Published monthly.
Subscription prices, per volume: Institutions $ 70.50, Individuals $ 22.00.
Second-class postage paid at New York, N.Y. USPS No. 043-250.
U.S. Mailing Agent: Expediters of the Printed Word Ltd., 527 Madison Avenue (Suite 1217), New York, NY 10022.
Synthese is published by D. Reidel Publishing Company, Voorstraat 479–483, P.O. Box 17, 3300 AA Dordrecht, Holland, and 190 Old Derby Street, Hingham, MA 02043, U.S.A.
Postmaster: please send all address corrections to: c/o Expediters of the Printed Word Ltd., 527 Madison Avenue (Suite 1217), New York, NY 10022, U.S.A.

LUCIA VAINA

FROM SHAPES AND MOVEMENTS TO OBJECTS AND ACTIONS

Design Constraints on the Representation

INTRODUCTION

In the ordinary pursuit of everyday life, humans display a great deal of knowledge about the world. We drink fom a stream, a cup, a glass, but we do not attempt to drink from a tree, a lamp, or a stick. We can not only recognize actions such as kicking, throwing, running, and so forth, but we can execute them ourselves and we anticipate their consequences. One of the most fascinating questions in the study of the brain is how does this come about? How is general knowledge about objects and actions represented in the brain? What are the information processing problems that are involved in these tasks, and that constrain the organization of such information? What seems so direct and effortless as acting appropriately in the world, turns out, on close consideration, to involve many rapid and complex processes, the details of which we are only beginning to glimpse.

One of the difficulties in embarking on the study of such questions has always been where to start from and what exactly to aim for. Are there any symbols with which we can assume confidently the brain deals and whose organization we can study? Until recently the study of language has provided almost the only route into the problem: words are undoubtedly manipulated by the brain, and investigations into this area (known as models of semantic memory or knowledge representation systems) whether from psychology or artificial intelligence have had a pronounced linguistic flavor. However, our representation of words already embodies specific knowledge about objects and actions in the world. This work is a first step towards the understanding of the nature of cognitive representations in that it starts from a more simple representation of the world, one which is nearer the raw material that our perceptual systems provide us with. Thus the question I shall ask here is this: What information must be attached to sensory information that will be relevant to the recognition of objects and actions, and what are the design constraints on

Synthese **54** (1983) 3–36. 0039–7857/83/0541–0003 $03.40

the representation of this information? The discovery of the constraints the world puts on a representation is important in that it provides an unequivocal answer to the fundamental question: what is represented and why is the representation that way?

While single modality representations provide information about shapes of objects and their movements, noises they make, and so on, they do not provide any explicit information about their possible functions. Thus, for example, if one is in a forest, by just hearing a noise, even as detailed a description as the auditory system can provide for it, one would not know that it is a bear, and the bear might attack; or, however detailed an analysis of the shape of a rock that vision might offer, we do not know only from vision that it could move when kicked. More generally, while the analysis remains in the domain of any particular sense, it cannot encompass information about use, purpose and function. However, the exigences of the real world demand that such information becomes available to the perceiver as rapidly as possible.

I shall examine the representational problems posed by adding simple notions of use, purpose, and function to the analysis of sensory information. (Here by *simple*, I mean having a direct reliance only on perceptual processes).

The input information is restricted here to vision. My purpose is to derive from the analysis of shapes of their movements cognitive properties that characterize objects and actions. I have chosen the visual modality for the following two reasons. Firstly, vision is one of the most important systems used in interacting with the world; in many animals, and in humans, vision is the process of discovering from images what is present in the world. Thus, vision can deliver quite quickly and accurately information about shapes of objects, their spatial organization, their texture and color. Secondly, the work in Vision of David Marr and of his co-workers at the MIT Artificial Intelligence Laboratory provide an adequate framework and a reliable starting point for the input to the Functional Representation. (As we will see later the functional representation relies quite directly on Marr and Nishihara's work on representing static shapes, and Marr and Vaina's work on representing moving shapes.) Thus starting from purely visual information, I shall be asking what are the recognizable (1) *actions* – purposeful movements whose consequences can be visually identified such as KICK, SALUTE, THROW, and (2) *objects* – visually describable shapes grouped by their function, or use in actions. For example, since a CUP, a BOOK, or a SMALL ROCK can be thrown, or kicked, at some level of abstraction they can be considered to constitute an action determined class such as SMALL-PHYSICAL-OBJECT.

Formulated in a different way, I ask what aspects of visually derived information enable humans to act and use objects for their basic survival needs? Answers to this question provide insight into what information we require from the perceptual system. I shall approach this problem by constructing what Marr and Poggio [56] have called a *computational theory*.

THEORETICAL PRELIMINARIES AND ASSUMPTIONS

In this section I shall describe the theoretical basis of my approach to the problem of deriving and representing functional information. I shall discuss: a) the levels of description of an information processing task, whose distinction is important for the development of a computational theory and b) the assumption of modularity, which allows the isolation of specific information processing tasks.

(a1) *Levels of Description*

Marr and Poggio [56] introduced three levels of abstraction at which an information processing task can be described. A complete understanding of a process requires an adequate description at all three levels.

The first level, the *computational theory*, is primarily concerned with what is being computed and why. On this level our purpose is to derive useful properties of objects and actions from images, and to isolate constraints that are at once powerful enough to allow the information processing task under study to be accomplished and are generally true of the biological world.

The second level, the *algorithm*, is concerned with how the problem defined by the computational theory can be solved. One must choose (1) a specific representation for the input and the output of the process and (2) an algorithm by which the transformation might be accomplished. Here the input representation consists of the set of representations produced by the visual processing of the world. I

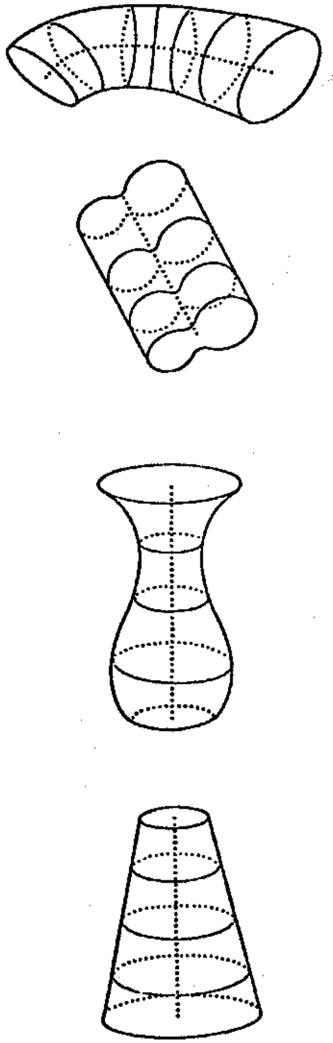

Fig. CONE. The definition of a generalized cone. In this article, it is the surface created by moving a cross-section along a given straight axis. The cross-section may vary smoothly in size, but its shape remains constant. We here show several examples. In each, the cross-section is shown at several positions along the trajectory that spins out the construction.

have restricted the input world to those shapes and movements that can be mathematically described by a generalized cone [7].

I choose this class of shapes because they have well-defined axes, and also many common shapes in the world are included in the scope of such a representation (all animals, many plants, man-made objects, and so forth).

Objects and actions are identified and represented as the output, and the algorithm must explain how this output can be obtained. The algorithm usually depends on both the computational theory and the actual hardware (brain or computer) on which the algorith is supposed to run.

This brings us to the third level, that of the device in which the process is physically realized. The same algorithm might be implemented in quite different technologies. For example, one can imagine humans performing simple arithmetic tasks, such as addition, in a way which is similar to computers, from left to right. But generally a given algorithm is better suited for some devices than others though this does not influence the nature of the problem as described at the level of the computational theory.

Although algorithms and mechanisms are in some cases empirically more accessible, the level of computational theory is the most critical from an information processing point of view. This is because it is crucial to know first what is computed and why that computation is needed for reaching a given goal. In the formulation of the goal psychological data can play an important role by establishing the competence of the human system at a specific information processing task. Once the goal has been formulated, the problem of achieving it can be addressed. The research objective is to formulate a theory that will account for the input-output relations in idealized conditions (Ullman [91]). The critical step in formulating a computational theory is the discovery of the constraints on the way the world is structured. These constraints should provide enough information to allow the processing to succeed, while maintaining as much generality as possible.

(b) *Modularity*

One method for achieving some understanding of a system as complex as the brain, is to isolate the information processing tasks it performs and separately analyze each of them. The following two requirements need to be fulfilled: firstly, the various tasks that the brain carries out must be separated into modular units that can be independently studied. Secondly, each task needs to be described at each of the three levels outlined earlier. The separation of the specific tasks solved by the brain requires the fundamental assumption that the brain is modular. That is to say, it consists of separate systems, each with its own intrinsic structure, designed to handle a particular kind of information processing task, with the whole system interacting in such a way as to create a complex of highly specialized structures. The assumption of modularity of the information processing systems is crucial for research tractability.

