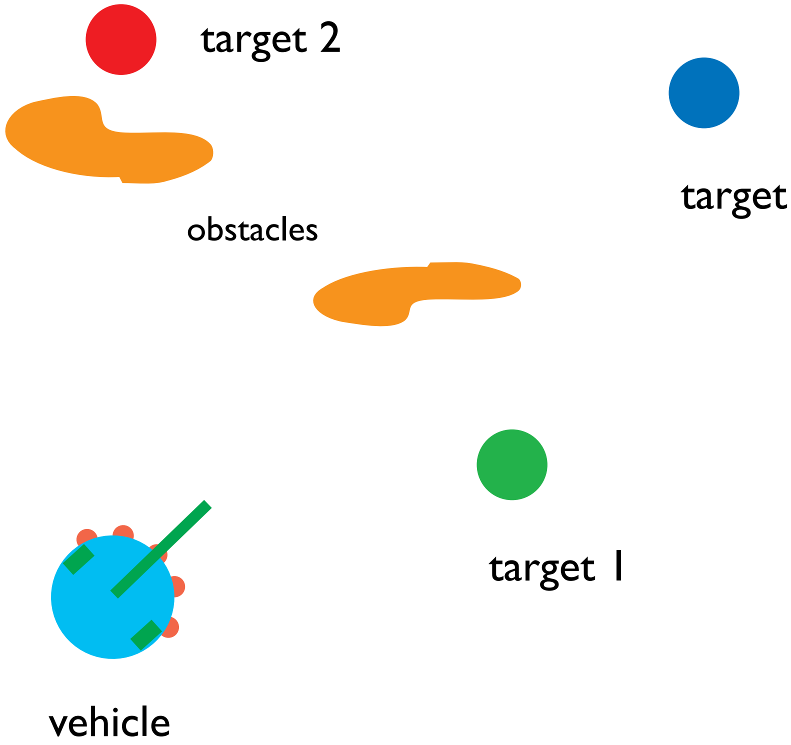



# Dynamic Field Theory: autonomy

Gregor Schöner
Institute for Neural Computation (INI)
Ruhr-University Bochum, Germany
dynamicfieldtheory.org
gregor.schoener@rub.de

# Sequences

- all actioning and thinking consist of sequences of movements, perceptual states, and inferences
- sometimes in a fixed order (routines, action patterns)
- but potentially highly flexible: serial order, productivity…

# Challenge

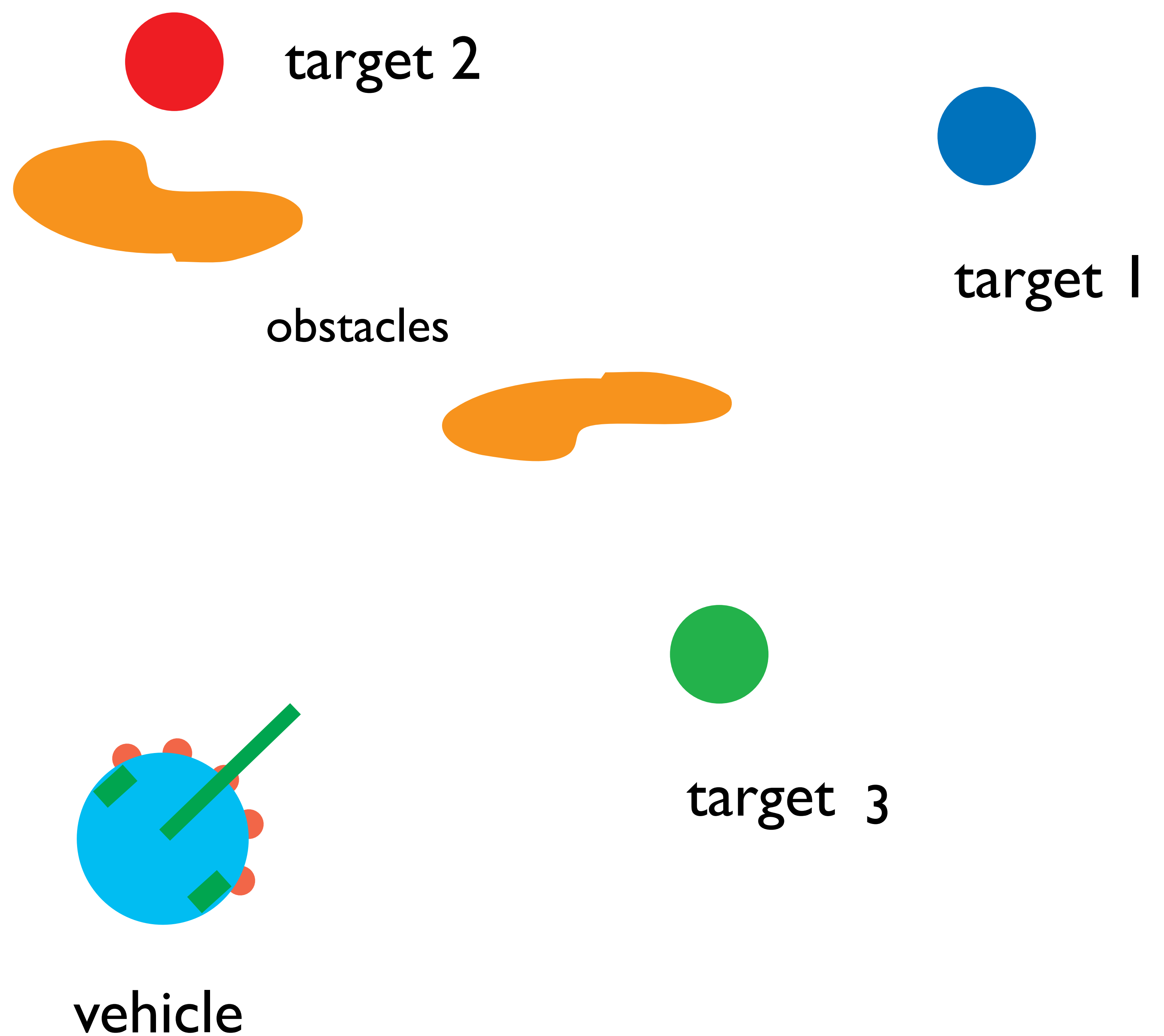
- DFT postulates that all neural states driving behavior/mental process are attractors
- that resist change…
- sequences require change…
- answer: induce an instability to access new attractor

# Sequence generation

- an illustrative example
- the neural/mathematical mechanism

# Illustration

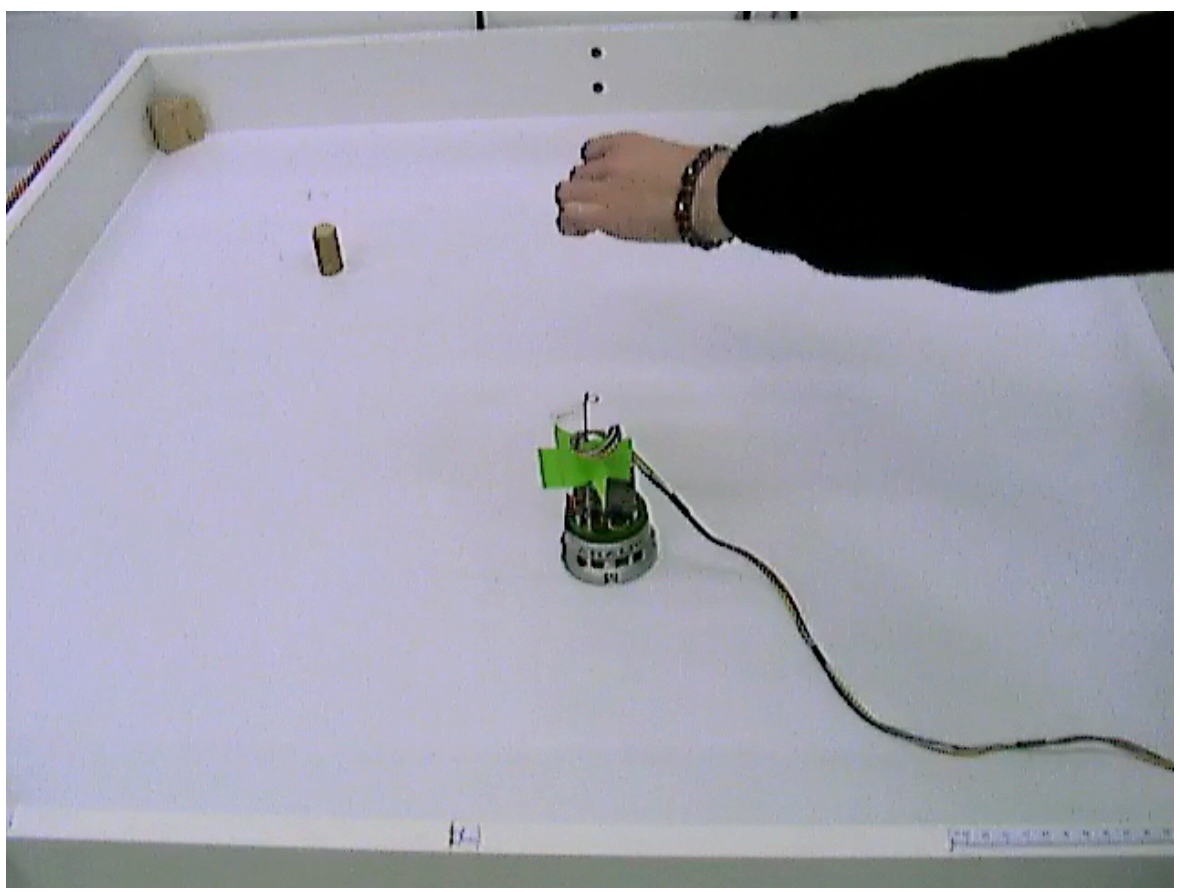
- search for objects of a given color in a given order
  - 1 blue
  - 2 red
  - green
- stably couple to objects once they are detected
- ignore objects when their turn has not yet come (distractors)



# Implementation as an imitation task

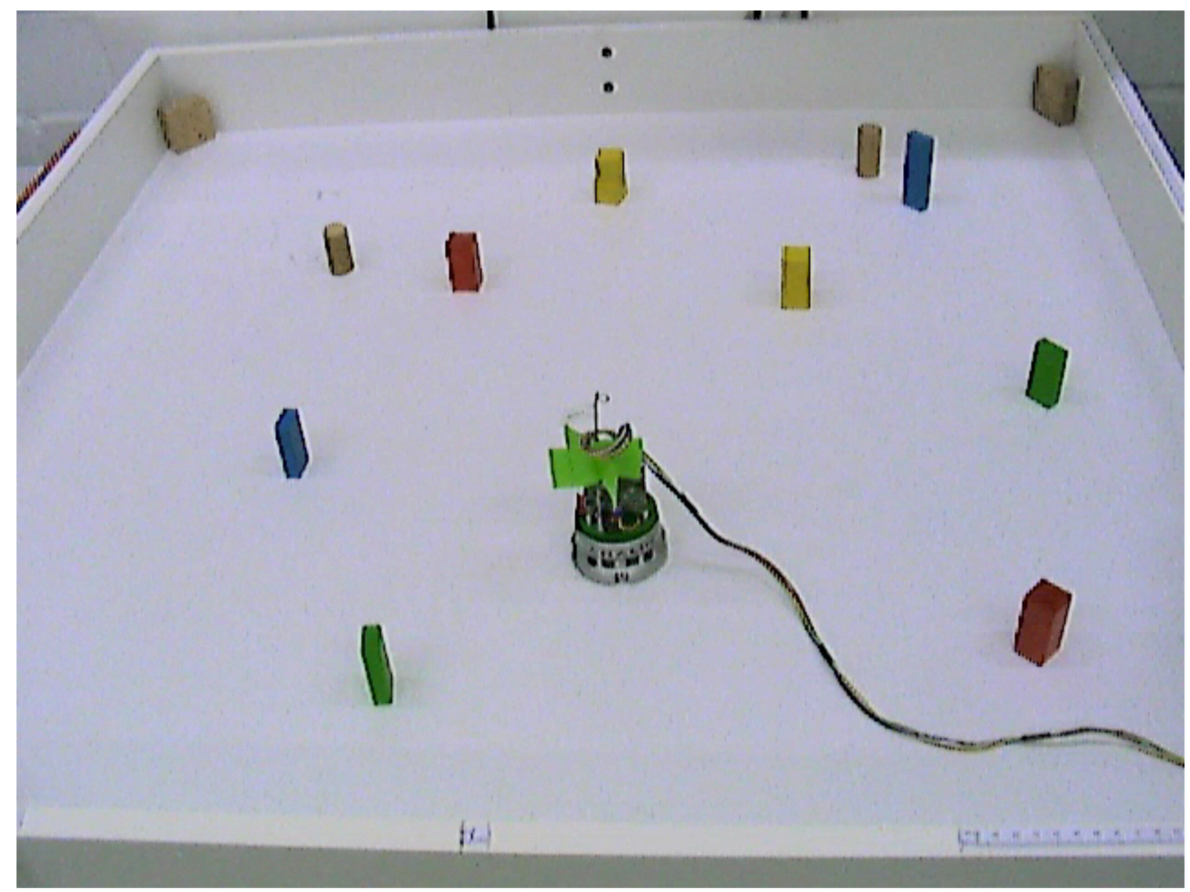
- learn a serially ordered sequence from a single demonstration

yellow-red-green-blue-red



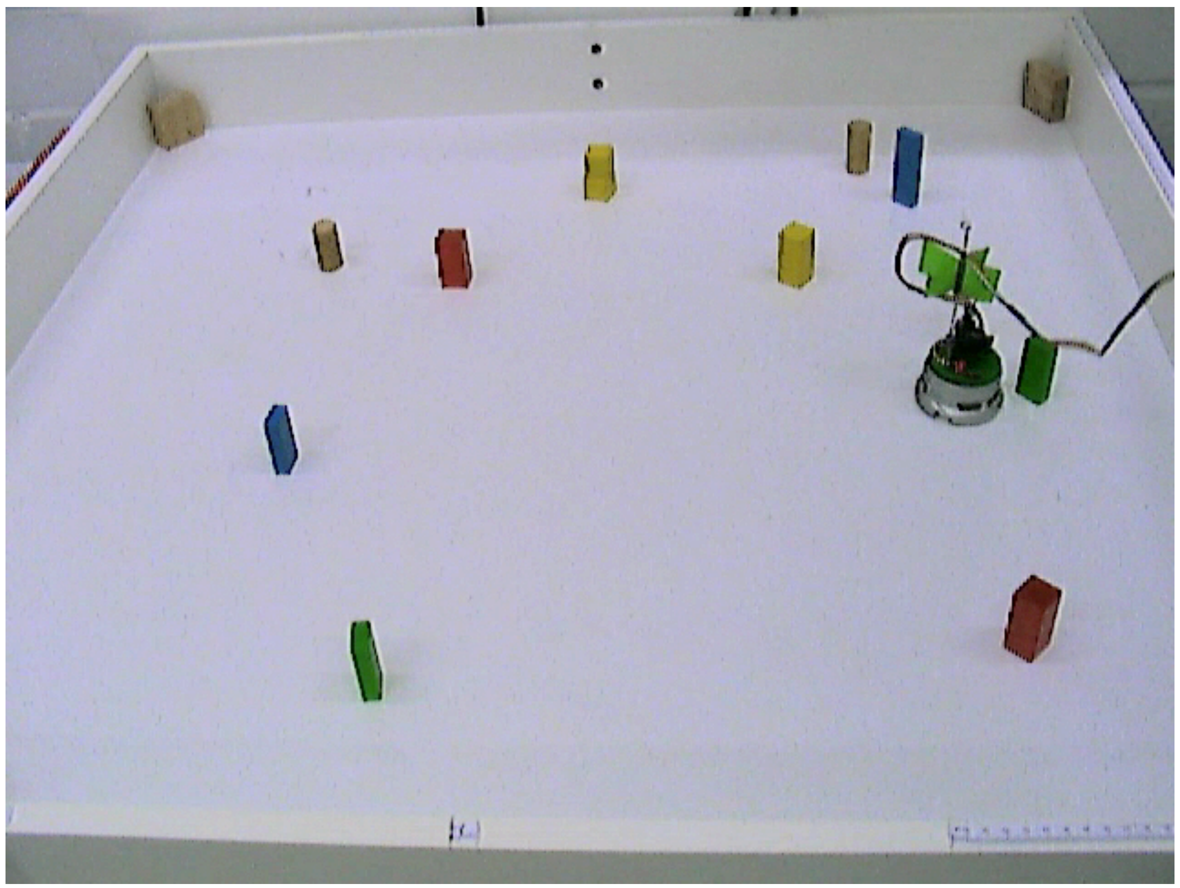
- perform the serially ordered sequence with new timing

