Deictic Codes for the Embodiment of Cognition

Ballard, Hayhoe, Pook, & Rao

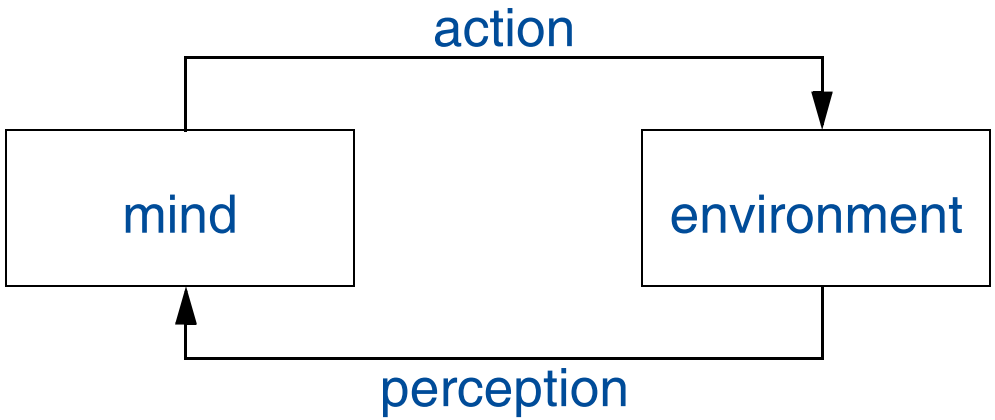
Saccadic Eye Movements

Versus smooth pursuit, vergence

200-300 milliseconds to plan and execute

Reveal attention, goals, cognitive operations

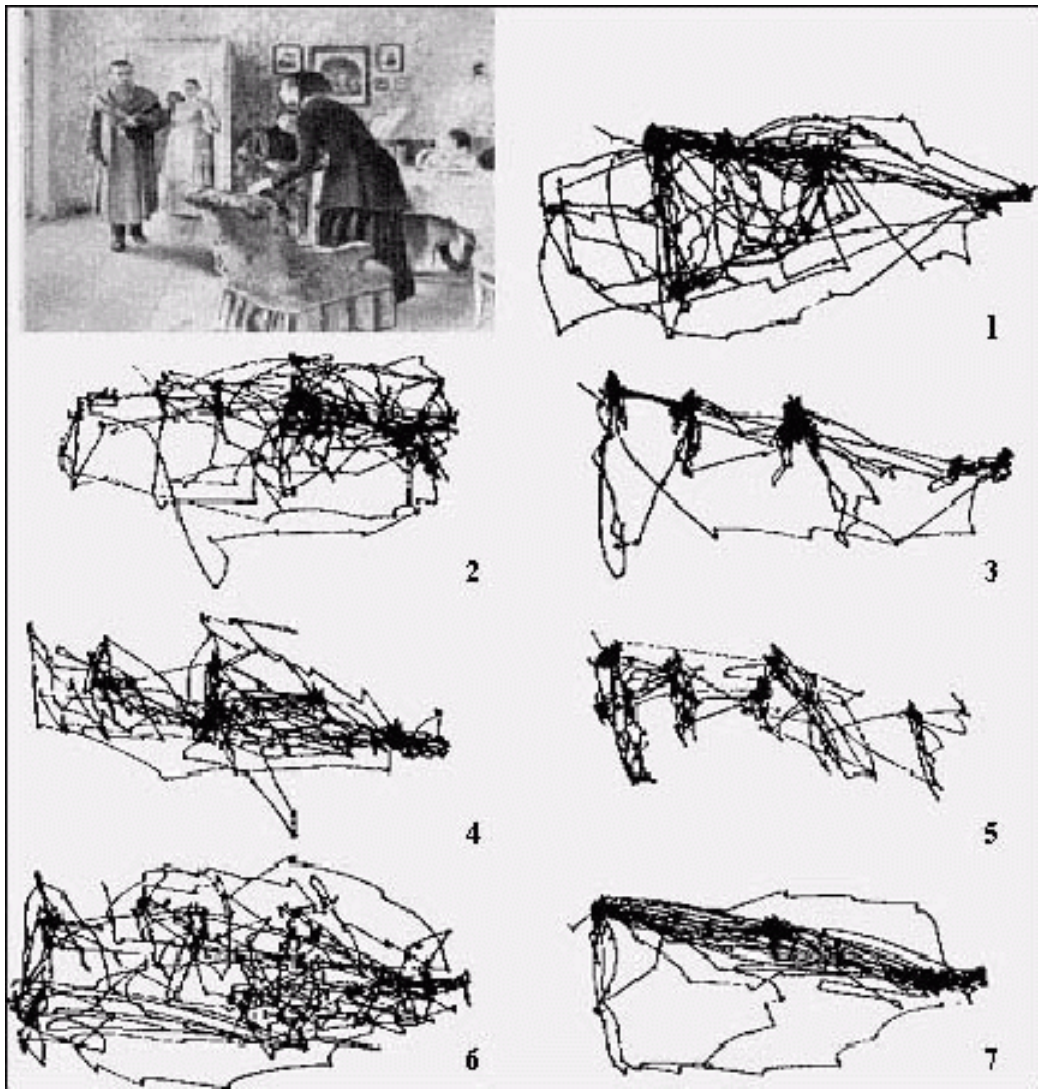
Finest grain behavior of perception-action cycle



