Deictic Codes for the Embodiment of Cognition

Dana H. Ballard, Mary M. Hayhoe, Polly K. Pook and Rajesh P. N. Rao
Computer Science Department
University of Rochester
Rochester, NY 14627, USA
({dana, mary, pook, rao}@cs.rochester.edu)

Short Abstract

To describe phenomena that occur at different time scales, computational models of the brain must necessarily incorporate different levels of abstraction. We argue that at time scales of approximately one-third of a second, orienting movements of the body play a crucial role in cognition and form a useful computational level, termed the *embodiment level*. At this level, the constraints of the body determine the nature of cognitive operations, since the natural sequentiality of body movements can be matched to the natural computational economies of sequential decision systems. The way this is done is through a system of implicit reference termed *deictic*, whereby pointing movements are used to bind objects in the world to cognitive programs. We show how deictic bindings enable the solution of natural tasks and argue that one of the central features of cognition, working memory, can be related to moment-by-moment dispositions of body features such as eye movements and hand movements.

Keywords: deictic computations; embodiment; working memory; natural tasks; eye movements; brain computation; binding; sensory-motor tasks; pointers.

Long Abstract

To describe phenomena that occur at different time scales, computational models of the brain must necessarily incorporate different levels of abstraction. We argue that at time scales of approximately one-third of a second, orienting movements of the body play a crucial role in cognition and form a useful computational level. This level is more abstract than that used to capture neural phenomena yet is framed at a level of abstraction below that traditionally used to study high-level cognitive processes such as reasoning. We term this level the *embodiment level*. At the embodiment level, the constraints of the physical system determine the nature of cognitive operations. The key synergy is that, at time scales of about one-third second, the natural sequentiality of body movements can be matched to the natural computational economies of sequential decision systems. The way this is done is through a system of implicit reference termed *deictic*, whereby pointing movements are used to bind objects in the world to cognitive programs. The focus of this paper is to study how deictic bindings enable the solution of natural tasks. We show how deictic computation provides a mechanism for representing the essential features that link external sensory data with internal cognitive programs and motor actions. In particular, we argue that one of the central features of cognition, working memory, can be related to moment-by-moment dispositions of body features such as eye movements and hand movements.

1 Embodiment

This paper is an attempt to describe the cognitive functioning of the brain in terms of its interactions with the rest of the body. Our central thesis is that intelligence has to relate to interactions with the physical world, and that means that the particular form of the human body is a vital constraint in delimiting many aspects of intelligent behavior.

On first consideration, the assertion that the aspects of body movements play a vital role in cognition might seem unusual. The tenets of logic and reason demand that these formalisms can exist independently of body aspects and that intelligence can be described in purely computational terms without recourse to any particular embodiment. From this perspective, the special features of the human body and its particular ways of interacting in the world are seen as secondary to the fundamental problems of intelligence. However, the world of formal logic is often freed from the constraints of process. When the production of intelligent behavior by the body-brain system is taken into account, the constraints of time and space intervene to limit what is possible. We will argue that at time scales of approximately one-third of a second, the momentary disposition of the body plays an essential role in the brain's symbolic computations. The body's movements at this time scale provide an essential link between processes underlying elemental perceptual events and those involved in symbol manipulation and the organization of complex behaviors.

To understand the motivation for the one-third second time scale, one first must understand the different time scales that are available for computation in the brain. Since the brain is a physical system, communicating over long distances is costly in time and space and thus local computation is the most efficient. Local computation can be utilized effectively by organizing systems hierarchically [Newell, 1990]. Hierarchical structure allows one to tailor local effects to the most appropriate temporal and spatial scales.[1] In addition, a hierarchical organization may be necessary for a complex system to achieve stability [Simon, 1962]. Newell has pointed out [Newell, 1990] that whenever a system is constructed of units that are composed of simpler primitives, the more abstract primitives are necessarily larger and slower. This is because within each level in a hierarchical system there will be sequential units of computation that must be composed to form a primitive result at the next level. In fact, with increasing levels of abstraction, the more abstract components run slower at geometric rates. This constraint provides a context for understanding the functioning of the brain and the organization of behavior by allowing us to separate processes that occur at different time scales and different levels of abstraction.

Consider first the communication system between neurons. Almost all neurons communicate by sending electrical spikes that take about one millisecond to generate. This means that the circuitry that uses these spikes for computation has to run slower than this rate. Let us use Newell's

[1] One can have systems for which local effects can be of great consequence at long scales. Such systems are termed chaotic [Baker and Gollub, 1990], but these systems cannot be easily utilized in goal-directed computation. The bottom line is that for any physical system to be manageable, it must be organized hierarchically.

Abstraction Level	**Temporal Scale**	**Primitive**	**Example**
Cognitive	2-3 sec	Unit Task	Dialing a phone number
Embodiment	*0.3 sec*	*Physical Act*	*Eye movement*
Attentive	50 msec	Deliberate Act	Noticing a stimulus
Neural	10 msec	Neural Circuit	Lateral inhibition
Neural	1 msec	Neuron Spike	Basic signal

Table 1: The organization of human computation into temporal bands (after [Newell, 1990] but with some time scales adjusted to account for experimental observations).

assumption that about ten operations are composed at each level. Then local cortical circuitry will require 10 milliseconds. These operations are in turn composed for the fastest "deliberate act," in Newell's terminology. A primitive deliberate act is then 100 milliseconds. A deliberate act would correspond to any kind of perceptual decision, e.g., recognizing a pattern, a visual search operation, or an attentional shift. The next level is the physical act. Examples of primitive physical acts would include an eye movement, a hand movement, or a spoken word. Composing these results is a primitive task, which defines a new level. Examples of this level would be uttering a sentence or any action requiring a sequence of movements, such as making a cup of coffee or dialing a telephone number. Another example would be a chess move. Speed chess is played at about 10 seconds per move.[2]

Newell's "ten-operations" rule is very close to experimental observations. Simple perceptual acts such as an attentional shift or pattern classification take several tens of milliseconds, so the 100 milliseconds estimate is probably within a factor of 2 or 3 of the correct value. Speeds of body movements show that eye movements take about 0.2-0.3 sec to generate, which is about 5 times the duration of a perceptual decision. One can see that the composition of tasks by primitive acts requires the persistence of the information in time. Thus the demands of task composition require some form of working memory. Human working memory has a natural decay constant of a few seconds, so this is also consistent with a hierarchical structure. Table 1 shows these relations.

Our focus is the one-third second time scale, which is the shortest time scale at which body movements such as eye movements can be observed. We argue that this time scale defines a special level of abstraction, which we term the *embodiment level*. At this level, the appropriate model of computation is very different from those that might be used at shorter or longer time scales.

[2]Another reason that the modeling methods here are different than the traditional symbol manipulation used in AI is that the time scales are much shorter, too short to form a symbol.

Computation at this level governs the rapid deployment of the body's sensors and effectors in order to bind variables in behavioral programs. As such this computation provides a language which represents the essential features that link external sensory data with internal cognitive programs and motor actions. In addition this language provides an interface between lower-level neural "deliberate acts" and higher-level symbolic programs. The ramifications of this view are several.

- Cognitive and perceptual processes cannot be easily separated, and in fact are interlocked for reasons of computational economy. The products of perception are integrated into distinct serial sensory-motor primitives, each taking a fraction of a second. This viewpoint is very compatible with Arbib's perception-action cycle [Arbib, 1981, Arbib et al., 1985, Fuster, 1989], but with the emphasis on (a) the 1/3 sec time scale; and (b) sensory motor primitives. For problems that take on the order of many seconds to minutes to solve, many of these sensory-motor primitives must be synthesized into the solution.

- The key constraint is the number of degrees of freedom, or variables, needed to define the ongoing cognitive programs. We argue that this is a useful interpretation of the role of working memory. Furthermore the brain's programs structure behaviors so as to minimize the amount of working memory needed at any instant. The structure of working memory and its role in formation of long-term memories has been extensively examined [Baddeley, 1986, Logie, 1995]. Our focus is different: the rapid accessing of working memory during the execution of behavioral programs.

- The function of the sensory-motor primitives is to load or bind the items in working memory. This can be done by accessing the external environment or long-term memory. Items are only bound for as long as they are needed in the encompassing task. In addition the contents of an item varies with task context, and may be only fragmentary portions of the available sensory stimulus.

1.1 Deictic Sensory-Motor Primitives

A ubiquitous example of a rapid sensory-motor primitive is the saccadic eye movement system. Saccadic eye movements are typically made at the rate of about three per second and we make on the order of 10^5 saccades per day. Eye fixations are at the boundary of perception and cognition, in that they are an overt indicator that information is being represented in cognitive programs. Attempts to understand the cognitive role of eye movements have focused either on the eye movement patterns, as did Noton and Stark in their study of "scanpaths" [Noton and Stark, 1971b] and Simon and Chase in their study of eye movement patterns in chess [Chase and Simon, 1973], or on the duration of fixation patterns themselves (e.g., [Just and Carpenter, 1976]). But as Viviani points out [Viviani, 1990], the crux of the matter is that one has to have an independent way of assessing cognitive state in addition to the underlying overt structure of the eye scanning patterns.

For that reason studies of reading have been the most successful [Pollatsek and Rayner, 1990], but these results do not carry over to general visual behaviors. Viviani's point is crucial: one needs to be able to relate the actions of the physical system to the internal cognitive state. One way to start to do this is to posit a general role for such movements, irrespective of the particular behavioral program. The role we posit here is *variable binding*, and it is best illustrated with the eye movement system.

Since humans can fixate an environmental point, their visual system can directly sample portions of three-dimensional space, as shown in Figure 1, and as a consequence, the brain's internal representations are implicitly referred to an external point. Thus neurons tuned to zero-disparity at the fovea refer to the instantaneous, exocentric three-dimensional fixation point. The ability to use an external frame of reference centered at the fixation point that can be rapidly moved to different locations leads to great simplifications in algorithmic complexity [Ballard, 1991].[3] For example, an object is usually grasped by first looking at it and then directing the hand to the center of the fixation coordinate frame [Jeannerod, 1988, Milner and Goodale, 1995]. For the terminal phase of the movement, the hand can be servoed in depth relative to the horopter by using binocular cues. Placing a grasped object can be done in a similar manner. The location can be selected using an eye fixation and then that fixation can be used to guide the hand movement. Informally, we refer to these behaviors as "do-it-where-I'm-looking" strategies, but more technically they are referred to as *deictic* strategies after [Agre and Chapman, 1987], building on work by [Ullman, 1984]. The word deictic means "pointing" or "showing." Deictic primitives dynamically refer to points in the world with respect to their crucial describing features (e.g., color or shape). The dynamic nature of the referent also captures the agent's momentary intentions. In contrast, a non-deictic system might construct a representation of all the positions and properties of a set of objects in viewer-centered coordinates, and there would be no notion of current goals.

Vision is not the only sense that can be modeled as a deictic pointing device. Haptic manipulation, which can be used for grasping or pointing, and audition, which can be used for localization, can also be modeled as localization devices. We can think of fixation and grasping as mechanical pointing devices, and localization by attention as a neural pointing device. Thus one can think of vision as having either mechanical or neural deictic devices: fixation and attention. This paper emphasizes the deictic nature of vision, but the arguments hold for the other sensory modalities as well.

1.2 The Computational Role of Deictic Reference

Although the human brain is radically different in many ways from conventional silicon computers, they both have to address many of the same problems. Thus it is sometimes useful to look at how

[3] This is a very different assertion than that of [Marr, 1982], who emphasized that vision calculations were initially in viewer-centered coordinates and did not address the functional role of eye movements.

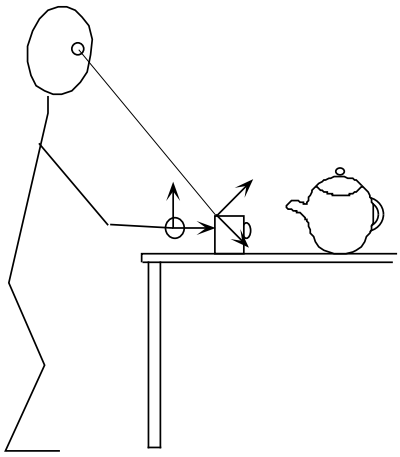


Figure 1: Biological and psychophysical data argue for deictic frames. These frames are selected by the observer to suit information-gathering goals.

problems are handled by silicon computers. One major problem is that of variable binding. As recognized by Pylyshyn [Pylyshyn, 1989] in his FINST studies, for symbolic computation it is often necessary to have a symbol denote a very large number of bits, and then modify this reference during the course of a computation. Let us examine how this is done using an artificial example.

Table 2 shows a hypothetical portion of memory for a computer video game[4] where a penguin has to battle bees. The most important bee is the closest, so that bee is denoted, or pointed to, with a special symbol "the-bee-chasing-me." The properties of the lead bee are associated with the pointer. That is, conjoined with the symbol name is an address in the next word of memory that locates the properties of the lead bee. In the table this refers to the contents of location 0001, which is itself an address, pointing to the location of beeA's properties, the three contiguous entries starting at location 0011. Now suppose that beeB takes the lead. The use of pointers vastly simplifies the necessary bookkeeping in this case. To change the referent's properties, the contents of location 0001 are changed to 1000 instead of 0011. Changing just one memory location's contents accomplishes the change of reference. Consider the alternative, which is to have all of the properties of the "the-bee-chasing-me" in immediately contiguous addresses. In that case, to switch to beeB, all of the latter's properties have to be copied into the locations currently occupied by beeA. Using pointers avoids the copying problem.

It should be apparent now how deictic reference, as exemplified by eye fixations, can act as a pointer system. Here the external world is analogous to computer memory. When fixating a location, the neurons that are linked to the fovea refer to information computed from that location.

[4]For technical details see [Agre and Chapman, 1987].

Changing gaze is analogous to changing the memory reference in a silicon computer. Physical pointing with fixation is a technique that works as long as the embodying physical system, the gaze control system, is maintaining fixation. In a similar way the attentional system can be thought of as a neural way of pointing. The center of gaze does not have to be moved, but the idea is the same: to create a momentary reference to a point in space, so that the properties of the referent can be used as a unit in computation. The properties of the pointer referent may not be, and almost always are not, all those available from the sensors. The reason is that the decision-making process is greatly simplified by limiting the basis of the decision to essential features of the current task.

Both the gaze control system and neural attentional mechanisms each dedicate themselves to processing a single token. If behaviors require additional variables, these must be kept in a separate system termed working memory [Baddeley, 1986, Broadbent, 1958, Logie, 1995]. Although the brain works on very different principles than a computer, the problem faced is the same. In working memory the references to the items therein have to be changed with the requirements of the ongoing computation. The strategy of copying that was used as a straw man in the silicon example is even more implausible here, as most neurons in the cortex exhibit a form of place coding [Barlow, 1972, Ballard, 1986] that cannot be easily changed. Thus it seems that at the one-third second time scale, ways of temporarily binding huge numbers of neurons and changing those bindings must exist. That is, the brain must have some kind of pointer mechanism.[5]

1.3 Outline

The focus of this paper is to explain why deictic codes are a good model for behavior at the embodiment level. The presentation is organized into three main sections.

- **Section 2** argues that the computational role of deictic codes or pointers is to represent the essential degrees of freedom used to characterize behavioral programs. Several different arguments suggest that there are computational advantages to using the minimum number of pointers at any instant.
- **Section 3** discusses the psychological evidence in favor of deictic strategies. Studying a simple sensory-motor task provides evidence that working memory is intimately involved in describing the task and that working memory is reset from moment to moment with deictic actions.
- **Section 4** discusses the implications of deictic computation in understanding cortical circuitry. A consequence of complex programs being composed of simpler primitives, each of

[5]Several mechanisms for such operations have been proposed [Koch and Crick, 1994, Buhmann et al., 1990, Shastri, 1993], but as of yet there is not sufficient data to resolve the issue.

