Integrating Imagery and Visual Representations *

B. CHANDRASEKARAN AND N. H. NARAYANAN
Laboratory for Artificial Intelligence Research
Department of Computer and Information Science
The Ohio State University
Columbus, OH 43210

The issue of propositional versus analogic representations for visual information has been debated extensively in cognitive psychology. In this paper we argue that issues arising from this debate can be effectively addressed by postulating a common mechanism underlying visual processing and mental imagery which, while representing information symbolically, can manipulate visual aspects of perceived objects in imaginal forms. The central components of this mechanism are representations resulting from visual perception, called visual representations, and a special-purpose architecture specific to the visual modality. The experience of mental imagery arises from the interpretation of visual representations by the modality-specific architecture. This architecture also allows operations specific to the visual modality to be performed on visual representations. Thus, the modality-specific architecture is the key that integrates the experience of analogic imagery and symbolic visual representations. We argue that there is no real opposition between a belief in the need for some form of analogic phenomena to explain imagery and a belief in the realization of such phenomena from symbolic structures, and that a middle course can be charted between the purely analogic and the purely propositional views while retaining the advantages of both.

1. INTRODUCTION

The representation of visual information and its relation to the phenomenon of mental imagery have attracted a great deal of attention from cognitive psychologists. There has been considerable debate (Anderson, 1978; Kosslyn & Pomerantz, 1977; Kosslyn, 1981; Pylyshyn, 1981) about postulating separate analogic representations for imagery as opposed to uniform propositional representations. The representation of visual information must be "picture-like", in some sense of the term, to one camp because they see evidence that some aspects of human reasoning are better explained by exploiting special properties of this modality. Scanning, relative distance estimation, and direction finding are examples of mental actions, explanations of which benefit from postulating a pictorial representation. For researchers subscribing to this view, what is important is that any ultimate proposal for mental representations should be able to explain this differential use of modality-specific operations. On the other side there are people who view propositions as the basic currency of mental representations. They believe that the use of propositional representations unite conceptual and perceptual processes. They view the advocacy of image-like representations as a rejection of their claims about the generality of propositional representations. They argue that the use of visual operations can be explained by propositional representations and processes as well. Many such arguments are given in (Anderson, 1978; Pylyshyn, 1981).

This is the dichotomy between the positions taken by proponents of analogic and propositional representations. But are these two views really mutually exclusive contenders for the position of the "ultimate" theory of visual representations? Maybe what is missing is the realization that a mechanism that preserves the significant properties of both analogic and propositional representations can account for the supportive experimental evidence presented by researchers on both sides of the issue. This is the insight that led to the ideas presented in the rest of this paper.

---

* Ideas in this paper have benefited from our discussions with Stephen Kosslyn. This research has been supported by DARPA & AFOSR contract # F-49620-89-C-0110 and by AFOSR grant # 890250. First author's email address is chandra@cis.ohio-state.edu.

2. TYPES OF VISUAL REPRESENTATIONS

In the literature, representations that result from actually perceiving objects are called perceptual representations and representations that underlie mental imagery are called imaginal representations. We refer to these collectively as visual representations in this paper. In this section three types of visual representations are described.

The analogic-propositional debate centered around two kinds of representations. One is called an analogic representation and the motivation for it stemmed from the phenomenon of mental imagery. While few deny the existence of mental imagery, the hypothesis that analogic representations underlie mental imagery has been questioned. Those who subscribe to this hypothesis, whom we call analogists, are impressed by what they see as evidence for subjects' preferential use of pictorial operations (e.g: scanning) on mental imagery and to account for this they ascribe to this representation certain properties that are usually associated with pictures. Such operations have a special status in analog representational theories, but not in propositional theories. Kosslyn (1981) presents a concrete explication of an analogic theory of visual representation. He proposes two levels of representations: one "deep" representation that is abstract and not experienced directly and a "surface" representation that supports mental imagery. A surface representation is pictorial in nature since it depicts an object by regions of activation in the visual buffer which is an analog medium. This two-level view of imagery (i.e., surface displays generated by comparison and/or transformation processes from a deep structure) is used in (Kosslyn & Schwartz, 1977; Kosslyn & Pomerantz, 1977; Pinker, 1980) as well.

