BEHAVIORAL AND BRAIN SCIENCES (1997) **20,** 723–767
Printed in the United States of America

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
Electronic mail: *dana@cs.rochester.edu; mary@cs.rochester.edu; pook@isr.com; rao@salk.edu www.cs.rochester.edu/urcs.html*

Abstract: To describe phenomena that occur at different time scales, computational models of the brain must incorporate different levels of abstraction. At time scales of approximately ⅓ of a second, orienting movements of the body play a crucial role in cognition and form a useful computational level – more abstract than that used to capture natural phenomena but less abstract than what is traditionally used to study high-level cognitive processes such as reasoning. At this "embodiment level," the constraints of the physical system determine the nature of cognitive operations. The key synergy is that at time scales of about ⅓ of a second, the natural sequentiality of body movements can be matched to the natural computational economies of sequential decision systems through a system of implicit reference called *deictic* in which pointing movements are used to bind objects in the world to cognitive programs. This target article focuses on how deictic bindings make it possible to perform natural tasks. Deictic computation provides a mechanism for representing the essential features that link external sensory data with internal cognitive programs and motor actions. 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: binding; brain computation; deictic computations; embodiment; eye movements; natural tasks; pointers; sensory-motor tasks; working memory.

1. Embodiment

This target article 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, meaning 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 ⅓ 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 ⅓ second time scale, one must first understand the different time scales that are available for computation in the brain. Because the brain is a physical system, communicating over long distances is costly in time and space and therefore local computation is the most efficient. Local computation can be used 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 (1990) has pointed out 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 1 millisecond to generate. This means that the circuitry that uses these spikes for computation has to run slower than this rate. If we use Newell's assumption

 0140-525X/XX $9.00+.10

that about 10 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 takes on the order of 100 milliseconds. A deliberate act would correspond to any kind of perceptual decision, for example, 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 10s of milliseconds, so Newell's 100 milliseconds probably overestimates the correct value by at most a factor of 2 or 3 (Duncan et al. 1994). Body movements such as saccadic eye movements take about 200–300 milliseconds to generate, which is about 5 times the duration of a perceptual act. At the next abstraction level, the composition of tasks by primitive acts requires the persistence of the information in time. Therefore, 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 ⅓ 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 call 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. Computation at this level governs the rapid deployment of the body's sensors and effectors to bind variables in behavioral programs. This computation provides a language that 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. There are several ramifications of this view:

1. Cognitive and perceptual processes cannot be easily separated, and are in fact 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 ⅓ 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.

2. 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. The brain's programs structure behaviors to minimize the amount of working memory needed at any instant. The structure of working memory and its role in the 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.

3. 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 bound only for as long as they are needed in the encompassing task. In addition, the contents of an item vary with task context, and are usually only fragmentary portions of the available sensory stimulus.

Table 1. *The organization of human computation into temporal bands*

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

Source: Adapted from Newell (1990), but with some time scales adjusted to account for experimental observations.

1.1. Deictic sensory-motor primitives

A primary example of a rapid sensory-motor primitive is the saccadic eye movement. Saccadic eye movements are typically made at the rate of about 3 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 & Stark 1971b) and Simon and Chase in their study of eye movement patterns in chess (Chase & Simon 1973), or on the duration of fixation patterns themselves (e.g., Just & Carpenter 1976). But as Viviani (1990) points out, 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 & 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.

Because humans can fixate on 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-

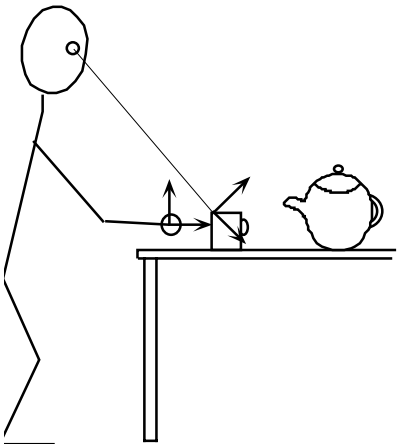

Figure 1. Biological and psychophysical data argue for deictic frames. These frames are selected by the observer to suit information-gathering goals.

Table 2. *A portion of computer memory illustrating the use of pointers*

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

Left: Reference is to beeA. Right: Reference is to beeB. The change in reference can be accomplished by changing a single memory cell.

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 & 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 that fixation can then 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 nondeictic 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 (Tsotsos et al. 1995). Thus, one can think of vision as having either mechanical or neural deictic devices: fixation and attention. This target article 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 from conventional silicon computers, they both have to address many of the same problems. It is sometimes useful therefore to look at how problems are handled by silicon computers. One major problem is that of variable binding. As recognized by 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] in which 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-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. 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

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

referent can be used as a unit in computation. The properties of the pointer referent may not be, and almost never are, 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 dedicate themselves to processing a single token. If behaviors require additional variables, they must be kept in a separate system called *working memory* (Baddeley 1986; Broadbent 1958; Logie 1995). Although the brain and computer work on very different principles, 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 (Ballard 1986; Barlow 1972) that cannot be easily changed. It seems therefore that at the ⅓ 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 purpose of this target article is to explain why deictic codes are a good model for behavior at the embodiment level. The presentation is organized into three main sections.

1. 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.

2. 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 is reset from moment to moment with deictic actions.

3. Section 4 discusses the implications of deictic computation in understanding cortical circuitry. A consequence of complex programs being composed of simpler primitives, each of 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 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 in a local region because the human eye has much better resolution in a small region near the optical axis, that is, the fovea. Over a region of approximately 1° to 2° of visual angle the resolution is better than in the periphery by an order of magnitude. 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 texts of reversed letters (see Fig. 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 because 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) and 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 & 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 Whitehead & Ballard 1991). This problem is solvable by computer simulation

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 & Anton 1987).

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

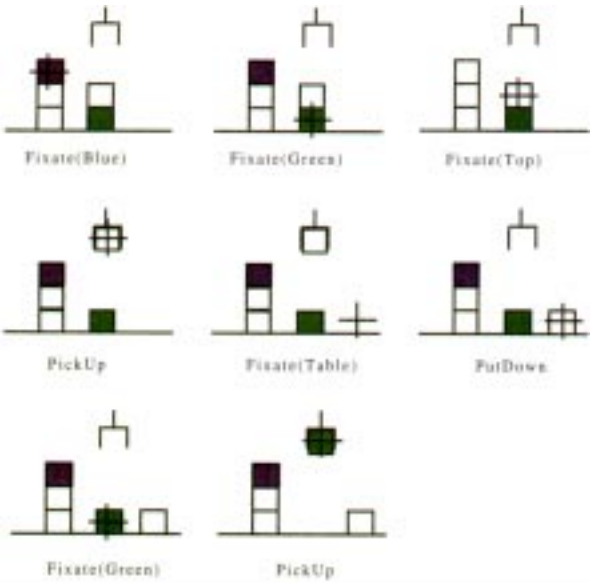

Figure 3. A graphical display from the output of a program that has learned the "pick up the green block" task. The steps in the program use deictic references rather than geometrical coordinates. For each stage in the solution, the plus symbol shows the location of the fixation point. (In the figure, blue appears as black and green appears as grey.)

using reinforcement learning (Whitehead & 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. Therefore, when fixating a blue block, it is 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) showed 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, because 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-fixating)* will transfer the gaze to the top of the stack currently fixated.

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 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

Table 3. *The sensory representation used to solve the blocks task*

Bits	Feature
1	red-in-scene
1	green-in-scene
1	blue-in-scene
1	object-in-hand
2	fixated-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

The representation consists of four global features and twelve features accessed by the fixation and attention pointers. The left column shows the number of bits used to represent each feature. The right column describes each feature.

Table 4. *The discrete set of actions used to solve the blocks task*

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

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.

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

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: the necessary state required to keep track of a complex task 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, it can be handled by explicit pointing such as fixation, or "neural" pointing such as "attention." Additional pointers require some additional representational mechanism, however. 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 reused 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 picking up a green block is changed to that of picking up a yellow block and that 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 to find behavioral programs that require only a small number of pointers.

A second reason for having only a small number of pointers is that this may be sufficient for the task. McCallum (1995) builds a "history tree" that 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 1995). These simulations suggest that impressive performance may be achievable with very little context.

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

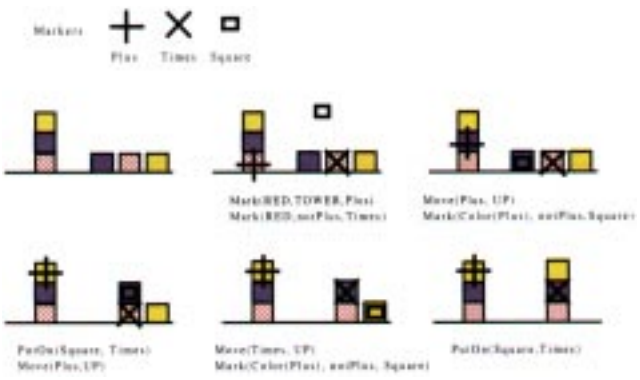

Figure 4. A tower of blocks can be copied using three pointers. At any instant the access to the sensorium is limited to the marked objects. The number of pointers is minimized by re-using pointers during the course of the behavior.

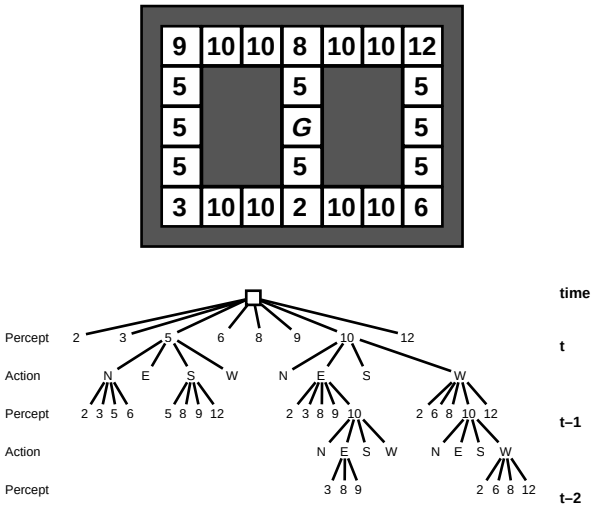

Figure 5. The different amounts of context used in decision making can require different amounts of working memory. (Top) Maze used in McCallum's reinforcement learning algorithm (McCallum 1995). The numerical codes indicate the walls surrounding each location. For example, a north and south wall is 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 (to simplify the figure, the actions stored at the leaf of the tree are not shown, but they can be inferred from the maze).

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

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].) We can see therefore 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 example, 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 (Andersen 1995; Jeannerod 1988; Soechting & Flanders 1989), 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 & Ballard 1994b). In teleassistance, a human operator is the "brain" to an otherwise autonomous dextrous robot manipulator. The operator does not control the robot directly, 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.[9]

The sign language is very simple. To help a robot 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.[10] With adequate dexterity and compliance in the robot manipulator, simply flexing its fingers toward the origin of that frame, coupled with a force control loop, suffices to form a stable grasp. Because 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 requires only 22% of the total time for executive control (indicated by the extent of the dark shaded areas in Fig. 6). Thus the pointers required to implement the state machine of Figure 7 are required for only 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.

Table 5. *Signs used in teleassistance experiment*

Sign	Meaning
POINT	While the operator points, the robot moves in the direction of the pointing axis, independent 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 his hand to define a grasp form and a new spatial frame centered on the palm.
HALT	Halting is used to punctuate the motor program.

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 (1992) 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. Furthermore, they suggest that both these functions may involve both dorsal and ventral streams. The point is that the *identification/location* dichotomy is more a functional than an architectural separation.

Implementation of deictic location and identification

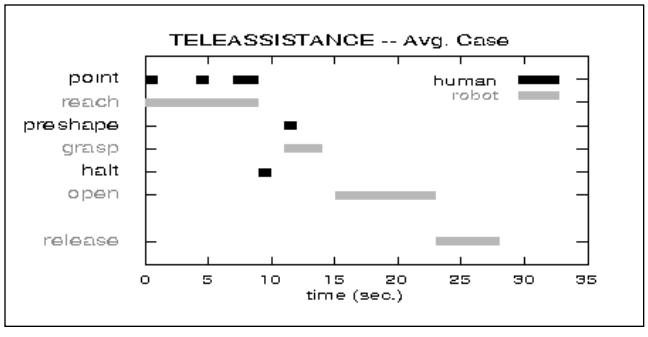

Figure 6. Results of control by teleassistance in the door-opening task. Bars show the time spent in each subtask. The teleassistance model shows the interaction between deictic commands that signal the low-level autonomous routines and the routines themselves.

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

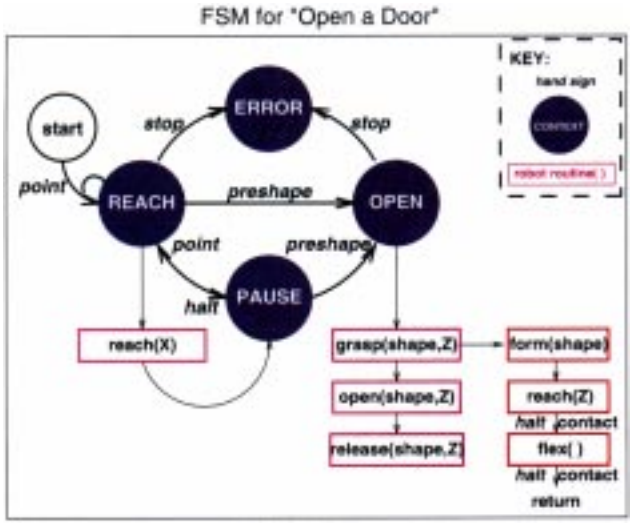

Figure 7. Simple state transition used for the interpretation of deictic signs. The deictic signs defined in Table 5 map directly onto transitions between states (shaded) of a simple control program. In a successful trial, a "point" sign from the operator results in the REACH state, which in turn triggers autonomous movement in the pointed direction. Next, a "halt" sign stops the robot, resulting in the PAUSE state. Finally, an appropriate "preshape" sign leads to the OPEN state, which triggers the autonomous routines that open the door.

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 simplify the general task into simpler identification and location tasks, thereby making it computationally more tractable. 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 having to find the image coordinates of only 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 (Rao & Ballard 1995; Swain & Ballard 1991). Therefore, 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 and section 4 describes how these operations might be implemented.

The concept of pointers changes the conceptual focus of computation from continuous to discrete processing. Such processing is centered around the momentary disposition of pointers that indicate 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. We can therefore 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 pointers 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 this conceptualization is in fact appropriate for human behavior. Because the eyes allow a natural implementation of deictic strategies, the question immediately raised is how humans actually use their eye movements in the context of natural behaviors. We designed a series of experiments to test the use of

Table 6. *The organization of visual computation into* identification/location *modules may have a basis in complexity*

		Models	
		One	Many
Image Parts	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?

Trying to match many image segments to many models simultaneously may be too difficult. The complexity of visual computation can be substantially reduced, however, by decomposing a given task into simpler deictic operations.

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

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: as in any natural behavior, the subjects organize the steps to compose the 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 the following sections. 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 into three areas, the *model,* the *resource,* and the *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.[11] Subjects copied the block pattern as described above, and were asked only to perform the task as quickly as possible. No other instructions were given, so as not to bias subjects toward 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 subpatterns 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 chose 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.

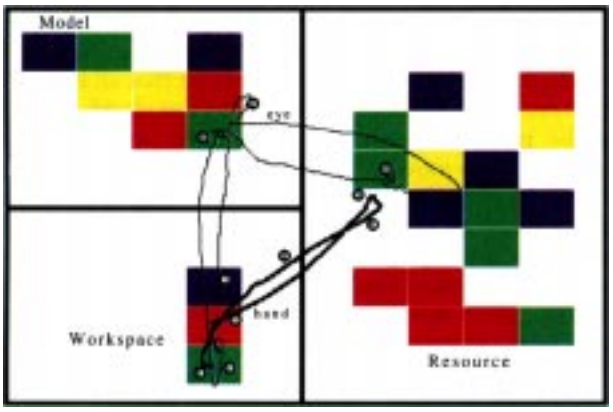

Figure 8. Copying a single block within the task. The eye position trace is shown by the cross and the dotted line. The cursor trace is shown by the arrow and the dark line. The numbers indicate corresponding points in time for the eye and hand traces.

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. Because 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 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, the result should be a "model-pickup-model-drop" sequence. If the color is known, a "pickup-model-drop" sequence should result; if the location is known, we should see a "model-pickup-drop" sequence. If both are known, there should be a "pickup-drop" sequence. In the data, "pickup-drop" sequences were invariably the last one or sometimes two blocks in the sequence. With respect to color and location, therefore, 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 are shown as the dark bars in Figure 9b. The lowest-memory "model-pickup-

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

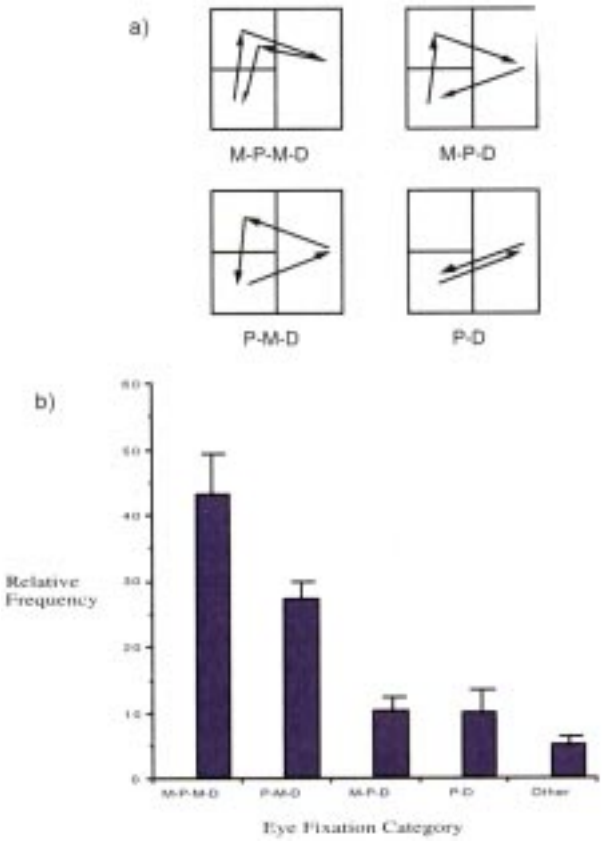

Figure 9. (a) The codes. "M" means that the eyes are directed to the model; "P" and "D" mean that the eyes and cursor are coincident at the pickup point and drop-off point, respectively. Thus for the PMD strategy, the eye goes directly to the resource for pickup, then to the model area, and then to the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

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. We take the frequent access to the model area during the construction of the copy 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 required.

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, especially because they have been instructed to complete the task as quickly as possible, and memorization saves time.

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

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

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, for the technical reasons discussed in section 2, the carrying cost of working memory is expensive compared to the cost of acquiring the information on-line, 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 is based on 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 appear instead 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. A control experiment argues that both of these alternate hypotheses are unlikely, however. 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 component shapes. In this case, subjects do abandon eye movements to the model. We conclude therefore that the movements to the model in the standard case are necessary and related to the individual properties of blocks.

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

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 10 as "Before Pickup," 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.[12] This is indicated by the zig-zag. The small arrow indicates the color change in the block. In another condition, shown in Figure 10 as "After Pickup," 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 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 each 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 millisecond) 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.

In the fixations following the pickup there is a large (129 millisecond) 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

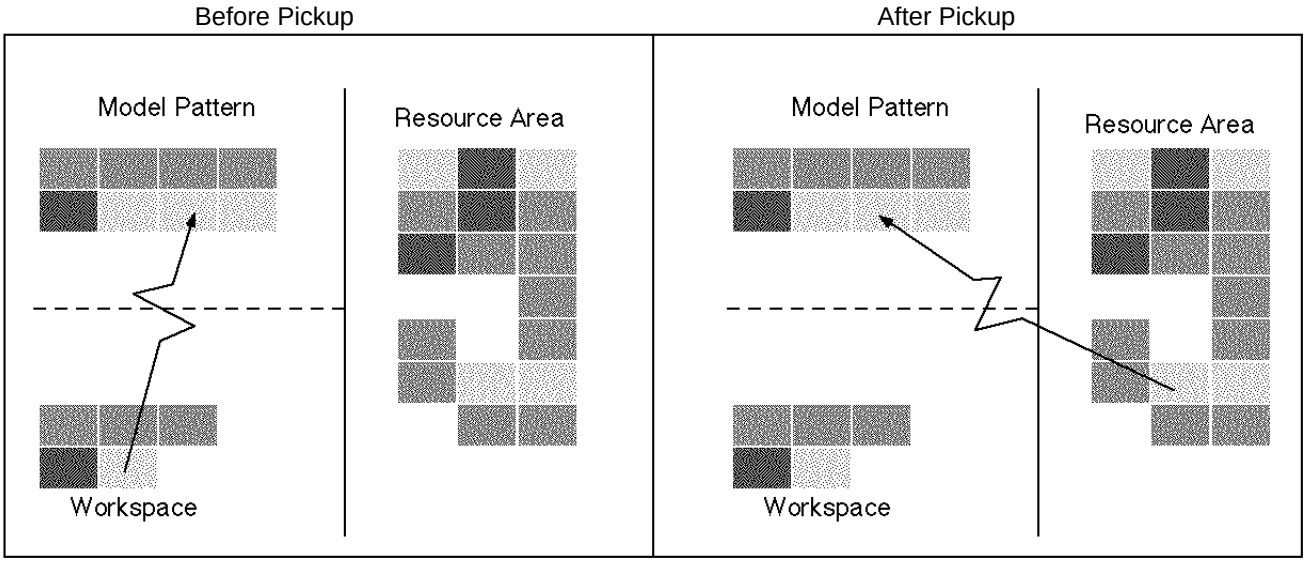

Figure 10. Two different experimental conditions for the color-changing experiment. (Left) The color of an uncopied model block is changed during a workspace-to-model saccade. (Right) The color of an uncopied model block is changed during a resource-to-model saccade.

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

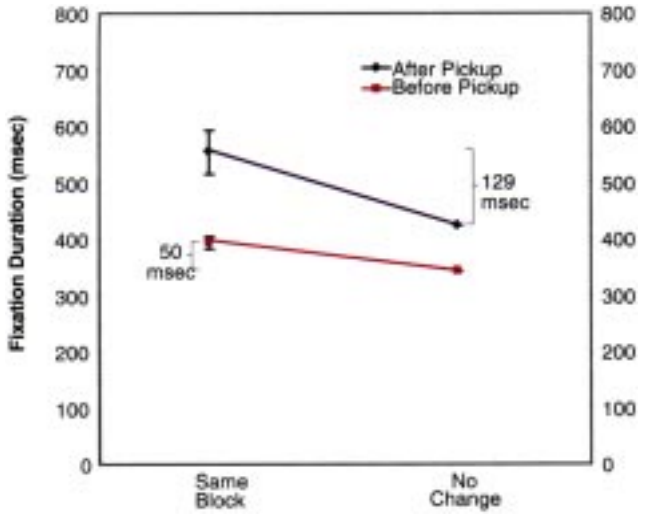

Figure 11. Results of the color-changing experiments, showing the fixation time when the target of the fixation was the next block moved compared to control trials where there was no change. For the "Before Pickup" condition (lower points) there is a small change in fixation time. The "After Pickup"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

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 this case 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 well be 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 (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 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 does not 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 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). In addition, to complete the task, we need 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

The experimental data in the previous section supports the notion of deictic computation using pointers, but does not address the issue of how the *referents* of these pointers are computed and maintained. In this section, we switch to a computational venue to suggest how this might be done. We also switch abstraction levels to talk about the faster operations occurring at the level of the deliberate act (Table 1).

Deictic actions suggest that computation is limited to just what is needed for the current point in the task. This is illustrated very dramatically by the blocks task of the previous section, particularly by the experiments that switch colors during saccades. These experiments suggest that (1) the brain seems to postpone binding color and relative location information until just before it is required, and (2) 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 precompute 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

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

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. This strongly suggests a functional view of visual computation in which different operations are applied at different stages during a complex task.

4.1. Functional routines

The hypothesis that vision must be functional relies crucially on the existence of some mechanism for spanning the space of task-dependent representations. An attractive way of achieving this is to compose complex behaviors from primitive routines. Thus, at the level of abstraction below the embodiment level (i.e., at the attentive or deliberate act 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.

Although the functional primitives (or "routines") could exist for any modality, we concentrate here 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. Therefore, the two primary 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 because 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 (Jones & Malik 1992; Rao & Ballard 1995; Rao et al. 1996; Wiskott & 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 & Adelson 1991) but very similar filters that have properties commensurate with those observed in the primate cortex can be learned from natural stimuli (Rao & Ballard 1996c). 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°, respectively. These are used for each of 3 color channels (intensity, red-green opponent, and yellow-blue opponent) and at 3 different scales, for a total of 81 filter responses for each image point. The exact number and composition of the filters is unimportant 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 & Ballard 1995), and therefore each location in the sensorium can be given a unique descriptor.