Yarbus (1961)

3 minutes viewing

1. free examination
2. material circumstances of family
3. ages
4. what had family been doing before visitor's arrival?
5. remember clothes
6. remember position of people & objects
7. how long had visitor been away



Deictic Strategies

deictic = pointing, showing

deictic strategy = "do it where I'm pointing/looking"

- constrains what object is to be acted on
- nondeictic strategy involves building a representation of entire scene

By always acting on the object at fixation, perception and motor control are simplified.

- Brings viewer-centered frame and object-centered frames into alignment
- Brings object into high resolution region of retina
- Action occurs with respect to a fixed perceptual reference frame
- Removes ambiguity by specifying what object/feature is to be acted upon [see *pointers* and *binding*]

Pointers

To understand computational role of deictic reference, need to understand the concept of *pointers*.

Real world: house addresses versus house occupants

Computer programs:

add 3, 47

add 0004, 0003

add B, C

0000	12
0001	19
0002	0
0003	47
0004	3
0005	1
0006	2
0007	2

Instead of pointing to locations on the street or in memory, deictic reference provides pointers to *locations in the world*.

Variable Binding

The mind needs to represent general knowledge without reference to specific objects.

If Rob intentionally hurts Mike, then Rob does not like Mike.
If Becky intentionally hurts Laura, then Becky does not like Laura.

If X intentionally hurts Y, then X does not like Y.

intentionally-hurt(X,Y) -> dislike(X,Y)

True not only of facts/inferences, but also *scripts* and *plans.*

Make breakfast by pouring cereal and milk into a bowl.

milk(M)	pour(M,B)
cereal(C)	pour(C,B)
bowl(B)	consume(contents-of(B))

X, Y, M, etc. are *variables* that can be *bound* to world objects.

Deictic strategies (fixation) bind a variable to world objects via locations in space (vs. use of working memory).

