ACTION-ORIENTED MEMORY SUBSERVING PERCEPTION[1]

Michael A. Arbib, Parvati Dev,
and Richard L. Didday[2]

Department of Computer and Information Science
University of Massachusetts
Amherst, Massachusetts 01002

Technical Report 70C-01

ABSTRACT

We argue that the brain be viewed as a layered computer, with long-term memory serving to ensure the correlation of "sensory features" in the "sensory layers" with "output feature clusters" in the "motor layers" which can determine action appropriate to objects in the environment; while short-term memory resides in maintained activity of "output feature clusters" which may be appropriate though uncorrelated with current sensory input. The slide-box and hologram metaphors are handled circumspectly.

---

[1]The investigations reported herein were supported in part by the Office of Naval Research under Contract Number N-00014-67-A-0044 to the Electrical Engineering Department of Stanford University. We are indebted to Richard Reiss for his thoroughgoing critique of an earlier version. This report will be published as two papers in Volume I of the Journal of Cybernetics: M. A. Arbib and R. L. Didday: The Organisation of Action-Oriented Memory for a Perceiving System: I. The Basic Model (Parts 1 & 3 of this report); and M. A. Arbib and P. Dev: The Organisation of Action-Oriented Memory for a Perceiving System II. The Hologram Metaphor (Part 2 and Appendix of this report.)

[2]Dr. Didday is with the Department of Mathematics and Statistics (Computer Science Section) Colorado State University, Fort Collins, Colorado 80521

1. INTRODUCTION

Let us start where another paper ended, by claiming that:

"... the following paradigms, all too often neglected in the cybernetics literature, must play a crucial role in future brain theory:

1. Theory must be Action-Oriented: e.g., studies of sensory processing must take into account the behavior of the animal, and the classification of sensory input implied by its actions.

2. Computation must be Distributed: e.g., the organism is committed on the basis of interaction between whole populations of simultaneously active neurons, rather than as a result of serial processing by a localised group of "executive" neurons.

3. The Brain is a Layered Computer, with Somatotopic Relations between Layers: this third observation is common knowledge to neurophysiologists and neuroanatomists - the time has come to incorporate it in our theories." [Arbib, 1971].

There our task was to pay tribute to Warren McCulloch by showing that these paradigms were implicit in his and Walter Pitts' 1947 paper "How We Know Universals". Here our task is to relate the above general thesis to a consideration of the memory structures required by any system which is to be said to perceive: A human knows that if he presses his hands in a downward direction at a fairly large velocity towards a rectangular surface (a table top) which he has sensed only visually, his hands will not go through the surface, and a noise will ensue. But it is no trivial task to program a computer to take so little visual information, and make predictions about the trajectory that it could get with its effectors, and the resultant feedback, constraints, and sound. Such predictions are the essence of perception, extrapolating from partial sensory information a great deal in many modalities besides those sensed, that is relevant to action.

Here, then, is the by now well-known idea of a long-term model of the world (see, e.g., Craik [1943], Gregory [1969]), something that tells

us that when we see surfaces of a certain kind, they are associated with certain textures, feelings, constraints on action, etc. We further posit that we have a short-term memory which contains information representing our model of the current state of the environment. Rather than every millisecond having to recognize the scene anew, we just notice discrepancies and use this greatly reduced information flow to update the short-term model which guides our activities. It is the long-term model that allows us to build up these short-term models using only partial information to gain access to far more information relevant to action. [This is an analysis-by-synthesis view of perception.] The real meaning of this distinction will be made clear in our discussion in Section 3.

Our crucial thesis is that perception is inseparable from memory, which in turn is meaningless without reference to the action of the organism - or, more properly, the interaction of the organism with the environment. Perception can be seen as the construction of a partially predictive internal (short-term) model, using long-term memory to incorporate information about action possibilities, and about sensory information from modalities besides those cuing. Perception is dynamic - both in that current information tends to be treated in the context of an existent short-term model, and also in that the extant model "unfolds" with time. [Whereas many designs of pattern-recognition machines give the classification of static patterns a primary role, human perception involves continual eye movements so that "static recognition" must be viewed as a "cross-section" of a dynamic process rather than the essence of perception.]

It must also be stressed that since the touchstone for perception is its effectiveness in providing information relevant to the guiding of

action, perception is not so much of "what" as it is of "where" and "what to do". This last characterization is not an arbitrary decomposition of the perceptual process - in fact the "what" of perception and the "where" of perception seem to be basically different functions mediated by different regions of the brain. A good amount of evidence (Schneider [1969], Ingle, Schneider, Trevarthen and Held [1969]) indicates that in mammals, midbrain structures are concerned with relating the animal's acts and sensory impressions to a body-oriented reference frame. This allows cortical structures to operate to recognize scenes largely free from small changes induced by actual body positions and movements.

Clearly, then, information - save at such high levels as involved in human linguistic activity - must be continually referrred to what we shall call the ACTION FRAME, namely, the frameof reference induced by the postural and effector mechanisms of the organism, the former serving to stabilize the frame offered by the latter. The repertoire of possible actions of an organism, and of possible questions it may ask, helps to determine the most appropriate neural representation of information.

In particular, the organism may encode information in different ways on different occasions depending on the questions prevailing at the time. This fact partly explains the discrepancy between the millions of nerve fibres carrying information to the brain, and the 20 bits per second that some researchers have used as the measure of human information-processing, since a great deal of input information may be used to generate a code which only contains that small amount of information currently required. One can test this by writing six words - some in script, some in hand-printing - and asking people to memorize the list. Most will recall the list perfectly as a list of words per se, but will not remember which were

printed and which were script - even though, presumably, the processes of word recognition work somewhat differently for script and printing. The brain must have responded momentarily to the difference in order to process the input appropriately, but this difference did not enter the memory "trace", which instead was "something" that encoded the words, or perhaps what they denoted, rather than the writing itself that you displayed to the subject.

Taking this one step further, one may suggest that in some learning situations, what we learn is not what a sensory input pattern is, but rather what is the appropriate question to ask about it. For example, a rat in a maze is not simply pairing a response with a stimulus - it is executing complex muscular activity while being bombarded with a mass of sensory input. A rat might perfectly well remember how he got to the food on a certain trial and still fail at the next - because he does not know whether it was a subtle smell, the texture of the floor, a direction, a patterning of his muscular activity, or a mark on the door that was significant. Thus, even with perfect recall of what happened at each trial, it might take many, many trials until the rat learned to disregard irrelevant stimuli and consistently use the cues used by the experimenter to locate the reward. Anything which could tell the rat which aspect of the information mattered would lead to almost instant learning - by directing attention, rather than conveying a specific message as to the location of food. Perhaps such "focusing of attention" is responsible for the "one-shot learning" some researchers have claimed to obtain by injecting brain extracts from trained animals into untrained animals. Less fancifully, much of the most basic usefulness of language resides in its ability to direct the attention of the listener.

2. TWO METAPHORS FOR A DISTRIBUTED MEMORY.

WARNING: The two metaphors for distributed memory which follow - slide-box and hologram - must not be taken literally.

HOMILY ON THE NATURE OF METAPHOR: A metaphor, by comparing a system to another which we may understand better, is designed to aid comprehension of the first system. But, besides properties which the two systems may share, there are many that they do not share. The reader must thus not believe that provocative properties of the second system are automatically shared by the first system - especially if we have not mentioned them. To say that "My love is like a red, red rose" does not imply that she will appreciate having the hose turned on her. Even for similes, so for metaphors. A good metaphor is a rich source for hypotheses about the first system, but must not be regarded as a theory of the first system.

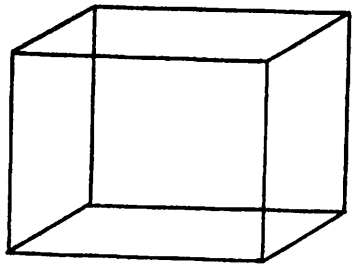
SUGGESTION: The reader who does not trust himself to avoid false importations from metaphors should proceed directly to Section 3, where we have tried to present our "pretheory" of memory in a way as unencumbered as possible by the metaphorical influences of its evolution. It seems appropriate therefore, that one of the more fruitful relics of this evolution - a mathematical treatment of some aspects of holographic memory - should appear as an appendix! The authors deny any claim that this paper suggests that "the brain is a slide-box" or "the brain is a hologram". What we do claim is the far weaker statement "Certain resemblances between some aspects of human memory and some techniques of movie-making and holography shaped the development of our theory; so that some readers may better appraise that theory if they are presented with the metaphors that helped shape it". Please read no more into our metaphor than that.

Perhaps some more insight into our notions of models of the world may be gained by a metaphor drawn from the making of movie cartoons. Drawing each frameindividually is too inefficient, and so instead use is made of the layering technique. One might go for a whole minute of the cartoon without the background changing, so one could draw it just once. In the middle ground, there might be a tree, say, about which nothing particularly changes during a certain period of time except its position relative to the background. It could thus be drawn on a separate layer,

which could then be displaced for succeeding frames. Finally, in the foreground, it may well be that one could draw most portions of the actors for repeated use, and then position the arms, facial expressions, etc., individually for each frame. The layers can then be photographed appropriately positioned in a slide-box for each frame, with only a few parameter changes and minimal redrawing required between each frame.

