P on the map then one immediately has access to information about the world at the position corresponding to P. If a line on the map corresponds to the equator, then one can easily determine which countries the equator passes through by following that line and looking at neighboring map regions.

Consideration of the forms of problem solving greatly facilitated for us by the pictures in these examples (and other obvious ones) suggests that an important characteristic of the drawings is that they are "spatially indexed":-

> **Informal Definition.** A ***spatially indexed*** data structure is a data structure with the following properties. There is a set of items, called ***"spatial elements"***, which correspond to some regions in space (or in some projection of space). These regions cover a volume or area of space, and none of the regions contains any of the others. Associated with each spatial element e there is some information I(e) concerning the spatial region corresponding to the element. There is an ***efficiently implemented primitive operation*** which, given access to one of the spatial elements, gives access to I(e). There is another ***efficiently implemented primitive operation*** which, given access to one of the spatial elements, gives access to those spatial elements (if any) which correspond to regions neighboring the region corresponding to the given element.*

---

* Spatial indexing bears much of the import of the term "analogical" when used in relation to data structures describing physical situations, but that term is notorious for being used in confusingly different ways in the cognitive science literature.

This definition serves to make our furniture plans, billiard-table diagrams and geographical maps special examples of spatially indexed data structures (the "implementation" being ourselves): we take the spatial elements to be points (or rather very small regions) on the drawing or map, and the two operations are the operation of seeing what is on the paper at such a point e (I(e) being the marks on the paper at e) and the operation of turning attention to points neighboring a given point. Note that neighborhood of regions is represented as neighborhood of diagram points. Equally, we might have a furniture plan or map implemented on a computer, by having a conventionally-implemented 2D array whose elements take over the role of points on a piece of paper. The addresses of the array elements can be taken to be the spatial elements. We then have the efficient operations of retrieving the contents of the addressed memory location (or series of locations), and of finding "spatially neighboring" addresses. Alternatively, and more abstractly, the spatial elements could be pairs of array subscripts. The efficient operations would be the operations of retrieving the value associated with a subscript pair, and of computing the subscript pairs "neighboring" a given subscript pair.

The space containing the regions associated with the spatial elements in a spatially indexed data structure need not be essentially two-dimensional, as it is in the furniture and map examples. More importantly, the information I(e) need not be of the pictorial, low-level sort found in a pictorial drawing. Bobrow (1975, p.5) has suggested the use of a data structure which consists of an array A isomorphic to an array of points in space, where each element e of A is associated with an arbitrarily complicated and abstract item of information I(e) (represented perhaps as a list structure) concerning the world region around e's corresponding spatial point. (Hayes (1974) and Minsky (1975) have made somewhat similar suggestions.) With suitable implementation, e.g. by giving A a conventional array implementation on a computer, the data structure would qualify as being spatially indexed.

There is no implication that a spatially indexed data structure should explicitly involve array-like entities or graphic media, or even that the set of "spatial elements" should correspond to regularly spaced regions of space. For instance, a set of logical formulae specifying the locations of some objects (by means of predications such as "at(block1, 3,5,7)", say) would be spatially indexed if implemented in ***some*** way such that the two required operations are efficient. The spatial elements are those coordinate tuples which appear in some formula, but these explicitly represented positions might be sparsely and irregularly scattered.

It certainly appears that for some types of problem solving the use of spatially indexed data structures (and in particular, pictorial ones) is especially convenient. The use of blatantly pictorial spatially indexed data structures has been studied a little in the artificial-intelligence literature (e.g. Funt (1977)) and in the cognitive-science literature (e.g. Kosslyn and Schwartz (1977), Kosslyn (1981)). However, the pictorial quality of the data structures in these studies is only part of the reason that they are convenient for certain types of manipulation. It is the spatially indexed quality of the implementation of the data structures which accounts for much of that convenience.

It is now claimed, without further argument, that the convenience of spatially indexed data structures for some types of cognition is sufficient to justify adopting the following assumption.

The Spatial Indexing Hypothesis

> a) Some of the data structures the brain constructs and manipulates for problem solving and other cognitive purposes are ***spatially indexed data structures*** holding information about physical situations. Some of the spatially indexed data structures the brain uses are ***"pictorial"***, in that the information associated with the spatial elements of the structure is at a low level (like the feature information at feature map elements).
>
> b) The physiological implementation of the array medium in which feature maps appear is such that the medium together with a feature map in it constitutes a spatially indexed data structure.

The assumption that the feature maps are in fact spatially indexed data structures is quite natural, and certainly the feature map proposals with which the author is familiar appear to assume spatial indexing in any reasonable implementation of the feature maps.

The third of our questions should be interpreted in the context of the spatial indexing hypothesis. The question becomes partly one of the way in which the brain's spatially indexed data structures fit in with other data structures that the brain may use. However, it is important to realize that it is not being claimed that the use of spatially indexed data structures exhausts the brain's methods for dealing with spatial information. The possibility of data structures which contain spatial information but are not spatially indexed is not excluded.

A fourth important assumption we make is that some of the data structures the brain manipulates are "propositional" in nature – that is, roughly, they are structures of predicative form where the predicates are at arbitrary levels of abstractness or concreteness.

The attention of the model schema is focussed on rapid, short-term cognitive processing, such as is involved in the interpretation of visual input, understanding of natural-language text, problem solving, etc. The paper has little to say about the nature and acquisition of long-term knowledge, and omits detailed consideration of such matters as the low-level transduction of perceptual input and the production of motor and other output. Another restriction of the scope of the discussion is that it concerns only normal, adult, human cognition.

Section 2 discusses a familiar continuum of graphic notations. Section 3 outlines the thinking that led to the formulation of a diagrammatic cognitive model, and defines the abstract notion of diagram which we use later in our cognitive model. Section 4 explains how a computational system could operate by the manipulation of net-like propositional diagrams. Section 5 outlines the extensions necessary to cope with diagrams with pictorial as well as propositional aspects. Section 6 states more precisely what is being claimed about human cognition. Section 7 discusses some justifications for studying the model schema and mentions some putative objections to it. Section 8 is the conclusion.

SECTION 2: THE DIAGRAMMATIC CONTINUUM

As presaged by the first sentence of the introduction, the cognitive model to be presented is intimately bound up with the notion of a diagram. We can conveniently use the term "diagram" to encompass all the forms of 2D graphic notation with which we are familiar in our everyday lives. Graphic notation includes, at one extreme, faithfully pictorial diagrams and, at another extreme, "propositional" diagrams. Two examples of "propositional" diagrams are a body of natural-language text and a network diagram such as might be found in a paper about semantic networks (see e.g. Findler (1979)).

We shall henceforth use the term "spatial (part of a) diagram" to mean (a part of) a diagram in which there is some natural correspondence between positions in the diagram and positions in space. As we noted in the furniture-plan and other examples, spatial diagrams are automatically spatially indexed, by virtue of the way we manipulate them.

Propositional diagrams are an extreme form of non-spatial diagram. Note that propositional diagrams can nevertheless deal with spatial matters. A pictorial diagram is an extreme case of a spatial diagram. We shall not attempt to define precisely what it means for a diagram to be pictorial.* A pictorial diagram can be a picture in some commonplace sense of some part of a physical world. It can depart dramatically from being photographically precise – it can be approximate, schematic, stylized and oversimplified. Thus a stick figure of a person can be classed as pictorial. An equator line on a map contributes to the map failing to be pictorial in any strong sense, because the line does not picture any irregularity in the world itself. The line does nonetheless correspond to a set of world points which are of special interest. We can therefore describe an equator line as being a pictorial feature of a map in some weak sense.

A non-pictorial spatial diagram is one where the pattern in a diagram region corresponding to a spatial region is interpreted as an abstract piece of information about the spatial region. The pattern may, for instance, be a natural-language fragment. Such a diagram therefore automatically mixes propositional and spatial aspects. It is a commonplace that the diagrams we use mix non-spatial (including propositional) and spatial (including pictorial) aspects in various and intricate ways. Maps, cartoons, advertisements, musical notation, diagrams used in physics problems and (semi-)pictographic writing provide ready examples. (Some such examples are discussed in Fodor (1975, e.g. p.190) and Schwartz (1981, especially p. 117). Some network diagrams have (schematic) pictures at or in nodes. For instance, in Raphael (1976, p.81) a search graph for the Towers of Hanoi puzzle is illustrated as a linked set of nodes, in the usual way, but each node contains a schematic picture of the three towers and the discs on them. Fig. 1 shows another example of mixed picture/net notation. Network diagrams used in scene-analysis work (see e.g. Hanson and Riseman (1978) and Winston (1975)) often have a spatial aspect in the sense that nodes are arranged in the diagram in a way roughly corresponding to the physical arrangement of the objects denoted by the nodes. Indeed, much of the usefulness of the diagrams would be lost if they were not spatial in this way.

---

* Goodman (1968) has pointed out the large cultural factors influencing the attribution of term such as our "pictorial".

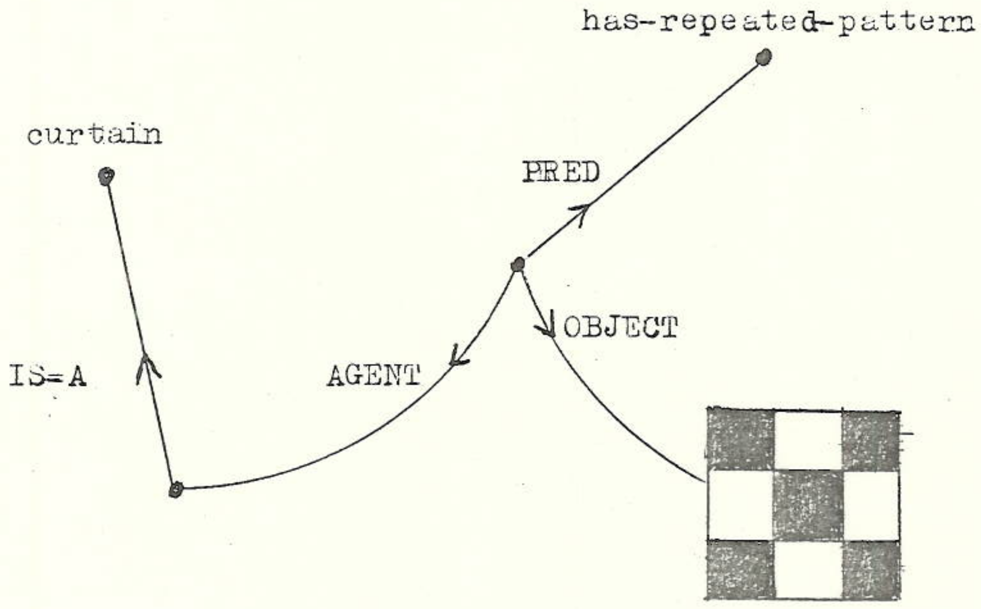

Fig. 1: A simple example of a mixed pictorial/propositional diagram.

Diagrams which are largely pictorial shade gracefully into more abstract diagrammatic notation. For example, a map of an underground railway system is more schematically pictorial than a map showing the true geographical relationships of the railway lines, but less abstract than a diagram in which stations are nodes and connecting railway lines are shown as links but there is no correspondence between the positioning of nodes in the diagram and real positioning of stations. The geographical-map/usual-map/net-diagram progression here is but a selection of three points on a continuum of possible diagrammatic representations of the railway system.

In partly spatial diagrams, positions and sizes of some patterns are directly related to spatial positions and sizes in the world. There is a closely related diagrammatic phenomenon where patterns whose relative diagrammatic position or size conveys abstract information. For example, in a "pie chart" showing where a university gets its money, different sectors of a disk (the "pie") correspond to different funding sources and have size proportional to the monetary contribution of that source. We can regard the pie diagram as being a ***"spatial analogue"*** of a non-spatial situation: a "spatial" measuring scale is being used to correspond in an obvious way to an abstract measuring scale. Other similar examples of spatial analogue are provided by x-y graphs of functions and by histograms. In network diagrams, the importance of a node can be indicated by the size of the node. The importance of a link can be inversely related to its length, so that for a given node, importantly related nodes are diagrammatically close to it. (Indeed, this importance/length relationship is discernable in net diagrams in the literature.)

