1. Introduction

In this paper I shall attempt to summarize and extend some of the arguments I have advanced against the use of the notion of mental image as an explanatory construct in cognitive psychology (e.g., in Pylyshyn, 1973). In the first part (sections 2-4) I shall review some of my reasons for preferring to speak of cognitive representations —such as those involved in memory and thinking—as structured descriptions (albeit rather different from the usual linguistic descriptions) instead of images. I shall try to show that the facts of human perception, storage, and retrieval argue against the view that what is involved in such cognitive activity is some iconic and uninterpreted sensory pattern, as is implied when we speak of images. The point is not that there is no such object as an image, only that an adequate theory of the mental representation involved in imaging will depict it as having a distinctly nonpictorial character. Although the main arguments in the first part of the paper will be directed at the question of how knowledge is represented in long-term memory, most of the points apply equally to those transient structures constructed during imaging and thinking.

In the second part of this paper (sections 5-8) I shall examine the growing trend (at least in psychology) of referring to something

Note: I wish to thank Jerry Fodor for his careful and critical reading of an earlier draft of this paper. Part of the work of writing this paper was done while I was a visiting faculty member at the Artificial Intelligence Laboratory of M.I.T. Discussions with various members of the laboratory were invaluable in clarifying many problems in my earlier thinking (which is not to say that no errors of confusions remain in the present version). Research reported herein was supported by the National Research Council of Canada, Operating Grant A4092.
called an analogical representation as a way of representing nonlin-
guistic information. I shall argue that much of the attraction of this
notion stems from a failure to recognize some fundamental differ-
ences between the objects of perception (i.e., the physical environ-
ment) and the objects of cognition (i.e., mental representations). In
this connection I shall examine some evidence frequently cited as
supporting an analogical view of mental representation—in particular,
experiments on such mental manipulations as "mental rotation" of
figures.

2. What Is a Mental Representation Like?

It is schemata, not images of objects, which underlie our pure sensible con-
cepts. No image could ever be adequate to the concept of a triangle in general.
. . . Still less is an object of experience or its image ever adequate to the em-
pirical concept; . . . The concept "dog" signifies a rule according to which
my imagination can delineate the figure of a four-footed animal in a general
manner, without limitation to any single determinate figure such as experience,
or any possible image that I can represent in concreto, actually presents.

Emmanuel Kant
(Critique of pure reason, 1781)

To begin, I shall give you an informal and somewhat discursive review of why I believe mental representations are appropriately thought of as a type of description. Consider what happens when a
scene is perceived and becomes assimilated into our store of knowl-
edge, and what happens when we later access this knowledge from
memory in recalling the scene.

It is useful to distinguish two phases of the process that intervenes
between the arrival of a proximal stimulus and its interpretation
and assimilation as knowledge. There is some reason to believe there
is an early phase in this process that has considerable autonomy—
i.e., it does not depend upon higher cognitive processes except in a
very general way, such as by adjusting peripheral receptors. Various
phenomena that appear early in life and seem to be resistant to learn-
ing—such as figure-ground separation, certain illusions, gestalt laws
of pragnanz, and perhaps some stereoscopic and temporal integra-
tion—may be identified with this phase. Roughly speaking, processes
such as that which Julesz (1971) refers to as "cyclopean vision" or
which Hochberg (1968) has characterized as the “mind’s eye” in perception may occur immediately after this stage.

David Marr (1975) has investigated the computational requisites of this lowest level of vision and has proposed a model of this process up to figure-ground isolation. The process first computes a rich description of the optical intensity level differences present in the image; this description is called the “primal sketch.” Marr then posits certain “non-attentive” groupings and first-order discriminations acting on the primal sketch. Higher-level knowledge and purpose are brought to bear on only very few of the decisions taken during this processing.

I would argue that from a computational point of view it is appropriate to treat this early semi-autonomous phase of vision as a special purpose transducer, which takes physical magnitudes as inputs and produces symbol structures as outputs. Regardless of the precise details of this phase (it is not clear, for example, whether Julesz, Hochberg, and Marr are describing precisely the same level), there is reason to believe that: (a) there is a semi-autonomous, preattentive phase in visual perception, (b) this phase is initiated by energy arriving at the sense organs, (c) only the output of this phase, and not intermediate steps, are available for further perceptual analysis, and (d) such cognitive processes as “noticing” and the assimilation of sensory patterns into cognitive structures take place after this phase. These characteristics are ones that one would expect a “wired-in” transducer to possess.

Because of the nature of this transducer itself, it may be excluded from the process of imaging, since there are literally no adequate physical stimuli—no light patterns—to which it can apply. It is not so obvious, however, that the output of the transducer—say the aggregated primal sketch—cannot be stored in memory or even generated in the process of imaging. I shall argue, however, that it is extremely unlikely that any preconceptual, preassimilated, or knowledge-independent data are stored or otherwise used in thinking or imaging.

Consider what happens to the transducer output as it is assimilated into some cognitive structure and stored in memory for later retrieval. First of all, we are clearly highly selective in what (and how) we notice. We have to be, since not only the scene itself but the transducer output is literally unlimited in its potential for interpretation.
So much is not controversial. But now let us look more closely at the *nature* of this selection or "noticing function." What follows is a sketch of some of the characteristics of the transformation that relates the output of a transducer and a memory representation of some event. Taken together, it seems to me, they provide a strong case against viewing cognitive (or memory) representations of perceptually acquired knowledge as consisting of unprocessed (unabstracted) records of transducer outputs. Note that this argument is being made not only against the view that memory consists of pictures, a view that may well be a straw man. It applies equally well if memory is thought to consist of collages of pictorial segments, sketches, dynamic motion pictures, holograms, encoded multidimensional intensity matrices, or any other form of record of a particular concrete event (the one-time output of the transducer). Properties of the transformation and of the memory representation that lead us to this view include the following.

(1) The transformation between transducer and memory does not simply produce a degradation of resolution (a blurring or a mapping of a coarser grid), since we clearly do not perceive (in any sense of that word) or remember something that is complete in all aspects but low in detail or in precision. As Bobrow (1975, p. 8) puts it, "Human visual memory does not seem to have (the) property of uniform extraction of detail, or of exhaustiveness."

(2) The transformation is not a continuous topological deformation of the pattern of stimulation. No continuous transformation results in such commonplace phenomena as, for example, failure to notice objects or relations in a scene, perceptual addition of features that were not there (e.g., "cognitive contours"), or noticing the "what" but not the "where" or "when" of scene contents. Furthermore, the radical manner in which perception is influenced by such things as motivation, expectation (e.g., see the review by Bruner, 1957), prior knowledge, or even stage of cognitive development, attests not only to the general malleability of perception, but to the high degree of stimulus-independent knowledge-based construction that goes into the mental representation. Although this simple point is frequently forgotten, it has been made repeatedly and eloquently in the psychological literature by people like Hochberg (1968),
Gombrich (1961), and Gregory (1974), and in the philosophical literature by people like Hanson (1953) and Goodman (1968).

(3) The representation of a scene contains many non-pictorial (and non-sensory) aspects—aspects that cannot realistically be said to be in the sense data at all. Examples of the latter are the perceived relation of causality (Michotte, 1963) or the relations of attack and defence on a chess board (Simon and Barenfeld, 1969). In fact I would argue that all relations are of this type: that there is no fundamental difference between the relations “is to the left of,” “is under attack by,” or “causes,” inasmuch as none of them is any more “directly in the scene” than any other. These are all abstract conceptual relations far removed from the output vocabulary of the transducer.