A similar strategy for obtaining a very economical description of what happens over a long period of time might be used in the brain, with a long-term memory (LTM) corresponding to a "slide file" and short-term memory (STM) corresponding to a "slide-box". The act of perception might then be compared to using sensory information to retrieve appropriate slides from the file to replace or augment those already in the slide-box, experimenting to decide whether a newly retrieved slide fits sensory input "better" than one currently in the slide-box. Also, part of the action of the organism in changing its relationship with the environment might be viewed as designed to obtain input which will help update the STM, by deciding between "competing" slides, as well as helping update the LTM, by "redrawing" and "editing" the slides.

Presumably, both evolution and development contribute to the criteria of when one slide is "better" than another. One might have a slide for "humming-bird at 30 feet" in the box (or, perhaps better - a slide for "humming-bird" positioned to represent a distance of 30 feet, etc.) and then the slide "insect at 10 feet" is popped up from LTM, and fits other sensory input better and so replaces the original slide. In a case such as the Necker cube, two equally good slides - cube pointing in, and cube pointing out - are available, and these alternate since there is no contextual cue to "lock in" either one.



Necker Cube

"Slide" is, of course, a bad name - "action-directing multi-modal audio-visual film-clip dynamic subroutine" might be better, but isn't! The "slide-box" is not a box into which static slides are inserted - rather, it is a mass of neural tissue lying athwart the channels which link the sensory and motor systems. "Putting in a slide" corresponds to activating this network, thus initiating transient wave-forms which change autonomously. Lying athwart the lines of communication, the "slide-box" fills the whole "postural-effector frame of reference" - and so a slide does not contain information from any single modality - rather we may use cues from one or more senses, or from feedback from motor activity, to "address" the activation of a wealth of multi-modal action-oriented information.

The above paragraph suggests that the danger with the slide-box metaphor is that it may force too rigid a view of neural activity. A good point about it is the way it emphasises that fine perceptual acts take place against a background - that the whole "scene" must always be filled in and we cannot recognise small details in a void. Further, present "slides" strongly color the choice of each addition.

At any rate, the metaphor - with its picture of many slides continually interacting and being updated, with no single locus of exclusive serial processing - does serve to emphasise our need to understand how

computations may take place in a highly parallel network of dynamically interacting subsystems. Such an understanding may also help us probe the effect of brain damage upon behavior. In 1929, Karl Lashley published a book "Brain Mechanisms and Intelligence" in which he reported that the impairment in maze-running behavior caused by removing portions of a rat's cortex did not seem to depend on what part of the cortex was removed, but only on how much was removed. He thus formulated two "laws": the "law" of mass action - that damage depended on the amount removed - and the "law" of equipotentiality - that every part of the brain can make the same contribution to problem-solving. Such data have seemed to many irreconcilable with any view of the brain as a precise computing network - but we may effect a reconciliation if we stress the notion of a computation involving the cooperation of many subroutines that are working simultaneously in parallel. Often, a computation can be effected by a subset of the routines. In general, removing subroutines will lower efficiency, though for some tasks the missing subroutines may be irrelevant, so that their removal saves the system from wasting time on them when other tasks are to be done. Robert White (personal communication) repeated Lashley's experiments, but rather than measure impairment by a single parameter, he judged wherein the impairment lay. One rat might perform poorly because of a tendency to turn left; another might be easily distractable; while a third might sit still most of the time, but find his way through the maze perfectly well whenever he could be "bothered" to try. Thus equipotentiality is really only valid if we use rather gross measurements of change in behavior - the underlying reality would seem

to be the removal of subsystems which can make quite different contributions to a given level of performance.

Turning to the relative insensitivity of the sensory pathways to lesions, we should note that as we move centrally in the input systems, each cell monitors a wider range of input. We have both convergence and divergence of information, so that each peripheral point contributes to the activity of many central cells, while, conversely, each central cell monitors many receptors. Thus the input information is both distributed and redundant. Even removal of fairly large pieces of the central pathways would still allow that at least partial information about most of the sensory periphery would be retained. Further, our ability to move our receptors - as in scanning a visual scene, or running our hand over a surface - allows us to make a "mosaic" of a total scene even if our input pathways are restricted in their peripheral range.

To consider our second metaphor, we start by making the following claims about human memory:

(a) Our memories are dynamic, and located in the "action frame" in that we, e.g., recall perceiving a room about us, rather than recalling a series of isolated two-dimensional perspectives.

(b) Humans may experience certain forms of gross brain damage without losing their memories (at least with respect to very rough tests - though localised damage to other sections of the brain may have gross effects, as with lesions of the hippocampus, which may destroy the ability to add new information to long-term memory).

(c) Memories of many different events can be stored in the same region as the brain. [This sounds reasonable, but can it be verified experimentally?]

This has suggested to a number of workers, including Karl Pribram [1969] that what we here insist is only a metaphor for the memory system may be found in the hologram, which is a form of photographic image with the following provocative properties:

(a) the image produced is three-dimensional, and may be viewed from many aspects;

(b) each part of the hologram can reproduce much of the entire image - in distinction to the sharp localisation of a conventional photograph - but resolution decreases as the area decreases;

(c) several images can be superimposed, and later recovered individually.

To understand properties (a) and (b) of real holograms, consider viewing an object through a window. The hologram may be regarded as a "freezing" of the light waves from the object as they impinge upon the window. We can thus see all aspects of the object, so long as we view it from an angle corresponding to standing in front of the window. Similarly, we can see the whole object, though we may have to move our eyes around more, as more of the window is covered.

To understand property (c), we need more technical information of the "freezing" process. A hologram takes a photograph using illumination from a laser - a "reference beam" of laser light, and another laser beam reflected from an object yield wavefronts of light which form an interference pattern on the photographic plate. The developed slide - called a hologram - can later be "read out" by shining a copy of the "reference beam" of laser light through it, which acts as if to restart the original wavefronts of light anew. The crucial point for (c) is that if we photograph several holograms on the one plate, but using a different reference

beam for each hologram, then when we illuminate the plate with a single reference beam, only the corresponding wavefront will be restarted, and so only the corresponding image will be reconstructed - the interference patterns for each of the other wavefronts will each cancel out with the wrong reference beam.

If we have two objects, numbered one and two, and record the interference pattern of the wavefront reflecting off both these, the reflection from each object acts as a reference beam for the other. The reflection from one object can be used to illuminate the hologram and "read out" the other object, thus gaining a primitive associative memory - if two objects have been photographed together, then one can be used to recall the other. [If we use what is called a Fourier transform hologram, object 1 need not be placed in a fixed position - object 2 will always be read out in the same position relative to object 1. This can be of great use in an information-retrieval system. Suppose that object 1 is an occurrence of a key word, say "automat", that we want to retrieve, and that object 2 is just a spot of light - so that when we illuminate the hologram with the word "automat", the output will be a relatively positioned spot of light. If we now illuminate the hologram with a whole page of type, every occurrence of "automat" will yield an appropriately positioned spot of light.]

The similarity between the (a)-(b)-(c) of memory and the (a)-(b)-(c) of the hologram has suggested that thinking of the brain in this context, in terms of propagation of waves of neural activity rather than in terms of the step-by-step computations of individual neurons, we could then imagine wavefronts of sensory excitation being frozen into a "neural hologram" which may be restarted whenever the memory of that experience

is called for. (Noting that the coherent beams of light used to make real holograms are generated by lasers, we should not make the mistake of the girl who asked in one of Pribram's classes "But doesn't the laser process destroy brain cells?" If the hologram metaphor is to have any real utility, then the reference beams, too, must comprise waves of neural activity.) "Ghost images" elicited by almost-reference beams might then provide a sort of associative memory.

Pribram suggests that there is a neural hologram - and so does not treat it as a metaphor - and that it is obtained as a result of interference on neurons between a pattern sent directly (impinging on near end of dendrites) and a slightly delayed pattern (impinging on far end of dendrites). He suggests that we test such a notion by analysing the mathematical rules for transformation from impulse to graded potentials, and vice versa, to see if it is holographic (and, in particular, invertible). He further seems to view consciousness as a reading-out of the hologram in the head - which may be just the homunculus fallacy, for if memory reconstructs visual input we may be back to "explaining" perception in terms of a little man sitting in a control room inside the head, monitoring neural messages from the periphery, and starting an infinite regress of smaller and smaller homunculi.

However, the real interest to the brain theorist of the (a) of holography should not be the three-dimensionality per se, but rather the fact that we have a record of a wavefront, rather than the static cross-section of conventional photography. Property (b) need not excite us since we expect it in the brain as a result of the redundancy obtained by the property we have already discussed of mutual convergence and divergence in neural pathways. Perhaps (c), the storage of many traces in the same

region, can give us the most fruitful ideas. In any case, we repeat that to profit from the metaphor, we must avoid too literal use of it.

We must not expect a neural hologram to share with the real hologram the exact mathematical nature of the transform from scene to record. The real hologram is essentially a spatial Fourier transform, so that each point records frequencies for the whole original in such a way that the transform can be inverted to reconstruct the complete wavefront from its transform. But it would seem to be of less value to the organism to reconstruct a visual input per se than to recall vital features of past experience. It is now well known that the early stages of an organism's input systems are preprocessors designed to extract from a sensory pattern features relevant to the activity of the organism. Thus, the appropriate notion for a neural hologram may not be the invertible Fourier transform, but rather the non-invertible FEATURE TRANSFORM in which a spatial array of intensities is replaced by a spatial array of features (and we stress that this is true of all modalities - not just the visual system). Note that each point of the feature transform corresponds to a large sensory field, with much overlap. Thus, portions of the feature transform may serve to encode larger portions of the original, albeit with lack of detail in certain features, even if there is a great clarity in others.