Address	Contents
0000	the-bee-chasing-me
0001	**0011**
0010	
0011	beeA's weight
0100	beeA's speed
0101	beeA's # of stripes
0110	
0111	
1000	beeB's weight
1001	beeB's speed
1010	beeB's # of stripes
1011	

Address	Contents
0000	the-bee-chasing-me
0001	**1000**
0010	
0011	beeA's weight
0100	beeA's speed
0101	beeA's # of stripes
0110	
0111	
1000	beeB's weight
1001	beeB's speed
1010	beeB's # of stripes
1011	

Table 2: A portion of computer memory illustrating the use of pointers. Left: Reference is to beeA. Right: Reference is to beeB. The change in reference can be accomplished by changing a single memory cell.

which involves sensory-motor operations, is that many disparate areas of the brain must interact in distinct ways to achieve special functions. Some of these operations bind parts of the sensorium and others use these bindings to select the next action.

2 Deictic Representation

Deictic representation is a system of implicit reference, whereby the body's pointing movements are used to bind objects in the world to cognitive programs. The computational role of deictic pointing is to represent the essential degrees of freedom used to characterize behavioral programs. This section shows how distilling the degrees of freedom down to the minimum allows simple decision making. The essential degrees of freedom can have perceptual, cognitive, and motor components. The perceptual component uses deictic pointing to define the context for the current behavioral program. The cognitive component maintains this context as variables in working memory. The motor component uses the working memory variables to mediate the action of effectors.

2.1 Deictic Models of Sensory Processing

The primary example of a deictic sensory action is fixation. There are a number of indications from human vision that fixation might have theoretical significance. Fixation provides high-resolution

Figure 2: A schematic of Kowler and Anton's experiment: Subjects reading text normally fixate words only once, but when the letters are reversed, each letter is fixated [Kowler and Anton, 1987].

in a local region since the human eye has much better resolution in a small region near the optical axis, i.e., the fovea. Over a region of approximately one to two degrees of visual angle the resolution is better by an order of magnitude than in the periphery. One feature of this design is the representation of local high acuity within a larger field of view. This makes it ideal as a pointing device to denote the relevant parts of the visible environment.

Given the high-resolution fovea, one might be tempted to conclude that the primary purpose of fixation is to obtain better spatial resolution. That certainly is an important consequence of fixation but is almost certainly not its only role. One indication of its computational role is given in a study by [Kowler and Anton, 1987], who measured fixation patterns while reading text of reversed letters (see Figure 2). In normal text, individual words are contained within the fovea and are fixated only once. With the reversed letters, however, individual letters were fixated, resulting in several fixations per word. Note that in this case the function of fixation cannot be for increased resolution as individual words can be resolved within a single fixation. It must be the case that fixation is serving some technical function in recognizing the reversed letters beyond that of improving spatial resolution. In this case, the letter is the appropriate pattern recognition unit. Other evidence for the importance of fixations in visual computations comes from the findings of [Schlingensiepen et al., 1986, Just and Carpenter, 1976], who showed that eye movements appear to be required for making same/different judgments of complex patterns. Investigations of chess playing [Chase and Simon, 1973] have also indicated that eye fixations are intimately related to spatial working memory.

Our contention is that in each of the above examples deictic primitives simplify complex behaviors, because each sensory-motor primitive defines the context for its successor using only the information immediately fixated or attended. This idea was tested in computer simulations, where an abstract hand-eye "robot" learned simple block manipulations.[6] For example, consider the problem of picking up a green block that has another block stacked on top of it, shown in Figure 3 (from

[6]Of course learning by repeated trials is not the way a human does this task. The point is rather that the reduced information used by a deictic strategy is *sufficient* to learn the task.

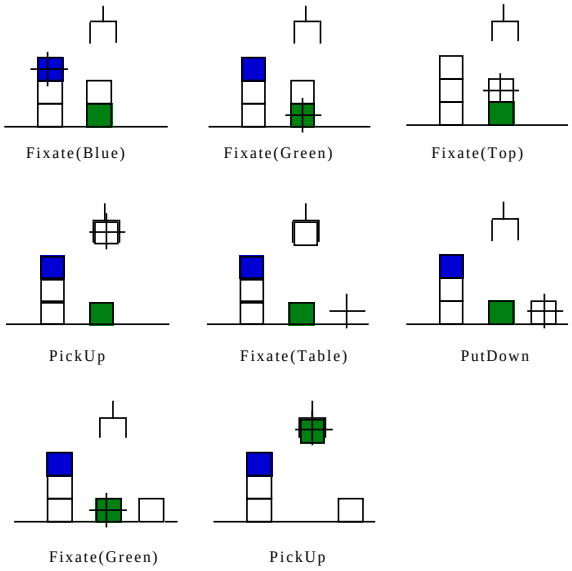


Figure 3: A graphical display from the output of a program that has learned the "Pick up the green block" task. The steps in the program use deictic references rather than geometrical coordinates. For each stage in the solution, the plus symbol shows the location of the fixation point.

[Whitehead and Ballard, 1991]). This problem is solvable by computer simulation using reinforcement learning [Whitehead and Ballard, 1990].[7] The result is the following sequence of actions or program:

```
Fixate(Green)
Fixate(Top-of-where-I'm-looking)
Pickup
Fixate(Somewhere-on-the-table)
Putdown
Fixate(Green)
Pickup
```

To work this program needs more than just the ability to fixate. The reason is that the deictic representation only uses the fixated objects to make decisions. Thus when fixating a blue block, it is

[7] Of the three current models of neural computation—reinforcement learning [Barto et al., 1990], neural networks [Hertz et al., 1991], and genetic algorithms [Koza, 1992, Goldberg, 1989]— reinforcement learning explicitly captures discrete, sequential structure as a primitive. As such it is a good model for integrating cognitive portions of a behavioral program with observed characteristics of the brain [Woodward et al., 1995, Schultz et al., 1995].

Bits	Feature
1	red-in-scene
1	green-in-scene
1	blue-in-scene
1	object-in-hand
2	attended-color(red, green, blue)
1	fixated-shape(block, table)
2	fixated-stack-height(0, 1, 2, >2)
1	table-below-fixation-point
1	fixating-hand
2	attended-color(red, green, blue)
1	attended-shape(block, table)
2	attended-stack-height(0, 1, 2, >2)
1	table-below-attention-point
1	attending-hand
1	fixation-and-attention-horizontally-aligned
1	fixation-and-attention-vertically-aligned

Table 3: The sensory representation used to solve the blocks task. The representation consists of four global features and twelve features accessed by the fixation and attention pointers. The leftmost column shows the number of bits used to represent each feature. The rightmost column describes each feature.

not clear what to do. If the block is at the top of the stack the next operation should be `PickUp`, but if it is on the table, the next instruction should be `Fixate(Green)`. [Whitehead and Ballard, 1990] have shown that this problem can be resolved by using an additional deictic mechanism in the form of a visual focus of attention. The complete sensory representation is shown in Table 3 and the repertoire of actions is shown in Table 4. This allows the program to keep track of the necessary context, as the attended object can be different in the two different cases. This in turn allows the task to be completed successfully.

In the program it is assumed that the instruction `Fixate(Image_feature)` will orient the center of gaze to point to a place in the image with that feature; the details of how this could be done are described in Section 4. These actions are context sensitive. For example, `Pickup` and `Putdown` are assumed to act at the center of the fixation frame. `Fixate(Top-of-where-I'm-looking)` will transfer the gaze to the top of the stack currently fixated.

Fixation-Relative Actions
PickUp
Drop
Fixate(Red)
Fixate(Green)
Fixate(Blue)
Fixate(Table)
Fixate(Top-of-where-I'm-looking)
Fixate(Bottom-of-where-I'm-looking)
Attention-Relative Actions
Attend(Red)
Attend(Green)
Attend(Blue)
Attend(Table)
Attend(Top-of-where-I'm-looking)
Attend(Bottom-of-where-I'm-looking)

Table 4: The discrete set of actions used to solve the blocks task. At each point in time the program has to select an action from this repertoire. The program is rewarded for finding a sequence of such actions that solves the task.

2.2 Adding Memory to Deictic Models

In the green block task, only two such pointers were required: "fixation" and "attention." For more complicated tasks, however, more pointers may be needed. For example, consider Chapman's example of copying a tower of colored blocks, each identified with a color, as shown in Figure 4 [Chapman, 1989]. To do this task, one pointer keeps track of the point in the tower being copied, another keeps track of the point in the copy, and a third is used for manipulating the new block that is part of the copy. The pointers provide a dynamic referent for the blocks that is action specific.[8]

The key advantage of the pointer strategy is that it scales well. Only three pointers are needed regardless of the tower height. This is the important claim of pointer-based behavioral programs: Complex tasks can be decomposed into small behavioral primitives, where the necessary state needed to keep track of the process can be represented with just the temporal history of a handful of pointers. In other words, our claim is that almost all complex behaviors can be performed with just a few pointers.

Now we can make the crucial connection between the computational and psychological domains. If a task can be solved using only one or two pointers, then this can be handled by explicit pointing, such as fixation, or "neural" pointing, such as "attention." However, additional pointers require some additional representational mechanism. A plausible psychological mechanism is that of working memory. Working memory items may be thought of as corresponding to computational pointers. A pointer allows access to the contents of an item of memory.

The pointer notation raises the issue of binding, or setting the pointer referent. This is because pointers are general variables that can be re-used for other computations. When are pointers bound? For a visual pointer one possible indication that it is being set could be fixation. Looking directly at a part of the scene provides special access to the features immediate to the fixation point, and these could be bound to a pointer during the fixation period. In all likelihood binding can take place faster than this, say by using an attentional process, but using fixation as a lower bound would allow us to bind several pointers per second with a capacity determined by the decay rate of the activated items.

Even though a problem can be solved with a small number of pointers, why should there be pressure to use the minimal-pointer solution? One argument for minimal-pointer programs can be made in terms of the cost of finding alternate solutions, which is often characterized as the credit assignment problem. To illustrate this problem consider two new tasks. Suppose that the task of

[8] A non-deictic way to solve this task would be to process the scene so that each block is cataloged and has a unique identification code. Then the copying task could be solved by the searching space of all possible relationships among coded items for the right ones. The problem is that this strategy is very expensive, as the number of possible relationships among different configurations of blocks can be prohibitively large. For all possible configurations of just 20 blocks, 43 billion relationships are needed! In contrast, a deictic strategy avoids costly descriptions by using pointers that can be re-assigned to blocks dynamically (possible mechanisms for this are described in Section 4).

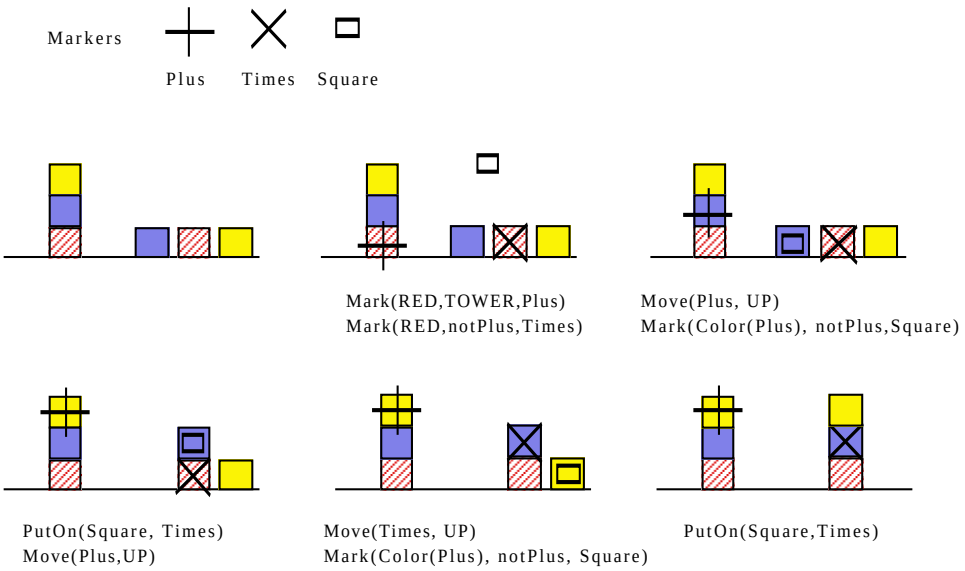


Figure 4: A tower of blocks can be copied using three pointers. At any instant the access to the sensorium is limited to the marked objects. The number of pointers is minimized by re-using pointers during the course of the behavior.

picking up a green block is changed to that of picking up a yellow block and the tower-copying task is changed so that the colors must be copied in reverse order. If the possibilities scale as a function of the number of required pointers, the sequence of actions for the first task is easier to discover. If we assume a sequential search model, such as that postulated in reinforcement learning models, then the cost of searching for an alternative solution to a problem could potentially scale as $(MV)^s$ where M is the number of pointers, V is the number of visual/manual routines, and s is the number of steps in the program. Thus the central problem may be just that the cost of searching alternatives scales badly with an increasing number of pointers. This may result in a tremendous pressure on finding behavioral programs that only require small numbers of pointers.

A second reason for only having a small number of pointers is that it may be sufficient for the task. McCallum [McCallum, 1995b] builds a "history tree" which stores the current action as a function of the immediate history of an agent's observations and actions. The idea of a history tree for a simple maze problem is illustrated in Figure 5. In a simple maze the agent must find a goal site but only senses the immediately surrounding four walls (or lack of them). Thus the actions at ambiguous states are resolved by additional history. McCallum has extended this learning algorithm to a model of highway driving and shown that the required number of features in short-term memory ranges from 2 to 14 [McCallum, 1995a]. These simulations suggest that impressive performance may be achievable with very little context.

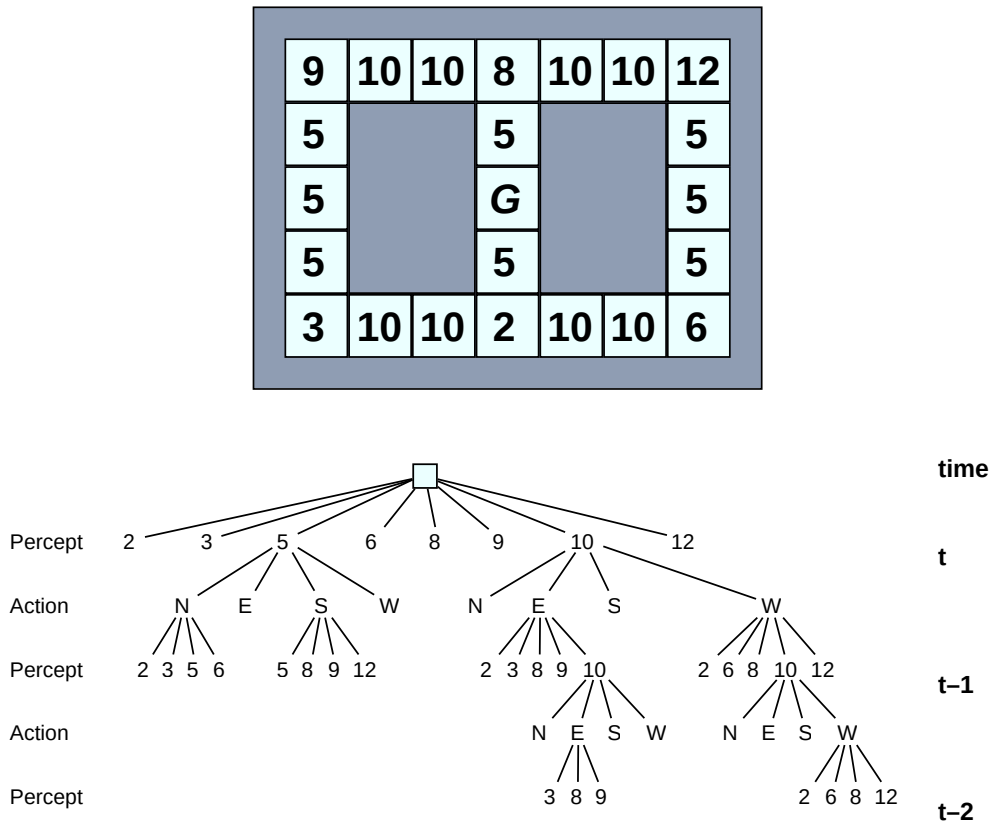


Figure 5: The different amounts of context used in decision making can require different amounts of working memory. (Top) Maze used in McCallum's reinforcement learning algorithm. The numerical codes indicate the walls surrounding each location. For example, a north and south wall is coded as a "10." (Bottom) After learning, the actions to take are stored in a tree that records the agent's previous history. The best action is stored at the leaf of the tree. For example, knowing what to do given the current percept is a "2" can be decided immediately (go North) but if the current percept is a "5," the previous history of actions and perceptions is required to resolve the situation (Some tree branches are not used e.g. the "E" below the "5").