(4) Although we often appear to go through a process of recalling an image of a scene and then of noticing or perceiving aspects of that image, this recall-reperceive sequence is extremely problematic. The fact that we can recall a scene, or part of a scene, by addressing aspects of the perceptually interpreted content of the scene argues that what we have stored is already interpreted and not in need of re-perception as we supposed. Retrieval of images is clearly hierarchical to an unlimited degree of detail and in the widest range of aspects. Thus, for example, I might image a certain sequence of events as I recall what happened at a conference session. Such images may be quite global and could involve a whole scene in a room over a period of time. But I might also image someone’s facial expression, or the substance of his remarks, or my reactions to the papers, or the approximate location of a questioner in the audience, without first calling up the entire scene. Such perceptual attributes must therefore be available as interpreted integral units in my representation of the whole scene. Not only can such recollections be of fine detail, but they can also be of rather abstract qualities, such as the mood of the assembly. Furthermore, when there are parts missing from one’s recollections, these are never arbitrary pieces of a visual scene. We do not, for example, recall a scene with some arbitrary segment missing, such as a torn photograph. What is missing is invariably some integral perceptual attribute or relation; for example, colors, patterns, events, or spatial relations (I might, for example, recall the
people who were in the front row without recalling exactly where they were sitting or what they were wearing). When our recollections are vague, it is always in the sense that certain perceptual qualities or attributes are absent or vague, not that there are geometrically definable pieces of a picture missing. All of the above suggest that one's representation of a scene must contain already differentiated and interpreted perceptual aspects. In other words, the representation is far from being raw and, so to speak, in need of "perceptual" interpretation. Because retrieval must be able to address perceptually interpreted content, the network of cross-classified relations must have interpreted objects (i.e., concepts) at its nodes.¹ This does not mean, of course, that what we retrieve cannot be further processed. The argument is simply that they are not subject to perceptual interpretation the way pictures are interpreted; by "perceptual" I refer to the processes of transduction and of interpretation or assimilation into cognitive conceptual structures.

Because the representation is so obviously selective and conceptual in nature, referring to it as an image—a term that has pictorial or projective connotations—is very misleading. Although there are some who have no objections to speaking of "conceptual images," I prefer the term "description" or "structural description" because this carries certain desirable connotations. For example, it implies that the representation is something that (a) must be constructed out of a vocabulary of available concepts (Kant's "Categories of understanding"), (b) bears a referential relation to the object it represents rather than a relation of "resembling," and (c) has its semantics defined by an accessing function that is not assumed to be the entire visual apparatus (I shall have more to say about (b) and (c) in section 6 below). The structured description approach also gives one a psychologically appropriate way of talking about the complexity of a representation. Such complexity is not a property defined over a material layout (e.g., extent, dimensionality, number of topologically defineable segments, etc.) but rather a property defined over a symbol structure (e.g., number of symbols, relations, etc. or, better still, number of nodes at different levels of a tree structure—some details for the latter measure have been proposed and successfully tested in a limited context by Palmer, 1974). Complexity, in others words, is a measure over a description in symbolic
or conceptual terms, not over a description in geometric or physical terms.

3. Some Illustrative Examples

To give an idea of what I believe can be gained by this approach to imagery, signalled as it were by the new terminology, I shall describe several phenomena—mostly ones observed in children (who incidentally have been shown to have particularly good "visual imagery" ability)—and then give an account of the phenomena in terms of the notion of "description."

Figure 1a schematically depicts some findings reported in Piaget and Inhelder (1956). When young children below the ages of four or five years are shown a colored fluid in an inclined transparent container and are later asked to draw (or to indicate by describing and pointing) what they saw, they typically indicate the fluid as being parallel to either the bottom or the side of the container. Two other related figural "errors" of reproduction or recognition that occur with young children are shown in Figures 1c and 1d. The first part of Figure 1c illustrates the well-known mirror image confusion common in children. Figure 1d (reported by Eve Clark, 1973) illustrates the following phenomenon. When young children are shown a small object being placed next to a container and are asked to imitate exactly the action they have just observed, they most frequently place the object inside the container. There are a number of other similar transformations that children systematically produce in imitating actions.

Such "errors" can be simply accounted for if we assume that children's internal vocabulary of descriptive concepts is limited or that the priorities they place on the use of such concepts differ from those of adults. For example, without a concept for the relation "is left of" or "is right of," no description of an asymmetrical figure is possible that distinguishes that figure from its mirror image (Figure 1c). Similarly, if a child lacks the concept of "geocentric level," his percept of the fluid in the inclined container may not be the same as an adult's. In such a case the nearest appropriate concepts (e.g., perpendicular, parallel) may be used, producing the observed errors. Of course such differences in the availability of figural concepts do not always produce a failure to make a distinction. In some cases
Figure 1. Examples of the conceptual nature of visual mental images. A: Typical recall error made by children who have not mastered the concept of "level" (Piaget & Inhelder, 1956). B: Common errors in children's drawings of a cube (adapted from Weinstein, 1974). C: Children are more likely to confuse a figure and its mirror image than a figure and a mis-oriented copy (e.g., Rock, 1973). D: When imitating the action of placing an object beside a container, a child is more likely to place the object inside the container (Clark, 1973). E: Chess masters' superior "visual recall" of chess positions holds only when the the positions are taken from real middle games (reported in Chase and Simon, 1973). F: When subjects are asked to synthesize a figure by mentally superimposing two given figures, their performance depends on the way the figure is decomposed (adapted from Palmer, 1974).
they result in a failure to perceive similarities. Thus, as illustrated in the second part of Figure 1c, children tend to treat a figure in a different orientation as a different figure. For example (see Rock, 1973), young children make fewer orientation generalization errors, confusing figures in different orientations in a discrimination learning experiment, than their older counterparts. (The apparent indifference of young children to the orientation of pictures and print may simply be due to their failure to recognize the importance of orientation to recognition—see Rock, 1973.) In our terms, for the same form in different orientations to be perceived as similar it would have to be represented in terms of appropriate orientation-independent concepts (e.g., relations like centripetal-centrifugal).²

The case of imitation is very similar. For what is considered to be mere mechanical imitation must be mediated by a memory representation, which, we have been arguing, depends on the availability of descriptive concepts. Of course imitation also depends on other factors such as preferred response strategies. Eve Clark (1973) found that if she asked a child to imitate the experimenter’s action of moving a small object and placing it beside a container, the child most frequently performed a similar movement but left the object inside the container (as depicted in Figure 1d). One might be inclined to say that the child saw an object being placed in some appropriate proximate relation to a container and constructed an internal representation that recorded this observation. In imitating, the child selects an action in his repertoire, according to some preferences such as discussed by Clark (1973), which is compatible with this representation. For the child, then, the observed and produced actions fall in the same equivalence class—as captured by its internal representation—just as a figure and its mirror image are in the same class because the child’s representation is conceptually less differentiated than that of the adult.

Figure 1b depicts a related situation in which children’s drawings of a cube deviate from those made by adults. Notice that relative to the more familiar perspective drawing, the children’s renderings are more faithful to a description of the cube (e.g., the angles are mostly 90°, faces are square and perpendicular to other faces, etc.). Weinstein (1974) found that older children produce hybrid drawings (such as the two on the second line) as they attempt to incorporate
the perspective conventions that the Western adult community did not adopt uniformly until the Renaissance (and that, conceivably, are considered the veridical rendering because they represent the way cameras operate). The point is that "optical projection onto a single stationary point of view" is only one of a large number of ways of transforming a mental representation of a three-dimensional object into a two-dimensional drawing. Since the mental representation is necessarily not identical with any of the drawings, it does not independently determine one of them as the unique "correct" rendering. The process of selecting from among the set of drawings compatible with the mental representation of the object must depend on other, probably learned and culturally specific principles—often collectively called "drawing skill."