The other type of representation is called propositional (and its advocates, propositionalists). It is made up of propositions. A proposition consists of symbols, but it is more than a mere collection of symbols. Propositions have identifiable predicate and argument constituents, bear truth values, and have rules of formation (Anderson, 1978). Thus a proposition has both a fixed syntactic form and a fixed semantic content. In the propositional view of mental representations, propositions encode knowledge about objects in a perceived scene and require interpretation for their semantic contents to be accessed by processes operating on them. Propositional theories have an underlying implication that the rules which govern how propositions are interpreted are independent of the perceptual or cognitive modality that produced the propositions. In other words, a general purpose mechanism that is not modality-specific is assumed to operate on propositional representations. For example, according to propositionalists, the propositions (*left-of A B*), (*smell-of C pungent*) and (*rate-of inflation high*) will be handled by mental processes in a uniform manner that does not reflect the distinction that one describes a visual attribute, another an olfactory attribute, and the third is a conceptual assertion. Thus the intrinsic property of propositional representations is their uniformity of representation and processing across modalities or "faculties".

There is a third type of representation, called a discrete symbolic representation. A discrete symbolic representation comprises structures of discrete or atomic symbols composed according to well-defined rules of formation. A symbol is just a token and, by itself, is devoid of any meaning. Its semantics derives from the *architecture* of the system it resides in and from *how* it is used by processes operating on it. Therefore it is conceivable that the same symbol may be interpreted differently by different architectures and processes. This is an important distinction. Consider the symbol "car"[1] appearing as the first element of a list. In an architecture designed specifically to execute Lisp programs, this symbol will trigger a process that extracts the first element of a list. In a different architecture the same symbol may have a different meaning. The meaning of a program (a program is nothing but a discrete symbol structure) is given by its operational semantics which specifies the operations performed and their effects on the inputs (which are also discrete symbol structures) and the compiler enforces this semantics. We propose

[1] "Car" is a Lisp command that extracts the first element of its argument, which should be a list.

that the discrete symbolic representation in the general sense is also a contender for the mental representation of visual information. The operational semantics of such a visual representation specifies the operations allowed by the visual modality-specific architecture and their effects on the representation and the architecture enforces this semantics.

Both propositional and discrete symbolic representations are made up of symbols and composed according to specific rules of formation. However, unlike propositional representations, a discrete symbolic representation does not necessarily need to have a truth value or have predicates and arguments as constituents. In other words, a proposition is a special type of discrete symbolic structure whose operational semantics is truth-preserving[2]. Thus discrete symbolic representations are more general than propositional representations. The only commitment entailed by the discrete symbolic representation is that it is composed, in a principled manner, of discrete symbols and interpreted consistently by the underlying architecture and processes operating on the representation. While the discrete symbolic representation has a fixed syntactic form, its semantic content is defined relative to the underlying architecture and the processes that operate on it. Properties exhibited by such a representation are not intrinsic to the representation itself, but stem from mechanisms (which may well be sensory-modality specific; this is one of the claims made later in this paper) that support and operate on the representation.

3. THE IMAGE REPRESENTATIONAL SYSTEM

We referred before to a debate in psychology between analogists and propositionalists about the nature of visual representations. The debate is often stated in terms of several "versus" formulations:

1. Analog versus symbolic: Some picture-like analog representations versus discrete symbolic representations.
2. Analog versus propositional: Picture-like representations versus propositional representations.
3. Sensory modality-specific representations and interpreters versus sensory modality-independent representations and a uniform interpreter.