A third reason for a small number of pointers may be that it reflects a balance between the natural decay rate of the marked items and the temporal demands of the task. In Section 1 we described the composition of cognitive tasks from components at a lower level, which we called "physical acts." To compose new behaviors in this way there must be some way of keeping the components active long enough to compose the new behavior. At the same time it seems likely that if the component items are active for too long they will no longer be relevant for the current task demands and may interfere with the execution of the next task. (The extraordinary case of Luria's patient S, whose sensory memory was excessively persistent, attests to such interference [Luria, 1968].) Thus we can see that the capacity of working memory will be a consequence of the natural decay rate of marked (activated) items and should reflect the dynamic demands of the task.

2.3 Deictic Motor Routines

Deictic variables (such as fixation on a visual target) can define a relative coordinate frame for successive motor behaviors. To open a door, for instance, fixating the doorknob during the reach defines a stable relative servo target that is invariant to observer motion. Use of a relative frame relevant to the ongoing task avoids the unwanted variance that occurs when describing movement with respect to world-centered frames. [Crisman and Cleary, 1994] demonstrate the computational advantage of target-centered frames for mobile robot navigation. In humans it is known that a variety of frames are used for motor actions [Soechting and Flanders, 1989, Andersen, 1995, Jeannerod, 1988] but the computational role of such frames is less studied. This section illustrates the computational advantages of deictic variables using simulations with robot hardware. We do this using a strategy we term *teleassistance* [Pook and Ballard, 1994b]. In teleassistance, a human operator is the "brain" to an otherwise autonomous dextrous robot manipulator. The operator does not directly control the robot but rather communicates symbolically via a deictic sign language shown in Table 5. A sign selects the next motor program to perform and tunes it with hand-centered pointers. This example illustrates a way of decoupling the human's link between motor program and reflexes. Here the output of the human subject is a deictic code for a motor program that a robot then carries out. This allows the study of the use and properties of the deictic code.

The sign language is very simple. To assist a robot to open a door requires only the three signs shown in Table 5. Pointing to the door handle prompts the robot to reach toward it and provides the axis along which to reach. A finite state machine (FSM) for the task specifies the flow of control. This embeds sign recognition and motor response within the overall task context.

Pointing and preshaping the hand create hand-centered spatial frames. Pointing defines a relative axis for subsequent motion. In the case of preshaping, the relative frame attaches within the opposition space [Arbib et al., 1985] of the robot fingers.[9] With adequate dexterity and compliance,

[9]Since morphology determines much of how hands are used, the domain knowledge inherent in the shape and frame position can be exploited. For example, a wrap grasp defines a coordinate system relative to the palm.

Sign	Meaning
POINT	While the operator points, the robot moves in the direction of the pointing axis, independently of world coordinates. Thus the robot reach is made relative to a deictic axis that the teleoperator can easily adjust.
PRESHAPE	A grasp *preshape* defines a new spatial frame centered on the palm of the hand. The operator preshapes her hand to define a grasp form and a new spatial frame centered on the palm.
HALT	Halting is used to punctuate the motor program.

Table 5: Signs used in teleassistance experiment.

simply flexing the fingers toward the origin of that frame coupled with a force control loop suffices to form a stable grasp. Since the motor action is bound to the local context, the same grasping action can be applied to different objects—a spatula, a mug, a doorknob—by changing the preshape.

The main features of the teleassistance strategy are that it can succinctly accommodate a range of natural variations in the task [Pook, 1995], but more importantly, it only requires 22% of the total time for executive control (indicated by the extent of the dark shaded areas in Figure 6). Thus the pointers required to implement the state machine of Figure 7 are only required for a small amount of time to initiate the lower-level primitives. The deictic signs may also be thought of as revealing how cognitive variables control human actions.

2.4 Deictic Strategies and the Identification/Location Dichotomy

We now discuss the referent of a visual pointer. A feature of visual cortex has been the separation of the primary feedforward pathways into dorsal and ventral streams. Initially these have been identified as separate processing streams, the "what" and "where" pathways of [Ungerleider and Mishkin, 1982]. More recently, Goodale and Milner have argued that a more appropriate division of labor might be in terms of identification of allocentric properties of an object, and the determination of its egocentric location in a scene [Goodale and Milner, 1992]. Furthermore they suggest that both these functions may involve both dorsal and lateral streams. The point is that the *identification/location* dichotomy is more of a functional than an architectural separation.

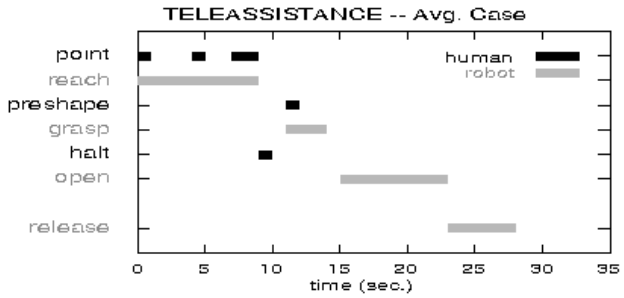


Figure 6: Results of simulating different forms of control in the door-opening task. Bars show the time spent in each subtask. The teleassistance model shows the interaction between deictic commands that signal the low-level routines and the routines themselves.

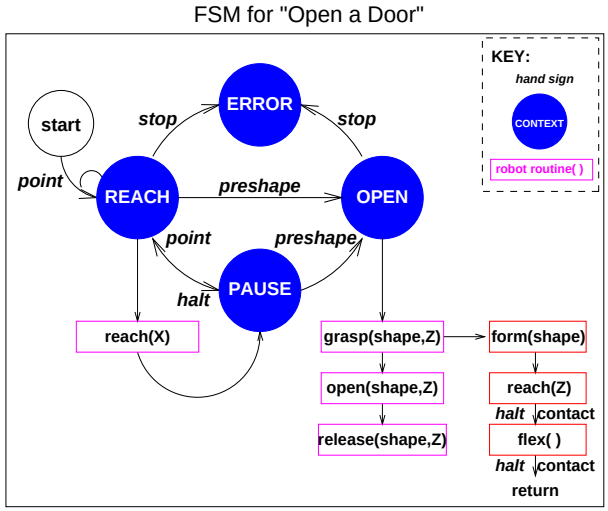


Figure 7: Simple state transition used for the interpretation of deictic signs. The deictic signs defined in Table 5 map directly onto transitions between states (shaded) of a simple control program.

Image Parts	Models	
	One	**Many**
One	**I. Deictic Access**: using a pointer to an object whose identity and location are known	**II. Identification**: trying to identify the object of a pointer referent
Many	**III. Location**: assigning a pointer a location	Too difficult?

Table 6: The organization of visual computation into *identification/location* modules may have a basis in complexity. Trying to match many image segments to many models simultaneously may be too difficult. The complexity of visual computation can however be substantially reduced by decomposing a given task into simpler deictic operations.

Implementation of deictic location and identification strategies lends support to the functional view. In computational terms, the suggestion is that the general problem of associating many internal models to many parts of an image simultaneously is too difficult for the brain to solve. Deictic strategies may make it computationally tractable by simplifying the general task into simpler identification and location tasks. These tasks either find information about location (using only one internal model) or identification (of only one world object whose location is known). Table 6 summarizes this view. A location task is greatly simplified by only having to find the image coordinates of a single model. In this task the image periphery must be searched; one can assume that the model has been chosen a priori. An identification task is greatly simplified by having to identify only the foveated part of the image. In this task one can assume that the location of the material to be established is at the fixation point; only the internal model data base must be searched.

Experimental tests of *identification/location* primitives on real image data confirm that this dichotomy leads to dramatically faster algorithms for each of the specialized tasks [Swain and Ballard, 1991, Rao and Ballard, 1995a]. Thus we can think of eye movements as solving a succession of location and identification subtasks in the process of meeting some larger cognitive goal. Section 3 shows that human performance of a sensory-motor task appears to be broken down into just such a sequence of primitive *identification/location* operations.

The concept of pointers changes the conceptual focus of computation from continuous processing to discrete processing. Such processing is centered around the momentary disposition of pointers that point to fragments of the sensory input such as the location or allocentric features of an object. Some actions change the location of a pointer and others compute properties or initiate movements

with respect to pointer locations. Thus we can interpret the *identification/location* dichotomy in Table 6 in terms of pointer operations. Identification can be interpreted as computing the perceptual properties of an active pointer referent at a known location. Location can be interpreted as computing the current location of an object with known properties and assigning a pointer to the computed location. This taxonomy emphasizes the functional properties of the computation as proposed by [Milner and Goodale, 1995].

3 Evidence for Deictic Strategies in Behavior

We began by positing the role of deictic actions as binding variables in deictic programs. Next we introduced markers as a general term to describe both variables in spatial working memory and current deictic variables for acquisition of visual information and initiation of motor routines. We now go on to examine whether in fact this conceptualization is appropriate for human behavior. Because the eyes allow a natural implementation of deictic strategies, the question immediately raised is how humans in fact use their eye movements in the context of natural behaviors. We designed a series of experiments to test the use of deictic strategies in the course of a simple task involving movements of the eyes and hand, and also visual memory. The task was to copy a pattern of colored blocks. It was chosen to reflect the basic sensory and motor operations involved in a wide range of human performance, involving a series of steps that require coordination of eye and hand movements and visual memory. An important feature of this task is that subjects have the freedom to choose their own parameters: the subjects organize the steps to compose the behavior as in any natural behavior. Another advantage is that the underlying cognitive operations are quite clearly defined by the implicit physical constraints of the task, as will become evident in what follows. This is important because definitive statements about the role of fixation in cognition are impossible when the internal cognitive state is undefined [Viviani, 1990].

3.1 Serialized Representations

The block copying task is shown in Figure 8. A display of colored blocks was divided up into three areas, the *model*, *resource*, and *workspace*. The model area contains the block configuration to be copied; the resource contains the blocks to be used; and the workspace is the area where the copy is assembled. Note that the colored blocks are random and difficult to group into larger shapes so they have to be handled individually. This allows the separation of perceptual and motor components of the task.[10] Subjects copied the block pattern as described above, and were asked

[10]The task has been studied using both real blocks and simulated blocks displayed on a Macintosh monitor. In this case the blocks were moved with a cursor. For the real blocks eye and head movements were monitored using an ASL head-mounted eye tracker that provides gaze position with an accuracy of about 1° over most of the visual field. The blocks region subtended about 30° of visual angle. In the Macintosh version of the task the display was

only to perform the task as quickly as possible. No other instructions were given, so as not to bias subjects towards particular strategies. A more detailed description of the experiments is given in [Ballard et al., 1995] and [Pelz, 1995].

A striking feature of task performance is that subjects behaved in a very similar, stereotypical way, characterized by frequent eye movements to the model pattern. Observations of individual eye movements suggest that information is acquired incrementally during the task and even modest demands on visual memory are avoided. For example, if the subject memorized and copied four sub-patterns of two blocks, which is well within visual memory limitations, one would expect a total of four looks into the model area. Instead, subjects sometimes made as many as 18 fixations in the model area in the course of copying the pattern, and did not appear to memorize more than the immediately relevant information from the model. Indeed, they commonly made more than one fixation in the model area while copying a single block. Thus subjects choose to serialize the task by adding many more eye fixations than might be expected. These fixations allow subjects to postpone the gathering of task-relevant information until just before it is required.

Figure 8 shows an example of the eye and hand (mouse) movements involved in moving a single block by one of the subjects. Following placement of the second block, the eye moves up to the model area, while at the same time the hand moves toward the blocks in the resource. During the fixation in the model area the subject presumably is acquiring the color of the next block. Following a visual search operation, a saccade is then programmed and the eye moves to the resource at the location of block three (green) and is used to guide the hand for a pickup action. The eye then *goes back* to the model while the cursor is moved to the workspace for putting down the block. This second fixation in the model area is presumably for the purpose of acquiring positional information for block placement. The eye then moves to the drop-off location to facilitate the release of the block.

The basic cycle from the point just after a block is dropped off to the point where the next block is dropped off allows us to explore the different sequences of primitive movements made in putting the blocks into place. A way of coding these subtasks is to summarize the eye fixations. Thus the sequence in Figure 8 can be encoded as "model-pickup-model-drop" with the understanding that the pickup occurs in the resource area and the drop in the workspace area. Four principal sequences of eye movements can be identified, as shown in Figure 9a. Since the decisive information is the color and relative location of each block, the observed sequences can be understood in terms of whether the subject has remembered either the color and/or the location of the block currently needed. The necessary assumption is that the information is most conveniently obtained by explicitly fixating the appropriate locations in the model and that the main preference is to acquire color or location information just before it is required. If both the color and location are needed—that is, have not been previously remembered—a "model-pickup-model-drop" sequence should result. If the color is

about 15° and eye position was recorded using a Dual Purkinje Image tracker that provides horizontal and vertical eye position signals with an accuracy of 10-15 min arc over about 15° range.

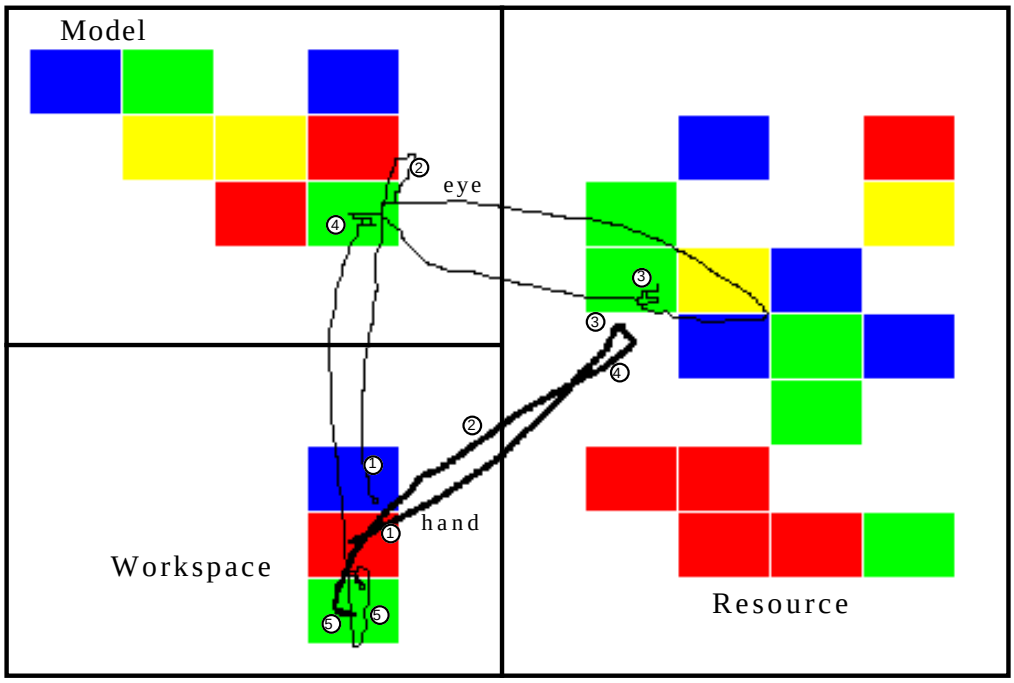


Figure 8: Copying a single block within the task. The eye position trace is shown by the cross and the dotted line. The cursor trace is shown by the arrow and the dark line. The numbers indicate corresponding points in time for the eye and hand traces.