Because children do not have so refined a vocabulary of descriptive figure-concepts as adults do, their mental representations may tend to be less differentiated than those of adults. In addition, children have not mastered adult conventions and various physical principles, so their reproductions and imitations are a source of more dramatic illustrations of the abstract conceptual or descriptive nature of mental representation. The principle is not by any means, however, confined to children. Consider the following two published results from adult subjects.

Chase and Simon (1973) describe a series of ingenious experiments on the visual memory of chess players. Chess masters are known to have a vastly superior memory for board positions than mediocre chess players. The question arises whether chess ability rests in part on exceptionally good visual imagery ability. Chase and Simon conclude, in effect, that such ability is very important but that it does not consist simply of a general visual imagery talent. The superior performance of chess masters in reproducing a board position after only a few seconds' exposure is manifest only when it is a true board position taken from an actual chess game. The difference between masters and duffers disappears when a random arrangement of the same pieces is used as the stimulus (see Figure 1e). Thus what appears to be strictly visual memory is extremely sensitive to chess-specific patterns. Chase and Simon conclude that the exceptional visual memory of chess experts derives from their very large repertoire of familiar chess configurations. Such "configurations"
are not defined simply by geometrical patterns, since they are independent of the shape, size, or color of chess pieces but are sensitive to nongeometrical relations such as attack, defence, control, etc., which may even be spatially nonlocal. In our terms chess masters may be said to have a rich internal vocabulary with which to construct a representation of the board. Their representation can thus be constructed rapidly and is also conceptually simple—i.e., it consists of a compact description constructed from a rich vocabulary appropriate to the game of chess. Even though the memory is of an apparently visual pattern, the particular internal representation constructed depends on nonvisual factors. As a further corroborating example Eisenstadt and Kareev (1975) have shown that the pattern of errors in the recall of a particular Go board configuration depends on whether a subject perceives it as a position in the game of Go or in the game of Gomoku. In this case geometrically identical patterns are shown to be represented differently depending on highly cognitive factors.

Another study illustrating the large conceptual component involved in visual imagery was reported by Palmer (1974). He presented subjects with two patterns (each with the same number of line segments), which they were to superimpose to yield a third synthesized pattern. The difficulty subjects experienced (measured in terms of both latency and accuracy) in synthesizing a particular pattern depended on the way it had been presented as subpatterns to be superimposed. Thus the subpatterns designated as “good” configurations (e.g., Figure 1e(ii)), which shared a larger number of major integrated substructures with the required figure (defined by Palmer in terms of a hierarchical description), were easier to synthesize.

Again, as in the other evidence cited above, Palmer found that in what appears to be fundamentally a visual imagery task it is the conceptual rather than some sort of graphic complexity that is the essential determiner of task difficulty. Apparently even when visual imagery seems clearly implicated, the underlying representation is best characterized as something more abstract and conceptual—i.e., what we have been calling a structured description. The point is not simply that there are tasks in which something beyond an iconic image is involved, but that even in cases in which visual images would appear to be the chief mode of representation, task complexity
measures lead one to recognize that what serves as the mental representation is highly cognitive. Furthermore, the most perspicuous way of talking about such representations is in terms of such notions as a vocabulary of internal concepts, compactness of descriptions couched in this vocabulary, and other locutions much more appropriate to descriptions than to pictures. I should emphasize, however, that in using the term “description” I am not referring to linguistic objects in the conventional sense. Such internal descriptions (e.g., those discussed in the next section) cannot be directly externalized as sentences. The reasons for this are, first, that the symbols involved may not have corresponding lexical labels in any natural language and, second, that the descriptive structures are not discursive in the sense that they must be scanned in a fixed sequence (as in the case with sentences). The primary reason for persisting in calling them “descriptions” lies in the way these representations are related to what they represent, as discussed in (a)-(c) in section 2 above.

4. Symbol Structures for Imagery

Those who are familiar with work in artificial intelligence will recognize that most computational data structures (e.g., semantic networks) have properties that make them suitable candidates for internal descriptions. Although they are all articulated symbol structures, most have very different formal properties from those of natural language or even of predicate calculus. For example, they contain flexible access paths among symbols that can be tailored to specific goals, they may designate procedures that can be evoked at appropriate times, and they may contain propositional forms that are asserted only in appropriate contexts, when bindings for free variables are provided. Although the formal properties of such descriptive systems as a class are not yet well understood (see Woods, 1975), it seems clear that they are at least a promising candidate as a formalism for internal representations, not only because of their descriptive power but also because of their structural flexibility. The latter quality is most important in the present context because we are presumably interested in finding a psychologically adequate form of representation as well as a logically adequate one. In particular we are interested in accounting for certain properties of natural intelligence. We would like to be able to give an account of why
certain tasks are easier than others (i.e., why some take less time, result in fewer errors), why certain types of systematic errors occur, how representations are transformed in memory and thought, etc. From this perspective there are good reasons for believing that different representational structures are used at different times and for different purposes. For example, representations that are temporarily constructed in the course of activity we call imaging have some properties not shared by those constructed during episodes we would describe as inner speech. Evidence for this need not rely solely on reports that these activities are accompanied by different subjective experiences. Various measures of access and manipulatory complexity (as assessed, for example, by reaction times) also supports this view. My contention is, however, that there is at present no good reason to reject the view that a common articulated descriptive system underlies all of these representations and that the apparent differences arise from such things as the particular vocabulary of symbols (i.e., designations of concepts and relations) that are used, accessibility paths that are set up among parts of the representation, and the particular operations that are evoked to process these symbol structures. For example, some characteristics of temporary symbol structures that have been developed to model aspects of imaging (as in the work of Baylor, 1972; Moran, 1973; Farley, 1974) include the following:

(1) Representations of physical objects and their attributes are individuated—i.e., individual objects are distinguished by distinct internal symbols, and attributes are often attached to them (i.e., attributes are accessible through these symbols). Thus individuals in such representations can be counted. For example, there would be no atomic symbol corresponding to “n windows.” Rather n distinct symbols would be generated one for each imagined window. These might even have to be related to one another by relations such as “above,” “to the right of,” etc.

(2) Spatial and temporal relations in such imaginal data structures are often found to provide particularly good access paths. For example, given an object in such an imaginal structure, it is easier to retrieve an object that is in the relation “next to” or “above” to it than in the relation “larger than” or “same color as” to it (e.g., see Collins and Quillian, 1969).
(3) When a temporary data structure corresponding to an image is constructed, many “default values” are included so that ready access is provided to some details not obviously relevant to the task at hand.

(4) In such workspace data structures we would not have quantifiers or logical connectives (i.e., we would not have an “image symbol structure” for the proposition of “all red blocks”), although sometimes prototypical patterns might be made to serve some of these functions.

(5) Common symbol systems, particularly those designating spatial relations, might be shared by various modalities and by the motor functions as well as the perceptual and image functions. This may explain why coordination is possible and why phenomena such as stimulus-response compatibility (Fitts and Seeger, 1953) and intramodality interference (Brooks, 1968) are observed.

(6) We might even postulate that certain operations performed on objects in the “imaginal workspace” are computational primitives. Suggestion (2) above can be thought of in this way—i.e., given a reference to an element, retrieving the element that designates some spatially adjacent object may be computationally cheap. In fact this could be the theoretical interpretation of the claim that images are “spatially organized”—the reason is not that image data structures are distributed in space (whatever that could mean), but rather that spatial relations such as adjacency can be used as access paths. One must be careful, however, in positing computational primitives for the image workspace. As I shall argue it later, it is very tempting to posit as primitive operations, processes that conceal a major part of what one is trying to explain (as occurs if we take metaphors such as the “mind’s eye” or “mental rotation” too literally).