yellow-red-green-blue-red

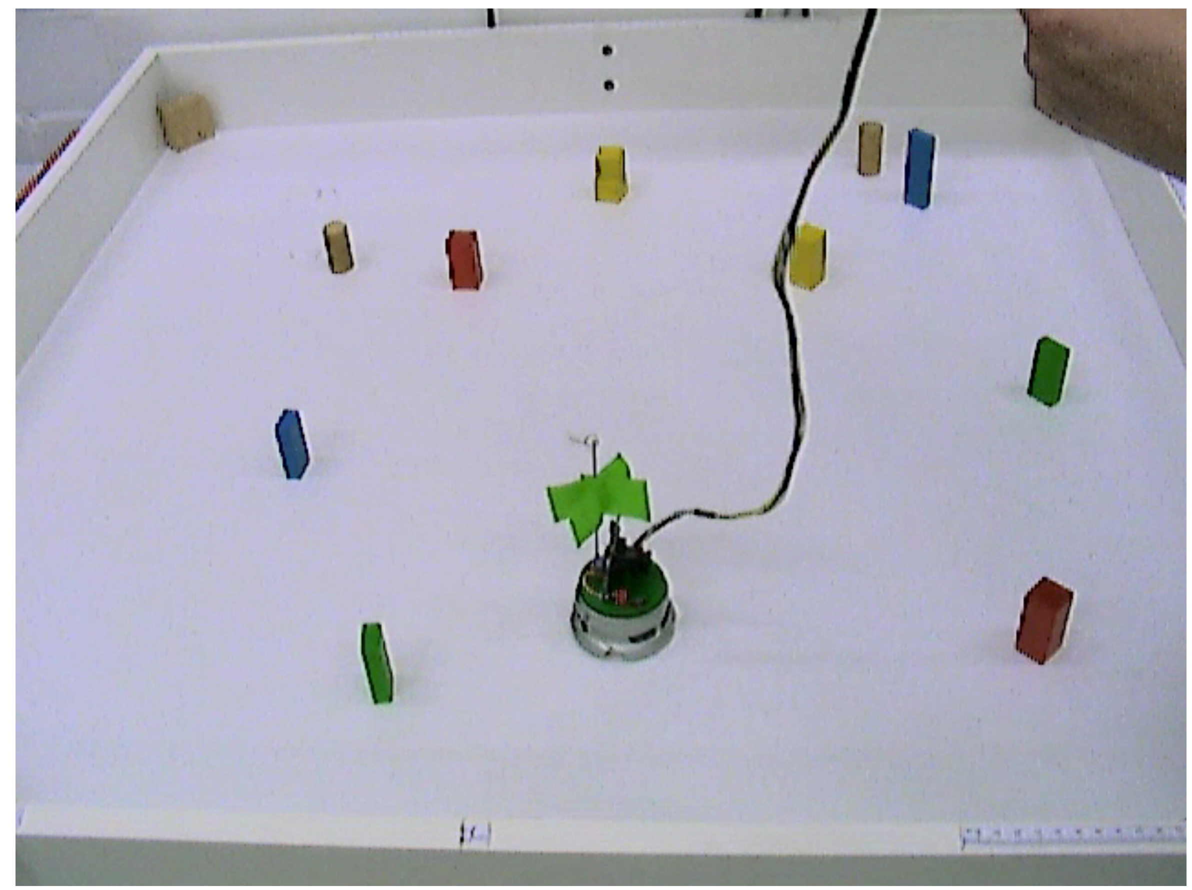


[Sandamirskaya, Schöner: *Neural Networks* 23:1163 (2010)]

red a distractor

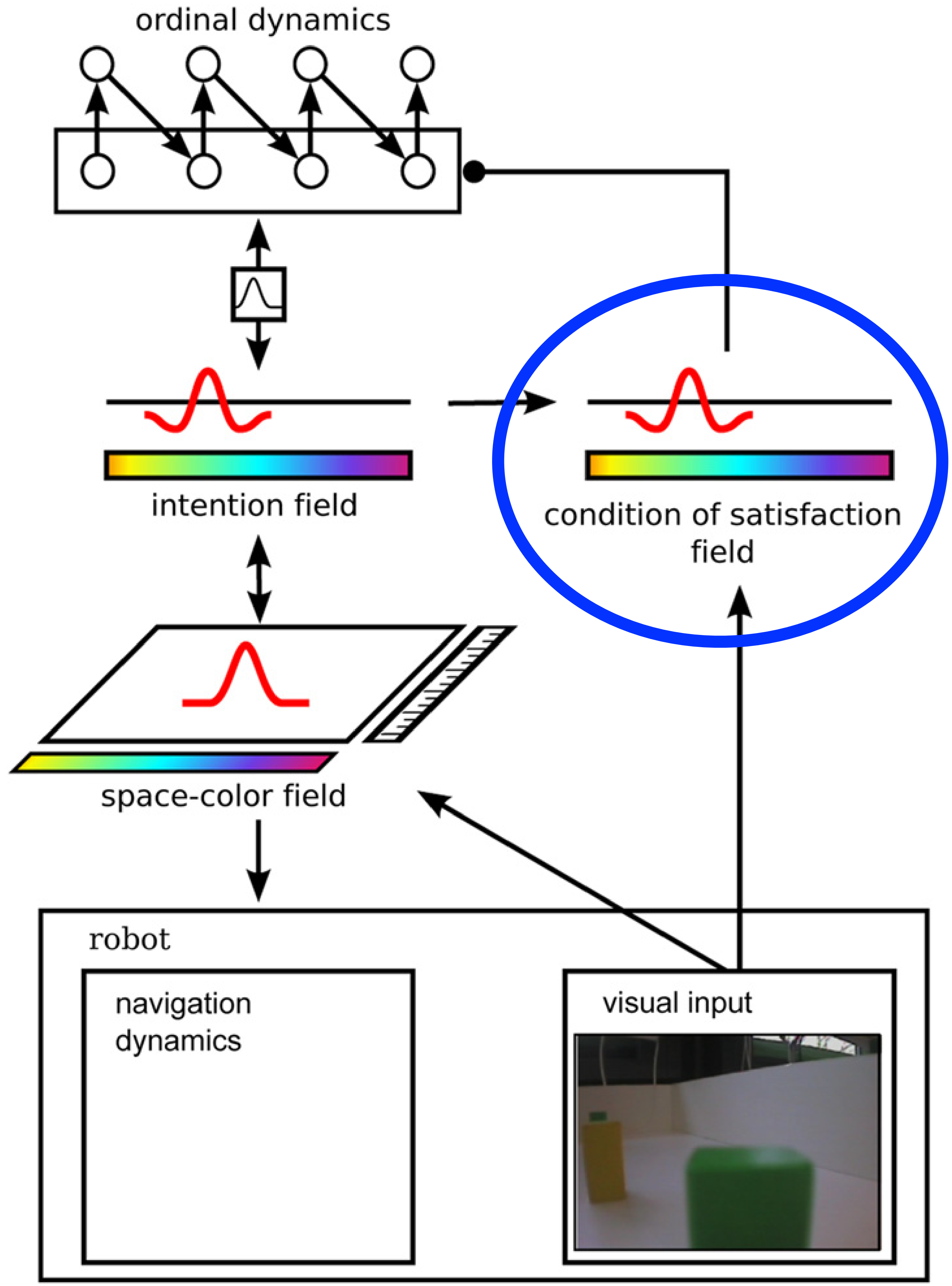


red a target



[Sandamirskaya, Schöner: *Neural Networks* 23:1163 (2010)]

# Condition of Satisfaction (CoS)

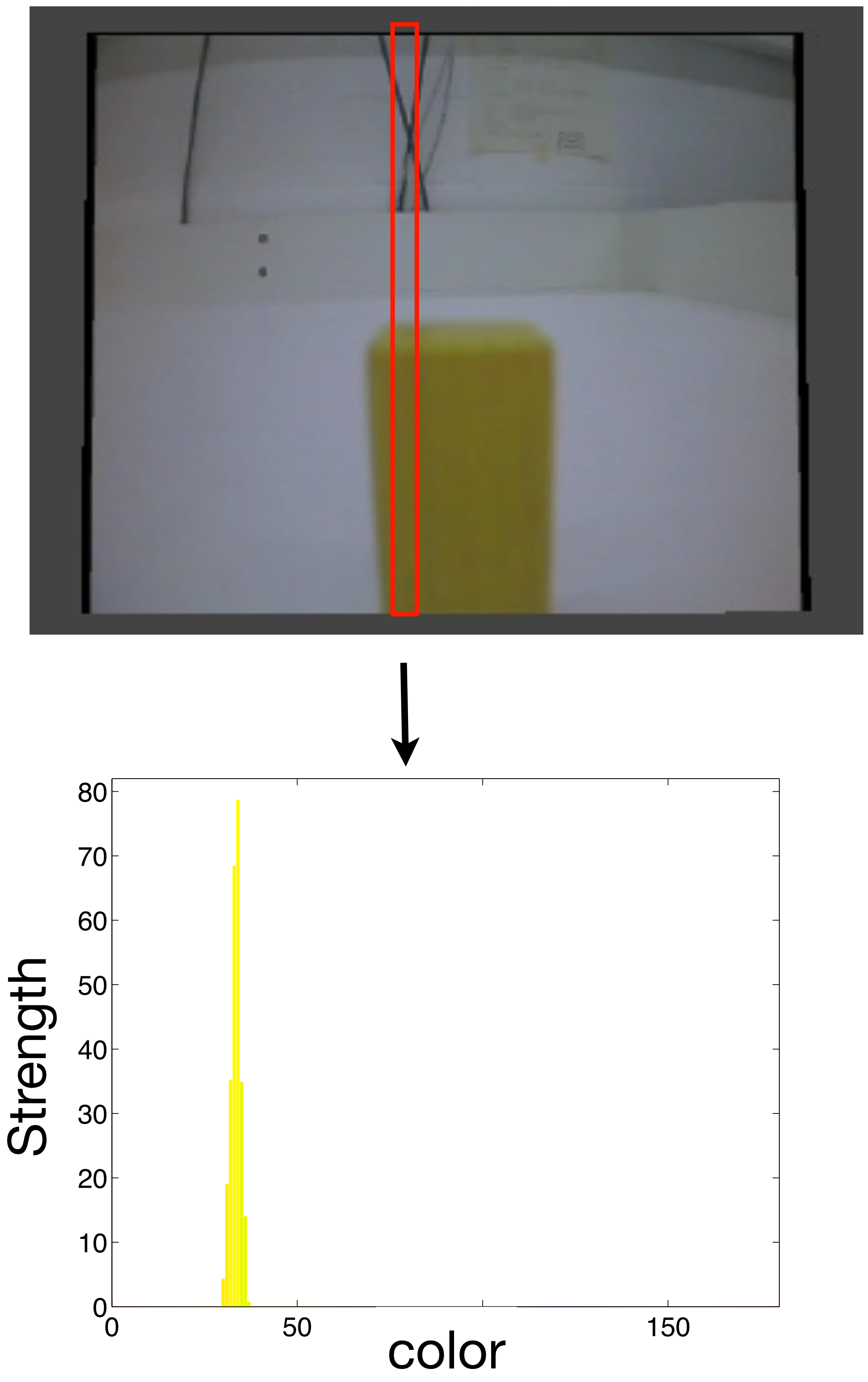



[Sandamirskaya, Schöner: *Neural Networks* 23:1163 (2010)]

# Visual search

- 2D visual input color vs. horizontal space
- intensity of input from a color histogram within each horizontal location