Examples of Deictic Reference in Perception

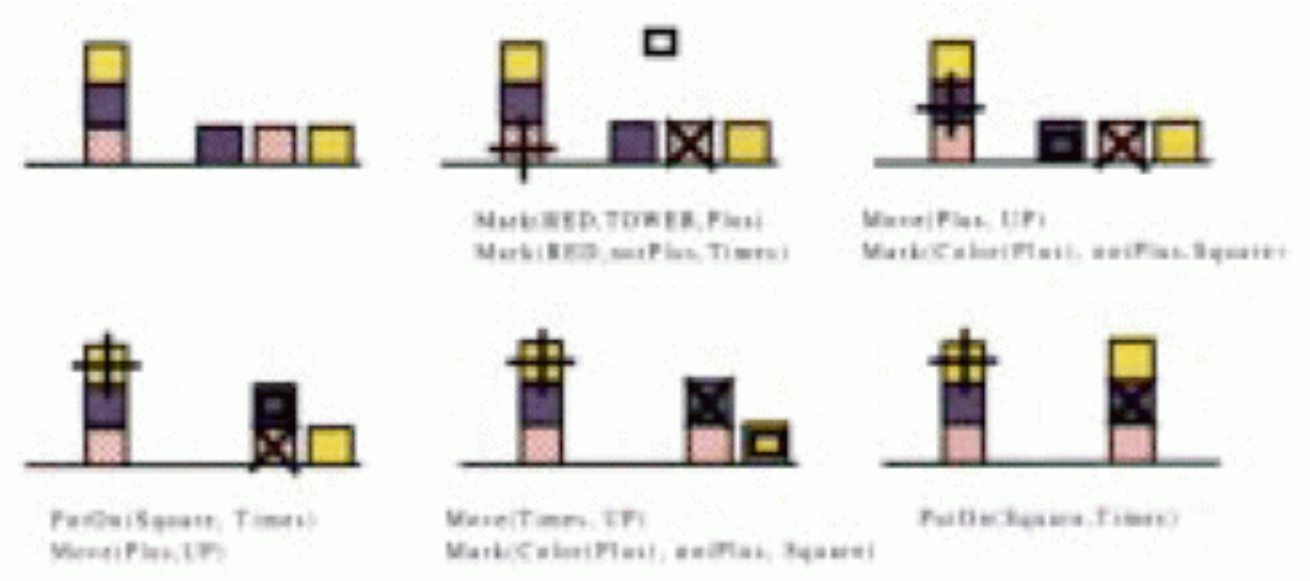
Human data

e.g., reading

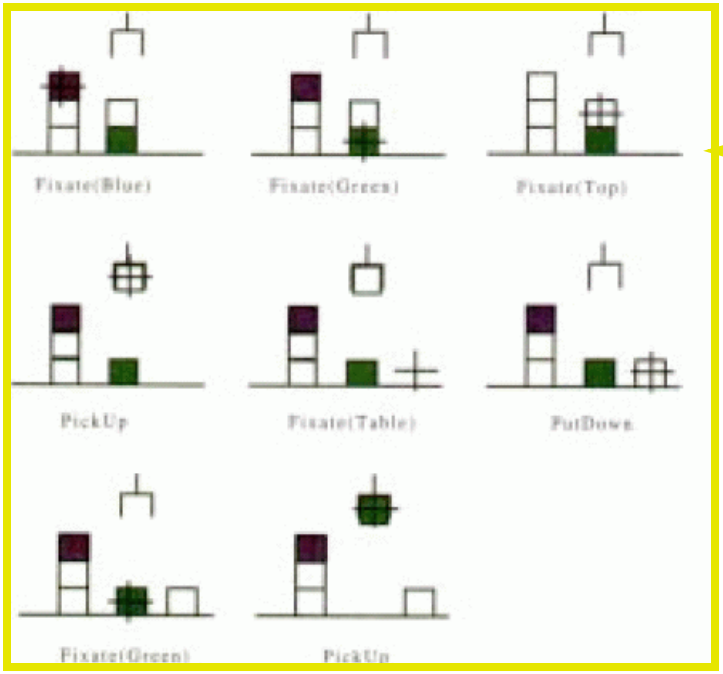
e.g., sign language

AI model (Chapman, 1989)

+ block in the template being copied
x block in copy being moved onto
o block being moved



Whitehead and Ballard Model



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

Model has two pointers: fixation and attention

Need to have more context for decision making than knowing what is at fixation

e.g., compare action from step 2->3 vs. from step 7->8

Whitehead and Ballard Model

Trained with reinforcement learning

RL learns a policy: mapping from states to actions

Trial and error learning

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

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)

Deictic Representation

Use the body's pointing movements to bind variables in cognitive programs to objects in the world.