But attempts to develop formal models of imagery are just beginning, and most of the story is yet to be told. Should it be possible to model all forms of cognition (in a manner that takes into account not only logical requisites but also psychological complexity evidence) in a single formalism, not only would be have achieved considerable theoretical parsimony, but we would also have made a significant contribution to bringing some integration to many classical philosophical puzzles of cognition.
I might remark that I have occasionally heard objections to such data structures on the grounds that the symbols are arbitrary atomic elements. As in mathematics, the symbol that designates some quality or some object is chosen for the theorist’s convenience; hence in computer science the symbol is usually a string of letters forming a mnemonic word or phrase. What troubles some people is the fact that one must at some point bridge the gap between the symbol and the world outside. Thus, even admitting that much of the representation is symbolic, they would prefer to have some nonarbitrary symbolic content in the representation. For example, one proposal is that a fragment of a representation of a checkered tablecloth might be expressed as something like: “(TEXTURE TABLECLOTH □),” where the third term is a piece of template that can both designate checkeredness and be used to identify this texture in some transducer (the same might be done for TABLECLOTH but perhaps not for the more abstract concept TEXTURE). But although this hybrid expression may look different from a standard data structure, this is a property only of the way we have chosen to display it. It is no different from (TEXTURE TABLECLOTH Q137) provided the atom Q137 is used consistently (a) when a reference to a checkered pattern is intended; (b) when a checkered pattern is detected by the transducer hardware; (c) when a verbal reference to checkered patterns is received or generated, etc. (The last condition is contingent on the system having “learned” the relation between Q137 and a verbal label such as “checkered.”) The nonarbitrariness of a symbol arises entirely from the system of symbols within which it occurs as well as the way in which input-output transducers are wired to translate between energy patterns and symbols.

5. Are Images Analogue?

Although enthusiasm for pictorial representations, which resemble what they represent, or for some kind of “sensory storage” may be waning (at least in the case of representations stored in long-term memory), many people do not subscribe to the view that articulated symbol systems are sufficient to account for many of the phenomena that have been studied by psychologists interested in imagery. These people feel we are forced into the position of admitting at least two
radically different types of representations—one to encompass articulated, verbal, or factual information, and the other to capture continuous, analogical, sensory, or wholistic types of phenomena implicated in imagery and perhaps other areas of cognition and thought. Within artificial intelligence the study of different types of representational systems is very much an active frontier, and it is impossible to rule out such a hypothesis—however vague the current notions of what constitutes an "analogue" representation. Whatever the outcome of this research, however, it nevertheless appears to me that the arguments and evidence that people have typically presented as favoring such nonarticulated representations have been far short of persuasive.

A wide variety of experimental phenomena has been cited in support of the claim that such nonarticulated imagery representations must be entertained. For example, there are experiments demonstrating differences in recall between concrete imaginable situations and abstract ones, and between performance under instructions to image vs. instructions to rehearse; experiments demonstrating confusion errors based on appearances as opposed to category membership; experiments showing intramodality interference during imaging; and more recently, experiments using reaction time measures, which show that relative difficulty of some tasks performed imaginarily mirrors the relative difficulty of such tasks performed perceptually, i.e., while examining actual displays (see Kosslyn & Pomerantz, 1977). In addition there are a number of ingenious experiments (mostly by Roger Shepard and his students; e.g., Shepard, 1975; Cooper & Shepard, 1973; Metzler & Shepard, 1974) also using reaction time measures, which suggest that mental manipulation of images involves carrying out a sequence (or possibly a continuum) of transformations paralleling those that would be carried out in manipulating real objects. For example, the time taken to determine that two figures are identical except for their relative orientation has been found to be a linear function of the angle between them. This effect has been demonstrated in a variety of ways, including asking a subject to prepare mentally for the second of the two figures at some prescribed angle. The preparation time appears to be the same linear function of angular deviation (Cooper & Shepard, 1973).
This is explained by saying that subjects rotate an image of the presented form at some constant rate.

Taken as a whole these studies have persuaded many psychologists that mental representations of objects, particularly in the visual modality, must be structurally isomorphic to the objects they represent. This is often phrased by saying that representations are analogical rather than descriptive or articulated and that they are transformed by wholistic analogue processes. The argument is often made that although it might be possible to fabricate an account of how such results could arise from articulated descriptive representations, such accounts are always post hoc and unnatural. Accounts based on positing the manipulation of internal analogues are invariably more natural and are independently motivated by the observation that the same laws of perception and transformation can be applied to the internal representation as are known to apply to external stimuli. My response to this is threefold.

(1) If one takes the position I outlined earlier—viz., that perception involves the construction of an internal description—then it should not be surprising that cognitive operations (e.g., judgments) occurring during perception bear some strong relation to cognitive operations occurring during imaging. On this account both involve the further processing of these internal descriptions.

Furthermore, it should not be surprising if operations upon internal representations show some systematic relationship to operations that would be carried out upon the corresponding objects in the world. We surely have some representation of physical operations as well as of objects. Our knowledge of what it means to manipulate objects derives at least in part from our experience in carrying out actions on real objects. Thus if someone asked me whether a piece of paper of a certain shape could be folded up to form a certain polyhedral form, I would not attempt to solve the problem by applying any arbitrary transformation to my representation of the paper. Instead I would go through a process of solving a series of subproblems, each of which involved answering the question, “What will happen to the shape if I make the following fold?” But this is far from being an argument for internal analogues, as many writers have claimed (e.g., Shepard, 1975). In any problem the solution
method I use depends on both the demands of the task (e.g., in this case only physically possible transformations are legitimate) and on the way my knowledge about such transformations is structured. Presumably my knowledge of folding consists of such facts as what happens to the shape of an object when a single fold is made in it, just as my knowledge of addition consists of such atomic facts as $2 + 3 = 5$, which I use in the solution of more complex problems. But notice that all this implies only that I solve the problem in stages by applying operations to representations. There need literally be nothing in common between my mental representation of folding and actual folding, other than that one can be used in certain situations to compute the effect produced by the other—i.e., to compute what could result from actually completing a fold. In fact I shall argue in the next section that theoretical adequacy will force the mental operation to be unlike the physical operation in certain critical respects, giving the theory that “unnaturalness” that bothers many people.

If there were a high degree of correspondence between operations in the world and mental operations (including comparable complexities and constraints on what could be performed), one might perhaps be justified in speaking of the mental activities as in some sense “analogue.” But the correspondence is highly partial: only certain aspects of some physical operations have correspondences. Although mental operations have few of the constraints that affect physical operations (i.e., it is easy to imagine physically impossible phenomena), they are also subject to many constraints for which there are no physical counterparts. There are countless simple operations that are impossible to imagine accurately. Sometimes we cannot keep track of all the relations. For example, imagine a familiar scene; now try to image it upside down, out of focus, viewed through a green filter, etc. Or sometimes we lack the tacit knowledge of the physical laws governing the phenomena. Examples of this are a child asked to imagine what will happen when a block is pushed over the edge of a table, or an adult asked to imagine the trajectory of a weight being dropped behind a screen. Ian Howard (1974) has discussed an interesting series of experiments in which he shows that adults’ “perceptual schemata” are often not consonant with the laws of physics. In fact in a recent ingenious experiment, using trick
3-D photography and motion pictures, he showed that (a) about half the college students he tested could not articulate the principle that fluid levels in a container remain horizontal as the container is tilted, and (b) those who could not articulate this principle could not recognize gross anomalies (up to 30° from horizontal) in fluid levels, whereas those who did articulate the principles were very accurate in their detection of anomalies (Howard, in press). The failure of “perceptual schemata” to be veridical has also been demonstrated for fluid levels by Thomas, Jamison, and Hummel (1973). The point is that even in perception, the detection of deviations from physical laws is far from automatic. Obviously in the case of imaging physical transformations, the ability to image the correct effect is highly dependent on what the subject knows and does not merely follow from the behavior of internal analogues. It is especially not a consequence of any intrinsic property of some analogue “medium,” as I shall argue presently. I shall return to the differences between physical and mental operations in sections 7 and 8 when I discuss mental rotation experiments.