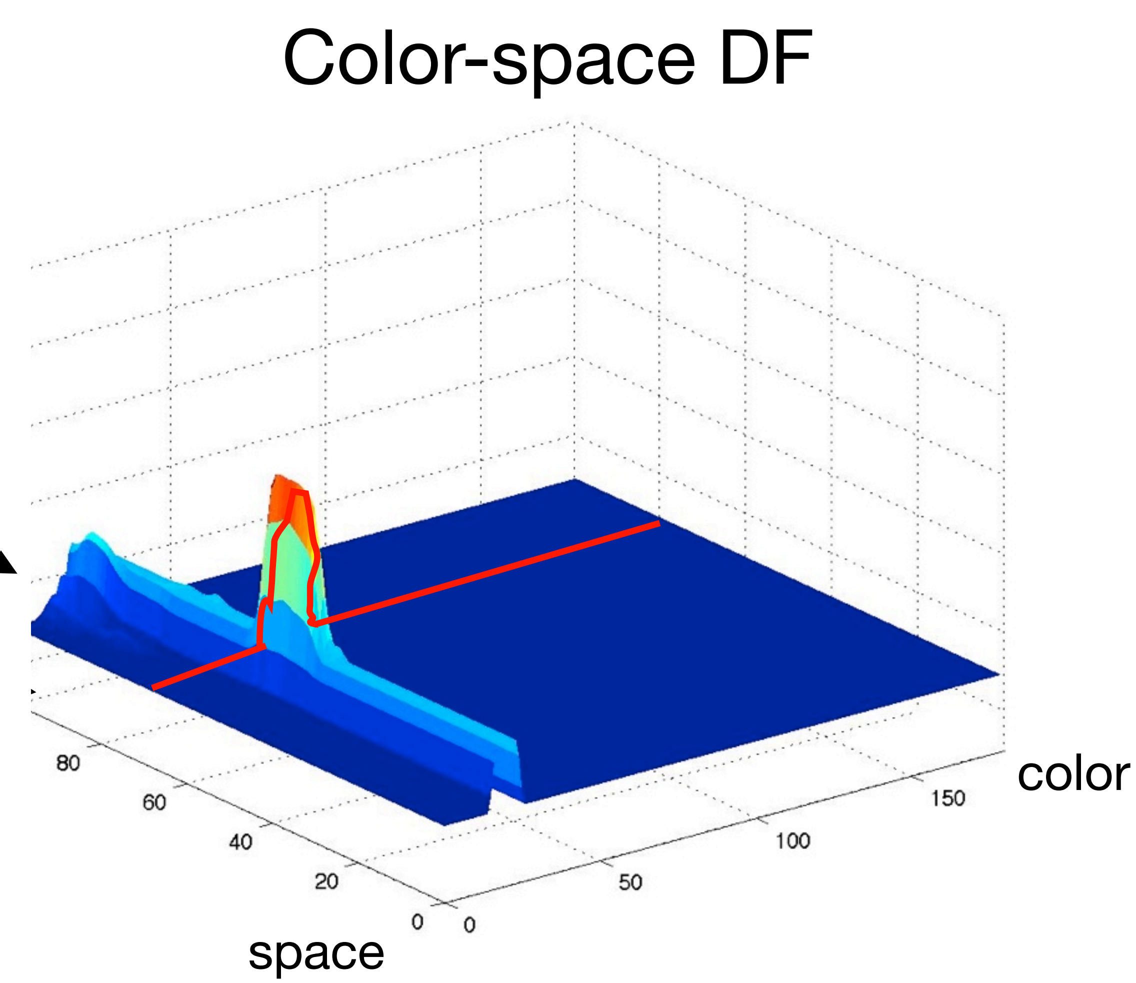
Camera image

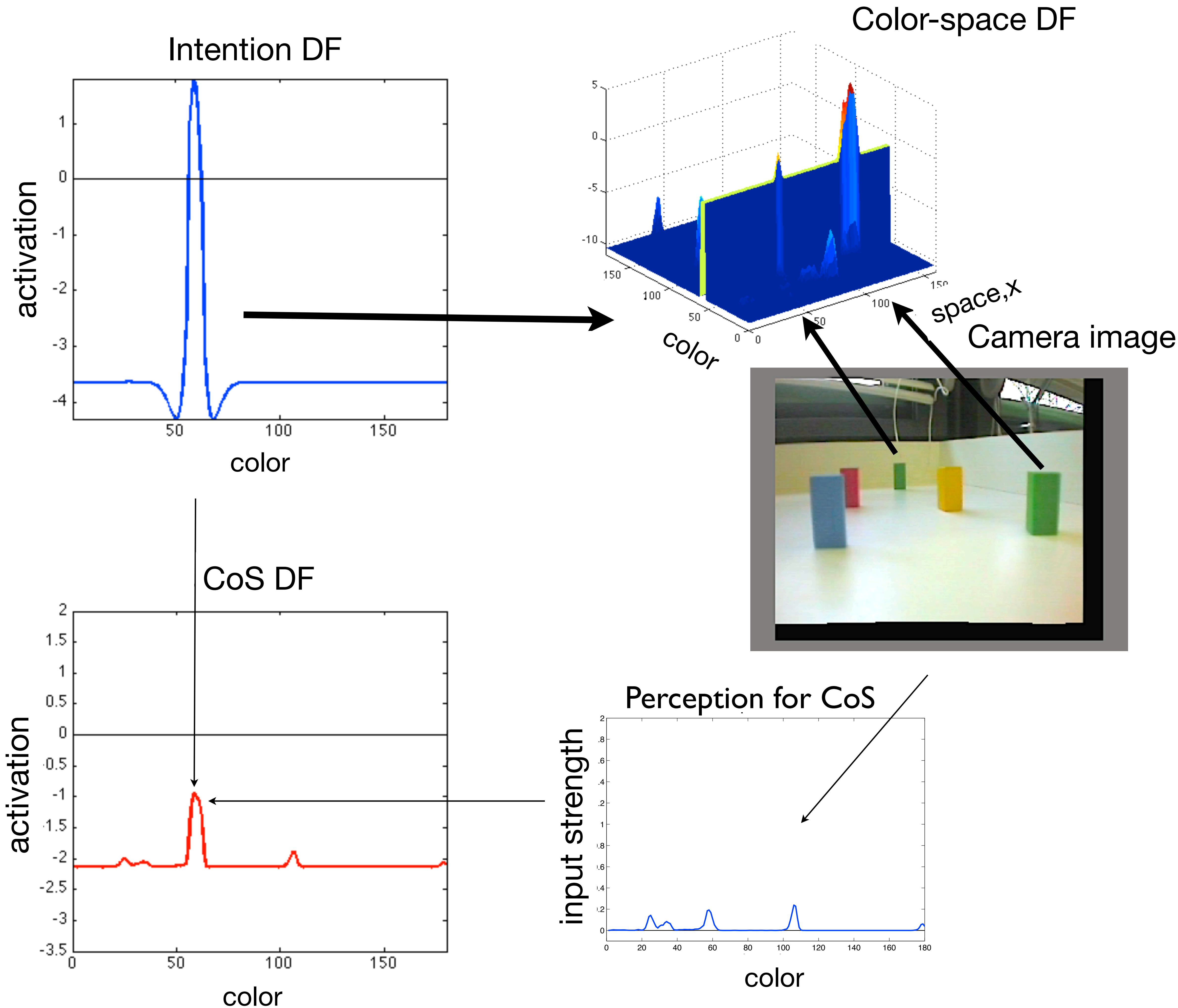


# Visual search

- current color searched provides ridge input into a color-space field

search cue



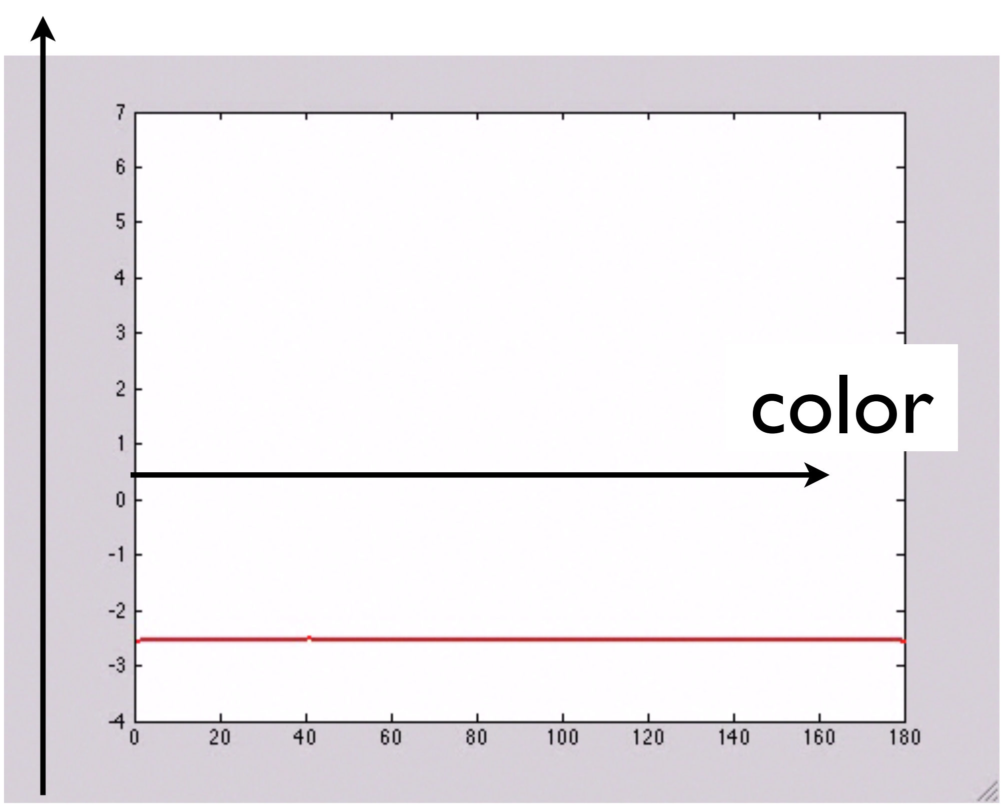


Color-space DF
Intention DF
activation
color
space,x
color
Camera image
CoS DF
activation
color
Perception for CoS
input strength
color

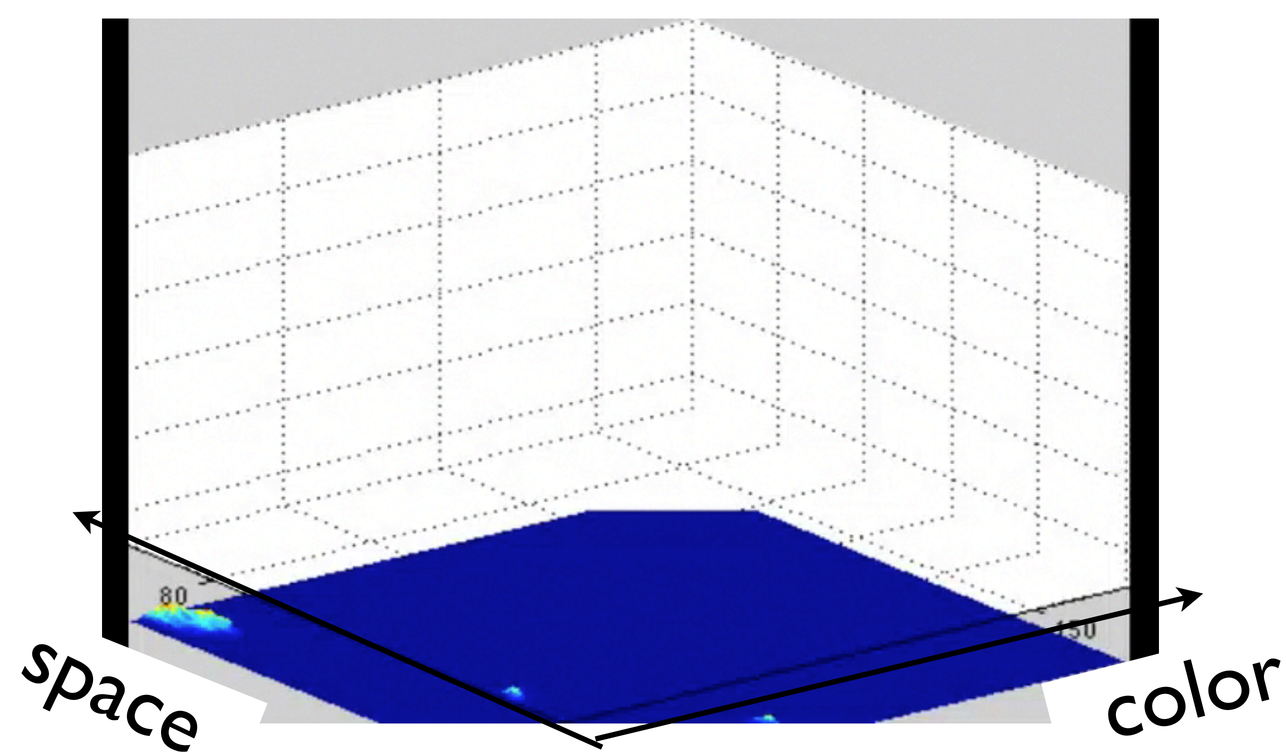
ordinal stack

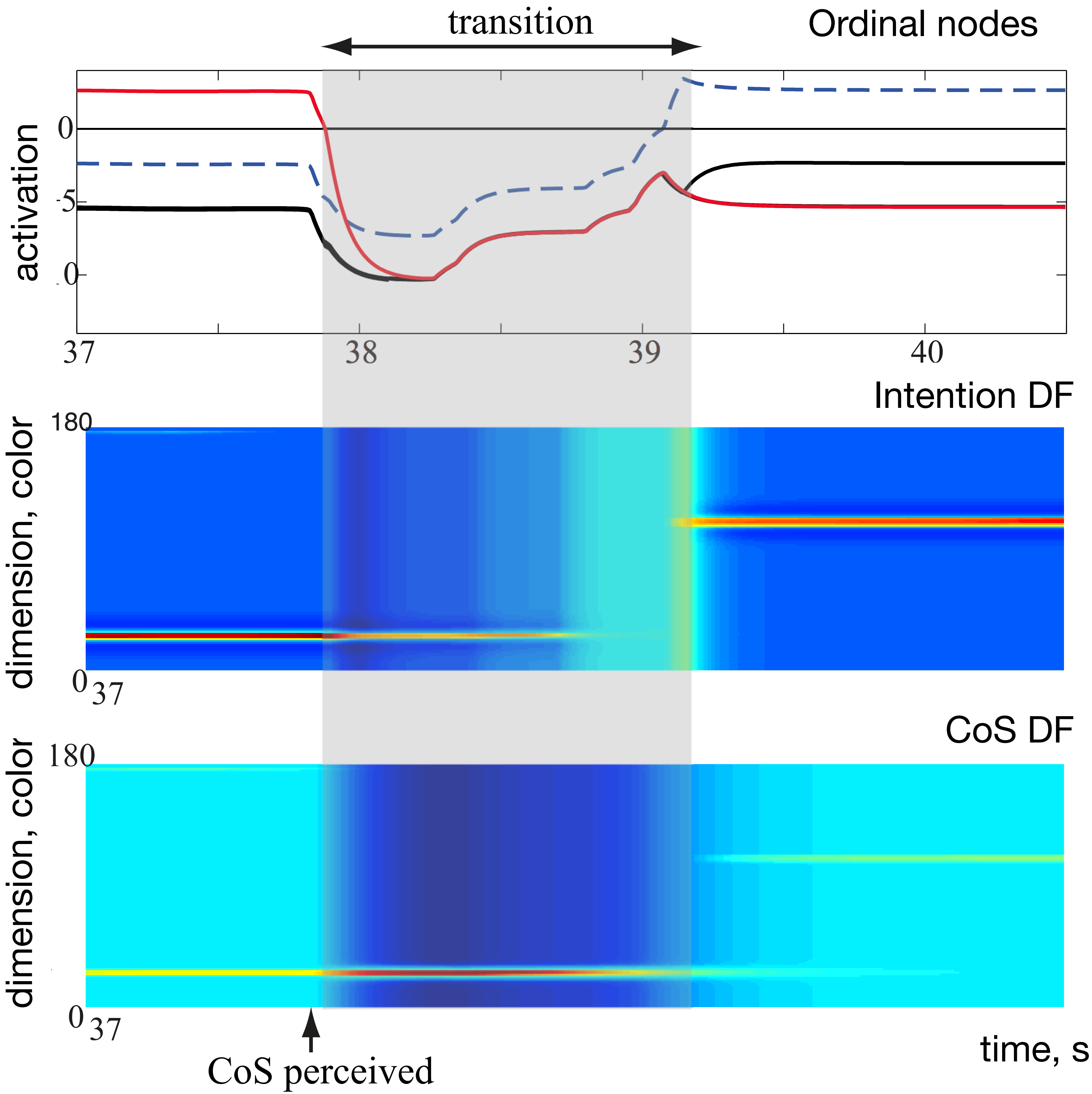
condition of satisfaction (CoS)

intentional state

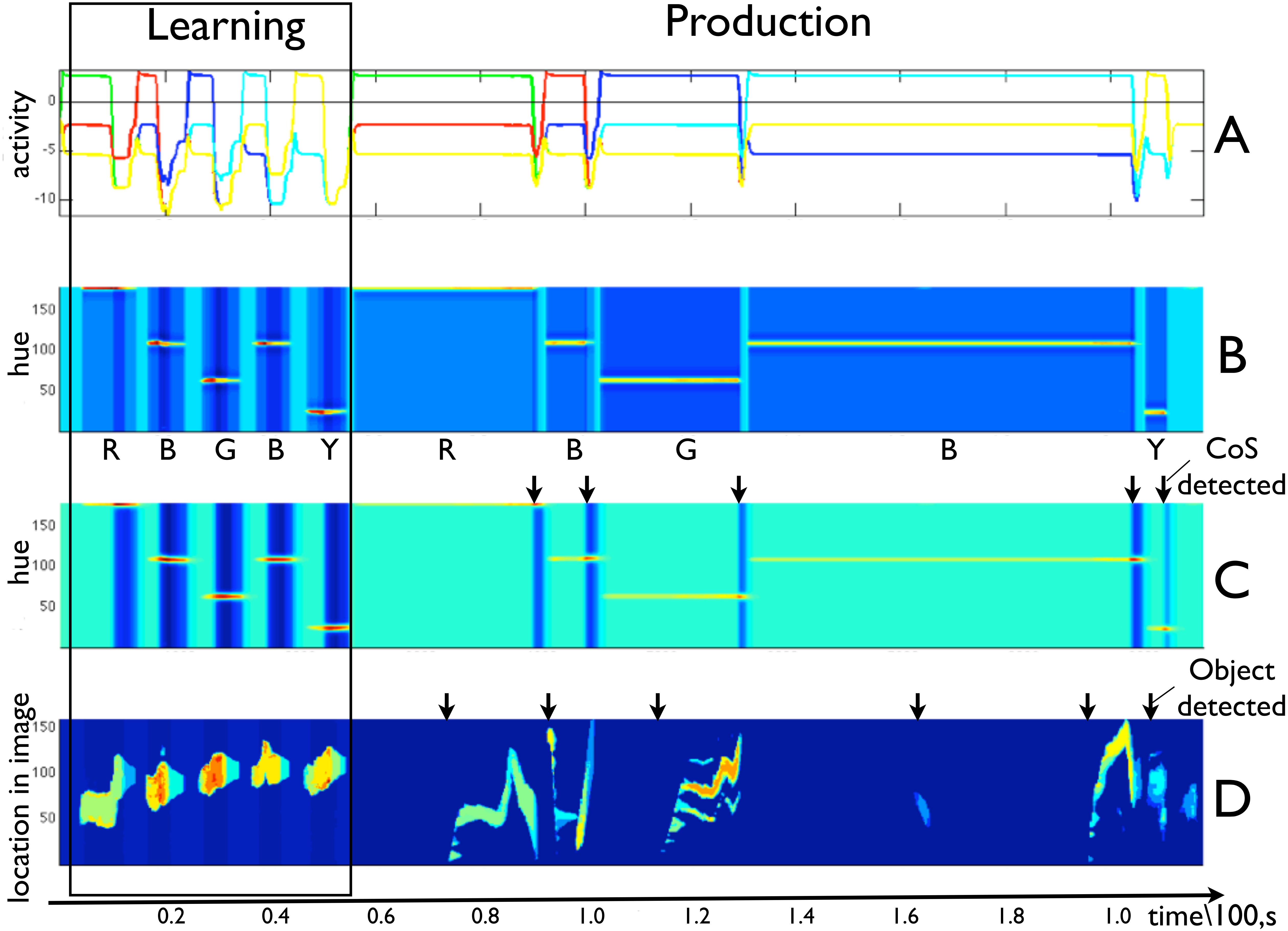



2D color-space field

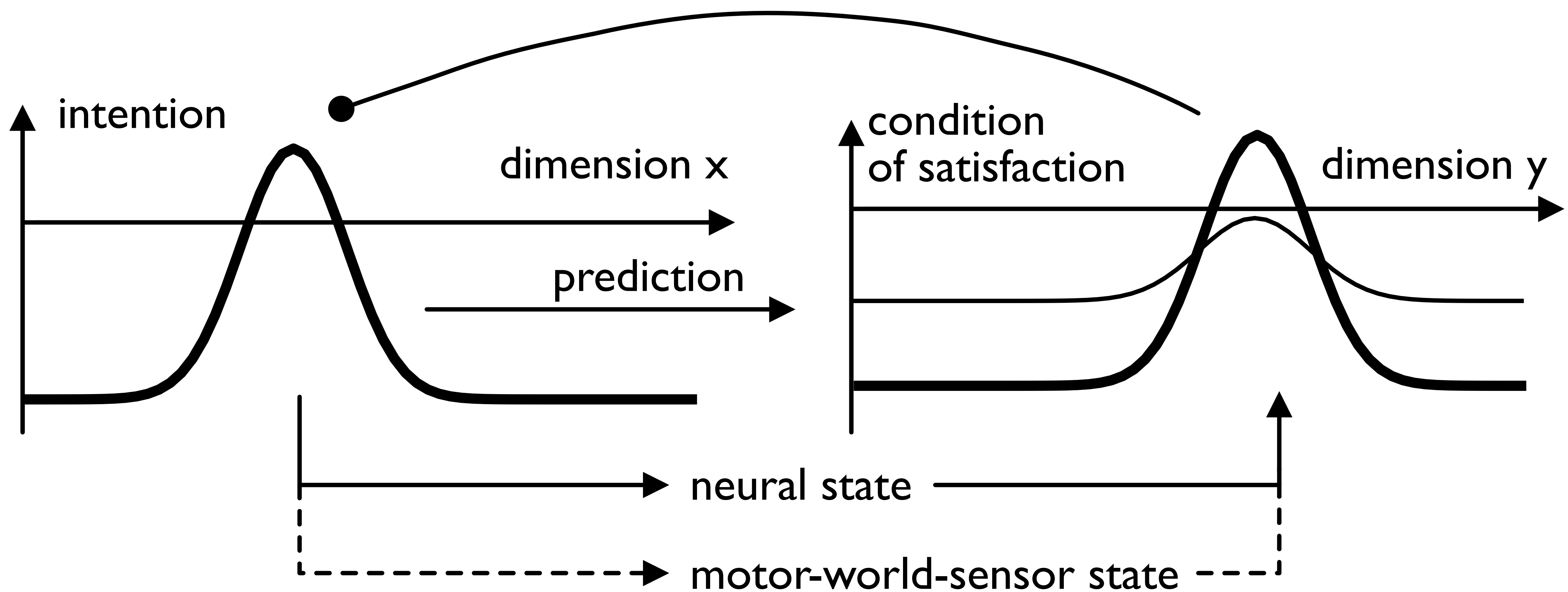




transition
Ordinal nodes
activation
0
-5
-10
37
38
39
40
Intention DF
dimension, color
180
0
37
CoS DF
dimension, color
180
0
37
CoS perceived
time, s