The type of spatial analogue in these examples is called "size analogue". The other main type is "position analogue". A familiar example of position analogue is the representation of time by space in a diagram. An ordinary calendar is a case in point. Although time is represented linguistically and numerically as well, much of the point of normal calendars would disappear if they did not have their analogue aspect. Relative spatial position also provides the analogue in Venn diagrams of sets. Here, blobs in the diagram represent sets, and their overlapping is an analogue of set intersection. Normally the size of blobs or overlapping regions is of no significance, but occasionally a variant of Venn diagrams is used in which the size of a region is roughly indicative of the size of the denoted sets. This variant thus incorporates size analogue as well. In tree diagrams (e.g. of family trees, taxonomy trees, part-of hierarchies) the nodes nearer the root are often placed higher in the diagram. This is an example of position analogue. This analogue is redundant when the hierarchy is completely defined by directed links drawn between nodes. However, the links often do not have direction marks on them, being considered to lead from their higher ends to their lower ends.

We have noted that diagrams in which position corresponds to spatial position are automatically spatially indexed. Clearly, there is a derived concept of "analogue-spatially indexed" which applies to parts of diagrams which are spatial analogue rather than spatial. Just as spatial indexing facilitates certain types of problem solving in a spatial domain, the analogue-spatial indexing facilitates some problem solving in non-spatial domains.

SECTION 3: A SKETCH OF AN ANSWER

Here we take a look at the thinking which led to the paper's answer to the three questions posed at the beginning of Section 1.

Recall that we are assuming that some of the data structures the brain uses are feature maps and that these structures are spatially indexed. Recall also that we drew an analogy between feature maps and pictorial diagrams (which, note, are also spatially indexed data structures, for us). Fig. 2 portrays the state of affairs, including the continuum of graphic notation noted above. Now consider that we are faced with the problem of where feature maps, and spatially indexed data structures in general, fit in with the whole range of data structure types in the brain. (This problem is part of questions Q2 and Q3.) But look at how pictorial diagrams, and spatial (therefore spatially indexed) diagrams in general, fit in with the totality of diagrams: they are simply special cases of the patterns of marks that we are capable of interpreting. What these observations suggested to the author was the following:

> ***it would be worth investigating the idea that the brain contains media analogous to external graphic media and that the brain's data structures — whether feature maps, other (internally generated) spatially indexed data structures, or propositional data structures — are "diagrams" on those media.***

This idea is the core of the cognitive model schema to be investigated. The author is the first to grant that the idea is by no means implied by the foregoing discussion.

An important motivation for studying the idea that the brain's data structures are "diagrams" is that it is very economical, in the sense that the assumption that the brain uses spatially indexed feature maps introduces diagram-like entities anyway, so we may as well see how far we can push the idea that the brain can deal with "diagrams" (patterns in "graphic media"). Of course, we must give some account of how the brain is meant to interpret and manipulate its "diagrams". Much of the rest of the paper is concerned with giving such an account. Note in particular that we must show that the "graphic media" can be implemented in such a way that spatial "diagrams" are indeed spatially indexed. (We remarked that ordinary spatial diagrams are automatically spatially indexed because they are "implemented on us". This does not mean that to get brain diagrams to be spatially indexed we have to assume a homunculus in the brain which looks at diagrams, and whose cognitive activities must in turn be explained, ...) We shall present a gross conjecture (in Section 6) about the way the "graphic media" of the brain are physiologically implemented. Thus question Q1 will be answered in gross outline. Furthermore, not only does the diagrammatic idea answer questions Q2 and Q3, but also it answers them in an elegant manner: the integration of all the different sorts of data structure used in cognition (according to the Data Structure Hypothesis) is as smooth and rich as is the integration of all the different sorts of diagram under the umbrella of graphic notation. (It is not, however, claimed that the brain's "diagrams" are necessarily very similar to external ones.) The desirability of a smooth integration of information types stretching from the pictorial to the propositional has been argued by Schwartz (1981, pp.127ff) and Fodor (1975, p.190), although of course those authors might disagree with our method of achieving that integration.

It is now time to say what the "graphic media" in the brain are meant to be, at an abstract level of description. Really all that we shall be doing is abstracting away from

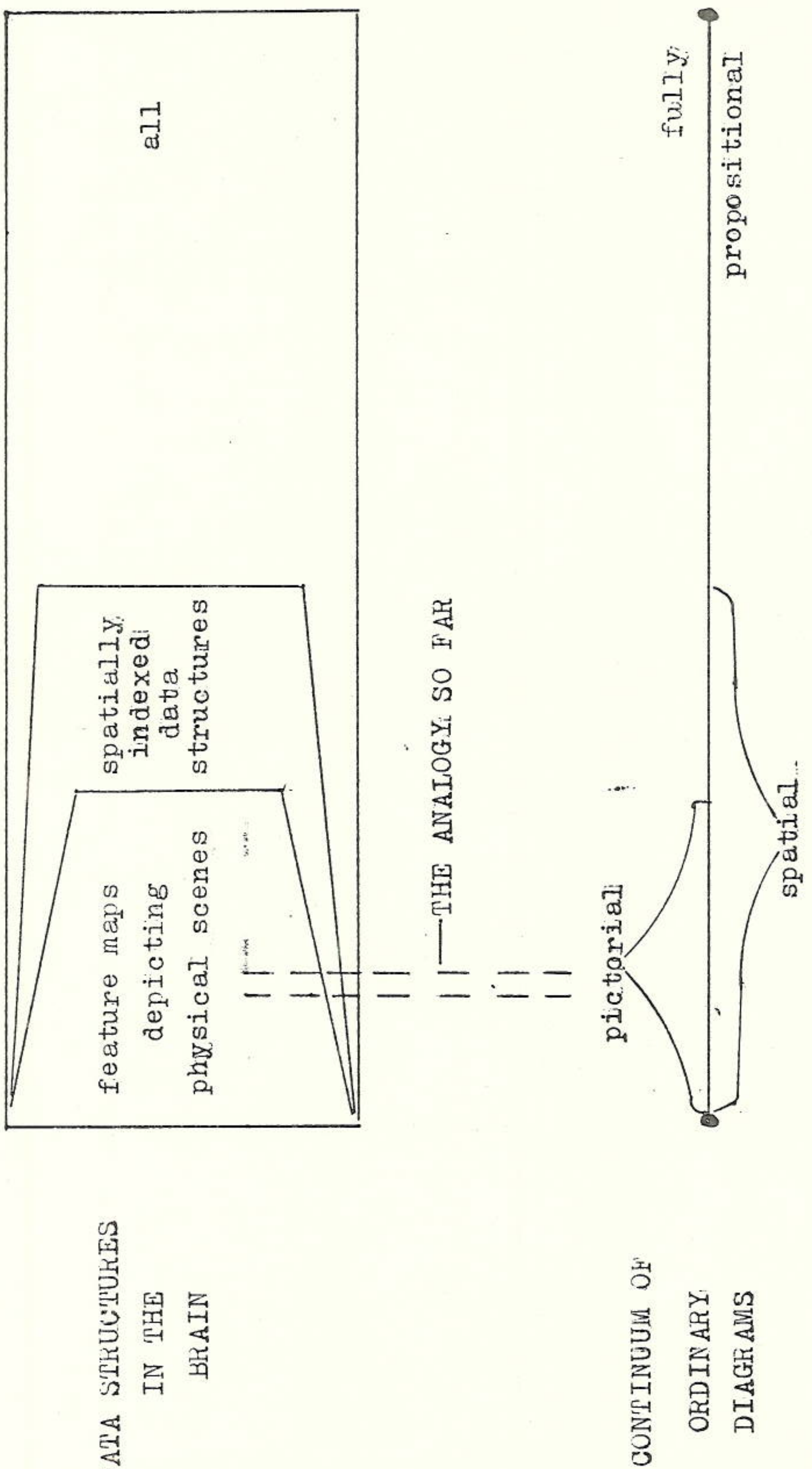

Fig.2: The depiction of the diagrammatic continuum is highly simplified.

any particular physical medium, and insisting on a medium made up of a finite number of elements. Henceforth, we shall talk about "imaginal matrices" instead of "graphic media" in the brain.

> **Informal Definition.** An ***imaginal matrix*** (IM) is an abstract, finite, n-dimensional rectangular array for some $n \geq 2$, where the array elements are called ***imagels,*** together with a set of of ***imaginal features*** (IFs) and, for each imaginal feature, a range of possible numerical values. An ***Image*** is an association of a value with each imaginal feature at each imagel in an IM (so that an Image is a pattern of IF values over an IM).

The point of the paper would not be affected by allowing other forms of pattern-holding matrix (e.g. matrices with hexagonal or polar "geometry"), but rectangularity is convenient for the purposes of simple discussion. The term "Image" will always appear with a capital letter to avoid confusion with other notions of "image" used elsewhere.

An IM is an ***abstract*** entity - if an IM were supposed to reside in a computer, brain or whatever, no assumption that it is manifested as some sort of physical array would be necessary. It is nevertheless convenient for heuristic purposes to imagine the imagels of an IM as forming a rectangular array in physical space. More particularly, we shall be concentrating on 2D IMs, and using the analogy of an IM as the points of a fine, invisible grid imposed on a piece of paper (or as the image points of a video screen). Then the IF values can be thought of as visible features such as intensities of marks, colors or whatever . (The choice of visible manifestation is arbitrary except in the case of an Image which is a feature map, the IFs in that case having a fixed interpretation in the heuristic. A feature map IF standing for a small line segment, say, would be illustrated by means of a small line segment on the paper.) An Image is then thought of as a diagram on the piece of paper (or screen). This analogy allows us to talk about, and illustrate, Images as if they were diagrams as discussed in Section 2. For instance, a pictorial Image is an Image which, with suitable visible manifestations for the IF values, can be said to correspond to a pictorial diagram.

SECTION 4: NETWORK IMAGE MANIPULATION

This section tries to clarify what it could mean to say that computation can take the form of the manipulation of propositional Images in 2D IMs. Such computation is part of the postulated activities of the brain. We shall return to pictorial and combined Images later. For definiteness, we take the particular example of the manipulation of conventional-network Images – without making any claim that precisely such Images are used in the brain, or indeed that one would want to build an artificial computing system using such Images. A conventional-network Image is an Image whose illustration on paper (by means of the heuristic mentioned in Section 3) is similar to the diagrams of conventional semantic networks found in such works as Findler (1979), Anderson and Bower (1973), etc. Thus a node in the Image is a small localized group of imagels for which some IF has a value distinctly different from the values for that IF in a surrounding region of imagels; a link is a similarly highlighted narrow ribbon or chain of imagels joining two nodes, with some "mark" on the ribbon to indicate the direction of the link; and a label on a link or node is some pattern adjacent to the link or node. We are, in effect, upgrading network diagrams from the status of visual aid to the status of bona fide data structure.

The precise sort of pattern which can be a node or link label is not of great importance at this point in the paper, but for convenience we illustrate our network Images by diagrams in which labels are natural-language fragments. It is worth noting that differnt sorts of node and link can be separated by using different IFs to "draw" them. We can conveniently summarize this by saying that nodes and links can be of different "colors". We do not expect, however, that there would be enough IFs to obviate the need for link labels.

For simplicity we shall start by assuming that there is just one IM, and that it is large enough to incorporate whatever network diagrams we might want the system to deal with. At the end of the section we shall introduce the modifications made necessary by having, instead, a number of finite IMs of a size such that a given network might have to span several IMs.

An important part of what is meant by propositional computation using Images is that the Images are ***"essential"*** to the computation – that is, it is not the case that Images are converted into or produced from non-Image data structures elsewhere in the system, the real computation being manipulations of those other data structures. Also, we assume that the mechanisms which interpret Images contain relatively little data of their own, and that what there is is of restricted sorts (e.g. data for holding control information in the course of doing a manipulation made up of a number of steps). Thus, most of the data-holding capability of the system is meant to reside in the IMs. This statement will become a little less vague later on.

In the rest of this section we look at the sorts of net manipulation a network-Image interpreter must perform and how it can perform them. We shall not, however, attempt to treat an example of the application of networks to any particular knowledge domain. Thus we shall be concerned entirely with general "sytactic" matters. The interpreters we shall consider examine networks in a fairly conventional manner. That is, they basically proceed by following paths along links in the network Image and take actions on the basis of the sorts of nodes, links and labels they detect. At a high level of description, the sorts of manipulations performed by currently implemented network systems are typified by the following operations: traversing specified sorts of paths; matching subnets; addition of subnets (linking them into the existing nets); deleting subnets; converting subnets to and from other forms (e.g. natural-language text); executing a piece of net acting as a declarative description of a process; and using the presence of certain sorts of subnet to trigger processes (a simple example of this is when a procedure is "attached" to a node). There is great variation among net-based systems as to which of the above operations are performed and as to the exact nature of the operations. However, we can discern a set of intermediate-level operations which are necessary or desirable to support the high-level operations listed:

- checking the presence of a particular label on a given link
- choosing an action on the basis of a particular label or on the basis of a particular simple configuration of labelled links based at a single node
- at a given node, finding a link (perhaps with a specified label) from the node
- at a given node, finding the label (if any) on a link to or from the node
- given a link, finding the node at a particular end of it (this allows traversing of links)
- remembering a traversal path so that it can be re-traversed later (perhaps in the opposite direction to the original traversal)
- deleting a node and the links emanating form it or going to it
- inserting a new node and linking it to existing nodes
- inserting and deleting links between existing nodes
- inserting, deleting and changing labels
- marking nodes and links as being of current interest.