4.1.1. Identification. The identification routine matches a foveal set of image features with a library of sets of stored model features (Fig. 16). The result is the model coordinates (or identity) of the best match. In the case of extracting the color of the material at the fixation point, the

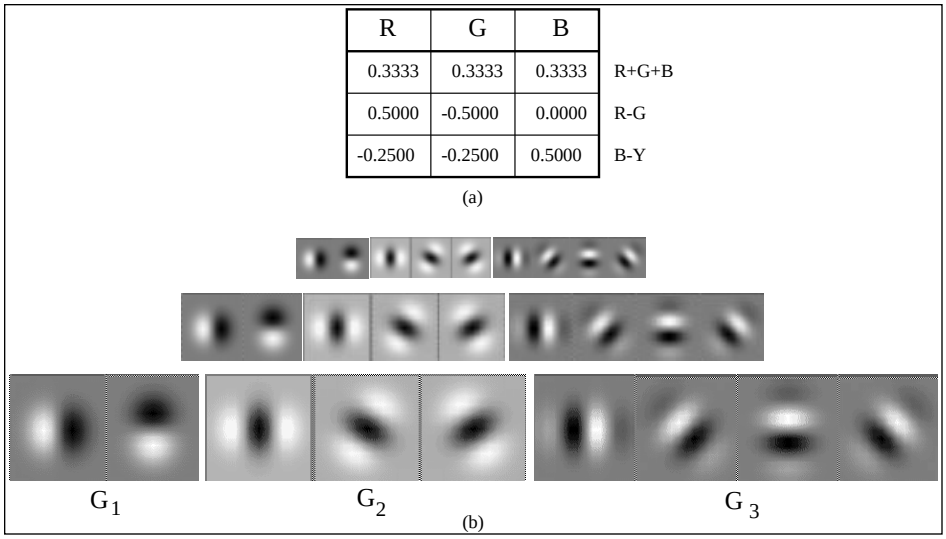

Figure 12. The spatiochromatic basis functions that uniquely define points in the image even under some view variation. Motivation for these basis functions comes from statistical characterizations of natural image stimuli (Derrico & Buchsbaum 1991; Hancock et al. 1992; Rao & Ballard 1996c). (a) shows the weights assigned to the three input color channels, generating a single achromatic channel (R+G+B) and two color-opponent channels (R-G and B-Y). (b) shows the nine "steerable" spatial filters used at three octave-separated scales for each of the three channels in (a) (bright regions denote positive magnitude whereas darker regions denote negative magnitude). At each scale, these nine filters are comprised of two first-order derivatives of a 2D photometric Gaussian (G_1), three second-order derivatives (G_2), and four third-order derivatives (G_3). Thus, there are 3 color channels, 3 scales per channel, and 9 steerable filters per scale, for a total of 81 filter responses characterizing each location in the image. These 81 spatiochromatic measurements can be thought of as the referent of a pointer.

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

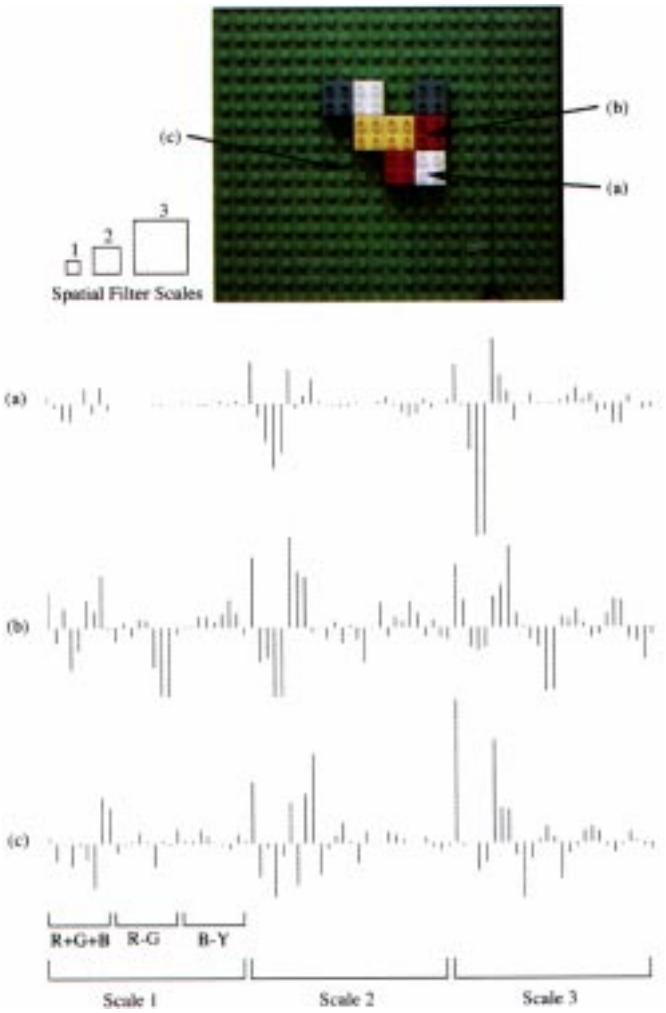

Figure 13. Using spatiochromatic filters to extract task-dependent properties. The blocks image used in the copying task is shown at the top. The three scales at which the filters of Figure 12 were applied to the image are shown on the left. (a) The filter responses for a location on a white block. Each individual filter, when convolved with the local intensities near the given image location, results in one measurement, for a total of 81 measurements per image location. The resulting 81-element vector can be viewed as the referent of a single pointer. Positive responses in the vector are represented as a bar above the horizontal, negative responses as a bar below the horizontal. As expected, the vector for the white block has many low responses caused by the opponent channel coding. (b) The filter response vector for a location on a red block. (c) The filter response vector for a location in the green background.

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.

4.1.2. Location. The location routine matches a given 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

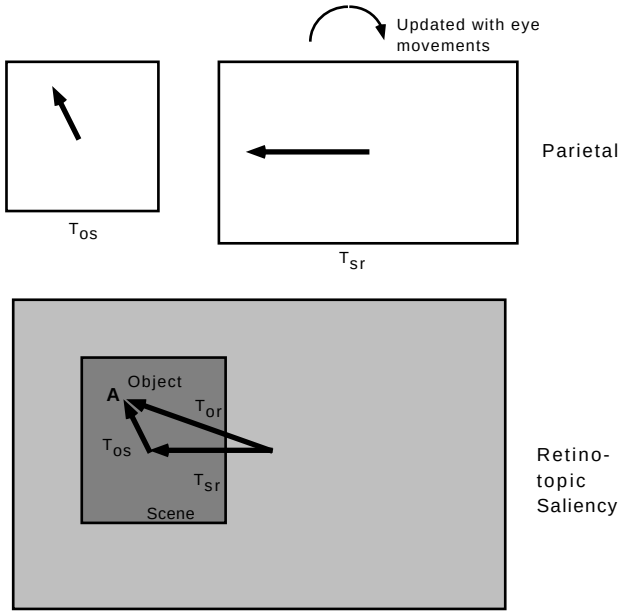

Figure 14. To represent the geometric relations of visual features three transformations are fundamental. The first describes how a particular depiction of the world, or scene, is related to the retinal coordinate system of the current fixation T_{sr}. The second describes how objects can be related 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}. Such an arrangement is computationally extremely efficacious. For example, in the case of the blocks task, representing the "resource" area (right side of the board) in T_{os} allows the search for new blocks of a particular color to be constrained to that area, regardless of current eye-position.

templates can be recorded. The particular location problem illustrated in Figure 15 is that of finding a block of a particular color in the resource area. 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 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.[13]

The location routine determines the transformation that describes the relationship between an object-centered reference frame and the current view frame represented by the fixation point. It is easy to demonstrate the importance of a third frame, however. 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 (Land & Lee 1994). 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. Experimental evidence for object-centered reference frames comes from studies of object-centered neglect in parietal patients (Behrmann & Moscovitch 1994) and from neurophysiological data indicating the existence of neurons sensitive to

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

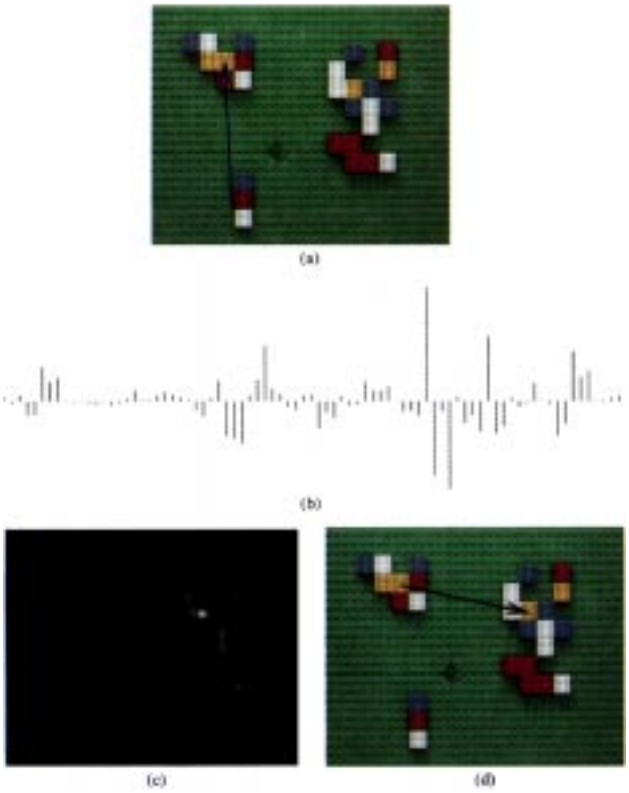

Figure 15. Using spatiochromatic filter representations for programming eye movements in the blocks task. (a) Fixating a yellow block in the model causes its filter responses (b) to be placed in working memory. (c) Those responses are then matched to filter responses at all locations in the resource 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 to pick up the yellow block in the resource area.

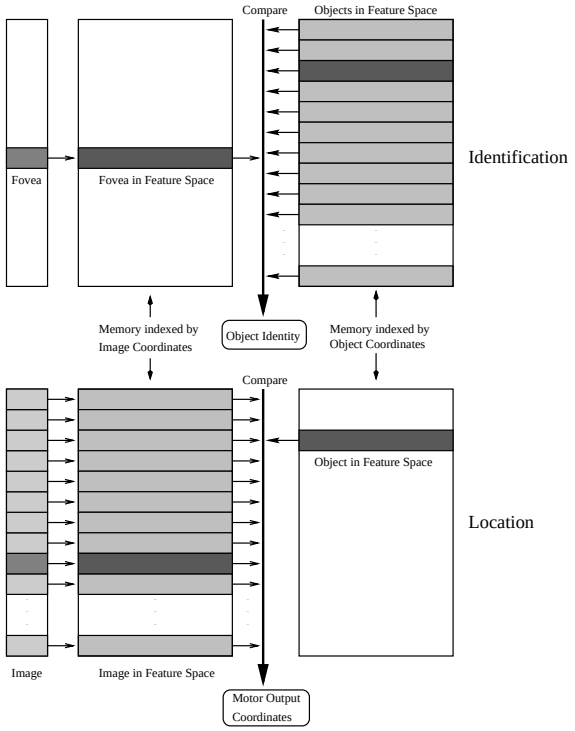

Figure 16. A highly schematized depiction of the identification/location control architecture (adapted from Rao & Ballard 1995). Neurons of the thalamus and visual cortex are summarized on the left in terms of a retinotopically-indexed memory of filtered representations; neurons of the infero-temporal cortex are summarized on the right as a model-indexed memory of filtered object representations. (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.

object-centered movements in the supplementary eye field (SEF) region of the primate cerebral cortex (Olson & Gettner 1995).

In the context of the blocks tasks, the factoring of retinotopic information into object-centered and scene-centered frames (Fig. 14) allows for temporary storage of remembered locations as well as task-dependent constraints that direct the eyes to appropriate points in the model, workspace, and resource areas. Task-dependent constraints are coded as transformations T_{os} and T_{sr}, as shown in Figure 14. Given explicit memory for these two transformations, a location relative to the scene can be placed in egocentric space by concatenating the two transformations. This works when the result is on the retina but also in the more general case where it may be hidden from immediate view. Support for such a strategy comes from simulations that show that it can model a variety of observed data from patients with lesions in the parietal cortex (Rao & Ballard 1996b).

As a specific example of how these frames may be used, consider the problem of finding a yellow block. Figure 15 shows how this could be done. When fixating the model (a), the filter response vector for a yellow block (b) is extracted and stored in working memory as a pointer referent. 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 retinal coordinates. The best match defines an oculomotor target for fixation in a "saliency map" (c). However, the search for the yellow block must be constrained to the resource area in the visual field. But how is the resource area delineated? An easy solution is to define it as a template with respect to the scene in T_{os}. Then, T_{sr} allows this template to be appropriately positioned in retinal coordinates.[14] This is done in (c) in the figure. Thus, the task-relevant reference frames allow 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).

4.2. Visual cortex and pointer referents

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 pointer referent may be iconic, and of considerable size. Now as a crude first approximation, let us think of the cortex as a form of content-addressable memory. This

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

allows the cortex to be modeled as holding the contents of a pointer.

Let us now speculate specifically on how the visual routines of the preceding section (that create pointer referents) could be implemented in terms of cortical architecture. In particular, consider the use of pointers to solve both *identification* and *location* problems. To implement the corresponding visual 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. In particular, 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 first 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 cortico-cortical feedback connections. The results of matching would nominally reside in the parietal cortex in the form of saliency maps denoting task-relevant spatial locations.

4.3. 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. If the job of the cortex is to hold the contents of pointers (as suggested in the previous section), additional *extracortical* circuitry is required to realize the different functionality involved in manipulating and using these pointer referents to solve a given task. For example, all of the following need to be done:

1. Sequencing in the task to determine what sensory processing is to be done;
2. Processing to extract task-dependent representations; and
3. Binding of these results to deictic pointers in working memory.

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

The *basal ganglia* comprise an extensive subcortical 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 (Schultz et al. 1995). 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. The implication therefore is that the monkey is learning to predict delayed reward and coding it via an internal dopamine messenger. The basal ganglia have extensive connections to the cortex that emphasize frontal and motor cortex (Graybiel & 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 represent 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 & 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 represent the 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, because 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 ⅓ second time scale one cannot think of parts of the brain in isolation. This point has also been extensively argued by

Fuster (1989; 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. Therefore, the kind of information held in working memory should reflect the kind of things for which it is used.

5. Discussion and conclusions

The focus of this target article 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 is different. Examination of computation at the embodiment level's ⅓ 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 direct their foci to localized parts of space quickly. 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 & 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 do 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 that fixation is used as a deictic pointing device to bind items dynamically in working memory. Section 4 examined how component low-level routines (that is, at the level of perceptual acts) might affect the referent of pointers. This formulation owes much to Milner and Goodale (1995) and provides a 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 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.* One 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 (McCallum 1994; Pook & Ballard 1994a; Whitehead & Ballard 1991).

4. *The simplification of sensory-motor routines.* Deictic codes lead to functional models of vision (Rao & Ballard 1995) 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 we can do without many of the products of the sensorium that we might have thought were necessary.

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.

5.1. The generality of the blocks task

One might think 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 & Stark 1971a) and chess (Chase & 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.

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

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 & Lee 1994). That is, the fixation patterns have predictive value for the driver's next action.

5.2. 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., Logie 1995; Norman & Bobrow 1975). 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 phenomenon 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 perceptuomotor 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 "preattentive" visual search apparently operates on information that 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. This is consistent with evidence by Johnston and Dark (1986) that selective attention can be guided by active schemata (of any kind).

Another implication of the ideas described here is that it is important to distinguish between events at different time scales. Events at the "embodiment" level reveal temporal structure of the ongoing cognitive program, whereas events at the level of the deliberate act reveal the temporal structure of the internal computational machinery subserving the cognitive program. It is therefore possible that the difficulty in estimating the timing of attentional processes (Chun & Potter 1995; Duncan et al. 1994; Ward et al. 1996) and in separating preattentive from attentive processes (Joseph et al. 1996; Wolfe 1996b) reflects ambiguity in the computational level tapped by different experimental paradigms. It is also possible that the process of "automatization" (Schneider et al. 1984) reflects a reorganization of the internal machinery, resulting in a transition from the embodiment level to the lower deliberate-act level.

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 we described herein shift the focus to the ongoing *process* of cycling information through working memory (Just & Carpenter 1992). From our 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. In keeping with 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.

5.3. 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. Critiques of this view have been presented by Churchland et al. (1994) and Ballard (1996). 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 (Brady 1981; Marr 1982). The emergence of an alternative approach (called active or animate vision; Aloimonos et al. 1988; Bajcsy 1988; Ballard 1991) that takes advantage of observer actions to minimize representations (Agre & Chapman 1987; Brooks 1986; 1991) 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, because color and location information appear to be ac-

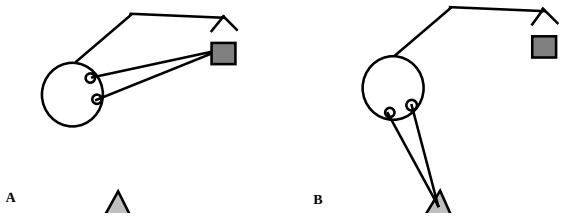

Figure 17. The fact that subjects can reach or look to targets outside their immediate field of view does not necessarily imply complete three-dimensional spatial representations (see discussion in the text).

quired separately, it appears that even in the attended regions the perceptual representation may be quite minimal. Therefore, 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 and Lévy-Schoen (1983) and 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. Wolfe's work on "post-attentive" vision is also consistent with this view (Wolfe 1996a).

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 4.1.2 in conjunction with the reference frames in Figure 14 (see also Hayhoe et al. 1992). The crucial constraint from our point of view, however, 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 compute in body-centered frames when required, but rather that such task-relevant frames are likely to be temporary.

5.4. 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. 1991; 1992), 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, independent 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 (Grimes & McConkie 1995; Irwin 1991; Irwin et al. 1990; Lachter & Hayhoe 1996). The saccade-updating experiment in section 3 supports the idea that the information retained across saccades depends crucially on the task. We would expect therefore that the recent findings of Duhamel et al. (1992), that neurons in the 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.

5.5. 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 the 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 (see also Gallant et al. 1995). This dependence of neural responses on task context can be mediated by the extensive feedback projections known to exist in the visual cortex.[15] Therefore, both striate and extra-striate visual cortical areas may in fact be computing highly specific task-relevant representations. Dramatic evidence for this kind of view of the cortex is provided by experiments by Maunsell (1993). In recordings from the 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. Therefore, although elegant techniques such as tachistoscopic presentations and backward masking have been used to isolate the feedforward pathway, revealing much about its structure (Wandell 1995), the difficulty in doing so 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. The very first visual measures are therefore necessarily dependent on the location of the fixation point which in turn depends on the current cognitive goal. Andersen and Snyder (Snyder & Andersen 1994; Stricanne et al. 1994) have also shown that the parietal cortex contains neurons sensitive to exocentric location in space. Motor cortex recordings also show the use

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

of exocentric or task-relevant frames (Helms-Tillery et al. 1991; Pellizzer et al. 1994; Tagaris et al. 1994).

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 uses 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 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, primarily 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.

NOTES

1. One can have systems for which local effects can be of great consequence at long scales. Such systems are termed chaotic (e.g., Baker & Gollub 1990), but they cannot be easily used in goal-directed computation. The bottom line is that for any physical system to be manageable, it must be organized hierarchically.

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

3. This is a very different assertion from 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.

4. For technical details see Agre and Chapman (1987).

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

6. 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.

7. Of the three current models of neural computation – reinforcement learning (Barto et al. 1990), neural networks (Hertz et al.1991), and genetic algorithms (Goldberg 1989; Koza 1992) – 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 (Schultz et al. 1995; Woodward et al. 1995).

8. A nondeictic way to solve this task would be to process the scene so that each block is catalogued and has a unique identification code. Then the copying task could be accomplished by searching the space of all possible relationships among coded items for the right ones. The problem is that this strategy is very expensive, because 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 reassigned to blocks dynamically. (Possible mechanisms for this are described in sect. 4.)

9. Very loosely, the analogy here is that the human operator is representing control at the level of the cortex and the midbrain, whereas the robot is representing control at the level of the spinal cord.

10. Because 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.

11. 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 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 a 10–15 min arc over about a 15° range.

12. Eye position was monitored by the Dual Purkinje Image eye tracker. Saccades were detected and the display updated within the 17 msec 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 msec. The saccades in this experiment typically lasted about 50 msec 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.

13. 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 & Malik 1992) have ameliorated these disadvantages. If the template is created dynamically and has a limited lifetime, then the first objection is less important because 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 (e.g., see Rao & Ballard 1995).

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

15. In terms of the computational model of section 4, both the identification and location routines crucially rely on the existence of cortico-cortical feedback connections (Rao & Ballard 1996a; 1996c).

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged

Are multiple fixations necessarily deictic?

Sally Bogacz

Department of Psychology, University of Maryland at College Park, College Park, MD 20742. **sb106@umail.umd.edu**

Abstract: The motor system might well use deictic strategies when subjects learn a new task. However, its not clear that Ballard et al. show this. Multiple eye-fixations may have little to do with deixis and more to do with the unfamiliarity of the task. In any case, deixis does not entail embodiment, since a disembodied Cartesian brain could use deictic strategies.

Ballard, Hayhoe, Pook & Rao align themselves with an increasingly popular approach to the mind, one that emphasizes the situatedness of cognition in the environment in contrast to the more intellectualist "disembodied" tradition of Descartes (sect. 1, para. 2). The question that I will explore in this commentary is whether the data that Ballard et al. present really bear upon this issue.

Ballard et al.'s main thesis is that at a time scale of one-third of a second, deictic strategies – presumably, the use of demonstratives such as "this" and "that" – play an essential role in the brain's symbolic computations by providing an efficient way of keeping track of relevant targets in the environment (sect. 1, para. 2; sect. 1.1, para. 2; sect. 1.2, para. 4). By contrast, most accounts of motor control either posit complex representations such as Rosenbaum et al.'s (1983) hierarchical "trees," or complex computations such as Gallistel's (in press) series of transforms. What make deictic strategies special is their simplicity, which derives from their capacity to control the location at which processing occurs (Ullman 1984). Only targets relevant to the task are processed, and this vastly simplifies computations (Ballard 1991). For example, if people used deictic strategies in sight-reading music, complex conceptual knowledge about a musical score would be unnecessary. Instead, a deictic strategy would mean that the representational content could be simplified to include only pointers to locally perceived notes and correlated finger movements.

Ballard et al.'s evidence for their deictic hypothesis comes from the "block copying" task described in section 3. The data are shown in Table 7 and Figure 9 and focus exclusively on eye-movement strategies used by the subject. We learn that subjects are more likely to use a multiple-fixation strategy drawing minimally on memory. But we also learn that using minimal memory is very costly in terms of how long it takes to perform the task: Table 7 indicates that a task with high-fixation strategy that uses minimal amounts of memory takes twice as long as one with a low-fixation strategy that uses lots of memory. This suggests that deictic strategies are not as efficient as memory strategies.

The authors explain the data in Table 7 by claiming that memory is more costly than acquiring information on-line (sect. 3.2, para. 2) because deictic strategies are able to simplify the computations substantially (sect. 1.1, para. 2; sect. 2.4, para. 2). An alternative explanation is possible, however: that the reason multiple fixations are correlated with slower performance on the task is *not* that the subject is using deictic strategies – a puzzling assertion because if deictic strategies simplify computation then subjects should be faster – but because the subject is doing something unfamiliar, that is, learning a new task that requires careful monitoring. Ballard et al. could help the reader to decide between these alternative hypotheses by presenting learning-curve plots that show eye-movement frequency and the time taken by each subject.

Thus, although Ballard et al.'s hypothesis has some plausibility – because it makes intuitive sense that the motor system would need to use a quick-and-dirty heuristic in order to react quickly to a changing environment – the evidence they present raises the concern that their multiple fixation results have little to do with deixis and could instead be explained by the obvious fact that the unfamiliar is apt to be more closely monitored than the familiar.

Even if it were shown that deictic strategies have psychological reality, Ballard et al.'s claim (sect. 1, para. 1) about embodiment seems entirely gratuitous: there is nothing about deixis *per se* that depends on embodiment; a disembodied, unsituated Cartesian brain – a brain in a vat – could use the very same deictic strategies of an embodied, situated one. The only difference would be that the deictic symbols of the disembodied, unsituated brain would simply fail to refer to anything. Thus, a further argument would have to be made to show either that Descartes was mistaken in thinking that cognition could be abstracted from its environment (the "situated" hypothesis) or that our cognition is constrained by the shape of the human body (the "embodied" hypothesis) as Ballard et al. claim.

ACKNOWLEDGMENT

I am indebted to Christopher Cherniak and Georges Rey for help in preparing this commentary.

Cognition without representational redescription

Joanna Bryson[a] and Will Lowe[b]

[a]*Laboratory for Cognitive Neuroscience and Intelligent Systems, Edinburgh University, Edinburgh EH8 9JZ, United Kingdom.*
joannab@dai.ed.ac.uk **www.ai.mit.edu/people/joanna/joanna.html**
[b]*Centre for Cognitive Science, Edinburgh University, Edinburgh EH8 9JZ, United Kingdom.* **will@cogsci.ed.ac.uk** **www.cogsci.ed.ac.uk/~will**

Abstract: Ballard et al. show how control structures using minimal state can be made flexible enough for complex cognitive tasks by using deictic pointers, but they do so within a specific computational framework. We discuss broader implications in cognition and memory and provide biological evidence for their theory. We also suggest an alternative account of pointer binding, which may better explain their limited number.

Ballard, Hayhoe, Pook & Rao point out in their conclusion that deictic coding is intimately connected to theories of intelligence that minimize representational content. As researchers in the field of reactive, behavior based artificial intelligence (JB) and modelling human semantic memory (WL), we consider Ballard et al.'s theory a valuable way to reconceptualize intelligence.