These systems can be divided as to whether the analysis they perform is *modality specific* or *modality unspecific*. Examples of modality specific analysis include the task of visual analysis, tactile analysis, and auditory analysis. It is clear that these different types of analysis must be taken at least some way before cross-modal interactions of any complexity could be useful. Clinical evidence from neurology suggests that these analyses proceed a substantial way before their combination ([22], [23], [94], [98]). For example, in vision a sophisticated representation of the shape and disposition of a viewed object can be derived by patients whose realization of the shape's use or purpose is severely impaired. One primary purpose of visual analysis is to facilitate the derivation of structural descriptions, shapes and movements of viewed objects from these images.

Each of these modality specific analyses poses its own self-contained representation problem. Such representations can be called *single modality* representations. Marr & Nishihara's 3-D model representation, which provides an arbitrarily detailed, object-centered description of a viewed object, provides one.

The descriptions supplied by the different single modalities are potentially complex, because they are capable of representing exhaustively all the information that can be acquired by that particular sense. Yet rich as these individual modality specific descriptions must be, we know from our own experience that the comprehension of what we see, touch and hear involves more than each one, and more even than their combination. For example, our comprehension of an object includes knowledge about its various uses and purposes, and of its name. Our comprehension of an action involves knowledge of what objects can participate in it and knowledge of its consequences. These are usually an action's most important aspects from the point of view of recognition and planning.

The organization and representation of this additional information involves a different category of analysis that is not modality specific. Preliminary suggestions about an appropriate form of this kind of representation, together with a discussion of some of the issues involved, can be found in Vaina [94] and Vaina and Greenblatt [95].

CONVENTIONAL APPROACHES

Visual perception and its relation to the structure of the environment and the higher cognitive structures kept busy the minds of psychologists and philosophers for a long time.

Thus, in linguistic semantics for example, many studied what the words or sentences correspond to ([42], [61]). How would one use them to partition the world ([103])? How do we talk about the world? How are words and objects related [67]?

Much progress was certainly made; yet we are still far from the understanding of how objects and actions are represented, and why the representation is that way. The problems with the previous approaches are of two kinds. Firstly, despite considerable efforts over a long period, the representation of *objects* and *actions* remained too dependent on high-level knowledge about the world which presupposes the answers we seek. Thus despite the lack of any precise formulation of its goal, the representation of objects and actions was pursued using the most different techniques employed in Artificial Intelligence or Psychology. The problem was that this representation was viewed as similar to "problem solving" and therefore it involved the testing and modifying "hypotheses" about objects and actions in each particular circumstance considered. Because of their specificity, any of the functional representations that deal with more than a toy-world must command a large number of such hypotheses and must be able to find and deploy the one that is needed. The goal of representing objects and actions is occluded by the additional problems raised by writing efficient programs. The comparison of "goodness" of representations becomes the comparison between the various programs and their control structures.

The second source of the difficulties comes from the input information. Most research utilizing visual or linguistic descriptions of the world invoked specialized knowledge about the scene being viewed in order to describe the objects expected in the scene (a classic example is Winograd's Shrdlu [101]). In visual analysis, the main effort was put into finding "segmentation" criteria for images. Vision seemed more a fancy device responsible for arranging for the right piece of knowledge to be available at the right moment during segmentation.

To *visual analysis* and *functional representation* as well, it can be objected the heavy reliance on the use of higher-level knowledge than that contained in the information that these representations aim to make explicit. The use of such information causes various handicaps in real world situations, making the recognition of unexpected, novel situations difficult. Yet, for this endeavor to be fruitful we must develop a system whose efficiency is close to that of humans, whose ability to cope with the environment goes far beyond their expectancies.

My main thrust is that a suitable *functional representation*, one which will embody an efficient and generalized representation of the world must be conceived independently of any high-level knowledge. The additional knowledge the representation will rely on, beside that obtained from vision, is the general knowledge about the physical world, such as the knowledge that objects exist in three dimensional space or that they are physically connected and so forth.

THE NATURE OF THE INPUT INFORMATION: A COMPUTATIONAL APPROACH TO VISUAL ANALYSIS

In this section, I shall present the computational work in vision of Marr and his co-workers, in order to firmly establish the nature of the modality specific processing which precedes the interpretation in terms of use, purpose, and function.

The starting point for vision is a gray-level intensity array, appropriate for approximating an image such as the world might cast upon the retinas of the eyes. The goal is to obtain a description that is well-suited for recognition of three-dimensional shapes dependent on the input array. This goal, however, is not attained at once. Marr and his group propose several steps for the analysis of an image. They postulated that perception begins with a transformation of the gray level array images into what they called *The Primal Sketch* of the image. The essential underlying assumption for both the design and the interpretation of the primal sketch is the way in which intensities change and the local geometry of those changes. The abrupt changes in intensity in the image reveal contour outlines and hence the shapes of objects in the visual world. The goal of the primal sketch is to provide explicit information about directions, magnitudes, and spatial extents of intensity changes present in the image. The motivation for this representation is that it has information about changes in intensity that is useful for processes such as stereopsis and motion perception, that is the spatial information which is more apparently useful to the viewer. Marr [49] pointed out that in order to generate the primal sketch, the intensity values have to be subjected to the kind of differential and detection analysis that is known to be carried out by the "X" and "Y" cells of the retina, and "simple" cells discovered in the visual cortex ([37], [38]). That is, the retinal ganglion cells represent a non-oriented second differential analysis of the image [67], which is found by the extraction of the overall pattern of spatial variation in light intensity by the ensemble of "simple" cells, with their edge and bar-shaped receptive fields. Marr and Hildreth [53] have worked out the mathematics underlying the computation of the primal sketch from an image. Despite its simplification relative to a gray-level array, the primal sketch of an image is typically a large collection of data. The next computational problem is that of its decoding. The traditional approach to machine vision assumes that the essence of the decoding is a process of segmentation whose purpose is to divide the image (or the primal sketch) into regions that are meaningful physical objects. But this would be an impossible problem to be solved by a "bottom-up" approach, without considering any high-level information. Marr and his group argue that the early stages of visual information processing ought instead to squeeze the "last possible ounce of information from an image before taking recourse to the descending influence of "high-level" knowledge about objects in the world." [55]

Horn [35] pointed out that the principal factors that determine the intensities projected upon the retina of the eyes, are (1) the illuminant, (2) the surface reflectance properties of the object viewed, (3) the shape of the visible surfaces of these objects, and (4) the vantage point of the viewer. Thus if at these early stages the visual processes operate autonomously, it can be expected that only information about these factors can be extracted. Early visual processing must be limited to the recovery of localized physical properties of the visible surfaces of the viewed object, for example, local surface dispositions (orientation, depth) and surface material properties (color, texture, reflectance). Examples of early visual processing are stereopsis (Marr [50], [52], Marr & Poggio [57] and Grimson [27], [28]), derivation of structure from motion (Ullman [91]), texture gradients [88], color, shading ([34],[35]), and so forth. More generally we know that vision provides several sources for information about surfaces in the visual world. Different as these techniques are, they have an important point in common: they rely only on information from the image rather than *a priori* knowledge about the world. The information they make explicit mirrors local properties of visible surfaces at arbitrary points in an image rather than depth or orientation associated with particular objects. Then the computational question that arises is how to go about seeking a representation of the visual scene that makes explicit the information these different processes can deliver. Thus a representation of surfaces in an image is sought that makes explicit their shapes and orientation. Marr and Nishihara proposed a specific representation which embodies this information, called the 2-1/2 D sketch. The goal of this stage of visual processing is primarily the construction of a representation that captures the surface orientation in a scene and tells the viewer (1) which of the contours in the primal sketch correspond to surface discontinuities and should therefore be represented in the 2-1/2 D sketch and (2) which contours are missing in the primal sketch and need to be inserted into the 2-1/2 sketch in order to bring it into a state consistent with the nature of the three-dimensional space. In addition to the surface geometry, the 2-1/2 sketch makes explicit other surface properties such as reflectance, color, texture, and so on.

All these levels, (the gray-level intensity arrays, the primal sketch, and the 2-1/2 D sketch) deal only with the discovery of the properties of surfaces in images. The final component of the visual processing theory concerns the application of visually derived surface information for the representation of three-dimensional shapes in a suitable way for recognition. Many of the representational issues posed by the shape recognition have evolved over a period of time. Thus,

Blum [8] and later Binford [7], recognized the importance of volumetric primitives for shape recognition. Agin [0] worked out the problem of deriving generalized cylinder descriptions from a scene using a laser ranging technique that computed a depth map. Nevatia [65] studied the use of the generalized cylinder specifications computed from the shape's visible surface along with their connectivity for recognition; but he did not make much use of the relative dispositions of their axes in a three-dimensional object centered coordinate frame. Hollebach [32] was the first to really consider the relative spatial relationships in the coordinate system defined by the generalized cylinder axis of a Greek vase; but his work was limited to single axis representations. Winston [102] and Minsky [62] addressed many important aspects of the organization of the representation, yet they did not deal with other design issues. Although all these researches contributed each in its own way to the problem of shape representation, they were far from offering a sufficiently complete solution. Marr and Nishihara [54] and Marr and Vaina [58] extended and integrated the previous work and proposed a three-dimensional representation for the recognition of static and moving shapes. They consider three criteria (as stipulated in [54] that such a representation should satisfy in order to account for the efficiency with which the human visual system recognizes 3-D objects:

Criterion 1 (Accessibility). The representation should be easy to compute from the pictorial image.