Note that to remember paths we need to be able to "keep a finger" on each of a finite set of nodes and/or links, where there is no a priori bound on the size of the set. We cannot get away with a fixed finite set of "fingers" (marks, pointer variables or the like).

What we should like to know eventually is how these intermediate-level operations can be implemented in terms of low-level operations which fiddle about with imagels' IF values, and moreover can be implemented with very little data-holding capability in the interpreter itself. Instead of treating this implementation problem in the general form at which it stands, we shall look at the particular precise case of the operations needed for a certain simple type of manipulation on a certain simple type of network Image. Also, we shall not go down quite to the low level mentioned above.

Section 4.1: A Simple Example

The type of network Image considered here is an arithmetic expression tree of the form illustrated in Fig. 3. The tree in the figure is a representation of the expression

$$(5 + 6*7) * (8*9 - 10)$$

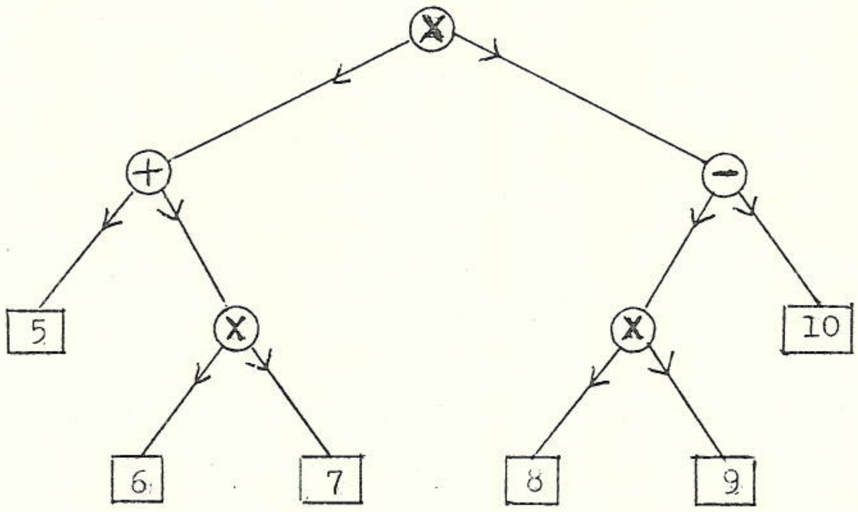

Fig. 3: The relative positioning of the nodes is insignificant.

The manipulation we consider is that of evaluating the tree, i.e. deriving the arithmetic value of the expression. The way this manipulation can be described at an intermediate level is in terms of the movement of numbers and certain markers about in the tree. These movements are derived from conventional methods of traversing tree structures implemented on computers. The markers are small patterns which can be placed next to nodes or links of the tree. There are four types of marker: '?', '!', '#', and '%'. (The quoted symbols will be used in figures and text, although in actuality the markers could be patterns of other forms. A simple option would be for the markers to be small blobs of four special, distinct "colors".) '?' and '!' markers are placed next to nodes, while '%' and '#' markers are placed next to the beginning of links. In addition, any '!' or '%' marker is the start of a link of a special sort which ends at some number. This special sort of link is called a "shows" link and the number is called the "showee" of the marker. ("Shows" links might be distinguished from tree links merely by being in a different "color".) A '?' next to an operation node (i.e. internal node) of the tree indicates that the subtree at that node is ready to be evaluated (but evaluation has not yet started). The '#' markers merely manage a flow of '?' markers down the tree. A '?' next to a leaf node is replaced by a '!' "showing" the number inside the node (i.e. there is a "shows" link from the '!' to the number). A '!' next to a node indicates that the subtree at that node has been evaluated, and that the result is shown by the '!'. '%' markers merely manage the passing of results up the tree. The action of applying an arithmetic operation to some numbers is initially considered to be indivisible, but a little later we shall see how that action might itself be decomposed.

The necessary manipulations of markers, numbers and links are carried out by a production system (Newell (1973), Newell and Simon (1972), Davis and King (1977), Waterman and Hayes-Roth (1978)) of the following informal nature. The productions are condition-action rules expressed in a constrained form of English. Productions fire one by one. If several productions are satisfied at any stage, an arbitrary one is allowed to proceed. Also, if this production is satisfied by several distinct Image parts, the production proceeds on the basis of an arbitrary one of these parts. The deletion of a '!' or '%' includes deletion of its shows link.

(Initialization)
A '?' next to the tree root.

(P1) **if** there is a '?' at a node with a non-'#' outlink
then insert a '#' on the beginning of this link.

(P2) **if** there is a '?' at an internal node with no non-'#' outlink
then delete this '?'.

(P3) **if** there is a '#'-link from a node with no '?'
then insert a '?' at the destination of this link;
delete this '#'.

(P4) **if** there is a '?' at a leaf node
then insert a '!' next to this leaf, showing the contents of this leaf;
delete this '?'.

```
(P5) if there is a '!' at a node with an inlink and no '%'-outlink
     then insert a '%' showing the showee of this '!'
            and next to the beginning of this inlink;
          delete this '!'.

(P6) if there is a  node containing a '+ ', with no '!' and with no non-'%' outlink
     then add the numbers shown by the '%' markers on the outlinks of this node;
          insert a '!' next to the '+ ' node
          and showing the result.

     <similarly for other arithmetic operations >

(P7) if there is a '!' next to a node with a '%'-outlink
     then delete the '%' on this link.

(P8) if there is a '!' at a node with no inlink and no '%' outlink
     then stop.
```

We should note an important primitive which is used in the productions' action parts. Each use of "this" in an action part betrays an implicit datum emanating from the production's condition and transmitted to the action part. This datum is the identity of the pattern (node, marker or link) which is meant to be denoted by the "this" phrase in the action part. The question is: where/how is this datum held? In fact, it is easy to hold the datum in the Image itself by marking the identified pattern in a certain way, namely by ***"highlighting"*** it with a special IF ("color"). Highlighting is done when a production is chosen for execution, and consists of giving an especially high value to the special IF at every imagel used in the pattern. The production's action part then "knows" the pattern simply as that pattern which is highlighted with the particular special IF. Similarly, highlighting can be used to perform feats of pattern-identity memory entirely within action parts.

A particular example will help to make the action of highlighting clearer. Consider production (P4). Suppose there is indeed a '?' mark at a leaf node and that (P4) is chosen for execution at this node. Then the '?' is highlighted with a special IF (IF1 say) and the leaf-node outline is highlighted with some special IF2. The sequence of actions taken by the production, at a lower level of description than before, could be as follows, where IF3 and IF4 are further special highlighting colors.

- delete any pattern highlighted by IF1
- insert a '!' pattern, highlighted by IF3, somewhere-close-to-and-outside any pattern highlighted by IF2
- highlight with IF4 any pattern contained within any pattern highlighted by IF2
- insert a shows link pattern from any IF3-highlighted pattern to any IF4-highlighted pattern
- remove all highlighting by IF2, IF3 and IF4.

IF1 and IF2 are used to keep track of ("point to", "mark") patterns discovered by the condition part. IF3 and IF4 are used to keep track of patterns created and discovered (respectively) by the action part itself. The use of "any" in these lower-level steps indicates that the production as a whole is not tied to any particular region in the IM and that the component primitive actions (insertions, deletions, etc.) are similarly not tied to specific regions.

We shall not look at how the necessary primitive pattern-recognition operations are performed – detecting nodes, links, markers and simple combinations of them. Note that such pattern-recognition occurs both in condition parts and in action parts. We shall later suppose (Section 6) that pattern-recognition of the sort required in this simple example is done by special "hardware" distributed over the IM.

The question naturally arises of whether operations pertaining to different productions can be performed in parallel – can production executions proceed in parallel? can condition testings proceed in parallel? etc. We would certainly want to look for detailed implementation techniques which would allow such parallelism. Problems are introduced, however, by the suggested highlighting techniques. There is potential for interference between the productions if different ones can involve highlighting with the same colors. Therefore the number of available IFs limits the extent of parallelism (of operations acting on a single IM) if highlighting is used.

Finally, comment is needed on the actual multiplication, etc. of numbers. These arithmetical operations were portrayed as being primitive operations in the production system . This portrayal was for simplicity of exposition – in fact the multiplications and so on can be broken down into Image manipulations. Assuming for definiteness that the numbers are in ordinary decimal notation (or rather, that they are linear sequences of pattern instances, one distinct pattern for each of the ten digits) a pair of numbers can be added together by "column-by-column" addition just as they would be by pen and paper (although it is not actually necessary for the numbers to be near each other in the IM.) Such column-by-column operations can be achieved by the highlighting and pattern insertion/deletion techniques referred to above – the details are straightforward. We are left with much simpler numerical primitives than the ones we assumed before. We now assume only that there are the primitives of adding together two digits and a carry, multiplying two digits, etc.

Section 4.2: General Pattern-Manipulation Primitives

Several network-Image examples have been worked out apart from the arithmetic-tree one, but they will not be detailed here. One involves networks which are representations of Lisp expressions, and the attendant production system acts as a Lisp interpreter. The Lisp example is managed using exactly the same sort of marker technique (though in a slightly more complicated form) as was used in the arithmetic example. In the Lisp example, paths of interlinked markers are set up to serve the purposes that stacks do in Lisp interpreters implemented on ordinary computing systems.

It is to be emphasized that the particular sort of network Image appearing in the examples which have been worked out are not claimed to be appropriate in putative real IM computing systems. They have been studied merely as exercises to clarify certain issues. One issue which has been clarified in considerable detail is that of the types of localized pattern examination and manipulation operation which are sufficient to support network Image systems (using a single large IM). In fact, we expect that the operations listed below are fundamental ones in the manipulation of all types of Image. In the following list, terms such as "specific pattern" and "specific position" are used and will be clarified in a moment.

- inserting a specific pattern (e.g. a marker) at a specific position
- deleting a specific pattern (e.g. box outline) from a specific position
- deleting from a specific region any pattern which happens to be there
- copying a specific pattern from one specific position to another
- copying from one specific region to another any pattern (e.g. a digit) which happens to be in the first region
- adding/deleting specific colors to/from existing specific patterns (e.g. to effect highlighting)
- adding/deleting specific colors to/from any pattern which happens to be in a specific position (for example to effect highlighting)
- following specific lines (e.g. network links) and outlines (e.g. of node boxes)
- using patterns to trigger complex processing (example when some production responds to the presence of a particular arc label).

A pattern is specified (for the purposes of the term "specific pattern") by shape and/or colors used, but not by position in the IM. A position or region is specified (for "specific position/region") by shape and/or colors used, by qualitative position within the IM, or by qualitative position relative to some specific pattern. "Qualitative position within the IM" incorporates characterizations like "tending to the left", "tending to the top" ("left", etc. being derived form our drawing analogy). Such gross IM-relative positioning is allowed so that structure can be imposed on an Image without the need for structuring patterns such as links, and is useful in some of the examples which have been studied in detail. "Qualitative position relative to a specific pattern" can mean "inside the specific pattern [assumed to be some sort of outline]", "close to but outside the specific pattern [which could be a small outline, blob or marker]", "close to and to the left of the specific pattern", and so on. The word "qualitative" has been used to underscore the idea that the precise position of patterns relative to each other or to the IM itself is meant to be essentially irrelevant to the Image manipulation mechanisms. (This point anticipates a postulate concerning IM systems in the brain – Section 6.)

"Following a line" was included in the above list of operations. A more precise description of the operation would be "access the other end of a specific line one of whose ends is already accessed", where the access might take the form of highlighting. We can actually cast line-following itself in terms of more general operations. Suppose there is a line one of whose end segments has been highlighted with color IF1. If we have the operation

- add a specific color to the whole of a pattern specified by being (partly or wholly)drawn with a specific color

then we can use it to highlight the whole of the line with a color IF2. This latter end is now detectable by virtue of being highlighted by IF2 but not by IF1.