Traditional AI considers as its core problem representational redescription from a perception model to an action model. The reactive approach claims to eliminate representation altogether, focussing instead on units of directly coupled perceptions and actions (Brooks 1991). This in fact transfers state from world models to control. The complex representation processing of traditional AI systems is, in a reactive system, precompiled as an intricate and highly customized program structure. However, this transformation appears to result in computational savings. Reactive robots run considerably more quickly and robustly than their traditional counterparts, using computational devices of much lower power.

The increase of power and the reduction of combinatoric complexity provided by even four or five deictic variables has also been demonstrated (e.g., Chapman 1989; Horswill 1995). Deictic variables allow us to combine the speed and reliability of reactive systems with some of the flexibility of symbol-based AI.

Minimal state implementations of intelligent control are also appealing because they imply correspondingly minimal storage

requirements for episodic memory. Binding data, including reference to the control context, could be the sort of indexical information that is stored in the hippocampal system (McClelland et al. 1995, pp. 451–52). If episodic memory is stored by reference, then remembering is a constructive process of rationalizing sparse information. This could explain many recall effects, such as the suggestibility of witnesses over long periods of coaching or the rapid confabulations of the deranged, without postulating complex representational redescription. For the witness, episodic memory might reference a perceptual routine that changes through learning over time. For the deranged, confabulation from inaccurate pointers may be just as fluent as a normal person engaged in explanation or reminiscence.

The aforementioned "perceptual routine" does not need complex representational manipulations. Work by Tanaka (1996), Perrett (1996), and others indicates that many supposedly high-level cognitive tasks such as recognizing shapes, faces (either general or individual), and even the direction of another's attention may be performed by cells sensitive to higher-order visual features. These cells are ordered topographically with smooth transitions in feature specificity in a way similar to orientation-specific cells in the visual cortices. Perrett's theory also allows an account of the mental rotation task that does not require complex representational transformations over time; consistent with Ballard et al.'s theory, it requires only a single neural pointer moving over precompiled structure.

Surprisingly, Ballard et al.'s discussion of temporal bands (sect. 1) makes no reference to the work of Pöppel (e.g., 1994) and colleagues in this area. Extensive reaction time studies point to a processing window of about 30 msec, the smallest interval in which two stimuli can be temporally ordered; events occurring within the interval are treated as simultaneous. Pöppel suggests that treating stimuli within the window as simultaneous allows the brain to normalize for differing sensory transduction times.

This research may also be relevant to the issue of pointer binding. Forty Hz brain waves have been implicated in perceptual processing (see Phillips & Singer, forthcoming, for a review). This frequency defines a sequence of system states of approximately 30 msec duration. Current neuronal theories of perception (von der Malsburg 1995) use synchronous oscillations to bind features together within system states.

If pointer binding is due to synchronous oscillation, we might also have a more biological explanation than those offered on section 2.2 for the limited number of available pointers. Oscillation rates are highly dependent on the electro-chemical properties of the nervous system. Only a handful of distinct phases within the 40 Hz oscillations can co-exist without mutual interference. This could constitute a neural constraint on the number of pointers simultaneously active.

Ballard et al.'s theory constitutes an advance toward an alternative understanding of intelligence based on immense tables of perceptual and motor skills tightly coupled with functional routines, where coherence emerges through dynamically bound deictic variables. There has long been a debate as to what extent our intelligence is constrained and affected by our biology. Perhaps these are some new answers.

Connecting perception to cognition

R. I. Damper

Cognitive Sciences Centre and Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, England.
rid@ecs.soton.ac.uk www.isis.ecs.soton.ac.uk/

Abstract: Following the "modularity" orthodoxy of some years ago, it has traditionally been assumed that there is a clear and obvious separation between perception and cognition. Close examination of this concept, however, fails to reveal the join. Ballard et al.'s contention that the two "cannot be easily separated" is consistent with nonmodular views of the way that symbol grounding might be achieved in situated systems. Indeed, the traditional separation is viewed as unhelpful.

Although the point is not made a very central part of their article, Ballard et al. (sect. 1) state "cognitive and perceptual processes cannot be easily separated, and in fact are interlocked for reasons of computational economy." Now, some years ago I presented an early version of our work on the computational modelling of categorical perception of stop consonants at a speech conference. (See Damper et al., submitted, for the latest report on this work.) During questioning and discussion, I was surprised to be told by a venerable and respected professor that we had "confused perception and cognition." Strangely, I did not feel confused, even though I had allegedly failed to recognise where perception ended and cognition began and had blurred the apparently important distinction between them.

Taking this criticism seriously as a reflection of my ignorance, I searched through the literature on my return from the conference, looking for enlightenment. At that time, I was able to find only one piece of work which took the division between perception and cognition seriously, rather than just assuming its existence in hand-waving fashion. This was the work of Samuel (1986) in the area of speech perception. He makes the telling point (p. 94) that "what we want to call 'speech perception' is not very well defined. Do we want to include any inferences that the listener draws while listening? What if the inferences are essential to the understanding of the next sentence, or word?" Samuel enumerates three classes of theories of lexical accessing – types I, II, and III – and then attempts to assess them critically. Theories of Types I and II are distinguished by "where we draw the line between perception and what follows (cognition?)," while type III theories involve several interactive levels of representation. In such highly interactive models, "perception and cognition are inseparable" (p. 95). Although the evidence is suggestive rather than conclusive, Samuel's feeling is that models somewhere between types II and III are best supported. The interactive nature of Type III models should be retained, but the lexical level should be considered "the last stage of perception" (p. 109).

Looking at the history of theories of speech perception (see Liberman, 1996, for a comprehensive treatment), the notion of separate perceptual and cognitive processes is seen to have had a profound influence. While the distinction between auditory and phonetic modes of perception is widely accepted (see Sawusch 1986), opinions vary over whether the phonetic component first makes itself felt at the perceptual or the cognitive stage. The latter is called the "horizontal" mode, and (Liberman 1996, p. 3) "assumes a second stage, beyond perception, where the purely auditory percepts are given phonetic names." The former is called the "vertical" mode, in which "there is a distinctly phonetic mode of perception, different in its primary perceptual representations from the auditory mode" (Liberman 1996, p. 307).

Against this background, why would one want to abolish the perceptual/cognitive interface? Of course, in the target article Ballard et al. give reasons based on the computational economy which can be gained from deictic codes. However, Harnad (1992) gives some additional reasons to do with the grounding of symbols in cognition. Considering the physical-symbol system hypothesis of intelligence, he states (p. 80) "any symbol system would require transducers and effectors to interact with the real world," and it is the symbol system itself which "would be doing the real cognitive work." Harnad emphasises that this view is modular – in very much the terms under discussion here – with the symbolic/cognitive module of primary importance while the transducer/perceptual module is "just implementation or I/O." But it is also homuncular, in that mental states are attributed to the symbolic module. By contrast, Harnad promotes a hybrid, nonmodular and nonhomuncular model that "cannot be decomposed into autonomous nonsymbolic components." Thus, symbols and

symbolic capacity are grounded "in the structures and processes that underlie robotic capacity." (See also Clark, 1987, for relevant commentary.)

This is surely entirely consistent with Ballard et al.'s contention (sect. 5, para. 1) that "deictic computation provides a mechanism for representing the essential features that link external sensory data with internal cognitive programs and motor actions." It is interesting, given his status as a pioneer symbolicist, to note that Newell 1990, pp. 159–60) writes: "one thing wrong with most theorizing about cognition is that it does not pay much attention to perception on the one side or motor behavior on the other. It separates these two systems out." Newell accepts that the loss from so doing is "serious – it assures that theories will never cover the complete arc from stimulus to response," but presents it as a regrettable necessity because "the sorts of considerations that go into perception and motor action seem too disparate to integrate." Ballard et al. do the field a service by indicating how this integration might occur. To the extent that separating perception and cognition is unhelpful, "confusing" the two (the long-standing charge against me) seems an entirely reasonable thing to do.

From double-step and colliding saccades to pointing in abstract space: Toward a basis for analogical transfer

Peter F. Dominey

INSERM U94 and CNRS Institut des Sciences Cognitives, 69676 Bron Cedex, France. **dominey@lyon151.inserm.fr**

Abstract: Deictic pointers allow the nervous system to exploit information in a frame that is centered on the object of interest. This processing may take place in visual or haptic space, but the information processing advantages of deictic pointing can also be applied in abstract spaces, providing the basis for analogical transfer. Simulation and behavioral results illustrating this progression from embodiment to abstraction are discussed.

Ballard et al. argue for use of deictic primitives, or pointers to objects in the world, as a means by which the nervous system exploits external structure, extracting data with respect to current points of interest, rather than constructing representations in arbitrary frames. Simulation and experimental work from our laboratory support this position at the embodiment and higher levels, and also suggest how deictic pointers to objects in abstract spaces might provide the basis for analogical reasoning.

Ballard et al. present the oculomotor saccade as a classic example of a deictic pointing mechanism. That the nervous system can prepare more than one saccade during a single reaction time likely reflects the dense temporal organization of visual events in the real world, and highlights the importance of accurate responses to these events. In this "double-step" saccade problem the retinal information that defines the site of the second target with respect to the current eye position or deictic pointer is "invalidated" by the first saccade. The retinal error is effectively an offset defining the movement required to attain a visual target, but the offset is only valid with respect to the pointer, that is the eye position, from which it was specified. Thus, the targeting saccade must start from that position, or the pointer must be updated to account for the intervening change in eye position. Results in "colliding saccade" studies from the Schlag laboratory (Schlag & Schlag-Rey 1990) indicate that this kind of pointer manipulation takes place at a relatively low level, likely in the brainstem. We have recently demonstrated by simulation how such a relatively low level system that compensates for neural transmission delays provides the neural basis for performing double-step saccades, and also explains the colliding saccade phenomena (Dominey et al. 1997a), illustrating an embodiment level computation that assures reliable interaction between the oculomotor sensor-effector system and the environment.

Although pointer updating that compensates for single eye movements may take place at a relatively low level, we (Dominey et al. 1995a) propose that more complex sequential behavior requires deictic binding manipulation at the level of interaction between cortex and basal ganglia, in agreement with the pointer manipulation scheme suggested by Ballard et al. In this model, a recurrent network corresponding to the primate prefrontal cortex encodes sequential state, and these states are bound to behavioral responses via learning-related plasticity in corticostriatal synapses. The recurrent state system thus maintains an ordered set of pointers whose referents are specified via learning in the corticostriatal synapses.

We studied sequence learning tasks in which a series of spatial targets are sequentially presented and must then be sequentially selected, by choice from among all the targets, in the correct order. We can simulate the task in two conditions: one in which the saccade choice to the current sequence element is made from the location of the previous element (deictic condition), and another in which saccades are made from an arbitrary central fixation (arbitrary condition). Simulation results demonstrate that the deictic condition is more effective in terms of number of trials to learn the task. Why? Consider the sequence ABCBDC in which each letter corresponds to a target in space. In the arbitrary condition, this sequence is ambiguous in the sense that not all elements have unique successors, for example B is followed by C and by D. Hence, the sequence cannot be learned as a set of associations, but requires more elaborate context encoding. In the deictic condition, the required context is explicitly provided by visual input. The sequence is executed as transitions or saccades between elements AB BC CB BD DC, and the ambiguity is removed as all of these transitions are unique. Problem complexity is reduced by use of deictic versus global reference, as the necessary context defining the previous sequence element is provided directly by visual input, rather than from memory.

This system is robust in simulating human and nonhuman primate behavior and neurophysiology results (Dominey 1997; Dominey & Boussaoud 1997; Dominey et al. 1995a), but it fails in analogical transfer in sequence learning, in which deictic pointers must refer to objects in an abstract space. If we define the surface structure of a sequence as the serial order of its elements and the abstract structure in terms of relations between repeating elements, then the sequences ABCBAC and DEFEDF have identical abstract structure (123213) and different surface structures, and are thus isomorphic. Humans learn such abstract structure and transfer this knowledge to new, isomorphic sequences (Dominey et al. 1995b; 1997b) displaying a simple form of analogical transfer. The model learns surface structure, but fails to learn abstract structure. We modified the model to represent sequences in terms of abstract rather than surface structure. Using the same cortico-striatal pointer manipulation mechanism with abstract rather than surface structure pointer referents, the modified model now displays human-like performance in learning abstract structure (Dominey et al. 1995b), and thus can provide the basis for analogical transfer.

A central process in analogical reasoning is the identification of structural isomorphisms between source and target objects or problems (Thagard et al. 1990). This requires the identification of structural relations (abstract structure) in the deictic frame of the source object that can then be mapped into the deictic frame of the target problem. Our simulation results indicate that the manipulation of deictic pointers in sensory (e.g., visual or haptic) domains extends naturally into abstract spaces where the referent objects correspond to abstract rather than surface structures, thus providing the basis for analogical transfer.

ACKNOWLEDGMENT

The author is supported by the Fyssen Foundation (Paris).

Deictic codes, embodiment of cognition, and the real world

Julie Epelboim

Center for the Study of Language and Information, Stanford University, Stanford, CA 94305-4115. **yulya@brissun.umd.edu**

Abstract: It is unlikely that Ballard et al.'s embodiment theory has general applicability to cognition because it is based on experiments that neglect the importance of meaning, purpose, and learning in cognitive tasks. Limitations of the theory are illustrated with examples from everyday life and the results of recent experiments using cognitive and visuomotor tasks.

Ballard et al.'s target article proposes that eye movements and other orienting motions are crucial to cognition on the "embodiment level," where they reduce the load on working memory by binding objects to variables in behavioral programs. This ambitious theory, which the authors claim applies to all cognitive processes, is supported empirically only by experiments in which subjects copied meaningless shapes consisting of colored blocks (see sect. 3). This block-copying task, however, is not representative of most cognitive and visuomotor tasks done in everyday life. It lacks both utility and meaning and does not provide an opportunity for learning. Consider the similar, but realistic and useful task of assembling a toy truck. You are given four wheels, two axles, the cab, the chassis, the truck bed, and a picture of the assembled truck. Even an average 3-year-old can assemble this truck without having to refer to the picture.

In more "advanced" toys, components may need assembly; for example, attaching doors and headlights, mounting a steering wheel inside the cab. But even here a picture of the truck may not be needed. Repeated picture-scanning is only required when the model has no meaning or utility, or when the building blocks are small with respect to the model's overall size (e.g., LEGO). Here, learning reduces the need to look at the finished product's picture. In Ballard et al.'s block-copying task, subjects usually stopped looking at the model by the end of each building sequence (sect. 3.2), implying that the meaningless pattern was being learned quickly. I believe that had the same model been built repeatedly, glances at the model would become fewer and would then disappear. Learning effects were not studied. They can, however, be predicted both intuitively and from published experiments using other tasks.

Geometry. Epelboim and Suppes (1997) recorded eye movements of subjects solving geometry problems posed with diagrams. A strong correspondence was observed between the eye movement pattern and the cognitive steps used to solve each problem. This finding supports embodiment theory, because repeated scanning of the diagram could have reduced the load on working memory. Further support comes from the fact that scanning was not limited to features visible in the diagram. The solution of problems usually required the visualization of structures not in the diagram. For example, in one problem, subjects had to connect (mentally) the ends of two intersecting line-segments and realize that the resulting imaginary triangle was equilateral. Expert geometers visualized this triangle and spent appreciable time scanning its interior. Later in the same session they encountered a problem that requiring a similar step, that is, connecting two line-segments to form an equilateral triangle. They started by looking at the line segments as before, but then recognized the similarity between the present to the earlier problem and solved the new problem without scanning the interior of an imagined triangle. So, recent experience with only one similar problem allowed these geometers to encapsulate individual steps into a single cognitive operation. This encapsulation resulted in both a different eye movement pattern and a faster solution than had been possible initially. It could have reduced the load on working memory, thereby reducing the need to scan the diagram. If this were the case, deictic strategies are of only limited use once tasks are learned, and, therefore are less wide-spread than Ballard et al.'s embodiment theory suggests, because most useful tasks in everyday life are not completely novel, or cease to be novel after limited practice.

Tapping target sequences. Learning effects were also observed by Epelboim et al. (1995), who asked subjects to tap sequences of 2, 4, or 6 targets on a work table in front of them. The sequence-order was indicated by the colors of small lights mounted on the targets. Each target configuration was tapped 10 times in a row. Subjects got faster with repetitions because (1) they made fewer eye movements to scan the work table surface and (2) they spent less time looking at each target. However, they always looked at each target just before tapping it, even after they learned the sequence. This "look-before-tap" strategy was used by all subjects, even in 2-target sequences in which memory-load was trivial. They also looked before tapping when they were permitted to tap in a self-selected order. Why? If the purpose of these eye movements was to reduce memory-load (if they were "deictic"), they should be observed less frequently in conditions with lighter memory-load. This was not observed, so the eye movements must have served a different purpose. A more plausible reason for this look-before-tap behavior was that subjects continued to need detailed visual information to guide their taps even after they had learned the targets' locations. Access to such information required bringing each target to the fovea. Two findings support this explanation: (1) subjects could not tap the targets with their eye closed, even after they had learned the pattern, and (2) they tapped slower (50–100 msec/target) when visual input was limited to the light produced by target LEDs.

Conclusion. Ballard et al.'s discussion of embodiment theory does not give sufficient weight to the capacity of humans to learn to organize information into hierarchical structures. This capacity is known to be an effective way of reducing the load on working memory. The applicability of the proposed theory to cognition in general is questionable, and will remain so until this oversight is corrected.

Embodiment is the foundation, not a level

Jerome A. Feldman

Electrical Engineering and Computer Science, University of California at Berkeley; International Computer Science Institute, Berkeley, CA 94704-1198. **jfeldman@icsi.berkeley.edu**

Abstract: Embodiment, the explicit dependence of cognition on the properties of the human body, is the foundation of contemporary cognitive science. Ballard et al.'s target article makes an important contribution to the embodiment story by suggesting how limitations on neural binding ability lead to deictic strategies for many tasks. It also exploits the powerful experimental method of instrumented virtual reality. This commentary suggests some ways in which the target article might be misinterpreted and offers other cautions.

The traditional view of the mind is as a processor of formal symbols that derive their meaning from a model theory, which is in turn mapped onto a world uniquely divided into objects. The focus of the interdisciplinary field called Cognitive Science is to replace this formulation with one where the central idea is an embodied mind in which meaning derives from the sensory, motor, and computational properties of the brain. *BBS* has played a leading role in this development by featuring articles like this one which speculate on how important behaviors might be realized neurally. The target article makes an important contribution to the embodiment story by suggesting how limitations on neural binding ability lead to deictic strategies for many tasks. Unfortunately, Ballard et al. appropriate the term "embodiment" to refer to a hypothesized granularity of action of about ⅓ second in a hierarchy. They also invent names for longer and shorter durations. There are two related problems with this wordsmithing. A minor problem is that readers might get the impression that the authors believe that, at longer time scales, cognition is not embodied and can be treated in

the old abstract way. More important, the idea of embodiment has been treated as the fundamental principle defining contemporary cognitive science for at least a decade (Johnson 1987). This more general usage of embodiment has grown increasingly important (e.g., the 1996 AAAI Fall Symposium on Embodied Cognition and Action, organized by Ballard among others). Embodiment seems the perfect term to characterize the modern view of cognition and it will lose its scientific value if people start using the word promiscuously.

One of the beauties of the target article is the methodology of instrumenting a virtual reality task, here block copying. Ballard et al. have exploited only part of the potential of their paradigm and some of their general conclusions seem too dependent on the fact that vision was very cheap. Imagine a variant where there was a (virtual) sliding door over the model which had to be held open (by the mouse) for visibility. This would significantly increase the cost of vision and people would switch to more memory-intensive strategies, as already happened in the large visual angle case. They would then almost certainly notice changes in the salient parts of the model. More generally, the article understates the use of non-deictic representations and problem solving strategies, typically relegating them to footnotes. The results from an example constructed to minimize the need for models hardly seems adequate to justify the general conclusion that human vision "may only create perceptual descriptions that are relevant to the current task" (sect. 5, para. 8). What about long-term visual memory? Ballard et al. apparently did not ask the subjects what they remembered about the experience, but we can be quite sure that it was not a series of saccades.

The rhythm of the eyes: Overt and covert attentional pointing

John M. Findlay, Valerie Brown, and Iain D. Gilchrist

Centre for Vision and Visual Cognition, Department of Psychology, University of Durham, Durham, DH1 3LE, England.

j.m.findlay@durham.ac.uk

www.dur.ac.uk/~dps0www2/cvvchomepage.html

Abstract: This commentary centres around the system of human visual attention. Although generally supportive of the position advocated in the target article, we suggest that the detailed account overestimates the capacities of active human vision. Limitations of peripheral search and saccadic accuracy are discussed in relation to the division of labour between covert and overt attentional processes.

Ballard et al. have argued convincingly that theories of cognitive function must be grounded in primitives which are related to sensorimotor neurobiology. They illustrate these principles in a novel and fascinating way using the example of visual attention and gaze control. This commentary will address the question of whether the known properties of the visual attention systems support their position.

The target article notes the extreme readiness with which the mobility of the eyes is brought into play during visual tasks. Several studies from our laboratory confirm this characteristic of visual behaviour. Walker-Smith et al. (1977) recorded eye scanning during tasks involving comparison of two nonfamiliar face photographs presented simultaneously. The observers scanned rapidly and frequently between the two faces, in a manner which was at times suggestive of a region by region comparison. A recent research programme has used a controlled search task in which a subject is presented with a display of items, all equidistant from fixation, and the task is to pick out a target from distractors (Brown et al. 1996; Findlay 1995; 1997). In this task also subjects move their eyes with extreme readiness even when, as discussed below, this is not the most efficient search strategy.

In these experiments we examined how information from peripheral vision might guide eye movements. In section 4.1.2 of the target article, this process is termed the "location" routine and a suggestion is made that it might "match a given set of model features with image features at all possible retinal locations." This suggestion presents a far too optimistic view of the capacities of active vision. Classical search theory (Treisman & Gelade 1980) requires that a sequential, attention demanding set of operations is needed for such a feature conjunction search. Only simple feature searches, such as the one involved in the article example of finding a block of a particular colour, can be carried out in parallel. Our work (Findlay 1995; 1997) has shown that similar limits occur in the processing which can guide a saccadic eye movement. However we have found somewhat more parallel processing than predicted by classical feature integration theory; for example, subjects can often locate a target defined by a colour shape conjunction within 300 msec from amidst a set of eight items (cf. Pashler 1987).

The rhythm with which the eyes move (between 2 and 4 fixations per second) seems to be a rather basic feature of visual cognition. An intriguing confirmation is provided by the experiment of Vitu et al. (1995) who asked subjects to scan meaningless "z-strings" of letters and found quite similar eye movement patterns to that of reading (see also Epelboim et al. 1994). How intrinsic is this rhythm? We argue next that there are constraints which limit the generation of either a more rapid, or a slower pace of events.

Although there may be physiological limits in the operation of the fixation/move reciprocal inhibition system (Munoz & Wurtz 1993), speed accuracy trade-off provides a plausible reason why saccades are not made at a faster pace. There is some evidence for such a trade-off even with saccades to a single target (Abrams et al. 1989; Kapoula 1984). However, when two or multiple potential targets are present, very clear evidence of such a trade-off appears. Short latency saccades are influenced by global aspects of the stimulation and only by delaying the saccade can accurate saccades be made (Findlay 1981; 1997; Ottes et al. 1985). The assumption implicit in the target article that the eyes can be directed readily and accurately needs considerable qualification.

What about a slower pace for saccadic eye movements, perhaps leaving more of a role for covert attentional shifts? Under some circumstances, it can be advantageous to delay moving the eyes. For example in one of our studies (Brown et al. 1996), the task was to search for a face image amidst a small set of visually similar but distinguishable distractors (inverted faces or scrambled faces) and to direct the gaze to the target. Naive subjects made an initial eye movement with a latency of 200–300 msec following stimulus presentation. These movements went to distractors with the same probability as to the target and so did not aid the task. With training however, subjects learned to forestall this compulsion to move the eyes and to generate a single direct saccade to the target following a longer delay (500–700 msec).

One account of the finding might be that the target is located with a covert attention scan before the eyes were moved (although benefits would also be expected with a longer fixation period if processing of all locations occurred in parallel). Why do the eyes seem so ready to move when there is a covert attentional system available? Our tentative answer is that in most naturally occurring situations, the limitations of peripheral vision (acuity reduction, Wertheim 1894; lateral masking, Toet & Levi 1992) preclude locating targets in the periphery with covert attention. In our search tasks, we enhanced the availability of information from the visual periphery by restricting displays to well segmented objects all equidistant from the eyes. Such artificial tasks are of low ecological validity and it may be that under most natural viewing conditions the use of covert attention shifts will not be worth the effort. This could explain why the eyes themselves seem to be the preferred deictic pointers.

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

A reader's point of view on looking

Martin H. Fischer

Institut für Psychologie, University of Munich, D-80802 Munich, Germany.
mfischer@mip.paed.uni-muenchen.de

Abstract: Questions about the validity of eye fixations in the blocks task as a memory indicator are discussed. Examples from reading research illustrate the influence of extraneous factors.