Criterion 2 (Scope and uniqueness). It should provide a description for a sufficiently large class of shapes; for each shape within its scope, the representation should provide a description that is unique from *any point of view.* Otherwise, if the description is to be used for recognition, one may encounter the difficult problem of whether two descriptions describe the same shape.

Criterion 3 (Stability and sensitivity). The representation should reflect the similarity between two shapes while also preserving the differences. As Sutherland [89] put it, it is important to be able to recognize both that a shape is a man and that the man is Jones or Smith.

Based on these criteria, Marr and Nishihara considered three aspects of a representation's design: (i) the representation's coordinate system, (ii) its primitives, which are the primary units of shape information used in the representation, and (iii) the organization the representation imposes on the information it describes. They concluded that for recognition, a shape representation (1) should be based on an object-centered rather than a viewer-centered coordinate system, (2) that it should include volumetric primitives, not just the type of surface primitive more easily derivable from images and (3) that it should impose a modular hierarchical organization on the description. These aspects of a shape representation are captured in their simplest form by the 3-*D model representation*, illustrated in Fig. HUMAN.

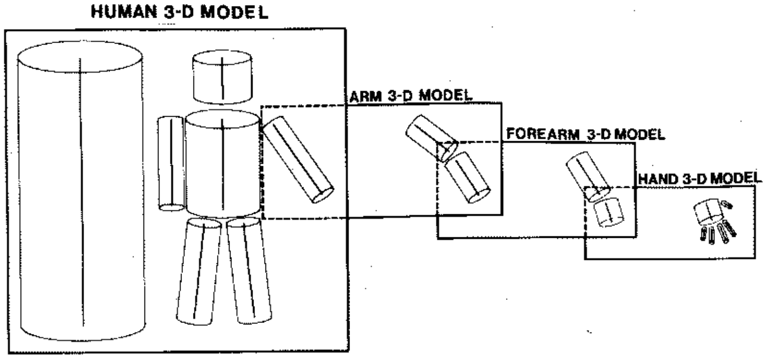

Fig. HUMAN. This diagram, taken from Marr and Nishihara, Figure 3, illustrates the organization of shape information in a 3-D model description. Each box corresponds to a 3-D model; with its model axis on the left side of the box and the arrangement of its component axes are shown on the right side. In addition, some component axes have 3-D models associated with them and this is indicated by the ways the boxes overlap. The relative arrangement of each model's component axes, however, is shown improperly, since it should be in an object-centred system rather than the viewer-centred projected used here (a more correct 3-D model is shown in Figure 2). The important characteristics of this type of organization are: (i) each 3-D model is a self-contained unit of shape information and has a limited complexity, (ii) information appears in shape contexts appropriate for recognition (the disposition of a finger is most stable when specified relative to the hand that contains it) and (iii) the representation can be manipulated flexibly. The approach limits the representation's scope however, since it will only be useful for shapes that have well-defined 3-D model decompositions.

The basic unit of this representation is the 3-*D model*, which consists of two parts. Firstly, an overall *model axis*, (shown on the left of each box in Fig. HUMAN) attached to which is a rough volumetric primitive (the cylinder) describing coarsely the size and orientation of the overall shape represented. Secondly, a collection of *component* axes, as shown on the right of each box, which give more detailed information about the spatial organization of the shape. Each component axis is also attached to a volumetric primitive (a cylinder here), and its location in space is defined relative to the principal axis of the model. The principal axis is the axis that has the most adjoining axes: for the human 3-D model, it would be the torso axis. Much of Marr and Nishihara's article is concerned with how this representation satisfies their criteria. Roughly, the scope of the representation is restricted to shapes that have a natural or canonical axis, as for example, defined by elongation or symmetry or even the gravitational vertical. Uniqueness is achieved largely because the representation is

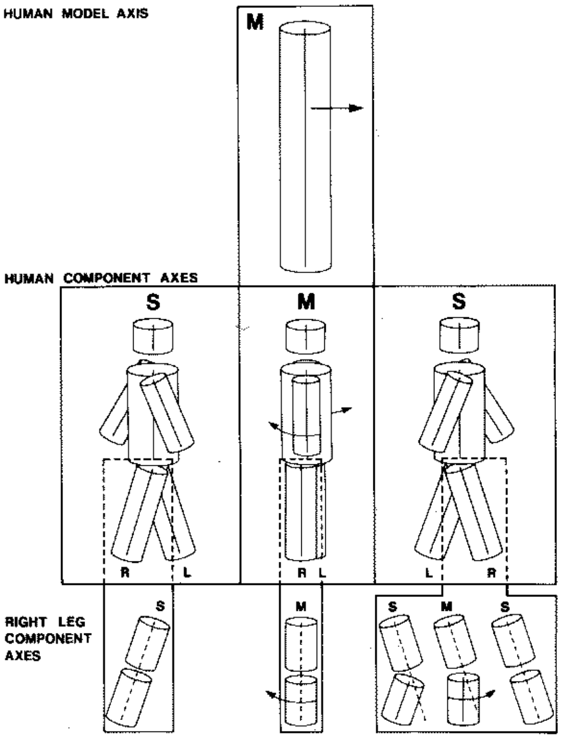

Fig. WALK. This figure represents the walking sequence (S denotes stationary states of the 3-D model and M ones in which motion occurs) at a level at which the shape is only coarsely described.

Three levels of representation are shown: (i) the overall motion of the walk, captured at the level of the human model axis, (ii) the swinging of the arms and legs captured in the motions associated with the component axes of the HUMAN 3-D model, and (iii) the motion of the knees of the non-load-bearing leg by attaching additional FORELIMB 3-D models to the representation, the motions of the feet during a step can be represented in a similar way.

object-centred. The trade-off between stability and sensitivity is accomplished by maintaining a description of a shape at a range of resolutions: to determine whether the shape is a man, one looks at the coarse topmost level; to decide whether the man is a bricklayer or a concert pianist, one looks at the 3-D model of his hands. Finally, although the accessibility issue has not been fully resolved, a start has been made on the problem of how to derive a shape's natural axis from an image (Marr [50]). There seems every reason to hope that as we expand our knowledge of how to derive shape information from images, the difficult problems posed by arbitrary vantage points will eventually yield to analysis. The representation scheme proposed for static shapes forms the point of departure for the representation of moving shapes (Marr and Vaina [58]). They study the problems associated with the representation of instantaneously moving shapes. Then they provide a representation for movements that are extended in time. The essential problem which this representation attempts to solve is how a stream of movement is decomposed into pieces, each of which is described separately. The basic idea is to segment the movement when a component axis, for example an ARM, starts to move relative to its local coordinate frame (the torso). So in figure WALK we see that the movement is divided into a sequence of the stationary states between each swing of the arms and legs, and the actual motion between the stationary points (relative to the torso, not the ground). Marr and Vaina call this representation state-motion-state (SMS).

PSYCHOLOGICAL RELEVANCE OF THE FUNCTIONAL REPRESENTATION

It is widely accepted by now that one of the most salient features of neuropsychology is that it provides methods for partitioning complex cognitive processes into their constituent components. The basic interest to the neuropsychologist is the pattern of dissociation and not the physical structure responsible for it. It is significant only that lesions somewhere selectively affect "syntactic" processing and not the "semantic", or they affect recognition of the function of objects in spite of good perception through single modalities ([22], [23], [1], [96], [97], [98]).

Thus the studies of selective deficits of visual recognition contributed a great deal to the basic hypotheses of this research. Lissauer [47] was the first to distinguish two stages: the act of conscious perception of a sensory impression – the apperceptive stage of perception- and the act of connecting the content of perception with meaning – the associative stage of perception. This dichotomy between perceptual processing and semantic processing has to some extent been confirmed by quantitative investigations of groups of patients with unilateral cerebral lesions. Patients with right hemisphere lesions were impaired on tasks which maximize perceptual analysis (e.g. colour matching, complex pattern matching), whereas patients with left-hemisphere lesions were impaired on tasks with a greater semantic component (e.g. matching colour to object, matching pictorial representation to object). The level of perceptual categorization is independent of language and verbal hypotheses and is also independent of semantic categorization. Evidence from single case studies show that perceptual categorization can be achieved independently of meaning or the significance of photographs of objects. Interestingly, experiments have shown (Warrington [97]) that severely agnosic patients had no greater difficulty in identifying unconventional views of objects than conventional, prototypical views, or patients who were totally unable to categorize visual objects semantically had no difficulties with perceptual matching tasks (Hecaen [29]). On the other hand, patients performing poorly on matching-by-physical-identity task, that is to allocate to the same class different aspects of the same object, a conventional view and an unconventional view strengthens the hypothesis that there is a perceptual stage in the object recognition which can independently be impaired. Thus one can speculate some as to the computations used by the system which achieves perceptual categorization. The fact that generally patients with damage to the right-posterior part of the brain are better in recognizing a conventional view of an object than a less conventional one, suggests than an abstract structured description of the object might be stored (Marr and Nishihara's 3-D representation for an object centered description is a plausible candidate). The less efficient the system the less able to tolerate deviations from the prototype.