Since we consider one of the primary functions of memory to be to augment current input in determining action, it seems appropriate to consider an ACTION SPECTRUM in considering the neural model, rather than the frequency spectrum of the real hologram - so that the features considered above will be those which enable the animal to react quickly and appropriately. As we shall emphasise in the next section, we shall regard the

pattern of receptor activity as being preprocessed to extract those sensory features likely to be of significance to the animal, which features are further processed to yield "output features" which "describe" an object in terms of what the organism can do to it. Rather than an invertible record from which the animal can reconstruct the original stimulus, we posit a non-invertible record from which it can construct an appropriate response.

In humans, the picture is obscured by the fact that one of our responses to a scene is to describe it verbally - we are suggesting that our use of language gives us the ability to reconstruct a description of the original stimulus from the action spectrum in a way denied to (though not without precedent in) lower animals. In other words, the recall of specific events that we view as a primary function of memory may actually be a phenomenon very recent in evolutionary terms.

It is tempting to try to understand the human brain by starting from that most human of facilities - the use of language. We shall instead stress principles of brain organisation which tend to be obscured by this approach. Here let us just note how the capacity to use language might arise as an elaboration of a more basic organisation. Of course, once language enters the picture it can modify behavior drastically - we do not wish to deny these basic differences, but would simply emphasise the importance of understanding the substrate from which they arise.

Our use of nouns reminds us of the importance of classifying objects. However, we often let our emphasis on language delude us into thinking that our interaction with the world must involve a verbal mediation of somewhat the following kind:

1. Seeing an object, name it

2. Use the name of the object to name the appropriate action

3. Act as designated

though this simple scheme is elaborated by such considerations as the need to weigh alternatives, etc. In any case, this approach yields the sort of block diagram shown in Fig. 1. Much Artificial Intelligence research has adopted this approach because they sought methods amenable to string-processing on conventional "serial" computers, rather than trying to understand the brain mechanisms underlying Real Intelligence.

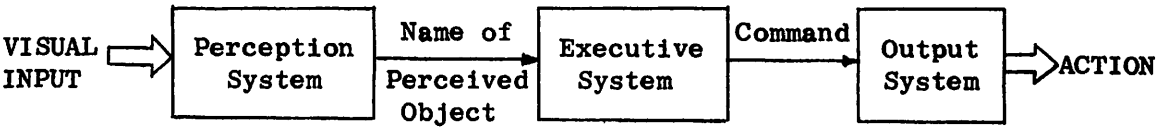


Figure 1.

Block Diagram for Verbal Mediation of Action.

The central point of Figure 1 is that all the input funnels down to a few well-chosen words which can be processed by some centralised executive to yield a command which can then control the musculature. Now, especially with a few feedback loops judiciously thrown in, this may well be a useful model for analysing much of human behavior. Our point, however, is that there exist sophisticated strategies for sensori-motor relations which do not require an arch-controller, and that we may do well to regard verbal mediation as providing a higher level - in a hierarchical structure such as shown in Fig. 2.

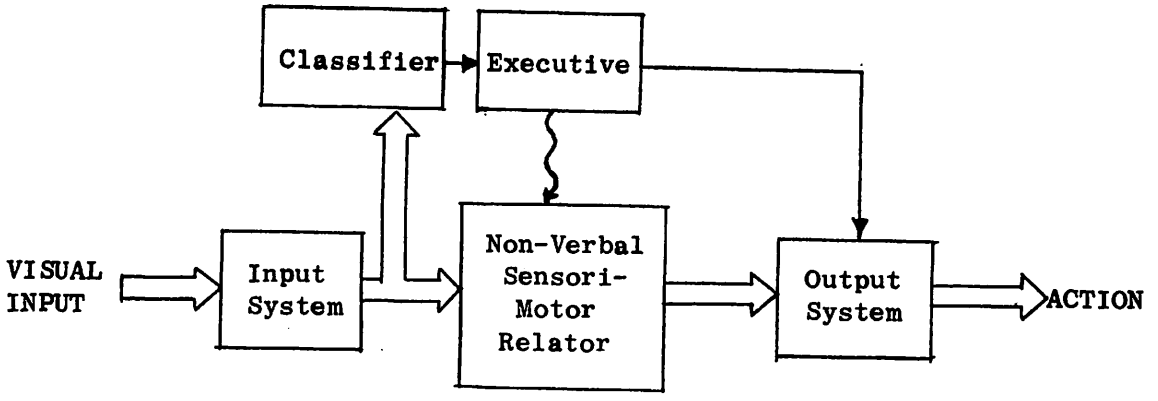


Figure 2.

A Hierarchical Approach

Here the executive can control the output directly, but need not intervene in much of intelligent behavior which is not explicitly verbal, leaving the task to lower centres, or simply biassing the underlying sensori-motor apparatus, which does not make explicit use, generally, of verbal coding of input to compute actions. It is thus anarchic, in that there is no head or centre which directs its computations. Rather, it is a distributed highly parallel computer. In the next section we shall ignore the "linguistic level" and instead explore the idea that the "meaning" of an input for an organism resides in the interactions that are appropriate with the object it represents - which actions depend not only on what an object is, but also on where it is.

With this, we have given enough clues to the genesis of the scheme we shall present in the next section; and have also seen the danger of using a metaphor when one does not distinguish it carefully enough from a theory. It is for this reason that we shall eschew in Section 3 all use of terms like "slide-box" and "hologram", but the critical reader may well wish to be alert for false importations which are not so simply banished by editorial decree.

3. THE GENERAL FLOW DIAGRAM

We suggest that the spinal cord is so organised that the natural patterns of stimulation that it receives from higher levels - brain stem and motorsensory cortex, for example - do not yield twitches of individual muscles but coordinated patterns of movement which may involve a number of muscles - as attempts to wiggle the middle toe alone clearly show. We stress that each muscle comprises a population of muscle fibres, so that large movements of the organism (as distinct from fine finger movements in human manipulation) involve activity in a population of motorneurons. We suggest that recognition of the current environment comprises a pattern of activation in frontal cortex, say (this is still speculative), which if allowed to play upon spinal centres would yield the attempted execution of a host of interactive behaviours of the organism each more or less appropriate to that environment. One of the central problems of nervous system "design" is thus to ensure that at most one of these behaviours is manifested at any one time, and that the one that is manifested is among the more appropriate for the current situation of the organism (cf. Didday [1971] for possible such circuitry in the frog).

We know that at some level the brain may be viewed as a layered computer with locus in a layer representing spatial location of the source of the stimulation which can activate cells in that locus; we suggest that in layers on the "output side", locus in the layer represents spatial location of the region in which interaction would ensue were the spinal centres to "execute" the behaviour controlled by that region. We may thus speak of "sensory layers" in which the environment is encoded as a spatially tagged array of features extractable in a few layers

of neural processing from the pattern of receptor responses to environmental energies; and of "motor layers" in which the environment is encoded as a spatially tagged array of "output features" or "motor subroutines" (beware: the computer metaphor is creeping in!) which can be transformed into coordinated movements by the musculature after a layer or two of neural processing in the spinal cord. The fact that clusters of "output features" may change locus in a highly correlated fashion corresponds to our segmentation of the world into objects.

Thus a crucial notion in our model is that of the OFC (output feature cluster) as the encoding of an object in the action frame - i.e., as a cluster of features that might be appropriate for interaction of the organism with the object. Whereas in lower animals the relationship between sensory features and motor features is essentially direct and innate (cf. the discussion of the frog tectum in Arbib and Didday [1971]), in mammals there is a great additional mechanism which allows development of relationships which have their bases in the past experience of the animal.

We thus model LTM (long term memory) as residing in the intervening network which enables the array of sensory feature activation to be segmented and transformed into a spatially tagged array of OFC's, in such a way that the OFC's are indeed appropriate for interaction with the objects yielding the sensory features from which they were obtained.

We thus do not distinguish at this stage genetically specified mechanisms from those which are learned, for since in general both are active in a perceptual situation, together they define the animal's long-term model of the world. We also defer for a later paper the discussion

of memory of events rather than skills. The reader, being warned that our definition of LTM is different from that of most other authors, will hopefully have no trouble with it.

We reiterate that an OFC will in general contain far more subroutines than the organism would be able to execute at any one time, and so much neural machinery must serve to commit the organism to only one set of compatible actions at any one time. Kilmer, McCulloch and Blum [1969] have suggested that the reticular formation (RF) chooses a gross mode, and Kilmer and McLardy [1970] go on to suggest that the hippocampus chooses acts within modes. Arbib and Didday [1971] have noted that an even finer analysis is necessary (and, incidentally, offer an alternative to the RF theory) - giving a mechanism whereby an environment containing several flies becomes encoded as a spatial array of "snap" commands, requiring further neural processing to ensure that, in general, only one of these features will get through to motor centres and that the frog will snap at only one of the flies.

We further suggest that STM (short term memory) comprises the totality of current activity of OFC's. It is LTM that allows the organism to generate an OFC from partial (probably ambiguous) sensory information about an object; it is STM that keeps appropriate OFC's activated even though they supply at most remote context for currently sensed objects, rather than providing OFC's simply for current input. A crucial property of STM must then be to continually relocate an OFC so that its location remains appropriate to the corresponding object, even should that object be moving. We shall stress that such relocations are relative and may be obtained by a combination of remappings of input (sensory) and output (motor)

pathways. The neural circuitry should be such that it is in general easier to detect that an activated OFC matches sensory features despite a discrepancy in locus, and relocate the OFC appropriately, than to use LTM to generate a whole OFC, little of which bears any direct relationship to current sensory features.