Ballard et al. interpret frequent eye fixations on the model area in their blocks task as a strong indication of memory updating. Their evidence, however, does not clearly demonstrate this updating role for eye fixations. Continuous monitoring of eye position necessarily yields large numbers of registered fixations. We need to know how often participants simply looked straight ahead (not fixating either zone) when the task-relevant areas were in the peripheral visual field, because the reduced number of fixations on the model in Ballard et al.'s "peripheral" condition could reflect the fact that the model was no longer located at a preferred resting position for the eyes. Similarly, fewer fixations per block in the "monochrome" condition might only reflect the overall shorter duration of an easier task. Before one interprets fixations as memory pointers, similar relative frequencies of fixations and transition patterns should be obtained for spatial permutations of the three areas.

Reading research shows that eye fixations are affected by extraneous limitations, such as the maximum speed of articulation. In reading aloud, long eye fixations occur because the voice must eventually catch up with the eyes (e.g., Levin 1979). Similarly, even the maximum speed of arm movements will leave ample time for nonfunctional eye fixations in the blocks task. Related to this, moving one's arms under speed instructions induces accompanying eye fixations to insure motor accuracy (e.g., Abrams 1992). To test whether fixations reflect memory limitations or accuracy demands, Ballard et al. might replicate the task with slow movements. Deictic coding predicts more fixations owing to pointer decay, and the accuracy hypothesis predicts fewer fixations than with fast movements.

Ballard et al. acknowledge the problem of how goal-directed saccades are programmed and why they are so accurate if memory is so poor. I have investigated memory for word locations to address the same issue in reading. The spatial arrangement of words was manipulated during sentence presentation and readers were asked to locate one target word from each previously read sentence using a mouse cursor (Fischer 1997). Surprisingly, location memory was limited to the most recent one or two words, and other locations were reconstructed from item memory. Nevertheless, readers make accurate long-range eye regressions to resolve syntactic or anaphoric difficulties (Rayner & Pollatsek 1989). The conflict can be resolved by considering the different time delays for eye regressions and for mouse cursor positioning. This observation demonstrates that different processing strategies are normally used when a task is performed on the "cognitive" and the "embodiment" scale (see Table 1 of the target article). In general, when a strategy applies across time domains, then it was probably induced by extraneous factors.

There is doing with and without knowing, at any rate, and at any level

Joaquín M. Fuster

Brain Research Institute, School of Medicine, University of California, Los Angeles, Los Angeles, CA 90024. **joaquinf@ucla.edu**

Abstract: Ballard et al.'s is a plausible and useful model. Leaving aside some unnecessary constraints, the model would probably be valid through a wider gamut of interactions between the self and the environment than the authors envision.

Ballard et al.'s general view of the interactions between an organism and the world around it is compatible with my own view of the perception–action cycle and its role in sequential goal-directed behavior (Fuster 1995): a functional hierarchy of sensory-motor circuits linking the environment to a corresponding hierarchy of sensory and motor structures that extends from the spinal cord to the neocortex. This organization allows the concomitant sensory-motor integration at several levels of abstraction, from the lowest to the highest. At the top of that neural hierarchy, the posterior association cortex and the prefrontal cortex serve the sensory-motor integration in the making of novel and extended sequences of behavior with overarching goals: the plans and programs. At the bottom, simple circuits integrate simple reactions and reflexes (primitives), the microcomponents of those plans and programs. The higher the level of integration, the more necessary it becomes to mediate cross-temporal contingencies. This necessity is maximal in complex sequences of behavior, speech included, where working memory and preparatory set, two basic functions of the dorsolateral prefrontal cortex, come into play.

Ballard et al. select an intermediate level of the hierarchy, the "embodiment level," and apply a plausible computational model to its operations. Their deictic symbolic codes, their pointers, guide the instantiations of the perception–action cycle at that intermediate level, as exemplified in performance of the blocks task. The authors base their rationale, as well as the parameters of sensory-motor integration and of their deictic pointing device, on the frequency of approximately three per second, the observable frequency of spontaneous scanning saccades. With this and other constraints, their model appears to do a good job in simulating computations at some unspecified neural level appropriate to "embodiment." It is possible, however, that by extending temporal parameters, adding working memory, and bringing in cortical structures, especially the prefrontal cortex, the model would apply also to higher levels of the hierarchy.

As Ballard et al. recognize, the generality of the model is uncertain; it is unclear to what extent it is suited to other hierarchical levels. However, the imposition of the time constraints seems somewhat unnecessary and ties the model to a restricted level of analysis. By use of arbitrary and more flexible temporal parameters, the model might apply to a wide range of integrative phenomena at multiple hierarchical levels. An even more serious constraint appears to be the sequential operation of the assumed deictic programs. This mode of operation seems to preclude computations at several hierarchical levels simultaneously, thus depriving the system of economy and flexibility. Yet the organism is capable of simultaneous computations at several cognitive levels. The assumption of serial processing is especially appropriate to selective attention and conscious behavior, presumed above the "embodiment level." Nonetheless, both serial and parallel processing support practically all cognitive operations in the interplay of the organism with its environment. Further, the separation between a level of cognitive awareness and the "embodiment level" appears somewhat contrived. In any event, it is reasonable to suppose that every level is endowed with its system of references, "pointers," symbols, and cognitive representations, and that it operates in its own particular time scale. Thus, what goes on at the "embodiment level" may be paradigmatic of what goes on at every other level.

In conclusion, the narrow choice of the "embodiment level," the assumption of serial processing, and the time constraints of that processing, may be unwarranted and too restrictive. They deprive the proposed model of the generality it probably has and of its potential value for explaining the functions of the neural structures involved in the perception-action cycle, which operates at multiple levels simultaneously, in and out of consciousness or the focus of attention.

Deictic codes for embodied language

Arthur M. Glenberg

Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706. **glenberg@facstaff.wisc.edu**

Abstract: Ballard et al. claim that fixations bind variables to cognitive pointers. I comment on three aspects of this claim: (1) its contribution to the interpretation of indexical language; (2) empirical support for the use of very few deictic pointers; (3) nonetheless, abstract pointers cannot be taken as prototypical cognitive representations.

A major claim made by Ballard et al. is that eye fixations bind variables to deictic pointers. The referent of the pointer – its current meaning, if you will – corresponds to task-relevant components of the object being fixated. I explore three aspects of this claim. First, deictic pointers give a partial understanding of how we interpret indexicals and gesture in language. Second, the computational efficiency of using few pointers is supported by experiments probing the structure of spatial mental models. Third, a cautionary note: abstract deictic pointers useful for process control should not be taken as prototypical of knowledge representation. Instead, the task-relevant representations that result from the binding are consonant with embodied, nonabstract accounts of language, memory, and meaning (Glenberg 1997).

Language understanding requires the interpretation of indexicals, such as "I," "it," "here," and "now." That is, these words must take a good deal of their meaning from the context. But even common nouns often require reference to a situation for correct interpretation. Thus, the interpretation of "the cow" (as a particular cow), "my pencil," and "insert tab *a* into slot *b*" (what is the orientation?) are indexical. Getting the right referent for such terms is often accomplished by pointing with the hand, the body (turning), the head, or the eyes (Clark 1996). All of these pointing devices get listeners to literally look at the object, and thus, in Ballard et al.'s terminology, to bind the object to deictic pointers or codes. The proposed characteristics of deictic codes have testable implications for language understanding. For example, deictic codes do not make all features of the object available, only those that are task relevant. Thus, mentioning "my pencil" and glancing toward it in the context of describing "color" makes available different information compared to, for instance, the word and glance when the context is "pointiness."

The efficiency and number of deictic pointers also have implications for work in language comprehension. Ballard et al. note that "The dynamic nature of the referent also captures the agent's momentary intentions. In contrast, a nondeictic system might construct a representation of all the positions and properties of a set of objects" (sect. 1.1); an image or scene can be conceptualized in a tremendous variety of ways, whereas task-dependent computation is an "economical way of computing . . . just that required by the task demands as they become known" (sect. 4). In section 5.3, Ballard et al. note that deictic codes challenge "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." These claims contrast with once-common assumptions about the nature of spatial mental models derived from text, namely, that these models are relatively complete spatial layouts (Glenberg et al. 1994; Mani & Johnson-Laird 1982; Rinck & Bower 1995). Two recent investigations of the properties of spatial mental models reveal that they may have properties similar to those noted by Ballard et al., namely, that they make use of few codes. Participants in Langston et al.'s (in press) study read or listened to short texts describing spatial layouts of 3–6 objects. Schematically a text might be, "A is to the left of B. C is to the left/right of B." By locating object C at different points but always relative to object B, Langston et al. varied whether object C was (in the reader's putative model) near or far from object A. Contrary to what would be expected from a full three-dimensional model, Langston et al. found no effects of "distance" between object C and object A on the availability of object A. This is just what would be expected if readers used a limited number of deictic pointers (sect. 2.2) that coded the relation between the observer and the object rather than spatial relations between all objects. Rinck et al. (1997), using a different procedure, also concluded that spatial models are not full, three-dimensional representations. Thus, in language comprehension as in perception, we seem to be able to get by with representing information as it is needed, rather than building complete models. This outcome is to be expected if language comprehension builds on other components of cognition (Glenberg 1997).

The third point of this commentary concerns the general role of abstract codes in cognition. Ballard et al. note the usefulness of abstract deictic codes in process control. Does that warrant an extension to the claim that all knowledge can be modeled as abstract codes, such as is common with semantic networks and semantic feature theories? The answer is no. There are severe problems associated with assuming abstract codes for all knowledge, including the symbol grounding problem (Harnad 1990), linguistic vagary (Barsalou 1993), and the extraordinary dependence of meaning on context (Shanon 1988). Although Ballard et al.'s pointers are abstract, binding creates representations that are specific to the current situation, not abstract. As Ballard et al. write, "visual cortical areas may in fact be computing highly specific task-relevant representations" (sect. 5.5). Thus, thought and action are likely to be controlled by the details of the situation at hand, not abstract semantic codes.

"Highly specific, task-relevant representations" are consistent with several components of my account of embodied memory (Glenberg 1997). For example, my "clamping," much like the binding of variables to deictic pointers, is claimed to result in a task-relevant representation of the current situation; but for me, task relevance is in terms of the actions afforded by the environment, given the body of the perceiver. These action affordances are meshed (combined) with actions from memory and language, to complete the conceptualization of the situation. Thus, when a tired person looks at a chair, he notes that it affords sitting. But, his memory of having seen someone else sit in the chair recently makes it that person's chair, and blocks his sitting. This meshing of actions from the clamped environment with actions from memory is possible because of the shared medium: actions for the body. When the chair's former occupant says, "Please sit here; I was just leaving," the conceptualization of the chair changes again. Interpretation of the indexicals "here" and "I" makes use of the deictic pointing system. The chair ("here") is clamped (bound to the pointer), and its action-based interpretation is changed by meshing it with the action-based interpretation of the sentence: the former occupant is leaving, making the action of sitting available again. Thus, task-relevant representations resulting from the binding of variables to deictic codes become a vehicle for interpreting language in context and for controlling action (e.g., sitting in the chair).[1]

NOTE

1. Requests for reprints may be directed to Arthur Glenberg, Department of Psychology, 1202 West Johnson Street, Madison, WI 53706, or glenberg@facstaff.wisc.edu.

Pointing the way to a unified theory of action and perception

Mel Goodale

Department of Psychology and Graduate Program in Neuroscience, University of Western Ontario, London, Ontario N6A 5C2, Canada. **goodale@uwo.ca www.uwo.ca/neuro/faculty/goodal.html**

Abstract: Deictic coding offers a useful model for understanding the interactions between the dorsal and ventral streams of visual processing in the cerebral cortex. By extending Ballard et al.'s ideas on teleassistance, I show how dedicated low-level visuomotor processes in the dorsal stream

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

might be engaged for the services of high-level cognitive operations in the ventral stream.

Ballard et al. make a persuasive argument that orienting movements, particularly saccadic eye movements, play a crucial role in cognition, providing a computational platform for cognitive operations. This is an idea with considerable merit. It joins nicely the fields of motor control and cognition – fields which too often are treated as separate realms of investigation. It also makes evolutionary sense by suggesting that mechanisms which evolved for the distal control of movement might have been co-opted (in both a literal and figural sense) for the computations underlying the cognitive life of the animal.

There are a number of other reasons to welcome this approach to understanding cognitive operations. For one thing, it blurs the distinction between perception and cognition. In contrast to the ideas of reconstructive theorists, Ballard et al.'s deictic coding approach does not require elaborate three-dimensional models of visual scenes to provide the raw material for cognitive operations. In deictic coding, a pointing action flags an object in the scene for cognitive operations. In this way, deictic coding is much more in tune with "active vision" ideas which emphasize the synergy between the world and the observer's actions. The need for elaborate visual representations is minimized.

Deictic coding also speaks to the issue of scene integration – how information is integrated and stored over successive fixations. Ballard et al.'s experiments suggest that the stored information about a scene can be quite sketchy. It would appear that long-term memory about the world, rather than the integration of detailed information across successive fixations, determines how we deal with the scene before us. A particular object in the scene is referenced when needed by marking it with fixation.

I am least happy with the way in which Ballard and his colleagues attempt to situate their model in the brain – not because I think what they say is necessarily wrong but because I think they don't really spell out exactly what it is they are saying. They flirt with the idea of mapping deictic coding and perceptual/cognitive functions (at least in the visual domain) onto the dorsal and ventral streams of processing that have been identified in the cortical visual pathways. They point out the computational utility of separating identification from localization, but stop short of accepting the division of labour originally proposed by Ungerleider and Mishkin (1982) for the ventral and dorsal streams – the "what" versus "where" hypothesis. Instead, Ballard et al. appear to opt for the distinction put forward more recently by David Milner and me (Goodale & Milner 1992; Milner & Goodale 1995). In our reformulation of the Ungerleider and Mishkin story, both streams are seen as processing information about the orientation, size, and shape of objects, and about their spatial relations. Each stream, however, deals with incoming visual information in different ways. Transformations carried out in the ventral stream permit the formation of perceptual-cognitive representations which embody the enduring characteristics of objects and their spatial relations with each other; those carried out in the dorsal stream, which utilize moment-to-moment information about the disposition of objects within egocentric frames of reference, mediate the control of a number of different goal-directed actions.

I said that Ballard et al. *appear* to opt for the Goodale and Milner story. In fact, they say that our account fits with theirs but they don't really say why. They spend quite a bit of time discussing the possible constraints on any neural implementation of their model, but their discussion of the role of the dorsal and ventral streams, primary visual cortex, the basal ganglia, and a number of other brain structures is quite free ranging and, in the end, noncommittal. They can scarcely be faulted for this however. Ballard and his colleagues are still developing their ideas about the ways in which cognitive operations are embodied in deictic codes. No one should expect them to offer more than a few tentative suggestions about implementation in brain. But it is hard to resist the impulse to embed the model in neural hardware. For what it's worth, here is my attempt to meld deictic codes with my own ideas about the division of labour in the dorsal and ventral streams.

Some clues as to how that implementation might work can be found in Ballard et al.'s invocation of teleassistance. In teleassistance, a human operator uses a deictic sign language to communicate with a robot that actually performs the required motor act on the marked goal object. Ballard et al. suggest that one can think of the human operator as representing control at the level of the cerebral cortex and midbrain, and the robot as representing control at the level of the spinal cord. Although I like the analogy of teleassistance, I see the distinction between the human operator and the robot, in the domain of vision at any rate, not as one between cortex and spinal cord but as one between the ventral and dorsal streams. The perceptual-cognitive systems in the ventral stream, like the human operator in teleassistance, move the deictic pointer onto different objects in the scene. When a particular goal object is flagged, dedicated visuomotor networks in the dorsal stream (in conjunction with related circuits in premotor cortex, basal ganglia, and brainstem) are activated to perform the desired motor act. The important point here is that the visuomotor networks in the dorsal stream not only mediate the eye movements involved in deictic coding but they also mediate the visually guided actions that are directed at the flagged goal object. The reference frames in the dorsal stream, as Ballard et al. point out, are egocentric while those in the ventral stream are relative or allocentric. This means that a flagged (foveated) object in the scene can be processed in parallel by both ventral and dorsal stream mechanisms – each transforming information on the retinal array for different purposes. The ventral stream uses largely foveal information for cognitive processing in goal selection while the dorsal stream uses both foveal and peripheral information for programming and guiding the goal-directed motor act.

The reader will have noticed that there are some glosses in this account. It is not clear, for example, how the ventral and dorsal streams interact in moving the deictic pointer (i.e., making a saccadic eye movement) from one location to another in the scene. Presumably, the saccades can be both "endogenously" and "exogenously" driven, in the sense that cognitive information about the scene or highly salient objects in the scene could both drive the pointer. Perhaps shared control by ventral and dorsal mechanisms (via, for example, frontal eye fields) could drive the endogenously driven saccades and collicular-dorsal stream mechanisms could drive the more exogenously elicited saccades. But of course, this is pure speculation. In any case, I believe that Ballard and his colleagues have given us a model that shows how low-level motor processes can be engaged for the services of high-level cognitive operations – a model that in the visual domain fits rather well with what we know about the functional characteristics of the two major streams of visual processing in cerebral cortex. As such, the model also points the way to understanding how the two streams of visual processing might interact.

Spatial perception is contextualized by actual and intended deictic codes

J. Scott Jordan

Department of Psychology, Saint Xavier University, Chicago, IL 60655.
jordan@sxu.edu

Abstract: Ballard et al. model eye position as a deictic pointer for spatial perception. Evidence from research on gaze control indicates, however, that shifts in actual eye position are neither necessary nor sufficient to produce shifts in spatial perception. Deictic context is instead provided by the interaction between two deictic pointers; one representing actual eye position, and the other, intended eye position.

Ballard et al. present evidence indicating that the spatial context for perception is provided by deictic pointers such as eye position.

Gaze control data, however, indicate that such physical pointing constitutes neither a necessary nor a sufficient condition for the presence of deictic perception.

Stevens et al. (1976), for example, demonstrated that shifts in deictic pointers do not necessarily require shifts in eye position. Their subjects made judgments about the perceived egocentric location of stationary objects under controlled experimental conditions involving total extraocular paralysis. As subjects attempted to look in different directions, the perceived location of objects in the visual world shifted (their deictic-pointer changed), apparently all at once, in the direction of the intended eye-movement, despite the fact that the eyes (i.e., the physical deictic pointers) did not move.

Matin et al. (1966) discovered that shifts in eye position are not sufficient to bring about shifts in deictic pointers. Specifically, their subjects made judgments about the location of a 6 msec flash of light that was presented as other subjects fixated the remembered location of a previously lit fixation target. Matin et al. found that the perceived location of the test flash varied inversely with changes in eye position. If actual eye position were sufficient to bring about shifts in deictic pointers, the described changes in eye position would have shifted the pointer values, and spatial constancy would have prevailed.

These experiments indicate that the deictic context provided by the gaze-control system reflects more than just actual eye position. Rather, it reflects an interaction between two deictic pointers, one representing actual eye position, the other representing intended eye position. One can test this notion by producing saccadic eye-movements in the dark across a point-light source blinking on and off at 120 Hz. One sees an array of flashes similar to that depicted in Figure 1. The first flash is displaced in the direction of the intended saccade, and subsequent flashes appear in locations closer and closer to the location of the blinking light (Hershberger & Jordan 1996). The first flash in the array appears in its displaced location roughly 80 msec prior to the onset of the saccade (Jordan & Hershberger 1994). Given that the spatial extension of the array is brought about by the sweeping motion of the retina, the perceived location of each flash appears to be the product of (1) a deictic pointer representing intended eye position, (2) a deictic pointer representing actual eye position, and (3) the retinal locus struck by the flash. In other words, the perceived location of the flashes occurs within the context provided by the moment-to-moment discrepancy existing between intended and actual deictic pointers.

This partitioning of deictic codes into actual and intended pointers in no way detracts from Ballard et al.'s deictic model. On the contrary, it allows one to model deictic pointers within a context of anticipatory, on-line control. Actual pointer values are continuously driven into correspondence with intended pointer values. Modeling deictic codes in this way will prove important when attempting to apply deictic-coding models to continuous movements such as oculomotor smooth pursuit or movement of the body as a whole, because such movements require continuous changes in deictic pointer values if spatial constancy is to be maintained.

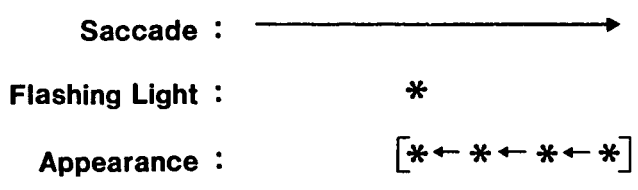

Figure 1 (Jordan). If you shift your gaze saccadically from the left to the right of a point light source blinking on and off at 120 Hz in an otherwise darkened room, you will see phi movement to the left within a phantom array that is displaced to the right (Hershberger 1987, reprinted by permission).

Evidence of such dynamic deictic codes can be found in Melzack's (1992) work on phantom limbs. Those who have lost a limb continue to experience the missing limb in spatial locations that are often consistent with on-going behavior. This indicates a rather permanent de-coupling of intended and actual deictic pointers: the deictic code for the intended location of the effector is present, but feedback regarding the effector's actual location, is not. It is telling that what one experiences about the missing limb is its intended location in space-time. The limb does not exist, yet one experiences it anyway, in coherent, anticipatory space-time locations. Elsewhere I have argued (Jordan 1997) that it is the control of these anticipatory "feels" of the body in space-time, not actual limb position, that constitutes what we refer to as volitional action. In this light, volitional action might be thought of as the control of anticipatory deictic codes.

On learning and shift (in)variance of pattern recognition across the visual field

Martin Jüttner

Institut für Medizinische Psychologie, Universität München, D-80336 Munich, Germany. **martin@imp.med.uni-muenchen.de**

Abstract: Ballard et al.'s principle of deictic coding as exemplified in the analysis of fixation patterns relies on a functional dichotomy between foveal and extrafoveal vision based on the well-known dependency of spatial resolution on eccentricity. Experimental evidence suggests that for processes of pattern learning and recognition such a dichotomy may be less warranted because its manifestation depends on the learning state of the observer. This finding calls for an explicit consideration of learning mechanisms within deictic coding schemes.

Scientific reasoning in general, and theories concerning the functioning of the brain in particular, rely on intended simplifications. Such simplifying assumptions are necessary because they allow one to reduce a given problem to a form where one can hope to find some analytic or at least computational solution which may then serve as a platform for further theoretical elaboration. The difficulty with this procedure is ensuring that nothing essential has been missed in these initial simplifications, and that the derived solution may therefore be regarded as a zero-order approximation to physical reality.

Ballard et al. adopt the basic principle of deictic coding from the field of computer science where pointers provide a standard technique for the manipulation of complex data structures. Their primary application of this principle to sensory processing is in the sequences of fixations and intervening saccades which accompany our visual exploration of the world. Here fixational periods are associated with a process of identification restricted to the foveal part of the visual field. Extrafoveal vision, in contrast, is thought to provide the context for detecting and localizing the target for the next saccade in the oculomotor sequence. It is this functional dichotomy between foveal and extrafoveal vision which makes up the simplifying assumption that allows the authors to establish a straight correspondence to their deictic coding scheme.

Underlying their view is the fact that spatial resolution in the fovea is an order of magnitude better than in the periphery, making the former "ideal as a pointing device to denote the relevant parts of the visible environment" (Ballard et al., sect. 2.1). However, characterizing visual performance in terms of spatial resolution is necessarily incomplete: it lacks the essential aspect of *form* as the basis of pattern recognition proper. From a paradigmatic viewpoint, this reduced concept of visual processing can be related to detection or discrimination tasks which are intrinsically characterized by a one-dimensional processing of stimulus information. By contrast, cognitive science has traditionally preferred to define pattern recognition as the ability to assign perceived objects to previously acquired categorical concepts (see, e.g., Bruner 1957; Rosch 1978). Such classifications in general require

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

simultaneous processing of stimulus information along multiple stimulus dimensions (Watanabe 1985). Ballard et al. tend to avoid this cognitive aspect of vision: for example, in the block-copying task the blocks were arranged in a random way, avoiding the emergence of form, or "gestalt" properties. Thus the perceptual process is forced to result in a series of *detections* concerning local color attributes. This raises the question of the extent to which the assumed functional dichotomy between foveal and extrafoveal vision persists if processes of pattern recognition become involved.

The distinction between tasks of discrimination or detection on the one hand and classification on the other seems to be more than merely an epistemological matter. Evidence comes from studies on discrimination and classification learning of grey level images (compound-sinewave gratings) across the visual field. As we had demonstrated earlier (Jüttner & Rentschler 1996), extrafoveally derived representations of class concepts are generally characterized by reduced perceptual dimensionality whereas foveally acquired representations are not. Recently, we compared performance in classification and discrimination learning in foveal and extrafoveal vision for the same set of stimuli (Jüttner & Rentschler, submitted). Whereas for a foveal presentation of the stimuli the learning speed for the two types of tasks was found to be equal, there was a clear dissociation in extrafoveal presentation. Learning duration for pattern classification now increased by a factor of five relative to the foveal condition, whereas it remained unaffected for pattern discrimination. Such a divergence suggests that internal representations underlying pattern classification and discrimination arise at distinct cortical levels in the brain and that the former are normally developed in an extremely narrow visual field limited to the fovea.