Throughout, we have made virtually no reference to the direction of links. This is because the direction of a link may be used, for example, in the decision as to whether to traverse the link, but the link direction is in principle independent of possible directions of traversal of the link.* Note that if a particular link is to be traversable in both directions, it is helpful to have a label at each end of the link (as opposed to having only one label on the link). Indeed, if this is done and the labels are variants of each other (e.g. AGENT and AGENT-1) then there is no need to have another device specifying the link direction.

Section 4.3: Systems using Several IMs

In later sections IM systems in which there are a number of IMs of moderately restrictive size will be of special interest to us. In the context of network Images, we can take "moderately restrictive" to mean that no more than a dozen or two nodes and attendant links can be fitted in. We assume that all the IMs are isomorphic (same dimensions, same IFs and value ranges).

Let us then assume that we wish to implement a conventional semantic network in the IMs of such a system. Some method is necessary for tying together pieces of network lying in different IMs. It is possible to imagine a system in which a pattern (or single imagel value) in an IM could serve as the address or name of some other IM; similarly, a certain sort of pattern or single imagel value in an IM could serve as the address of an imagel in that very IM. When we come to the cognitive model as such (in Section 6) we shall avoid these naive addressing techniques. There is an alternative: a mechanism which ***associates*** sufficiently similar patterns in different IMs, i.e. notices that a pattern appears (perhaps with some limited variation) in two or more IMs. Suppose for instance that a node in (a network Image in) one IM must be equated with a node in another IM. The equation could be established by labelling the two nodes with the same label, perhaps drawn from a special class of labels. In other systems there could be other interpretations placed on the presence of the same label in two IMs.

We may therefore propose that a desideratum in an IM computing system is an ***association primitive*** with the following characteristics. When the primitive is applied to a sufficiently simple pattern in an IM, the primitive highlights sufficiently similar patterns in other IMs, provided that in each IM the pattern is in a sufficiently simple context. The IF used in the highlighting is a parameter to the primitive. Node and link labels should pass the test of being sufficiently simple (even if they are as complex as natural-language words, in our drawing heuristic). "Sufficiently similar" should allow only a reasonably small amount of distortion: the similarity is meant to be a spatial one – so for instance two pieces of net Image which were isomorphic as nets would fail to associate unless they were laid out in IM "space" in approximately the same way. A difference in the IFs used for the patterns to be associated is allowed; indeed, we could envisage parameters to the primitive indicating what range of IFs are allowed in associated patterns. Some variation in the IF values is also allowed. The "sufficiently simple context" proviso is included to relieve the primitive from the requirement that it notice patterns which are intricately embedded in other patterns.

---

* This point is often obscured in the literature, especially when there is an unstated assumption that links not only have abstract structuring significance but also denote pointers in a computer implementation – see Woods (1975) and Brachman (1979) for discussion of such confusions.

A very important requirement we place on an IM system incorporating the association primitive is that the primitive should be efficiently implemented. That is, the time taken by the primitive to find all the associated patterns should be comparable to the time taken, say, to detect the presence of a single primitive pattern (e.g. node, link, marker in a network) or to follow a link. ("Comparable", not "equal", is definitely intended here – perhaps a factor of about one order of magnitude could be allowed.)

Despite the "sufficiently simple pattern" criterion, it is ***not*** assumed, for the purposes of usefully applying the primitive to a pattern, that the pattern has to be one which is recognized by the primitive pattern-recognition mechanisms tied to IMs. For example, an inter-IM association label might not be one which is actually detectable (as anything other than merely ***some*** label) by any production.

There is no reason to forbid the primitive to operate within individual IMs as well as between them. This gives us an intra-IM structuring method over and above the use of adjacency and links. An advantage of intra-IM pattern association over the use of link patterns is that the overall comlexity of Images is reduced, and structuring within IMs becomes more unified with structuring between IMs. However, the local complexity of IMs may be increased (more labels are needed) and the time taken to "follow the link" may be greater. These drawbacks would force a compromise between the use of association and the use of explicit links. One advantage of using an associative technique and not an IM (or imagel) addressing technique is that patterns can be moved around within and between IMs without disturbing structure in any way.

The possibility of using associative techniques (or "content addressing") in computers has been known for a long time, and in fact some computers use them – this subsection is merely pointing out that such techniques are also useful in an IM system.

Section 4.4: The Meaning of Nodes

The meaning of a node (the effect its presence has on the interpreter in the net system) is defined by two things: the connections of the node to other nodes, and the labels or other information "in" or "at" the node. In some cases the connections alone are sufficient to determine the meaning. For instance, in Fig. 4 the hit-instance node (the one at the start of the PRED link) is known to be a hit instance node precisely by virtue of the PRED link, and all other information about the node needed by the particular system at hand is given by the other links adjacent to the node. However, some nodes do need labels (or other special attached information) – for instance, we may (depending on the precise nature of the total system) need the label "hit" in the figure to allow the interpreter to bring appropriAte processes into play.

One of the most important type of link from a node is a PRED link (or, similarly, an OCCURRENCE-OF, SUPERCONCEPT, IS-A, etc. link). Although such links can appear explicitly in network Images, there is an alternative technique which may make more sense, at least in simple sorts of network. Instead of a PRED link as in Fig. 4 we could have a label association as in Fig. 5(a) or Fig. 5(b). Here we are simply replacing a link by a pattern association. Indeed, in a multi-IM system we would tend to have to use an association instead of a link in any case. (Equally, the generic concept taking part in, say, an IS-A relationship might not ***have*** a node anywhere to represent it, in which case we take the label on the instance node to serve to "associate nodes to implicit procedural

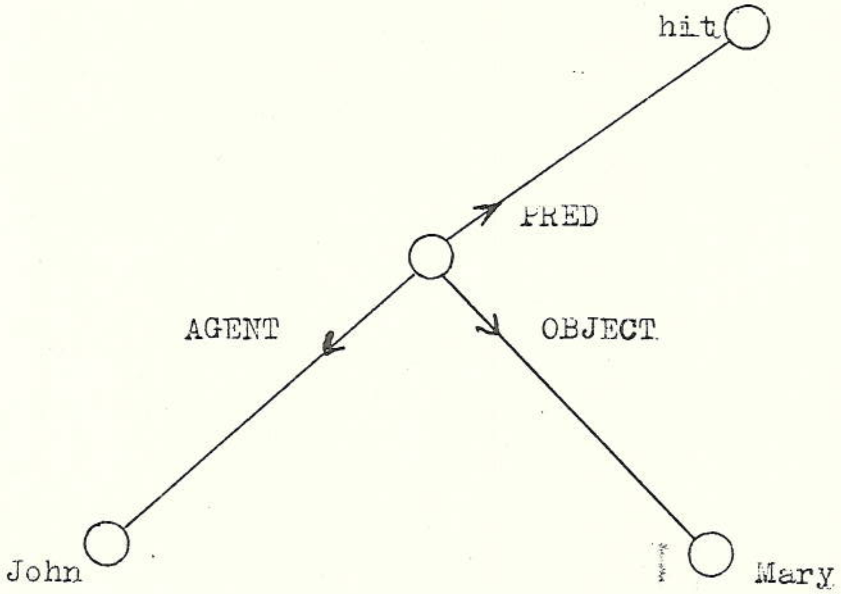

Fig. 4: A conventional-network Image or diagram.

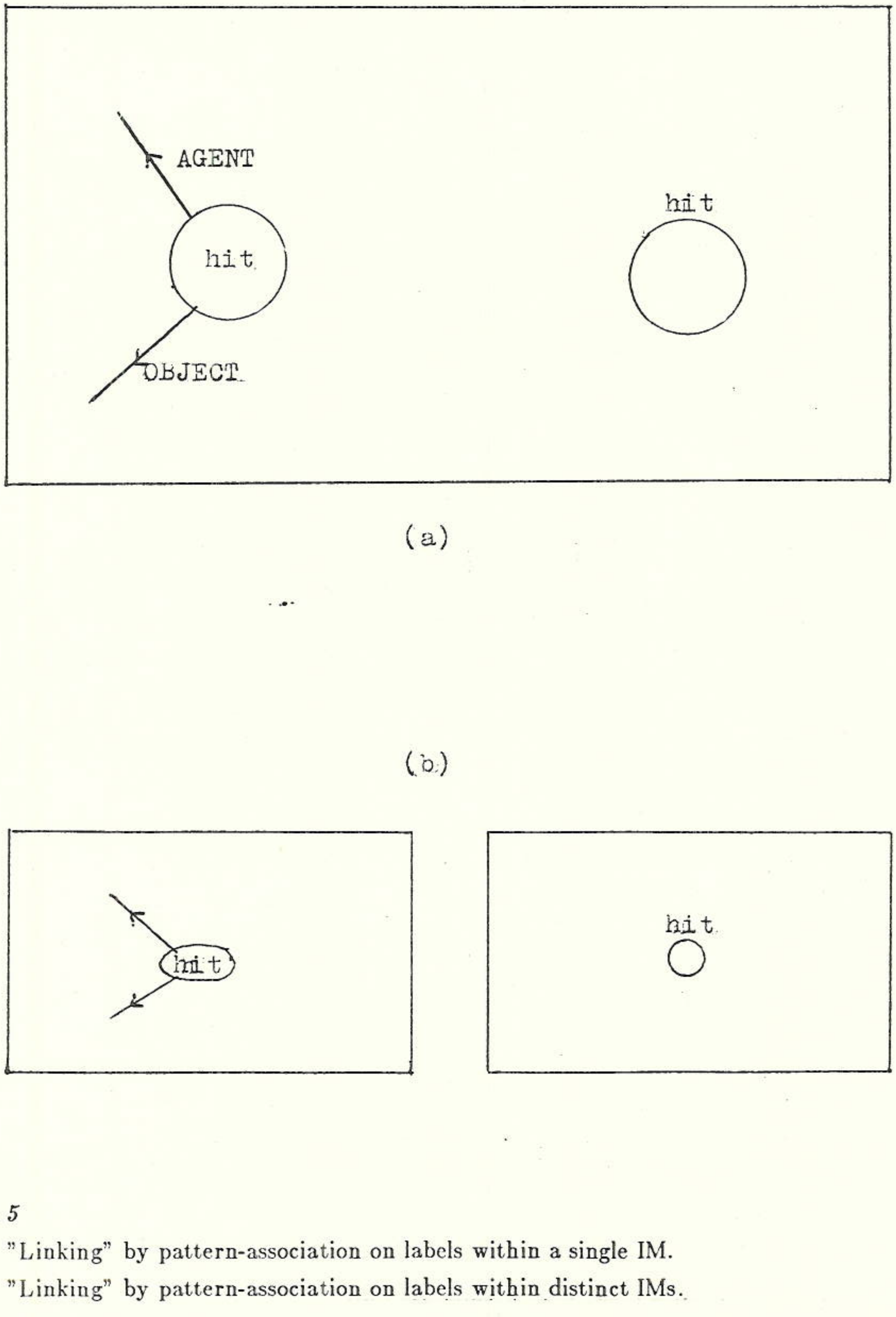

Fig. 5

(a) "Linking" by pattern-association on labels within a single IM.

(b) "Linking" by pattern-association on labels within distinct IMs.

knowledge".)

Note that in Fig. 5 the two "hit" labels are differently placed with respect to their nodes. Some differentiation is needed between the labelling used at the two nodes, because the connection we are establishing is directional.

Section 4.5: Locality

A significant characteristic of most of the network-Image manipulations described in this Section is that are based on ***"local"*** operations. That is, the basic manipulations used each deal with just a small portion of a single network Image (e.g. a node and a few links). These local manipulations are directly inspired by the local manipulations currently used in implementations of network systems in conventional computers. The locality is, however, not something we demand in general. The association primitive is an example of a basic mechanism which is non-local, both because an application of the primitive is concerned with many Images, or many parts of a single Image, and because there is no requirement that the associated patterns be small.

SECTION 5: PICTORIAL AND GENERAL IMAGE MANIPULATION

We now look at how pictorial Images could be manipulated in an IM computation system. The first point to emphasize is that an IM which manipulates pictorial Images is not excluded from manipulating other sorts of Image, such as propositional ones. In particular, the system may contain ways of producing a propositional Image summarizing a pictorial Image and of producing a pictorial Image from some prescription in a propositional Image. For instance, consider a "blocks-world" system (see e.g. Winston (1975)) where blocks can be of different sizes and shapes. Let us assume it is useful to have pictorial representations of situations for the purposes of some of the problem solving done by the system. But let us also assume that it may be convenient to have propositional representations of the qualitative spatial relationships (e.g. left-of) and support relationships in situations. Therefore, a change in a pictorial Image may be guided by a change in a propositional Image summarizing the same situation, and a change in a propositional Image may be required by a change in a pictorial Image. Of course, some means of associating parts of propositional Images with parts of pictorial Images is necessary. This issue will be taken up later.