We emphasise again, then, that an OFC is a pattern of neural activation representing an object in terms of the organism's possible interactions with it. [The same object may elicit different OFC's on different occasions - a cat may activate the output feature "stroking" on one occasion, and the output feature "pushing off" on another - we may say that the transformation from object via sensory features to OFC's yields a polythetic classification.] STM's status as a collection of data on what to do with the environment may be considered equivalent to having an internal model of the environment. An OFC is thus not a simple recoding of (in the case of visual sensation) two-dimensional patterns. For example, Fig. 3 shows two perspectives of a cube which are not intertransformable by motions in the plane, but which may nonetheless elicit the same OFC, appropriate to (probably among many other things) grasping a cube, save, e.g., for differences in a parameter reflecting the different angle at which the thumb-finger opposition would have to approach a real cube were it responsible for one of those two-dimensional projections.

The position and orientation of an object relative to the organism's surfaces determines in part what environmental energies will impinge upon the receptor surface of the organism. Let us pick a standard position X of the object. At any time t it has a position in absolute space corresponding to a displacement and rotation $S_e(t)$ - so that the

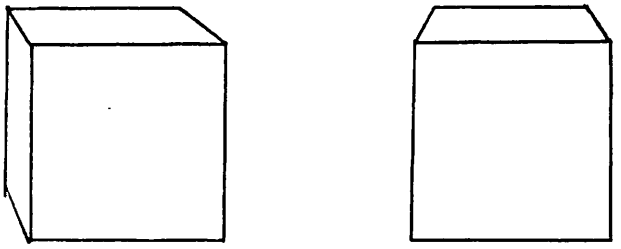


Fig. 3.

Two Perspectives of the Cube not Intertransformable by Motions in the Plane. One has 7 vertices and 9 lines; the other 6 vertices and 7 lines; yet this discrepancy in sensory features is banished as soon as we recognise them both as perspectives of cubes.

resultant distribution of matter in space may be described by $S_e(t) \cdot X$. The position of the receptor serface at time t then yields a projection $S_r(t)$ so that the actual neural representation will be determined by the two-dimensional pattern $S_r(t) \cdot S_e(t) \cdot X$. [We are talking of visual receptors here - a slightly modified discussion would handle other modalities as well.]

Now it is true that an adult may think in terms of abstractions, or plan his movements from the outside as if imagining himself moving on a map, but we now want to consider a more fundamental "egocentric" form of perception in which the organism views objects not in terms of some absolute spatial framework but in terms of position relative to the perceiver, i.e., in terms of coordinates induced by the reach of the hands, direction relative to that in which the body is facing, number of steps the object is away, etc. Size may then be perceived to the extent that we can scale our movements for appropriate interaction, orientation to the extent that we can turn our head to centre our gaze on an object, and so on. We might, for instance, say that we perceive the lines on a page of ruled paper to the extent that we can not only position a pencil on them, but also adjust

our handwriting size to the spacing of the lines. In other words, our theory must include a transformation $S_m(t)$ which scales muscular activity to the task at hand.

Given the important role in our theory of the generation of transformations to match motor output to sensory input, it is appropriate to recall here from [Arbib, 1971] a slight generalisation of Pitts and McCulloch's [1947] "uniform principle of design for reflex mechanisms which secure invariance under an arbitrary group G". However, where the original model only considers tranformations of sensory features (such as rotation of a visual perspective in the plane) we shall also be interested in transformations of output features which represent all possible motions of an object in space, since the organism needs data (though we shall see below that they need not all be neurally encoded) on the relative position of objects if it is to successfully interact with them. On the sensory side, we might want to generate a transformation T that will bring a sensory pattern φ to the middle of the visual field. [Pitts and McCulloch consider the case where T is implemented as a motion of the eyes, but a case can be made for internal remappings as well.] On the motor side, we might want to generate a transformation T that will modify the parameters of an OFC so that the hand will close about the object, and not adjacent empty space.

In either case, the generalised Pitts-McCulloch scheme uses an error reduction scheme to find the transformation T which will transform a given pattern to a standard form $T\varphi = \varphi_0$.

We associate with each pattern φ an n-dimensional "error vector" with the property (to be relaxed in a significant fashion below) that

$E(\varphi) = 0$ if and only if φ is in standard form. Further, we introduce a mapping $\mathcal{W}$ which associates with each error some transformation which can reduce it, i.e.,

$$\mathcal{W} : \mathbb{R}^n \to G \quad \text{is such that} \quad \|E[\mathcal{W}(E(\varphi)) \cdot \varphi]\| \leq \|E(\varphi)\|$$

for all patterns φ, and with equality only in case φ is in standard form.

The student of artificial intelligence will notice here a forerunner of the Difference-Operator method of Newell, Shaw and Simon's [1959] GPS (General Problem Solver), with two crucial distinctions: (i) GPS only employs a finite set of "error-vectors" or "differences"; (ii) the operator $\mathcal{W}_D$ associated with a difference D does not always successfully reduce the difference - thus a procedure known as "heuristic search" is required to find which sequence of operators among a number of apparently likely candidates, will indeed remove the initial difference. It is this less-than-ideal, but more realistic, situation that we shall have in mind when we allude to heuristic search below. Meanwhile, Fig. 4 shows a discrete-time system which will find, in the ideal case, the transform which will reduce it to standard form.

Of course, the "hard work" in both GPS and the Pitts-McCulloch scheme is to find the appropriate error measure E and error-reducing map $\mathcal{W}$ to so generate transformations that continual use of error feedback will eventually bring the initial pattern to its standard form.

Let us present one more mechanism before attempting an overall schematic for a perceiving system. If an animal has an object in its visual field, then the same pattern of retinal stimulation can be produced in

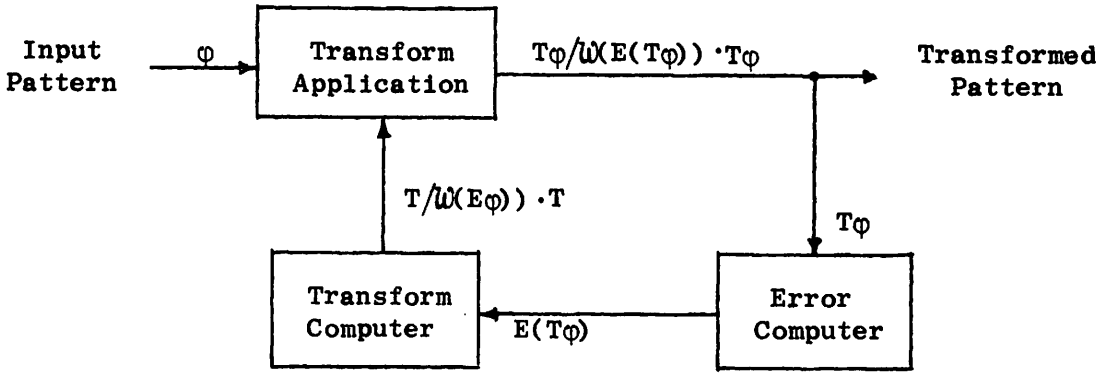


Fig. 4.

A Generalisation of the Pitts-McCulloch Scheme for Transforming a Pattern to Standard Form.

The transform application box is memoryless - input pattern φ and transform T at its input yield transformed pattern $T\varphi$ at its output. The error computer box is memoryless - an input pattern at its input yields the corresponding error at its output. The transform computer box is a sequential machine - if its state at time t is the transform T, and its input at time t is the error vector e, then its new state and output at time $t + 1$ will both be the transform $\omega(e)\cdot T$. [From Arbib, 1971].

two distinct cases:

Case 1: the animal advances one metre towards the object; and

Case 2: the object advances one metre towards the animal.

Despite this identity of retinal stimulation, in a hostile environment such as that which shaped our own evolution, only in Case 2 does the object have such importance that the animal must orient to, and identify, the source of the retinal stimulation. This is graphic evidence of the important notion that - while it is change in stimulation that carries information for a perceiving system - changes produced by the animal's own movements must be "downgraded" to some extent lest they mask important changes in the environment.

This is a massive problem for a system such as a mammal which has delicately controllable receptors which bear many-dimensional relationships to similarly complex effector surfaces. We suggest that this

requires mechanisms which operate conceptually like those shown in Fig. 5.