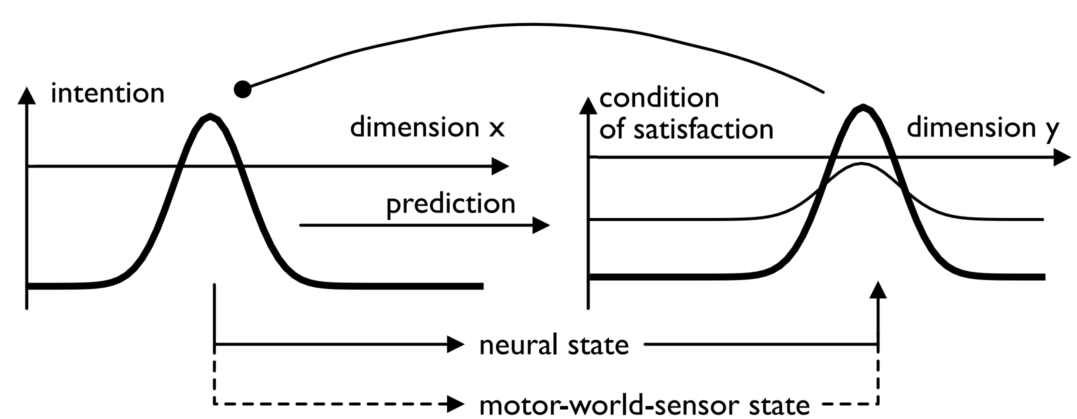

Learning
Production
activity
0
-5
-10
A
hue
150
100
50
B
R
B
G
B
Y
R
B
G
B
Y
CoS
detected
hue
150
100
50
C
Object
detected
location in image
150
100
50
D
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
1.0
time\100,s

# Mathematical mechanism



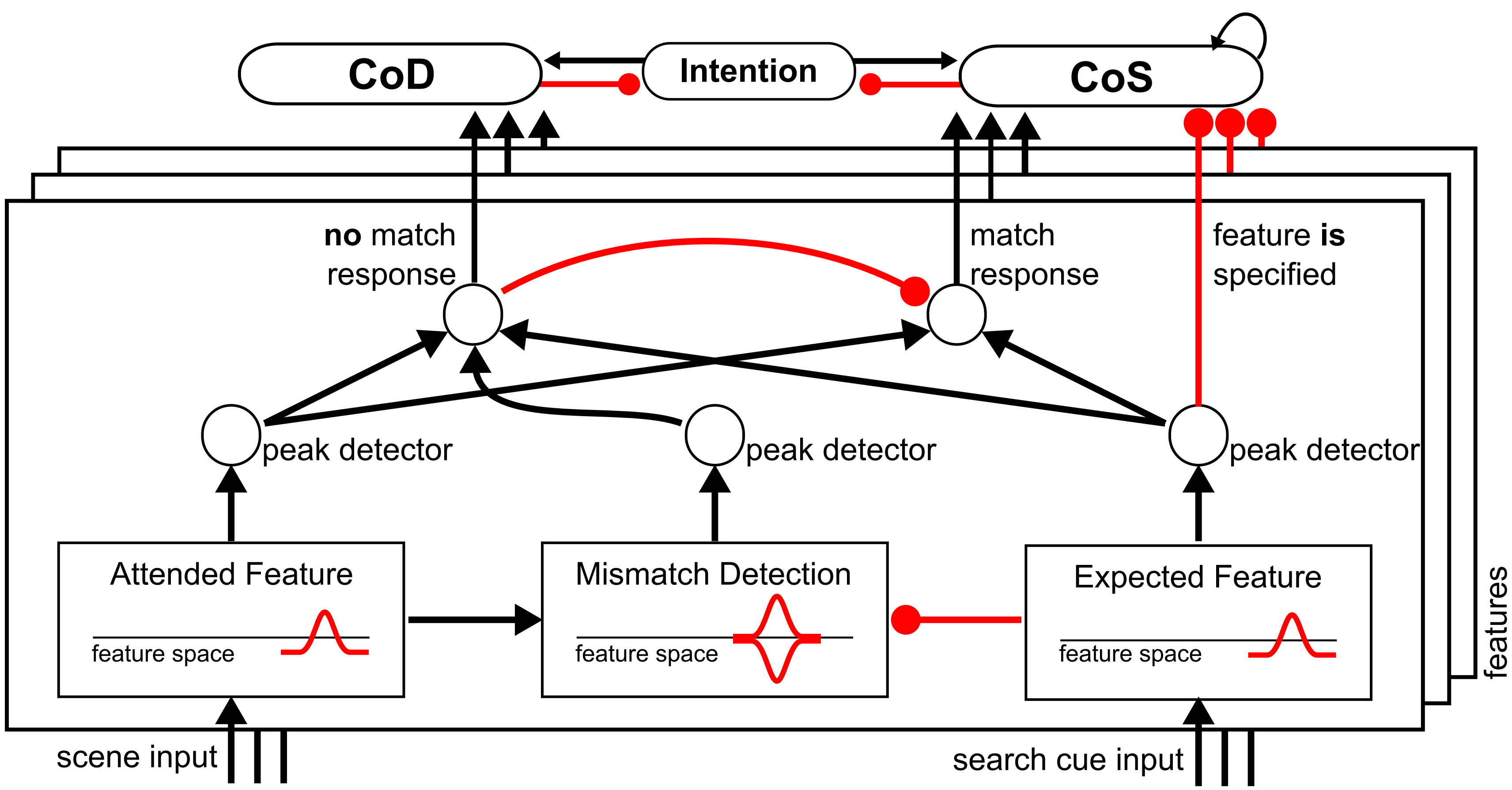
# Sequence of instabilities

- the CoS is pre-shaped by the intention field, but is in the sub-threshold state
- until a matching input pushes the CoS field through the detection instability
- the CoS field inhibits the intention field that goes through a reverse detection instability
- the removal of input from the intention to the CoS field induce a reverse detection instability
- both fields are sub-threshold



# Generalization

- match-detection => CoS
- mis-match (or change) detection => CoD (condition of dissatisfaction)




[Grieben, Schöner, *CogSci* 2021]

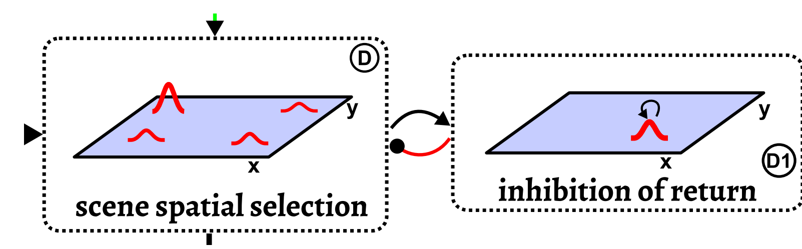
Roadmap

# How is the next state selected?

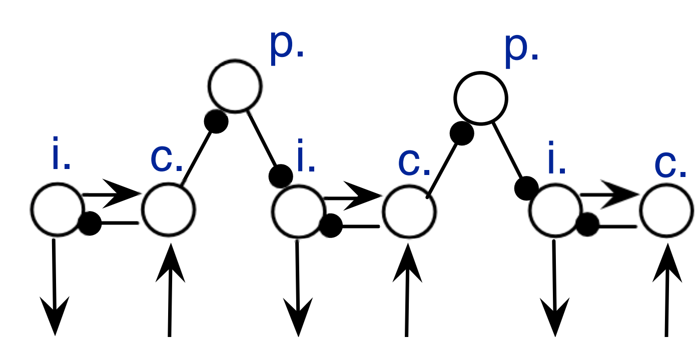
- once the current state has been de-activated…
- three notions
  - gradient-based selection
  - chaining
  - positional representation
- an illustration

# How is the next state selected?

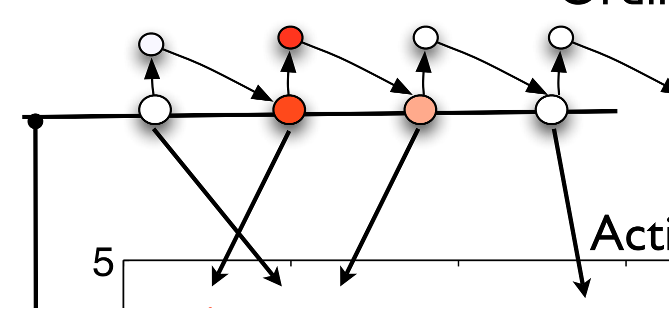
- once the current state has been deactivated…
- 3 notions (~Henson Burgess 1997)




- 1 gradient-based selection



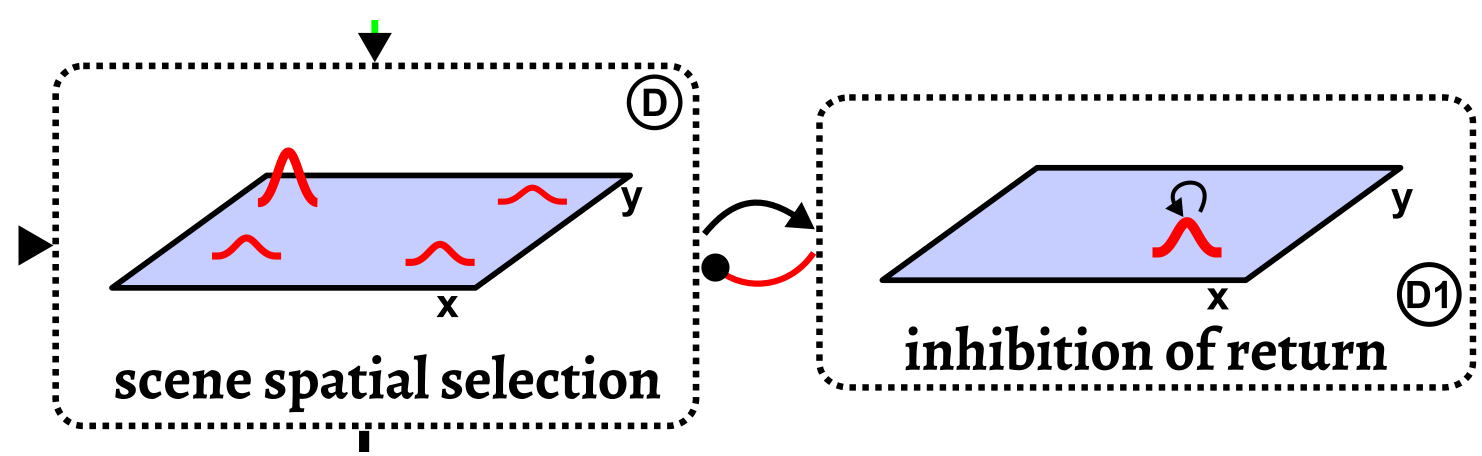

- 2 chaining
- 3 positional representation



# Gradient-based

- a field/set of nodes is released from inhibition once the current state is deactivated…
- a new peak/node wins the selective competition based on inputs…
  - e.g. salience map for visual search
  - e.g. overlapping input from multiple fields..
- return to previous states avoided by inhibition of return

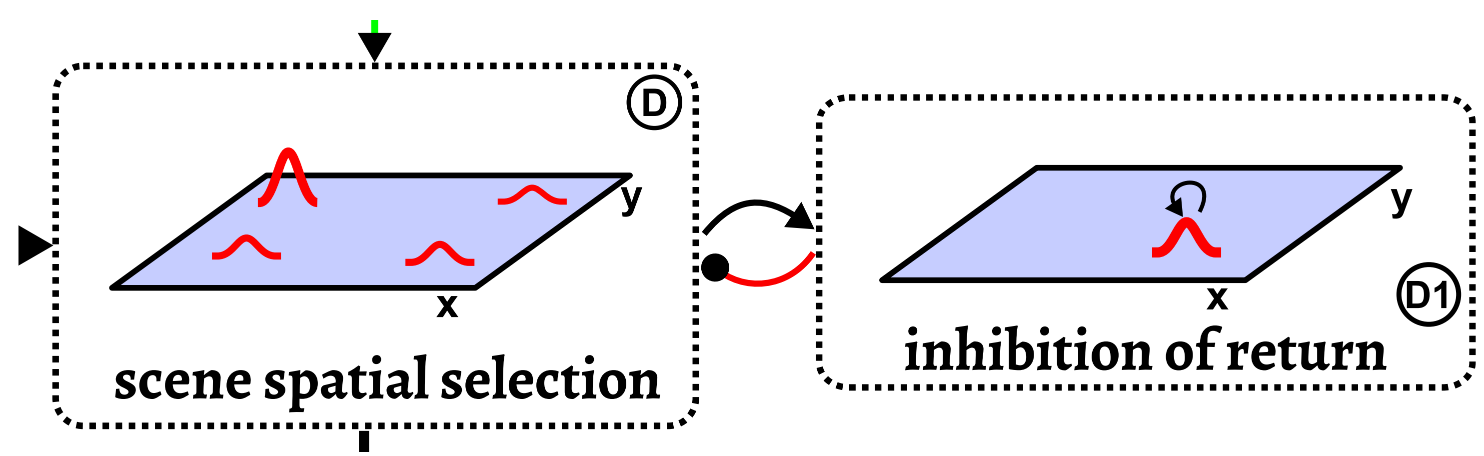
[Grieben, Schöner, *CogSci* 2021]



# Gradient-based

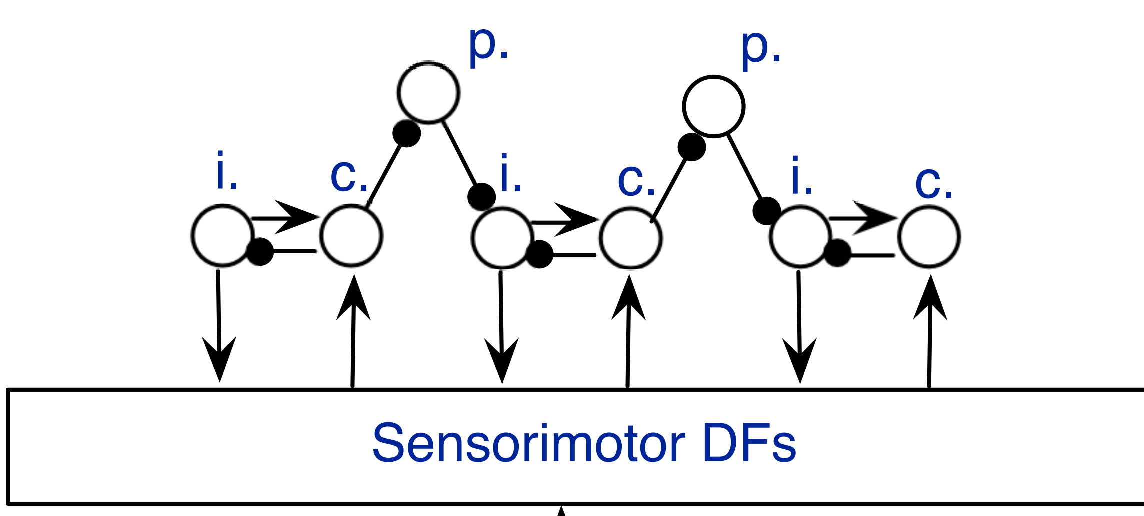
- this is used in many of the DFT architectures
  - visual search
  - relational grounding
  - mental mapping

[Grieben, Schöner, *CogSci* 2021]