What does this buy cognition?

- Flexibility of programs with variables
- Provide a common focus to perception, decision making, and action
- Scales well

  e.g., Chapman model will work with stacks of arbitrary height

Why a small number of pointers?

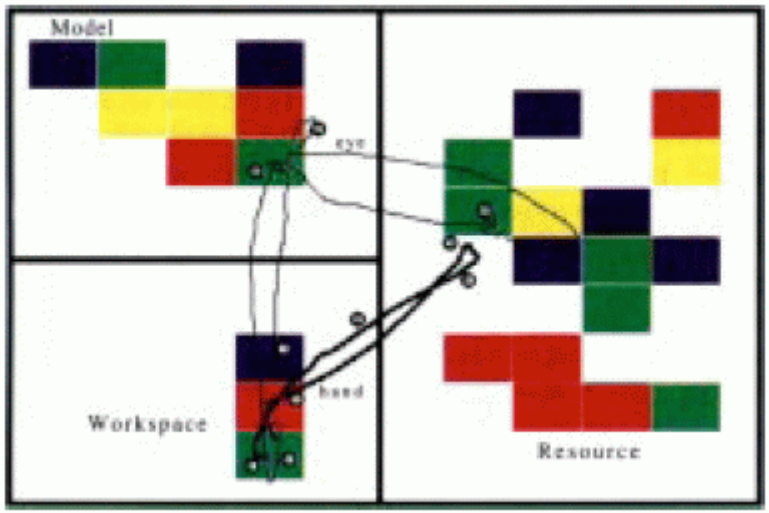
(visual STM ~ 3 items)

- most tasks don't require a lot of contextual information (McCallum 1995)
- (reinforcement) learning scales poorly with the # of pointers: $(MV)^s$, where s=steps in program, M = # pointers, V = # action

Experiment: Do People Use Fixation as a Deictic Pointing Device?

Task

copy a pattern of colored blocks



Continuum of possible strategies

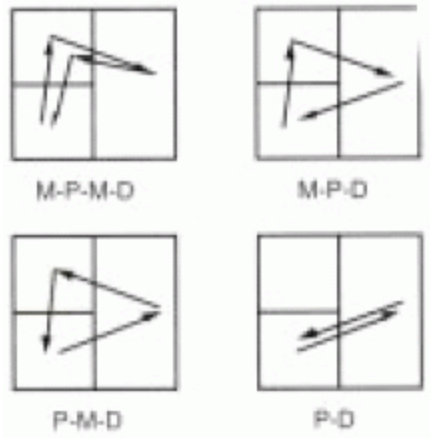
- View model, construct memory image of model; reconstruct shape in workspace using memory representation.
- Rather than constructing memory representation, refer to model in display.