In the literature on this debate, discrete symbolic representations are equated with propositional representations, i.e., 1 and 2 above are generally supposed to make the same distinctions. However, as we just discussed, propositional representations are a special form of discrete symbolic representations. Additionally, because of the uniformity of interpretation that applies to propositional representations, propositionalists are necessarily led to a sensory modality-independent view of representation. To the extent that analogic representations involve interpretations that give a privileged status to some operations over others depending on the sensory modality, there is a genuine opposition between analogists and propositionalists.

The burden of this paper is that, on the contrary, there is no real opposition between a belief in the need for some form of analogic representations and a belief in the realization of such a representation from symbolic structures. The facile equation of propositional representations and symbolic representations has, in our view, set up a false opposition in 1 above. The root of this misconception arises from the assumption that symbolic representations necessarily need to run on general purpose computers. We argue that in fact one can have analogic, i.e., visual modality-specific, representations which are also in principle symbol structures. The key to this possibility is that representations for different sensory modalities are run on interpreters which provide privileged operations that are specific to that modality. Of course, these representations, to the extent possible and relevant, need to be coordinated with corresponding representations in

---

2 Note that logic has two distinct uses. One is as a representational language and the other as a metalanguage in which a representational theory can be described and analyzed. Logic as a programming language is an example of the former and logic specification of program semantics is an example of the latter. Here we are concerned only with the former use of logic in propositional representations.

other sensory modalities as well as the general cognitive architecture.

We propose that properties exhibited by visual representations stem from an underlying mechanism, which we call the Image Representational System (IRS). The IRS contains a special purpose architecture that is specific to the visual modality. Visual representations reside in this architecture and the architecture interprets the representations in a way that allows visual operations (e.g., scanning, relative position estimation etc.) to be performed on them. It is this interpretation that gives rise to mental imagery. Thus the analog nature of mental imagery arises from the interaction between visual representations and the visual modality-specific architecture.

The central components of the IRS are the visual representations that result from perception, called Image Symbol Structures (ISS), the underlying specialized (to the visual modality) architecture in which these structures reside, the interpretation of ISS (which gives rise to mental imagery) that the architecture produces, and the visual operations that it provides. An ISS is a hierarchical discrete symbolic[3] representation, similar in spirit to the 3-D sketch of Marr and Nishihara (1978). It is the end product of visual processing that starts from the retinal image.

The ISS has both syntactic form and semantic content. When an ISS is interpreted by the architecture, its semantic content can be experienced as mental imagery. These interpretations as well as the ISS itself function as inputs to high level visual processes for object recognition etc. When an object represented in the ISS is recognized and labelled, that facilitates the evocation of non-visual (conceptual) knowledge about it. The visual modality-specific architecture of the IRS also provides basic operations such as scanning on the ISS. These have concomitant effects on its interpretation as well. Visual processes can manipulate the Image Symbol Structures by invoking these operations. Fig. 1 illustrates[4] this role of the IRS. A mechanism like the IRS can account for the preferential use of visual operations on mental images based on underlying non-analogic representations.

4. THE IMAGE SYMBOL STRUCTURE

Visual perception, according to the theory of Marr and Nishihara (1978), consists of the transformation of the primal sketch obtained from the retinal image into a $2^1/_2$-D sketch and then into a 3-D sketch which feeds into shape and object recognition processes. The most interesting aspect of this theory is its use of parametrized volumetric primitives in the 3-D sketch. Our conception of the Image Symbol Structure has been inspired by Marr's theory. The ISS is defined to be the internal representation of a perceived real world scene. It is a hierarchical compound structure made up of primitives.

Primitives of shape, texture, color and other visually perceivable attributes are assumed to be available to the IRS. Each primitive is parametrized. For example, a shape primitive may have its relevant dimensions as parameters. These parameters are not absolute measurements, but have relative meaning within an internal reference frame. That is, they serve as yardsticks that facilitate attribute value comparisons with other primitives of the same type.