Strategy	Time (Sec)	Memory Items
MPMD	3	
PMD	2.5	color
MPD	2.0	offset
PD	1.5	color and offset

Table 7: Speed vs. memory tradeoffs observed in the block-copying task.

known, a "pickup-model-drop" sequence should result; if the location is known, a "model-pickup-drop" sequence should result. If both are known, a "pickup-drop" sequence should result. In the data, "pickup-drop" sequences were invariably the last one or sometimes two blocks in the sequence. Thus with respect to color and location, the "model-pickup-model-drop" sequences are memoryless, and "model-pickup-drop," "pickup-model-drop," and "pickup-drop" sequences can be explained if the subjects are sometimes able to remember an extra location and/or color when they fixate the model area.

Summary data for seven subjects is shown as the dark bars in Figure 9b. The lowest-memory "model-pickup-model-drop" strategy is the most frequently used by all the subjects, far outweighing the others. (The figure shows data collected using the Macintosh. The same pattern of strategies is also reliably observed with real blocks and hand movements, with as many as 20 subjects [Pelz, 1995].) Note that if subjects were able to complete the task from memory, then a sequence composed exclusively of "pickup-drops" could have been observed, but instead the "pickup-drop" strategy is usually used only near the end of the construction. The frequent access to the model area during the construction of the copy we take as evidence of incremental access to information in the world in the process of performing the task. As the task progresses, the pointer referents, color and location in this case, are reset as the new information is called for.

3.2 Minimal Memory Strategies

The time required for each strategy when the target is in view is revealing. The time tallies are shown in Table 7, along with the putative memory load for color and location for each strategy. What is seen is that the lower memory strategies take longer. This is not too surprising, as the number of fixations goes down if items can be memorized. However, it is surprising that subjects choose minimal memory strategies in view of their temporal cost, particularly as they have been instructed to complete the task as quickly as possible, and memorization saves time.

The reluctance to use working memory to capacity can be explained if such memory is expensive to use with respect to the cost of the serializing strategy. Our experiments suggest that working memory is expensive compared to the cost of acquiring the information on-line, for the technical

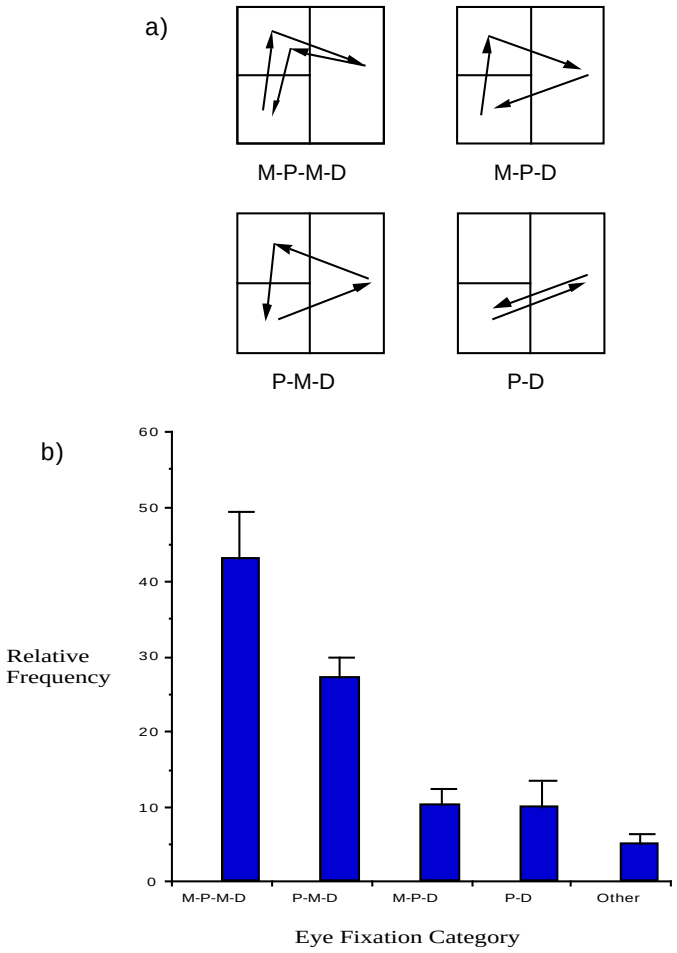


Figure 9: (a) The codes. "M" means that the eyes are directed to the model; "P" and "D" mean that the eyes and cursor are coincident at the pickup point and drop-off point, respectively. Thus for the PMD strategy, the eye goes directly to the resource for pickup, then to the model area, and then to the workspace for drop-off. (b) The relative frequency of the different strategies for seven subjects.

reasons discussed in Section 2, so that low memory strategies are preferred. This hypothesis would predict that, if the cost of the on-line acquisition of information could be increased relative to that of memorizing, the balance of effort should shift in the direction of increased memorization. To test this, the cost of on-line acquisition was increased by moving the model and copy from their previous position underneath one another to eccentric positions separated by 70°. Under these conditions subjects use memory more, as reflected in fewer eye movements to the model area. The number of eye movements decreases from an average of 1.3 per block to 1.0 per block. Thus eye movements, head movements, and memory load trade off against each other in a flexible way.

The analysis makes the assumption that the individual blocks are primitives for the task. This implies that the eye movements back to the model primarily serve to obtain properties of individual blocks. An alternate explanation is that the extra movements to the model area are in fact not essential but instead appear because of some other effect that is not being modeled. One such explanation is that the eyes move faster than the hand so that there will be extra time to check the model in a way that is unrelated to the properties of individual blocks. Another is that working memory is cheap but unreliable, so that subjects are checking to compensate for memory errors. However, a control experiment argues that both of these alternate hypotheses are unlikely. In the control, conditions were identical to the standard experiment with the exception that all the blocks were one color. This allows the subject to chunk segments of the pattern. There was a dramatic decrease in the number of eye movements used to inspect the model area: 0.7 per block in the monochrome case versus 1.3 per block in the multicolored case. (The control of separating the model and workspace also argues against unreliable memory. The increased time of transit would argue for more fixations given unreliable memory, but in fact fewer fixations were observed.) A closer inspection of the individual trials in the monochrome case reveals that subjects copy subpatterns without reference to the model, suggesting that they are able to recognize these composite shapes. In this case, subjects do abandon eye movements to the model. Thus we conclude that the movements to the model in the standard case are necessary and related to the individual properties of blocks.

3.3 The Role of Fixation

Performance in the blocks task provides plausible evidence that subjects use fixation as a deictic pointing device to serialize the task and allow incremental access to the immediately task-relevant information. However, it is important to attempt a more direct verification. A way to do this is to explore exactly what visual information is retained in visual memory from prior fixations by changing various aspects of the display during task performance. Changing information that is critical for the task should disrupt performance in some way. In one experiment we changed the color of one of the uncopied blocks while the subject was making a saccade to the model area following a block placement as shown in Figure 10. The changes were made when the subject's eyes crossed the boundary between the workspace and model area. Fixations in the workspace area are

almost invariably for the purpose of guiding the placement of a block in the partially completed copy. If the subject follows this placement with a saccade to the model area, the implication is that the subject currently has no color information or location information in memory and is fixating the model to acquire this information. It is not clear on what basis saccades to the model area are programmed, although they tend to be close to the previous fixation. In the first condition, illustrated in Figure 10a, the color of a block in the model was changed during the saccade to the model following block placement in the workspace, when the subject was beginning to work on the next block.[11] This is indicated by the zig-zag. The small arrow indicates the color change in the block. In another condition, shown in Figure 10b, the change was made after the subject had picked up a block and was returning to the model, presumably to check its location. In both conditions the changed block was chosen randomly from among the unworked blocks. A change occurred on about 0.25 of the fixations in the model area, and patterns where changes occurred were interleaved with control patterns where there were no changes.

Data for three subjects are shown in Figure 11. We measured the total time the subject spent fixating in the model area under different conditions. On the right of the figure is the fixation duration for the control trials, where no color changes occurred. On the left is the average fixation duration on trials when the changed block was the one the subject was about to copy. The lower line corresponds to when the change was made at the start of a new block move, that is, on the saccade to the model area following placement of the previous block and preceding pickup of the current block. This is the point in the task where subjects are thought to be acquiring color information. The upper line shows data for trials when the change was made following a pickup in the resource area. At this point in the task we hypothesized that the subject was acquiring relative location information for guiding block placement.

In the fixations preceding pickup there is only a small (50 milliseconds) increase in fixation duration for changes preceding pickup, even when the changed block is the target of the saccade. It suggests that block color is not retained in visual memory from previous model fixations, even though the subject has made multiple fixations in the model area prior to the change. The target selection involved in programming the saccade into the model does not appear to involve the acquisition of color information at the targeted location, and this function occurs during the fixation in the model area. This implies that rather minimal information is retained from the immediately prior fixation, and is consistent with the suggestion that fixation is used for acquiring information just prior to its use.

[11] Eye position was monitored by the Dual Purkinje Image eye tracker. Saccades were detected and the display updated within the 17 millisecond limit set by the refresh rate of the monitor. Display updates were performed seamlessly through video look-up table changes. All events were timed with an accuracy of 1 millisecond. The saccades in this experiment typically lasted about 50 milliseconds, and changes almost always occurred before the end of the saccade. This was verified by measuring the display change with a photodetector and comparing this with the eye position signal from the tracker.

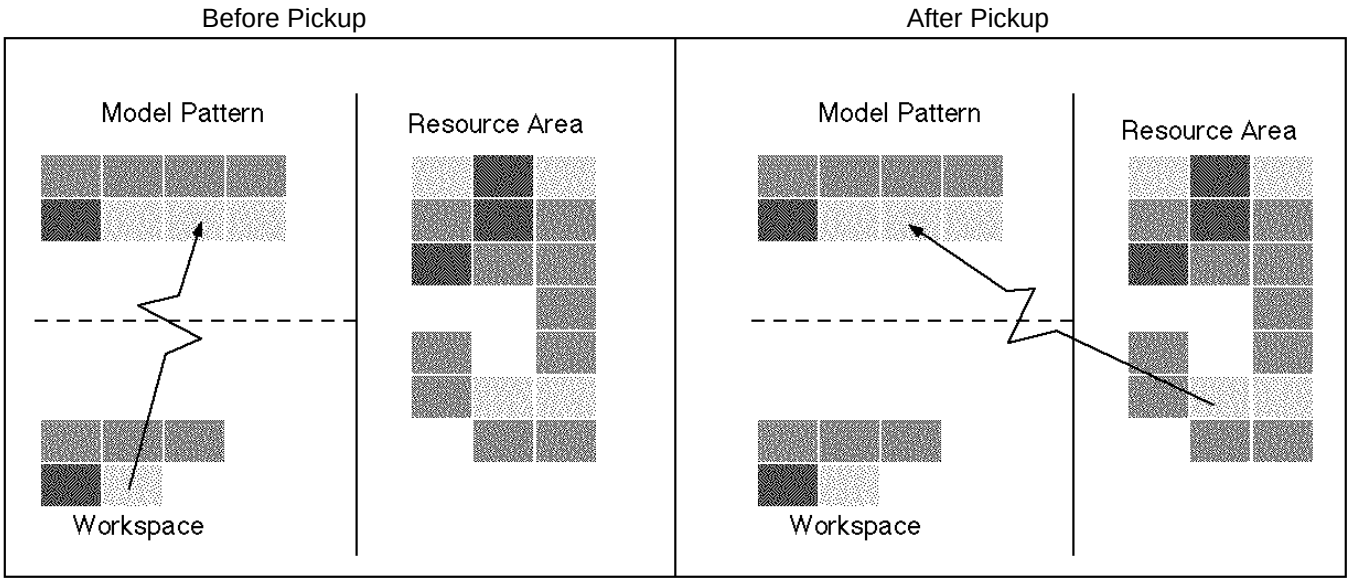


Figure 10: Two different experimental conditions for the color-changing experiment. In (a) the color of an uncopied model block is changed during a workspace-to-model saccade. In (b) the color of an uncopied model block is changed during a resource-to-model saccade.

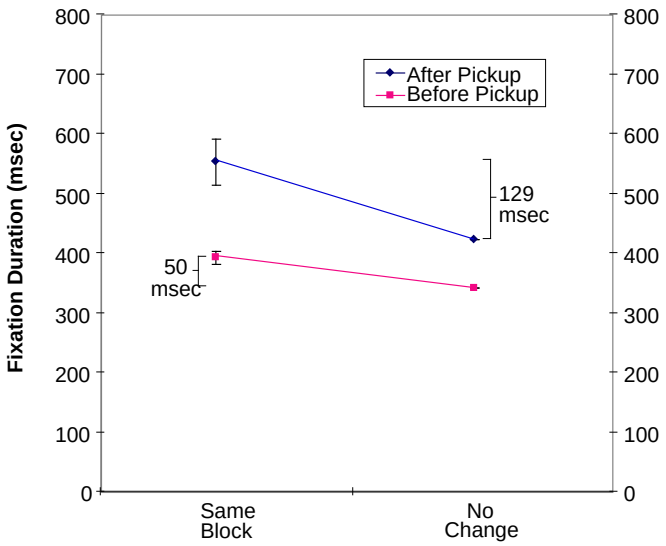


Figure 11: Results of the color-changing experiments, showing the fixation time when the target of the fixation was the next block moved compared to control trials where there was no change. For the condition in Figure 10a (lower points) there is a small change in fixation time. The condition in Figure 10b (upper points) shows a more dramatic increase in fixation time when the target block's color has changed.

In the fixations following the pickup there is a large (129 milliseconds) increase in fixation duration. Our interpretation is that the color information has been retained since it is now task-relevant and that the additional cost reflects changes that need to be made to the control program. This also validates fixation duration as a sensitive measure of the ongoing computations. Despite the increased fixation duration, in most instances subjects were not aware of the change, but fixated a neighboring block of the color of the block they were holding and placed it in the appropriate new location. The longer time spent in the model area in Figure 11b partly reflects this additional fixation, and partly reflects that individual fixations are longer (see [Bensinger et al., 1995]). The important finding is that the information retained from prior saccades is determined by what is currently relevant for the task.

3.4 Implications

It appears that human performance in tasks like these can be well characterized as a sequence of deictic instructions based on a small number of primitive operations. This obviates the need for complex memory representations. These results support the computational interpretation of the limitations of human working memory. Rather than being thought of as a limitation on processing capacity, it can be seen as a necessary feature of a system that makes dynamic use of deictic

variables. The limited number of variables need only be a handicap if the entire task is to be completed from memory; in that case, the short-term memory system is overburdened. In the more natural case of performing the task with ongoing access to the visual world, the task is completed perfectly. This suggests that a natural metric for evaluating behavioral programs can be based on their spatio-temporal information requirements.