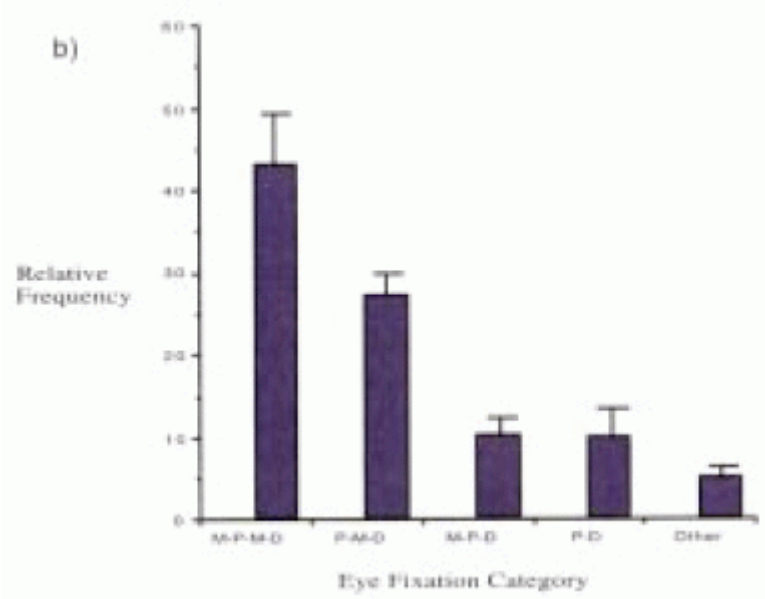
Possible eye movement patterns

M: model
P: pickup point
D: drop off point

Continuum of memory usage:

1. MPMD (use image for obtaining color and location)
2. MPD / PMD (use image for color, memory for location; or vice versa)
3. PD (use memory for color and location)





Subjects use lower memory strategies despite higher time cost.

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

Claim: cost of using working memory (effort, reliability) is greater than cost of making additional fixations (time)

Test

When model and copy are separated by 70 deg visual angle, number of eye movements decreases from 1.3 per block to 1.0 per block.

Further studies of this sort in Gray et al. (2006) – reading for next week

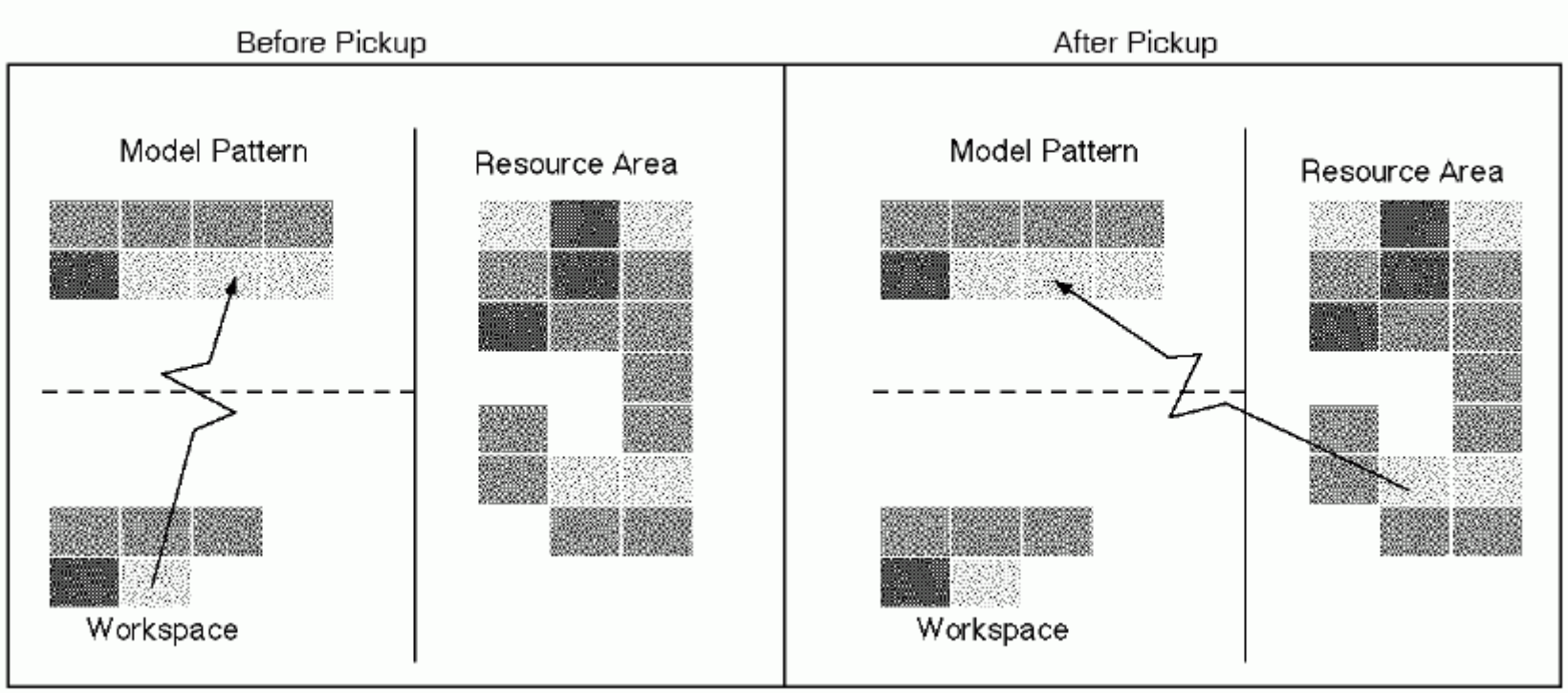
Do Deictic Strategies Avoid the Need for Storage of Information from Prior Fixations?