On the semantic categorization task, which is matching by functional identity, patients with disturbances at the perceptual stage of recognition continue to be impaired, and it was shown that their impairment could not be attributed to any left-hemisphere deficits.

The semantic categorization hypothesis has been advanced in the context of studies of visual agnosia (associative agnosia). In visual agnosia, defined as the failure to identify objects by naming or functional description in spite of adequate processing at a sensory level, investigation of semantic categorization skills has brought to light interesting phenomena ([97], [29], [96], [46], [47]). The key point is that agnosic patients are able to access some levels of semantic information, but that information is insufficient for the object's precise identification. Not only are they unable to name an object and demonstrate its use, but they also cannot remember ever having seen the object before. Poor recognition is usually limited to the visual sphere and appropriate responses occur when the patient is allowed to handle the object or hear it in use.

One could argue that perceptual categorization and semantic categorization are serially organized, because as we have seen in the discussion above, successful perceptual categorization does not require accurate semantic categorization; but conversely, impaired perceptual analysis preclude accurate semantic or functional categorization.

The studies of the disturbance called by Lissauer associative visual agnosia and of the deficits in perceptual categorization provide much insight in the nature of the representation whose goal is object recognition. For actions, the best place to look is the way in which recognition of gestures and pantomime is done by the brain damaged population. Most interesting is the fact that the literature on aphasia from the middle 1800s includes references on the ability of aphasic patients to indicate, by means of pantomime, their awareness of certain things which they cannot verbalize, as for example how to use an object which they cannot name. The brief historical review which follows attempts to bring into focus the relevant issues. Impaired pantomime recognition as a correlate of aphasia has been noted frequently. Finkelburg [20] was the first to attribute this nonverbal defect to "assymbolia", a general disturbance of symbolic thinking in which verbal and nonverbal activities are equally impaired. Similarly Jackson [39] and Head [30] viewed pantomime recognition defects as being the result of a determinant conceptual system which also affected linguistic functioning. Although there is a long standing controversy about whether there is an autonomic prelinguistic conceptual system or not, I shall not discuss this issue here. I adopt Finkelburg's point that gestural or pantomimic disturbance is a defect in the symbolic representation of movements without a coexisting disturbance of movement per se. In a similar vein Liepmann [46] puts forward the idea that there is a dissociation between the idea of the movement and its motor execution. On imitation some movements might be restored but the whole performance is still defective. This includes difficulties in the purposeful manipulation of real objects as well as actions with pretended objects. In an article that is classical by now, Goodglass and Kaplan [25] observed that aphasics frequently used their hands or their fingers to represent an intended object, as in "hammering" with the fist as the hammer. This phenomenon, termed Body-Part as Object was also observed to predominate in children in preschool age. This observation gives us support for the existence of a prelinguistic Functional Representation of objects and actions, in which objects are categorized by their use in actions and not by their appearance. This phenomenon of Body Part as Object is a quite novel and unpredicted by-product of the lessening of the gestural ability. Goodglass and Kaplan make the supposition that using the body part as an object aphasics evade the difficult task of reproducing a movement sequence outside of the concrete context which ordinarily elicits it. It offers the reality of acting on an object and we can conjecture that it permits a more vivid experience of the affective experience of the pretended action. However, the main thing to be retained from this for the Functional Representation is the hypothesis that in the brain there is such a module that deals with the representation of uses of objects in actions and with actions categorized by their consequences. This module, I believe, although closely interfaced with a module which categorizes objects and actions perceptually, is separate from it and can function on its own.

DESIGN CONSTRAINTS ON THE REPRESENTATION OF OBJECTS AND ACTIONS. ASSUMPTIONS FOR THE DESIGN OF CONSTRAINTS

The functional representation will have primitives such as *actions* and *objects* that can act or be acted upon. We proposed to call our representation *Functional Representation* (*FR*). Constraints upon this representation are based on the requirement that its goal is to provide descriptions of objects, actions and their interrelations that efficiently serve the needs and purposes of everyday life. These are largely shaped by the external world and one's relation to it at the time.

The first assumption is that from the point of view of the *functional representation* the external world consists of a set of descriptions each written in a representation specific to a different sensory modality. As we shall see this essentially casts the main structure of the representation I propose. The way in which incoming information is expressed to the *functional representation* is determined by the structure of each single modality representation, which in turn will have been shaped by the need to be able to represent and recognize efficiently the type of information for which they were designed.

Whereas the first assumption concerns the nature of the information from which the FR must be derived, the second assumption concerns the information that the FR is designed to make explicit.

The Functional Representation is designed to make explicit the relations between objects, actions, and their consequences for the purpose of efficiently recognizing, or constructing recipes for, actions.

The effect of these two assumptions is essentially to define the FR and its associated information processing tasks as a module whose incoming information is purely sensory and whose output deals with actions and their immediate consequences.

The third assumption is that the FR does not duplicate the detailed information delivered by the different sensory modalities, but rather that easy access to that modality's specific information is maintained. In addition, if more information about some aspect of the sensory information is required, computations can be initiated to obtain it. For example, if it should become important during the analysis of an action to know how the subject's right leg is moving, the request from FR will cause the visual module to deliver a suitably detailed description. This assumption is essentially one of economy; it prevents information from being duplicated unnecessarily and it limits the amount of computation to roughly the amount needed.

CONSTRAINTS ON THE FUNCTIONAL REPRESENTATION DETERMINED BY THE INPUT INFORMATION

These constraints are determined by the three criteria which Marr and Nishihara gave for evaluating the adequacy of a representation. As they have been discussed earlier in the paper I am assuming that they are known, and I shall discuss how they apply to the *functional representation.*

Accessibility. This constraint requires that the description of objects and actions be computed from the 3-D representation. In other words, a way has to be found for an efficient matching between the information computed by the 3-D and 2-1/2 sketch representations and the information manipulated by the *functional representation.* This criterion induces an organization by *parts* in the FR. A 3-D model extracted from an image offers a shape description that is *volume-based, modular* and *object-centered.* Each module is a 3-D model in itself. Looking at Figure HUMAN we see that the complete 3-D model is in fact a hierarchy of 3-D models shown as extending down to the level of FINGERS. Under this hierarchical scheme any component of a shape can be treated as a shape in itself, and thus the final description of the shape can be carried down to any arbitrary level of detail.

These characteristics of the *input representation* suggest the notion of *part* for the FR. This is a well-defined notion, because to each module we uniquely assign a part that directly corresponds to a part of the physical object in the world (because it is volumetric) and does not depend on any particular view (object-centered). For example, the 3-D model LEG from the HUMAN 3-D model is a part in the FR represented by the atom LEG. The parts in the FR inherit the hierarchical organization from the visual description. And, as in the visual 3-D representation, from a part (3-D model) one can recover the whole object (shape).

This important property of organization by parts is based on the physical connectivity of physical objects in the world. For example, if we see on the trunk of a tree four paws and we recognize that they are a bear's paws, we should immediately know that there is a bear there, and that the bear might attack. The categorization by parts is still descriptional. Yet, the goal of the FR requires a functional categorization as well. Thus, the parts should have their own, independent description by use. For example, LEG is a part in FR as it corresponds to the 3-D model LEG delivered by vision, and it is connected with the other parts of the physical object it belongs to. But LEG should have an independent functional representation, characterized by its use for support. In this representation, a chair leg and a mammal leg, for example, should be grouped together. In a more detailed description, some LEGS (of aminals, for example) are used for moving. Thus, whereas the representation by part is induced by the *shape description*, the functional representation of parts is induced by their use in *action*.

Stability and Sensitivity. This criterion asks how well the representation makes explicit the information that really matters. Thus in representing an action like KICK one only has to deal with the part LEG (corresponding to the 3-D model primarily involved in the movement). However, if further information should be necessary for recognizing an action, it can be easily obtained. For example, information about the position of the foot (is the boy kicking the ball toward you, or toward the window?), or the joint angle between the two components of the LEG can be computed when required for the FR. However, the important issue is that this induces another kind of action categorization: namely, categorization by the module that is involved in producing the movement underlying the action. Expressed in the FR terms, this means categorization by the part of the physical object that does the action. But, which 3-D model (part) should we choose when more than one is involved? In LOB for example, the arm, the hand, and the fingers all are involved in the action. Because the parts are hierarchically organized, and the hierarchy goes from general to particular, we can make the convention to always choose the first level in the hierarchy relevant to the action described. This convention is motivated by a principle of efficiency which says to interpret as soon as possible, and by a principle of economy which means relying on the smallest amount of information sufficient for achieving a given task. Both of these principles are corollaries of the stability and sensitivity criterion.