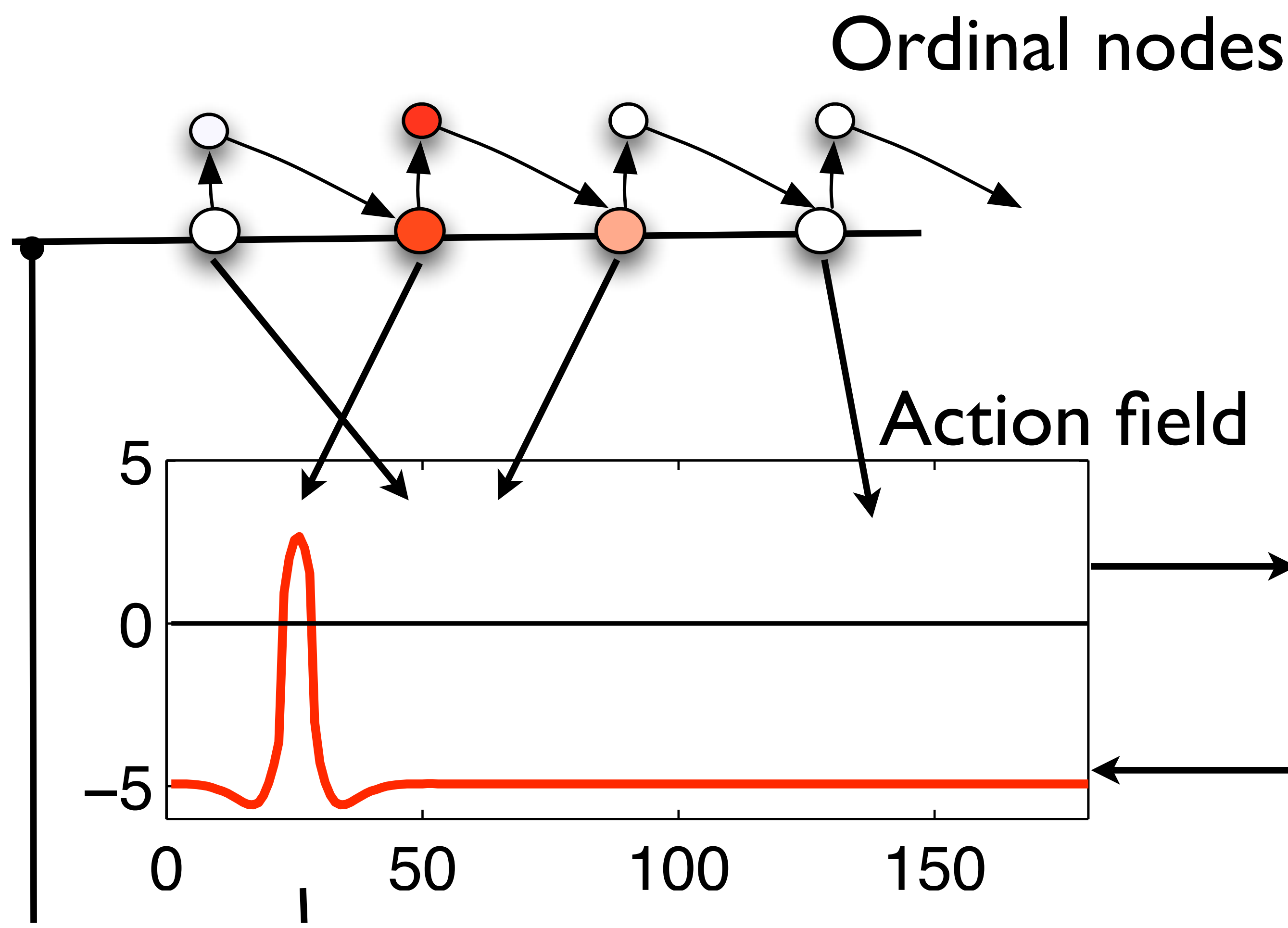
# Chaining

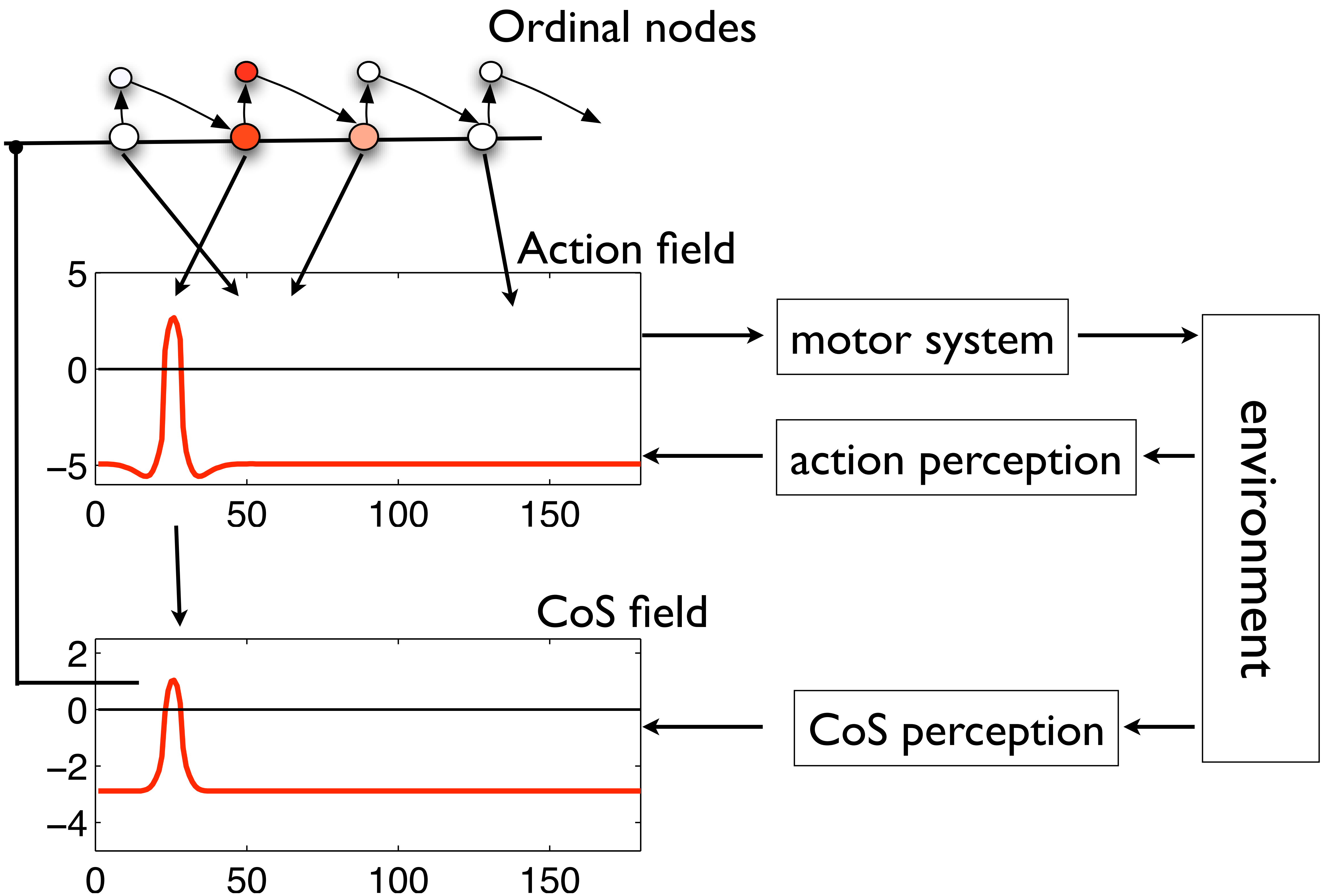
- for fixed sequences…
  - e.g. reach-grasp
  - fixed order of mental operations… e.g. ground reference object first, then target object
- less flexible (e.g.. when going through the same state with different futures)
- could be thought to emerge with practice/habit from the positional system



# Positional representation

- a neural representation of ordinal position is organized to be sequentially activated…
- the contents at each ordinal position is determined by neural projections from each ordinal node…



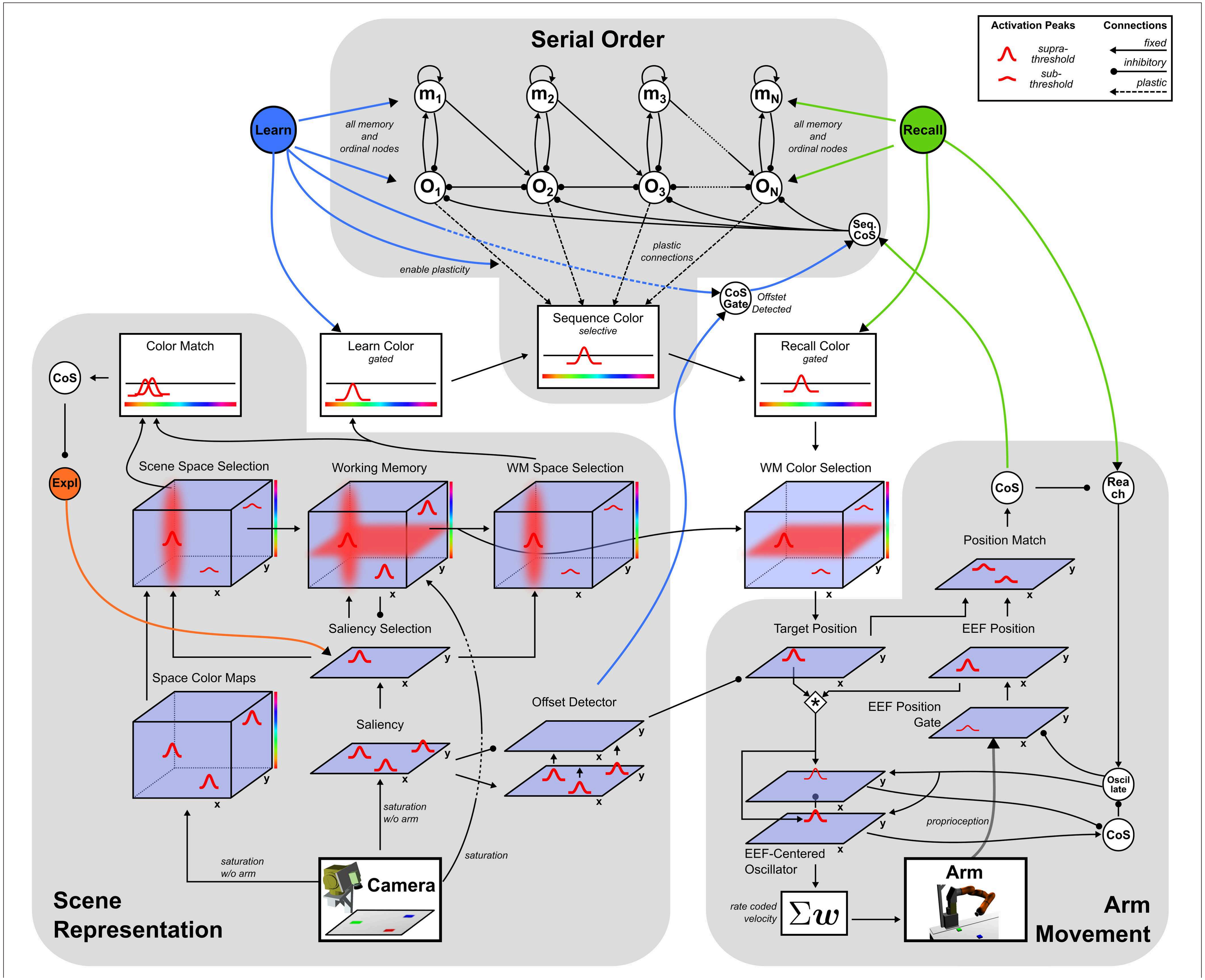



[Sandamirskaya, Schöner: Neural Networks 23:1163 (2010)]

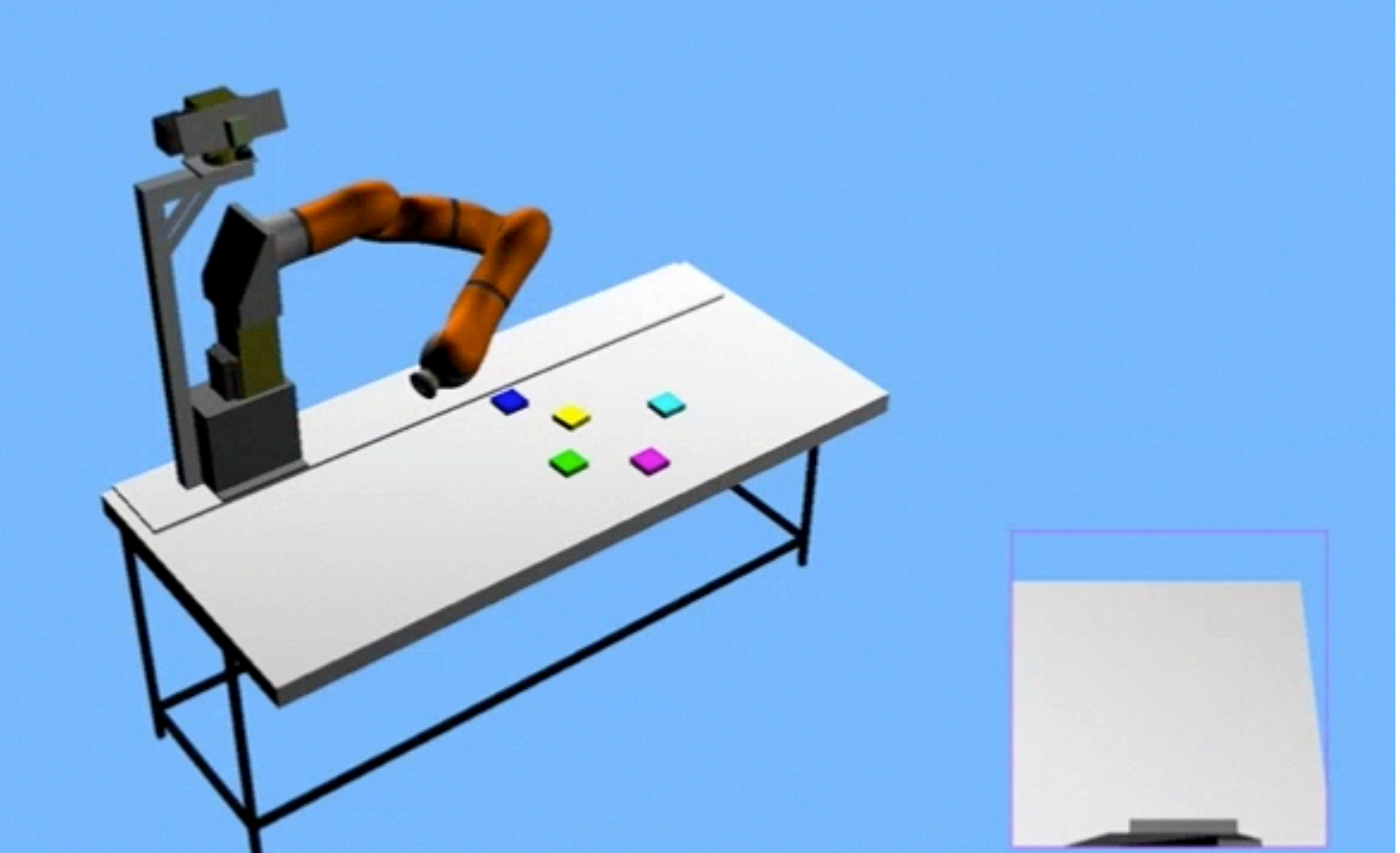
# Positional representation

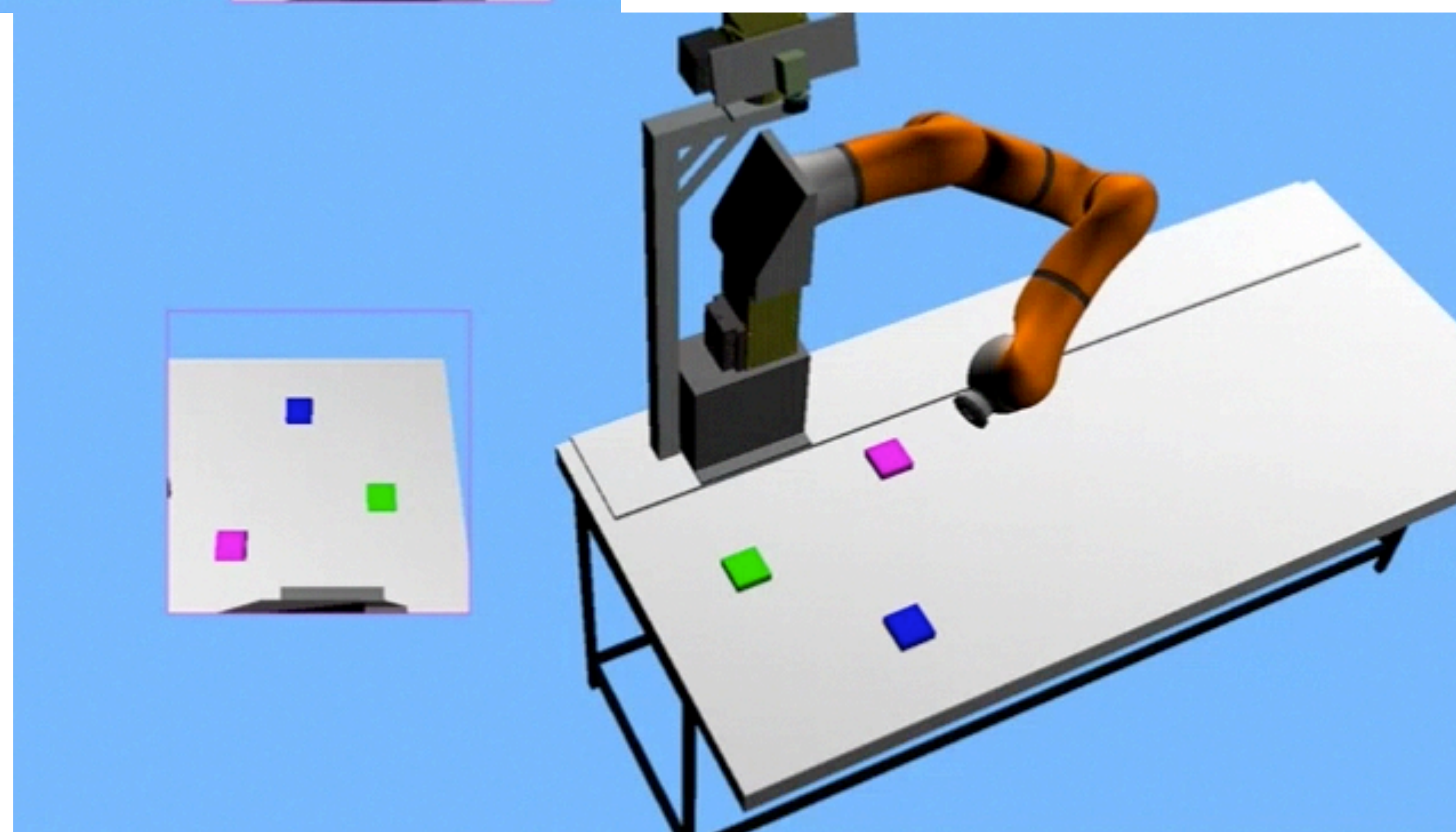
- essentially chaining with flexible contents
- good for fast learning of sequences…
  - e.g. imitation
  - a Hippocampus function?
- but: must have potential synaptic links to many representations…
- => such ordinal systems must exist for sub-representations… embodiment effects…