Hence, with respect to the *learning* of categorical concepts for recognition, Ballard et al.'s assumed primacy of foveal vision appears to receive additional empirical support. However, concerning the aspect of *spatial generalization,* that is, the application of previously learned concepts across the visual field, the situation is different. Using the same paradigm as in the learning studies, we have shown that observers are able to generalize class concepts that they have acquired at one particular retinal location to other retinal sites (Jüttner et al. 1996). The finding of a partial generalization of learned categorical concepts stands in contrast to the failure to obtain shift-invariance concerning the discrimination of random binary patterns (Nazir & O'Regan 1990). However, it is compatible to results concerning character recognition across the visual field (Strasburger et al. 1994; Strasburger & Rentschler 1996). These studies demonstrated that, on the one hand, the recognition field for (numeric) characters clearly dissociates from the field of detection as defined by classical perimetry. On the other hand, for supra-threshold stimuli, the former extends to eccentricities up to 45 deg. If numerals are regarded as examples of familiar pattern classes, such a finding is in line with the idea of shift-invariance for the application of learned categorical concepts. Further support comes from observations concerning the transsaccadic processing of visual information (Jüttner & Röhler 1993; Jüttner 1997). Here it was shown that extrafoveal categorical information at the location of the saccade goal distinctly influences the way postsaccadic foveally obtained stimulus information is interpreted.

In summary, taking into account processes of pattern learning and recognition, the functional dichotomy between foveal and extrafoveal vision assumed by Ballard et al. loses some of its apparent clarity. Whereas such a dichotomy is supported with respect to the learning of categorical concepts, it seems less warranted in the spatial generalization of acquired concepts. Both factors together emphasize the importance of learning mechanisms in human cognition, an issue which is still absent in the current version of the approach. An appropriate extension of the theory would therefore be promising.

Rediscovering Turing's brain

Alex Kirlik

Center for Human-Machine Systems Research, School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205. **kirlik@chmsr.gatech.edu**

Abstract: The embodied AI paradigm represents a distinct shift toward an ecological perspective on intelligent behavior. I outline how Ballard et al. have made a promising advance in expanding the seat of intelligence to include sensory and motor systems, but they have not gone far enough. Sustained growth toward truly generalizable accounts of intelligent systems will also require expanding the locus of intelligence to include the environmental structure participating in intelligent behavior.

> It should be noted that from now on "the system" means not the nervous system but the whole complex of the organism and the environment. Thus, if it should be shown that "the system" has some property, it must not be assumed that this property is attributed to the nervous system: it belongs to the whole; and detailed examination may be necessary to ascertain the contributions of the separate parts.
>
> W. Ross Ashby, 1952

> Control lies not in the brain, but in the animal-environment system.
>
> James J. Gibson, 1979

It is impressive and heartening to see the convergence of the Embodied-AI paradigm, as demonstrated by Ballard and his colleagues, and the ecological or interactionist paradigm, as reflected in the observations of Ashby (1952) and Gibson (1979). According to Ballard and his colleagues, "intelligence has to relate to interactions with the physical world" (sect. 1, para. 1); when we attribute intelligence solely to "the brain's cognitive programs" we make a category error. This agrees with the ecological perspective that attributions of properties such as control, intelligence, and the like are only appropriately applied to the organism–environment system as a functional unit. Ballard et al. take a step in this direction by expanding the basis of intelligence to include not only "cognitive programs" but also the peripheral sensory and motor systems that provide the bridge between cognition and the external world. This systems perspective on intelligence represents a distinct shift toward an ecological or interactionist approach and bodes well for both Embodied-AI and ecological psychology.

Embodiment, however, is still a tenuous concept in current computational modeling, as reflected by Ballard et al.'s own surprising statement that "the tenets of logic and reason . . . demand that intelligence can be described in purely computational terms without recourse to any particular embodiment" (sect. 1, para. 2). To an ecological theorist, however, embodiment is part and parcel of intelligence, for if as Ballard et al. say, "intelligence has to relate to interactions with the physical world," then of course the world itself must participate in the functional description of intelligence. That is, if intelligence is ultimately to be measured in terms of adaptation to an external environment, then the structure of that environment in part determines (and defines) what it means to be an intelligent system. This should not really be a new idea to robotics researchers: the field is characterized by a number of so-called "intelligent" systems that perform well in some specified class of environments but fail miserably in others. Even roboticists would find it difficult to describe a soda can-gathering robot trying to play tennis as demonstrating any meaningful properties of intelligence at all.

Why has including the functional contribution of the external world in the description of intelligence come only so recently to computational modeling? The answer to this question can be found in an early category error in which internal symbol manipulation was designated as the sole seat of intelligence.

Turing's brain. Hutchins (1995a) has provided the best (to my knowledge) historical analysis of the early misconception that gave

rise to cognitive science's preoccupation with internal symbol manipulation as the sole seat of intelligent behavior, and my treatment will parallel his. Hutchins cites Dennett (1991) to explain the origin of the symbol processing view of mind in Alan Turing's introspections of how he went about a mathematical proof:

> "What do I do," he must have asked himself, "when I perform a computation? Well, first I ask myself which rule applies, and then I apply the rule, and then write down the result, and then I ask myself what to do next, and . . . ." (Dennett, 1991, cited in Hutchins, 1995a, p. 361)

The symbol processing model of mind originated in the goal of replicating within a computer the manual symbol manipulation activities performed by a person, in tasks such as theorem proving. But something was lost in the translation, as Hutchins notes:

> Originally, the model cognitive system was a person actually doing the manipulation of the symbols with his or her hands and eyes. The mathematician or logician was visually and manually interacting with a material world. . . . The properties of the human in interaction with the symbols produce some kind of computation. But that does not mean that computation is happening inside the person's head. (1995a, p. 361)

Clearly, something is happening inside a person's head while performing a computation with pencil and paper, and this "something" may well involve the use of internal representation. But it would be a category mistake to claim that this "something" is identical to the computation that is actually taking place in the entire visual, cognitive, motor, pencil, paper, and symbol system (i.e., Ashby's "whole complex of the organism and environment"). Cognitive science originally made this mistake, however, in its failure to follow Ashby's dictum that "if it is shown that the 'system' has some property, it must not be assumed that this property is attributed to the nervous system." As a result, computational modeling has a history of trying to describe solely internally what is typically done in dynamic interaction with the external world.

Ballard and his colleagues are rediscovering the actual system for which the symbol processing model was invented: cognition, eye, hand, and world. In doing the "detailed examination," Ashby called for a correct assessment of the contributions of these separate components to the overall function of intelligence. Ballard et al. are rediscovering the original function of Turing's brain in partialling out the parts of the overall computation originally done by Turing's hands, eyes, pencil, and paper. In pushing this line of research further, they and others will no doubt discover a wide variety of additional environmental structure routinely participating in the production of intelligent behavior. In doing so, they should give as much attention to building generalizable models of the external world as they give to building generalizable models of internal processing. Only research systematically addressing both these contributions to intelligent behavior, and their interaction, will bring forth a truly general understanding of intelligent systems.

Beyond embodiment: Cognition as interactive skill

Paul P. Maglio

IBM Almaden Research Center, San Jose, CA 95120.
pmaglio@almaden.ibm.com www.almaden.ibm.com/cs/people/pmaglio

Abstract: The target article makes a compelling case for the idea that agents rely on the world for external memory in fast-paced perceptual tasks. As I argue, however, agents also rely on the external environment for computational hardware that helps to keep cognitive computations tractable. Hence the external world provides not only memory for computations involving perceptual system actions, but it provides domain-level actions that figure in cognitive computations as well.

My only gripe with Ballard et al. is that they do not go far enough in blurring the traditional boundary between processing that occurs inside and outside the head. In this short space, I will argue that in cognitive computation, the external environment functions not merely as external memory – providing pointer referents for internal computation – but also as critical computational hardware that people rely on to help keep cognitive computations simple. People routinely take external, domain-level actions for their informational effects and to keep computational costs low.

Data cited in the target article support the claim that agents use the world as a kind of short-term memory for tasks involving a high degree of visual processing (e.g., Ballard, et al. 1995; O'Regan 1992), but I think this is only part of the story. To see why, consider how people get better at Tetris (Kirsh & Maglio 1994; Maglio 1995; Maglio & Kirsh 1996). In playing this videogame, people maneuver shapes that fall from the top of the computer screen and land in specific arrangements. To score points and to delay the game's end, the player must successfully place pieces on the pile of shapes that have already landed. The shapes are rotated and laterally moved so their fall controls landing orientation and position. The rate at which shapes fall speeds up over time, so players must act ever more quickly to keep pace. Thus, as expected, in our Tetris studies we found that improving means getting faster. But we did not expect to find that improving sometimes means backtracking more in the task environment (Maglio 1995; Maglio & Kirsh 1996). We discovered that as players get better, they regularly rotate falling shapes beyond their final orientation; hence they must undo (or backtrack over) these extra rotations. Why do skilled players display such apparently errorful behavior? By investigating the computational demands of several models of Tetris expertise, I found that even for a skilled perception model of expertise (e.g., Chase & Simon 1973), backtracking is adaptive because it can help constrain the cognitive and perceptual problems that need to be solved (Maglio 1995). The perceptual computation – matching patterns of falling shapes to patterns of shapes that have already landed – is done more efficiently by serial search than by fully parallel pattern recognition. In particular, it turns out to be computationally easier to match falling shapes one orientation at a time than to match them in all orientations at once.

I think the Tetris findings cohere with Ballard et al.'s results. One consequence of our shared view is that serial processing (e.g., interposing eye movements or external rotations between internal computational steps) is computationally more efficient than parallel processing (taking in all the information at once and calculating a plan) because of the high cost of internal storage and use of partial results. Basically, agents use external actions to save internal computational resources. As mentioned, however, the idea that agents can rely on the world to provide an external short-term memory is only part of what is going on. Eye movements are active from the point of view of the visual system – they change focus and act on the agent's perceptual input – but they are passive from the perspective of the external task environment. Extra rotations made by skilled Tetris players, however, are active in the task environment, though they change the perceptual input in much the same way that eye movements do. Thus, extra rotations change the stimulus but are done for their computational effect. In this case, the world functions not as a passive memory buffer – holding information to be picked up by looking – but the agent in interaction with the world functions as a working memory system, that is, as an *interactive visuospatial sketchpad.* The task environment itself provides part of the computational mechanism agents use to solve a perceptual problem, namely a rotation operation.

It is no surprise that people routinely offload symbolic computation (e.g., preferring paper and pencil to mental arithmetic, Hitch 1978), but it is surprising to discover that people routinely offload perceptual computation as well. Tetris players set up their external environments to facilitate perceptual processing much as gin rummy players physically organize the cards they have been dealt (Kirsh 1995), and much as airline pilots place external markers to help maintain appropriate speed and flap settings (Hutchins 1995b). For Tetris, setting up the environment occurs on a faster

time scale, suggesting a very tight coupling between perception and task-level information-gathering actions (Kirsh & Maglio 1994).

Pointing to see?

Brendan McGonigle

Department of Psychology, Laboratory for Cognitive Neuroscience and Intelligent systems, University of Edinburgh, Edinburgh EH8 9JZ, Scotland.
ejua48@ed.ac.uk www.ed.ac.uk/îsl/lab.home.html

Abstract: Ballard et al. neatly demonstrate the tradeoff between memory and externalised serial deictic functions which help the agent achieve economy. However, the target article represents an implementation which does not seem to reveal the hidden level of cognitive control enabling tasks, contexts, and the agent's own needs to specify and recruit information from the visual layout.

In harmony with various forms of the "active" agent thesis, Ballard et al. draw our attention to the data reducing, externalising role of deictic codes operating in visual search domains. Using the world itself as a scratch pad to off-load some of the memory obligations which could otherwise be troublesome is a trick which evolution seems to have used often. Even body orientation to a hidden food source seems to provide an effective mnemonic as Walter S. Hunter (1913) and others after him were to discover in the delayed response task.

Pointing with a limb also purchases focus and close attention to local detail. As many a primate worker has discovered, monkeys lock into visual information in the neighbourhood of where they put their fingers and when using an exocentric environmental frame of reference rather than a body centred reference retain contextually well specified adaptations to prismatically induced rearrangement over long periods (Flook & McGonigle 1978) again showing the ready retention of motor learning when there are external reference frames (McGonigle & Flook 1978).

Pointing with the eye seems like a straight extension of this at a more subtle level of functioning and there is indeed a suggestive logic that scanpath-based markers or landmarks simplify the memory and the encoding problems, acting as pointers and reducing the possible number of ways in which interrogation could take place. Under these conditions the layout is converted from a simultaneous presentation to a serial, sequential one paced by a clock which constrains the flow of input and forms an interface between cognition and the world. The dynamics are revealed in a trade-off between the run time and the relative investment in memory. And the economy motive behind the thinking here runs fully consistently with the way cognitive systems in general are being considered as governed by an economy motive which attempts to achieve the most behaviour for the least effort (Anderson 1990; McGonigle & Chalmers 1996).

However, the functional layer involving visual deixis has problems deriving from the source of top-down control as specified in the hierarchy. What exactly directs vision from the top and actively recruits information from the layout? The deictic visual pointer level seems too local and myopic a mechanism to deliver the transition between looking at to looking for – a long-standing issue in the domain of human development. Russian work, Zinchenko et al. (1963) for example – not cited by Ballard et al. – indicates that children confronted with identification tasks where they must identify an object intra- or inter-modally from a set following its interrogation either haptically or visually tend to be as good as their interrogation strategies; young children have long fixation times and grasp objects in a palmar grasp; older children who succeed in identification take a sequential search attitude in both modalities, tracing outlines with their fingers in the haptic task, and with their eyes (movements) in visual interrogation. Whilst cognitive state and the eye movements are well correlated in this domain, it is hardly likely however, that eye movements develop patterns of behaviour on their own. As Ballard et al. themselves emphasise, changing the task demands, even in the same context, will alter the criteria and the focus of what information needs to be recruited. And the layout itself cannot arbitrate on this matter. Instead, the process of perceptual learning and extraction of defining features, which will control the hunting movements of the eye as it sweeps an array, requires extensive experience of how objects vary with reference to one another in sets and collections, and what a particular task demands by way of "level of processing" (McGonigle & Jones 1978). These, however, are control matters which are not specified here, and it isn't clear either what kind of agent is necessary to run such a system in the first place. For the most part, the target article records what I take to be an implementation level specification, presuming a competent, perceptually sophisticated human subject, but leaving matters of genesis, and the role of higher level control for another day.

Nevertheless, I found much to agree with in the authors' characterisation of the brain as hierarchical, with various levels performing according to different time scales. To develop these ideas further needs more research into serial executive control which can both specify cognitive state and map eye movement. The sorts of search and ordering tasks using touch screens we have been developing at Edinburgh over the past 8 years or so offer extensive analyses of the hierarchical layers of organisation and the timing functions demanded by various visual sorting and ordering routines needed to control large search spaces (McGonigle & Chalmers 1996). Again, economy and resource optimisation is a key player, particularly in keeping memory demands low. In spatial search tasks, for example, it is clear that *Cebus apella* use external, locative features of the array as memory reducing processes following vectors of dots in principled self-organised and untutored series, rather than searching randomly and then relying on brute force memory to support exhaustive search requirements. Combining these search techniques with the eye movement analyses of Ballard et al. suggest productive new avenues for research into the way cognitive systems evolve the mental and physical pointers for efficient, optimised search in well embodied cognitive adaptations.

Embodiment, enaction, and developing spatial knowledge: Beyond deficit egocentrism?

Julie C. Rutkowska

Cognitive & Computing Sciences, University of Sussex, Brighton BN1 9QH, United Kingdom. **julier@cogs.susx.ac.uk**
www.cogs.susx.ac.uk/users/julier/index.html

Abstract: Traditional cognitivism treats a situated agent's point of view in terms of deficit egocentrism. Can Ballard et al.'s framework remedy this characterization? And will its fusion of computational and enactivist explanations change assumptions about what cognition is? "Yes" is suggested by considering human infants' developing spatial knowledge, but further questions are raised by analysis of their robot counterparts.

Accounts of embodiment often amount to little more than locating the agent in an environment. From the perspective of traditional cognitivism, this is a problematic thing, since it appears to tie the agent to a limiting spatio-temporal "point of view." This commentary suggests that "egocentrism" offers a misleadingly negative characterization of the inherent subjectivity that is at the core of adaptive functioning, and considers how far Ballard, Hayhoe, Pook, and Rao's seriously embodied variety of situatedness supports this position.

Treated at Ballard et al.'s embodiment level, movements like fixation are more than overt indices of covert cognition; they become an essential part of the cognitive process. Especially interesting is Ballard et al.'s application of classical computational

notions to their in-depth account of embodiment's place in situatedness and the deictic representations that it supports. Their analysis of how the embodiment level's fixation behaviors allow neurons to "refer" to an external point, effecting temporary variable bindings in behavioral programs, illustrates a de-centralized treatment of cognition through extension of program-governed processes to overt behavior, and representation as selective correspondence with the environment – not substitution for it through model-like re-presentations. Such assumptions move Ballard et al.'s framework towards the enactivist paradigm for understanding cognition, which seeks foundations in self-organization through the subject's sensory-motor activities, rather than in pre-given external information or internal models that it associates with the programs of traditionally cognitivist computational accounts (Varela et al. 1991).

How far may this computation–enaction rapprochement contribute to changing views of cognition? Focus on tasks like stacking blocks invites standard objections that this looks fine for low-level abilities but cannot scale up to "real" cognition. One way out of this is to consider what develops in domains traditionally thought to involve a qualitative transition from sensory-motor activity to "real" representation and cognition. A suitable candidate is our understanding of spatial knowledge.

Applied to human infants' spatial understanding, Ballard et al.'s embodiment level supports a promising alternative to the Piagetian picture of declining reliance on an action-based, egocentric, and often misleading appreciation of object location towards an objective representation of space as a container in which the self and its activity hold no privileged place. Contemporary work, while dispensing with the action focus, reinforces such assumptions by proposing a shift from (subjective) egocentric to (objective) allocentric strategies for coding location. Egocentric codes (e.g., "it's on my right") are considered of limited potential in a changing world; any movement of subject or of object invalidates them, making successful reaching for visible objects or search for hidden ones problematic. By way of contrast, allocentric codes that relate position to a surrounding spatial framework (e.g., "it's at a landmark") are considered objective and invariant with the subject's activities.

Perhaps advance requires us to retain behavior as the key to developing spatial understanding, but shift focus toward what subjects can achieve through increasing ability to exploit their gaze control and fixation. In the case of eye-hand coordination, a "do-it-where-I'm-looking" hand movement strategy can be enabled by a deictic code like "the-thing-I'm-fixating," which is neither subjective nor objective in the traditional sense. Since its referent automatically updates with the subject's gaze activity, it is viewer-oriented without being restrictively viewer-centered like an egocentric spatial code; and it is simultaneously object-centered without involving an objective description of the thing to which it refers or its spatial position. As far as the outcome of development is concerned, such mechanisms make sense of the view that increased use of landmarks when searching for objects may not involve abandoning self-referent coding in favor of purportedly more objective spatial codes (Bremner 1989). Instead, landmarks may aid fixation during infants' movements, supporting updating of what remains a self-referential code. Details of Ballard et al.'s embodiment analysis of fixation may considerably clarify how landmarks could work in this sense, rather than through reference to specific locations in an exhaustive representation of 3-D space. In particular, detailed comparisons with infancy studies are merited by proposals for localizing objects beyond the immediate visual field by combining transformations that relate objects to scenes and scenes to retinal co-ordinates of current fixation.

A reservation about the current presentation of Ballard et al.'s framework stems from how computational notions inform discussion of deictic representation, supporting a more traditional view of agent–environment systems than seems warranted by the framework's enactivist possibilities. The embodiment level is clearly shown to facilitate efficiency and speed of real-time processing. Yet its relation to the environment looks much like selection of pre-given environmental properties. For example, deictic primitives are said to "refer to points in the world with respect to their crucial describing features (e.g., color or shape)" (sect. 1.1).

Artificial systems that acquire spatial skills suggest tighter integration of sensory and motor contributions to situated activity. For example, the reverse engineering of an arena-centering robot finds no useful characterization of its performance in terms of sensors coming to detect any task-related invariant property of sensory stimulation; no distinction is found between input and output units in the sensory-motor controller that genetic algorithms enable it to "evolve" (Husbands et al. 1995). Sensory-motor co-variation rather than any version of environmental feature selection may be at stake, consistent with the view that sensors are not measurement devices; rather than encoding information about states of an agent in its environment, variation in sensor signals may depend on dynamics of agent–environment interaction (Smithers 1994). Grounds for retiring computational notions in favor of dynamic systems approaches? First, it will be essential to consider whether co-variational phenomena might effectively be captured by highlighting dimensions of animate vision that address how behavioral states like fixation and visual following may play a constructive role in constraining interpretation of input data such as optical flow (Ballard 1989).

Pointing with focussing devices

Wolfram Schultz

Institute of Physiology, University of Fribourg, CH-1700 Fribourg, Switzerland. **wolfram.schultz@unifr.ch www.unifr.ch/inph/**

Abstract: Evolutionary pressure selects for the most efficient way of information processing by the brain. This is achieved by focussing neuronal processing onto essential environmental objects, by using focussing devices as pointers to different objects rather than reestablishing new representations, and by using external storage bound to internal representations by pointers. Would external storage increase the capacity of cognitive processing?

Ballard et al.'s "It is always a good idea to give the brain less to do" (sect. 5) comes close to being a central tenet of this review. The brain is an extreme energy consumer, and there are limits to what the body may spend for a central system controlling its interaction with the environment. Evolution would certainly select for the most efficient way of energy consumption in the brain. This is the underlying assumption of the present theory.

An obviously efficient way to spend energy is to process only those events that are absolutely crucial at a given moment, and to use support devices with lower energy demands. It is not always necessary to process all components of the environment if mechanisms exist that allow missing components to be rapidly acquired. Taking the example of the eye, the limited area of highly concentrated photoreceptors and the low convergence to retinal ganglion cells restrict the energy-consuming processing largely to the fovea. The retinal periphery only transmits a sketchy overview image to the brain, which requires less energy for processing. However, the fovea can be rapidly moved by saccades to previously peripheral objects when they need to be processed with high resolution. This makes it possible to establish an accurate visual representation of the important components of the environment despite low acuity peripheral vision.

The obvious reduction in energy consumption is not limited to the retina and primary visual centers. It would also be useful to reduce the processing of changes of the visual scene at higher visual centers. As is well known to computer programmers, pointers would allow reference to be moved to a different event without recalculating the entire scene. Ballard et al. hypothesize that foveation may in fact constitute such a pointer mechanism,

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

and that saccadic eye movements serve as pointing devices. This would reduce energy consumption in brain centers involved in cognitive processing beyond the immediate sensory processing of visual inputs and would constitute a second way to reduce energy consumption besides the focussed processing by the fovea. Rather than simply compensating for low acuity peripheral vision, saccadic eye movements make it possible to reduce energy consumption by serving as pointers for cognitive operations.

Saccadic eye movements precede limb movements and even the activity of prime mover muscles. We have seen this in monkeys (1) performing in reaction time tasks in which an external imperative stimulus elicited the behavioral reaction (Schultz et al. 1989) and (2) with self-initiated movements in the absence of external imperative stimuli (Romo & Schultz 1990). Very impressively, the subject rapidly shifts its eyes to the target for the later limb movement irrespective of the movement being elicited by an external stimulus or following a deliberate decision. Only after that is the limb movement executed. The fact that this sequence also occurs with spontaneous eye movements suggests that subjects can deliberately select the target to be addressed by the pointer. This is instrumental in maintaining a high degree of liberty for choosing the objects to be processed.

Retinal organization is only one example of focussed processing and pointing. A more cognitive example is attention. As with the retina, the "attentional spotlight" can be moved as a kind of mental pointer to a salient environmental stimulus following an orienting reaction, or it can be deliberately directed to a known point of interest on the basis of cognitive representations. Mentally focussing attention on a particular object in the environment restricts the processing to that object and allows one to process it in great detail. Surrounding objects are largely neglected but can easily be processed in great detail by rapidly moving the attentional focus onto them. Although focussing implies that information from outside the area of focus is excluded or less well processed, deliberately moving the focus of selective attention would compensate for that effect and allow the deliberate choice between all objects in the environment.

The use of retinal and mental pointers allows further reductions in energy consumption by using external information storage. As the block-copying task revealed, the brain may not establish a complete representation of even the most crucial environmental objects. A representation of the color of the block to be copied was only maintained until the corresponding block was acquired, and subsequently only positional information was processed. Such fractionated representations exist in situations in which external storage is conceivably less energy-consuming than internal storage. The example of making saccades more expensive by moving the blocks farther away shows that the degree of completeness of internal representations can vary according to the respective efforts. Apparently, the brain adopts in each situation a processing strategy that uses the least amount of energy-consuming central representations.

Would the parsimonious use of representations simply lead to a reduction in energy consumption imposed by evolutionary selection, or could it also constitute a mechanism for increasing mental capacities? Specifically, would the increased use of pointers and external information storage extend the capacity of working memory and attentional processors beyond the usual 6–7 chunks? Further experiments might explore situations in which the use of pointers and external storage devices is quantitatively varied and the maximal processing capacity assessed.

Real and virtual environments, real and virtual memory

Gary W. Strong

Interactive Systems Program, National Science Foundation, Arlington, VA 22230. **gstrong@nsf.gov** **www.cise.nsf.gov/iris/isppdhome.html**

Abstract: What is encoded in working memory may be a content-addressable pointer, but a critical portion of the information that is addressed includes the motor information to achieve deictic reference in the environment. Additionally, the same strategy that is used to access environment information just in time for its use may also be used to access long-term memory via the pre-frontal cortex.