Vision informs us in detail about the shape structure that induces parts and their relations in the FR. Yet we have seen in the previous section that objects participating in actions can also be thrown, kicked, dropped, pushed, and so forth. Different objects can participate in the same action, and the same object can participate in different actions (often even without changing the role). What does this mean for the representation of objects? How can the representation capture the fact that a *stick*, an object with a certain visual

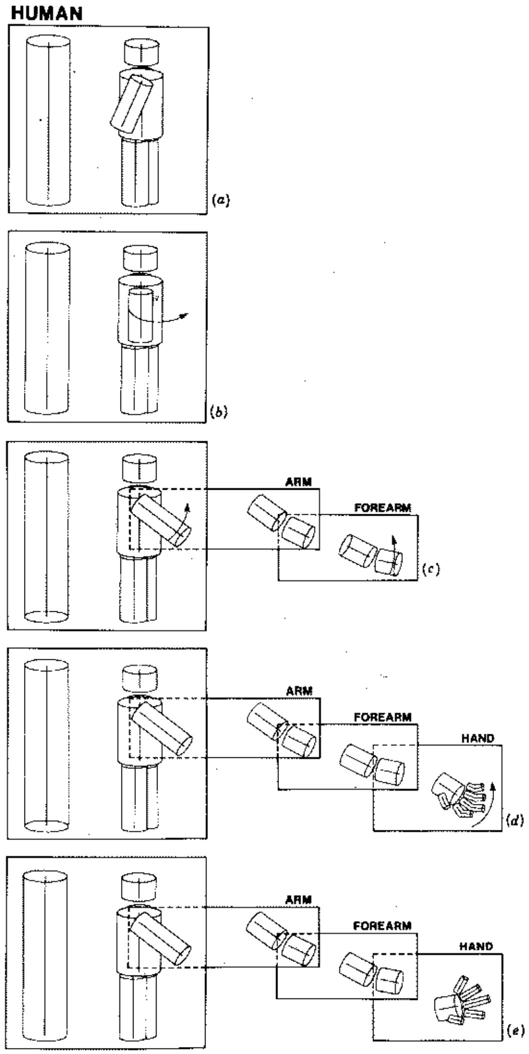

Fig. LOB. The sequence of movements that constitute the overall movement LOB. To begin with, (a) the man is static, about to begin the movement. In (b) the arm swings, but there is no motion at the lower levels of description. In (c) motion appears down in the FOREARM module, the ARM module is still static, and the HUMAN module displays the same motion as before. In fact, we have also included some positional information in the HUMAN module, as well as its motion, for the arm is shown further on in its rotation than in (b), roughly specifying the position at which the hand joint in the FOREARM module starts to move. Next, in (d) the fingers move in the HAND module, as the thrower releases the projectile. Finally, in (e) the movement ends and all 3-D models become static again.

description, can be thrown, burned, or broken? How can the same object participate in actions with such different requirements? From the point of view of the representation, each of these different uses often needs to be expressed in separate hierarchies. For example, a hierarchy will capture aspects of a stick necessary for throwing it (its size, position, shape, weight), while another might deal with combustion (is it wood? is the wood wet?) is it hard or soft? is it damp or rotten?). To answer these questions information only about the shape's geometry would not tell us enough. We need to obtain information from the 2-1/2D representation that tells us about texture, shading, and so forth. This requires a direct communication between the FR and the 2-1/2D representation as well.

Scope and Uniqueness. This criterion asks about the class of *objects* and *actions* for which the representation is designed and, furthermore, whether the members of a class have a well-defined description in the representation. The objects described by FR are those whose shape is captured in the visual representation I have considered for an input.

What are the actions? I propose to derive the actions from the changes that a shape can undergo. The phrase "change of state" has a precise meaning, because the possible changes have to fall within the scope of the underlying visual representation. The changes that a shape admits depend upon its description in the 3-D representation. These induce four classes of actions according to whether the change is in (1) the state (position, orientation, velocity) of the model axis in the model's external coordinate frame (walking, running, rolling, and so forth), (2) the state (position, orientation, velocity) of the components axes of the 3-D coordinate frame (such as throwing, saluting, kicking, nudging, blowing-away and so forth), (3) the parameters associated with the shape primitives in the 3-D model. Examples of these actions would be: squashing, stretching, inflating, bending, etc., and (4) the primitives in the 3-D model by which the representation describes a piece of matter as opposed to simple changes in the value of the parameters associated with the given primitives such as: (a) when the change is due to physical disconnection or breaking of the components of a shape as in actions like break, fractionate, shatter, slice, and (b) when the change is due to assembling sub-shapes together to make a new shape as in glue, combine, and so forth. These categories are not totally distinct, they

may overlap to some extent. For example, if one sands down a peg, one produces a change in its shape of type (3), as it becomes shorter and perhaps thinner. In addition, however, one could view each tiny grain of sawdust as a separate, new 3-D model, created by the sanding process and this puts the action into the class (4). Having carved the classes of actions the representation describes, the next step is to refine them in such a way as to provide uniqueness.

This implies introducing some criteria that allow a discrimination between a THROW and a HURL, for example. Both are actions in the Category (2), they involve the same 3D-model. Both involve an object and have the same general consequence from the object point of view as well, namely to cause the object involved to perform a movement which falls in the category (1). Yet they are different actions. HURL is a kind of THROW but more intense. Can vision capture this difference, and if so, are those the types of differences that the FR needs? Indeed, if one focuses on the arm that performs the action, one could see that the angles between joints are much sharper in HURL than in THROW. For satisfying the condition of uniqueness more information is needed. It is usually obtained from the lower modules involved in the movement (i.e., one looks at the values of the parameters associated with the primitives, and compares them in different cases). This information is transmitted to the FR in form of properties of actions such as intensity and so forth. Although vision deals with various tolerances for the values of the angles between axes instead of the direction numbers, FR is even more approximate when it codifies this information as properties of actions. A further detail can be obtained from focusing on the object participating in the action. Thus, is the particular THROW the one appropriate for a javelin, or for a frisbee?

TABLE THROW

THROW DROP
LOB PITCH
BOWL CURL
SPEAR-THROW HURL
SHOT-PUT DISCUS-THROW
HAMMER-THROW FLING
TWO-HANDED-OVERARM-THROW CABER-TOSS

In the object representation, the different hierarchies induce a multiple description of the object by its use in different actions. The properties of objects derived from the 3D representation or the 2-1/2D representation of shapes are those that inform us about the geometry of the object, such as shape, size, width, thickness, or location and orientation, weight and the material of which it is made. But in an accurate representation, one needs to know more than that. If, for example, I want to move an object, it might be important to know if it would break by throwing, or if kicking a stone might move it, or stepping on an ice cream cake might change its perfect round shape. Thus new properties need to be added to the previously mentioned set. These properties will be derived from (a) the general changes a shape can undergo and (b) the fact that the shape *does* the action (is an agent), or the action *is done* to it. Thus if a change occurs at the 3-D model level, and the shape produces it, the shape is *movable* (i.e., all animals are movable). If the shape undergoes the change, we shall say that it is *removable* (i.e., a rock, a book, but not a mountain). Or if the change occurs in the value of the parameters associated with the primitivies, the object is *deformable.* If the primitives themselves change, the object is *fragile.* Some interesting questions arise: are the properties of an object given in a fixed order or do they depend on the object's use? Does one know off hand that a cup is breakable, regardless of whether one contemplates throwing it? To answer these questions experimental data from neuropsychology will be of relevance [Vaina and Goodglass (in prep)].

FUNCTIONAL CONSTRAINTS ON THE REPRESENTATION

Perhaps one of the most striking features of the way information is organized at the higher level is that humans and other animals are not constrained to purely sensory similarities. The shapes in Figure CHAIRS, are all very dissimilar yet they are all recognized as chairs. All of the items in the Table THROW refer to visually distinct actions, yet they all are different types of *throw*. How and why does this come about? I argue that the critical difference between the types of organization or classification provided by vision and those acting at the next level is that perceptual classifications are organized to admit efficient descriptions of a shape or movement, whereas at the next stage, the primary organization is centered around their *use, purpose,*