# Serial order demonstrated/enacted




[Tekülve et al., Frontiers in Neurorobotics (2019)]





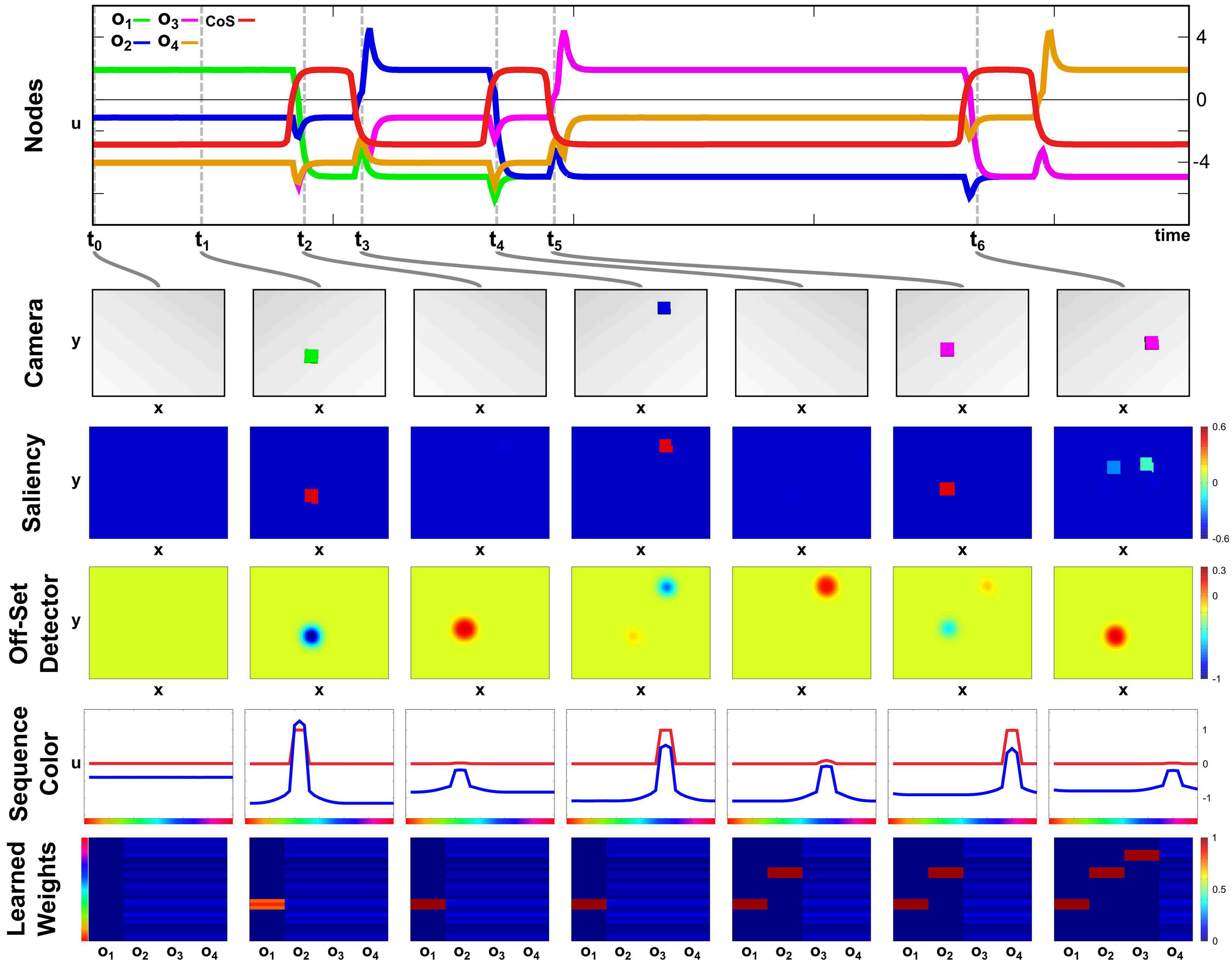



**FIGURE 5 |** Time course of learning a three element sequence with varying presentation time.

# Time course of attention selection and building of scene memory

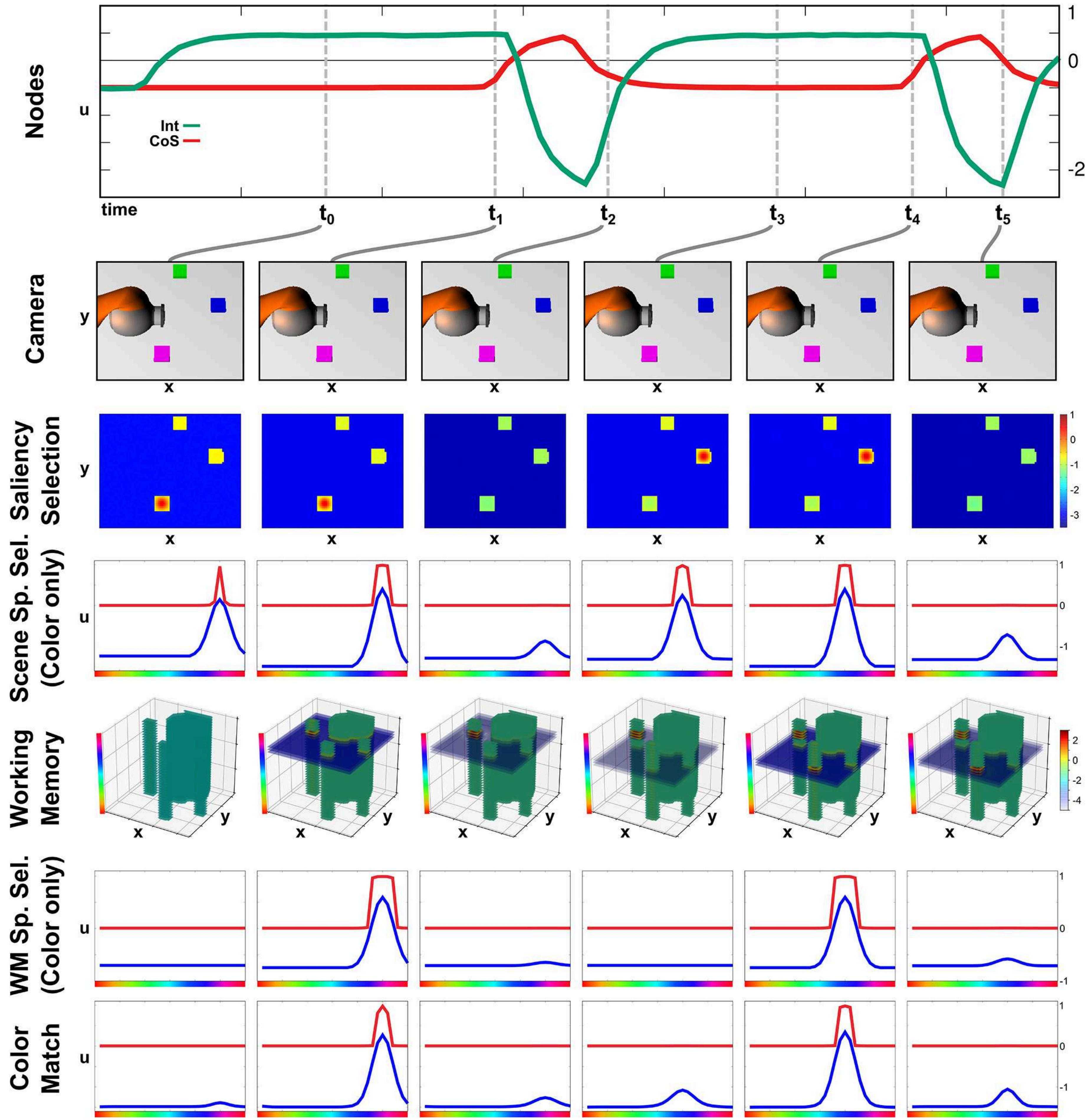



FIGURE 4 | Time course of building a scene memory.

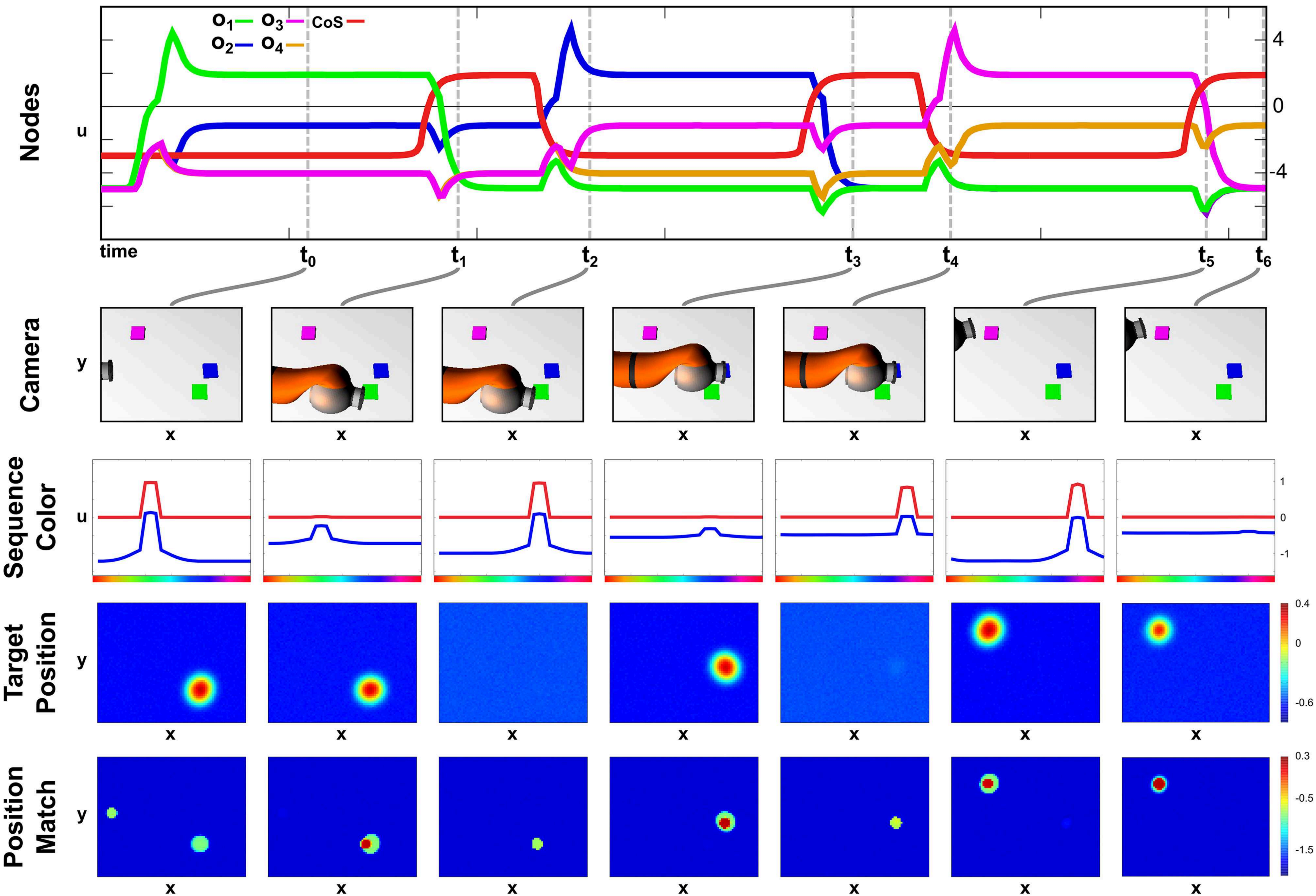
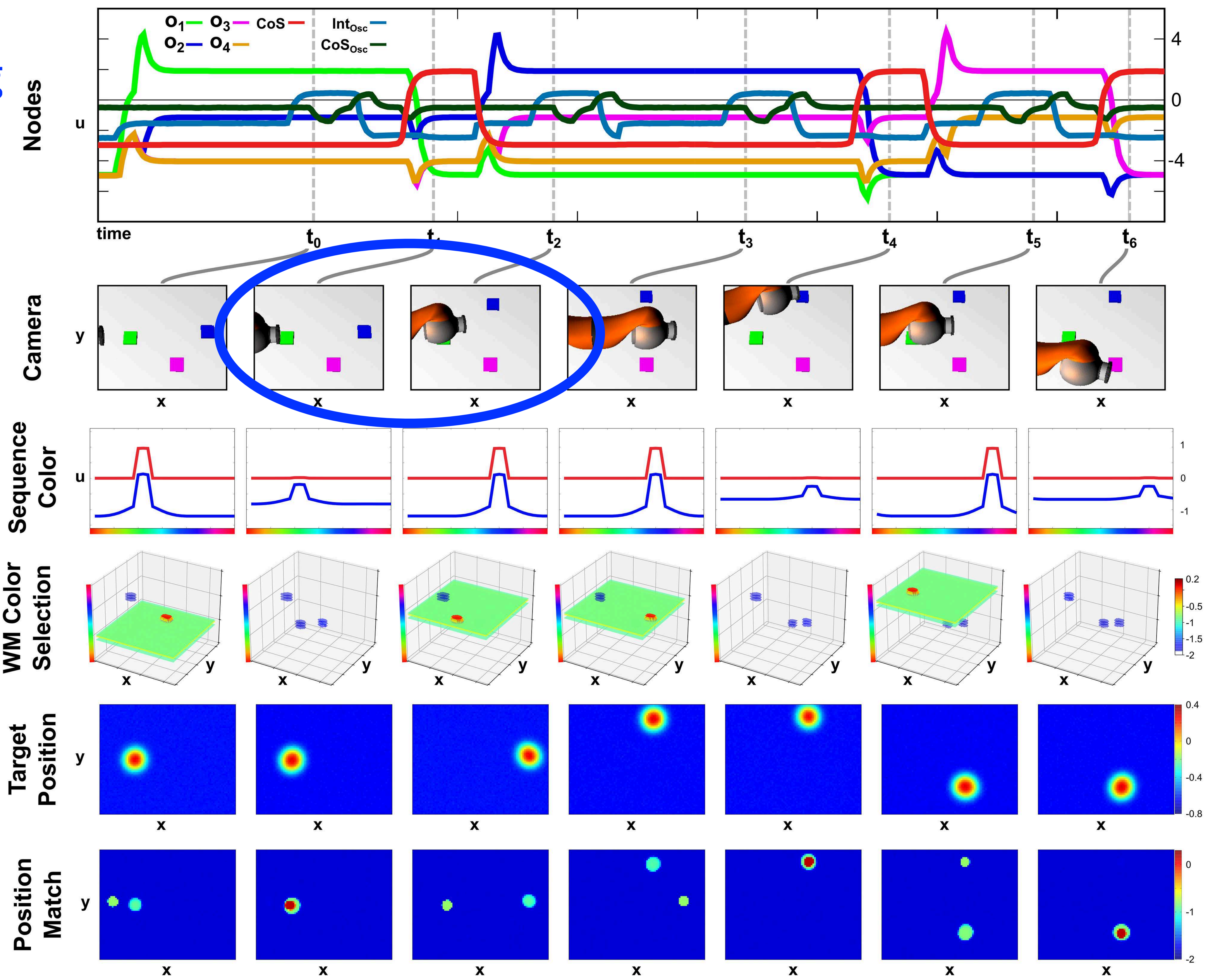



**FIGURE 6 |** Time course of recalling a three element sequence through pointing at colored objects.

online updating



Roadmap

# How far does such autonomy take us?

- the concept of *intentionality* to guide the building of an embodied cognitive architecture
- two directions of fit and the CoS
- an illustration

# How does the mind emerge from neural processes?

- What do I mean by "mind"?
- *Intentionality* = the capacity of nervous systems to generate mental states that are *about things in the world*
  - *things* may include an organism's own body
  - *things* may include the nervous system's own states

# Two *directions of fit* of intentional states (according to John Searle)

- *world-to-mind*: the world must match the intentional state to fulfill that state's *condition-of-satisfaction* (CoS)
  - => the motor flavor of intentionality
- *mind-to-world*: the intentional state must match the state of the world to fulfill the CoS
  - => the perceptual flavor of intentionality