(2) Although there are similarities between cognitive operations in perception and in imaging, there are also some outstanding differences that may be more revealing of the underlying processes. For example, the order of scanning and the sorts of things that can be “noticed” in imaging are much more constrained than in perception. The reason for this is partly that a scene has a stable independent existence and can be reexamined at will to produce new interpretations. In contrast, the construction of an internal description from stored knowledge can hardly be divorced from its interpretation. While some reinterpretation is certainly possible, it surely is more like the derivation of new entailments from the stored knowledge than like the discovery of new aspects of an environment by the usual visual means. Discovering even moderately novel readings from a mental image such as those required to find simple embedded figures in a pattern, have been shown to be exceedingly difficult (Reed, 1974). Another important property of an image that distinguishes it from perception is that it is quite limited in its content. This limitation, however, does not appear to depend on any simple measure of geometrical complexity so much as on conceptual, or what I would call descriptive, complexity. The latter in turn varies depending on
the availability of appropriate concepts for describing the display, as I have already argued.

(3) My third reaction to the arguments for analogical representations based on the parallels between imaging and perception is the following. I maintain that the reason why structured descriptions and the computational processes that go with them appear unnatural is precisely that they are an earnest attempt to make explicit the detailed structure of the entire cognitive system involved in imagery, down to the level of mechanically realizable processes. It seems that naturalness of theoretical accounts of imagery can be gained by sweeping a large part of the puzzle under one of two rugs: we can attribute some of the phenomena to unexplained properties of the "mind’s eye" or some other interpreting process, or we can attribute some of them to intrinsic properties of the analogical representational medium. I shall suggest that both of these moves involve us in the game of obscurum per obscurus, an unreasonable price to pay for naturalness.

6. Properties of the Mind’s Eye

I shall begin this sketch by giving the following caricature of a class of arguments for “analogical” or “direct” or “presentational” representations. Consider the parallel between the pairs “organism-environment” and “mental process-representation” (Figure 2a, b).

The system depicted in Figure 2a must surely have many properties in common with the system depicted in Figure 2b; otherwise thought would be irrelevant to action, and our chances of survival would be negligible. From this one is tempted to say that representations and the objects they represent must have much in common. Beginning with this innocent remark we are irresistibly and imperceptibly drawn towards the fatal error of attributing more and more of the properties of the environment, as described by the physical sciences, to the representation itself. If I were permitted to misappropriate other people’s terms slightly, I might call this the tendency to commit the “stimulus error” after Titchner or to succumb to the “objective pull” after Quine. It is in the failure to emphasize the fundamental differences between the mental object, which we call the representation, and the physical object (i.e., the two right-hand elements in Figure 2) that we run ourselves into various traps. The
physical object has a stable existence, its transformations are governed by natural laws, and it is open to as many readings or interpretations as are compatible with the cognitive powers of its perceiver. The representation, on the other hand, is already an interpretation or reading given to the object by an act of conceptualization, and any transformations of the representation are determined by cognitive operations that may or may not bear any relation to the laws

Figure 2. Because the object-organism relation depicted in A must share some functional properties with the "mental image"-"mind's eye" relation depicted in B, we can be seduced into attributing many object properties to the image and many perception properties to the mind's eye.
of physics. These may seem like rather obvious differences, but I shall argue that the failure to keep them in focus has been behind some of the arguments for analogical representations.

The temptation to draw the external world inside the head leads to a classical dilemma of imaginal representation: if the representation is too similar to the world it represents, it is of no help in apprehending the world, since it merely moves the problem in one layer; but if it is too dissimilar, then how can it represent the world at all? This apparent dilemma turns on the use of the word “similar,” which surely is appropriate only when two things are examined by the same process (e.g., when they are both viewed). This, however, is a gratuitous assumption that underlies and confuses much of the discussion of representation.

For example, one frequently hears that a “nonverbal” representation preserves the structure of the environment it represents. Such preservation of structure is taken by many to be a defining characteristic of analogical representations. Aaron Sloman (1971, pp. 216-217) makes the following comparison in contrasting analogical representations with a predicate calculus formulation, or what he calls Fregean systems:

In an analogical system properties of and relations between parts of the representing configuration represent properties and relations of parts in a complex configuration, so that the structure of the representation gives information about the structure of what is represented. . . . By contrast, in a Fregean system there is basically only one type of ‘expressive’ relation between parts of a configuration, namely the relation between ‘function-signs’ and ‘argument-signs’. . . . For example, the denoting phrase ‘the brother of the wife of Tom’ would be analyzed by Frege as containing two function-signs ‘the brother of ( )’ and ‘the wife of ( )’ and two argument-signs ‘Tom’ and ‘the wife of Tom’ as indicated in ‘the brother of (the wife of (Tom))’. Clearly the structure of such a configuration need not correspond to the structure of what it represents or denotes.

Now this may sound like a reasonable claim until one tries to interpret the phrase “the structure of X.” All the phrase can mean is that some function (which I have called the “semantic interpretation function”) can give X an interpretation as a structure. There is literally nothing intrinsic in any object that can be called its “structure.” Sloman’s distinction is empty unless we are told which of an unlimited number of potential structures it refers to. For example,
Sloman notes that in the above example the sign "Tom" is part of the sign "the wife of Tom," whereas in the situation represented, the thing designated by the sign "Tom" is certainly not part of the thing designated by the sign "the wife of Tom." Thus, he argues, the structure of the representation does not reflect the structure of what is represented. But the point is that so long as the function that interprets the phrase shows no inclination to attend to what you and I might call the relation "is a part of," or, if it does attend to such a relation, does not identify it with a similarly named (but in fact quite different) relation in the world, the problem does not arise. In this case "is a part of" is simply not a signifying relation. In fact it would be correct to say that from the point of view of the "semantic interpretation function" such a relation does not exist. Thus it is true that a potential relation in the representation does not signify a relation in the world. But neither does the relation "is heavier than" in a picture represent a relation in the scene depicted—e.g., the part of a picture corresponding to a cloud may weigh more than the part of the picture representing a building, but so long as the interpretation function does not attend to relative weights, this remains irrelevant. Or to take a less farfetched example, the relation among areas in a Mercator map projection is not signifying, although a compass direction is.3

Thus discussions about the nature of mental representations should really be discussions about representational systems consisting of the pair "representation" and "semantic interpretation function." Furthermore, my earlier claim that representations are descriptions should more properly be put as a claim that representations function as descriptions—i.e., they are related to the objects they represent in the way sentences are related to the objects they describe (i.e., via an interpretation or something like Wittgenstein's "laws of projection") rather than the way photographs are related to the objects they picture (i.e., via laws of optics and principles of projective geometry).

Occasionally writers have recognized the importance of the semantic interpretation function. In one recent case it has been used to resolve a long-standing philosophical puzzle relating to the indeterminateness of images. This is a problem that Dennett (1969) considers a serious difficulty with the imagery view. Recently, however,
Fodor (1975, p. 191) has argued that such indeterminateness is not problematic for a pictorial view of images. Fodor’s argument rests on a recognition that the content of a mental representation is always relative to some interpretation. This is precisely the point we have been discussing above (as well as in Pylyshyn, in press). Let us look briefly at Fodor’s argument that what we call images can reasonably be understood as indeterminate.