These results also support the role of foveating eye movements suggested in Section 2. Since Yarbus's classic observations [Yarbus, 1967], saccadic eye movements have often been thought to reflect cognitive events, in addition to being driven by the poor resolution of peripheral vision. However, making this link has proved sufficiently difficult so as to raise questions about how much can be learned about cognitive operations by inspecting the fixation patterns [Viviani, 1990]. As discussed above, one of the difficulties of relating fixations to cognitive processes is that fixation itself doesn't indicate what properties are being acquired. In the block-copying paradigm, however, fixation appears to be tightly linked to the underlying processes by marking the location at which information (e.g., color, relative location) is to be acquired, or the location that specifies the target of the hand movement (picking up, putting down). Thus fixation can be seen as binding the value of the variable currently relevant for the task. Our ability to relate fixations to cognitive processes in this instance is a consequence of our ability to provide an explicit description of the task. In previous attempts to glean insight from eye movements (e.g., viewing a scene or identifiying a pattern), the task demands are not well specified or observable.

We can now reexamine the computational hypothesis illustrated in Table 6 in the light of our observations of human performance in this task. We can think of task performance as being explained by the successive application of three operations of the kind illustrated there. Thus, a model fixation will acquire visual properties (color, relative location) at the location pointed to by fixation (cf. the *identification* box, in Table 6). This will be followed by a visual search operation to find the target color in the resource, or the putdown location in the workspace (the *location*), saccade programming to that location, and visual guidance of the hand to the fixated location (the *deictic access* box). To complete the task, we need, in addition, the operation of holding a very small number of properties of the model pattern in working memory, and programming the ballistic phase of the hand movement.

4 Deictic Strategies and Cerebral Organization

Deictic actions suggest the use of functional models that compute just what is needed for the current point in the task. This is illustrated by the blocks task very dramatically, particularly by the experiments that switch colors during saccades. These experiments suggest (1) that the brain seems to postpone binding color and relative location information until just before it is required, and (2) that the information bound to a pointer during a fixation is just the useful portion (e.g., a color or relative location) of that available.

An obvious reason for timely task-dependent computation is that its products are so varied that the brain cannot pre-compute them. Consider all the information that one might have to know about an image. An economical way of computing this, perhaps the only way, is by tailoring the computation to just that required by the task demands as they become known. In the blocks task, at one point in time subjects need a block of an appropriate color, and at another point they need to know where to put that block. At both of these times they are fixating the model area in the same place. The key difference is that they need to apply a different computation to the same image data. During the first fixation subjects need to extract the color of the next block; during the second fixation they need the relative offset of the next block with respect to the model pattern.

The hypothesis that vision must be functional depends on some mechanism of spanning the space of task-dependent representations. One way of doing this is to compose primitive routines. Thus at the level of abstraction below the embodiment level one can think of a set of more primitive instructions that implement the binding required by the embodiment level. In this section we describe how these primitives might work and how their functionality might map onto brain anatomy.

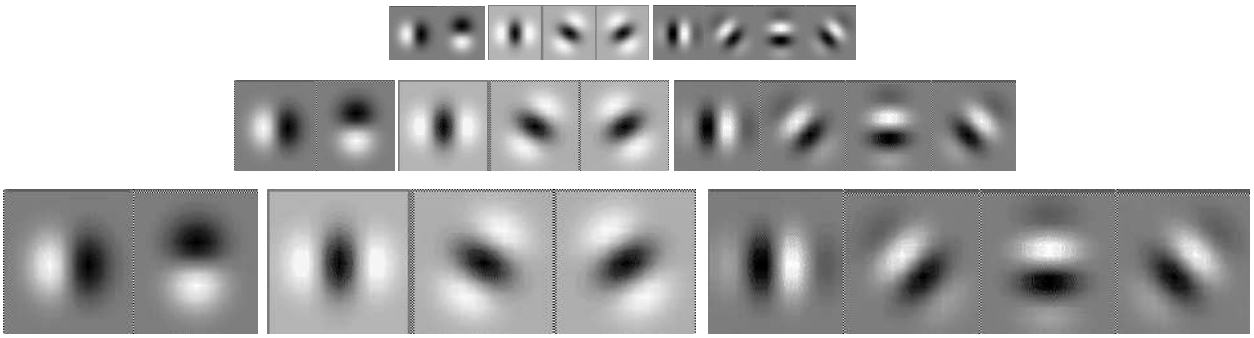
4.1 Functional Routines

Although the functional primitives could exist for any modality, let us concentrate on vision. Visual routines were first suggested by [Kosslyn, 1994] and [Just and Carpenter, 1976], but [Ullman, 1984] developed the essential arguments for them. Ullman's visual routines had a graphics flavor; in contrast, our main motivation is to show how visual routines can support the two principal requirements of deictic computation. These are simply the *identification* and *location* subtasks described in Section 2.4. Thus the two important visual routines are: (1) the ability to extract the properties of pointer locations (identification); and (2) the ability to point to aspects of the physical environment (location).

The task of specifying visual routines would seem to pose a conundrum as the principal advantage of task-dependent routines is to be able to minimize representation, yet there must be *some* representation to get started. The base representation that we and others have proposed [Rao and Ballard, 1995a, Rao et al., 1996, Jones and Malik, 1992, Wiskott and von der Malsburg, 1993] is a high-dimensional feature vector. This vector is composed of a set of basis functions that span features such as spatial frequency and color as well as scale. For the demonstrations used here the steerable filters are used [Freeman and Adelson, 1991] but very similar filters that have properties commensurate with those observed in the primate cortex can be learned from natural stimuli [Rao and Ballard, 1995b]. The filters are shown in Figure 12. They consist of first, second and third derivatives of a Gaussian intensity profile rotated in increments of 90, 60, and 45 degrees, respectively. These are used for each of three color channels (intensity, red-green opponent, and yellow-blue opponent) and at three different scales, for a total of eighty-one filter responses for each image point. The exact number and composition of the filters is unimpor-

R	G	B	
0.3333	0.3333	0.3333	R+G+B
0.5000	-0.5000	0.0000	R-G
-0.2500	-0.2500	0.5000	B-Y

(a)



(b)

Figure 12: The spatiochromatic basis templates that uniquely define points in the image even under some view variation. Shown are the filters used in the computer simulations. There are three color channels, three scales per channel, and nine steerable filters per scale, for a total of 81 filters for each point in the image.

tant for the algorithms, but the structure of the ones we use is motivated by cortical data. The advantage of high-dimensional feature vectors is that they are for all practical purposes unique [Kanerva, 1988, Rao and Ballard, 1995a], and thus each location in the sensorium can be given a unique descriptor.

4.2 Identification

Object identification matches a foveal set of image features with all possible model features. The result is the model coordinates of the best match. In the case of extracting the color of the material at the fixation point, the responses of the color components of the filters can be compared to model prototype colors. Figure 13 suggests how this can be done by showing actual filter responses for three points in the color display—two blocks and a background point. The scale is chosen so the largest filter diameter is approximately the size of a block. What the figure shows is that the three response vectors differ significantly, making the extraction of the color of a block an easy problem.

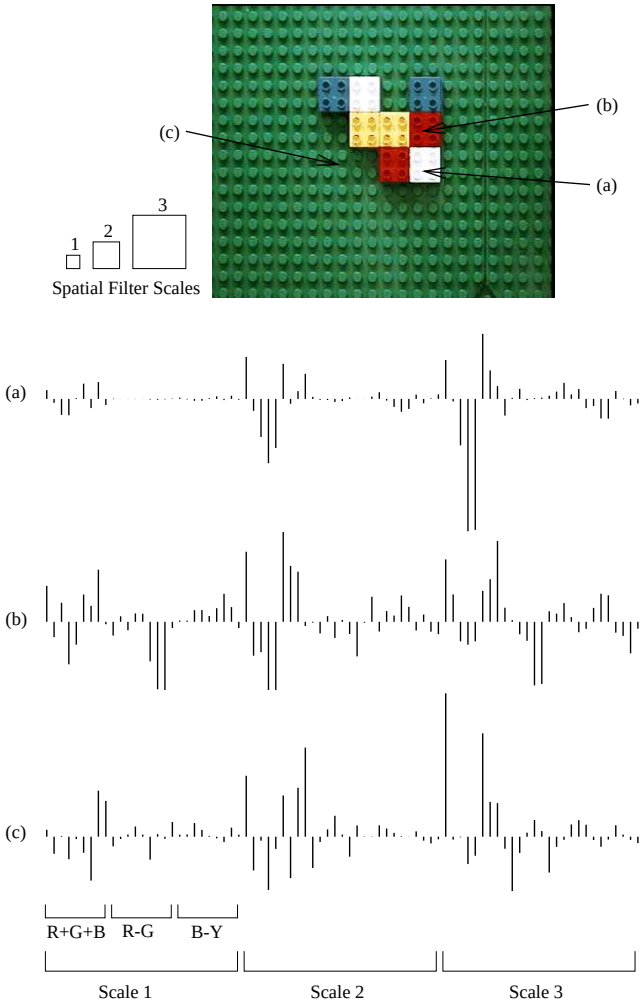


Figure 13: Using spatiochromatic filters to extract task-dependent properties. The blocks image used in the copying task is shown at the top. (a) The filter responses for a white block. Positive responses are represented as a bar above the horizontal, negative responses as a bar below the horizontal. The white point has many low responses due to the opponent channel coding. (b) The filter responses for a red block. (c) The filter responses for a point in the green background.

4.3 Location

Object location matches a localized set of model features with image features at all possible retinal locations. The result is the image coordinates of the best match. By fixating at a point, the responses of the basis templates can be recorded. The location problem we consider is to find a block of a particular color in the resource area (Figure 15). Alternately, one can consider the problem of returning gaze to a point after the gaze has gone elsewhere, when the features of the remembered point are accessible via working memory and the point is still in view. In both these cases, the location of the point relative to the current fixation can be determined by matching the remembered features with the features at all the current locations.[12]

The strategy of matching a remembered template to retinotopic template responses works well but needs additional structure to reflect the demands of the task. This additional structure also handles the case when the target is not in view. Basically the requirement is for a structure that remembers relative locations with respect to a given scene.

The filter matching algorithm determines the transformation that describes the relationship between an object-centered reference frame and the current view frame represented by the fixation point. However, it is easy to demonstrate the importance of a third frame. In reading, the position of letters with respect to the retina is unimportant compared to their position in the encompassing word. In driving, the position of the car with respect to the fixation point is unimportant compared to its position with respect to the edge of the road. In both of these examples the crucial information is contained in the transformation between an object-centered frame and a *scene* frame [Hinton, 1981]. Figure 14 shows these relationships for the image of the letter "A" depicted on a television display.

This additional structure allows for temporary storage of remembered locations as well as task-dependent constraints that direct the eyes to the appropriate points in the model, workspace and resource areas. The memory codes constraints as transformations T_{sr} and T_{os}, as shown in Figure 14. Given explicit memory for these two transformations, a location relative to the scene can be placed in space by concatenating the two transformations. This works when the result is on the retina but also in the more general case when it may be hidden from view. Support for this strategy comes from simulations that show it can model a variety of observed data from patients with parietal lesions [Rao and Ballard, 1996].

As a specific example of how these frames are used, consider the problem of finding a yellow block. Figure 15 shows how this can be done. When fixating the model (a), the filter code for a

[12]Template matching can be vulnerable to lighting changes; it is vulnerable to transformations such as scale and rotation; and its storage requirements can be prohibitive. However, recent developments [Buhmann et al., 1990, Jones and Malik, 1992] have ameliorated these disadvantages. If the template is created dynamically and has a limited lifetime, then the first objection is less important, as lighting conditions are likely to be constant over the duration of the task. As for the second and third objections, special compact codes for templates can overcome these difficulties (for example, see [Rao and Ballard, 1995a]).

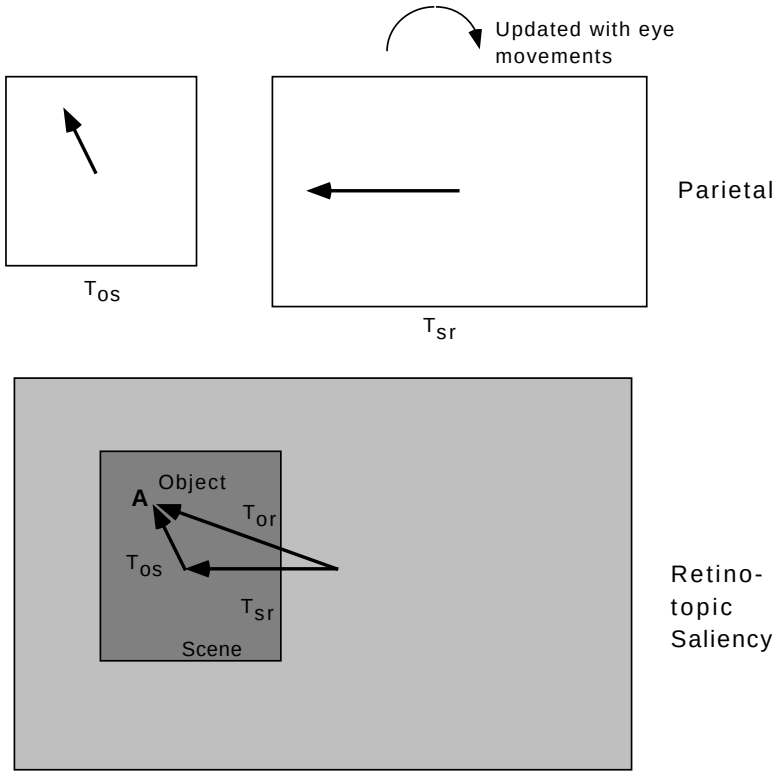


Figure 14: To represent the geometric relations of visual features three transformations are fundamental. The first describes how a particular depiction of the world, or scene, is related to the retinal coordinate system of the current fixation(T_{sr}). The second describes how objects can be related to the scene (T_{os}). The third, which is the composition of the other two, describes how objects are transformed with respect to the retina (T_{or}).

yellow block (b) is extracted and its features are stored in working memory. Later, at the moment the eyes are required to fixate a yellow block in the resource area, the remembered responses are compared to the current responses in the retinal coordinates. The best match defines a motor target in a "saliency map" (c) that can be fixated. However, the search for the yellow block must be constrained to the resource area. But where is the resource area? It is easy to define as a template with respect to the scene in T_{os}. Then T_{sr} allows this template to be appropriately positioned in retinal coordinates.[13] This is done in (c) in the figure. Thus there is a simple mechanism for including only the yellow blocks in the resource area as targets. The best match defines the target point, as shown in (d).

4.4 Visual Cortex and Pointer Referents

Let us now speculate specifically on how the visual routines that support pointer computation could be implemented in terms of the cortical architecture. In particular, consider the use of pointers to solve both *identification* and *location* problems. To implement the routines, feature vectors are stored in two separate memories, as shown in Figure 16. One memory is indexed by image coordinates, as depicted by the rectangle on the left-hand side of the figure. The other memory is indexed by object coordinates, as depicted by the rectangle on the right-hand side of the figure. This highly schematized figure suggests the overall control pathways. The neurons of the thalamus and visual cortex are summarized in terms of retinotopically-indexed banks of filtered representations (at multiple scales) at each retinotopic location, as shown on the left-hand side of the figure. The neurons of the infero-temporal cortex are summarized on the right-hand side of the figure as model-indexed banks of filtered object representations.

Consider the identification problem: the task context requires that properties at the fovea be compared to remembered properties. This could be done in principle by matching the features of the foveated location with features currently pointed to by working memory. To do so requires the operations depicted in the upper part of Figure 16. Now consider the converse problem: the task context requires that gaze be directed to a scene location with a remembered set of features. This could be done in principle by matching the remembered set of features with features currently on the retina. To do so requires the operations depicted in the lower part of Figure 16. The remembered set of features can be communicated for matching with their counterparts on the retinotopically-indexed cortex via feedback connections. The results of matching would nominally reside in the parietal cortex in the form of saliency maps denoting task-relevant spatial locations.