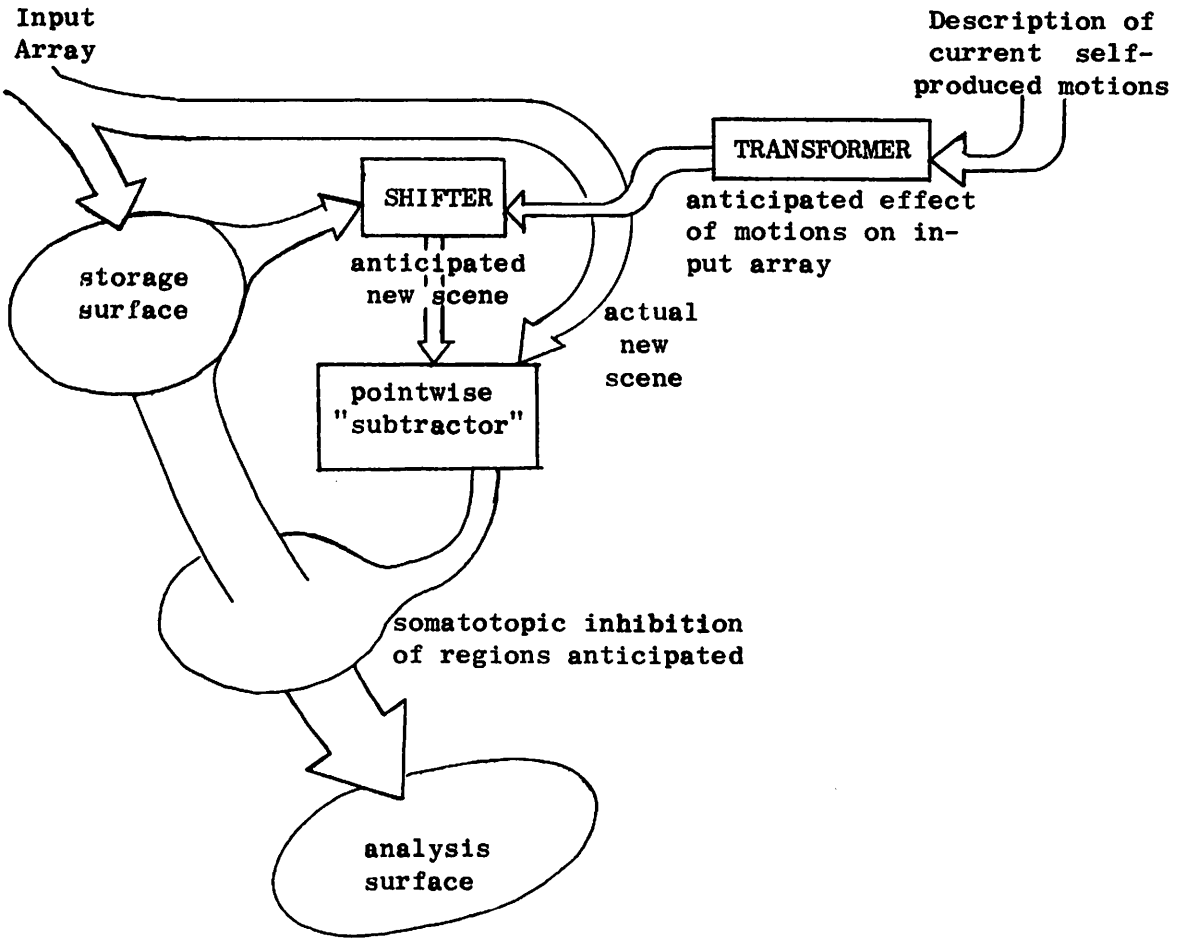


Fig. 5.

Schematic for Ignoring Inessential Changes in Sensory Input.

Figure 5 shows the input array about to enter a "storage surface" where it will soon replace the last such input. The input array also enters a mechanism which compares, in parallel, points in the input array with points in the "anticipated new scene". This anticipated scene is one computed by altering the last input scene (held on the "storage surface") in a way determined by the self-produced motions the system is currently carrying out. Regions in which the new scene is changed due only to self-produced motions will thus match the corresponding regions

of the "anticipated new scene" and inhibit the corresponding regions of the pathway leading from the "storage surfact" to the first"analysis surface". Regions in which the new scene does not match the anticipated scene are not inhibited as they lead to the "analysis surface" and are exactly those regions to which the system must attend. We emphasise that Fig. 5 is a tentative formulation - the "transformer" and "shifter" are themselves highly somatotopic devices, and so transform different parts of the array in different ways. Further, it is very much an open question in our minds as to what extent the processes involve direct efferent control of the receptors, and to what extent the "output feature clusters" of the "analysis surface" themselves feed back to control much of the activity of the "transformer" and "shifter".

The orienting reaction is not to be considered one of the output features representing an object though it does contribute to the spatial location of the eventual OFCs in the somatotopic action frame of STM. Orienting allows the fine details of OFC generation, and so precedes it, or at least its refinement. In animals with foveal vision, we might regard the cortical "pattern recognisers" as being "time-shared" (to use the computer metaphor again) with the orienting mechanism and cortical shifts of interest controlling its allocation (see Arbib and Didday [1971] for some speculation on how this might be done) - as if figuring out what is there requires more "machinery" than noting a discrepancy from what is expected. Man is thus midway between a frog - no fovea, uniform rather limited processing of visual input - and a superman - all fovea, uniform sophisticated processing of a large visual field.

With these schemes as background, we may now turn to Fig. 6, where we have drawn a crude block diagram for a system which incorporates much

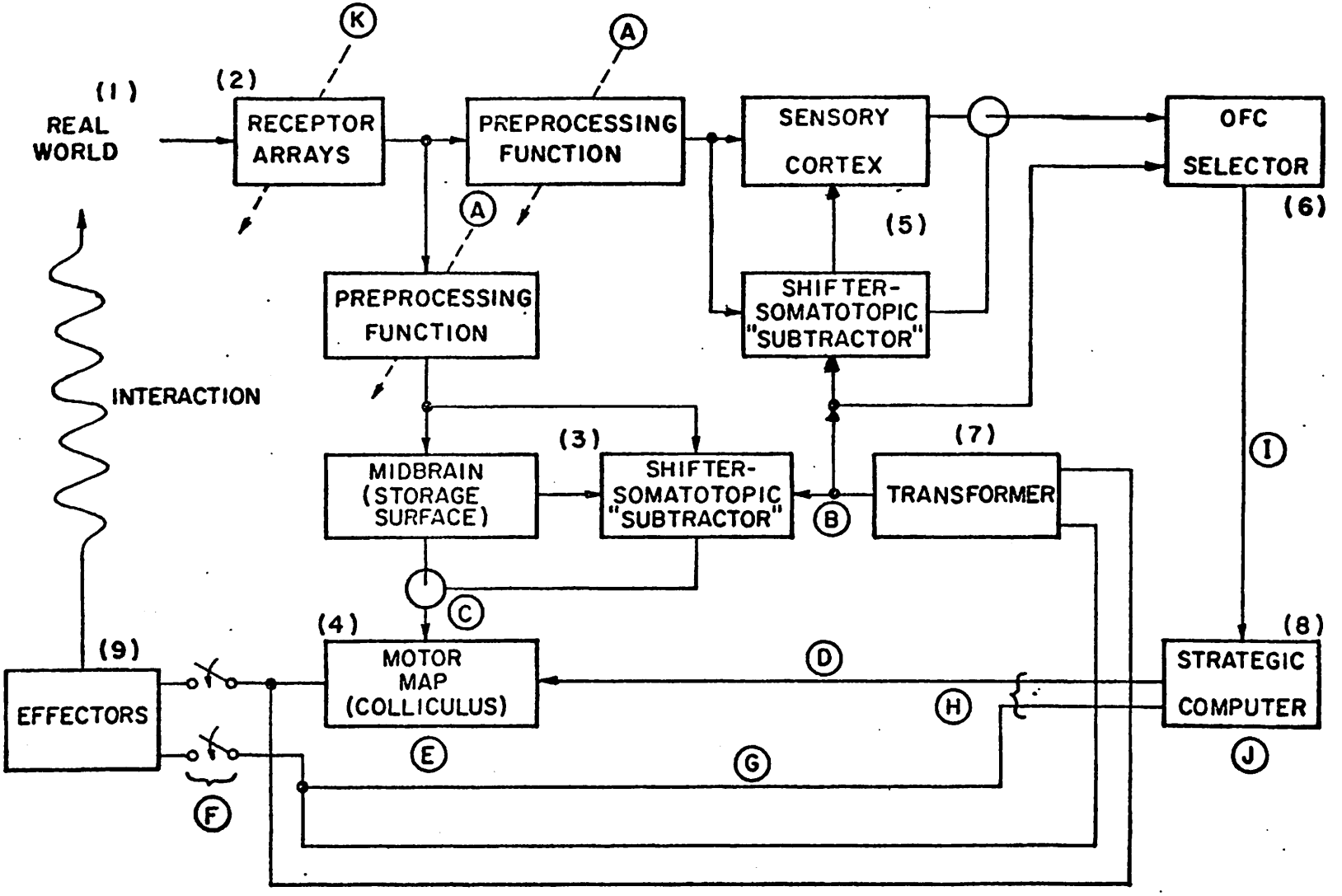

(1)
REAL WORLD
(2)
RECEPTOR ARRAYS
PREPROCESSING FUNCTION
SENSORY CORTEX
OFC SELECTOR
(6)
(5)
PREPROCESSING FUNCTION
SHIFTER-SOMATOTOPIC "SUBTRACTOR"
INTERACTION
(3)
(7)
MIDBRAIN (STORAGE SURFACE)
SHIFTER-SOMATOTOPIC "SUBTRACTOR"
TRANSFORMER
(4)
(8)
(9)
MOTOR MAP (COLLICULUS)
STRATEGIC COMPUTER
EFFECTORS
A
B
C
D
E
F
G
H
I
J
K

Fig. 6.

Legend for Figure 6.

A. under system control

B. "effect" of chosen OFC

C. orientations to feature center of gravity

D. grosser movement commands

E. includes remapping for receptor, effector position

F. final control over whether to "unleash" the OFC

G. pyramidal tract - direct cortical control of fine movements

H. chosen OFC

I. potential OFC's

J. decides among OFC's on basis of goals; sequences OFCs in PLANS

K. position under system control

of the "mammalian strategy" to take account of the above considerations - and others that we shall explicate as we proceed. The reader will certainly appreciate that many other structures could embody the same functions. Further, we shall see that despite its many boxes, it still presents too crude a caricature of action-oriented perception. Nonetheless, we hope that by studying it we may attain a more sophisticated level of discussion of perception. In cases where conflict arises, the logical flow of information should be considered to override physiological assignments.