The ISS is structured as a hierarchy of descriptions with levels that decrease in grain size,

---

[3] Quite independent of the debate about visual representations being analogic or propositional, there is another ongoing debate about mental representations in general being symbolic versus connectionist. In this paper we argue that visual representations can be discrete symbolic while preserving analogic properties. This argument can be extended to the connectionist position as well. The aspects relevant to our proposal are the informational content of the representations (Chandrasekaran, Goel, & Allemang, 1988), how this visual information is organized (e.g., structured as hierarchical composites built up from primitives of visual attributes, see the following section), and how the representations get interpreted by the architecture.

[4] A similar depiction was originally suggested to us by Bruce Flinchbaugh.

or alternately, increase in resolution. The description at each level is made up of appropriately parametrized primitives corresponding to objects delineative at that level's resolution. The topmost level describes the image coarsely while the lowest level describes it in terms of the finest details captured during perception. At each intermediate level more details get added to the descriptions of image components from the level above. The ISS encodes only the intrinsically visual (and therefore internally visualizable) aspects of a scene. Non-visualizable aspects (for example, knowledge about the weight of an object) are part of the conceptual knowledge associated with the scene and not part of the ISS. Also, the ISS is neutral with respect to recognition; it represents objects in terms of visual attributes, but does not "name" or label the objects that it represents.

We term those parts of an ISS that together correspond to an object or a delineative part of an object, an S-percept ("symbolic percept"). An ISS is thus made up of multiple S-percepts and an S-percept may itself be composed of other S-percepts. It is essentially a description that contains (only) visual aspects of a delineative object or its part, in terms of parametrized primitives. For example the S-percept corresponding to an apple will consist of primitives that describe its shape, color, shiny texture etc., but not its taste or nutritional value. In addition to parametrized primitives an S-percept also contains descriptions of spatial relations among the primitives.

Each S-percept has a corresponding mental image that results from its interpretation. This entity is called an A-percept ("analogic percept"). An A-percept exhibits all visual attributes and spatial orientations described by the corresponding S-percept. For instance, an A-percept corresponding to an S-percept of the form {cylinder (diameter:2, length:4), color(red)} will be the mental image of a red cylinder of diameter 2 units and length 4 units. Thus the S-percept of an object is a symbolic entity whereas the A-percept of the object is an imaginal entity, and the two may be thought of as two sides of the same coin.

Thus the ISS may be viewed as an internal representation of a perceived scene that functions as a symbolic description as well as an algorithm for the composition of a mental image by the visual modality-specific architecture. The following analogy should explicate how ISS, S-percepts, and A-percepts are related. Consider a robot standing beside bins containing cubes, cylinders and other geometric objects of various dimensions and colors. Assume that it is possible to write a description of any structure made up of these geometric objects in an abstract language that the robot can interpret. Upon loading such a description into the robot it is capable of picking up objects of appropriate shape, size and color and building that structure. Conversely, if a structure made up of the geometric objects is provided, the robot can produce a description of that structure in the abstract language. Assume that if some component in this structure (or the abstract description) is removed or replaced with another, it is possible for the robot to sense the change and correspondingly change the abstract description (or the structure). This situation is similar to the mechanism comprising IRS and ISS. The robot is analogous to the special architecture of the IRS, the abstract description is analogous to an ISS, parts of the abstract description are analogous to S-percepts, the geometric objects are analogous to A-percepts, and the structure built by the robot is analogous to the interpretation of an ISS.

Operations performed on S-percepts (A-percepts) have concomitant effects on A-percepts (S-percepts). The mental image of a scene results from accessing the ISS corresponding to that scene and bringing into the IRS a description of the scene from the appropriate level in the ISS hierarchy. Some examples of basic operations performed on mental images are changing the relative position of an object, enlarging or zooming in on an object, and scanning. In the IRS these are not analogic operations. Rather, changing the position of an object can be achieved by appropriately modifying spatial relations among S-percepts in the ISS. Zooming in on an object corresponds to selectively bringing in a more detailed (lower level) description of the corresponding S-percept from the ISS hierarchy. Scanning is the process of moving one's fixation point from object to object in the mental image and bringing in more material (from the

ISS) as the fixation point approaches the mental horizon (Pinker, 1980).