What makes my stick figure an image of a tiger is not that it looks much like one (my drawings of tigers don’t look much like tigers either) but rather that it’s my image, so I’m the one who gets to say what it’s an image of. My images (and my drawings) connect with my intentions in a certain way; I take them as tiger-pictures for purposes of whatever task I happen to have in hand. Since my mental image is an image, there will be some visual descriptions under which it is determinate; hence there will be some question whose answers I can ‘read off’ the display, and the more pictorial the display is the more such questions there will be. But, in the case of any given image, there might be arbitrarily many visual properties which would not be pictured but, as it were, carried by the description under which the image is intended.

This is an important and relevant observation. But notice what it has done to the notion of an image. The image has lost its essential quality. It has become an object that must be read via an intention and that can be read in many different ways. In other words, it contains forms or symbol tokens exactly as does a structured description. What makes it pictorial, according to Fodor, is that there are many properties that can be “read off,” presumably with low computational cost. But this is precisely what happens when we enrich a structured description by making it more elaborate and detailed. The advantages of thinking of this as the elaboration of a description, rather than of the image being more pictorial (apart from the vagueness of the notion of “degree of pictorialness”), are that (a) this interpretation gives recognition to the fact that the elaboration is done within the constraints of available concepts rather than by the addition of arbitrary pictorial fragments; (b) “reading off” becomes a well-defined symbol-matching operation rather than involving all of the perceptual apparatus driven by intentions; (c) no matter how much elaboration of detail is carried out, there will always be an arbitrarily large amount of indeterminateness in the resulting representation (it will always fail to be determinate with respect to
some aspects which are not only determinate in the scene but which no picture would leave uncommitted), and furthermore, as noted in section 2 (1), the representation is not homogeneous in the amount of determinateness of various aspects; (d) this interpretation discourages the view, invariably associated with the term "image," that there exists an object that is interpreted the way a scene is interpreted (i.e., visually), that has a stable simultaneous existence so that it can be scanned perceptually for new readings, and that inherits certain intrinsic properties from the material medium in which it is embedded (e.g., rigidity under various transformations—see the discussion below).

Replacing "images" with "images under descriptions," as Fodor does, frees the term from many of the philosophical problems that plagued it in the past. The trouble with this move is precisely the problem of making clear the sense in which images under descriptions are to be distinguished from descriptions. Fodor (p. 190) puts it this way: "Images under description share their nondiscursiveness with images tout court. What they share with descriptions is that they needn't look much like what they represent." Thus discursiveness seems to be the crucial property. But, as we have seen, symbol structures are not discursive in the sense that sentences are—i.e., they need not be read in a prescribed order. The order of "scanning" is determined by the accessing algorithm and makes use of the relations that are the access paths of the structured description, just as a visual scan of an image would presumably be determined by the intentions of the perceiver together with something like peripheral vision. This amounts to saying that we have yet to see a viable distinction among images, images under descriptions, and structured descriptions when any of these is embedded within a representational system—i.e., when paired with the appropriate semantic interpretation function. Given the advantages of the nonpictorial description option mentioned above, I see no reason to abandon this approach, which at least has some theoretical exemplars in current computational models.

It is important to keep in mind the central role the semantic interpretation function plays in the whole issue of representation. One of the reasons why imagistic representations appear so natural is that they can literally resemble the objects they depict, just as we might think of the contents of our recollections as resembling the recalled
situation. But this can be paraphrased as saying that the relation between images (or pictures) and their designata is clear when the semantic interpretation function in both cases is nothing less than the whole of intelligent human perception. That this way of characterizing representations is plagued with difficulties has been amply discussed by Wittgenstein (1953), Goodman (1968), Dennett (1969), Fodor (1975), and others, so I shall pass up the opportunity to add my comments. However, there are also more subtle errors based on tacit assumptions regarding the nature of the interpretation function. For example, there are frequent claims that certain kinds of information are "directly available" in an analogue representation and need merely to be "read off," as opposed to being computed from a descriptive representation. But as we have seen above, such claims are not about the merits of one form of representation as opposed to another, but about which aspects of a situation are explicitly built into the representation in advance and which types of operations are primitive in the semantic interpretation function.

It has long been recognized in computer science that there is a trade-off between the complexity of data structures and the complexity of algorithms for processing them. For example, at one extreme one of the simplest forms of representation is a list of propositions in the predicate calculus. However, extracting answers from such a representation requires a combinatorially explosive theorem-proving system. At the other extreme are some exhaustively cross-referenced data networks from which most of the more frequent questions can be answered by pattern-matching and graph-processing techniques. The difference is that in one case the work is done when new information is entered, whereas in the other case it is done at the time information is retrieved. For the psychologist, choosing some intermediate ground between these is at least partly an empirical issue, since he wishes to model the accessing complexity exhibited by human cognition. The trap here is that representations appear "natural" in proportion to the intelligence attributed to the accessing function. The most natural representation (the picture in the head) requires a full-fledged homunculus for its interpretation. Few psychologists would opt for this alternative. Next in attractiveness comes the wholistic analogue. What type of interpreting func-
tion this requires is not clear, but one that is sometimes hinted at would simply compute some similarity metric, such as implied by Quine’s “quality space,” or a function that recognizes something like Wittgenstein’s “family resemblances.” Such a function could thus indicate how similar some whole configuration was to, say, a prototypical one. Unfortunately we know nothing about how such similarity metrics can be “wholistically” computed. Even if we dropped the “wholistic” requirement, no one has been able to show general dimensional characteristics of similarity. In fact what evidence there is, such as the failure to find dimensions of generalization or dimensions of similarity in multidimensional scaling of structured stimuli (Shepard, 1964), suggests that a dimensional approach to similarity is probably doomed to failure.

As we depart from these direct or analogue representations and build more complex articulated descriptions, we find we can get away with somewhat better understood symbol-processing algorithms. What we lose in naturalness of representation we gain in approaching realizable systems. Since we are still far from an adequate overall model of imagery, it is not a closed issue as to whether we will eventually run into fundamental difficulties. But at least the problems are out in the open—in all their unnatural nakedness—where they can be examined, rather than hidden in metaphors, such as that imagery involves perception.

7. Properties of the Medium

In the first place, I declare to you, sir, that when one has only confused ideas of thought and of matter, as one ordinarily has, it is not to be wondered at if one does not see the means of solving such questions.

Leibniz

*New essays on the understanding*, 1704

The second rug under which people have attempted to hide some of the puzzle of representation has been the representational medium itself. This approach is often taken in attempting to account for certain mental operations performed on representations. Before describing this approach I should like to describe a problem known as the “frame problem,” which researchers in artificial intelligence have studied in the context of robot-planning, since it illuminates a relevant point.
Suppose we have a robot that has perceptual and motor capabilities and can be directed to move about, grasp objects, and generally follow a simple sequence of commands while observing what is going on around it. Such a robot has no difficulty with inconsistencies in its world model, since it merely observes what happens and updates its knowledge base. Sooner or later we would want to be able to give the robot more general goals that would require it to plan out an effective series of actions in advance. This, it turns out, is a qualitatively very different task from the one it has been performing. For now there is a problem of consistency. After each planned action the robot must, in effect, recompute its representation of the entire state of the world, since it must take into account all the possible effects of the action on every aspect of the environment. Such a recomputation may in fact involve referring to the laws of physics. The problem of reasoning about actions, in contrast with merely acting, gets us into a very difficult set of problems stemming from the interdependence of actions. A number of approaches to this so-called frame problem have been proposed (see Simon, 1972; McCarthy & Hayes, 1969; Raphael, 1971). All of them appear somehow to be unduly complex and unnatural. It is clear, however, that part of their unnaturalness rests on the fact that a great deal of knowledge must explicitly be brought to bear in reasoning about actions that we are not aware of using and that indeed we may not have to use when we operate directly on the world. In the latter case relevant interactions are given to us for free by the environment. In the case of reasoning, however, the relations are not free. We must in some way explicitly build in the knowledge regarding what effects do and do not follow from any action.5