Objects in the real world (1) may be considered to modulate environmental energy flow. The animal's receptor arrays (2) register projections of this flux along the different sensory "modalities" and in space. The visual receptors, tactile receptors of the hands, etc., may be moved with respect to the rest of the animal's body. Receptor inputs project to two major systems, one midbrain (3) and one cortical (5). The midbrain projection suffers local (pointwise) preprocessing and enters a mechanism like that illustrated in Fig. 5. Regions which contain stimulation differing from that expected emerge to produce orientation movements through the motor map structures (4). This route is that described by Pitts and McCulloch [1947] as causing the eyes to bring objects to a "standard position" for the cortical recognisers. We leave a more detailed discussion of this interaction for later, and turn to the cortical route.

Here, sensory inputs reach another mechanism like that of Fig. 5 which allows those portions of the pattern which have temporal changes unpredicted by the self motion of the organism to pass (6) to the OFC selector (pattern recogniser, object hypothesis generator) which analyses

the firing pattern spatially. This works because the same unpredicted regions will (through the orienting mechanisms (3) and (4)) bring the receptor surfaces to face these information-bearing regions, each in its turn. The OFC selector also receives a measure of what current movements are doing from (7) so that the selection of OFCs will be appropriate to the context of ongoing activity. Decisions made with respect to "goals of the organism" (the "phlogiston" of the present discussion, alas) choose one of the potential OFCs offered by (6). This strategic computer (8) implements the chosen OFC through two major routes. One corresponds to the pyramidal tract which controls fine movements which take place in the context of grosser movements mediated by the other route. This second route addresses the effectors through the midbrain-maintained body-centered reference frame.

Schneider [1969] describes behavioral changes caused by selective ablations of the visual system of hamsters. Destroying visual cortex (breaking our block diagram at (5)) yields animals capable of orientation (locating) but not recognition. Destroying the superior colliculus (region (3) of our diagram) results in an animal which can discriminate patterns should its eyes accidentally rest on the pattern in standard position for the OFC selector, but is incapable of appropriate orienting movements.

Imagine a person presented with a 34 mm slide projection of an unknown scene - which is (say) upside-down. This case requires extensive analysis of one subpart of the environment. The observer must first realise that whatever it is he is looking at is "upside-down", then must either position his receptors (2) (perhaps by standing on his head) or introduce a shift in the output motor map (4) which will be reflected by

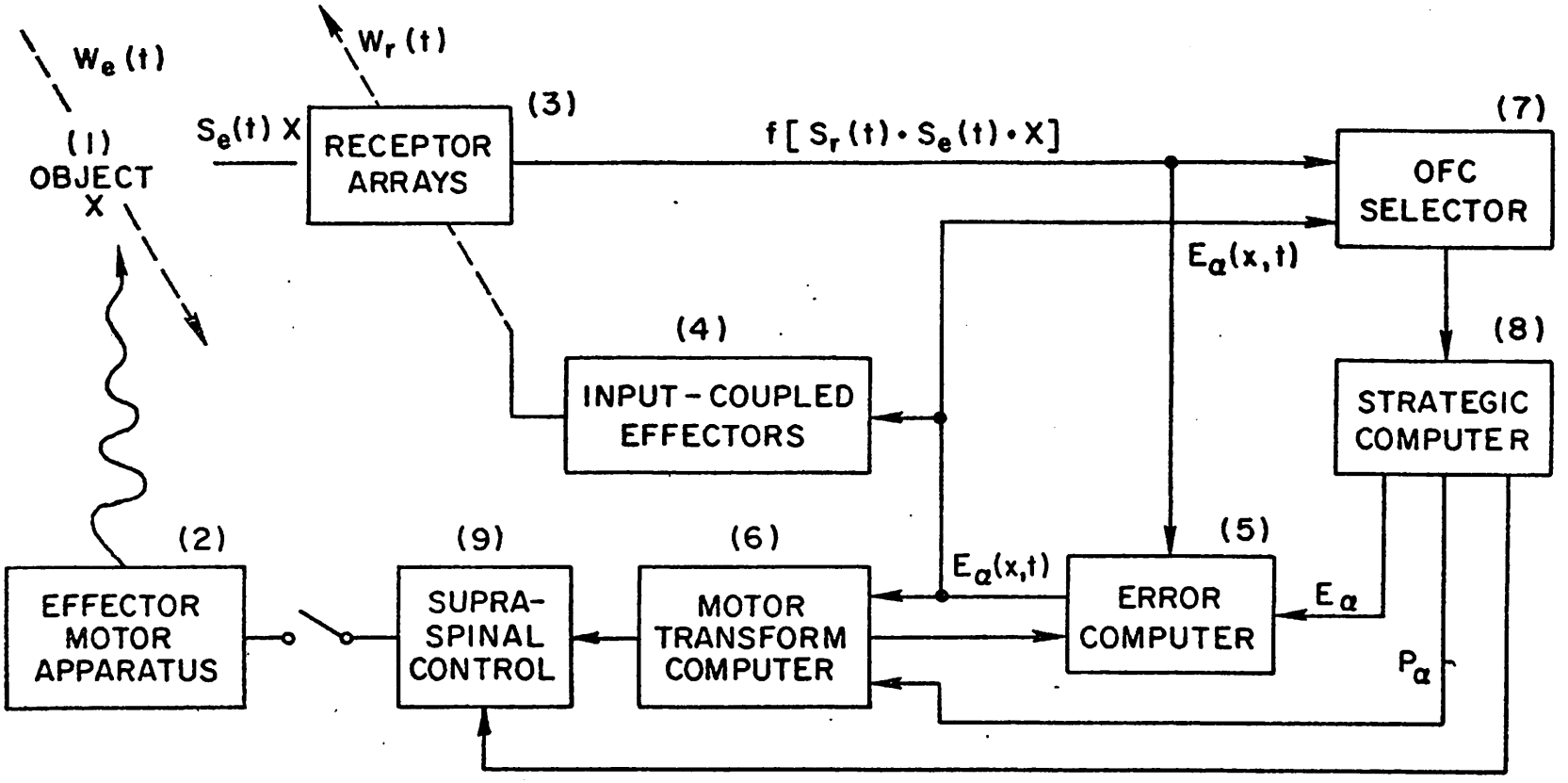

$W_e(t)$
$W_r(t)$
(1)
OBJECT
X
$S_e(t)$ X
(3)
RECEPTOR
ARRAYS
$f[S_r(t) \cdot S_e(t) \cdot X]$
(7)
OFC
SELECTOR
$E_\alpha(x,t)$
(4)
INPUT - COUPLED
EFFECTORS
(8)
STRATEGIC
COMPUTER
(2)
EFFECTOR
MOTOR
APPARATUS
(9)
SUPRA-
SPINAL
CONTROL
(6)
MOTOR
TRANSFORM
COMPUTER
$E_\alpha(x,t)$
(5)
ERROR
COMPUTER
E_α
P_α

Fig. 7.

the transformer (7) in the parametric adjustment of OFCs (6). Then he will proceed to make further transformations with the aim of bringing the slide into a standard form for whatever the scene depicts. This will entail receptor positioning, altering the motor map (and perhaps effector positioning) and the preprocessing functions so that the perceptual system will be "tuned" to whatever OFC hypothesis is being tried. This means that the boxes we have so blithely labelled "OFC selector (6)" and "strategic computer (8)" must be complex, and closely integrated to the other parts of the system.

Let us now concentrate on this aspect of perception which is incompletely filled in for lack of space in Fig. 6. For this further discussion refer to Fig. 7, which illustrates in more detail this role, and which generalises the Pitts-McCulloch [1947] scheme of our Fig. 4.

To ease the transition from Fig. 6 to Fig. 7, notice that for simplicity, sections (3) and (5) of Fig. 6 have been left out of Fig. 7, that the transformer (7) has been included in the motor map mechanism (4) and the influence of these two on OFC selection is indicated by the line in Fig. 7 from the "motor transform computer" (6) to the "error computer" (5). The error computer (5) of Fig. 7 was not drawn in Fig. 6, where it was included in the OFC selector (6) and strategic computer (8). "Supraspinal control" (8) has been added to exemplify the translation of an OFC, by stages, into individual muscle fibre commands.

Thus, in Fig. 7 we have decomposed the OFC generating and choosing mechanisms into those parts explicitly involved in finding a correct transformation and controlling movement by OFC parameterization from the rest of the scheme, and have hidden the anatomical distinction of the two different pathways seen in Fig. 6. In reading the discussion of Fig. 7,

the following should be noticed: Fig. 6 presumes that the system may always be said to be perceiving the scene in that the OFCs which would cause "let's walk around that thing and maybe we can see what in the world it could be" may be considered to be a percept of an object as something to walk around. Figure 7 separates out these initial exploratory parts of movement and receptor settings, making explicit how those processes could lead to "fuller" perception of an object ("oh, that's upside-down - it's a picture of the Washington Monument") essential for naming.

The object (box (1) in Fig. 7) has at any time some position represented by the operator $S_e(t)$, the change $\dot{W}_e(t)$ in which may be occasioned both by the autonomous motion of the object, and by the effects of the motor apparatus (2) upon the object. The position of the receptor arrays (3) yields an excitation pattern $f[S_r(t) \cdot S_e(t) \cdot X]$. The change $\dot{W}_r(t)$ in input array positioning is induced by the input-coupled effectors (4) which act - much as in the Pitts-McCulloch scheme - on the basis of an error vector supplied by the error computer (5). Similarly, the motor transform computer (6) generates the appropriate setting of supraspinal control on the basis of this error computer.