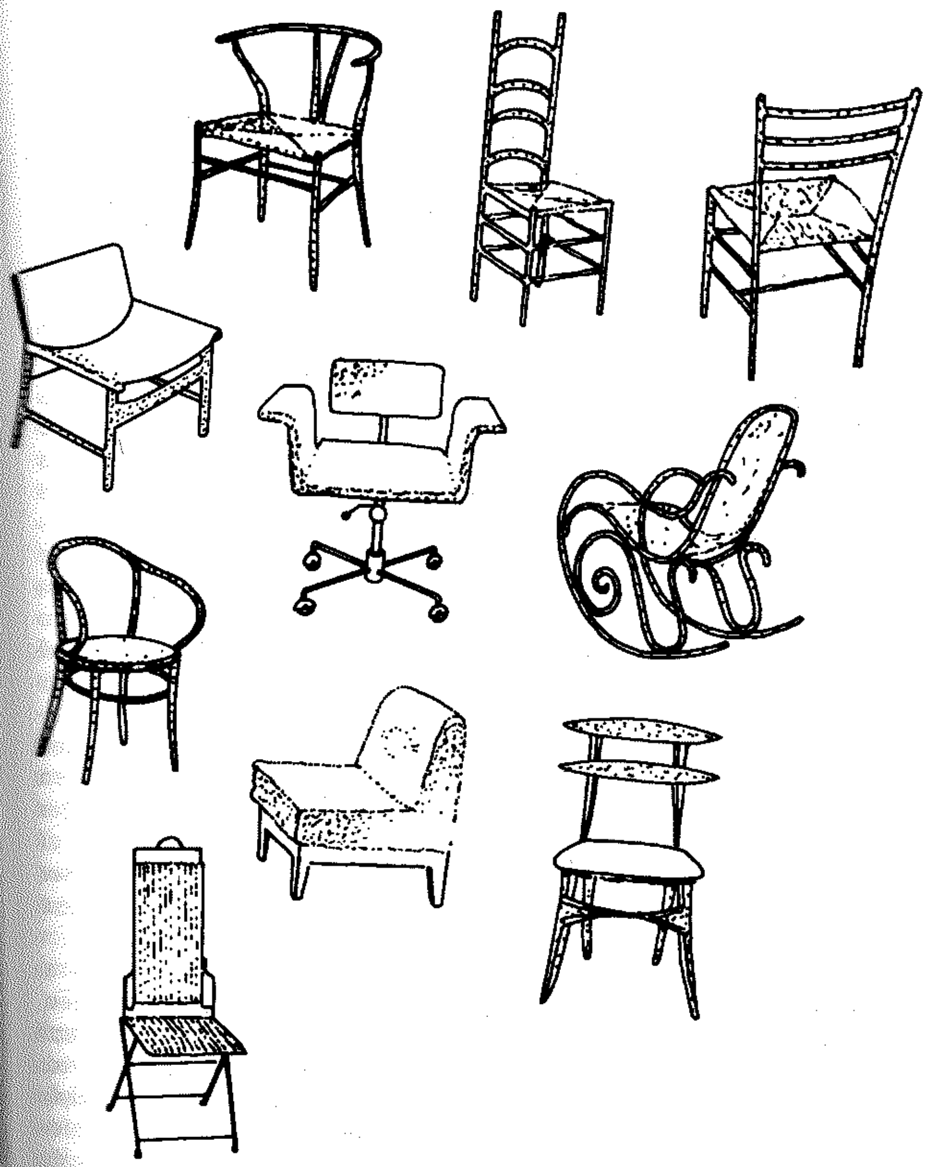

Fig. CHAIRS. What does it take to be a chair?

and *function*. Thus the critical new step involved here is the classification of shapes (objects) by their uses in actions (movements), and the classification of movements (actions) by their application to different objects (shapes).

Even these rather general considerations impose powerful constraints on the way information should be organized in the representation and on how to access it. We shall illustrate this point with the following three examples.

(1) *Categorical Organization and Multiple Descriptions*

In order to achieve flexibility and generality in the representation of actions, and at the same time maintain reasonable economy of computation, the representation of an object must include some guide as to whether or not it can be included in an action. For example, one can throw a ball, a small rock, a small animal, or a stick, but it is unreasonable to suppose that the representation of "throw" should include a precise list of all things that one can throw.

From the point of view of organizing behavior, one needs a more abstract representation of the things that can be thrown, not just a list of items that one can exactly throw. Thus, if the purpose is to represent objects for the efficient use in actions, a critical aspect of the representation of an object is that it be categorized with respect to these actions. Thus one might invent the class SMALL-PHYSICAL-OBJECT initially defined by whether an object can be thrown; and the representation of all subsequently encountered objects will then have to make explicit whether or not they can be included in this category. Further economy is is possible whenever the prerequisites of two actions happen to overlap or coincide: for example, the classes for CARRY and for THROW are very similar, and an efficient representation scheme will make use of this fact.

Thus we see that the requirement for efficient use of objects in actions induces strong constraints on the form of the representation. Each object must first be categorized in several ways, governed ultimately by the range of actions in which it can become involved. We can express this by saying that each object will have multiple descriptions, of a rather unspecific sort. Secondly, the representational machinery for actions on the whole will relate nonspecific rather than specific description of objects, although the process can start with, or can at any moment specialize to, very specific descriptions.

(2) *Access from the General to the Particular*

The second and somewhat surprising example is that the access path between representational items in a memory, whose purpose is the efficient assembly of recipes for action, should be organized from the general to the particular rather than the converse. The reason for this is simple: a critical aspect of the representation of an object is that it includes several "coarse" descriptions of it, ultimately because they are useful categories when assembling actions. For example, a DUCK is described as an animal, a bird, and perhaps a mallard; but we can describe the duck as a SMALL-PHYSICAL-OBJECT, or as FOOD and so forth. New information about a specific object will be attached to the specific levels of such a representation. For example, the name or color of a particular duck will be attached to a level near the DUCK end rather than to FOOD. However, such details are unimportant for the assembly of actions concerning ducks (e.g., killing, cooking). To example them early would merely encumber the processing with unnecessary detail. It is the more general categories that first need to be accessed.

One can see the same idea from another point of view. Suppose that one is looking around for something to throw (perhaps at the duck). Vision can, for example, deliver descriptions of nearby objects to arbitrary levels of detail, but if one wants to know whether the viewed object can be thrown, such detail is only confusing. What is wanted is a level of visual description that accesses the cognitive representation at the level of categorization, (describing whether the object can be thrown or not) and this level of description is relatively coarse.

Thus the primary organization of the access path in an action based representation should lead from the general to the particular. When accessed with a particular visual description (or even some kind of label) the first information to be elicited should be rather general, and only later should very specialized information about the particular object be recovered.

(3) *Collective Physical Objects*

Basic to the representation we are seeking shall be the notion of part of a physical object, that corresponds to a 3-D model in its shape description. Because of the connection of the physical object, its parts are always related. The physical world contains some items that are like physical objects, in that they are characterized by a part-whole relation. However, the connection among them is functional rather than physical. This functional basis for grouping different items together is derived from reasons of efficiency for actions in which they are involved. I shall call this *collective physical objects*. Some examples are room, kitchen, tool box and sŏ forth. (A kitchen, for instance, contains various objects used in cooking, such as pots, pans, perhaps a refrigerator, and a stove.)

DISCUSSION

In approaching the representation of objects and actions as a problem in Information Processing, I have stressed two points. Firstly, the nature of the understanding we seek should be clearly stated at the level of a computational theory. The critical act in formulating computational theories is the discovery of valid constraints on the way the input world is structured – constraints that provide sufficient information to allow the processing to succeed and to clearly define the goal of the representation. The discovery of constraints that are valid and universal leads to results about the representation that have the same quality of permanence as results in the natural sciences.

There are two kinds of constraints on the Functional Representation. One set of constraints on the Functional Representation comes from the constraints valid for the specific sensory systems which serve as an input. In other words, the Functional Representation uses as an input the information from the world already processed in the specific way characteristic to the sensory systems which serve as an input (I took Vision as a specific example). The second set of constraints is specific to the Functional Representation, such as grouping objects in *functional categories* for example, which presupposes that the same object can be used in actions which at some degree of generality are semanticaly similar. Another example is that the same object can be used in a set of very different actions each using another aspect of the object; this leads to the representation by *multiple description.*

The second point I have tried to make is that the overall framework for the functional processing uses only that visual information, shapes, and movements that are useful for the recognition of objects and actions. Thus each element of the Functional Representation embodies only the sensory information that is relevant for the use of objects in actions and for the action's end result. The demands of the world surrounding us are such that often it is essential to act very quickly. In other words the functional representation is organized by a principle of efficiency which says: compute as little as possible and interpret as soon as possible. For the representation to satisfy the principle of *efficiency* and to provide *reliability* of recognition it seems natural that the information should be accessed from the general to the particular. Thus, for example, if my goal is to chase away a cat who is too near a pot of sourcream, it is enough to pick up almost any small physical object and throw it at the cat. Whether the object is a fork, a spoon, a piece of soap or a glass it is not important for this purpose: it is enough to recognize it as a small physical object. Yet if my goal is to eat the sourcream (after I have rescued it from the cat), I need to find a spoon; a glass or a piece of soap would not be of any help.

Massachusetts Institute of Technology-Center for Cognitive Science and Boston University-Computer Science

ACKNOWLEDGEMENTS

I specially thank Keith Nishihara for many hours of discussions about this work and Ellen Hildreth for so carefully reading through an earlier version of the manuscript. I thank them both for helping to clarify the language and the thought. I thank Morris Halle, Noam Chomsky, Joan Bresnan, Marvin Minsky, Shimon Ullman, Kent Stevens, Dan Dennett and Patrick Winston for their comments on the earlier versions of this paper. The article describes research done at the Cognitive Science Center of the Massachusetts Institute of Technology under the NSF grant NSF-1 ST 811 6549.

BIBLIOGRAPHY

[0] Agin, Gerald J.: 1972, *Representation and Description of Curved Objects*, Stanford Artificial Intelligence Project, Memo AIM-173, Stanford University.