Ballard et al. present a convincing argument for a cognitive-perceptual processing strategy that off-loads information storage requirements by means of a postponed binding between short-term, working memory pointers and information that exists at spatio-temporal locations in the external environment. Such a Gibsonian strategy for minimizing internal representation stands in opposition to traditional symbol-processing accounts of elaborate rule-governed planning and behavior, and is more in line with recent contextual accounts such as those of Agre and Chapman (1987) and Brooks (1991). In addition, working memory is seen by Ballard et al. as a collection of pointers to information, each loaded by a single fixation or focus of attention and persisting over several subsequent fixations or foci. It is not made clear in the target article whether what is in working memory consists of convergence zones (Damasio 1989), pointer addresses similar to the computer addresses Ballard et al. present in some of their early examples, or spatial references associated with pointing movements of the perceiver. In a content-addressable memory, convergence zones and pointers may amount to the same thing. Spatial references associated with pointing movements of the body (i.e., deictic references), however, may require working memory to specify deictic behavior, as in the neural network of Strong and Whitehead (1989) that binds orienting movements, or "tags," to perceptual feature information to create temporary assemblies of minicolumns.

It may be incorrect to restrict the proposed strategy to environmentally derived information. Just in time binding may not only apply to location and perception of environmentally available information. Studies of the pre-frontal cortex and its connectivity with the rest of the brain (Goldman-Rakic 1987; Miyashita 1995; Ungerleider 1995) suggest that there are at least two pathways for visually derived information to reach the prefrontal cortex, one parietal and one temporal, with dense feedback projections along each pathway. Goldman-Rakic (personal communication) has suggested that the pre-frontal cortex may support the creation of a "surrogate environment" in which behavior must be based on plans with respect to information not currently available in the external world. From this one can conjecture that the postponed binding strategy of Ballard et al. may function just as well with respect to a virtual environment as with the physical one. Internal images could be spatially indexed (and internal phonological streams temporally indexed) using the same mechanisms as those postulated to occur with external scenes. In fact, the pre-frontal cortex would not have to represent an entire scene or stream but could be collecting an appropriate point of view just in time, in coordination with the indexing process.

Hence there may be at least two different ways to bring information to bear "just in time" on our behavior, as both Ballard et al. and Calvin (1996) suggest. One, by keeping information in the real world until it is needed. The other may be by accessing long-term memory with the aid of the prefrontal cortex just when needed. Both processes could operate in parallel, with an appropriate trade-off between external and internal processes according to the quality of the long-term memory. In this view of internal, long-term memory plays the role of a virtual environment that we access using the same tools we use to access information in the environment, but with much less physical constraint. This allows

us to contradict the present environment if experience has shown that the truth lies deeper than what appears before our eyes.

(The opinions contained herein are those of the author and not of the National Science Foundation.)

On the variety of "deictic codes"

Boris M. Velichkovsky

Unit of Applied Cognitive Research, Dresden University of Technology, D-01062 Dresden, Germany. **velich@psy1.psych.tu-dresden.de**
physik.phy.tu-dresden.de/psycho/cogsci.html

Abstract: Eye movements play a variety of roles in perception, cognition, and communication. The roles are revealed by the duration of fixations reflecting the quality of processing in the first line. We describe possible roles of eye fixations in different temporal diapasons. These forms of processing may be specific to sensorimotor coordinations. Any generalization to other domains should be cautious.

Ballard et al. present a refreshing review with nice experiments. Their main approach is to make neurocognitive modeling more flexible and integrated by postulating the "embodiment" level at a time scale of about 300 msec. In my commentary I will argue that there are several "embodiment" levels – at, but also below and above 300 msec. Second, I will elaborate on the task-dependency of processing emphasized by Ballard et al. and will show that some of their conclusions may only be valid for specific groups of the tasks.

First of all, with respect to the domain proper – investigation of visual exploration and problem solving as reflected in eye movements – Ballard et al.'s approach does not take into account qualitative differences among fixations. Individual fixations may be much longer as well as much shorter than the average value of 300 msec; this variation accordingly seems to have some functional consequences. For instance, we have recently investigated deixis in a more traditional sense than the one described by Ballard et al., that is, not in the service of visual information processing but in its referential function for interpersonal communication (Velichkovsky 1995). The experimental task was interactive tutoring in which experts could use their gaze to disambiguate verbal instructions given to novices. The deictic role was taken over by the group of exceptionally long fixations, with duration more than 500 msec. Thus, there may well be a relatively high-level deictic mechanism for binding another person's attention – a possible locus for "joint attention" effects in human cognition and communication.

In another study, we attempted to clarify the functional role of fixations in a more systematic way, considering them as events evoked by the presentation of a stimulus (Velichkovsky et al. 1997). The data show that there are extremely short fixations less than 100 msec; their role is to provide low-level support for orienting reaction as they demonstrate a very fast "ON-OFF" type of behavior. The group of fixations around 120–250 msec is largely a response to stimulus presentation and slowly diminishes in the course of the presentation. It is interesting to note that the number of these fixations correlates positively with attention to visual features of material; we accordingly called this most numerous group "perceptual fixations." The next group – from 250 to 450 msec – demonstrates behavior very similar to perceptual fixations with the difference that their reactions times are longer. Again, the key to the interpretation of these fixations is that their number grows when instruction emphasizes semantic categorization of the material – a numerical concurrence with the well-known involvement of P300 and N400 brain Event-Related Potential in semantic processing (e.g., Rugg & Coles 1995). Finally, the group of extremely long fixations – identified in our previous research as "communicative fixations" – again plays a role which is not specific to the processing of stimulus information per se. The number of fixations grows toward the end of presentation and during interstimulus intervals as if they were expectancy fixations correlated with the end of every specific experimental period.

My second comment is that although Ballard et al. emphasize the task-dependence of information processing, they seem to overgeneralize their paradigm being based on eye movement data. This is illustrated in their discussion of the unimportance of 3-D representations. Indeed, in saccadic programming for exact fixations on points in a landscape or, say, in a Gibsonian gradient, only proximal relationships are relevant. A version of 2-D representation is sufficient. The importance of 3-D representations in human perceptual experience and cognition can not be denied, however; it is well documented by investigations of metacontrast and motion (Velichkovsky & Van der Heijden 1994), brightness (Albright 1994), attention (Hoffman & Mueller, submitted) and imagery (Neisser & Kerr 1973). In other words, visual processing for sensorimotor coordination as in the paradigmatic task can be different from the type of processing underlying perception and cognition. This is the message of studies by Bridgeman (1991), Wong and Mack (1981), as well as Milner and Goodale (1995). In the same vein, in higher-order processing extrapolations on the basis of eye movement data are not necessary conclusive for phenomenal perception and cognition (Perner 1997, in press).

Together with other behavioral and neuropsychological data (Bernstein 1996; Challis et al. 1996; Fischer & Weber 1993), these results show the existence of a whole hierarchy of brain mechanisms. These mechanisms can probably be differentiated not only by the temporal parameters of complicated brain events but also by the duration of eye fixations. Although this conclusion coincides with Ballard et al.'s general premises, there seems to be a lot of room for further specifying the roles for the "deictic codes."

ACKNOWLEDGMENT

Thanks are due to Bruce Bridgeman, Marc Pomplun, Andreas Sprenger, and Pieter Unema for discussion of these issues.

Pointers, codes, and embodiment

Robert A. Wilson

Cognitive Science Group, Beckman Institute, University of Illinois, Urbana-Champaign, Urbana, IL 61801. **rwilson@uiuc.edu**
www.beckman.uiuc.edu/groups/cs/people/wilson.html

Abstract: This commentary raises three questions about the target article: What are pointers or deictic devices? Why insist on deictic codes for cognition rather than deixis simpliciter? And in what sense is cognition embodied, on this view?

Two of Ballard et al.'s crucial claims are that (1) a significant part of cognition uses deictic strategies, using pointers to reduce computational load; and that (2) this deictic reliance constitutes a way in which cognition is embodied. The general views in the target article are ones with which I find myself in sympathy; this commentary asks three questions about the central notions in play: deixis, coding, and embodiment.

What are pointers? "Deictic" is an adjective used chiefly to characterize uses of language where a significant part of the meaning is conveyed through contextual demonstration of some type. Hence, indexicals and demonstratives are paradigmatic deictic devices in natural language, since using them appropriately involves interaction between language and the context in which that language is used. Understanding the use of deictic devices, such as "I," "you," and "this," involves an investigation of the world beyond the speaker (at least, *qua* speaker). Competent language users know that "I" refers to the speaker, but to know the specific reference of "I" on particular occasions requires examining the context to see which person is the speaker. We can think of these linguistic, deictic devices as pointers, but clearly they are not the only types of pointers there could be.

The general idea of a pointer might seem straightforward: it is a device that indicates or reveals something not by explicitly representing that thing but by pointing to it. Hence pointers might play

a crucial representational role in cognition without themselves being representations of the things represented. Enter the idea that eye movements – or, more particularly, visual fixations – are a kind of deictic device, a kind of pointer (target article, sect. 1.1). And enter the idea that apart from this kind of "mechanical" pointer, there can also be "neural" pointers, such as *attention* (sect. 2.2). But in addition to these perceptual aspects to deictic cognition, there is also a deictic role for working memory and motor program execution through the operation of variable binding (sects. 2.2, 2.3).

My initial question can now be rephrased: Is there a univocal use of "pointer" in all of these cases? While fixation and attention play the same sort of functional role that indexicals and demonstratives play – that is, they connect up representations (mental or public) with the aspects of the world represented – it is not clear how this is true of working memory except insofar as it processes inputs from perception. (In this connection, I found problematic the example in Table 2 on the use of pointers in computer memory.) Moreover, insofar as motor instructions are deictic (e.g., grasp an object with *this* shape), they would seem to do so via a *perceptual* loop. So it would seem that the deictic nature of cognition and motor programs is, at best, derivative from that of perception. Is this a welcome consequence of Ballard et al.'s views, or does it involve a misinterpretation of what they mean by a "pointer"?

***Why deictic* codes?** These concerns aside, deictic strategies may indeed play a more important role in cognition than many have thought, but why emphasize deictic *encoding?* Consider eye movements and rapid visual fixation. Fixation, conceived of as a deictic sensory action, does make computational sense, but why should this action – something one does with one's body – be conceived in terms of the notion of encoding? One virtue of deictic representation is that it avoids a regress problem that hyper-representationalist views of mental (especially linguistic) processing face by recognizing that not all parts of a representational process need be explicitly representational. But by emphasizing "deictic codes" for cognition, Ballard et al. lose this (and related) advantages. This is not to suggest that there is *no* encoding or internal representation in cognition, only that deictic cognition is best construed as showing that not all of cognition is internal, encoded representation.

A consideration of linguistic indexicals and demonstratives may help make the point more clearly. What marks them off as deictic devices in communication is that they do not encode their complete sense; that is instead provided by an encodable linguistic rule (e.g., "I" refers to the speaker) together with unencoded information that lies in one's environment.

In what sense "embodiment"? The notion of embodiment is introduced in the target article (sect. 1) as a distinct *level* at which cognitive processing takes place, marked off chiefly by its temporal distinctness (0.3 sec). It is embodied because it involves a distinctive computational primitive, the "physical act," such as an eye movement. And it is deictic because of the role those acts play in cognition. As we have seen, however, other deictic strategies are used in cognition, such as attention and grasping; and as the authors seem to acknowledge in Table 1, these operate at very different time scales. The previous question concerned the sense in which all three are deictic; the current question concerns the sense in which they constitute a distinct level of cognitive processing.

There is a different and more radical sense in which cognition is embodied that runs counter to the idea of a distinct level of embodiment but accords with a closing thought in the target article. In their conclusion (sect. 5), Ballard et al. note that the idea of leaving noncrucial information in the environment to be picked up just as it is needed for cognitive processing carries with it obvious computational advantages. Our basic body movements provide the crucial link in the execution of this strategy. We, along with many other creatures, may indeed have evolved this general informational short-cut. But our adoption of this strategy also involved constructing informationally enriched environments: public signs, permanent symbols, written languages. What makes us cognitively distinctive is our reliance on continual interaction with these enriched environments. Cognition is embodied and deictic further "up" than the 0.3 second rule suggests.

Authors' Response

Pointing the way

Dana H. Ballard, Mary M. Hayhoe, Polly K. Pook, and Rajesh P. N. Rao

Computer Science Department, University of Rochester, Rochester, NY 14627 **dana@cs.rochester.edu; mary@cs.rochester.edu; pook@isr.comm; rao@salk.edu www.cs.rochester.edu**

Abstract: The majority of commentators agree that the time to focus on embodiment has arrived and that the disembodied approach that was taken from the birth of artificial intelligence is unlikely to provide a satisfactory account of the special features of human intelligence. In our Response, we begin by addressing the general comments and criticisms directed at the emerging enterprise of deictic and embodied cognition. In subsequent sections we examine the topics that constitute the core of the commentaries: embodiment mechanisms, dorsal and ventral visual processing, eye movements, and learning.

R1. The enterprise

The majority of respondents appear to agree that the time to focus on embodiment has arrived and that a disembodied approach to human cognition is unlikely to yield satisfactory results. **Bogacz** calls this "an increasingly popular approach to the mind," but despite pioneering work by Lakoff (1987), Clark (1997), and **Glenberg,** the very difficult task of relating embodiment to neural structures in a detailed manner remains in its infancy. In his recent *BBS* target article, Glenberg (1997) makes an excellent case for the embodiment of memory, collecting a wealth of evidence for such an approach and against earlier "propositional" approaches. However, the level of exposition in that article is precomputational, whereas, as noted by **Kirlik,** we are concerned with understanding the detailed computational mechanisms that the brain uses for implementing such schemes. The goal of the target article is to suggest initial steps in this admittedly difficult process.

Feldman defines embodiment as the essence of cognitive science and suggests that we over-specialize the term by limiting it to the one-third second time scale. **Fuster** and **Wilson** also suggest that the one-third second scale is too specific to be meaningful. However, there is a sense in which this specialized use captures an essential feature of human behavior. Given much less than 0.3 second, computation is decidedly neural because there is not enough time to communicate with the outside world except in the case of primitive reflexes. Given much more than 0.3 second, for example, 10 seconds, there is time to plan behavior by running internal simulations. Once again, computation becomes predominantly neural, albeit for a different reason. Thus, the 0.3 to 1 second time scale is the time when behaviors usually interact with the world. This is just

enough time to look up what to do next, before one has to do it. Incidently, as **Bryson & Lowe** point out, we are certainly remiss in not mentioning the work of Pöppel (1994), who has been a pioneer in pointing out temporal constraints on brain processing capabilities.

Fuster also questions the sense in which the embodiment level is fundamental, specifically challenging our claim that slower time scale structures are built on top of the embodiment level and are therefore bound to its results. This separation of levels is illustrated in a set of block copying experiments (sect. 3 in the target article) in which subjects were asked what they remembered (Hayhoe et al. 1997). In these experiments, the color of uncopied blocks was changed when the subjects were making saccadic eye movements to the model area. Subjects did not notice these changes, but when they were told of the change and asked how many blocks had changed color, they gave a modal response of one, whereas the modal number of blocks that had actually changed color was seven. Their response makes sense if the slower "verbal" level only has access to the ongoing state of working memory and thus must report the variable's value as represented by the ongoing task. In the same way, one can understand the illusion of a stable world. This illusion is certainly not easily computed from the fractionated picture presented to the brain by the one-third second saccadic fixations, with a one degree fovea. More likely, the illusion is produced by a confluence of two factors: (1) the circuitry of awareness is much slower than that of deictic orienting structures, and (2) awareness is heavily focused on the state used to keep track of ongoing tasks.

Thus, in response to **Fuster,** the illusion of perceptual stability is a byproduct of the brain's processing hierarchies and is an indication of a central commitment rather than a loose constraint that could in principle be arbitrarily extended to other levels. Some of the questions raised by **Fuster** may have their origins in the fact that there are fundamentally different ways to think about cortex. In one, which is perhaps the most widely held, the cortical hierarchies are viewed as different neural levels capable of notionally independent processing for mediating many different behaviors. In contrast, our view is that at the one-third time scale, the cortex can be seen as a coherent distributed memory that is the source of pointer referents. A pointer may refer to abstractions of stimuli currently available predominantly from the dorsal and ventral visual streams or to hypothetical cases that primarily use the prefrontal cortex, as in the case of internal simulation (**Dominey** and **Strong** make similar suggestions).

Feldman also suggests that the limited perception of changing blocks during saccades cannot be the whole story. A variant of his suggestion of a sliding door over the model has been experimentally tested by Bensinger. The model was made invisible when subjects were placing blocks in the workspace. In this case, the number of looks to the model was comparable to the normal case of a completely visible model, suggesting that subjects resisted a memorization strategy. As to the claim that subjects would then "almost certainly notice changes in the salient parts of the model," this would have to be tested. Current experiments by Rensink et al. (1996) and Grimes and McConkie (1995) indicate that when such tests are done, subjects invariably notice changes less than expected.

Damper blurs the distinction between perception and cognition, rightly reminding us that controversy about this stems from earlier work by Newell (1990) and Harnad (1990), as well as his own experimental work (submitted). The need to span abstraction levels arises naturally in speech because it contains natural units – phonemes and sentences – each with time scales of their own. The target article suggests that underlying the perception/cognition distinction is the fact that brain computation conforms to different abstract models at different time scales.

Our view is that at the level of "cognition" an underlying behavioral program manipulates variables stored in short term memory, which is the way the program keeps track of where it is in its execution cycle. These variables can be thought of as pointers to larger sensorimotor referents. The time scale of cognition is typically of the order of several seconds, but because focus is on the world, the contents of pointers, which are computed much more quickly, are highly relevant. Similarly, "perception" typically concerns the referent of a pointer that can be computed in much shorter time scales, from 80 to 300 milliseconds. Nonetheless, experiments in perception often involve longer time scales, for example, when verbal reports that take seconds to generate are used as a measure. Thus the focal problem is that the words "cognition" and "perception," which have served us so well for so long, do not factor easily into the brain's underlying machinery, in that they refer to concepts that span levels of representation that are distinct in space and time.

Several commentators, notably **Maglio, Wilson,** and **Kirlik,** suggest that the target article did not go far enough in characterizing the extent to which the world participates in task-directed action. However, our neglect here was more one of emphasis, in that the target article stressed the temporal dynamics of the binding mechanisms for interacting with the world rather than attempting to describe the complete story. In particular, **Maglio** is right in that the target article's account of pointers does not do full justice to the variety of ways in which the world is used to store the partial results of computations. His own work with Tetris players shows that players rotate blocks as early as possible, presumably because it is easier to match visibly rotated pieces. (Incidently, the overrotation observed may have a simpler interpretation than suggested. If subjects had to wait for the perceptual match after each rotation before rotating further, this could be slower than decoupling the matching process from the rotating process and then going back when detecting a correct match.)

R2. Embodiment mechanisms

Wilson registers some difficulty with the distinction between explicit and implicit pointing. These are covered in the target article but perhaps not sufficiently. When gaze is focused on a specific point, the "neural pointer" is unnecessary because the body's physical orientation is subserving the requisite function. However, when a pointer is required to a region of space that is not currently foveated, a neural pointer must be recruited. The obvious candidate for the genesis of neural pointers is the hippocampus because it has been implicated in long term potentiation and skill transfer. Thus, our contention, which has also been voiced by others, is that the hippocampus has special circuitry for reactivating the structures that were activated during physical pointing.

Wilson also suggests that the connection of pointers with short-term memory and language is tenuous. Carpenter and Just (1989), however, have shown that college students with reduced working memory have more trouble understanding embedded sentences, thus relating working memory to pointers in the parse tree. This suggests the possibility of linguistic deictic reference, which one can think of as naming pointers in a parse tree. **Wilson** also questions our use of "codes." Codes are used in our model at many levels. First, pointers are codes in that they are impoverished references that point to larger structures. Second, the structures to which they point are also codes when viewed in the context of visual routines. In this case, the neural structures subserve task-directed goals that are only peripherally related to the full fidelity of the external world.

The block copying task is special in that the scaling of the problem encourages the use of overt eye movements to accomplish the task. In fact, the value of the blocks task is precisely that such a normally covert strategy is made overt. We assume with **Epelboim** that it is possible to copy or build an object from memory. In that case, internal deictic strategies may still be used. Studies of human learning indicate that performance in a perceptual task can be improved overnight via consolidation during rapid eye movement (REM) sleep (Karni et al. 1994). The use of REMs during consolidation suggests that subjects may retain a more or less literal encoding of the task that includes its deictic features.

The central finding of our experimental work is that subjects seem to be structuring working memory in order to simplify the ongoing program. In support of this, **Schultz** makes the connection with metabolism. The brain consumes 20% of the body's energy but is only 2% of the total body weight. Thus, keeping a huge library of stored programs, while no doubt expensive, is apparently worth the cost. Allman and colleagues (1993), in comparing omnivorous primates with their leaf-eating counterparts, show that the former have significantly larger brains (after factoring out the effect of body weight), presumably to store the exigencies of their more complex behaviors.

To clarify this hypothesis, Figure R1 compares two hypothetical cases, showing the instantaneous working memory load in each case. In the upper figure, three blocks are memorized at once. The lower figure shows the working memory load for the model-pickup-model-drop (MPMD) case, which was the modal eye movement sequence observed experimentally in the block copying task. The lower case is distinguished by having dramatically less transient state. Thus, our primary suggestion is that the brain is set up to store programs as rote behaviors and, in the process of their automatization, favors descriptions with the fewest free parameters. Thus, rather than needing to add working memory to our model, as suggested by **Fuster,** it is clear that working memory is already present as a central governing feature.

The embodiment level does not address the problem of what happens at even more basic levels. For example, how do the representations used by the visual routines get formed? This is addressed in Rao and Ballard (1997) but not in the target article. However, knowing the details of the representation at these finer scales will not cause the embodiment levels to be revised drastically. **Rutkowska** suggests that work with primitive learning robots raises problems for the deictic account of behavior, in that such

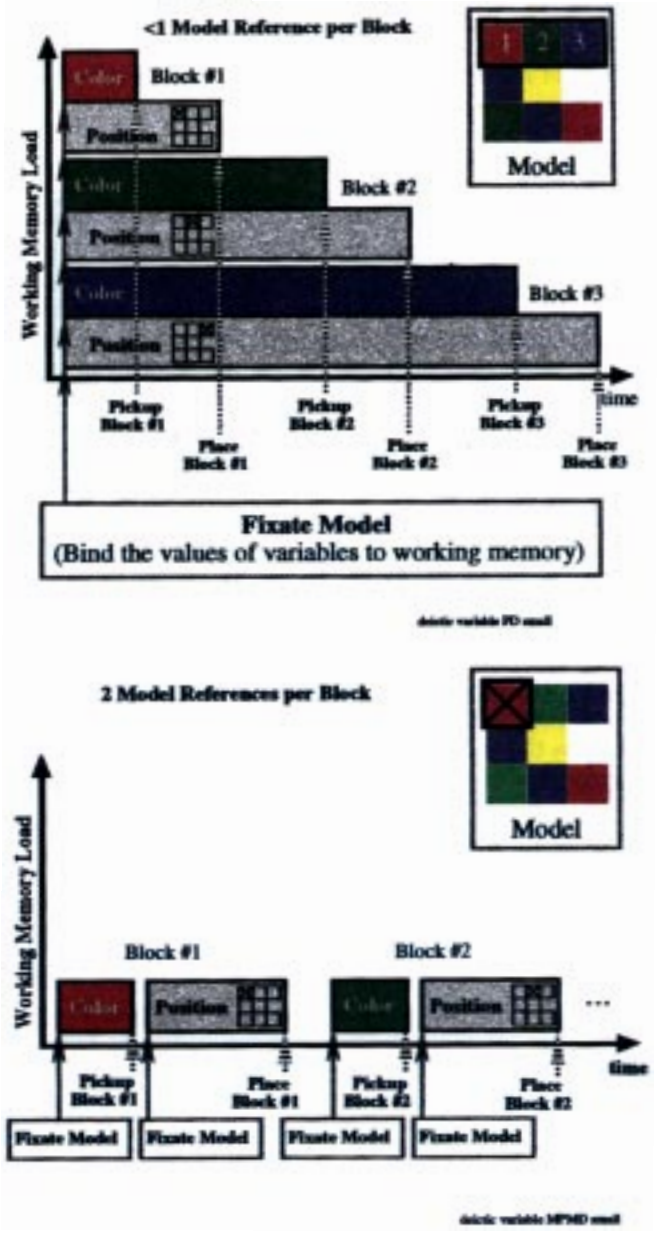

Figure R1. Working memory in the block copying task. A fundamental finding of the block copying task (sect. 3 in the target article) has been that working memory in tasks is typically used below capacity. This is illustrated here schematically (top). In copying a block pattern, if subjects memorized the position of three blocks at a time, they would have to keep that information in working memory until that portion of the copy was completed (bottom). Instead, subjects prefer to acquire the information on an individual block-by-block basis as they go about the copying task, thereby minimizing their instantaneous working memory load.

robots do not "pre-select" their environmental properties, as is implied by our feature vector. However, (1) the feature vector in the current model is a straw man, standing in for the products of a low-level learning algorithm (see Rao & Fuentes 1996 or Rao & Ballard 1996) and (2) such robots, which are already extremely deictic in their orienting responses, may become even more so and in addition will need working memory as they take on more demanding tasks.