Not only may propositional Images be used to describe the same block configurations as are depicted in pictorial Images, but other sorts of propositional Image might be manipulated as part of the computation required to manipulate the pictorial Images or their propositional counterparts. For example, an Image might (propositionally) represent a plan of action which is to be taken. Again, a propositional Image might contain a rule which says what should happen to a pile of blocks if the bottom one is removed. The general technique of guiding manipulation of pictorial Images by manipulation of propositional Images is an IM version of "cognitive penetration" into imagery, as discussed by Pylyshyn (1981).

We can envisage pictorial Images being manipulated in much the way the network Images of Section 4 are. For example, a movement of an Image portion picturing a block could be done by highlighting the block subImage and the desired destination region, and then applying a copy primitive and a deletion primitive like those in Section 4. (Note here that the copy and deletion primitives need not be sensitive to nature of the particular pattern being dealt with.) Of course, we are likely to need basic pattern-recognition mechanisms which are capable of detecting patterns picturing blocks. This pattern-recognition is likely to have to detect subImages which are considerably larger than those recognized in network-Image systems (e.g. nodes, labels). Note that once a block subImage has been detected, it can be marked or highlighted in such a way as to obviate the need for further pictorial pattern-recognition on that block.

It would be convenient to have a movement operator which had the same effect as a combined copy and delete, but more efficiently implemented. This operator would also be useful in network-Image systems. Another useful operator would be a movement operator which is given a direction and a distance as parameters rather than a destination region. With a small distance value this operator could be used to simulate continuous motion of pictured objects. In certain problems it might be necessary to keep on the lookout for collision of objects which are being subjected to simulated motion. We could therefore propose that there is a basic mechanism which detects adjacency and overlapping of subImages. Such a mechanism is in any case useful in the case of non-pictorial

Images. If an Image on an IM is a spatially indexed data structure, then the implementation of that mechanism would be efficient.

At present it appears that, given methods for manipulating pictorial Images and for manipulating propositional Images, there is no major difficulty in devising methods for manipulating Images at intermediate points on the continuum of diagrammatic notation. The major new factor is that the processes working on Images must be able to distinguish between pictorial, propositional and other types of Image aspect.

We saw earlier that an IM system solving blocks-world problems might manipulate both pictorial and propositional Images describing the same situation. If this were so there would have to be a way of identifying parts of a pictorial Image with parts of a propositional Image. There are at least two simple ways in which this might be done. Let us assume that the propositional Image is of network form. Then object subImages in the pictorial Image could be annotated with labels of some sort which are also used to label nodes in the propositional Image. We appeal here to the pattern-association primitive of Section 4.3. The second method is for the position of a node in the propositional Image to correspond to the position of the corresponding object in the pictorial Image. But this method suggests we could sidestep the whole issue by having Images which contain superimposed pictorial and propositional patterns, as illustrated in Fig. 6. The two superimposed Images might or might not use disjoint sets of IFs.

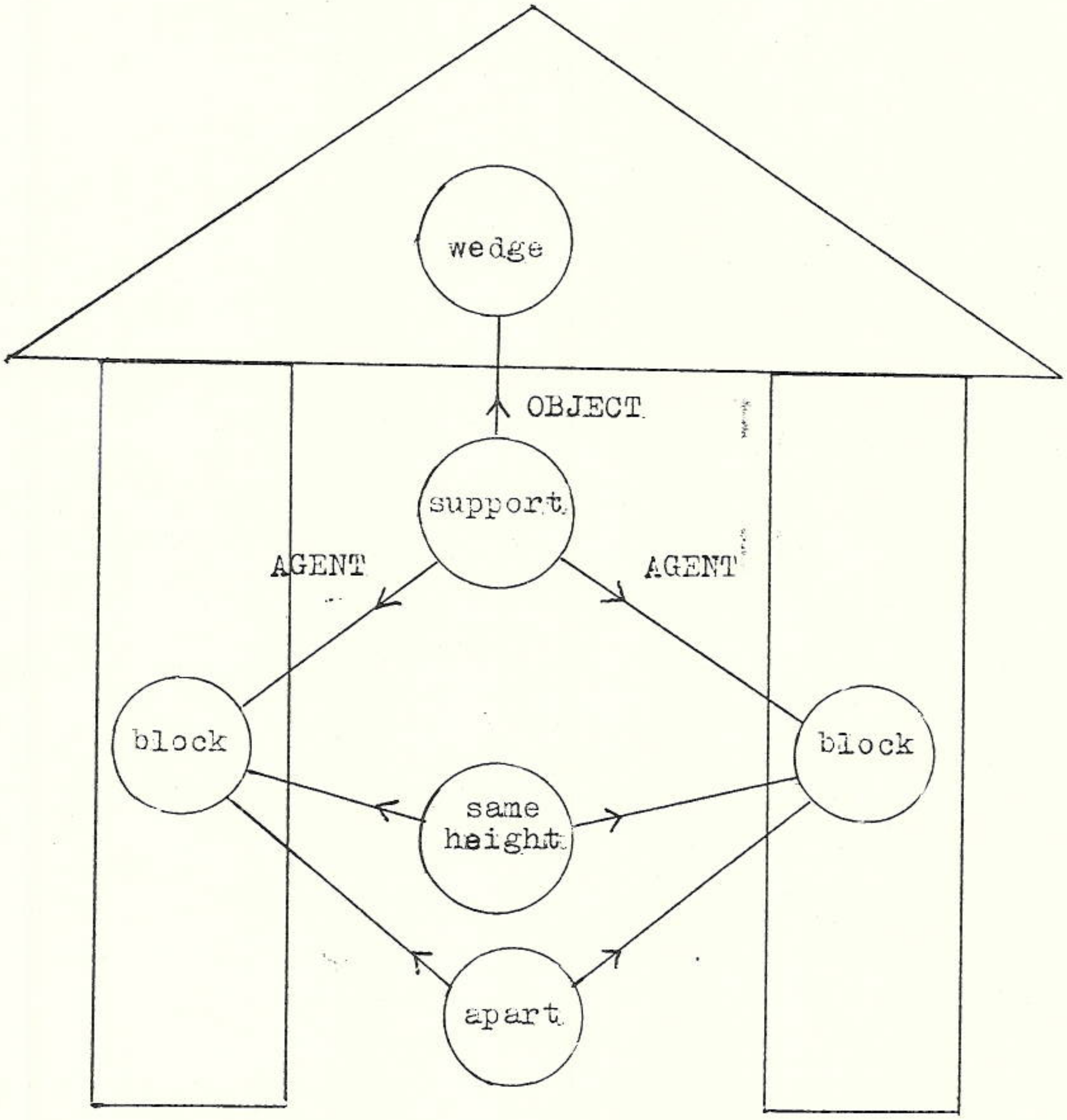

Fig. 6: A mixed pictorial/propositional Image. The network pattern is itself spatially laid out. There are no IS-A or PRED links, in the spirit of Section 4.4.

SECTION 6: IMAGINAL MATRICES IN THE BRAIN

This section describes the main claim of the paper in more detail.

Main Postulate

> The entire machinery that the brain has for the rapid internal manipulation of temporary data structures (in vision, natural-language understanding, problem solving, general thought processes) is an IM production system, where there are many IMs and they are all 2D and isomorphic. The system is subject to the constraints mentioned in the rest of this section. The Images can be pictorial, propositional, or elsewhere on the continuum of "diagrams". The Images have no necessary relationship to conscious imagery.

The Main Postulate is not to be taken to mean that all the sorts of Image suggested by analogy with the diagrams in Section 2 can be manipulated. That section served only to clarify the nature of the axis on which brain Images are conjectured to lie.

Implementational Postulate

a) Imagels (in the same or different IMs) are implemented as disjoint physical entities in the brain. Let us call these entities "physical imagels". Physical imagels may be disjoint neuron assemblies of some sort, but there are other possibilities. For each IF there is a different respect in which the physical imagel can be activated. The degree of activation implements the value of the IF at the imagel.

b) The basic pattern-recognition applied to IMs by the production system is achieved by distributed processing networks with a rough functional similarity to those discussed by by Hebb (1949), Hinton (1981), Feldman and Ballard (1981), Anderson et al (1977), and numerous others.

c) The abstract neighborhood of imagels within individual IMs is implemented in such a way that the basic pattern-recognition operations and the primitive Image-manipulation operations in Section 4.2 are efficient.

(It *may* well be that Part (c) is best satisfied by having abstract neighborhood of imagels implemented as proximity of physical imagels. It may even be that the physical imagels implementing an IM are arranged as a physical array, but this is by no means a necessary assumption and we shall not make it.)

Suppose we assume that highlighting is the standard technique used by productions for "accessing" patterns and regions. We similarly define a production to have "access" to an imagel group if that group is highlighted in a way detected by that production. Part (c) of the Implementational Postulate then immediately implies that if a production has access to an imagel group it has efficient access to the patterns supported by the imagels in the group. Part (c) also implies that if a production has access to an imagel then it can gain access to neighboring imagels (by effecting a spread of highlighting). Now recall the definition of spatial indexing in Section 1. We see that Part (c) of the

Implementational Postulate ensures that, if an IM is currently being interpreted on the basis that localized imagel groups correspond to spatial regions, then the IM together with its Image is a spatially indexed data structure. The spatial elements e are imagel groups, and the information I(e) takes the form of the pattern(s) supported by the imagels in group e.

An important consequence of part (b) is that ***not*** all the temporary data structures in the brain can conveniently be said to be Images: the pattern of excitation of the processing elements in the pattern-recognition mechanisms can be viewed as a temporary data structure lying outside IMs. However, it is legitimate to say that the data structures involved in the pattern-recognition mechanisms account for but a small proportion of all the temporary data structures manipulated in the brain, and that furthermore these data structures are of a highly specialized sort compared to the range of data structures which can appear as Images. Suggestions about what patterns in IMs are recognized by the basic pattern-recognition mechanisms are made below.

Section 6.1: General Nature of the Brain's IM System

We now place quite strong constraints on the sort of IM production system that the brain is conjectured to use. Some of the constraints were anticipated in previous sections.

Basic Parameters

The precise number of IMs, size of IMs, number of IFs, and value-ranges of IFs in the brain are parameters to the model schema being presented. As an initial guess, there are a few dozen IMs, an IM has size N by N where N is no more than a couple of hundred, and there are a dozen or two IFs. This paper makes the simplifying assumption that all IMs are isomorphic and use the same IFs*. The set of IMs is fixed during adult life.

No Addressing

There is no way of gaining access to an imagel or to an IM by means of a precise "address" which is some subImage or just the value of some imagel. The no-addressing restriction greatly influences the conjectures we can plausibly make about the nature of the Images used in the brain.

Spatial Tolerance

Consider two states of the whole IM set (that is, two possible sets of Images, one Image per IM) where the states only differ by small displacements and scalings of subImages such that the adjacency and containment relationships of patterns are maintained. Then the response of the production system to the two states is essentially the same.

---

* That is, for any two IMs there is a 1-1 correspondence between the IF sets such that corresponding IFs have the same value-ranges. We also assume there is an inter-IM pattern-copy operation which respects the IF correspondence.

For example, if all the Images were networks as in Section 4 and the corresponding Images in the two states only differed by small variations in the spatial positions of the nodes and links, then the manipulations the productions perform affect the networks in structurally the same way. Actually, in this example there is no necessary restriction to small differences between the Images (although only small scalings of node sizes and link thicknesses may be tolerable). If, however, the rough positions of nodes within an IM were important (cf. the hierarchy diagrams mentioned in Section 2) then large differences would not be tolerated.

In later more detailed models it may be appropriate to recast spatial tolerance as a mathematically rigorous form of continuity of the function from Images to production effects.

IF-Value Tolerance

A property similar to spatial tolerance is that two patterns which differ only by small variations in the IF values used have essentially the same effect on the production system. A related assumption is that individual IF values do not carry significant information. This is intended to exclude the possibility, say, that a set of data structures is given a numerical encoding, so that a single IF value could represent a complex piece of information. It is definitely the patterns of IF values which are intended to be the information bearers at the level of description used in this paper.