[1] Albert, M. L., Avinoam, R. and Silverberg, R.: 1975, 'Associative visual agnosia without alexia', *Neurology* **25**, 322–26.

[2] Anderson, J. R.: 1976, *Language, Memory and Thought*, Erlbaum Press, Hillsdale, N.J.

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

[4] Badler, N.: 1976, 'The conceptual description of physical objects', *American Jr. of Comp. Ling.*

[5] Badler, N.: 1975, *Temporal Scene Analysis: Conceptual Description of Object Movement*, Toronto, Department of Computer Science, *Technical Report*, 80.

[6] Bell, A. and Quillian, M. R.: 1971 'Capturing concepts in a semantic net', E. L. Jacks (ed.), *Associative Information Techniques*, Elsevier, New York.

[7] Binford, T. O.: 1971, 'Visual perception by computer', presented to the *IEEE Conference on Systems and Control*, Miami in December 1971.

[8] Blum, H.; 1973, 'Biological shape and visual science' (Part I), *J. Theor. Biol.* **38**, 205–87.

[9] Bobrow,'D. and Collins, C.: 1975, *Representation and Understanding Studies in Cognitive Science*, Academic Press, New York-San Francisco.

[10] Brachman, R.: 1977, *A Structural Paradigm for Representing Knowledge*, Ph.D. Thesis, Division of Engineering and Applied Physics, Harvard University.

[11] Carhamazza, A. and Zurif, E.: 1976, 'Dissociation of algorithmic and heuristic processes in Language Comprehension: Evidence from Aphasia, *Brain and Language* **3**, 572–82.

[12] Cercone, N. and Schubert, L.: 1975, 'Toward a state-based conceptual representation', *IJCAI* **4**,

[13] Chafe, W. L.: 1973, 'Language and memory', *Language* **49**, 261–81.

[14] Collins, A. M. and Quillian, M. R.: 1972, 'How to make a language user,' E. Tulving and W. Donaldson (eds.), *Organization of Memory*, Academic Press, New York.

[15] Collins, A. M. and Quillian, M. R.: 1969, 'Retrieval time for semantic memory', *Journal of Verbal Learning and Verbal Behavior* **8**, 240–48.

[16] Collins, A. M. and Quillian, M. R.: 1972, 'Experiments on semantic memory and language comprehension', in L. W. Gregg (ed.), *Cognition in Learning and Memory*, Wiley, New York.

[17] Craik, F. I. and Lickhart, R. S.: 1972, 'Levels of processing: A framework for memory research', *Journal of Verbal Learning and Verbal Behavior*, 671–84.

[18] Fahlman, S.: 1979, *NETL: A system for Representing and Using Real-World Knowledge*, The MIT Press, Cambridge.

[19] Fillmore, C.: 1968, 'The case for case', In E. Bach and R. Harms (eds.), *Universals in Linguistics*, Holt, Rinehart and Winston, Inc., Chicago.

[20] Finkelburg, F.: 1870, Niederrheinische Gesellschaft, Sitzung vom 21. Berl. Ann. Wschr, VII, 449–59, 460–62.

[21] Gentner, D.: 1975, 'Evidence for the psychological reality of semantic components: the verbs of possession', in D. Norman and D. Rumelhard, *Explorations in Cognition*, Freeman, San Francisco.

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

[23] Geshwind, N.: 1965, 'Disconnection syndrome in animal and man', *Brain* **88**, 237–94.

[24] Goodglass, H., Klein, B., Carey, P., and Jones, K.: 1966, 'Specific semantic word categories in aphasia', *Cortex*, 74–89.

[25] Goodglass, H. and Kaplan, E.: 1963, 'Disturbance of gesture and pantomime in aphasia', *Brain* **86**, 703–20.

[26] Grice, H. P.: 1968, 'Utterer's meaning, sentence meaning, and word-meaning', *Foundations of Language* **4**, 1–18.

[27] Grimson, W. E. L.: in preparation, *A Computer Implementation of a Theory of Human Stereo Vision.*

[28] Grimson, W. E. L. and Marr, D.: 1979, 'A computer implementation of a theory of human stereo vision', *Proceedings of ARPA Image Understanding Workshop*, SRI, pp. 41–45.

[29] Hecaen, H., Goldblum, M. C., Masure, M. C., and Ramiers, A. M.: 1974, 'Une nouvelle observation d'agnosie d'objet. Déficit de l'association, ou de la categorization specifique de la modalité visuelle?' *Neuropsychologia* **12** 447–64.

[30] Head, H.: 1926, *Aphasia and Kindred Disorders of Speech*, Cambridge University Press, London.

[31] Hendrix, G. Thompson, C. and Slocum, J.: 1973b, 'Language processing via canonical verbs and semantic models', *IJCAI* **3**, 262–69.

[32] Hollerbach, J. M.: 1975 'Hierarchical shape description of objects by selection and modification of prototypes', *M.I.T.-AI Laboratory*, TR-346.

[33] Hollerbach, J. M.: 1980, 'A Recursive Lagrangian formulation of manipulator dynamics and a comparative study of dynamics formulation complexity', IEEE Transactions on Systems, Man, and Cybernetics, pp. 730–36.

[34] Horn, B. K. P.: 1974, 'Determining lightness from an image', *Computer Graphics and Image Processing* **3**, 277–99.

[35] Horn, B. K. P.: 1975, 'Obtaining shape from shading information', in *The Psychology of Computer Vision*, ed. Winston, P. H., 115–55. New York: McGraw-Hill.

[36] Horn, B. K. P.: 1975, 'Kinematics, statics and dynamics of two-D manipulators', *M.I.T. A.I. Laboratory*, W.P. 99.

[37] Hubel, D. H. and Wiesel, T. N.: 1962, 'Receptive fields, binocular interaction and functional architecture in the cat's visual cortex', *J. Physiol.* **166**, 106–54.

[38] Hubel, D. H. and Wiesel, T. N.: 1968, 'Receptive fields and functional architecture of monkey striate cortex', *J. Physiol. (Lond.)* **195**, 215–43.

[39] Jackson, H.: 1878, 'On afflictions of speech from disease', *Brain* **1**, 304–30.

[40] Johansson, G.: 1973, 'Visual perception of biological motion and a model for its analysis', *Perception & Psychophysics* **14**, 201–11.

[41] Katz, J. J.: 1972, *Semantic Theory*, Harper and Row, New York.

[42] Katz, J. J. and Fodor, J. A.: 1963, 'The structure of semantic theory', *Language* **39**, 170–210.

[43] Kintsch, W.: 1974, *The Representation of Meaning in Memory*, Erlbaum Press, Hillsdale, N.J.

[44] Labov, W.: 1973, 'The boundaries of words and their meanings', in C. J. Bailey and R. (eds.), *New Ways of Analysing Variations in English*, Georgetown Pr., Washington, D.C.

[45] Lakoff, G.: 1972, 'Hedges: A study in meaning criteria and the logic of fuzzy concepts', in *Papers from the Eight Regional Meeting, Chicago Linguistic Society*, University of Chicago, Linguistics Dept, Chicago.

[46] Liepman, H.: 1900, 'Das krankheitshild der Apraxie (motorische Assymbolie)', *Mtschr. Psychiat* **8** 15–44, 182–97.

[47] Lissauer, H.: 1889, 'Ein Fall von Seelenblindheit nebst Beitrage zur Theorie derselben', *Arch. F. Psychiat. u. Nervenkr.* **21**, 220–70.

[48] Loftus, E. F.: 1974, 'Activation of semantic memory', *American Journal of Psychology* **86** 331–37.

[49] Marr, D.: 1976, 'Early processing of visual information', *Phil. Trans. R. Soc. Lond. B.* **275**, 483–524.

[50] Marr, D.: 1978, 'Representing visual information', *AAAS 143rd Annual Meeting, Symposium on Some Mathematical Questions in Biology*, published in *Lectures on Mathematics in the Life Sciences* **10**, 101–80. Reprinted in *Computer Vision Systems*, ed. A. R. Hanson and E. M. Riseman, 1979, pp. 61–80, Academic Press, New York. Also available as M.I.T. AI. Lab. Memo 415 (1977).

[51] Marr, D.: 1979, 'Visual information processing: the structure and creation of visual representations', *Phil. Trans. R. Soc. Lond. B.*

[52] Marr, D.: 1982, *Vision*, Freeman, San Francisco.

[53] Marr, D. and Hildreth, E.: 1980, 'Theory of edge detection', *Proc. R. Soc. Lond. B.* **207**, 187–217.

[54] Marr, D. and Nishihara, H. K.: 1978, 'Representation and recognition of the spatial organization of three-dimensional shapes', *Proc. R. Soc. Lond. B.* **200**, 269–94.

[55] Marr, D. and Nishihara, H. K.: 1978, 'Visual information processing' Artificial intelligence and the sensorium of sight', *Technology Review* **8**, 2–22.

[56] Marr, D. and Poggio, T.: 1977, 'From understanding computation to understanding neural circuitry', *Neurosciences Res. Prog. Bull.* **15**, 470–88.