Test

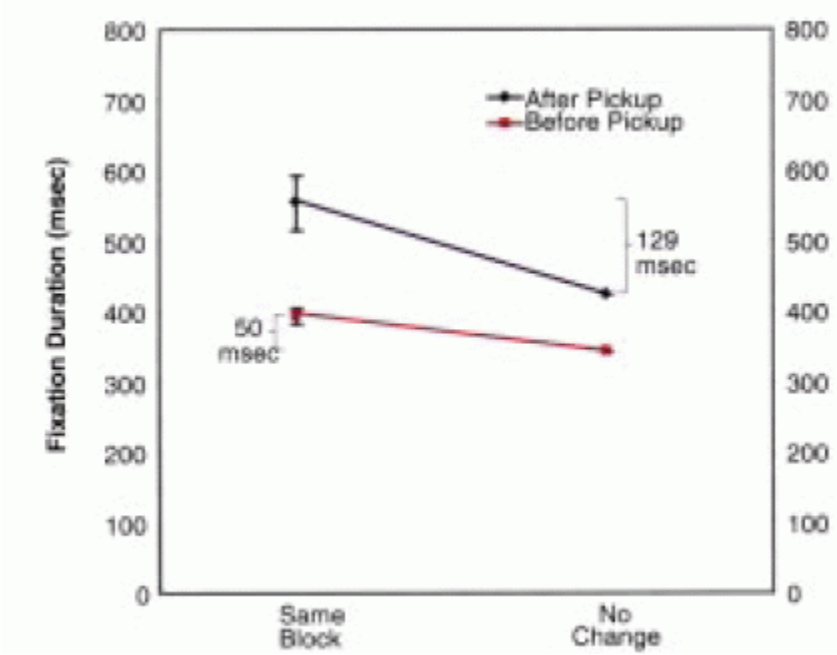
Change information during saccade: color of one uncopied block when moving eyes to model area following block placement.

Change *before* or *after* pickup

If change doesn't slow processing, it must not have been stored in memory.



Results



Before pickup: “Block color is not retained in memory” (reflected by 50 ms difference)

After pickup: “Color info has been retained since it is now (immediately) task relevant” (reflected in 129 ms difference)

Information retained from prior saccades is determined by what is currently relevant for task.

Avoids need for complex memory representations

Summary

It's too hard for the brain to build an internal representation of complex visual scenes.

Involves forming representation of many location-color bindings, and many block-to-block relative location bindings.

Each binding is a complex relational memory element, Relation(X,Y), not just an atomic element such as X or Y

Deictic strategies have many advantages

- provide a simple means of variable binding, and thereby maintaining a reference across time and space.
- relieve demands on internal memory by leaving important information out in the world.
- result in simplified programs via variables, and simplified learning (credit assignment)

Role of body configuration in cognition

"On the time scale of 1/3 sec, the momentary disposition of the body plays an essential role in the brain's symbolic computations."

"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."