[13] In [Rao and Ballard, 1996], it is argued that T_{sr} is updated with respect to self (eye/head/body) movements as well as task-relevant scene movements.

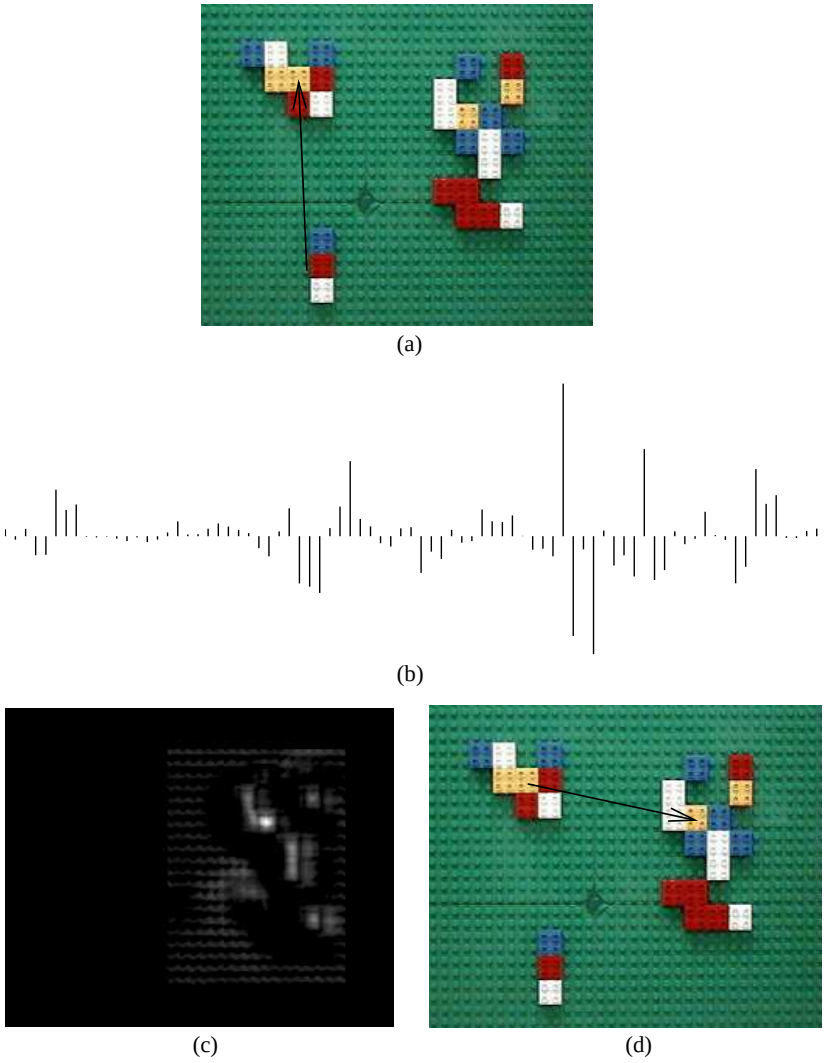


Figure 15: Using spatiochromatic filter representations for programming eye movements in the blocks task. (a) Fixating a yellow block in the model causes its filter responses (b) to be placed in working memory. (c) Those responses are then matched to filter responses at all locations in the resource area. The result is a "saliency map" that specifies possible yellow block locations. Saliency is coded by intensity; brighter points represent more salient locations. (c) An eye movement can then be generated to the most salient point in order to pick up the yellow block in the resource area.

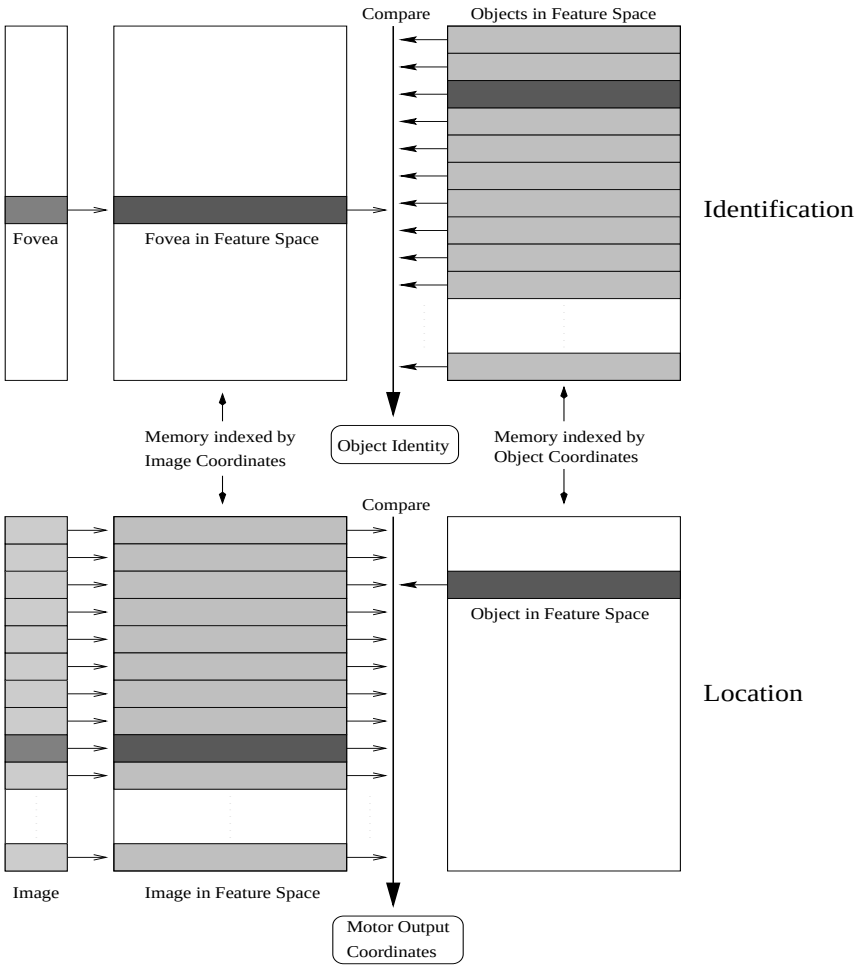


Figure 16: A highly schematized depiction of the identification/location control architecture (from [Rao and Ballard, 1995a]). (Upper) To identify an object, the features near the fovea are matched against a data base of iconic models, also encoded in terms of features. The result is decision as to the object's identity. (Lower) To locate an object, its features are matched against retinal features at similar locations. The result is the location of the object in retinal coordinates. This may be augmented by a depth term obtained from binocular vergence.

4.5 Basal Ganglia and Pointer Manipulation

Although the brain's various subsystems are far from completely understood, enough information about them is available to at least attempt to piece together a picture of how the functionality required by pointers might be implemented. Section 2 developed the idea of minimal representations with respect to the minimal *number* of pointers to describe how pointers factor into computation at the embodiment level, but one should not think that the information represented by a particular pointer is small. This concerns the *contents* of a pointer or the information that is the pointer referent. The content of a pointer may be iconic, and of considerable size. Now to a crude first approximation, let us think of the cortex as a kind of content-addressable memory. As such the cortex can be modeled as holding the contents of a pointer. However, if the job of the cortex is to hold the contents of pointers, additional *extracortical* neural circuitry is required to realize the different functionality, such as visual routines, that manipulate these contents. For example, all of the following needs to be done:

1. The sequencing in the task to determine what sensory processing should be done;
2. The processing to extract task-dependent representations;
3. The binding of these results to deictic pointers in working memory.

Thus although the logical place for most of the detailed processing such as filter template matching is in the retinotopically-indexed areas of cortex, other areas are needed to implement the control structure associated with a pointer-manipulation task. In order to develop this further, let us briefly review some important ideas about a key subsystem.

The *basal ganglia* is an extensive sub-cortical nucleus implicated in the learning of program sequences [Houk et al., 1995]. Independently, [Strick et al., 1995] and [Miyachi et al., 1994] have shown that neurons in the basal ganglia that respond to task-specific sub-sequences emerge in the course of training. Schultz has shown that basal ganglia neurons learn to predict reward. When a monkey initially reaches into an enclosed box for an apple, these neurons respond when the fingers touch the apple. If a light is paired with the apple reward in advance of reaching into the box, the same neurons develop responses to the light and not to the actual touching. These neurons are dopaminergic; that is, they are responsible for one of the main chemical reward systems used by the brain [Schultz et al., 1995]. Thus the implication is that the monkey is learning to predict delayed reward and coding it via an internal dopamine messenger. The basal ganglia has extensive connections to cortex that emphasize frontal and motor cortex [Graybiel and Kimura, 1995].

The point is that the functionality that supports different aspects of a pointer-related task requires the coordinated activity of different places in the brain, but broadly speaking, the crucial information about program sequence is represented in the basal ganglia. In recent studies of Parkinson's patients, a disease associated with damage to the basal ganglia, such patients performing a

task very like the blocks task have revealed deficits in working memory [Gabrieli, 1995]. Consider the interaction of vision and action. If vision is in fact task-dependent then it is very natural that a basal ganglia deficit produces a working memory deficit. The properties of cells in the caudal neostriatum are consistent with a role in short-term visual memory and may participate in a variety of cognitive operations in different contexts [Caan et al., 1984]. [Kimura et al., 1992] have also suggested that different groups of neurons in the putamen participate in retrieving functionally different kinds of information from either short-term memory or sensory data. The suggestion is that this circuitry is used in a functional manner with cortex producing the referent of basal ganglia pointers. The basal ganglia represents motor program sequencing information; the visual cortex represents potential locations and properties. This role is also supported by Yeterian and Pandya's comprehensive study of projections from the extrastriate areas of visual cortex, showing distinctive patterns of connectivity to caudate nucleus and putamen [Yeterian and Pandya, 1995].

Thus the disparate purposes of saccadic fixations to the same part of visual space in the blocks task can be resolved by the basal ganglia, which represents essential programmatic temporal context on "why" those fixations are taking place and "when" this information should be used. Another specific example makes the same point. Experiments in cortical area 46 have shown that there are memory representations for the next eye movement in motor coordinates [Goldman-Rakic, 1995]. However, this representation does not necessarily contain the information as to when this information is to be used; that kind of information is part of a motor program such as might be found in the basal ganglia. For this reason the anatomical connections of the basal ganglia should be very important, as they may have to influence the earliest visual processing.

The essential point of the above discussion is that the behavioral primitives at the embodiment level necessarily involve most of the cortical circuitry and that at the one-third second time scale one cannot think of parts of the brain in isolation. This point has also been extensively argued by Fuster [Fuster, 1989, Fuster, 1995]. Thus our view is similar to [Van Essen et al., 1994] in that they also see the need for attentional "dynamic routing" of information, but different in that we see the essential need for the basal ganglia to indicate program temporal context. Van Essen et al.'s circuitry is restricted to the pulvinar. [Kosslyn, 1994] has long advocated the use of visual routines, and recent PET studies have implicated the early visual areas of striate cortex. The routines here are compatible with Kosslyn's suggestions but make the implementation of visual routines more concrete. Other PET studies [Jonides et al., 1993, Paulesu et al., 1993] have looked for specific areas of the cortex that are active during tasks that use working memory. One interesting feature of the studies is that widely distributed cortical areas appear to be involved, depending on the particular task [Raichle, 1993]. This is consistent with the distributed scheme proposed here, where the cortex holds the referent of the pointers. This would mean that the particular areas that are activated depend in a very direct fashion on the particular task. It also raises the question of whether particular brain areas underlie the well established findings that working memory can be divided into a central executive and a small number of slave systems: the articulatory loop and visuo-spatial scratch pad [Baddeley, 1986, Logie, 1995]. It should be the case that working memory

can be differentiated in this way only to the extent that the tasks are differentiated, and to the extent visual and auditory function involve different cortical regions. Thus the kind of information held in working memory should reflect the kind of things it is used for.

5 Discussion and Conclusions

The focus of this paper has been an abstract model of computation that describes the interfacing of the body's apparatus to the brain's behavioral programs. Viewing the brain as hierarchically organized allows the differentiation of processes that occur at different spatial and temporal scales. It is important to do this because the nature of the computations at each level are different. Examination of computation at the embodiment level's one-third second time scale provides a crucial link between elemental perceptual events that occur on a shorter time scale of 50 msec, and events at the level of cognitive tasks that occur on a longer time scale of seconds. The importance of examining the embodiment level is that body movements have a natural computational interpretation in terms of deictic pointers, because of the ability of the different sensory modalities to quickly direct their foci to localized parts of space. As such, deictic computation provides a mechanism for representing the essential features that link external sensory data with internal cognitive programs and motor actions. Section 2 explored the computational advantages of sequential, deictic programs for behavior. The notion of a pointer was introduced by [Pylyshyn, 1989] as an abstraction for representing spatial locations independent of their features. Pylyshyn conceived the pointers as a product of bottom-up processing [Trick and Pylyshyn, 1996], and therein lies the crucial difference between those pointers and the deictic pointers used herein. Deictic pointers are required for variables in a cognitive "top-down" program. Section 3 presented evidence that humans indeed use fixation in a way that is consistent with this computational strategy. When performing natural tasks subjects make moment-by-moment tradeoffs between the visual information maintained in working memory and that acquired by eye fixations. This serialization of the task with frequent eye movements is consistent with the interpretation of working memory whereby fixation is used as a deictic pointing device to dynamically bind items in working memory. Section 4 examined how component low-level routines (that is, at the level of perceptual acts) might effect the referent of pointers. This formulation owes much to [Milner and Goodale, 1995] and provides a very concrete model of how their psychophysical data could arise from neural circuitry.

Deictic codes provide compact descriptions of the sensory space that have many advantages for cognitive programs:

1. *The facilitation of spatio-temporal reference.* Sequential computations in cognitive tasks make extensive use of the body's ability to orient. The chief example of this is in the ability of the eyes to fixate on a target in three-dimensional space, which in turn leads to simplified manipulation strategies.

2. *The use of "just-in-time" representation.* Deictic representations allow the brain to leave important information out in the world and acquire it just before it is needed in the cognitive program. This avoids the carrying cost of the information.

3. *The simplification of cognitive programs.* A way of understanding programs is in terms of the number of variables needed to describe the computation at any instant. Deictic pointers provide a way of understanding this cost accounting. Identifying working memory items with pointers suggests that temporary memory should be minimized. It simplifies the credit assignment problem in cognitive programs as described in Section 2 [Whitehead and Ballard, 1991, McCallum, 1994, Pook and Ballard, 1994a].

4. *The simplification of sensory-motor routines.* The use of deictic codes leads to functional models of vision wherein the representational products are only computed if they are vital to the current cognitive program. It is always a good idea to give the brain less to do, and functional models show that much of the products of the sensorium that we might have thought were necessary can be done without.

Deictic codes can lead to different ways of thinking about traditional approaches to perception and cognition. At the same time the models described herein are formative and need considerable development. The ensuing discussion tackles some of the principal issues that arise from this view.

The Generality of the Blocks Task

It might be thought that the main line of evidence for deictic codes comes from the blocks task and that the serial nature of that task, as well as its specificity, is sufficiently constrained so that the results are an inevitable consequence rather than a more general principle. The blocks task represents an approach to studying natural tasks where data are taken in a natural setting over many different applications of eye movements and hand movements. This approach to evaluating the use of eye movements has also been used in the study of recognition [Noton and Stark, 1971a] and chess [Chase and Simon, 1973], and even though the underlying task was not as constrained in those settings, the overall pattern of sensory-motor coordination would suggest that deictic strategies are used in those cases also. The eye movements in chess have been observed to fixate pieces that are important to the current situation. Simon and Chase suggested that the purpose of these might be to obtain patterns that would be used to access chess moves that were stored in a tree, even though they could not say what the patterns were or comment on the exact role of individual eye movements. Nonetheless one can say that it is very plausible that here too eye movements are used to extract a given pattern stored as the contents of a pointer and that the contents of several pointers are temporarily stored in working memory. Studies of eye movements during driving reveal very specific fixations to targets that are germane to the driving task [Land, 1994]. That is, the fixation patterns have predictive value for the driver's next action.