# From the logical definition of intentionality to neural processes

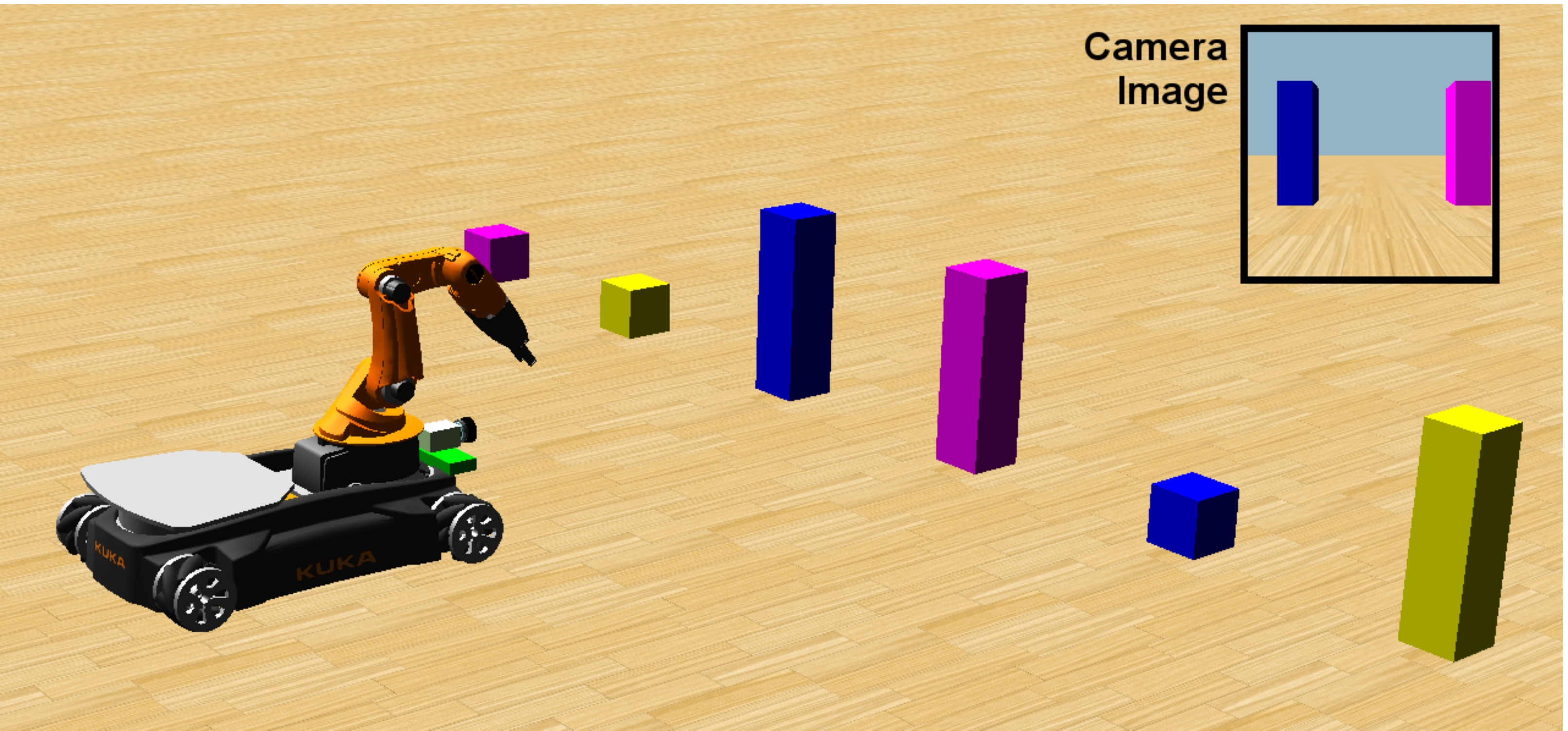
- *CoS of world-to-mind* (motor) intentionality
  - control the sequential unfolding of actions
  - intention critical to initiate actions
  - CoS is critical to terminate action intentions
- *CoS of mind-to-world* (perceptual) intentionality
  - the intentional state itself must match the state of the world => is its own CoS… arises with the intentional state
  - the match is a property of the process
  - possibility of error (e.g. mis-perception)

# Searle's six *psychological modes*

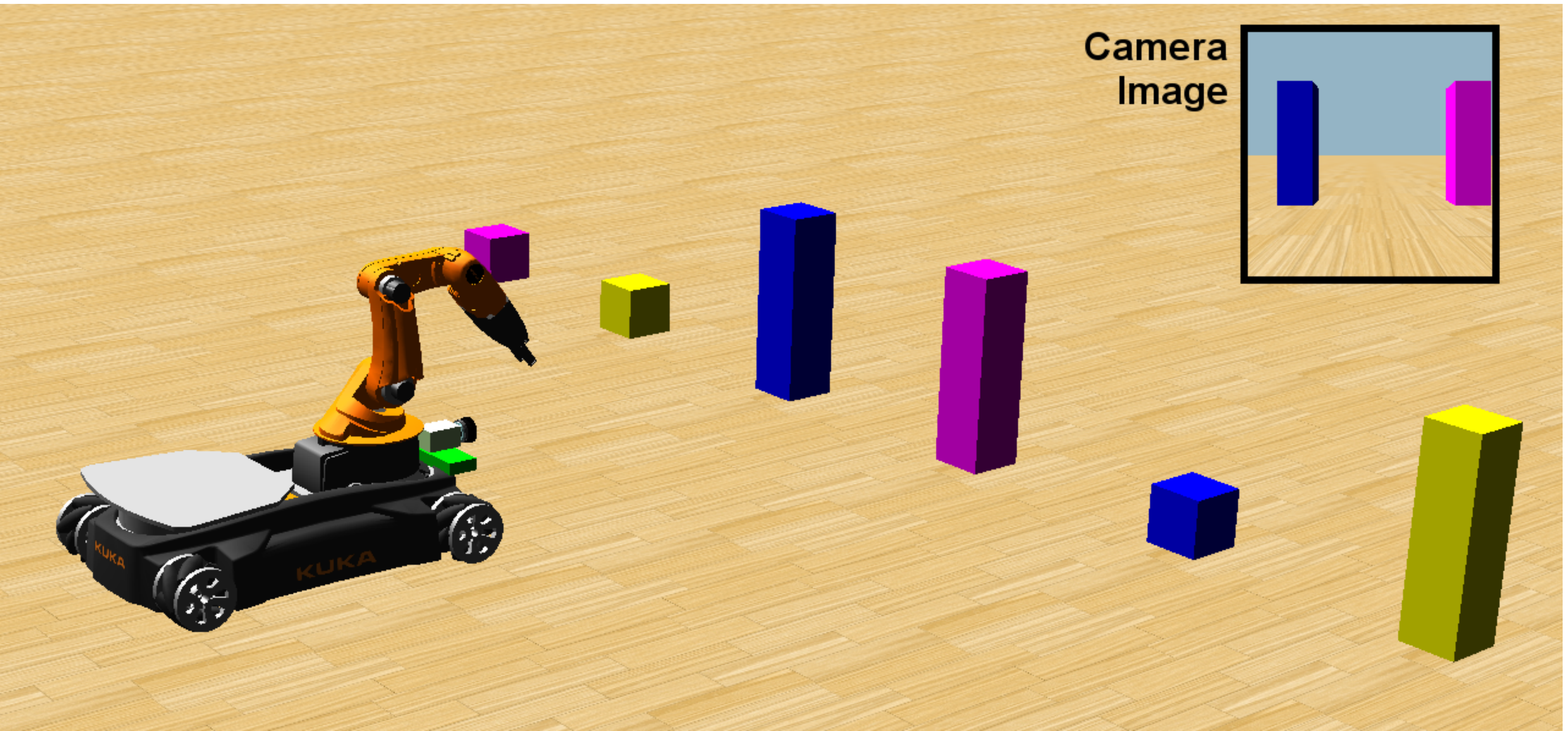
- *mind-to-world*
  - *perception*
  - *memory*
  - *belief*
- *world-to-mind*
  - *intention-in-action*
  - *prior intention*
  - *desire*

- as a heuristic for building cognitive architectures …
- that reflect the sensory-motor basis of cognition

# Illustration: a neural dynamic intentional agent in a simple world



# Scenario: intentional agent in simple world

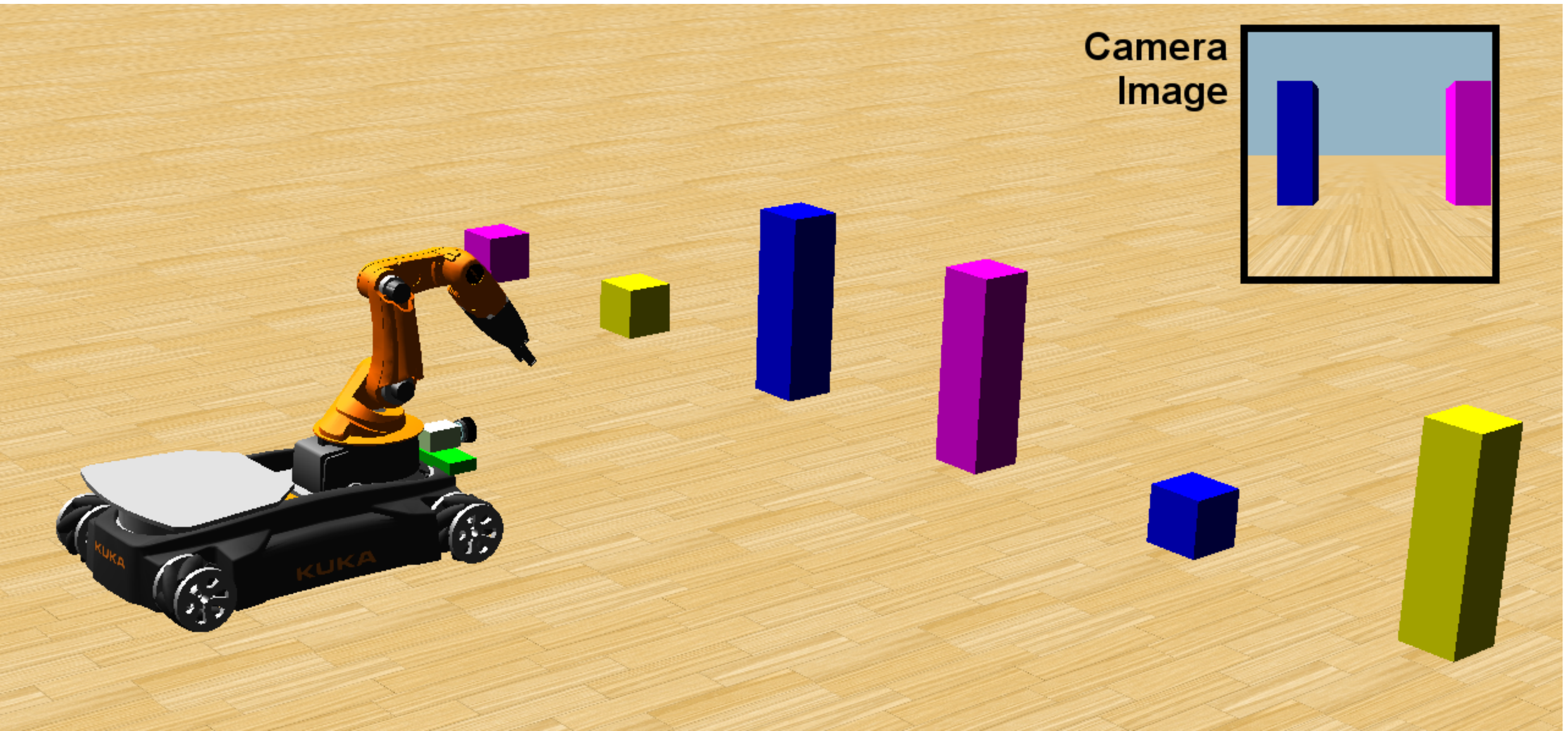



- world
  - colored objects (small)
  - paint buckets (tall)
  - vehicle with arm
- perception
  - see color/height feature
  - sense position, arm, paint in gripper
- intention in action
  - move in 1D
  - reach to take up paint
  - reach to apply a coat of paint

# Scenario: intentional agent in simple world

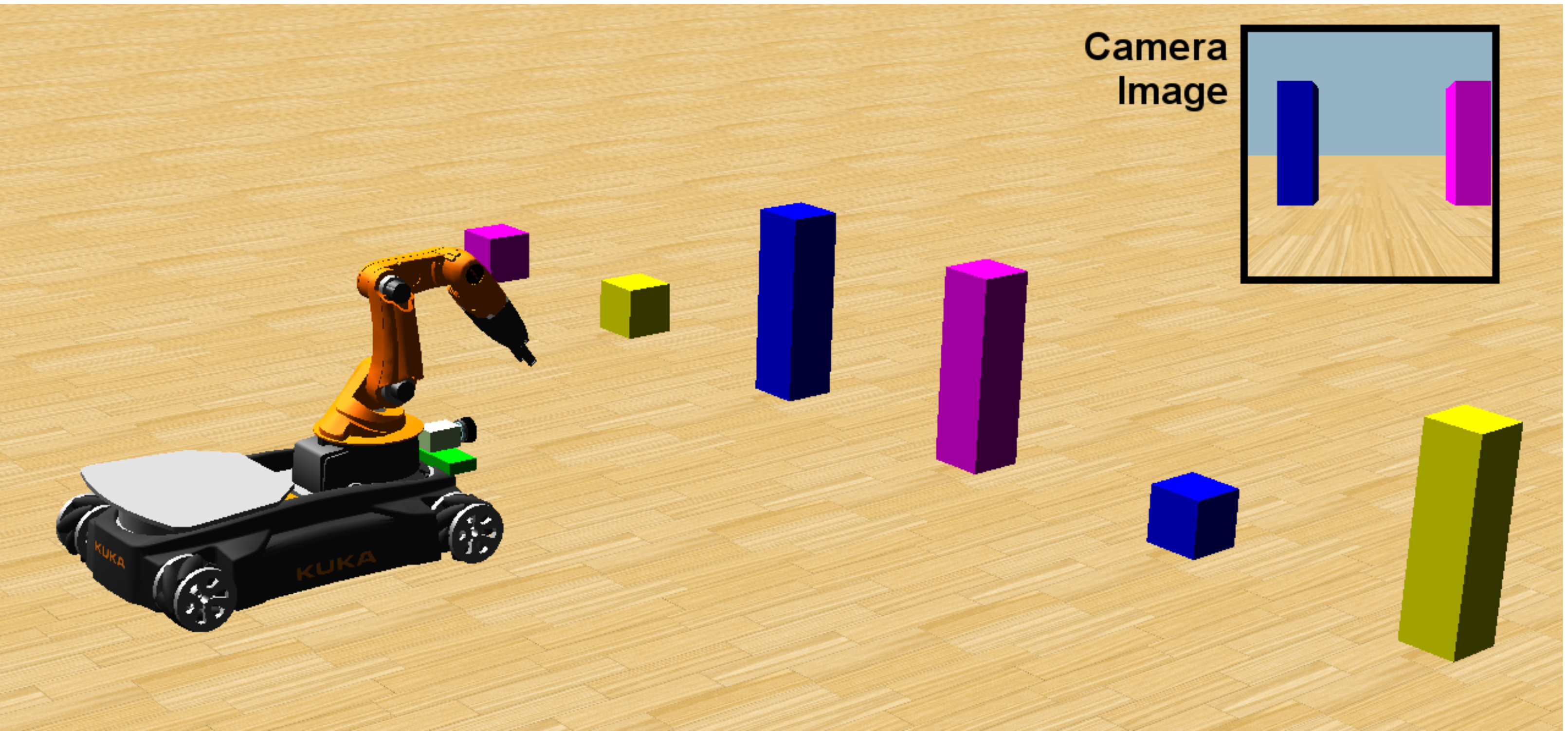
- memory
  - of the scene (feature maps)
- prior intentions
  - search to paint
  - search to load paint
  - reach to apply paint
  - move to a recalled location