Images are the Essence

A strong assumption throughout this paper is that Images are ***essential*** to the computation in an IM system. That is, there is no other system of data structures of which Images are merely recodings and in which the real computation takes place. This assumption explains the intent of the term "entire" in the first sentence of the Main Postulate.

Simplicity and Locality of Productions

The productions are simple and local in the following sense. The condition parts mostly check for the presence or absence of simple combinations of patterns which are small compared to the size of an IM and which are recognizable by the basic pattern-recognition mechanisms. The testing of a condition part does not affect Images (except in so far as the processing of a condition part, or the selection of the action part for execution, may cause changes such as highlighting). The action part of a production is a simple sequence of primitive operations, with no branching or looping. (These primitive operations include input/output to other systems in the brain - e.g. vision preprocessors and motor systems - as well as Image-manipulation and pattern-recognition operations.) There are no local variables in productions – their effect must be achieved in other ways, such as by highlighting or by other placement of special marks in Images.

An exception to the locality of condition parts is that the manipulation of pictorial Images may require response to patterns convering a large region of an IM. It is also conjectured that such matters as parallelism of lines, symmetry of patterns, and repetition of patterns are aspects of Images which can be detected by the basic pattern-recognition mechanisms, even when the patterns involved in the symmetry and repetition cases cannot in themselves be recognized by those mechanisms.

Connexions between Productions and IMs

Productions are envisaged as being, by and large, tied to individual IMs, and productions tied to different IMs may generally execute in parallel. In implementation terms we therefore envisage a multiplication of mechanisms – e.g. in a network system each IM has its own mechanisms for detecting nodes. The tying of a production to a single IM means that its Image execution and manipulation is confined to that IM. We allow a relatively small number of productions to be permanently tied to more than one IM. Also, a production which is tied to a single IM may dynamically select another IM to serve as temporary storage. For this grabbing of storage to make sense, we attach to each IM an ***activation strength:*** an inspectable measure of how intensively it is currently in use. The lower the activation strength, the more likely the IM is to be grabbed. These measures are a small amount of data in excess of that contained in Images themselves, and therefore detract slightly from the "purity" of the IM system.

Distinguishability of IMs

It is possible for different IMs to be used in distinctly different ways and for their productions to be distinctly different. For example, different IMs might be used to hold very different types of network, requiring different sorts of processing. In such a system, therefore, the information provided by an Image (that is, the meaning of an Image) is not completely defined by the sort of pattern the Image is – we have to bring in the identity of the IM the Image occupies. Now, one could push the distinguishability of IMs to an absurd extreme, by having as many distinguishable IMs are there are different data structures that we want the system to manipulate! Let us say there are M such data structures. Let each IM consist of a single imagel with a single IF with two possible values, 0 and 1. Let there be just one production working on each IM; this production is triggered by its IM being "on" (that is, the IF value is 1) and acts by switching its IM off and switching some IM on. Clearly, such a degenerate IM system violates the spirit of IM systems. A less degenerate IM system is obtained by having only $\log_2 M$ switch-like IMs, allowing any number of them to be on at once, and allowing productions which can detect the state of an arbitrary number of switches. There are then M possible states of the whole IM set. Such a system still violates the spirit of IM systems. (It should be noted that this system is similar to Hebb-like neural assembly systems.) The important point revealed, however, is that there is a continuum of possible types of IM system going from the degenerate sorts just mentioned down to the more genuine sorts discussed previously. Different systems on this continuum allow differing amounts of information to be encoded in the distinguishability of IMs. We assume that the brain's IM system encodes relatively little information by IM distinguishability. A small amount of such encoding is probably unavoidable in practice.

Activation Strengths of Productions

Another sort of non-Image data which we allow if not taken to extremes consists of the activation strengths of productions. We envisage that one of the things a production can do is to specify that a certain set of productions are to be made more active or less active. The condition parts of productions which are more active are more likely to be satisfied, other things being equal. This feature is included so that the production system can be in different processing modes at different times. The intent is, however, that these modes define what processing is to occur only in a gross sense, and are not used to impose fine control over production executions.

Creation and Deletion of Productions

It is not proposed that productions can create and delete productions. We do, however, allow the existence of gross mechanisms, over and above productions themselves, which can produce or destroy productions. For instance, there may exist a strength-decay mechanism and a stipulation that a production with strength zero is destroyed. There may be a mechanism which automatically generalizes sufficiently similar productions to produce a new, more general production. We shall not dwell on these possibilities in this paper.

In a number of proposed semantic network systems (e.g. Norman and Rumelhart (1975), Mylopoulos and Levesque (1979)) pices of network can be executed as programs. Thus there is nothing to prevent some of the important high-level processing in an IM system being defined by explicit Image information rather than by the nature of the productions. The productions would serve to interpret the explicit Image "programs". (Indeed, this is what happens in the Lisp-interpreter exercise mentioned in Section 4.2.) In a system containing executable Images, productions can of course create and destroy executable Images (which may themselves represent productions).

The Nature of the Brain's Images

It is not the purpose of this paper to fix the precise class of Images the brain's IM system can manipulate. We can, however, make some general, plausible conjectures:-

a) Images at many different points on the pictorial-spatial-propositional axis are extensively used in various different types of problem-solving. In particular, spatial analogue is extensively used.

b) In propositional (parts of) Images, the main ***structuring primitives*** are:-
   - adjacency of patterns (much as in the adjacency of a node and a link, or of a label and a node; we also include pattern containment as a form of adjacency)
   - linking of patterns by line-like or ribbon-like patterns (such as the links in network-Images)
   - pattern association within and between IMs (see Section 4.3).

c) The adjacency and linking in (b) are used, in particular, to associate pictorial subImages and propositional subImages.

d) Network Images which are in a gross sense similar to the diagrams of semantic networks in the literature are extensively used, as are networks which are partly pictorial or spatial. (The nature of labels in networks receives some discussion in Section 7.)

e) The patterns which can be recognized by the basic pattern-recognition mechanisms include: blobs, lines, circles, and other simple shapes (e.g. for nodes); outlines of and stylized sketches of commonplace physical objects*; and patterns used for labelling purposes.

---

* "Stylized sketches" can be patterns similar to the feature map patterns arising from visual perception of physical objects.

It would be wrong to imagine that the class of Images usable in the brain's IM system constitutes a language in any sense which implies that there is a precise semantics of it in terms of the world outside the particular brain concerned. There is indeed a precise semantics — the precise effect that the presence of Images has on the brain. If this is to serve any purpose, of course, we expect some correspondence between certain Images and the external world and we expect that the manipulations of those Images reflect possible changes in the world. However, it may well be that that correspondence and reflection is only approximate, incomplete and oversimplified. In sum, it is a mistake to assume that the data structures manipulated by the brain need ***represent*** anything except in a loose, approximate sense.

Long-term Storage of Images

IMs hold only short-term (very temporary) data structures. We assume that the activation strength of an IM decays very rapidly unless productions are continually at work on it. The question therefore arises of the form that long-term knowledge takes. Certainly, we already have a rough characterization of long-term knowledge: namely, as the set of productions. Here we propose that there exist also a dynamically-changing, very large population of ***"long-term IMs"*** (LTIMs) whose function is to hold Images for periods much longer than is possible in the short-term IMs (STIMs) which we have considered so far. The role of LTIMs is roughly described by the following postulates.

(1) STIMs are implemented in such a way that Image interpretation and manipulation is efficient but the brain does not have enough physiological resources to contain enough STIMs for long-term storage purposes. LTIMs are implemented in such a way that enough of them can be present, but Image interpretation and manipulation within them is impractical.

(2) LTIMs can be permamently grouped together, to allow storage of data structures which do not fit into one LTIM.

(3) Some LTIMs are permanently associated with productions. A production can cause the Image in an LTIM permanently associated with the production to be copied into a specific region in an STIM.

(4) A production can cause an LTIM group to be created and some Images or specific subImages to be copied into the LTIMs in the group.

(5) Just as in the case of STIMs, each LTIM has an ***activation strength***. High-activation LTIMs continuously compete to have their Images transferred into low-activation STIMs. This competitive tranfer mechanism is, aside from the possibility of productions changing the activation strengths of STIMs and of LTIMs, independent of and concurrent with the workings of the production system.

(6) A production cannot change the Image in an existing LTIM. It is possible that LTIMs are occasionally destroyed by mechanisms independent of the production system (e.g. decay mechanisms).

(7) There is an LTIM/STIM counterpart to the inter-STIM pattern-association primitive: namely, when an STIM and an LTIM contain sufficiently similar patterns (in sufficiently similar contexts) and the LTIM's activation-strength is sufficiently high, the activation strength of the LTIM may be boosted (thus making it more likely that the LTIM's Image will be loaded into an STIM). This association primitive

probably works much more slowly than does the inter-STIM association primitive.

We need not commit ourselves to the presence of a direct inter-LTIM pattern-association primitive of any sort, although it would be advantageous. If such a primitive is not present, indirect inter-LTIM association can occur via the STIM-LTIM pattern-association primitive together with the inter-STIM pattern-association primitive.

If the brain does manipulate network-like Images, then only temporary networks (e.g. parts encoding viewed scenes or natural-language input) can lie in STIMs, since the Images in STIMs decay very quickly unless continuously operated upon. However, LTIMs can include network-like Images, and these Images can be "linked" to nets in STIMs by virtue of (3) and (7). For example, a node label can be detected by a production which then (as in (3)) loads into an STIM some permanent piece of net which holds information which is meant to be keyed by the label.

Section 6.2: The Link to Perception

We now turn to question (Q2) posed at the beginning of the Introduction:

(Q2) How does the manipulation of data structures derived from perceptual input fit in with the brain's data structure manipulation generally?

The data structures derived directly from visual input are the feature maps, as proposed in The Feature Map Hypothesis for Vision in the Introduction. It should come as no surprise (especially in view of the discussion in Section 3) that we now propose that feature maps are just pictorial Images with a special origin.

Vision Postulate

During visual perception, some of the brain's STIMs are the receptors of retinal stimulation which has been preprocessed into feature maps. That is, Images which are feature maps are continually being generated from retinal stimulation. The examination, manipulation and transformation of feature map Images is but a special case of the general operations of the IM production system, and IMs which are used to hold feature maps can also be used for the manipulation of internally-generated Images (which need not be at all spatial).

It is still assumed that all STIMs (and now LTIMs) are isomorphic and use the same set of IFs. We must therefore suppose that there are at least as many IFs as are needed for the purposes of feature maps. In fact, there are probably several more. The IFs which are used in feature map Images to encode features of retinal stimulation do not thereby necessarily have a special interpretation in Images which are not feature maps. However, an internally-constructed pictorial Image depicting a situation S can be similar to a feature map Image resulting from looking at S.

The Vision Postulate explains two aspects of our model schema which may have been puzzling the reader: the insistence on 2D IMs and the insistence on a collection of IMs rather than a single much larger IM. The explanation is the desire to have a uniform system in which externally generated and internally generated Images are essentially of the same type. The sources of this desire are in turn explained in Section 7.

SECTION 7: DISCUSSION

The claim is that the IM-system proposal for the brain provides a workable answer to the three questions that we posed in the Introduction. Part of the purpose of this section is to argue that this answer is, moreover, a good, though as yet schematic, one. At the end of the section we mention a few putative objections to the model.

Much of our case rests on the elegance and economy resulting from (a) postulating a continuum of data structures which includes "pictorial images" and propositional data structures merely as extreme cases, while (b) reducing the problem of the implementation of the brain's temporary data structures (question Q1) to that of implementing certain simple pattern-holding media (the IMs) and certain simple, powerful and general primitive operations working on those media.

A very important part of (b) is that a solution is given to the problem of ***frequent, temporary, not-explicitly-foreseen, not-previously-encountered association of information***. This association of information should be one of the main concerns of any theory which purports to explain cognition in terms of data structure manipulation, and to explain data structure manipulation in terms of physiological mechanisms. In our model the association largely takes the form of

- the adjacency of patterns within IMs
- the association of patterns within and across IMs by means of the pattern-association primitive of Section 4.3.
- and the explicit linking of patterns within IMs (as in a network-Image system).

Even if other parts of the theory turn out to be wrong, the idea that association of information takes these forms in regular networks of imagel-like entities may still be important and substantially correct.