Fig. 2 shows an example that illustrates the inherent correspondence between S-percepts and A-percepts. It is, however, meant only as an analogy. Consider a two-dimensional array of two-valued logic elements with light bulbs attached to each element so that an element has value 1 if and only if its bulb is on and an element has value 0 if and only if its bulb is off. A pattern stored in this array may then be viewed either as an image or as a set of multi-dimensional bit vectors. Now interface this array with symbolic processes through a "vector set<->symbol" convertor that can translate between parametrized two-dimensional shape descriptions like "rectangle(2,3)" - meaning a rectangle of breadth 2 units and length 3 units - and bit vector sets. Similarly interface the array through a "photoreceptor-actuator" matrix with pattern-manipulating processes. This matrix is capable of sensing images displayed on the array as well as actuating bulbs in the array in response to input from the pattern-manipulating processes. Then the symbolic (pattern-manipulating) processes can input a symbolic description (image) to the vector set<->symbol convertor (photoreceptor-actuator matrix) and a corresponding bit vector set will be loaded onto the array which can then be sensed by the photoreceptor-actuator matrix (vector set<->symbol convertor) and the resulting image (symbolic description) can be sensed and modified by the pattern-manipulating (symbolic) processes. The modified image (symbolic description) can then be fed back into the symbolic (pattern-manipulating) processes in a similar fashion. This is an example of a mechanism that consists of a representation (bit vector set) residing in a specialized architecture (consisting of the array, photoreceptor-actuator matrix and the vector set<->symbol convertor) which generates symbolic descriptions and image-like interpretations of the represented entity and also acts as a two-way channel between symbolic and pattern-manipulating processes. In this example the symbolic shape descriptions are analogous to S-percepts and the patterns depicted on the array are analogous to A-percepts. The bit vectors are analogous to the primitives that S-percepts are composed of.

5. RELATED WORK AND DISCUSSION

Work by Kosslyn and Schwartz (1977) on two dimensional mental images and by Pinker (1980) on 3-D images report similar models. Kosslyn (1981) provides a cognitive theory that utilizes two distinct representations - deep non-pictorial representations and surface representations (patterns) in a visual buffer. There are similarities and differences between this and our model. The ISS and A-percepts may be viewed as deep and surface representations respectively. The IRS can have a component similar to the visual buffer as a medium for A-percepts. However, the ISS is a hierarchical multi-resolution representation which consists of object descriptions composed of visual primitives, as opposed to propositional encodings in the form of lists and literal encodings. Also, an architecture specialized for visual perception is central to our theory.

Pylyshyn (1981; 1984) has also made significant contributions to the imagery debate. He uses the concepts of tacit knowledge and cognitive penetrability to argue against a purely "analogue" position. The IRS is not in opposition to these concepts. In fact, some cognitive processes that operate on ISS or A-percepts may be (and some may not be) cognitively penetrable and thus influenced by tacit knowledge, overt instructions etc. However, we postulate that the interpretation of the ISS by the visual modality-specific architecture and the basic operations that the architecture provides on ISS and A-percepts are not cognitively penetrable and that these are in fact properties of the functional architecture (as defined by Pylyshyn) of visual perception and mental imagery.

The IRS does not provide for a separate analogic representation. However, the preferential treatment of modality-specific operations that analogists require is preserved in the IRS. On the other hand, while the IRS does not confirm to propositionalists' belief in a uniform mechanism for all of cognitive and perceptual processing and instead advocates modality-specific architectures, propositions are in fact one type of discrete symbolic representation and thus

advantages of a uniform representation that a propositional theory provides are preserved in the IRS as well.