Bryson & Lowe draw attention to the possibility of synchronous firing, suggested by von der Malsburg (1995)

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

as a way of binding disparate referents in cortex. This possibility has been extensively analyzed by Shastri and Ajjanagadde (1993). The problem with this hypothesis has been the lack of direct evidence for its use in this role. It may be that synchronous firing reflects natural onsets in the stimuli (Singer & Gray 1995) or that synchrony is used in subtle ways that are different from pointer binding Abeles et al. (1994).

R3. Dorsal and ventral streams

Goodale makes a very useful analogy between teleassistance and his and Milner's distinction between dorsal and ventral streams. The point of their model is that both streams cooperate in visuomotor computations, but with different foci. **Goodale** suggests that we are vague in making the connection between their work and the model presented in the target article. This may be a fault of the presentation, although the visual routines section shows explicitly how dorsal and ventral computations might be organized. To identify an object, the processing area is limited to the foveal region. Thus, the circuitry in the dorsal stream must target this foveal region to task-relevant locations so that the primary object recognition routines in the ventral stream can match the foveal feature vector. Thus, both streams cooperate to match the object centered features. To locate an object, the remembered "top-down" description of the object is propagated along the feedback pathways in the ventral stream, where it is correlated with the "bottom-up" sensory feature vectors at multiple scales. This correlation process produces a salience map that is assumed to reside in the dorsal pathway. The salience map can be used to define correlation targets for saccadic targeting. Here, the primary focus is the salience map in the dorsal stream, but the computation that produces this result begins in the ventral stream.

The tasks studied by **Jüttner** are interesting but might be explained in terms of the dorsal/ventral dichotomy. It is possible that learning complex concepts requires the fovea, as supported by the work of Jüttner (1993; 1996) and others, but that spatial generalization uses what the target article calls a location mechanism for computing salience maps and saccadic targeting. Given a target description stored in the ventral stream, the ability to correlate that target over the retinal array could be the core of a generalization process.

R4. Eye movements

Findlay et al. make a number of very important points that refine our simple model of saccadic targeting. Similar refinements are supported by the recent experiments of Hooge (1996) and by our own experiments (Rao et al. 1997; Zelinsky et al. 1997). In the latter, subjects were shown a small picture of an object and then had to report its presence or absence in a large scene shown afterward. Under these circumstances, the eyes were noticeably affected by the background as in **Findlay et al.**'s (1995) experiments. However, the search pattern was not serial by object or token but instead appeared to be a sequence of successively parallel searches using finer grained information (Geisler & Chou 1995; Palmer et al. 1993). When subjects were given a preview of the scene before seeing the target, they could make a saccade directly to the target. This suggests that parallel search mechanisms can be augmented with information about spatial location. The structure of that memory is suggested in section 4 of the target article.

Another important point raised is that delaying the onset of the saccade may allow subjects to saccade directly to the target. In our model, this can happen in circumstances without spatial memory where the target is computed from features via a multiscale correlation process assumed to be carried out in the visual cortical hierarchies. This computation is relatively slow, taking more time than is needed to generate eye movements. Thus, eye movements in the model are made whenever they can be, using the current and possibly intermediate results of the cortical correlation process. Similarly, the distinction between overt and covert attention that **Findlay et al.** make may be less than has previously been believed. It is possible that eye movements are not made in a given task because (1) the system responsible for making a decision pertaining to the task can decide faster than the time needed to generate an eye movement, and (2) only peripheral information may suffice for the task. For example, in the context of the targeting model presented in the target article, a peak in the salience map can be used to judge directly the presence of a target object in the periphery using a signal-to-noise criterion rather than executing an overt saccade to the location containing the salience peak. This would correspond to making a decision based on covert attention.

A number of commentators cautioned that eye movements are not just used for "cognitive" or memory operations but, as shown in both the blocks task and **Epelboim**'s tapping task (1995), are used for hand-eye coordination as well. Incidentally, the fact that **Epelboim**'s subjects are slower using LEDs only is predicted by the model in Rao et al. (1997). The sparse lighting makes it difficult to use a scene frame for priming spatial locations.

Fischer makes the excellent observation that, in coordinating fixations with arm movements, the fixation durations may be adjusted to interface the eye movement system with the rest of the cognitive program. The important point, however, is that the direction of gaze is still an important indicator of processing. The dwell time (fixation duration) may be influenced by several factors. In fact, **Fischer**'s observation is borne out in the blocks task. The fixations in the workspace and resource precede the hand by approximately 1 second, indicating that they are "waiting" to guide the hand. Our prediction is that if the hand movements were slowed the number of model fixations would remain constant but the dwell times would increase.

With regard to **Fischer**'s own experiments, the disparity between distal fixations in parsing and the spatial memory for his task may reflect the unfamiliar nature of the task. His subjects must index using word information only, whereas parsers generally use grammatical referents set up by an overlearned parsing program.

Regarding more minor points, (1) subjects do not use a neutral resting point in the "FAR" condition where the condition and model are widely separated. All fixations appeared to be purposeful. However, the workspace is placed halfway between the resource and model, so perhaps this is still unsatisfactory. (2) Our measure of "fixations per block" was specifically used to normalize for the shorter time of the monochrome task.

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press

Velichkovsky makes the point that fixation durations may be variable for a variety of reasons (e.g., saccade programming constraints, motor integration [reading text aloud], or cognitive operations). One certainly has to take all these factors into account in interpreting latency. **Velichkovsky** also suggests that visual operations use three-dimensional (3D) representations. There is no doubt that the brain can perform a number of 3D operations; but more to the point, deictic primitives allow the brain to economize in many cases by avoiding the need to build costly 3D representations. This is because the effects of 3D can be realized functionally in terms of visual routines that are fundamentally two-dimensional, together with behaviors. A well known example supporting such a claim is the set of experiments by Lee (1976) showing that subjects use $\dot{\tau}$ to estimate time-to-contact when stopping cars.

One of the hallmarks of research using pointers in natural tasks is that the full complexity of natural environments is used. Under these conditions, **Jordan**'s actual and intended pointers coincide, so there is no need to make his distinction. **Jordan**'s examples are interesting, but each involves a special situation. Again, the key distinction we would keep in mind is the crucial difference between a pointer and its referent. The referent is just the information needed for the current point in an ongoing task. This may be a color or offset as in the blocks task or it may be a visual direction as in **Jordan**'s examples. To actually compute a referent involves a visual routine, and, as **Jordan** points out, there are a number of conditions under which the computations of visual direction can go awry.

R5. Learning

We agree with the commentators, particularly **Epelboim,** who see learning as an absent but essential feature of any deictic account of behavior. In fact, McCallum (1996) has extended his learning algorithm to use deictic encodings and has had dramatic success in a simulated driving task. In collaboration with Prof. Gerhard Sagerer at the University of Bielefeld in Germany, we have also studied the effects of learning using a copying task similar to the blocks task. Subjects repeatedly built copies of model objects constructed from German Baufix toy parts consisting of bars, screws, wheels, nuts, and bolts (Magnuson et al. 1997). These parts subtend approximately 1 degree of visual angle during manipulation and are thus ideal for eye tracking studies.

As **Epelboim** and **Bogacz** predict, eye movements dropped dramatically during repeated assemblies of the same model object. In particular, the number of eye movements to the model were reduced by a factor of almost four between the first and the eighth assembly. Even at the eighth assembly, however, they were still greater than one per part, suggesting that the subjects still found deictic strategies helpful. By the twenty-fifth assembly of the object, the construction can be memorized. The main suggestion would be that deictic strategies meld with learning algorithms in the following way. Given that a primary goal of the brain is to develop and store programs that have predictive value, programs that have no or few free parameters are likely to be more efficient. Deictic reference is a way of holding these parameters until the brain can find a way to encode them more compactly. One way to achieve this is to discover features that work under low resolution so that the program can still be run even though the features are in the visual periphery. Hence, the motivation for **Jüttner**'s results. Learning occurs in the fovea but can later transfer to the periphery. The particular advantage of the Baufix assembly task and of using objects with approximately 8 to 30 fairly large parts is that strategies that might ordinarily be covert with respect to individual steps are made overt by the particular demands of the task geometry. Thus, one can catch a glimpse of the brain's programming mechanisms at work.

McGonigle points out that learning in development has been studied already and provides hints as to the development of deictic programs. **McGonigle** also finds that in the target article there is not enough development of the control of deictic programs. We and others have studied this, but it was not included in the target article except for a cursory account of Whitehead's (1991) and McCallum's (1995) respective work in reinforcement learning. As noted by Barto et al. (1991), Montague et al. (1996), and **Schultz,** the huge problem in programming the brain is in dealing with delayed rewards. An agent must choose an action at the current time instant based on the action's value in getting a reward in the future. The target article shows how the memory needed to negotiate a maze can be learned in the form of a tree of current and previous actions and perceptions. Once learned, at any given moment, the agent uses its previous perceptual-motor history to select the best next action. McCallum (1996) has extended this to a system of deictic actions for driving and has shown that in simulation, an agent can learn to negotiate a highway of slower and faster moving vehicles.

Dominey also agrees that learning is important and suggests that deictic pointers may play a crucial role in internal simulations in the prefrontal cortex. However, in his example of disambiguating the sequence ABCBDC, he maintains that "the required context is explicitly provided by visual input." It is hard to see how this could be the case. The context is visual but implicit, and to disambiguate the sequence, a parser has to form the disambiguating pairs AB BC and so on. McCallum's (1996) work in learning is of interest here because it provides an algorithm for deciding on the right amount of past context needed for correctly disambiguating such perceptual sequences.

Bogacz indicates that deictic strategies should be faster and that the reason for the fixations is the unfamiliarity of the task. Deictic strategies *are* faster than the strategy in which the subject is forced to memorize the model. In addition, it is precisely the unfamiliarity of the task that, in our view, motivates the usefulness of deictic encoding. **Bogacz** assigns the brain the task of monitoring, but it is not at all clear what this is. A more detailed model might start to have the level of detail described in section 4 of the target article.

Strong makes the excellent point that the role of internal memory may be to stimulate what might happen, and proposes that prefrontal cortex has the machinery to support such simulations, a suggestion also made by **Dominey.** The key use of such simulations would be to predict the outcome of future strategies based on current knowledge. As **Strong** suggests, this is not inconsistent with deictic strategies, but overt deictic strategies are just part of the story. Covert deictic strategies are possible, too. **Bogacz** makes this point in her criticism of overt deictic strategies, but in response, we once again note that our experiments

were designed to force the deictic nature of the task to be at its most explicit, in order to readily characterize its distinctive properties. In addition, internal deictic strategies are likely to be extensions and elaborations of more primal overt deictic strategies. Thus, in response to **Bogacz,** it is extremely hard to see how a Cartesian "brain in a vat" approach to cognition could yield any meaningful results because the normal brain depends crucially on sensorimotor interactions with the external world for developing its internal representations.

References

Letters *a* and *r* appearing before authors' initials refer to target article and response respectively.

Abeles, M., Prut, Y., Bergman, H. & Vaadia, E. (1994) Synchronization in neuronal transmission and its importance for information processing. *Progress in Brain Research* 102:395–404. [rDHB]

Abrams, R. A. (1992) Coordination of eye and hand for aimed limb movements. In: *Vision and motor control*, ed. L. Proteau & D. Elliott. Elsevier Science. [MHF]

Abrams, R. A., Meyer, D. E. & Kornblum, S. (1989) Speed and accuracy characteristics of saccadic eye movements: Characteristics of impulse variability in the saccadic system. *Journal of Experimental Psychology: Human Perception and Performance* 6:529–43. [JMFi]

Agre, P. E. & Chapman, D. (1987) Pengi: An implementation of a theory of activity. *Proceedings of the American Association for Artificial Intelligence* 87:268–72. [aDHB, GWS]

Albright, T. D. (1994) Why do things look as they do? *Trends in Neurosciences* 17(5):175–77. [BMV]

Allman, J. M., McLaughlin, T. & Hakeen, A. (1993) Brain structures and life-span in primate species. *Proceedings of the National Academy of Sciences USA* 90:3559–63. [rDHB]

Allport, A. (1989) Visual attention. In: *Foundations of cognitive science*, ed. M. I. Posner. MIT Press. [aDHB]

Aloimonos, J., Bandopadhay, A.& Weiss, I. (1988) Active vision. *International Journal of Computer Vision 1*(4):333–56. [aDHB]

American Association for Artificial Intelligence (AAAI) (1996) Fall symposium on embodied cognition and action. Technical Report FS 95–05. [JAF]

Andersen, R. A. (1995) Coordinate transformations and motor planning in posterior parietal cortex. In: *The cognitive neurosciences*, ed. M. S. Gazzaniga. MIT Press. [aDHB]

Anderson, J. (1990) *The adaptive character of thought*. Lawrence Erlbaum. [BM]

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

Arbib, M. A. (1981) Perceptual structures and distributed motor control. In: *Handbook of physiology – the nervous system II: Motor control*, ed. V. B. Brooks. American Physiological Society. [aDHB]

Arbib, M. A., Iberall, T. & Lyons, D. (1985) Coordinated control programs for movements of the hand. In: *Hand function and the neocortex*, ed. A. W. Goodman & I. Darian-Smith. Springer-Verlag. [aDHB]

Ashby, W. R. (1952) *Design for a brain*. Chapman & Hall. [AK]

Baddeley, A. (1986) *Working memory*. Clarendon Press. [aDHB]

Bajcsy, R. (1988) Active perception. *Proceedings of the Institute of Electrical and Electronics Engineers* 76:996–1005. [aDHB]

Baker, G. L. & Gollub, J. P. (1990) *Chaotic dynamics: An introduction*. Cambridge University Press. [aDHB]

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

(1989) Reference frames for animate vision. In: *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, ed. N. S. Sridharan. Morgan Kaufmann. [JCR]

(1991) Animate vision. *Artificial Intelligence* 48:57–86. [aDHB, SB]

(1996) On the function of visual representation. In: *Problems in perception*, ed. K. A. Akins. Proceedings, Simon Fraser Conference on Cognitive Science, February 1992. Oxford University Press. [aDHB]

Ballard, D. H., Hayhoe, M. M. & Pelz, J. B. (1995) Memory representations in natural tasks. *Journal of Cognitive Neuroscience* 7(1):66–80. [aDHB, PPM]

Barlow, H. B. (1972) Single units and cognition: A neurone doctrine for perceptual psychology. *Perception* 1:371–94. [aDHB]

Barsalou, L. W. (1993) Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In: *Theories of memories*, ed. A. C. Collins, S. E. Gathercole & M. A. Conway. Lawrence Erlbaum. [AMG]

Barto, A. G., Sutton, R. S. & Watkins, C. J. (1990) Sequential decision problems and neural networks. In: *Advances in neural information processing systems 2*, ed. D. S. Touretzky. Morgan Kaufmann. [aDHB]

(1991) Learning and sequential decision making. In: *Learning and computational neuroscience*, ed. M. Gabriel & J. W. Moore. MIT Press. [rDHB]

Behrmann, M. & Moscovitch, M. (1994) Object-centered neglect in patients with unilateral neglect: Effects of left-right coordinates of objects. *Journal of Cognitive Neuroscience* 6(1):1–16. [aDHB]

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

Bernstein, N. A. (1996) *Dexterity and its development*. Lawrence Erlbaum. [BMV]

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

Bremner, J. G. (1989) Development of spatial awareness in infancy. In: *Infant development*, ed. A. Slater & J. G. Bremner. Lawrence Erlbaum. [JCR]

Bridgeman, B. (1991) Complementary cognitive and motor image processing. In: *Presbyopia research*, ed. G. Obrecht & L. W. Stark. Plenum Press. [BMV]

Broadbent, D. E. (1958) *Perception and communication*. Oxford University Press. [aDHB]

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

(1991) Intelligence without reason. *Artificial Intelligence* 47:139–59. [aDHB, JB, GWS]

Brown, V., Huey, D. & Findlay, J. M. (1996) Face detection in peripheral vision. *Perception* 25(Suppl.):A89. [JMFi]

Bruner, J. (1957) On perceptual readiness. *Psychological Review* 64:123–57. [MJ]

Buhmann, J. M., Lades, M. & von der Malsburg, C. (1990) Size and distortion invariant object recognition by hierarchical graph matching. *Proceedings of the Institute of Electrical and Electronics Engineers International Joint Conference on Neural Networks* (Vol. II). IEEE Neural Networks Council, San Diego, CA. [aDHB]

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

Calvin, W. H. (1996) *How brains think*. Basic Books. [GWS]

Carpenter, P. A. & Just, M. A. (1989) The role of working memory in language comprehension. In: *Complex information processing: The impact of Herbert A. Simon*, ed. D. Klahr & K. Kotovsky. Lawrence Erlbaum. [rDHB]

Challis, B. H., Velichkovsky, B. M. & Craik, F. I. M. (1996) Levels-of-processing effects on a variety of memory tasks. *Consciousness and Cognition* 5(1/2):142–64. [BMV]

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

Chase, W. G. & Simon, H. A. (1973) Perception in chess. *Cognitive Psychology* 4:55–81. [aDHB, PPM]

Chun, M. M. & Potter, M. C. (1995) A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance* 21(1):109–27. [aDHB]

Churchland, P. S., Ramachandran, V. S. & Sejnowski, T. J. (1994) A critique of pure vision. In: *Large-scale neuronal theories of the brain*, ed. C. Koch & J. L. Davis. MIT Press/Bradford Books. [aDHB]

Clark, A. (1987) Being there: Why implementation matters to cognitive science. *Artificial Intelligence Review* 1:231–44. [RID]

(1997) *Being there: Putting brain, body, and world together again*. MIT Press. [rDHB]

Clark, H. H. (1996) *Using language*. Cambridge University Press. [AMG]

Crisman, J. & Cleary, M. (1994) Deictic primitives for general purpose navigation. *Proceedings of the American Association for Artificial Intelligence Conference on Intelligent Robots in Factory, Field, Space, and Service (CIRFFSS)*, March. [aDHB]

Damasio, A. R. (1989) Time-locked multi-regional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33:25–62. [GWS]

Damper, R. I., Harnad, S. & Gore, M. O. (submitted) A computational model of the perception of voicing in initial stops. *Journal of the Acoustical Society of America*. [RID]

Derrico, J. B. & Buchsbaum, G. (1991) A computational model of spatiochromatic image coding in early vision. *Journal of Visual Communication and Image Representation* 2(1):31–38. [aDHB]

Dominey, P. F. (1997) Influences of temporal organization on transfer in

sequence learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, in press. [PFD]

Dominey, P. F., Arbib, M. A. & Joseph, J. P. (1995a) A model of cortico-striatal plasticity for learning oculomotor associations and sequences. *Journal of Cognitive Neuroscience* 7(3):311–36. [PFD]

Dominey, P. F. & Boussaoud, D. (1997) Behavioral context is encoded in recurrent loops of the frontostriatal system. *Cognitve Brain Research*, in press. [PFD]

Dominey, P. F., Schlag, J., Schlag-Rey, M. & Arbib, M. A. (1997a) Colliding saccades evoked by frontal eye field stimulation: Artifact of evidence for an oculomotor compensatory mechanism? *Biological Cybernetics* 76:41–52. [PFD]

Dominey, P. F., Ventre-Dominey, J., Broussolle, E. & Jeannerod, M. (1995b) Analogical transfer in sequence learning: Human and neural-network models of fronto-striatal function. *Annals of the New York Academy of Science* 769:369–73. [PFD]

(1997b) Analogical transfer is effective in a serial reaction time task in Parkinson's disease: Evidence for a dissociable sequence learning mechanism. *Neuropsychologia* 35(1):1–9. [PFD]

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

Duncan, J., Ward, R. & Shapiro, K. (1994) Direct measurement of attention dwell time in human vision. *Nature* 369:313–14. [aDHB]

Epelboim, J., Booth, J. R. & Steinman, R. M. (1994) Reading unspaced text: Implications for theories of reading eye movements. *Vision Research* 34:1735–66. [JMFi]

Epelboim, J., Steinman, R. M., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C. J. & Collewijn, H. (1995) The function of visual search and memory in sequential looking tasks. *Vision Research* 35:3401–22. [JE]

Epelboim, J. & Suppes, P. (1997) Eye movements during geometrical problem solving. *Proceedings of the Cognitive Science Society*. [JE]

Findlay, J. M. (1981) Local and global influences on saccadic eye movements. In: *Eye movements, cognition and visual perception*, ed. D. F. Fisher, R. A. Monty & J. W. Senders. Lawrence Erlbaum. [JMFi]

(1995) Visual search: Eye movements and peripheral vision. *Optometry and Vision Science* 72:461–66. [rDHB, JMFi]

(1997) Saccade target selection in visual search. *Vision Research* 37:617–31. [JMFi]

Fischer, B. & Weber, H. (1993) Express saccades and visual attention. *Behavioral and Brain Sciences* 16:553–610. [BMV]

Fischer, M. H. (submitted) Poor memory for word locations in reading. [MHF]

Flook, J. P. & McGonigle, B. (1977) Serial adaptation to conflicting prismatic arrangement effects in monkey and man. *Perception* 6:15–19. [BM]

Freeman, W. T. & Adelson, E. H. (1991) The design and use of steerable filters. *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence* 13(9):891–906. [aDHB]

Fuster, J. M. (1989) *The prefrontal cortex: Anatomy, physiology, and neuropsychology of the frontal lobe*, 2nd ed. Raven Press. [aDHB]

(1995) *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate.* MIT Press/Bradford Books. [aDHB, JMFu]

Gabrieli, J. (1995) Contribution of the basal ganglia to skill learning and working memory in humans. In: *Models in information processing in the basal ganglia*, ed. J. C. Houk, J. L. Davis & D. G. Beiser. MIT Press/Bradford Books. [aDHB]

Gallant, J. L., Connor, C. E., Drury, H. & Van Essen, D. (1995) Neural responses in monkey visual cortex during free viewing of natural scenes: Mechanisms of response suppression. *Investigative Ophthalmology and Visual Science* 36:1052. [aDHB]

Gallistel, C. (in press) Coordinate transformations in the genesis of directed action. In: *Handbook of cognition and perception: Cognitive science*, ed., D. E. Rumelhart & B. O. Martin. [SB]

Geisler, W. S. & Chou, K.-L. (1995) Separation of low-level and high-level factors in complex tasks: Visual search. *Psychological Review* 102(2):356–78. [rDHB]

Gibson, J. J. (1979) *The ecological approach to visual perception.* Houghton-Mifflin. [AK]

Glenberg, A. M. (1997) What memory is for. *Behavioral and Brain Sciences* 20(1):1–19. [AMG]

Glenberg, A. M., Kruley, P. & Langston, W. E. (1994) Analogical processes in comprehension: Simulation of a mental model. In: *Handbook of psycholinguistics*, ed. M. A. Gernsbacher. Academic Press. [AMG]

Goldberg, D. E. (1989) *Genetic algorithms in search, optimization, and machine learning.* Addison-Wesley. [aDHB]

Goldman-Rakic, P. S. (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: *Handbook of physiology – The nervous system V*, ed. F. Plum. American Physiology Society. [GWS]

(1995) Toward a circuit model of working memory and the guidance of voluntary motor action. In: *Models in information processing in the basal ganglia*, ed. J. C. Houk, J. L. Davis & D. G. Beiser. MIT Press/Bradford Books. [aDHB]

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

Graybiel, A. M. & Kimura, M. (1995) Adaptive neural networks in the basal ganglia. In: *Models in information processing in the basal ganglia*, ed. J. C. Houk, J. L. Davis & D. G. Beiser. MIT Press/Bradford Books. [aDHB]

Grimes, J. & McConkie, G. (1996) On the insensitivity of the human visual system to image changes made during saccades. In: *Problems in perception*, ed. K. A. Akins. Oxford University Press. [arDHB]

Hancock, P. J. B., Baddeley, R. J. & Smith, L. S. (1992) The principal components of natural images. *Network* 3:61–70. [aDHB]

Harnad, S. (1990) The symbol grounding problem. *Physica* D 42:335–46. [AMG]

(1992) Connecting object to symbol in modelling cognition. In: *Connectionism in context*, ed. A. Clark & R. Lutz. Springer-Verlag. [RID]

Hayhoe, M. M., Bensinger, D. G. & Ballard, D. H. (1997) Visual memory in natural tasks. *Vision Research*. [rDHB]

(in press) Task constraints in visual working memory. *Vision Research*. [rDHB]

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

Hayhoe, M. M., Lachter, J. & Möller, P. (1992) Spatial memory and integration across saccadic eye movements. In: *Eye movements and visual cognition: Scene perception and reading*, ed. K. Rayner. Springer-Verlag. [aDHB]

Helms-Tillery, S. I., Flanders, M. & Soechting, J. F. (1991) A coordinate system for the synthesis of visual and kinesthetic information. *Journal of Neuroscience* 11(3):770–78. [aDHB]

Hershberger, W. (1987) Saccadic eye movements and the perception of visual direction. *Perception & Psychophysics* 41:39. [JSJ]

Hershberger, W. A. & Jordan, J. S. (1996) The phantom array. *Behavioral and Brain Sciences* 19:552–55. [JSJ]

(in press) The phantom array: A peri-saccadic illusion of visual direction. *The Psychological Record*. [JSJ]

Hertz, J., Krogh, A. & Palmer, R. G. (1991) *Introduction to the theory of neural computation*, vol. 1, Santa Fe Institute Studies in the Sciences of Complexity Lecture Notes. Addison-Wesley. [aDHB]

Hinton, G. E. (1981) A parallel computation that assigns canonical object-based frames of reference. *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, August: 683–85. [aDHB]

Hitch, G. J. (1978) The role of short-term working memory in mental arithmetic. *Cognitive Psychology* 10:305–28. [PPM]

Hoffman, J. E. & Mueller, S. (submitted) Visuo-spatial attention in three-dimensional space. [BMV]

Hooge, I. T. C. (1996) Control of eye movements in visual search. PhD thesis, University of Utrecht, Netherlands. [rDHB]

Horswill, I. D. (1995) Visual routines and visual search. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, August. [JB]

Houk, J. C., Davis, J. L. & Beiser, D. G. eds. (1995) *Models in information processing in the basal ganglia.* MIT Press/Bradford Books. [aDHB]

Hunter, W. S. (1913) The delayed reaction in animals and children. *Behavior Monographs* 2.