The postulate that IMs in the brain are 2D and that there is a multiplicity of them calls for some justification. Recall from the Introduction that the basic assumption underlying question Q2 (concerning the relationship of perception to cognition as a whole) was that perception consists of decoding feature maps. Let us accept that feature maps are Images in 2D IMs as defined in Section 3. Suppose now that we were to propose an IM system as the basis of non-perceptual cognition, but that we did not a priori commit ourselves to the system having only 2D IMs or to having a multiplicity of them. It would then, anyway, be a good move for the sake of a simple and uniform model to restrict initial attention to systems in which all IMs are alike (and are therefore 2D because of the feature map assumption) and in which there must be a multiplicity of IMs (because by the feature map assumption an IM is constrained in size). We are thus using a methodological appeal to simplicity to justify the adoption of many, 2D IMs, with feature maps just being Images with a special origin. Furthermore, the necessary interpretation and manipulation operations may be considerably simpler in the case of 2D IMs than in the case of higher-dimension ones.*

---

* Another possible justification is that it would surely be easier to explain the evolution of human cognition from perception in lower organisms if we accept the feature-map-oriented IMs of our model than if we adopt a more elaborate system.

Why is it that we propose that the brain uses Image notations at intermediate points on the pictorial-propositional axis? The first point to note is that we are virtually forced into such a proposal by the very vagueness of the terms "pictorial" and "propositional". There is no particular line we can reasonably draw near either end of the axis. This point follows from the considerations of Section 2. A more concrete argument is simply that intermediate notations are useful. Some earlier sections support this, but we can conveniently summarize some of the points here.

Suppose we do accept that the brain finds it useful to manipulate pictorial Images in the course of some types of problem solving, and to manipulate propositional Images in other types of problem solving. It is then virtually immediate that it will find it useful to manipulate Images which are "mixed" – partly pictorial and partly propositional. This is simply because in some problems aspects of a situation described in a single Image may best be described pictorially whereas other aspects may best be decribed propositionally. Also, suppose even that it superficially appears appropriate to use two separate Images, one pictorial and one propositional, to describe a physical situation. Recalling the discussion of blocks-world problems in Section 5, we see that the desirability of a simple means of associating parts of the two Images suggests that the propositional Image be laid out pictorially, and even that it be superimposed on the pictorial Image.

Recall that it is claimed that spatial indexing accounts for a lot of the usefulness of pictorial diagrams, and that there are forms of non-pictorial diagram which are also useful as a result of their spatial indexing (or perhaps analogue-spatial indexing in the terminology of Section 2). It follows by analogy that non-pictorial spatial Images and analogue-spatial Images will be useful in some types of problem solving.

The ability to plan optimal routes along paths in space can easily be envisaged to have evolved into an ability to plan searches in search spaces portrayed as network Images. We assume the network is laid out in such a way that the length of a link is approximately proportional to the cost of traversing the link. The path-finding techniques which were evolved for spatial tasks could be used to find a set of search paths which have a good chance of being of minimal cost. These paths could then be examined in more detail by non-spatial computation. This suggested planning technique is like one mentioned by Sloman (1972). An intersting toy example of its potential is provided by Amarel's (1968) demonstration that certain diagrams of the search spaces in generalized Cannibals and Missionaries problems rapidly suggest to a human viewer, by virtue of their spatial layout, that a solution does or does not exist and what rough solution strategy should be followed.

Venn diagrams provide a paradigmatic example of how spatial-analogue diagrams can facilitate simple logical deductions. It is conceivable that some such technique (perhaps using Images similar to diagrams of the "mental models" suggested by Johnson-Laird (1980)) is used in the brain's IM system.

A common theme in the examples presented is that manipulations which can be conjectured to have arisen originally for use on pictorial Images can also be used on other sorts of Image. For instance, the mechanisms which notice adjacency and containment of pictorial patterns in IMs (and certainly such noticing is important in pictorial simulation of world activity) can be deployed to notice adjacency and containment of non-pictorial patterns (and certainly such noticing is fundamental in the network-Image manipulations of Section 4). Such unity of operation is also possible at higher levels of

Image manipulation/interpretation. For instance, a production system segment which was capable of finding routes along paths over physical terrain could be deployed for finding routes through a propositional network-Image. On a more general plane, mechanisms which notice similarities and regularities in pictorial Images could be used for the same purpose in non-pictorial Images.

Language in Thought?

The Vision Postulate renders perfectly plausible the following conjecture, which may at first sight be startling: that ***"natural-language fragments could appear as data structures in internally-generated Images"***. Or, more precisely: internally generated Images can contain interpretable patterns which are similar to those appearing in feature maps when a person is looking at natural-language text. Such patterns could, for instance, act as labels in network Images. Given the Vision Postulate, it would be surprising if we ***forbad*** such use of "natural-language fragments", since of course the IM system of anyone who can read must be able to recognize and interpret natural-language fragments in feature maps. To be sure, the difficulty of manipulating an Image goes up with the complexity of the natural-language fragments within it, and the interpretation of those fragments may entail the manipulation of other data structures (in other IMs). However, assuming that in the normal literate adult the commoner words are recognized by the basic pattern-recognition mechanisms, fragments which are single words do not pose any major problem deriving specifically from the fact that they are words as opposed to some other sort of subImage. Labels in network Images have to be patterns of some sort, and they might just as well be "words" as anything else. In fact, the use of words constitutes a measure of economy, since words must be interpretable in any case for the purposes of reading. Also, translation of data structures to and from pure natural-language form for the purpose of external communication is likely to be simplified.

Of course, it is not being suggested that natural-language fragments appear in internally generated Images of adult illiterates or young children (except in that the fragments may happen to lie in feature maps which have been transferred to LTIMs and later read back into STIMs). Also, this paper refrains from speculating about how usual it is for natural-language fragments to be used in internally generated Images, even in the case of normal adult literates. It is certainly not being suggested that such things as network labels ***must*** be natural-language fragments.

The proposal that textual natural-language fragments can appear for general cognitive purposes is a version of the idea that language is used in a literal way in thought. However, our proposal only makes a statement about unconscious thought (recall the note about consciousness in the Main Postulate in Section 6). Also, the IM model's conjectured use of language is considerably more sophisticated and flexible than that envisaged by a proposal in which there are functional components of the brain which are constrained to manipulate only natural-language fragments. There is no reason, of course, why on occasion an Image could not be entirely made up of natural-language fragments. Finally, the possible use of "auditory" natural-language fragments is touched on briefly in the next Section.

Some Objections

We now turn to brief consideration of some arguments which might be levelled at the model. One we can reject very quickly is the suggestion that the IM model involves an infinite regression because Images must be "seen" by some homunculus whose cognitive abilities are presumably explained in terms of Images, ... The Images in our theory are not "seen" in any sense stronger than that in which the bit strings in a computer are "seen" by the hardware which interprets and manipulates them.

Four putative objections which the author intends to discuss elsewhere are:

(1) that the number of LTIMs needed puts an implausibly large information-storage load on the brain (this argument is linked to objections by Pylyshyn (1973) that it would be implausibly expensive for the brain to hold a large stock of visual images)

(2) that a memory based on LTIMs would result in unprincipled loss of information from long-term Images (again, this is based on objections by Pylyshyn (1973, 1978b), and also by Palmer (1975), to visual-imagery theories)

(3) that the fact that blindness does not appear to impede general cognitive abilities may prove to be fatal to the model, in view of the strong link in the model between visual processes and the rest of cognition

(4) and that if, as seems natural, there is in fact a fairly close link between the visual images which appear to be present in consciousness and the pictorial Images the brain is supposed to manipulate, then it remains to be explained why we cannot have conscious visual Images similar to the propositional Images using the brain's own private notation (if, indeed, we cannot).

We can make the following brief observations about these objections. Concerning (1), so little is known about the physiological information-storage methods and capacity of the brain that the objection is without sensible foundation. Concerning (2), it has not been adequately shown that forgetting does not involve unprincipled loss of information, and in any case that objection involves a totally unwarranted assumption that ***non***-image information items would themselves necessarily be implemented in such a way as to avoid unprincipled loss of information. Concerning (3), it is not clear to the author whether there exist forms of blindness whose nature would dictate, in a reasonable, detailed IM model, postulation of failure of some physiological machinery essential to the IM system; so it is difficult to assess the importance of this objection. And concerning (4), the problem is but our version of, and is no more puzzling than, a general unsolved problem for ***any*** theory which is based on the Data Structure Hypothesis and in which not all of the supposed data structures are manifested in some direct sense in conscious thoughts. The problem is precisely this: why ***are*** not all the data structures so manifested?

SECTION 8: CONCLUDING REMARKS

The IM model schema has been put forward as an answer to the three questions Q1 - Q3 posed in the introduction. Each of these questions was asked in the context of a certain strong assumption. Naturally, there is no implication that the model follows inescapably from these assumptions. Our answer is admittedly only a model schema, not a precise model, and does not of course answer the three questions to the ideal degree of detail. The schema has many parameters which will need to be given "values" for a precise model to be forthcoming and therefore for the results of psychological and physiological experiment to be relevant. Some of the parameters are:

- the size of IMs
- the number of short-term IMs
- the number of IFs and the nature of their value spaces
- the precise nature of the production system
- the speed of operation of the primitive operations
- the precise nature of the pattern-association mechanism
- the precise class of Images used in IMs
- a precise set of tasks for the system to perform.

Perhaps the first predictions which would be testable by psychological experiment would concern the capacity of short-term memory and the dependence of rates of processing on complexity of the data structures involved – this rate being partially governed by the size and number of short-term IMs, the extent to which data structures fit into single IMs, and the amount of spurious inter-IM pattern-association which interfers with the desired processing.

The issue of the precise physiological implementation of IMs and productions has been largely uncoupled from more abstract issues of the nature of the Images used and of their use. Therefore, we could proceed in the relatively near future to specify a simplified cognitive model which is precise in its abstract aspects but leaves the question of detailed physiological implementation open. Conversely, we could pursue a parallel investigation into how IMs and productions could be physiologically implemented (or, at least, implemented in terms of idealized physiological mechanisms, e.g. simple pseudo-neurons) so as to facilitate the sort of primitive operations we discussed in Section 4.2 and 4.3, while leaving open the question of whether these operations are ultimately destined to be the correct ones. We may therefore justifiably claim that the IM model schema is very rich in suggestions for further interesting and feasible research. This richness and detail constitutes in itself a major incentive for studyiong the IM model, and the research generated would probably remain of interest even if the schema were eventually shown to be erroneous. We may admit that the IM model is highly speculative, while remarking that the degree of speculation is ***no higher*** than that of previous attemts to describe the general information-processing machinery of the brain.

The author is embarking on a series of computer-simulation exercises whose aim is to show in detail how an IM system can perform useful computation. An early exercise (on which work has started) will be a computer simulation of an IM system which acts as a diagrammatic form of the production system language PSG (Newell (1973)). (The list structures which PSG productions work on will become abstract Images.) This exercise will do much to clarify the feasibility of the basic pattern-recognition and manipulation operations suggested in Section 4. Another rather similar exercise (which is also already under way) is the computer simulation of an IM implementation of the programming language Lisp. Aiming in a somewhat different direction, a later exercise will be to simulate an IM system capable of simple pictorial event simulation (manipulating pictures of physical situations to effect simple problem solving). A succeeding exercise will add limited abstract annotational elements to the pictures. The last of the currently envisaged series of exercises will be the simulation of an IM system which manipulates a certain network-Image notation and can cope with the inclusion of limited pictorial elements.

The claim that the schema will fruitfully repay further stduy is a methodological justification for it. The other main justification is again a methodological one: thet the deployment of the basically simple idea that "the brain's data structures are diagrams" simultaneously does all of the following. First, it allows "pictorial images" and propositional data structures merely to be different points on a unified notational axis which also includes interesting forms of hybrid notation. Second, it is the start of a physiological account of how pieces of information (in the form of subImages) can be brought into purely temporary and unforeseen association. This is a fundamental capability of computer-implemented information-processing systems, and has been transported to a brain model without the need to transport the (numerical) location addressing used in conventional computers. Third, the schema has the desirable property that it incorporates the perceptual feature maps in a principled (not ad hoc) way.

The schema postulates that most of the data manipulated in cognition is encoded in patterns in IMs. As we observed, however, some information is contained in other places and therefore detracts from the "purity" of the IM system. The main impurities are: that the basic (perhaps Hebb-like) pattern-recognition mechanisms acting on IMs can be seen as manipulating data structures; that the identity of an IM in which a given Image resides may be important in the case of some IMs; and that IMs and productions have activation strengths. It is claimed that these forms of data are ***merely*** impurities because they are relatively uninformative compared to the Images themselves.