Our proposal is also related to the question about how a purely syntactic system, such as a Turing Machine, can make connections to semantics, except in an arbitrary way. For example, this question lies at the heart of Searle's Chinese Room argument which seeks to show that computer programs cannot understand the meanings of symbols they manipulate. The IRS shows how modality-specific architectures preserve some aspects of the semantics of the world in such a way that the symbols can be seen to be "grounded" in perception as Harnad (1988) points out. This issue is described in more detail in (Chandrasekaran & Narayanan, in press).

6. CONCLUSION

In this paper we provide a description of how visual knowledge can be represented within the computational framework of discrete symbolic representations in such a way that both mental images and symbolic thought processes can be explained. Thus we answer the question "how can a percept that appears in the mind's eye as an image be symbolic?" by saying that it can indeed be so, given that a special purpose architecture, providing privileged visual operations, underlies visual representations. We propose a mechanism called an Image Representational System that provides interpretations of and visual modality-specific operations on symbolic visual representations, called Image Symbol Structures. An Image Symbol Structure is a hierarchical multi-resolution structure composed of S-percepts. S-percepts are made up of parametrized symbolic primitives of visual attributes such as texture and color. S-percepts represent delineative objects or parts of objects seen. A-percepts are interpretations of S-percepts that give rise to mental imagery. S-percepts and A-percepts may be viewed as dual facets of a single entity, namely, information about a perceived object . The proposed architecture affords dual perspectives (symbolic and imaginal) on visual representations and similar mechanisms may underlie human visual perception and mental imagery.

REFERENCES

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

Chandrasekaran, B., Goel, A., & Allemang, D. (1988). Connectionism and information processing abstractions: the message still counts more than the medium. *AI Magazine*, 9:4, 24-34.

Chandrasekaran, B., & Narayanan, N. H. (in press). The dual nature of visual representations. (Technical Report). Laboratory for Artificial Intelligence Research, Department of Computer & Information Science, Ohio State University, Columbus, OH.

Harnad, S. (1988). Mind, machine, and Searle. *Journal of Experimental and Theoretical Artificial Intelligence*, 1, 5-27.

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

Kosslyn, S. M., & Pomerantz, J. R. (1977). Images, propositions, and the form of internal representations. *Cognitive Psychology*, 9, 52-76.

Kosslyn, S. M. (1981). The medium and the message in mental imagery: a theory. *Psychological Review*, 88, 46-66.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional shapes. *Proceedings of the Royal Society*, 200, 269-294.

Pinker, S. (1980). Mental imagery and the third dimension. *Journal of Experimental Psychology: General*, 109, 354-37.

Pylyshyn, Z. W. (1981). The imagery debate: analogue media versus tacit knowledge. *Psychological Review*, 88, 16-45.

Pylyshyn, Z. W. (1984). *Computation and cognition: towards a foundation for cognitive science*. Cambridge, Mass.: MIT Press.