It seems to me that the notion of an analogue representation medium is in part an attempt to get this information for free again. Consider the claim that data on the time-course of mental rotation (e.g., Shepard, 1975) argues that the process is analogue (since, as some people have been known to ask innocently, “How can you rotate a data structure through its intermediate positions?”). This carries the implication that once we start a rotation the medium will take care of maintaining the rigidity of the total pattern and carry along all the parts for us—just as the laws of physics take care of this for us in the real environment. But, as in the frame problem,
we are overlooking the fact that the person must know what will and will not happen to the bottom part when the top part starts to rotate. In a descriptive structure this is precisely what makes “mental rotation” appear awkward and computationally unduly costly. But this is unavoidable unless we have an analogical modeling medium which intrinsically follows the laws of physics.\(^6\) Unless we are willing to ascribe such laws to brain tissue (which, to some extent, is what Gestalt psychologists attempted to do), we are stuck with locating knowledge of such laws explicitly in some part of the total representation or in what I have called the semantic interpretation function (which does not, incidentally, preclude such knowledge from being a distributed computation attached to the data structure itself). If we admit this, however, we lose one of the main attractions of the “analogical medium” gambit. For now actions such as rotations must be accounted for by cognitive operations that are themselves not prima facie analogue, since they must in turn refer to knowledge about what happens to forms under certain transformations (we shall return to the notion of image rotation in the next section). Observations of children by Piaget as well as the experiments by Howard cited in section 5 (1) show that when such “operational knowledge” is not available, imagining actions does not lead to veridical conclusions, the supposedly analogical nature of the representation notwithstanding.

The phenomenon of attributing to the intrinsic nature of a representation some of the crucial aspects that need to be taken into account (because these are so intuitively obvious to the theorist) is not confined to analogical representations. Woods (1975) has recently shown that we frequently commit the same oversight in the case of semantic networks. For this reason it is important to attempt to simulate a significant portion of cognition by machine (although even here the existence of such built-in functions as an arithmetic processor may create the illusion that we get magnitudes for free—i.e., we need not account for how they are mentally represented).

8. What Is Rotated in Mental Rotation?

The mental rotation example raises a number of other related problems worth exploring. Let us suppose that an empirically adequate computational model of some cognitive process—say, for compari-
son of rotated forms, as in the Shepard (1975) experiments—is developed. What then would be the status of a description of the comparison process which used phrases such as “the image is mentally rotated”? There are several ways of approaching this question. One is to say that no rotation in fact takes place, since all behavioral data are accounted for by a model which contains no rotating entities. The only thing conceivably left to explain is why the subject reports “rotating an image.” This question might then be approached by an analysis such as that provided by Dan Dennett in his paper in this volume, in which he explores the source of introspective reports about cognitive processes.

A second approach is to say that “rotation” is the name we give to the result of a certain subprocess within the model that at a more microscopic level may or may not be carried out by a discrete symbolic computation. In other words, the “mental rotation” account is simply a description in a higher level language of a certain computation that takes place in the cognitive system. Furthermore, it could be argued that this is not just any arbitrary higher-level description but one that is particularly appropriate because (a) it is consonant with subjects’ reports of what they do and (b) it alone accounts for the empirical constraints on the transformations applied to the representation—i.e., of all the logically possible ways of solving the comparison problem by transforming the symbol structure that is the mental representation of the stimulus, only the one describable as “rotation” (or, in other contexts, as “enlargement” or some such equally plausible pictorial manipulation) offers a natural account of the empirical regularities. Thus “rotation” is more than a convenient global description of the computation involved; the term has additional explanatory power because it captures the significant generalization, as the linguist would say, which underlies the empirically observed, as opposed to the logically possible, transformations.

In fact one might even cast the computational model or simulation in a higher-level language that had ROTATE as one of its primitive operations. This would make the computational model, in a sense, isomorphic to the imagery account (although presumably more complete in its detail and not necessarily analogue in any strong sense).
Such a defence of the "image rotation" account, if correct, would reduce the distinction between the imagery approach and the artificial intelligence approach. It would not entirely eliminate the distinction, since the latter group, not satisfied with an explanation that rests on the statement that comparisons require a subject first to "rotate an image," demands a more complete explication of the entire process. It would, however, eliminate some of the arguments over what is meant by words like "rotation" used in a technical theoretical sense.

Unfortunately such a translation from the language of images and rotations leaves some residual difficulties. In particular there are reasons for resisting the use of global operations like ROTATE as computational primitives. Presumably any operation that is a computational primitive need not be decomposed (i.e., no new understanding of the underlying psychological process is gained by expressing the primitive operation in terms of still smaller steps, even though such an analysis may be required in order to get it to execute on some particular device or perhaps to relate it to neurophysiology). But if the operation is to be treated as a single computational step, then surely the amount of computational resources (time and memory capacity) the operation uses must be independent of the context in which it is used. In particular the amount of computational resources used by a primitive operation should be independent of the representation to which it is applied.

Thus it should take $t$ seconds to "rotate" a representation by $\Theta$ degrees regardless of what it is a representation of. If this were not the case, then we should further want to know what made one stimulus faster and another slower to rotate, which would be tantamount to asking what process takes place within the primitive operation ROTATE.

This approach reflects a general phenomenon in cognitive psychology. In constructing theories we often have the option of either postulating a large number of independent processes or else postulating a smaller number of more primitive interacting processes. In the former case variety of computations is accomplished by a variety of elementary processes, whereas in the latter case the variety comes from the way a few primitive processes interact. Given the option, the latter approach is usually preferred as providing the more power-
ful explanation—i.e., as being able to capture more significant generalizations with fewer theoretical entities.

Although such discussion rapidly gets into some deeper issues concerning the appropriate level of description of psychological processes, which cannot be discussed in general terms in this paper, the particular example of rotation should be clear enough. If the empirical evidence were compatible with the existence of a primitive cognitive operation for rotating a percept or an imaginal representation (which proceeded at some fixed rate of so many degrees per second), then it would be useful to speak of image-rotation as a description of part of the cognitive process. If, on the other hand, we have to qualify this description by saying that the cognitive process appears to be like a rotation of 360° per second for this kind of figure but of 60° per second for that kind of figure, or that some parts of a figure behave as if they were rotated but others do not, or that only certain kinds of figures can be subjected to rotation or only certain kinds of properties are contained in the rotated figure, then we have lost the most important currency this term had. It then no longer acts like a primitive cognitive operation, since we are forced to expose the underlying computations covered by the term in order to account for the qualifying conditions. In addition it no longer has the virtue of distinguishing between permissible and impermissible transformations on representations, since clearly more is going on than simple rotation. So the usefulness of the image rotation account turns on a set of empirical questions.