Although the precise nature of the Images manipulated in a brain is not specified by the model schema, we have conjectured that some of them are ***broadly*** similar to the network diagrams one finds in the literature. An interesting consequence of this is that it elevates net diagrams from the status of illustrations for the benefit of network researchers to the status of bona fide data structures. We could even conjecture that part of the reason we find network diagrams useful externally is precisely that they are reasonably similar to network "diagrams" used internally, so that mechanisms which interpret network Images can also be deployed in the interpretation of feature maps arising from perception of external network diagrams. We might go on from here to suggest that our fondness for many other particular sorts of external notation (at various points on the notational axis) is to be explained partly by a broad similarity to Image notation used internally. It is far too early to evaluate the worth of these conjectures, and their falsity would not disconfirm the IM model schema. One point to observe is that the sorts of external notation we use may influence the sorts of Image we manipulate. Exposure to a particular form of graphic notation could train the IM system in the manipulation of similar notation in Images. Indeed, the suggested use of natural-language fragments is a case in point.

There are major topics on which this paper has been silent but which will have to be tackled eventually. Some of the most important of these are: the detailed use of IMs in the understanding of spoken and written language; the inclusion of mechanisms to organize motor activity; and the detailed use of IMs in perception. The Vision Postulate is but a special case of a braoder but more tentative idea about the incorporation of perception as a whole. This broader conception has it that auditory and tactile perception, and some proprioception, also operate by preprocessing receptor stimulation into Images (the IMs associated with different modalities being non-isomorphic) and that these IMs are as much an integral part of the whole IM system as are the vision-oriented IMs of this paper. ad This conjecture generates an expansion of the class of Images plausibly to be suggested as occurring in the brain. In particular, it becomes possible to suggest the use of auditory natural-language fragments as well as graphic natural-language fragments in Images, thus getting over the problem that graphic ones are only functional in literate brains.

The radical and speculative nature of the model should not blind us to the fact that the computational mechanisms proposed for the brain are but a natural extrapolation from those used in computers. Instead of a linear store (of locations) we have a group of 2D stores (made up of imagels); instead of bit strings in locations we have IF values at imagels; instead of the association of information by location adjacency, location addressing and content addressing (bit-string association), we have association of information by imagel and pattern adjacency, linking by line-like patterns, and "pattern-association"; and instead of the basic method of decoding the bit-string from a single location, we have the basic, sustantially local pattern-recognition mechanisms. This computer analogy, together with the current trend in hardware development, then renders it not inconceivable that artificial IM systems implemented on special parallel hardware will some day be of use.

REFERENCES

Amarel, S. On representations of problems of reasoning about actions. In D. Michie (Ed.), ***Machine Intelligence 3,*** Edinburgh: Edinburgh University Press, 1968.

Anderson, J.A., Silverstein, J.W., Ritz, S.A. and Jones R.S. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. ***Psychological Review,*** 1977, *84, 413-451.*

Anderson, J.R. Arguments concerning representations for mental imagery. ***Psychological Review,*** 1978, *85,* 249-277.

Anderson, J.R. Further arguments concerning representations for mental imagery: A response to Hayes-Roth and Pylyshyn. ***Psychological Review,*** 1979, *4,* 395-406.

Anderson, J.R. and Bower, G.H. ***Human Associative Memory.*** Washington, D.C.: Winston, 1973.

Barnden, J.A. The integrated implementation of imaginal and propositional data structures in the brain. ***Procs. 4th. Cognitive Science Conf.,*** Ann Arbor, Michigan, 1982.

Barrow, H. and Tenenbaum, J.M. Recording intrinsic scene characteristics from images. In Hanson and Riseman (1978).

Baylor, G.W. Program and protocol analysis on a mental imagery task. ***Procs. 2nd. Int. Joint Conf. on Artificial Intelligence,*** London, 1971.

Block, N. (Ed). ***Imagery.*** Cambridge, MA: Bradford Books, MIT Press, 1981.

Bobrow, D.G. Dimensions of representation. In D.G. Bobrow and A. Collins (Eds.), ***Representation and understanding.*** New York: Academic Press, 1975.

Bower, G.H. Mental imagery and associative learning. In L.W. Gregg (Ed.), ***Cognition in learning and memory.*** New York: Wiley, 1972.

Brachman, R.J. On the epistemological status of semantic networks. In Findler (1979).

Chase, W.G. and Clark, H.H. Mental operations in the comparison of sentences and pictures. In L.W. Gregg (Ed.), ***Cognition in learning and memory.*** New York: Wiley, 1972.

Chase, W.G. and Simon, H.A. The mind's eye in chess. In W.G. Chase (Ed.), ***Visual information processing.*** New York: Academic Press, 1973.

Cooper, L.A. and Shepard, R.N. Chronometric studies of the rotation of mental images. In W.G. Chase (Ed.), ***Visual information processing.*** New York: Academic Press, 1973.

Davis, R. and King, J. An overview of production systems. In E. Elcock and D. Michie (Eds.), ***Machine Intelligence 8,*** Chichester: Ellis Horwood, 1977.

Dennett, D.C. ***Brainstorms: Philosophical essays on mind and psychology.*** Cambridge, MA: MIT Press, 1981.

Feldman, J.A. and Ballard, D.H. Computing with connections. Technical Report TR72, Computer Science Dept., University of Rochester, New York, April 1981.

Findler, N.V. (Ed). ***Associative networks.*** New York: Academic Press, 1979.

Fodor, J.A. ***The language of thought.*** New York: Crowell, 1975.

Fodor, J.A. ***Representations.*** Cambridge, MA: Bradford Books, MIT Press, 1981.

Funt, B.V. WHISPER: A problem-solving system utilizing diagrams and a parallel processing retina. In ***Procs. 5th. Int. Joint Conf. on Artificial Intelligence,*** Cambridge, MA, 1977.

Goodman, N. ***Languages of art: An approach to a theory of symbols.*** Indianapolis: Bobbs-Merrill, 1968.

Hanson, A.R. and Riseman, E.M. (Eds). ***Computer vision systems.*** New York: Academic Press, 1978.

Hayes, J.R. On the function of visual imagery in elementary mathematics. In W.G. Chase (Ed.), ***Visual information processing.*** New York: Academic Press, 1973.

Hayes, P.J. Some problems and non-problems in representation theory. ***Procs. Summer Conf. of the Society for the Study of Artificial Intelligence and Simulation of Behaviour,*** University of Sussex, July, 1974.

Hayes-Roth, F. Distinguishing theories of representation: A critique of Anderson's "Arguments Concerning Mental Imagery". ***Psychological Review,*** 1979, ***86,*** 376-382.

Hebb, D.O. ***Organization of behaviour.*** New York: Wiley, 1949.

Hebb, D.O. Concerning imagery. In J.M. Nicholas (Ed.), ***Images, perception, and knowledge.*** Boston: Reidel, 1977.

Hinton, G.E. Some demonstrations of the effects of structural descriptions in imagery. ***Cognitive Science,*** 1979, ***3,*** 231-250.

Hinton, G.E. Shape representation in parallel systems. In ***Procs. 7th. Int. Joint Conf. on Artificial Intelligence,*** Vancouver, 1981.

Hubel, D.H. and Wiesel, T.N. Brain mechanisms in vision. ***Scientific American,*** 1979, ***241,*** 150-163.

Johnson-Laird, P.N. Mental models in cognitive science. ***Cognitive Science,*** 1980, ***4,*** 71-115.

Kieras, D. Beyond pictures and words: Alternative information-processing models for imagery effects in verbal memory. ***Psychological Bulletin,*** 1978, ***85,*** 532-554.

Kolers, P.A. and Smythe, P.E. Images, synbols and skills. ***Canad. J. Psychol.,*** 1979, ***33,*** 158-184.

Kosslyn, S.M. The medium and the message in mental imagery: A theory. In Block (1981).

Kosslyn, S.M., Pinker, S., Smith, G.E. and Schwartz, S.P. On the demystification of mental imagery. In ***The Behavioral and Brain Sciences,*** Vol. 2, Cambridge University Press, 1979. Also in Block (1981).

Kosslyn, S.M. and Schwartz, S.P. A simulation of visual imagery. ***Cognitive Science,*** 1977, ***1,*** 265-295.

Marr, D. Early processing of visual information. ***Phil. Trans. Royal Soc.,*** 1976, ***275,*** No. B942.

Minsky, M. A framework for representing knowledge. In P.H. Winston (Ed.), ***The psychology of computer vision.*** New York: McGraw-Hill, 1975.

Mylopoulos, J. and Levesque, H. A procedural semantics for semantic networks. In Findler (1979).

Neisser, U. ***Cognition and reality.*** San Francisco: Freeman, 1976.

Newell, A. Production systems: models of control structures. In W.G. Chase (Ed.), ***Visual information processing.*** New York: Academic Press, 1973.

Newell, A. and Simon, H.A. ***Human problem solving.*** Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Norman, D.A., Rumelhart. D.E. and the LNR Research Group. ***Explorations in cognition.*** San Francisco: Freeman, 1975.

Paivio, A. ***Imagery and verbal processes.*** New York: Holt, Rinehart and Winston, 1971.

Paivio, A. Images, propositions and knowledge. In J.M. Nicholas (Ed.), ***Images,*** *perception, and knowledge.* Boston: Reidel, 1977.

Paivio, A. On weighing things in your mind. In P.W. Jusczyk and R.M. Klein (Eds.), ***The nature of thought: Essays in honor of D.O. Hebb.*** Hillsdale, N.J.: Lawrence Erlbaum, 1980.

Palmer, S.E. Visual perception and world knowledge. In Norman and Rumelhart (1975).

Palmer, S.E. Fundamental aspects of cognitive representation. In E. Rosch and B.B. LLoyd (Eds.), ***Cognition and categorization.*** Hillsdale, N.J.: Lawrence Erlbaum, 1978.

Peterson, M.J., Peterson, L.A. and Ward-Hull, C. Rehearsing and maintaining mental matrices. ***J. Verbal Learning and Verbal Behavior,*** 1977, ***16,*** 371-381.

Peterson, M.J., Thomas, J.E. and Johnson, H. Imagery, rehearsal, and the compatibility of input-output tasks. ***Memory and Cognition,*** 1977, ***5,*** 415-422.

Posner, M.I. Coordination of internal codes. In W.G. Chase (Ed.), ***Visual information processing.*** New York: Academic Press, 1973.

Pribram, K.H. ***Languages of the brain.*** Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Pylyshyn, Z.W. What the mind's eye tells the mind's brain: A critique of mental imagery. ***Psychological Bulletin,*** 1973, ***80,*** 1-24.

Pylyshyn, Z.W. What has language to do with perception? Some speculations on the Lingua Mentis. In ***Procs. 2nd. Workshop on Theor. Issues in Natural-Language Processing,*** 1978a.

Pylyshyn, Z.W. Imagery and Artificial Intelligence. In C.W. Savage (Ed.), ***Minnesota studies in the philosophy of science,*** Vol. IX, 1978b.

Pylyshyn, Z.W. Validating computational models: A critique of Anderson's indeterminacy of representation claim. ***Psychological Review,*** 1979, ***86,*** 383-394.

Pylyshyn, Z.W. The imagery debate: Analog media versus tacit knowledge. In Block (1981).

Raphael, B. ***The thinking computer.*** San Francisco: Freeman, 1976.

Richardson, A. ***Mental imagery.*** New York: Springer, 1969.

Schwartz, R. Imagery – there's more to it than meets the eye. In Block (1981).

Simon, H.A. What is visual imagery? An information-processing interpretation. In

L.W. Gregg (Ed.), ***Cognition in learning and memory.*** New York: Wiley, 1972.

Simon, H.A. On the forms of mental representation. In C.W. Savage (Ed.), ***Minnesota studies in the philosophy of science,*** Vol. IX, 1978.

Sloman, A. Interactions between philosophy and artificial intelligence. ***Artificial Intelligence,*** 1971, ***2,*** 209-225.

Sloman, A. Analogical representations. ***AISB European Newsletter 19,*** March 1975.

Waltz, D.L. Understanding scene descriptions as event simulations. In ***Procs. 18th. Meeting of the Association for Computational Linguistics,*** 1980.

Waterman, D.A. and Hayes-Roth, F. ***Pattern-directed inference systems.*** New York: Academic Press, 1978.

Winston, P.H. ***Artificial Intelligence.*** Reading, MA: Addison-Wesley, 1977.

Woods, W.A. What's in a link: Foundations for semantic networks. In D.G. Bobrow and A. Collins (Eds.), ***Representation and understanding.*** New York: Academic Press, 1975.