It is in discussing box (7) that we encounter new stubleties and something of a "chicken and egg problem". The problem is that we cannot find <u>the</u> appropriate standard form for an input pattern unless we have some idea of what object it represents - yet one of the aims of transforming the pattern to standard form is to identify the underlying object. We have suggested that an object is recognised when appropriate OFCs (as defined with respect to the organism's goals through the strategic computer box) are available to the system for interacting with it. With

each such OFC we associate an error criterion - an OFC is deemed to be appropriate for interaction with an object underlying an input pattern only when that pattern has been transformed to a form in which designated components (here is the relaxation we mentioned above) of the error vector are sufficiently close to zero, in which case the remaining components of the error vector serve as parameters for execution of the program. The error criterion must thus be determined both by the current goal of the system as well as the system's estimate of the class of objects to which the correctly scanned object belongs.

The OFC selector (7) conducts something akin to an heuristic search though at the level of choice of error criterion rather than of error reduction. Having generated the hypothesis that the appropriate OFC for controlling interaction is P_α, it provides the corresponding error criterion E_α to the error computer (5) whose output, the error vector $E_\alpha(X,t)$, is used by (4) and (6) to update the receptor position $S_r(t)$ and motor transformation $S_m(t)$. Box (7) keeps track of the resultant change in error to decide whether or not the current choice P_α is adequate. We assume that it and the environment are appropriately related so that a designated component $E^1_\alpha(X,t)$ becomes sufficiently close to zero to indicate that the current program P_α is applicable.

The strategic computer will allow a particular motor act to be executed only when $E^1_\alpha(x,t)$ is sufficiently small - it then sets up an "interpreter level" subroutine on the basis of the current program choice P_α and the parameters provided by the remaining components $E^2_\alpha(x,t)$ of the error vector. On the basis of this subroutine and error information, the motor transform computer (6) provides appropriate scaling information $S_m(t)$ - appropriateness being determined by its being fed back to the

error computer (5) and thus helping reduce $E^1_\alpha(x,t)$ towards zero. Finally, supraspinal control (9) interprets the "subroutine" provided by the strategic computer (8) using the information $S_m(t)$ provided by the motor transform computer (6).

Note that this scheme does not demand that we keep a complete neural record of the transformation T as is suggested by the version of the Pitts-McCulloch scheme we showed in Fig. 4, but allows motion of the motor apparatus to change the relative projection of the pattern, so that no absolute transformation is relevant. For example, Roberts [1967] has suggested that a cat, in moving its head to fixate a mouse, has concomitantly adjusted its motor apparatus ready for the spring.

The point is that the brain does not keep a record of all abstract transformations - rather it carries out operations which ensure that the relationship between organism and object is suitable for interaction.

In considering the changes $\mathcal{U}$ that can be made in any one unit of time of operation of the system, we shall assume that they comprise only a small subset of the set of possible transformations to which they belong. By definition, they must form a set of generators for that set. [In more detailed models, we would have to let the time unit in the loop be variable, and have some sort of tolerance relation on transformations.]

A crucial topological point here is that small displacements in space yield small changes in locus or level of excitation.

This ends our general discussion of how a perceiving system would actually use an action-oriented memory system to compute its motor activity. The appendix gives a preliminary sketch of how the "OFC selector"

might be actually designed using holographic principles. However, our attempts to use the general framework presented here to build upon our study of neural networks underlying action in the frog (Didday [1971], Arbib and Didday [1971]) to obtain a detailed theory of mammalian action-oriented memory has only just begun.

REFERENCES

1. M. A. Arbib (1971): "How We Know Universals: Retrospect and Prospect" Mathematical Biosciences, in press.

2. M. A. Arbib and R. L. Didday (1971): "What the Frog's Eye Tells the Frog", in press.

3. K. Craik (1943) The Nature of Explanation, Cambridge University Press.

4. R. L. Didday (1971) "A Possible Decision-Making Role for the 'Sameness' and 'Newness' Cells in the Frog", Brain Behaviour and Evolution, in Press.

5. J. J. Gibson (1968), The Senses Considered as Perceptual Systems, Allen and Unwin.

6. R. L. Gregory (1969) "On How so Little Information Controls so Much Behaviour," in Towards a Theoretical Biology, 2. Sketches (C. H. Waddington, Ed.) Edinburgh University Press.

7. D. Ingle, G. E. Schneider, C. B. Trevarthen and R. Held (1967), "Locating and Identifying: Two Modes of Visual Processing (A Symposium)", Psychologische Forschung, 31, Nos. 1 and 4.

8. W. L. Kilmer, W. S. McCulloch and J. Blum (1969) "A Model of the Vertebrate Central Command System," International Jour. Man-Machine Studies 1, 279-309.

9. W. L. Kilmer and T. McLardy (1970) "Hippocampal Circuitry", American Psychologist 25, 563-566.

10. K. Lashley (1929) Brain Mechanisms in Intelligence, Chicago University Press.

11. A. Newell, J. C. Shaw and H. A. Simon (1959) "Report on a General Problem-Solving Program", Proc. International Conf. Info. Proc., Paris: UNESCO, 256-264.

12. K. Pribram (1969) "The Neurophysiology of Remembering", Scientific American, January.

13. W. H. Pitts and W. S. McCulloch (1947): "How We Know Universals: The Perception of Auditory and Visual Forms", Bull. Math. Biophys. 9, 127-147.

14. T. Roberts (1967) The Neurophysiology of Postural Mechanisms, London, Butterworths.

15. G. E. Schneider (1969) "Two Visual Systems", Science, 163, pp. 895-902.

APPENDIX ON HOLOGRAPHY

This appendix has two purposes - to fill in some of the mathematical detail of our sketch of optical holography in Section 2; and then to suggest some ways to build from the hologram metaphor towards a theory of LTM and STM.

To produce a "real" optical hologram, a beam of laser light (i.e., coherent light) interferes with another laser beam that has been reflected from an object, and this interference pattern is recorded on a film plate. Let the function $O(x,y)$ describe the light reflected by the object, and $R(x,y)$ the reference beam which interferes with the object beam, as observed at the film plate. Then the intensity distribution incident on the film plate will be the squared amplitude of their sum:

$$I = [O(x,y) + R(x,y)] \cdot [O(x,y) + R(x,y)]^* \quad \text{(where } ^* \text{ indicates complex conjugate)}$$

$$= OO^* + OR^* + RO^* + RR^* \quad \text{(in shorthand notation)}$$

If the amplitude transmittance of the developed plate is proportional to the intensity, then the amplitude transmittance

$$t = OO^* + OR^* + RO^* + RR^*.$$

For reconstruction, a beam identical to the original reference beam is used. Therefore, output amplitude is

$$\text{output} = tR = OO^*R + OR^*R + RO^*R + RR^*R$$

If the reference beam is a plane wave,

$$R(x,y) = ae^{-j\frac{2\pi \sin\theta}{\lambda} \cdot x}$$

where **a** is the amplitude of the wave and θ is the angle of incidence on the plate.

then $R.R^* = a^2$

and $R.R = ae^{-j\frac{2}{\lambda}(2.\sin\theta)x}$

so that our above equation tells us that

$$\text{output} = (|O|^2 + a^2) \cdot ae^{-jk\sin\theta x} + O \cdot a^2 + O^* \cdot a^2 e^{-jk(2\sin\theta)x}$$

The first term is the zero order component. The second term is the original object wave, scaled by a constant and placed at θ degrees on one side of the zero order beam. The third term is the conjugate object wave, placed at θ degrees on the other side of the zero order. One of the arts of holography is to so arrange the apparatus that only the reconstructed object wave is viewed.

Instead of recording an object wave $O(x,y)$ with a plane reference wave $R(x,y)$, consider two object waves $O_1(x,y)$ and $O_2(x,y)$ recording each other. Then

$$t = O_1O_1^* + O_1O_2^* + O_2O_1^* + O_2O_2^*$$

and if O_1 is used for reconstruction, the output is:

$$O_1O_1^*O_1 + O_1O_2^*O_1 + O_2O_1^*O_1 + O_2O_2^*O_1 = |O_1|^2 \cdot O_1 + |O_1|^2 \cdot O_2 + O_1O_2^*O_1 + |O_2|^2 O_1$$

Note that, in the first two terms, O_1 and O_2 are reproduced (but with some amplitude modulation or distortion introduced by $|O_1|^2$ - the nearer O_1 is to a plane wave, the less this distortion is). This gives

primitive association memory where the input is simply one object wave but the output recalls an associated object.

It can be seen that even part of one of the object waves can be used to reconstruct that object completely as well as to recall the other. Suppose the wave O_1 is composed of two parts O_{1a} and O_{1b}, then we can consider the hologram to be a redording of three object waves O_{1a}, O_{1b} and O_2. Now, if O_{1a} is supplied as the reconstruction wave, O_{1b} and O_2 can be recalled. Thus a part of an object wave can be used to recall the whole object wave. However, the associated noise is increased.

The object wave $O(x,y)$ can be any form of representation of the object. If $O(x,y)$ is observed infinitesimally close to the object, there is a one-to-one correspondence between points on the object wave and points on the object. When $O(x,y)$ is observed infinitely far from the object, it is found to be the Fourier Transform of the object. Again, if the object is in the front focal plane of a lens, its Fourier Transform is observed at the back focal plane.