[57] Marr, D. and Poggio, T.: 1979, 'A computational theory of human stereo vision', *Proc. R. Soc. Lond. B.* **204**, 301–28.

[58] Marr, D. and Vaina, L.: 1982, 'Representation and recognition of the movement of shapes', *Proc. R. Soc. Lond. B.* **214**, 501–524.

[59] Martin, W. A.: 1979, 'Roles, co-descriptors, and the formal representation of quantified English expressions', in MIT-LCS/TM-139.

[60] Meyer, D. E.: 1970, 'On the representation and retrieval of stored semantic information', *Cognitive Psychology*, **1**, 242–49.

[61] Miller, G. and Johnson-Liard, P.: 1976, *Language and Perception*, Harvard University Press., Belknap Press, Cambridge, Mass.

[62] Minsky, M.: 1975, 'A framework for representing knowledge, in P. H. Winston (ed.), *The Psychology of Computer Vision*, McGraw-Hill, New York, 211–77.

[63] Minsky, M.: 1980, 'K-Lines: A Theory of Memory', *Cognitive Science* **4**, 117–35.

[64] Morton, J.: 1969, 'A functional model of memory', in D. Norman (ed.), *Models of Human Memory*.

[65] Nevatia, R.: 1974, 'Structured descriptions of complex curved objects for recognition and visual memory', in Standord Artificial Intelligence Project, *Memo AIM*-250, Stanford University.

[66] Norman, D. and Rumelhart, D. *Explorations in Cognition*, Freeman, San Fransicso.

[67] Quine, W. V.: 1960, *Word and Object*, M.I.T. Press, Cambridge, Mass.

[68] Richter, J. and Ullman, S.: 1980, 'A model for the spatio-temporal organization of X and Y-type ganglion cells in the primate retina', MIT-Laboratory; Memo. 573.

[69] Rieger, C.: 1975a, 'Conceptual overlays: a mechanism for interpretation of sentence meaning in context', *IJCAI* **4**.

[70] Rieger, C.: 1975b, 'One system for two tasks: A comonsense algorithm memory that solves problems and comprehends language', *Working Paper* **114**, M.I.T.-AI Lab.

[71] Rips, L. J.: 1975, 'Quantification and semantic memory', *Cognitive Psychology* **7**, 307–40.

[72] Rips, L. J., Shoben, E. J. and Smith, E. E.: 1973, 'Semantic distance and the verification of semantic relations', *Journal of Verbal Learning and Verbal Behavior* **12**, 1–20.

[73] Rips, L. J., Smith, E. E. and Schoben E. J.: 1975, 'Set-theoretic and network models reconsidered: a comment on Hollan's "Features and semantic memory', *Psychological Review* **83**, 156–57.

[74] Rosch, E. H., Simpson, C. and Miller, R. S.: 1976, 'Structural bases of typicality effects', *Journal of Experimental Psychology: Human Perception and Performance*.

[75] Rosch, E.: 1975, 'Cognitive reference points', *Cognitive Psychology* **1**, 532–47.

[76] Rosch, E.: 1973, 'On the internal structure of perceptual and semantic categories', in T. E. Moore (ed.), *Cognitive Development and Acquisition of Language*, Academic Press, New York.

[77] Rosch, E.: 1974, 'Universals and cultural specifics in human categorization', in R. Breslin and W. Lanner (eds.), *Cross-cultural Perspectives on Learning*, Sage Press, London.

[78] Rumelhart, D. E., Lindsay, P.H. and Norman, D.: 1972, 'A process model for longterm memory', in E. Tulving and Donaldson (eds.), *Organization and Memory*, Academic Press, New York.

[79] Schank, R.: 1972, 'Conceptual dependency: A theory of natural language understanding', *Cognitive Psychology* **3**, 552–631.

[80] Schank, R.: 1975, *Conceptual Information Processing*, North Holland, Amsterdam.

[81] Schank, R. and Abelson, R.: 1977, *Scripts, Plans, Goals and Understanding*, Lawrence Erlbaum, Hillsdale, N.J.

[82] Schank, R.: 1980, 'Language and memory', *Cognitive Science* **4**, 243–85.

[83] Silver, W. M.: 1981, 'On the representation of angular velocity and its effect on the efficiency of manipulator dynamics computation', *M.I.T.-AI Laboratory Memo.* 62'.

[84] Simmons, R.: 1973, 'Semantic networks: their computation and use in understanding English sentences', R. Schank and K. Colby, (eds.), *Computer Models of Thought and Language*, Freeman, San Francisco.

[85] Smith, E. E.: 1976, 'Theories of semantic memory', In W. K. Estes (ed.), *Handbook of Learning and Cognitive Processes*, vol. 6, Erlbaum Press, Potomac.

[86] Smith, E. E., Shoben, E. J. and Rips, L. J.: 1974, 'Structure and process in semantic memory: A feature model for semantic decisions', *Psychological Review* **81**, 214–41.

[87] Stent, G. S.: 1979, 'Cerebral hermeneutics', invited address at the Neuroscience meeting, Atlanta.

[88] Stevens, K. A.: 1979, 'Surface perception from local analysis of texture and contour', *M.I.T.-AI Laboratory*, Ph.D. Thesis.

[89] Sutherland, N. S.: 1979, The representation of three-dimensional objects', *Nature* **278**, 395–98.

[90] Tulving, E.: 1972, 'Episodic and semantic memory', in E. Tulving and W. Donaldson (eds.), *Organization and Memory*, Academic Press, New York.

[91] Ullman, S.: 1979, '*The Interpretation of Visual Motion*, M.I.T. Press, Cambridge, Mass.

[92] Vaina, L.: 1980, 'Towards a computational theory of semantic memory', *M.I.T.-AI Laboratory Memo*. 564.

[93] Vaina, L.: forthcoming, 'Towards a computational theory of semantic memory', in L. Vaina and J. Hintikka (eds.), *Cognitive Constraints on Communication: Representations and Processes*, Reidel, Dordrecht, Holland.

[94] Vaina, L.: (in prep), 'A visually based Functional Representation. A computational approach to neuropsychology'.

[95] Vaina, L. M. and Greenblatt, R.: 1979, 'The use of thread memory in anomic aphasia and concept learning', *MIT-AI Laboratory, Working Paper* 195.

[96] Warrington, E. K.: 1975, 'The selective impairment of semantic memory', in *Q. J. Exp. Psychol.* **27**, 635–57.

[97] Warrington, E. K. and Taylor, A. M.: 1973, 'The contribution of the right parietal lobe to object recognition', in *Cortex* **9**, 152–64.

[98] Warrington, E. K. and Taylor, A. M.: 1978, 'Two categorical stages of object recognition', in *Perception* **7**, 695–705.

[99] Waters, R.: 1979, 'Mechanical arm control', *M.I.T.-AI Laboratory Memo*. 549.

[100] Whitehouse, P., Caramazza, A. and Zurif, E.: 1978, 'Naming in aphasia: Interacting effects of form and function', in *Brain and Language* **6**, 63–74.

[101] Winograd, T.: 1972, *Understanding Natural Language*, Academic Press, New York-London.

[102] Winston, P. H.: 1975, 'Learning structural descriptions from examples', in *The Psychology of Computer Vision*, McGraw-Hill Book Comp, New York.

[103] Wittgenstein, L.: 1953, *Philosophical Investigations* (trans. by G. E. Anscombe, Blackwell, Oxford.

[104] Wernicke, C.: 1874, *Der aphasische Symptomencomplex*, Cohen and Weiger, Breslau.

JAMES A. MCGILVRAY

TO COLOR

The proper route to colors is via color in representings ("coloring"), rather than color in representeds (colored objects). "Coloring" can explain many of the features of colored objects, and allows us to locate color-properties as properties of sensings. I shall explain and defend this view in three stages. I start with an effort to map the logical space of color expressions in ordinary language. I then discuss critically Land's theory of color, to show how one scientific theory of color can suggest how to restructure this logical space. Finally, I construct a coherent philosophical account of color which shows that colors are best understood as modes of sensing. The result is a constructivist view of color which makes the colored object an "appearance".

1. "ORDINARY" LOGICAL SPACE

The logical space of color in ordinary language is not really coherent, although it appears to be so in normal epistemic contexts. Here, for instance, colors seem to be persistent properties of enduring physical objects, and common sense realism appears to be a proper account of color properties and their bearers. In abnormal contexts, however – situations which involve illusions, hallucinations, von Guericke's shadow, Yarbus' color washouts, and the like[1] – colors seem to be episodic properties, somehow generated "in the head (mind)". With these contexts uppermost, one could be convinced of the truth of phenomenalism, indirect realism, or idealism. The problem is compounded by the fact that, even in normal contexts, some elements in the logical space of color are unsettled (color-order can be specified in several compatible and non-equivalent ways) and others are opaque (homogeneity of color is not easily analyzed). The elements involved in the logical space of color seem to need a principled re-arrangement, an account more adequate than either common sense realism or phenomenalism.

Synthese **54** (1983) 37–70. 0039–7857/83/0541–0037 $03.40