Husbands, P., Harvey, I. & Cliff, D. (1995) Circle in the round: State space attractors for evolved sight robots. *Robotics and Autonomous Systems* 15:83–106. [JCR]

Hutchins, E. (1995a) *Cognition in the wild.* MIT Press. [AK]

(1995b) How a cockpit remembers its speeds. *Cognitive Science* 19:269–88. [PPM]

Irwin, D. E. (1991) Information integration across saccadic eye movements. *Cognitive Psychology* 23:420–56. [aDHB]

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

Jeannerod, M. (1988) *The neural and behavioural organization of goal-directed movements.* Clarendon Press. [aDHB]

Johnson, M. (1987) *The body in the mind: The bodily basis of reason and imagination.* University of Chicago Press. [JAF]

Johnston, W. A. & Dark, V. J. Selective attention. *Annual Review of Psychology* 37:43–75. [aDHB]

Jones, D. G. & Malik, J. (1992) A computational framework for determining stereo correspondence from a set of linear spatial filters. *Proceedings of the 2nd European Conference on Computer Vision.* [aDHB]

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S. & Mintun, M. A. (1993) Spatial working memory in humans as revealed by PET. *Nature* 363:June. [aDHB]

Jordan, J. S. (submitted) Recasting Dewey's critique of the reflex-arc concept via Vandervert's anticipatory theory of consciousness: Implications for theories of perception. [JSJ]
Jordan, J. S. & Hershberger, W. A. (1994) Timing the shift in retinal local signs that accompanies a saccadic eye movement. *Perception & Psychophysics* 55(6):657–66. [JSJ]
Joseph, J. S., Chun, M. M. & K. Nakayama, K. (submitted) Attentional requirements in a 'preattentive' feature search task. [aDHB]
Just, M. A. & Carpenter, P. A. (1976) Eye fixations and cognitive processes. *Cognitive Psychology* 8:441–80. [aDHB]
(1992) A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99(1):122–49. [aDHB]
Jüttner, M. (1997) Effects of perceptual context of transsaccadic visual matching. *Perception & Psychophysics*, in press. [MJ]
Jüttner, M. & Rentschler, I. (1996) Reduced perceptual dimensionality in extrafoveal vision. *Vision Research* 36:1007–21. [MJ]
(submitted) Uniqueness of foveal vision revealed by classification learning. [MJ]
Jüttner, M., Rentschler, I. & Unzicker, A. (1996) Shift-invariance of pattern recognition in the visual field? *Perception* 25(Suppl.):1. [rDHB, MJ]
Jüttner, M. & Röhler, R. (1993) Lateral information transfer across saccadic eye movements. *Perception & Psychophysics* 53:210–20. [rDHB, MJ]
Kanerva, P. (1988) *Sparse distributed memory.* MIT Press/Bradford Books. [aDHB]
Kapoula, Z. (1984) Aiming precision and characteristics of saccades. In: *Theoretical and applied aspects of oculomotor research*, ed. A. G. Gale & F. W. Johnson. Elsevier. [JMFi]
Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M. & Sagi, D. (1994) Dependence on REM sleep of overnight improvement of a perpetual skill. *Science* 265:679–82. [rDHB]
Kimura, M, Aosaki, T., Hu, Y., Ishida, A. & 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* 89:473–77. [aDHB]
Kirsh, D. (1995) The intelligent use of space. *Artificial Intelligence* 73:31–68. [PPM]
Kirsh, D. & Maglio, P. (1994) On distinguishing epistemic from pragmatic action. *Cognitive Science* 18:513–49. [PPM]
Koch, C. & Crick, F. (1994) Some further ideas regarding the neuronal basis of awareness. In: *Large-scale neuronal theories of the brain*, ed. C. Koch & J. L. Davis. MIT Press/Bradford Books. [aDHB]
Kosslyn, S. M. (1994) *Image and brain.* MIT Press/Bradford Books. [aDHB]
Kowler, E. & Anton, S. (1987) Reading twisted text: Implications for the role of saccades. *Vision Research* 27:45–60. [aDHB]
Koza, J. R. (1992) *Genetic programming: On the programming of computers by means of natural selection.* MIT Press. [aDHB]
Lachter, J. & Hayhoe, M. M. (1996) Memory limits in integration across saccades. *Perception and Psychophysics*, in press. [aDHB]
Lakoff, G. (1987) *Women, fire, and dangerous things: What categories reveal about the mind.* University of Chicago Press. [rDHB]
Land, M. F. & Lee, D.N. (1994) Where we look when we steer. *Nature* 369:742–44. [aDHB]
Langston, W., Kramer, D. C. & Glenberg, A. M. (in press) The representation of space in mental models derived from text. *Memory and Cognition.* [AMG]
Lee, D. N. (1976) A theory of visual control of braking based on information about time-to-collision. *Perception* 5:437–59. [rDHB]
Levin, H. (1979) *The eye-voice span.* MIT Press. [MHF]
Liberman, A. M. (1966) *Speech: A special code.* Bradford Books/MIT Press. [RID]
Logie, R. H. (1995) *Visuo-spatial working memory.* Lawrence Erlbaum. [aDHB]
Luria, A. R. (1968) *The mind of a mnemonist: A little book about a vast memory.* Harvard University Press. [aDHB]
Maglio, P. P. (1995) The computational basis of interactive skill. Doctoral dissertation, University of California, San Diego. [PPM]
Maglio, P. P. & Kirsh, D. (1996) Epistemic action increases with skill. In: *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society.* Lawrence Erlbaum. [PPM]
Magnuson, J. S., Sagerer, G., Hayhoe, M. M. & Ballard, D. H. (1997) The role of fixations in task automatization. *Investigative Ophthalmology and Visual Science* (Suppl.). [rDHB]
Mani, K. & Johnson-Laird, P. N. (1982) The mental representation of spatial descriptions. *Memory and Cognition* 10:181–87. [AMG]
Marr, D. C. (1982) *Vision.* W. H. Freeman. [aDHB]
Matin, L., Pearce, D., Matin, E. & Kibler, G. (1966) Visual perception of direction in the dark: Roles of local sign, eye movements, and ocular proprioception. *Vision Research* 6:453–69. [JSJ]
Maunsell, J. H. R. (1993) Oral presentation, Woods Hole Workshop on Computational Neuroscience, September. [aDHB]
McCallum, A. K. (1995) Reinforcement learning with selective perception and hidden state. Ph.D. thesis, Computer Science Dept., University of Rochester. [aDHB]
(1996) Learning to use selective attention and short-term memory in sequential tasks. In: *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior (SAB).* MIT Press. [rDHB]
McCallum, R. A. (1994) Reduced training time for reinforcement learning with hidden state. *Proceedings of the Eleventh International Machine Learning Workshop (Robot Learning).* [aDHB]
(1995) Instance-based utile distinctions for reinforcement learning. *Proceedings of the Twelfth International Machine Learning Conference.* Morgan Kaufmann. [arDHB]
McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. (1995) Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* 102(3):419–57. [JB]
McGonigle, B. & Chalmers, M. (1996) The ontology of order. In: *Critical readings on Piaget*, ed. L. Smith. Routledge. [BM]
McGonigle, B. & Flook, J. (1978) Long-term retention of single and multistate prismatic adaptation by humans. *Nature* 272:364–66. [BM]
McGonigle, B. & Jones, B. T. (1978) Levels of stimulus processing by the squirrel monkey: Relative and absolute judgements compared. *Perception* 7:635–59. [BM]
Melzack, R. (1992) Phantom limbs. *Scientific American* 266:120–26. [JSJ]
Milner, A. D. & Goodale, M. A. (1995) *The visual brain in action.* Oxford University Press. [arDHB, MG, BMV]
Miyachi, S., Miyashita, K., Karadi, Z. & Hikosaka, O. (1994) Effects of the blockade of monkey basal ganglia on the procedural learning and memory. *Abstracts, 24th Annual Meeting, Society for Neuroscience* (153.6):357. [aDHB]
Miyashita, Y. (1995) How the brain creates imagery: Projection to primary visual cortex. *Science* 268:1719–20. [GWS]
Montague, P. R., Dayan, P. & Sejnowski, T. J. (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience* 16(5):1936–47. [rDHB]
Munoz, D. P. & Wurtz, R. H. (1993) Fixation cells in monkey superior colliculus. I. Characteristics of cell discharge. *Journal of Neurophysiology* 70:559–75. [JMFi]
Nakayama, K. (1990) The iconic bottleneck and the tenuous link between early visual processing and perception. In: *Vision: Coding and efficiency*, ed. C. Blakemore. Cambridge University Press. [aDHB]
Nazir, T. A. & O'Regan, J. K. (1990) Some results on translation invariance in the human visual system. *Spatial Vision* 5:1–19. [MJ]
Neisser, U. & Kerr, N. (1973) Spatial and mnemonic properties of memory images. *Cognitive Psychology* 5:138–50. [BMV]
Newell, A. (1990) *Unified theories of cognition.* Harvard University Press. [aDHB, RID]
Norman, D. A. & Bobrow, D. G. (1975) On data-limited and resource-limited processes. *Cognitive Psychology* 7:44–64. [aDHB]
Noton, D. & Stark, L. (1971a) Eye movements and visual perception. *Scientific American* 224:34–43. [aDHB]
(1971b) Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research* 11:929. [aDHB]
Olson, C. R. & Gettner, S. N. (1995) Object-centered direction selectivity in the macaque supplementary eye field. *Science* 269:985–88. [aDHB]
O'Regan, J. K. (1992) Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology* 46:461–88. [aDHB, PPM]
O'Regan, J. K. & Lévy-Schoen, A. (1983) Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research* 23:765–69. [aDHB]
Ottes, F. P., Van Gisbergen, J. A. M. & Eggermont, J. J. (1985) Latency dependence of colour-based target vs. nontarget discrimination by the saccadic system. *Vision Research* 25:849–62. [JMFi]
Palmer, J., Ames, C. & Lindsey, D. (1993) Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance* 19:108–30. [rDHB]
Pashler, H. (1987) Detecting conjunction of color and form: Re-assessing the serial search hypothesis. *Perception and Psychophysics* 41:191–201. [JMFi]
Paulesu, E., Frith, C. D. & Frackowiak, R. S. J. (1993) The neural correlates of the verbal component of working memory. *Nature* 362:342–45. [aDHB]
Pellizzer, G., Sargent, P. & Georgopoulos, A.P. (1994) Motor cortex and visuomotor memory scanning. *Abstracts, 24th Annual Meeting, Society for Neuroscience* (403.12):983. [aDHB]
Pelz, J. B. (1995) Visual representations in a natural visuo-motor task. Ph.D.

thesis, Brain and Cognitive Sciences Department, University of Rochester. [aDHB]

Perner, J. (199X) The meta-intentional nature of executive functions and theory of mind. In: *Language and thought*, ed. P. Curruthers & J. Boucher. Cambridge University Press. [BMV]

Perrett, I. D. (1996) View-dependent coding in the ventral stream and its consequences for recognition. In: *Vision and movement: Mechanisms in the cerebral cortex*, ed. R. Caminiti, K.-P. Hoffman, & F. Lacquaniti. HFSP (Human Frontier Science Program). [JB]

Phillips, W. A. & Singer, W. (in press) In search of common cortical foundations. *Behavioral and Brain Sciences* 20(4). [JB]

Pollatsek, A. & Rayner, K. (1990) Eye movements and lexical access in reading. In: *Comprehension processes in reading*, ed. D. A. Balota, G. B. Flores d'Arcais & K. Rayner. Lawrence Erlbaum. [aDHB]

Pook, P. K. (1995) Teleassistance: Using deictic gestures to control robot action. Ph.D. thesis and TR 594, Computer Science Department, University of Rochester, September. [aDHB]

Pook, P. K. & Ballard, D.H. (1994a) Deictic teleassistance. *Proceedings of the Institute of Electrical and Electronics Engineers International Conference on Intelligent Robots and Systems*, Munich. [aDHB]

(1994b) Teleassistance: Contextual guidance for autonomous manipulation. *Proceedings 12th National Conference on Artificial Intelligence (American Association for Artificial Intelligence).* [aDHB]

Pöppel, E. (1994) Temporal mechanisms in perception. *International Review of Neurobiology* 37:185–202. [rDHB, JB]

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

Raichle, M.E. (1993) Neuropsychology – the scratchpad of the mind. *Nature* 363(6430):583–84. [aDHB]

Rao, R. P. N. & Ballard, D. H. (1995) An active vision architecture based on iconic representations. *Artificial Intelligence* 78:461–505. [aDHB]

(1996a) A class of stochastic models for invariant recognition, motion, and stereo. NRL TR 96.1, National Resource Laboratory for the Study of Brain and Behavior, Computer Science Department, University of Rochester, June. [aDHB]

(1996b) A computational model of spatial representations that explains object-centered neglect in parietal patients. In: *Computational neuroscience* '96, ed. J. Bower. Plenum Press, in press. [aDHB]

(1997) Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation* 9(4):721–63. [arDHB]

Rao, R. P. N. & Fuentes, O. (1996) Learning navigational behaviors using a predictive sparse distributed memory. In: *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior (SAB).* MIT Press. [rDHB]

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M. & Ballard, D. H. (1996) Modeling saccadic targeting in visual search. In: *Advances in neural information processing systems 8*, ed. D. Touretzky, M. Mozer & M. Hasselmo. MIT Press. [aDHB]

(1997) Eye movements in visual cognition: A computational study. Technical Report 97.1, National Resource Laboratory for the Study of Brain and Behavior, Department of Computer Science, University of Rochester, March. [rDHB]

Rayner, K. & Pollatsek, A. (1989) *The psychology of reading.* Prentice-Hall. [MHF]

Rensink, R., O'Regan, J. & Clark, J. (1996) To see or not to see: The need for attention to perceive changes in scenes. *Investigative Ophthalmology and Visual Science* (Suppl.)37:213. [rDHB]

Rinck, M. & Bower, G. H. (1995) Anaphora resolution and the focus of attention in situation models. *Journal of Memory and Language* 34:110–31. [AMG]

Rinck, M., Hähnel, A., Bower, G. & Glowalla, U. (1997) The metrics of spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:622–37. [AMG]

Romo, R. & Schultz, W. (1990) Dopamine neurons of the monkey midbrain: Contingencies of responses to active touch during self-initiated arm movements. *Journal of Neurophysiology* 63:592–606. [WS]

Rosch, E. (1978) Principles of categorization. In: *Cognition and categorization*, ed. E. Rosch & B. Lloyd. Lawrence Erlbaum. [MJ]

Rosenbaum, D. A., Kenny, S. & Derr, M. A. (1983) Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance* 9:86–102. [SB]

Rugg, M. D. & Coles, M. G. H. (1995) *Electrophysiology of mind: Event related brain potentials and cognition.* Oxford University Press. [BMV]

Samuel, A. G. (1986) The role of the lexicon in speech perception. In: *Pattern recognition by humans and machines*, vol. 1, *Speech perception*, ed. E. C. Schwab & H. C. Nusbaum. Academic Press. [RID]

Sawusch, J. R. (1986) Auditory and phonetic coding of speech. In: *Pattern recognition by humans and machines*, vol. 1, *Speech perception*, ed. E. C. Schwab & H. C. Nusbaum. Academic Press. [RID]

Schlag, J. D. & Schlag-Rey, M. (1990) Colliding saccades may reveal the secret of their marching orders. *Trends in Neuroscience* 13:410–15. [PFD]

Schlingensiepen, K.-H, Campbell, F. W., Legge, G. E. & Walker, T. D. (1986) The importance of eye movements in the analysis of simple patterns. *Vision Research* 26:1111–17. [aDHB]

Schneider, W., Dumais, S. T. & Shiffrin, R. M. (1984) Automatic and control processing and attention. In: *Varieties of attention.* Academic Press. [aDHB]

Schultz, W., Apicella, P., Romo, R. & Scarnati, E. (1995) Context-dependent activity in primate striatum reflecting past and future behavioral events. In: *Models in information processing in the basal ganglia*, ed. J. C. Houk, J. L. Davis & D. G. Beiser. MIT Press/Bradford Books. [aDHB]

Schultz, W., Romo, R., Scarnati, E., Sundström, E., Jonsson, G. & Studer, A. (1989) Saccadic reaction times, eye–arm coordination and spontaneous eye movements in normal and MPTP-treated monkeys. *Experimental Brain Research* 78:252–67. [WS]

Shanon, B. (1988) Semantic representation of meaning: A critique. *Psychological Bulletin* 104:70–83. [AMG]

Shastri, L. (1993) From simple associations to systematic reasoning. *Behavioral and Brain Sciences* 16(3):417–94. [aDHB]

Shastri, L. & Ajjanagadde, V. (1993) From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences* 16(3):417–94. [rDHB]

Simon, H. A. (1962) The architecture of complexity. *Proceedings of the American Philosophical Society* 26:467–82. [aDHB]

Singer, W. & Gray, C. M. (1995) Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience* 18:555–86. [rDHB]

Smithers, T. (1994) On why better robots make it harder. In: *From animals to animats 3*, ed. D. Cliff, P. Husbands, J.-A. Meyer & S. W. Wilson. MIT Press/Bradford Books. [JCR]

Snyder, L. H. & Andersen, R. A. (1994) Effects of vestibular and neck proprioceptive signals on visual responses in posterior parietal cortex. *Abstracts, 24th Annual Meeting, Society for Neuroscience* (525.1):1278. [aDHB]

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

Stevens, J. K., Emerson, R. C., Gerstein, G. L., Kallos, T., Neufeld, G. R., Nichols, C. W. & Rosenquist, A. C. (1976) Paralysis of the awake human: Visual perceptions. *Vision Research* 16:93–98. [JSJ]

Strasburger, H., Rentschler, I. & Harvey, L. O., Jr. (1994) Cortical magnification theory fails to predict visual recognition. *European Journal of Neuroscience* 6:1583–88. [MJ]

Strasburger, H. & Rentschler, I. (1996) Contrast-dependent dissociation of visual recognition and detection fields. *European Journal of Neuroscience* 8:1787–91. [MJ]

Stricanne, B., Xing, J., Mazzoni, P. & Andersen, R.A. (1994) Response of LIP neurons to auditory targets for saccadic eye movements: A distributed coding for sensorimotor transformation. *Abstracts, 24th Annual Meeting, Society for Neuroscience* (65.1):143. [aDHB]

Strick, P. L., Dum, R. P. & Picard, N. (1995) Macro-organization of the circuits connecting the basal ganglia with the cortical motor areas. In: *Models in information processing in the basal ganglia*, ed. J. C. Houk, J. L. Davis & D. G. Beiser. MIT Press/Bradford Books. [aDHB]

Strong, G. W. & Whitehead, B. A. (1989) A solution to the tag assignment problem for neural networks. *Behavioral and Brain Sciences* 12:381–433. [GWS]

Swain, M. J. & Ballard, D.H. (1991) Color indexing. *International Journal of Computer Vision* 7(1):11–32. [aDHB]

Tagaris, G., Kim, S.-G., Menon, R., Strupp, J., Andersen, P., Ugurbil, K. & Georgopoulos, A. P. (1994) High field (4 Telsa) functional MRI of mental rotation. *Abstracts, 24th Annual Meeting, Society for Neuroscience* (152.10):353. [aDHB]

Tanaka, K. (1996) Inferotemporal cortex and object recognition. In: *Vision and movement: Mechanisms in the cerebral cortex*, ed. R. Caminiti, K.-P. Hoffman & F. Lacquaniti. HFSP (Human Frontier Science Program). [JB]

Thagard, P., Holyoak, K. J., Nelson, G. & Gochfeld, D. (1990) Analog retrieval by constraint satisfaction. *Artificial Intelligence* 46:259–310. [PFD]

Toet, A. & Levi, D. M. (1992) Spatial interaction zones in the parafovea. *Vision Research* 32:1349–57. [JMFi]

Treisman, A. & Gelade, G. (1980) A feature integration theory of attention. *Cognitive Psychology* 12:97–136. [JMFi]

Trick, L. M. & Pylyshyn, Z. W. (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. [aDHB]

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N. & Nuflo, F.

(1995) Modeling visual attention via selective tuning. *Artificial Intelligence* (Special Issue on Vision) 78:507–45. [aDHB]
Ullman, S. (1984) Visual routines. *Cognition* 18:97–157. [aDHB, SB]
Ungerleider, L. G. (1995) Functional brain imaging studies of cortical mechanisms for memory. *Science* 270:769–75. [GWS]
Ungerleider, L. G. & Mishkin, M. (1982) Two cortical visual systems. In: *Analysis of visual behavior,* ed. D. J. Ingle, M. A. Goodale & R. J. W. Mansfield. MIT Press. [aDHB, MG]
Van Essen, D. C., Anderson, C. H. & Olshausen, B. A. (1994) Dynamic routing strategies in sensory, motor, and cognitive processing. In: *Large-scale neuronal theories of the brain,* ed. C. Koch & J. L. Davis. MIT Press/Bradford Books. [aDHB]
Varela, F. J., Thompson, E. & Rosch, E. (1991) *The embodied mind.* MIT Press/Bradford Books. [JCR]
Velichkovsky, B. M. (1995) Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics and Cognition* 3(2):199–222. [BMV]
Velichkovsky, B. M., Sprenger, A. & Unema, P. (1997) Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem." In: *Human–computer interaction: INTERACT '97,* ed. S. Howard, J. Hammond & G. Lindgaard. Chapman & Hall. [BMV]
Velichkovsky, B. M. & Van der Heijden, A. H. C. (1994) Space as reference signal? Elaborate it in depth! *Behavioral and Brain Sciences* 17(2):337–38. [BMV]
Vitu, F., O'Regan, J. K., Inhoff, A. W. & Topolski, R. (1995) Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception and Psychophysics* 57:352–64. [JMFi]
Viviani, P. (1990) Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In: *Eye movements and their role in visual and cognitive processes,* ed. E. Kowler. Reviews of oculomotor research V4. Elsevier Science. [aDHB]
von der Malsburg, C. (1995) Binding in models of perception and brain function. *Current Opinion in Neurobiology* 5:520–26. [rDHB, JB]
Walker-Smith, G. J., Gale, A. G. & Findlay, J. M. (1977) Eye movement strategies involved in face perception. *Perception* 6:313–26. [JMFi]
Wandell, B. (1995) *Foundations of vision science: Behavior, neuroscience, and computation.* Sinauer, Sunderland. [aDHB]
Ward, R., Duncan, J. & Shapiro, K. (1996) The slow time-course of visual attention. *Cognitive Psychology* 30:79–109. [aDHB]
Watanabe, S. (1985) *Pattern recognition: Human and mechanical.* Wiley. [MJ]
Wertheim, T. (1894) Über die indirekte Sehschärfe. *Zeitschrift für Psychologie und Physiologie der Sinnesorgans* 7:121–87. [JMFi]
Whitehead, S. D. & Ballard, D. H. (1990) Active perception and reinforcement learning. *Neural Computation* 2(4):409–19. [aDHB]
(1991) Learning to perceive and act by trial and error. *Machine Learning* 7(1):45–83. [arDHB]
Wiskott, L. & von der Malsburg, C. (1993) A neural system for the recognition of partially occluded objects in cluttered scenes: A pilot study. *International Journal of Pattern Recognition and Artificial Intelligence* 7:935 – 48. [aDHB]
Wolfe, J. M. (1996a) Post-attentive vision. *Investigative Ophthalmology and Visual Science* Suppl. 37:214. [aDHB]
(1996b) Visual search. In: *Attention,* ed. H. Pashler. University College London Press. [aDHB]
Wong, E. & Mack, A. (1981) Saccadic programming and perceived location. *Acta Psychologica* 48:123–31. [BMV]
Woodward, D. J., Kirillov, A. B., Myre, C. D. & Sawyer, S. F. (1995) Neostriatal circuitry as a scalar memory: Modeling and ensemble neuron recording. In: *Models in information processing in the basal ganglia,* ed. J. C. Houk, J. L. Davis & D. G. Beiser. MIT Press/Bradford Books. [aDHB]
Yarbus, A. L. (1967) *Eye movements and vision.* Plenum Press. [aDHB]
Yeterian, E. H. & Pandya, D. N. (1995) Corticostriatal connections of extrastriate visual areas in rhesus monkeys. *Journal of Comparative Neurology* 352:436–57. [aDHB]
Zelinsky, G., Rao, R. P. N., Hayhoe, M. M. & Ballard, D. H. (in press) Eye movements reveal the spatio-temporal dynamics of visual search. *Psychological Science.* [rDHB]
Zinchenko, V. P., Chzhi-Tsin, S. & Taralonov, A. (1963) The formation and development of perceptual activity. *Soviet Psychology and Psychiatry* 2:3–12. [BM]

https://doi.org/10.1017/S0140525X97311613 Published online by Cambridge University Press