The Fourier Transform is a very natural transformation in optics, and is particularly useful in holography because of its shift invariant property. In computer generated holograms, however, other transforms like the Hadamard and the Haar are often used, simply for ease of computer implementation. Other completely different transforms may be considered, such as the feature transform mentioned in Section 2.

How distributed the memory on the hologram is depends on the transform used. In a Fourier Transform hologram, each point on the hologram represents frequency information from every part of the object and the hologram is completely distributed. When $O(x,y)$ represents an object

wave very close to the object - i.e., when the identity transformation is used - there is a point-to-point correspondence between object and object wave, and the hologram is not at all distributed.

An important restriction on most holograms is that accurate alignment of the hologram and the reconstruction wave is necessary. However, if the object wave $O(x,y)$ is the spatial Fourier Transform of the object (this Fourier Transform is obtained at the back focal plane of the lens if the object is at the front focal plane,) then the reconstruction wave for this hologram may be shifted from the position of the original reference wave. The image is still obtained but it is shifted by the same amount.

A property of photographic film is that transmittance decreases with exposure time. Over a certain range of exposure, the relation between the two is linear, but for overexposure, saturation occurs and transmittance is uniformly low. A hologram, made so as to use this fact, can act as a filter to accentuate those features of an object that make it different from a set of similar objects.

When taking an optical Fourier Transform of an object, most of the light is found to be concentrated in the lower frequencies. Hence, the difference between two objects resides, mostly, in the shape of the spectrum at the higher frequencies.

Take a set of similar objects, e.g., the letters of the alphabet or photographs of faces, all normalised with respect to size and record their Fourier Transforms, one after the other, on a film plate such that points on the film that have received more than a certain amount of light are all overexposed, so that overexposed points correspond to those frequencies

that are common to most of the objects. Now let the Fourier Transform of an object, similar to those in this set, impinge on the filter. All frequencies that the object has in common with those of the set will be filtered out. The frequencies that are left over go to build up those features in which the object differs from the objects in the set - a suggestive mechanism when we consider implementing the "subtactor" of Fig. 5.

When it comes to trying hologram-like techniques in designing neural nets to implement some of the schemes in Section 3, the reference beam should itself be appropriate to the memory it triggers. We saw that different images may be keyed by using different reference beams, but that in real holograms, these are chosen arbitrarily. We would suggest that in a neural analogue of a hologram (if such exists) the reference beam would comprise a sampling of ongoing neural activity, both peripheral (so that the animal can recall information about what happened when it did something similar) and central (so that we may recall thoughts related to our present ones). We could then have as a special case that suggested by Pribram (cf. Section 2) in which the reference beam is simply a somewhat earlier version of the present input - this sort of loop explains perfectly temporal recall of sequences, with each piece of information reading in the next at time of storage, and with each action being elicited by its predecessor at time of read-out. This is somewhat like classical association theory - save that the discussion in Section 3 emphasises the importance of structural decompositions of what an associationist might view as a mental unity.

We should also stress that not all information should be stored. Efferent control may operate to filter incoming information - in other

words, the reference beam may act to negate input information rather than store it, somewhat along the lines sketched in Fig. 5. (Note, too, that in acquiring a skill, "input" information for the neural hologram may be from central activity - "remember this idea" - or from proprioceptors - "remember how this feels".) Neural activity may elicit enough of a "band" of memories to force certain generalisations upon the organism - a given stimulus may activate all the experience related to certain modes of activity. Conversely, if an experience is "ordinary", it will be stored with a reference beam that can later elicit only noise (cf. our discussion of saturating filters above) so that current input will then predominate. [Why store it if it is ordinary? Perhaps it may be simpler to "store" such things irretrievably than carry out detailed computation on every item to decide whether or not it should be stored retrievably.]

One notion of a neural hologram which might serve as an apt metaphor to guide certain aspects of brain research may be presented in Fig. 7 - which may be viewed as a research proposal rather than a polished theory, as attested to by the following sketchy discussion: Feedback plays a crucial role in this system, since retrieved memories help determine the reference beam, allowing systematic searches through memory and maintaining those portions of STM that are not consequences of the control continually exerted by the current sequence of inputs. Dreaming is then a form of activity almost completely under the control of the previously retrieved memories - it is thus free from the control of "reality", but still possesses local continuity. When awake, too great a mismatch between input and current short-term memory induces a drastic recomputation that normally remains untriggered in dreaming.

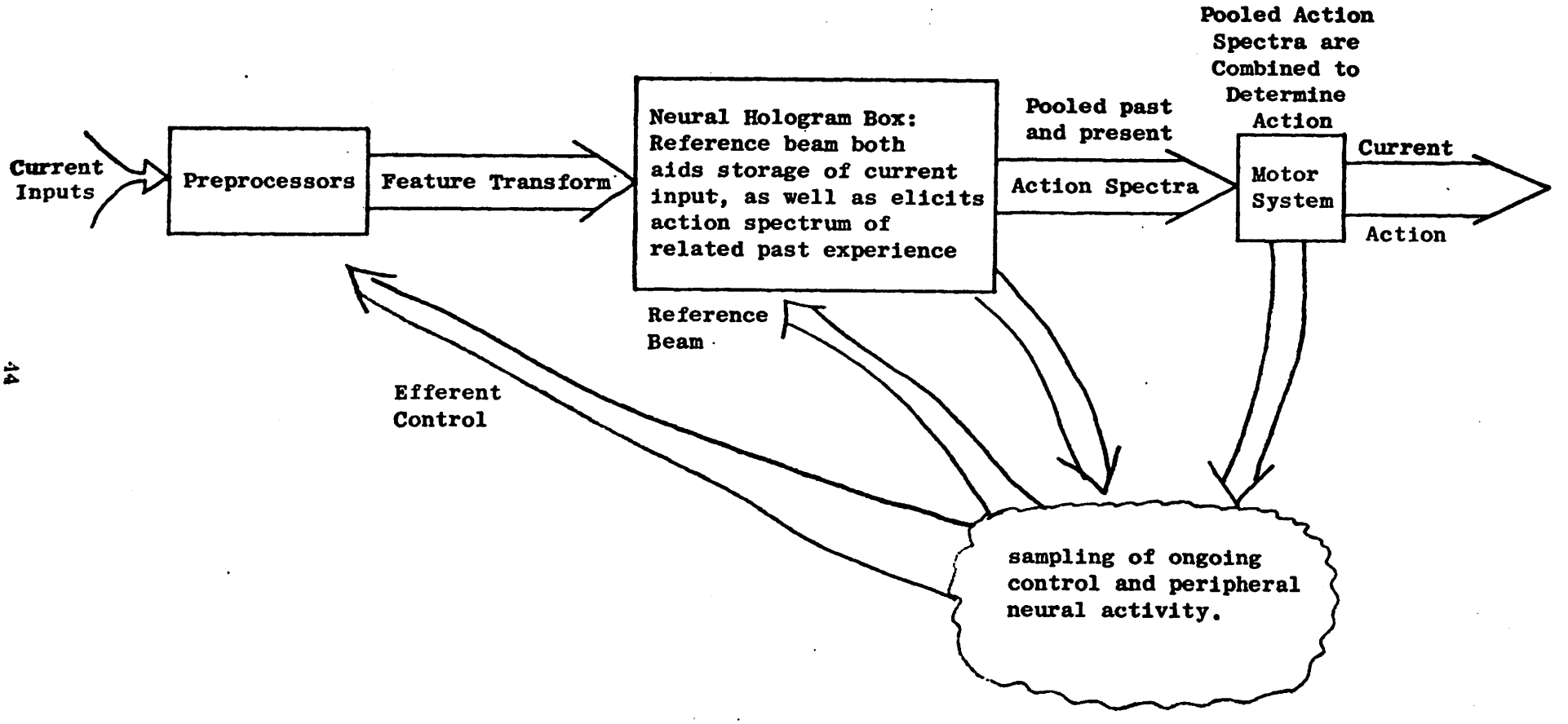


Fig. 8. An Extended Hologram Metaphor.

The input system extracts action-relevant features, and acts under efferent control to extract data relevant to our current ideas - among other things, data can be used here to address information required to reduce mismatch between short-term memory and what is relevant in the current environment. The reference beam can be used both to retrieve information, and to store further information.

The idea of focusing attention, or trying to concentrate to get something back, depends on the lower box to enhance computing certain features in the reference beam, specifically designed to get further information. We get into a loop of tighter and tighter interrogation to finally retrieve needed information. In the simplest case the reference-beam loop is just a delay, yielding recall of temporal sequences, because in this case a sequence is stored by using what happens at time n to encode what happens at time n+1, and so on.

In the above scheme, short-term memory may correspond to current activity around the loops, while long-term memory may correspond to changes in the connectivity of the neural hologram box. Humans with certain types of hippocampal damage cannot transfer information from short-term to long-term memory - perhaps the hippocampus corresponds to the "exposure control" for our neural hologram box. However, noting that the mathematics of complex valued waves required for optical holography differs drastically from the mathematics of neural spike trains, we see that it will be difficult enough to describe interference patterns for such wave trains, let alone model the analogue of exposing photographic film.

Irrespective of the conjectures we might make about functional block diagrams for human memory, neural holography should provide a useful metaphor if we avoid the temptation to use it literally (e.g., recreating visual input) but instead exploit the idea of portions of a wave activity helping recreate the whole wavefront, with different cuing waves allowing multiple storage in a region of brain tissue. This should serve as an antidote to the word-by-word view of memory we get from thinking about digital computers or the linguistic abilities of humans.