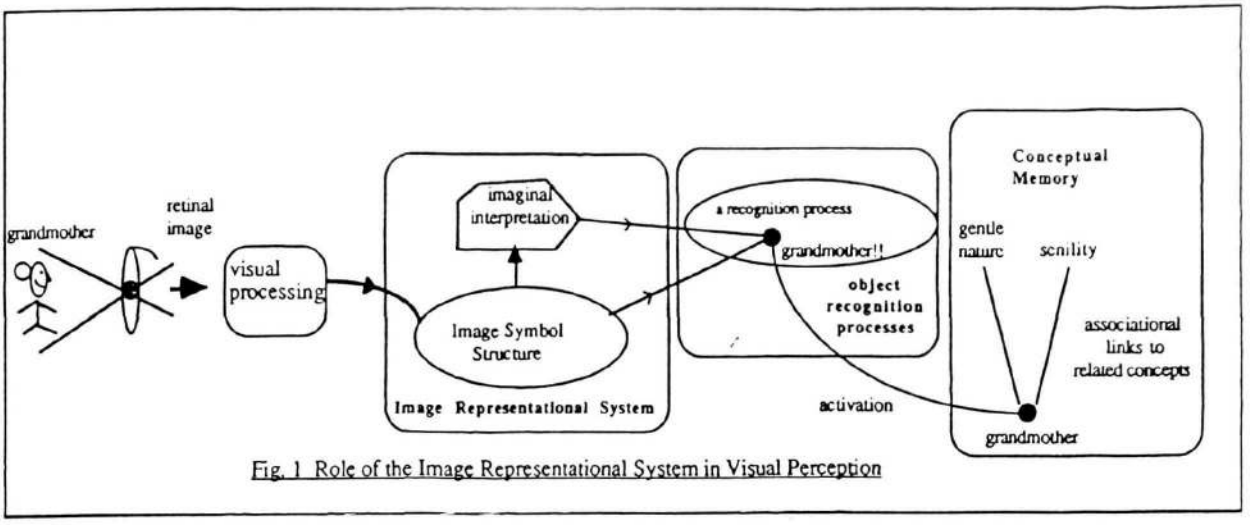

Fig. 1 Role of the Image Representational System in Visual Perception

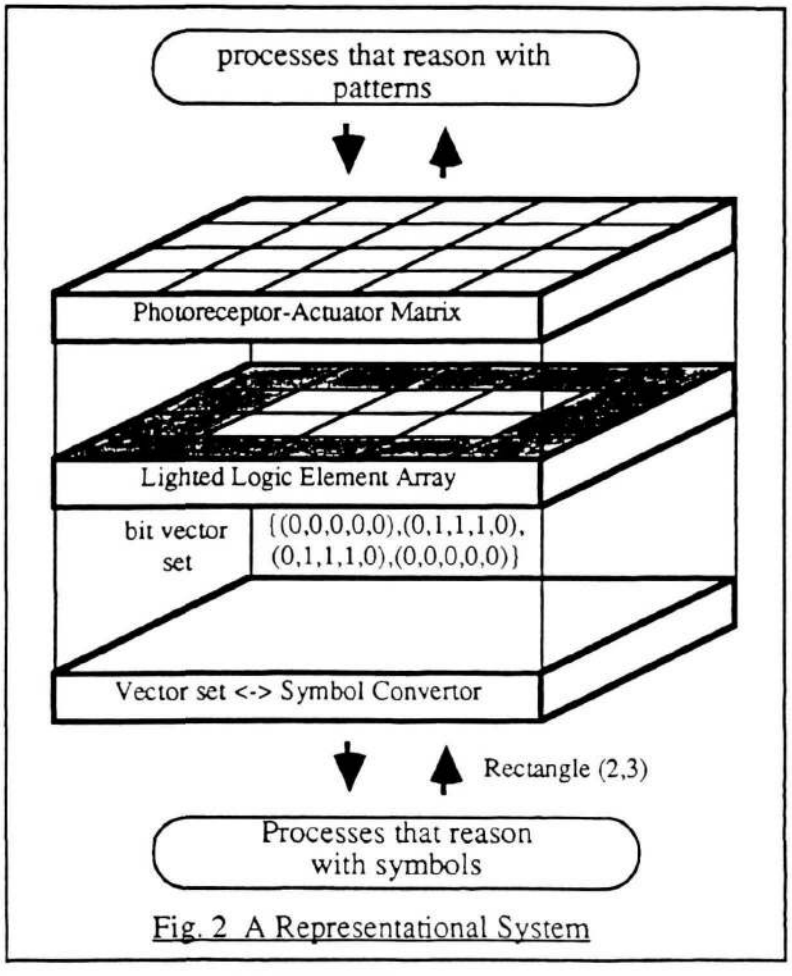

Fig. 2 A Representational System