The Role of Working Memory and Attention

Deictic codes lead us to a different way of thinking about working memory and attention. Traditionally, cognitive operations have been thought of as being fundamentally constrained by some kind of *capacity* limits on processing (see, e.g., [Just and Carpenter, 1992, Norman and Bobrow, 1975, Logie, 1995]). Similarly, attention has been viewed as some kind of limited mental resource that constrains cognitive processing. However, as [Allport, 1989] has pointed out, viewing attention as a limited resource may be little more than a redescription of the phenomena and does not explain why the limitations exist. Instead, viewing attention as a pointer gives its selective nature a computational rationale. In considering attentional limitations, or selectivity, Allport argues that some kind of selectivity is essential for coordinated perceptuo-motor action. (Thus an eye movement requires some kind of visual search operation to select a saccade target.) This is very compatible with the ideas developed here, in which attention is a pointer to parts of the sensorium that is manipulated by current task goals. This view explains the paradox that "pre-attentive" visual search apparently operates on information which has undergone some kind of segmentation, thought to require attention. This makes sense if we think of selective attention as being the extraction of the information relevant for the current task, and this may be a low-level feature or a high-level semantically-defined target. Evidence by [Johnston and Dark, 1986] that selective attention can be guided by active schemata (of any kind) is consistent with this.

A similar shift in viewpoint can be obtained for working memory. The structure of working memory has been considered largely from the perspective of the contents of the memory [Baddeley, 1986, Logie, 1995]. The experiments herein shift the focus to the ongoing *process* of cycling information through working memory. From this perspective, the capacity limits in working memory can be seen not as a constraint on processing, but as an inevitable consequence of a system that uses deictic variables to preserve only the products of the brain's computations that are necessary for the ongoing task. On this view, interference in dual task experiments will depend on the extent to which the different tasks compete for the particular brain circuitry required for task completion. Thus the important separation between the phonetic and visual components of working memory can be seen as a consequence of the way they are used in natural behaviors rather than an architectural feature. The conception of working memory as the set of currently active pointers also leads to a very simple interpretation of the tradeoffs between working memory load and eye movements, in which fixation can be seen as a choice of an external rather than an internal pointer.

Separate Perception and Cognition?

An interpretation of brain computations in terms of binding variables in behavioral programs blurs the distinction between perception and cognition, which have traditionally been thought of as different domains. It also challenges the idea that the visual system constructs a three-dimensional

model of the scene containing its parts as components with detailed location and shape information for each of the parts, and that the products of the perceptual computation are then delivered up to cognitive mechanisms for further processing. A description and critique of this view has been presented by [Churchland et al., 1994]. The idea of an elaborate scene model is perhaps clearest in the computer vision literature, where until recently the goal of the models has been primarily one of reconstruction [Marr, 1982, Brady, 1981]. The emergence of an alternative approach (called "active" or animate vision [Brooks, 1986, Brooks, 1991, Agre and Chapman, 1987, Ballard, 1991]) that takes advantage of observer actions to minimize representations forms the foundation for the ideas presented here.

There is still more ambiguity about the way perceptual representations in humans are conceived. On the one hand, a difference between processing of attended and unattended information is clearly acknowledged, and the limitations set by working memory are recognized as fundamental. On the other hand, it is often implicitly assumed that the function of perception is to construct a representation of the arrangement and identities of objects in a scene. We are inclined to think of perception as fundamentally parallel and cognition as fundamentally serial. However, the intimate relation between fixations and the serialized acquisition of information required for task completion presents a challenge for our understanding of the nature of perceptual experience. In the block-copying task described in Section 3, manipulations on a given block are largely independent of the information acquired in previous views. This suggests that it is unnecessary to construct an elaborate scene description to perform the task and that there is only minimal processing of unattended information. In addition, since color and location information appear to be acquired separately, it appears that even in the attended regions the perceptual representation may be quite minimal. Thus human vision may indeed reflect the computational economies allowed by deictic representations and may only create perceptual descriptions that are relevant to the current task. A similar suggestion has been made by [Nakayama, 1990] and by O'Regan [O'Regan and Lévy-Schoen, 1983, O'Regan, 1992]. O'Regan suggested that only minimal information about a scene is represented at any given time, and that the scene can be used as a kind of "external" memory, and this is indeed what our observers appeared to do.

The ability to reach for objects that are out of view is often cited as evidence that people do use complex three-dimensional models, but the embodiment structure of spatial working memory allows an alternate explanation. The key is to separate the ability to use three-dimensional information, which people obviously do, from the need to build elaborate temporary models, which is extremely difficult if not impossible. Figure 17 shows the deictic explanation of the reaching example. A subject places an object into spatial working memory at a certain point (A), perhaps by marking it with fixation. Then at a later time, while fixating a new object, the original object can be grasped using the mechanisms of Section 4.3 in conjunction with the reference frames in Figure 14 (See also [Hayhoe et al., 1992]). However, the crucial constraint from our point of view is that the object must have been inventoried into spatial working memory. To be represented, the object has to use a pointer from the spatial working memory budget. It is not so much that the brain cannot

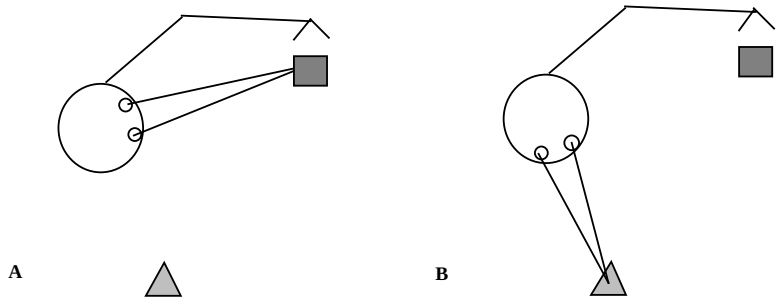


Figure 17: The fact that subjects can reach or look to targets outside their immediate field of view does not necessarily imply complete three-dimensional spatial representations (see discussion in the text).

compute in body-centered frames, but rather that the main focus is on the temporary nature of task-relevant frames.

Perceptual Integration?

The issue of the complexity of the visual representation is often confused with the issue of whether visual information is integrated across different eye positions. An internal scene representation is usually assumed to reflect information acquired from several fixations, so evidence that visual information can be integrated across eye movements has been seen as important evidence for this kind of view. However, the issues are really separate. Humans can clearly integrate visual information across eye movements when they are required to do so [Hayhoe et al., 1992, Hayhoe et al., 1991], and some ability to relate information across time and space seems necessary for coordinated action. It seems most natural to think of visual computations as extracting information in any of a range of reference frames as required by the task, independently of the particular eye position. At the same time, a number of studies reveal that the complexity of the representations that can be maintained across several eye positions is limited [Irwin, 1991, Irwin et al., 1990, Lachter and Hayhoe, 1996, Grimes and McConkie, 1995]. The saccade-updating experiment in Section 3 supports the idea that the information retained across saccades depends crucially on the task. Thus we would expect that the recent findings of [Duhamel et al., 1992], that neurons in macaque parietal cortex will fire in anticipation of a stimulus reaching their receptive fields at the end of a saccade, would be dependent on the nature of the task and the monkey's behavioral goals. One way of reconciling the fragmentary nature of the representations with the richness of our perceptual experience would be to suppose that our everyday experience of the world reflects events over a longer time scale than those revealed by fixations. This is suggested by the finding in Section 3 that task performance is affected by changes that are not perceptually salient.

The Ubiquitous Nature of Context

Our focus is a conception of the computations performed by the brain in disparate areas as devoted to the computation of currently relevant sensory-motor primitives. The functional model that has been proposed must be reconciled with the extensive neural circuitry of visual cortex, which has traditionally been investigated as a bottom-up, stimulus-driven system. The suggestion here is that even the activity of areas like V1 may be context-dependent. This dependence of neural responses on task context can be mediated by the extensive feedback projections known to exist in the visual cortex.[14] Thus both striate and extra-striate visual areas may in fact be computing highly specific task-relevant representations. Dramatic evidence for this kind of a view of cortex is provided by experiments by [Maunsell, 1993]. In recordings from parietal cortex of the monkey, he found that cells sensitive to motion would respond differently to identical visual stimuli depending on the experimental contingencies and whether or not the monkey anticipated that the stimulus would move. In a tracking experiment [Anstis and Ballard, 1995] found that a visual target such as a junction could not be pursued when "perceived" as two sliding bars, but could be pursued when "perceived" as a rigid intersection. Elegant techniques such as tachistoscopic presentations and backward masking have been used to isolate the feedforward pathway, and revealed much about its structure [Wandell, 1995], but the difficulty in doing this speaks to the more natural condition of using ongoing context (via top-down feedback).

The case that the brain uses external referents as implied by a deictic system is bolstered by observations that the neurons in visual cortex have logical zero measures such as zero disparity and zero optic flow. These relative measures are properties of the exocentric fixation point. Thus the very first visual measures are necessarily dependent on the location of the fixation point which in turn depends on the current cognitive goal. Andersen and Snyder [Stricanne et al., 1994, Snyder and Andersen, 1994] have also shown that parietal cortex contains neurons sensitive to exocentric location in space. Motor cortex recordings also show the use of exocentric or task-relevant frames [Tagaris et al., 1994, Pellizzer et al., 1994, Helms-Tillery et al., 1991].

In summary, we have presented a model at a level of abstraction that accounts for the body's pointing mechanisms. This is because traditional models that have been described at very short or very long time scales have proved insufficient to capture important features of behavior. The embodiment model utilizes large parts of the brain in a functional manner that at first might seem at odds with the seamlessness of cognition, but can be resolved if cognitive awareness and embodiment models operate at different levels of abstraction and therefore different time scales. One way to understand this is to use conventional computers as a metaphor. Consider the way virtual

[14] The location routine that mediates top-down visual search (as illustrated in Figure 15) crucially relies on the existence of feedback connections in the visual cortex as does the dynamic identification routine described in [Rao and Ballard, 1995b].

memory works on a conventional computer workstation. Virtual memory allows the applications programmer to write programs that are larger than the physical memory of the machine. Prior to the running of the program, it is broken up into smaller pages, and then at run time the requisite pages are brought into the memory from peripheral storage as required. This strategy works largely because conventional sequential programs are designed to be executed sequentially, and the information required to interpret an instruction is usually very localized. Consider now two very different viewpoints. From the application programmer's viewpoint, it appears that a program of unlimited length can be written that runs sequentially in a seamless manner. But the system programmer's viewpoint, wherein pages are rapidly shuffled in and out of memory, is very different, predominantly because it must explicitly account for the functionality at shorter time scales. It is just this comparison that captures the difference between the level of cognitive awareness and the level we are terming embodiment. Our conscious awareness simply may not have access to events on the shorter time scale of the embodiment level, where the human sensory-motor system employs many creative ways to interact with the world in a timely manner.

Acknowledgments

This research was generously supported by the National Institutes of Health under Grants 1-P41-RR09283-1 and EY05729, and by the National Science Foundation under Grants IRI-9406481 and CDA-9401142.

References

[Agre and Chapman, 1987] Agre, P. and Chapman, D. (1987). Pengi: An implementation of a theory of activity. In *Proc., AAAI-87*, pages 268–272, Seattle, WA.

[Allport, 1989] Allport, A. (1989). Visual attention. In Posner, M., editor, *Foundations of Cognitive Science*, pages 631–682. MIT Press, Cambridge, MA.

[Andersen, 1995] Andersen, R. (1995). Coordinate transformations and motor planning in posterior parietal cortex. In Gazzaniga, M., editor, *The Cognitive Neurosciences*, pages 519–532. MIT Press, Cambridge, MA.

[Anstis and Ballard, 1995] Anstis, S. and Ballard, D. (1995). Failure to pursue and perceive the motion of moving intersection and sliding rings. *Investigative Ophthalmology and Visual Science*, 36(4):S205.

[Arbib, 1981] Arbib, M. (1981). Perceptual structures and distributed motor control. In *Handbook of Physiology—The Nervous System II: Motor Control*, pages 1449–1480. American Physiological Society, Bethesda, MD.

[Arbib et al., 1985] Arbib, M., Iberall, T., and Lyons, D. (1985). Coordinated control programs for movements of the hand. In *Hand Function and the Neocortex*, pages 135–170. Springer-Verlag, Berlin.

[Baddeley, 1986] Baddeley, A. (1986). *Working Memory*. Oxford Clarendon Press.

[Baker and Gollub, 1990] Baker, G. and Gollub, J. (1990). *Chaotic Dynamics: An Introduction*. Cambridge University Press, New York, NY.

[Ballard, 1986] Ballard, D. (1986). Cortical connections and parallel processing: Structure and function. *Behavioral and Brain Sciences*, 9(1):67–120.

[Ballard, 1991] Ballard, D. (1991). Animate vision. *Artificial Intelligence Journal*, 48:57–86.

[Ballard et al., 1995] Ballard, D., Hayhoe, M., and Pelz, J. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7(1):66–80.

[Barlow, 1972] Barlow, H. (1972). Single units and cognition: A neurone doctrine for perceptual psychology. *Perception*, 1:371–394.

[Barto et al., 1990] Barto, A., Sutton, R., and Watkins, C. (1990). Sequential decision problems and neural networks. In Touretzky, D., editor, *Advances in Neural Information Processing Systems 2*. Morgan Kaufmann, San Mateo, CA.

[Bensinger et al., 1995] Bensinger, D., Hayhoe, M., and Ballard, D. (1995). Visual memory in a natural task. *Investigative Ophthalmology and Visual Science*, 36(4):S14.

[Brady, 1981] Brady, J. (1981). Preface—the changing shape of computer vision. *Artificial Intelligence*, 17(1–3):1–15.

[Broadbent, 1958] Broadbent, D. (1958). *Perception and Communication*. Oxford University Press, Oxford.

[Brooks, 1986] Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE J. Robotics and Automation*, 2:14–22.

[Brooks, 1991] Brooks, R. (1991). Intelligence without reason. Technical Report AI Memo 1293, Massachusetts Inst. of Technology.

[Buhmann et al., 1990] Buhmann, J. M., Lades, M., and von der Malsburg, C. (1990). Size and distortion invariant object recognition by hierarchical graph matching. In *Proc., IEEE IJCNN, (Vol. II)*, pages 411–416, San Diego, CA.

[Caan et al., 1984] Caan, W., Perrett, D., and Rolls, E. (1984). Responses of striatal neurons in the behaving monkey. 2. Visual processing in the caudal neostriatum. *Brain Research*, 290:53–65.

[Chapman, 1989] Chapman, D. (1989). Penguins can make cake. *AI Magazine*, 10(4):45–50.

[Chase and Simon, 1973] Chase, W. and Simon, H. (1973). Perception in chess. *Cognitive Psychology*, 4:55–81.

[Churchland et al., 1994] Churchland, P., Ramachandran, V., and Sejnowski, T. (1994). A critique of pure vision. In Koch, C. and Davis, J., editors, *Large-Scale Neuronal Theories of the Brain*, pages 23–60. MIT Press (A Bradford Book), Cambridge, MA.