Although there is not a great deal of evidence bearing on the kinds of possibilities raised above, there is some reason to believe that the hypothetical examples cited may very well be the case. In the first place, the ability to “mentally rotate” a presented figure clearly depends on the nature and complexity of that figure. For example, people find it almost impossible to recognize faces from inverted photographs (Rock, 1973) by performing a “mental rotation.” The difficulty here does not seem to be associated with such geometrical attributes of the figure as its extent or the number of its components or attributes, as one might expect if the percept were thought of as some sort of iconic display. Numerous experiments have shown that the ability to recall a display (Chase and Simon, 1973), to construct an image mentally from a description (Moran, 1973), or to
synthesize a composite figure mentally from component subfigures (Palmer, 1974) depends on the conceptual or descriptive complexity of the figures. Although I know of no published studies showing that the ability to manipulate images (e.g., by rotation) depends on conceptual complexity, this would certainly be a reasonable expectation, given, for example, the instability of complex images and the variable difficulty in retrieving different kinds of information from apparently clear initial images that are subjected to different transformations (e.g., in such tasks as the Guilford Spatial Visualization test—see Baylor, 1972).

In the second place, there is reason to believe that apparent rates of mental rotation depend on conceptual complexity. Metzler and Shepard (1974) found that line drawings of simple three-dimensional figures were rotated at only 60° per second, whereas Cooper and Shepard (1974) obtained a rate of 360° per second with letters of the alphabet. Hochberg and Gellman (1977) also report evidence that the apparent rate of rotation does depend on figural complexity, and in particular on the presence of “landmark features” in the figures. In fact, where salient landmark features are absent (as with certain patterns of filled and empty dots), no evidence of rotation is found (Hochber & Gellman, 1976). I have also obtained data (Pylyshyn, 1977) showing that apparent rates of rotation of line drawings are sensitive to such factors as practice and the type of discriminations that are to be performed on the rotated figures, suggesting that what passes for rotation in such experiments is not simply a rigid angular transformation of a gross iconic image.

There are, of course, many other proposals that could be made to account for data such as those cited here. These include proposals for various hybrid models involving iterating over features of the figure using some kind of relaxation method. Such iterations could involve small local rotations, rotations of a skeletal frame followed by partial reconstruction of the figure, or even iterations over descriptions with no obvious analogue (in the sense of continuous spatial function) processes involved. In the absence of a model worked out in detail, as well as of additional experimental analyses of factors affecting rotation, it is not clear how such proposals would fare. In any event it would seem that a major part of the evidence cited in support of “mental rotation” will be accounted for by computational
processes of various kinds and not by properties of some analogue medium.

In conclusion let me reiterate that I do not claim to have made an argument against analogical modes of representation, and still less am I satisfied that semantic networks, procedures, etc. are adequate to handle all forms of knowledge. I have simply tried to argue that many of the reasons people have for jumping on the "analogical" (whatever that may be) bandwagon are insufficient. Furthermore, we are so far from understanding the semantics of discrete data structures (as Woods has cogently argued) that any mass movement to abandon them, or even augment them with something radically different is at the very least premature.

Notes

1. One might respond to Kant's objection that "concepts without percepts are blind" by pointing out that (a) 'concepts' in our sense can refer to an equivalence class of transducer outputs - i.e., they may correspond to perceptual patterns; and (b) nodes need not be iconic or sense-resembling in order to represent percepts (see the last paragraph of section 4 for more on this point).

2. It is also worth pointing out another advantage of thinking of such figures as being represented by structured descriptions. This approach resolves an old psychological puzzle of why figures' shapes remain perceptually invariant, for both adults and children, when we view them lying down or with our heads at an angle: if the figures are described in relation to their background the description remains the same.

3. This is not to suggest that no useful distinction can be made between analogical and Fregean or between pictorial and discursive representations. The point here is that one cannot make the distinction by simply examining the representation itself. One must, in addition, know something about how it is being used or interpreted. In fact the notion of isomorphism between representations is not a useful one. A more useful notion of isomorphism is the one that appears in algebra - i.e., isomorphism between systems.

4. When it is sometimes claimed that a painting or sketch can be abstract, this surely means that it can be interpreted to bear an abstract relation to the object depicted. The picture itself is never abstract or vague. But this simply means that two-dimensional displays can sometimes, in certain respects, do the work of descriptions.

5. This is not to suggest that people solve the frame problem as it is described. In fact there is good reason to believe that our ability to plan, anticipate, etc. is rather limited precisely because we cannot bring all relevant facts to bear. The point is merely to argue that when we do anticipate successfully we bring to bear multifarious knowledge, including tacit knowledge of physical laws. In some cases we can design representational structures in such a way that certain consequences appear to follow without explicit appeal to stored principles. For example, by choosing a list representation for objects related by a total ordering and by examining the list serially we seem to obtain the transitivity property of such relations as a by-product. Finding such representations is an important goal in building efficient computational models. From a psychological point of view, however, it should be noted that (at least in this example) a commitment to transitivity is made along with
the decision to place particular objects on a particular list, and this decision (and hence the representation of transitivity) may then simply fall outside the domain of what is being modeled. But in general we shall want to model the implicit knowledge of such principles. The need to model intellectual structures explicitly arises because of the kind of independence of properties of thought and properties of the world that we see in cognitive development in children as well as in the nonveridical nature of perceptual schemata discussed in section 5 (2).

6. The more general problem, one that vexed Leibniz in the above quotation is that the most tempting way to represent property P is to attribute P to the representation. However, when P is a physical predicate and we are dealing with mental representations, we must guard against reifying the physical world in the mind. The seductiveness of applying physical predicates (e.g., those pertaining to physical magnitudes) to images appears to be almost irresistible. For example, in a recent response to my critique of imagery, Kosslyn and Pomerantz (1977, p. 13) begin by carefully noting that images themselves are neither large nor small but that they only “register size in the same way that the corresponding representations evoked during perception register size.” Now this unexceptionable position does not itself say anything about the form of the representation. However, in each case in which they find that “imaginal” accounts are more “natural” than “propositional” accounts, this is the case precisely because a literal interpretation of terms like “large-small” or “near far” is being applied to images projected onto a hypothetical screen. The Kosslyn and Pomerantz paper is one of most carefully argued expositions of the imagery position, one that emerged after considerable communication with the author of the present paper. The reader is invited to consult the Kosslyn and Pomerantz paper for a revealing sample of how precipitous the “analgoical” slope can get and how arguments in psychology can slip past one another in recycling classical philosophical puzzles.

7. Whether or not even at a pretheoretical level, the relevant phenomena are best described as “rotation” or something else, such as serial piece-by-piece analysis of where relevant portions of a figure would be were a rotation of the object actually carried out, is an empirical question. Although adequate fine-grain data bearing on this question are not available at present, such tentative evidence as introspective reports (e.g., gathered in our laboratory) and preliminary eye-movement evidence mentioned by Metzler and Shepard (1974) suggest that serial scanning and recomputation may in fact be a better description of the processes occurring in the comparison experiment than wholesale rotation. Recently detailed monitoring of eye-movements by Just and Carpenter (1976) has confirmed that there are several distinct phases to the rotation task, many of which clearly involve piecemeal search and comparison operations. In certain cases, however, such as closing one’s eyes and imagining a rotating object, rotation may be the appropriate phenomenological description, so we shall stick with this for the time being.

8. There remains of course the serious methodological problem of empirically estimating the computational complexity of an operation. Presumably response latency arises from several sources that may interact in various ways. This, however, is a problem for everybody’s theory, and we are here simply taking the current first-order view that variation in reaction time directly reflects changes in computational complexity of the primary operation. The next step would involve a theory of how such factors as attention and memory load interact with latency—i.e., a theory of computing under limited resources.

References


Hanson, V. R. *Patterns of discovery*. Cambridge: Cambridge University Press, 1953.


Kosslyn, S. M. & Pomerantz, J. R. Imagery, propositions and the form of internal representation. *Cognitive Psychology*, in press.


Pylyshyn, Z. Factors influencing the apparent rate of “mental rotation.” Research Bulletin no. 412, Department of Psychology, University of Western Ontario, 1977.