[Crisman and Cleary, 1994] Crisman, J. and Cleary, M. (1994). Deictic primitives for general purpose navigation. *Proc., AIAA Conf. on Intelligent Robots in Factory, Field, Space, and Service (CIRFFSS)*.

[Duhamel et al., 1992] Duhamel, J., Colby, C., and Goldberg, M. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255:90–92.

[Freeman and Adelson, 1991] Freeman, W. and Adelson, E. (1991). The design and use of steerable filters. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 13(9):891–906.

[Fuster, 1989] Fuster, J. (1989). *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*. Raven Press, New York, second edition.

[Fuster, 1995] Fuster, J. (1995). *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*. MIT Press (A Bradford Book), Cambridge, MA.

[Gabrieli, 1995] Gabrieli, J. (1995). Contribution of the basal ganglia to skill learning and working memory in humans. In Houk, J., Davis, J., and Beiser, D., editors, *Models in Information Processing in the Basal Ganglia*, pages 277–294. MIT Press (A Bradford Book), Cambridge, MA.

[Goldberg, 1989] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.

[Goldman-Rakic, 1995] Goldman-Rakic, P. (1995). Toward a circuit model of working memory and the guidance of voluntary motor action. In Houk, J., Davis, J., and Beiser, D., editors, *Models in Information Processing in the Basal Ganglia*, pages 131–148. MIT Press (A Bradford Book), Cambridge, MA.

[Goodale and Milner, 1992] Goodale, M. and Milner, A. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15:20–25.

[Graybiel and Kimura, 1995] Graybiel, A. and Kimura, M. (1995). Adaptive neural networks in the basal ganglia. In Houk, J., Davis, J., and Beiser, D., editors, *Models in Information Processing in the Basal Ganglia*, pages 103–116. MIT Press (A Bradford Book), Cambridge, MA.

[Grimes and McConkie, 1995] Grimes, J. and McConkie, G. (to appear, 1995). On the insensitivity of the human visual system to image changes made during saccades. In Akins, K., editor, *Problems in Perception*. Oxford University Press, Oxford, UK.

[Hayhoe et al., 1991] Hayhoe, M., Lachter, J., and Feldman, J. (1991). Integration of form across saccadic eye movements. *Perception*, 20:393–402.

[Hayhoe et al., 1992] Hayhoe, M., Lachter, J., and Möller, P. (1992). Spatial memory and integration across saccadic eye movements. In Rayner, K., editor, *Eye Movements and Visual Cognition: Scene Perception and Reading*, pages 130–145. Springer-Verlag, New York.

[Helms-Tillery et al., 1991] Helms-Tillery, S., Flanders, M., and Soechting, J. (1991). A coordinate system for the synthesis of visual and kinesthetic information. *J. Neuroscience*, 11(3):770–778.

[Hertz et al., 1991] Hertz, J., Krogh, A., and Palmer, R. (1991). *Introduction to the Theory of Neural Computation*, volume 1 of *Santa Fe Institute Studies in the Sciences of Complexity Lecture Notes*. Addison-Wesley.

[Hinton, 1981] Hinton, G. (1981). A parallel computation that assigns canonical object-based frames of reference. In *International Joint Conference on Artificial Intelligence*.

[Houk et al., 1995] Houk, J., Davis, J., and Beiser, D., editors (1995). *Models in Information Processing in the Basal Ganglia*. MIT Press (A Bradford Book), Cambridge, MA.

[Irwin, 1991] Irwin, D. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, 23:420–456.

[Irwin et al., 1990] Irwin, D., Zacks, J., and Brown, J. (1990). Visual memory and the perception of a stable visual environment. *Perception and Psychophysics*, 47:35–46.

[Jeannerod, 1988] Jeannerod, M. (1988). *The Neural and Behavioural Organization of Goal-Directed Movements*. Clarendon Press, Oxford.

[Johnston and Dark, 1986] Johnston, W. and Dark, V. (1986). Selective attention. *Annual Review of Psychology*, 37:43–75.

[Jones and Malik, 1992] Jones, D. G. and Malik, J. (1992). A computational framework for determining stereo correspondence from a set of linear spatial filters. In *Proc., 2nd European Conf. on Computer Vision*.

[Jonides et al., 1993] Jonides, J., Smith, E., Koeppe, R., Awh, E., Minoshima, S., and Mintun, M. (1993). Spatial working memory in humans as revealed by PET. *Nature*, 363.

[Just and Carpenter, 1976] Just, M. and Carpenter, P. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8:441–480.

[Just and Carpenter, 1992] Just, M. and Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1):122–149.

[Kanerva, 1988] Kanerva, P. (1988). *Sparse Distributed Memory*. Bradford Books, Cambridge, MA.

[Kimura et al., 1992] Kimura, M., Aosaki, T., Hu, Y., Ishida, A., and Watanabe, K. (1992). Activity of primate putamen neurons is selective to the mode of voluntary movement: Visually-guided, self-initiated or memory-guided. *Experimental Brain Research*, 89:473–477.

[Koch and Crick, 1994] Koch, C. and Crick, F. (1994). Some further ideas regarding the neuronal basis of awareness. In Koch, C. and Davis, J., editors, *Large-Scale Neuronal Theories of the Brain*, pages 93–109. MIT Press (A Bradford Book), Cambridge, MA.

[Kosslyn, 1994] Kosslyn, S. (1994). *Image and Brain*. MIT Press (A Bradford Book), Cambridge, MA.

[Kowler and Anton, 1987] Kowler, E. and Anton, S. (1987). Reading twisted text: Implications for the role of saccades. *Vision Research*, 27:45–60.

[Koza, 1992] Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA.

[Lachter and Hayhoe, 1996] Lachter, J. and Hayhoe, M. (1996). Capacity limits in integrating information across saccades. *Perception and Psychophysics*. In press.

[Land, 1994] Land, M. (1994). *Nature*, 369.

[Logie, 1995] Logie, R. (1995). *Visuo-Spatial Working Memory*. Lawrence Erlbaum Associates, Publishers, Hillsdale, NY.

[Luria, 1968] Luria, A. (1968). *The Mind of a Mnemonist: A Little Book about a Vast Memory*. Harvard University Press, Cambridge, MA.

[Marr, 1982] Marr, D. (1982). *Vision*. W.H. Freeman and Co., Oxford.

[Maunsell, 1993] Maunsell, J. (1993). Presentation, Woods Hole Workshop on Computational Neuroscience.

[McCallum, 1995a] McCallum, A. K. (1995a). *Reinforcement learning with selective perception and hidden state*. PhD thesis, Department of Computer Science, U. Rochester.

[McCallum, 1994] McCallum, R. (1994). Reduced training time for reinforcement learning with hidden state. In *Proc., 11th Int'l. Machine Learning Workshop (Robot Learning)*, New Brunswick, NJ.

[McCallum, 1995b] McCallum, R. A. (1995b). Instance-based utile distinctions for reinforcement learning. In *The Proceedings of the Twelfth International Machine Learning Conference*. Morgan Kaufmann Publishers, Inc.

[Milner and Goodale, 1995] Milner, A. and Goodale, M. (1995). *The Visual Brain in Action*. Oxford University Press, Oxford.

[Miyachi et al., 1994] Miyachi, S., Miyashita, K., Karadi, Z., and Hikosaka, O. (1994). Effects of the blockade of monkey basal ganglia on the procedural learning and memory. In *Abstracts, 24th Annual Meeting, Society for Neuroscience*, page 357 (153.6), Miami Beach, FL.

[Nakayama, 1990] Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. In Blakemore, C., editor, *Vision: Coding and Efficiency*, pages 411–422. Cambridge University Press.

[Newell, 1990] Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press, Cambridge, MA.

[Norman and Bobrow, 1975] Norman, D. and Bobrow, D. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7:44–64.

[Noton and Stark, 1971a] Noton, D. and Stark, L. (1971a). Eye movements and visual perception. *Scientific American*, 224:34–43.

[Noton and Stark, 1971b] Noton, D. and Stark, L. (1971b). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11:929.

[O'Regan, 1992] O'Regan, J. (1992). Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian J. Psychology*, 46:461–488.

[O'Regan and Lévy-Schoen, 1983] O'Regan, J. and Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, 23:765–769.

[Paulesu et al., 1993] Paulesu, E., Frith, C., and Frackowiak, R. (1993). *Nature*, 362:342–345.

[Pellizzer et al., 1994] Pellizzer, G., Sargent, P., and Georgopoulos, A. (1994). Motor cortex and visuomotor memory scanning. In *Abstracts, 24th Annual Meeting, Society for Neuroscience*, page 983 (403.12), Miami Beach, FL.

[Pelz, 1995] Pelz, J. (1995). *Visual representations in a natural visuo-motor task.* PhD thesis, Department of Brain and Cognitive Sciences, University of Rochester.

[Pollatsek and Rayner, 1990] Pollatsek, A. and Rayner, K. (1990). Eye movements and lexical access in reading. In Balota, D., d'Arcais, G. F., and Rayner, K., editors, *Comprehension Processes in Reading*, pages 143–164. Lawrence Erlbaum Associates, Publisher, Hillsdale, NJ.

[Pook, 1995] Pook, P. (1995). *Teleassistance: Using deictic gestures to control robot action.* PhD thesis, Department of Computer Science, U. Rochester. Also appeared as TR 594.

[Pook and Ballard, 1994a] Pook, P. and Ballard, D. (1994a). Deictic teleassistance. In *Proc., IEEE/RSJ/GI Int'l. Conf. on Intelligent Robots and Systems*, Munich, Germany.

[Pook and Ballard, 1994b] Pook, P. and Ballard, D. (1994b). Teleassistance: Contextual guidance for autonomous manipulation. In *Proc., 12th Nat'l. Conf. on Artificial Intelligence (AAAI)*, Seattle, WA.

[Pylyshyn, 1989] Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32:65–97.

[Raichle, 1993] Raichle, M. (1993). The scratchpad of the mind. *Nature*, 363.

[Rao and Ballard, 1995a] Rao, R. and Ballard, D. (1995a). An active vision architecture based on iconic representations. *Artificial Intelligence*, 78:461–505.

[Rao and Ballard, 1995b] Rao, R. and Ballard, D. (1995b). Dynamic model of visual memory predicts neural response properties in the visual cortex. Technical Report National Resource Laboratory for the Study of Brain and Behavior Technical Report 95.4, University of Rochester. Submitted for journal publication.

[Rao and Ballard, 1996] Rao, R. and Ballard, D. (1996). A computational model of spatial representations that explains object-centered neglect in parietal patients. Submitted for conference publication.

[Rao et al., 1996] Rao, R., Zelinsky, G., Hayhoe, M., and Ballard, D. (1996). Modeling saccadic targeting in visual search. *Advances in Neural Information Processing Systems (NIPS 95).* To appear.

[Schlingensiepen et al., 1986] Schlingensiepen, K.-H., Campbell, F., Legge, G., and Walker, T. (1986). The importance of eye movements in the analysis of simple patterns. *Vision Research*, 26:1111–1117.

[Schultz et al., 1995] Schultz, W., Apicella, P., Romo, R., and Scarnati, E. (1995). Context-dependent activity in primate striatum reflecting past and future behavioral events. In Houk, J., Davis, J., and Beiser, D., editors, *Models of Information Processing in the Basal Ganglia*, pages 11–27. MIT Press (A Bradford Book), Cambridge, MA.

[Shastri, 1993] Shastri, L. (1993). From simple associations to systematic reasoning. *Behavioral and Brain Sciences*, 16(3):417–494.

[Simon, 1962] Simon, H. (1962). The architecture of complexity. In *Proc., American Philosophical Society, Vol. 26*, pages 467–482.

[Snyder and Andersen, 1994] Snyder, L. and Andersen, R. (1994). Effects of vestibular and neck proprioceptive signals on visual responses in posterior parietal cortex. In *Abstracts, 24th Annual Meeting, Society for Neuroscience*, page 1278 (525.1), Miami Beach, FL.

[Soechting and Flanders, 1989] Soechting, J. and Flanders, M. (1989). Errors in pointing are due to approximations in sensorimotor transformations. *Journal of Neuroscience*, 62:595–608.

[Stricanne et al., 1994] Stricanne, B., Xing, J., Mazzoni, P., and Andersen, R. (1994). Response of lip neurons to auditory targets for saccadic eye movements: A distributed coding for sensorimotor transformation. In *Abstracts, 24th Annual Meeting, Society for Neuroscience*, page 143 (65.1), Miami Beach, FL.

[Strick et al., 1995] Strick, P., Dum, R., and Picard, N. (1995). Macro-organization of the circuits connecting the basal ganglia with the cortical motor areas. In Houk, J., Davis, J., and Beiser, D., editors, *Models in Information Processing in the Basal Ganglia*, pages 117–130. MIT Press (A Bradford Book), Cambridge, MA.

[Swain and Ballard, 1991] Swain, M. and Ballard, D. (1991). Color indexing. *Int'l. J. Computer Vision*, 7(1):11–32.

[Tagaris et al., 1994] Tagaris, G., Kim, S.-G., Menon, R., Strupp, J., Andersen, P., Ugurbil, K., and Georgopoulos, A. (1994). High field (4 Telsa) functional MRI of mental rotation. In *Abstracts, 24th Annual Meeting, Society for Neuroscience*, page 353 (152.10), Miami Beach, FL.

[Trick and Pylyshyn, 1996] Trick, L. and Pylyshyn, Z. (1996). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*. In press.

[Ullman, 1984] Ullman, S. (1984). Visual routines. *Cognition*, 18:97–157. Also in S. Pinker (Ed.), *Visual Cognition*. Cambridge, MA: Bradford Books.

[Ungerleider and Mishkin, 1982] Ungerleider, L. and Mishkin, M. (1982). Two cortical visual systems. In Ingle, D., Goodale, M., and Mansfield, R., editors, *Analysis of Visual Behavior*, pages 549–585. MIT Press, Cambridge, MA.

[Van Essen et al., 1994] Van Essen, D., Anderson, C., and Olshausen, B. (1994). Dynamic routing strategies in sensory, motor, and cognitive processing. In Koch, C. and Davis, J., editors, *Large-Scale Neuronal Theories of the Brain*, pages 271–299. MIT Press (A Bradford Book), Cambridge, MA.

[Viviani, 1990] Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In Kowler, E., editor, *Eye Movements and their Role in Visual and Cognitive Processes. Reviews of Oculomotor Research V4*, pages 353–383. Elsevier.

[Wandell, 1995] Wandell, B. (1995). *Foundations of Vision Science: Behavior, Neuroscience, and Computation*. Sinauer, Sunderland.

[Whitehead and Ballard, 1990] Whitehead, S. and Ballard, D. (1990). Active perception and reinforcement learning. *Neural Computation*, 2(4):409–419.

[Whitehead and Ballard, 1991] Whitehead, S. and Ballard, D. (1991). Learning to perceive and act by trial and error. *Machine Learning*, 7(1):45–83.

[Wiskott and von der Malsburg, 1993] Wiskott, L. and von der Malsburg, C. (1993). A neural system for the recognition of partially occluded objects in cluttered scenes: A pilot study. *IJPRAI*, 7:935–948.

[Woodward et al., 1995] Woodward, D., Kirillov, A., Myre, C., and Sawyer, S. (1995). Neostriatal circuitry as a scalar memory: Modeling and ensemble neuron recording. In Houk, J., Davis, J., and Beiser, D., editors, *Models of Information Processing in the Basal Ganglia*. MIT Press (A Bradford Book), Cambridge, MA.

[Yarbus, 1967] Yarbus, A. (1967). *Eye Movements and Vision*. Plenum, New York.

[Yeterian and Pandya, 1995] Yeterian, E. and Pandya, D. (1995). Corticostriatal connections of extrastriate visual areas in rhesus monkeys. *Journal of Comparative Neurology*, 352:436–457.