  …

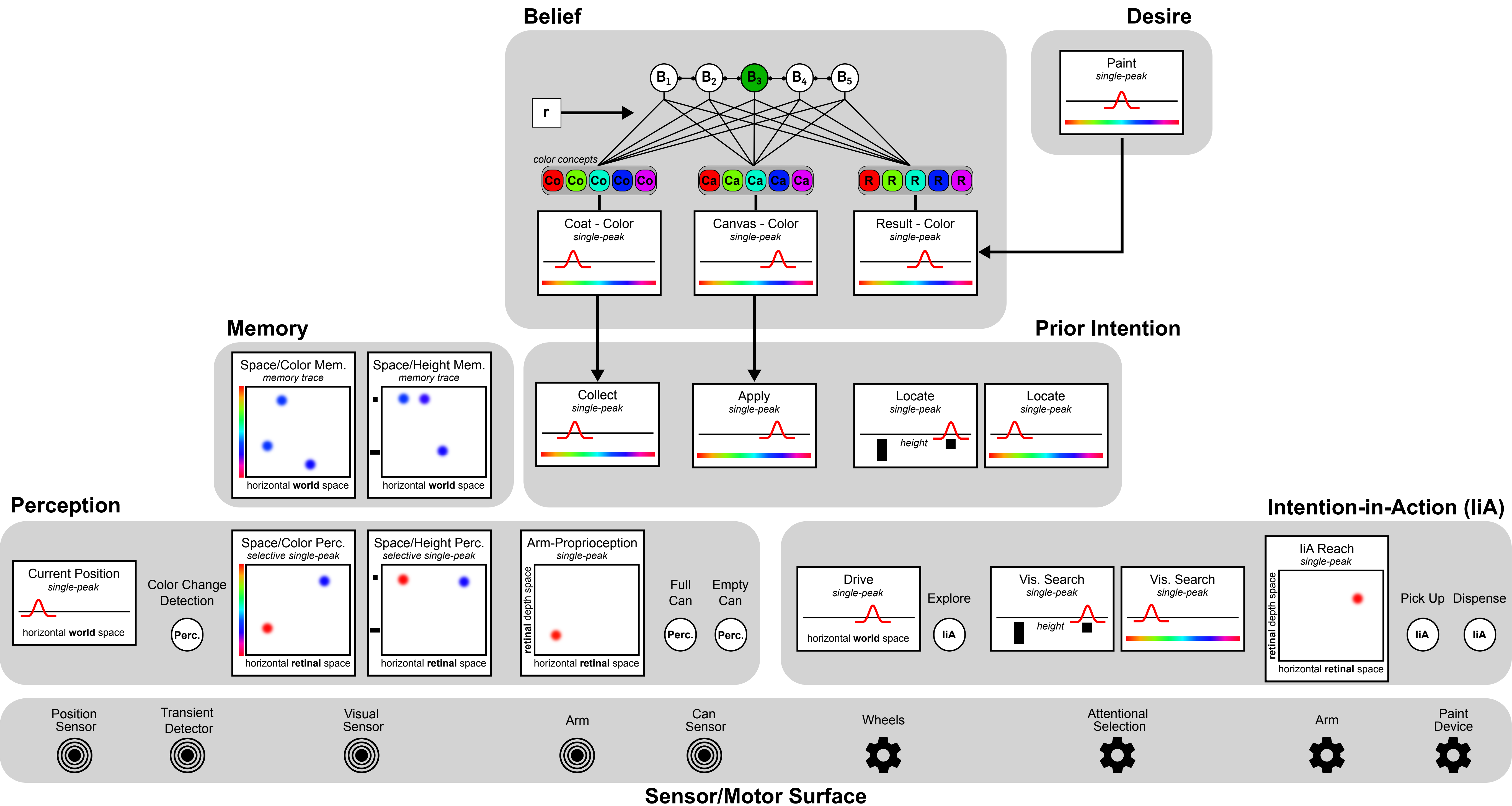


# Scenario: intentional agent in simple world

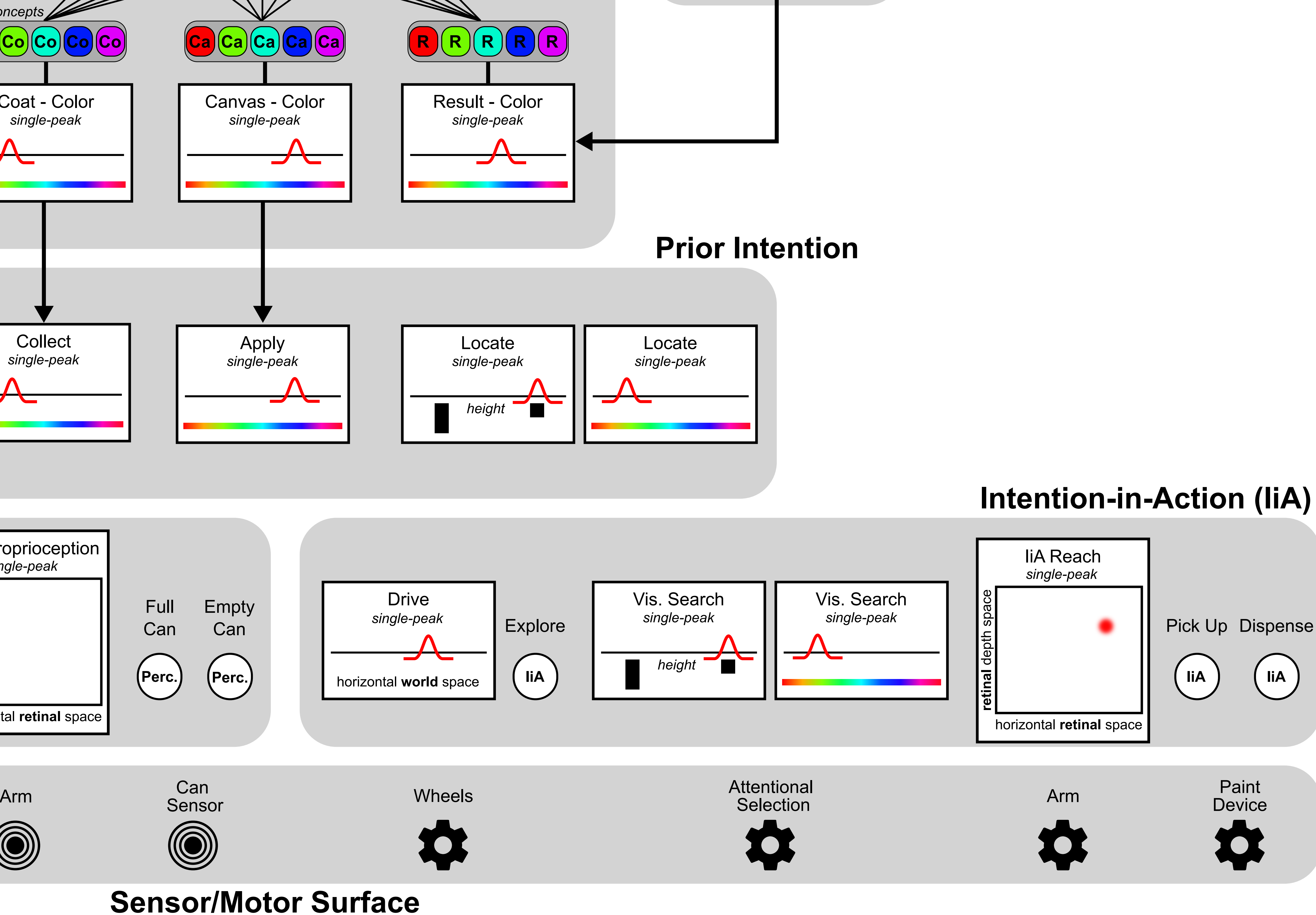
- beliefs
  - rules that link color concepts: which paint on which canvas generates which outcome color
- desires
  - for particular colors



# Neural dynamic architecture

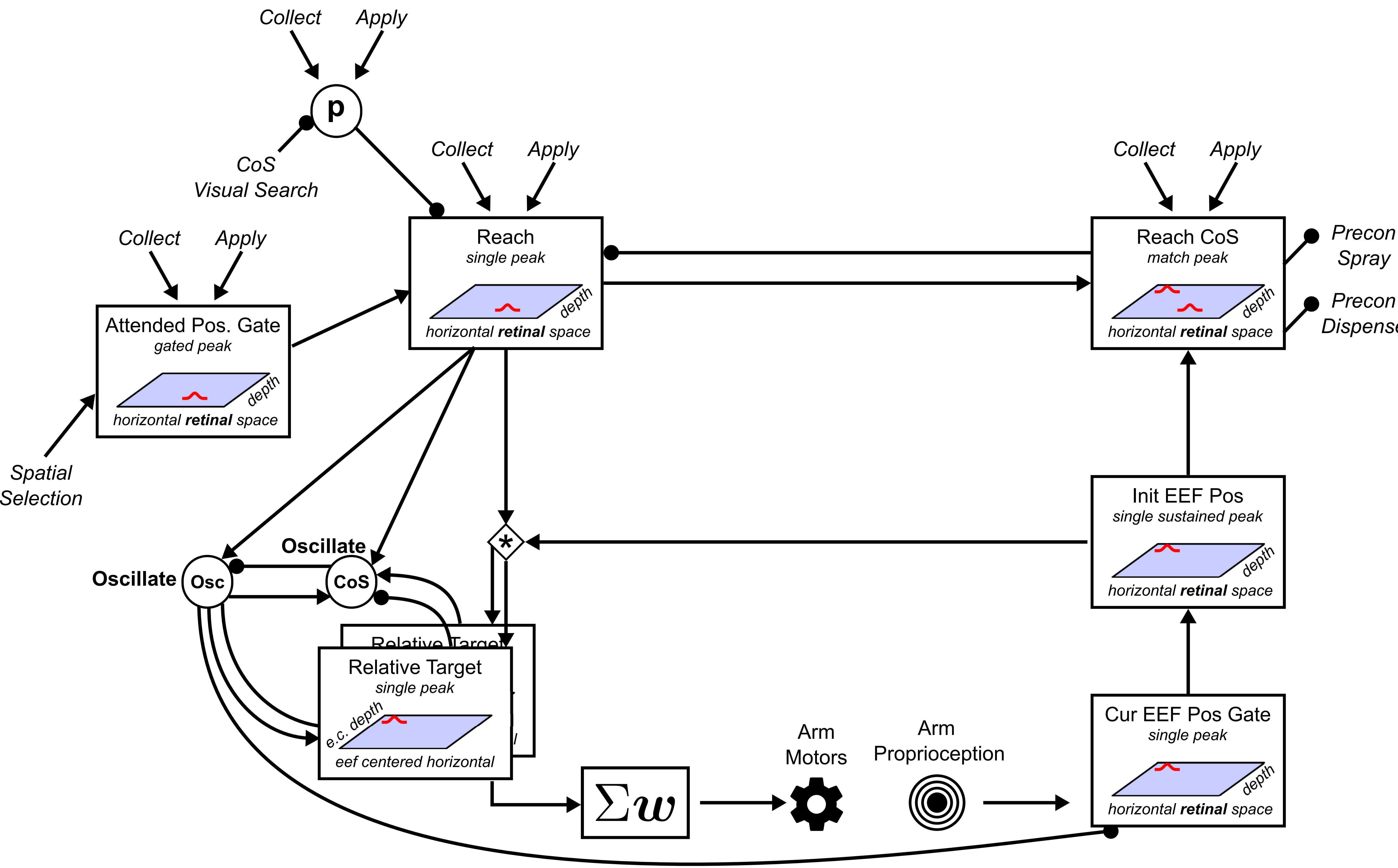


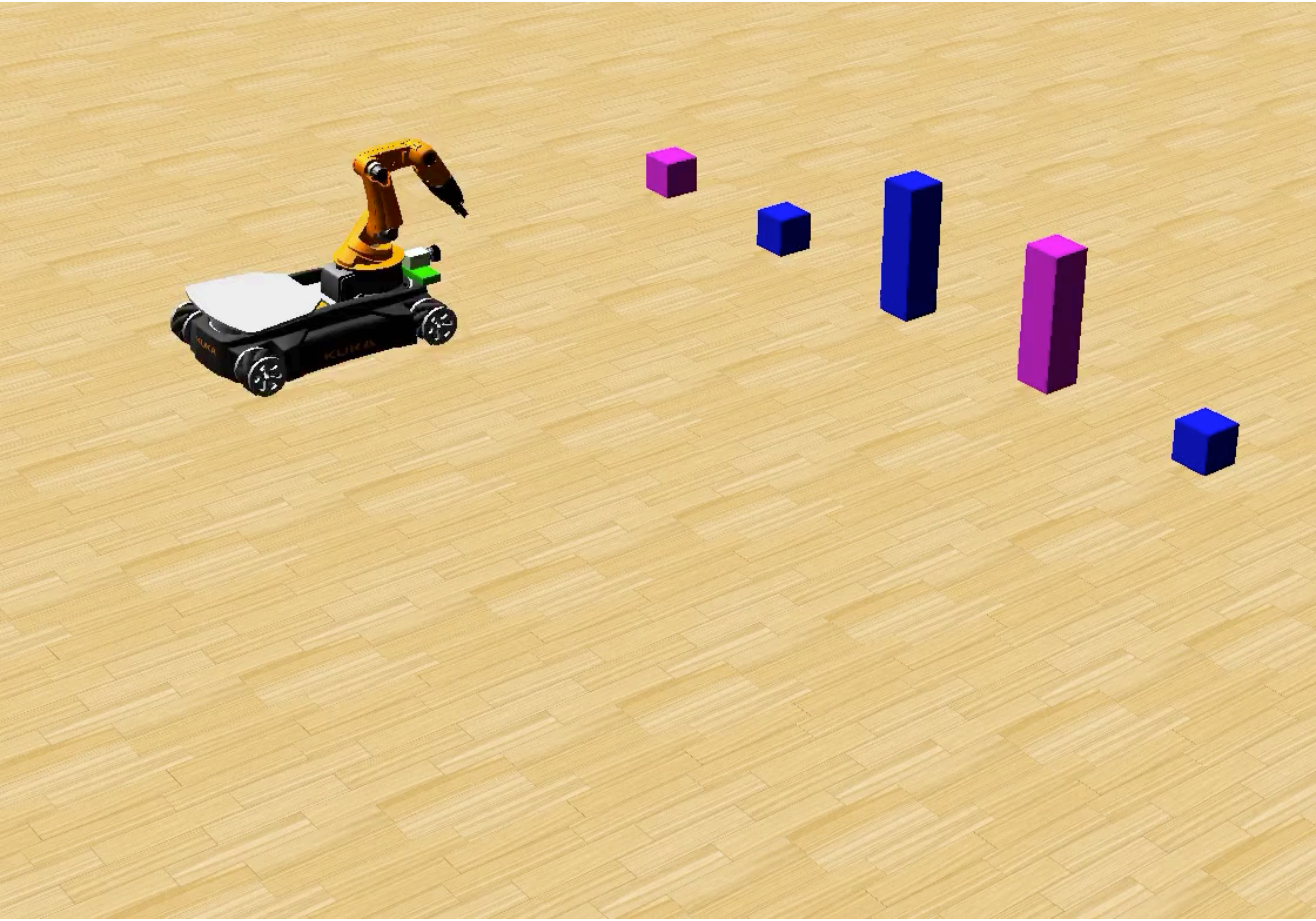

[Tekülve, Schöner: IEEE Trans Cog and Dev Sys (2020)]

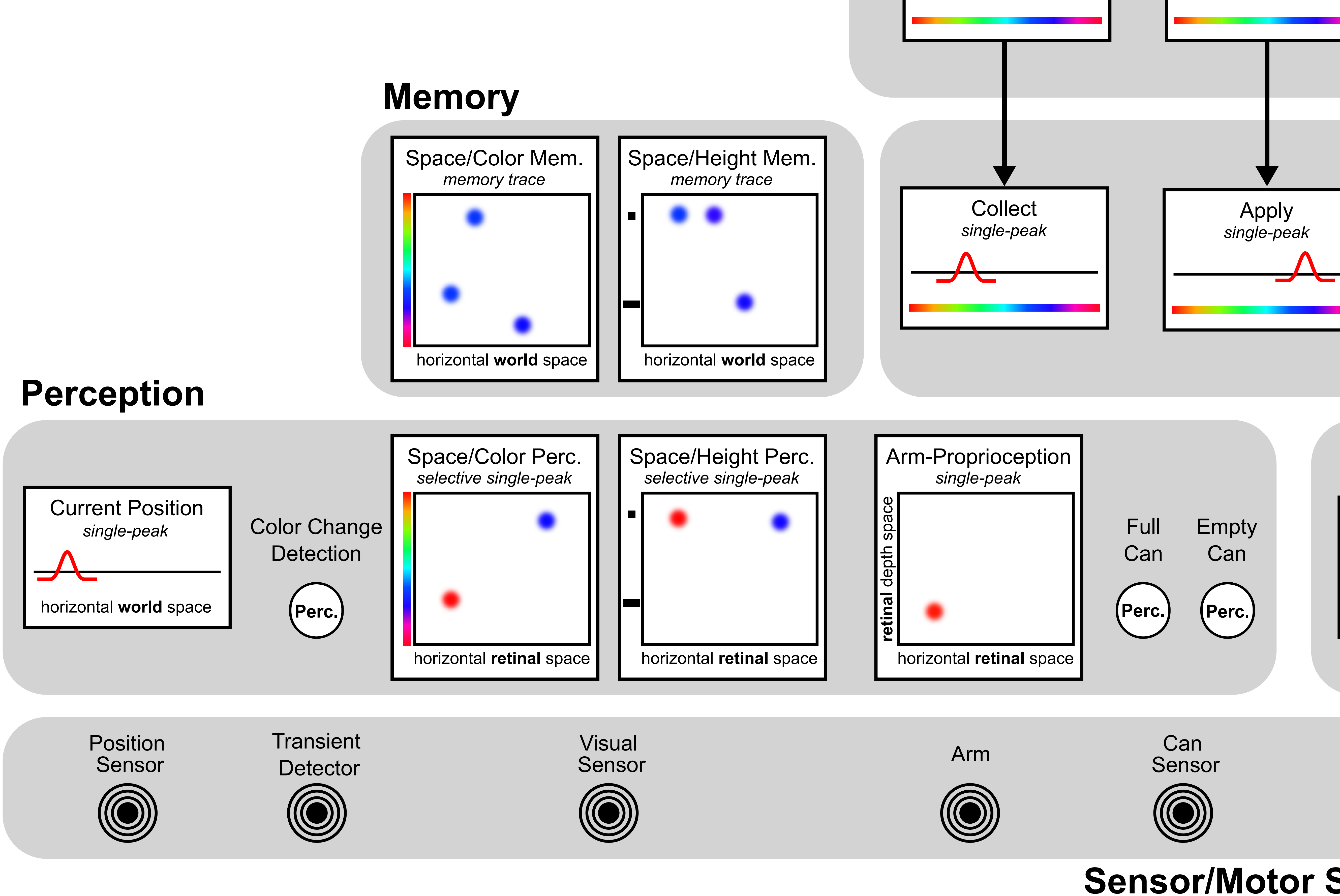


oncepts
Co Co Co Co
Ca Ca Ca Ca Ca
R R R R R
Coat - Color
single-peak
Canvas - Color
single-peak
Result - Color
single-peak
Prior Intention
Collect
single-peak
Apply
single-peak
Locate
single-peak
height
Locate
single-peak
Intention-in-Action (IiA)
roprioception
ngle-peak
tal retinal space
Full Can
Empty Can
Perc.
Perc.
Drive
single-peak
horizontal world space
Explore
IiA
Vis. Search
single-peak
height
Vis. Search
single-peak
IiA Reach
single-peak
retinal depth space
horizontal retinal space
Pick Up
Dispense
IiA
IiA
Arm
Can Sensor
Wheels
Attentional Selection
Arm
Paint Device
Sensor/Motor Surface

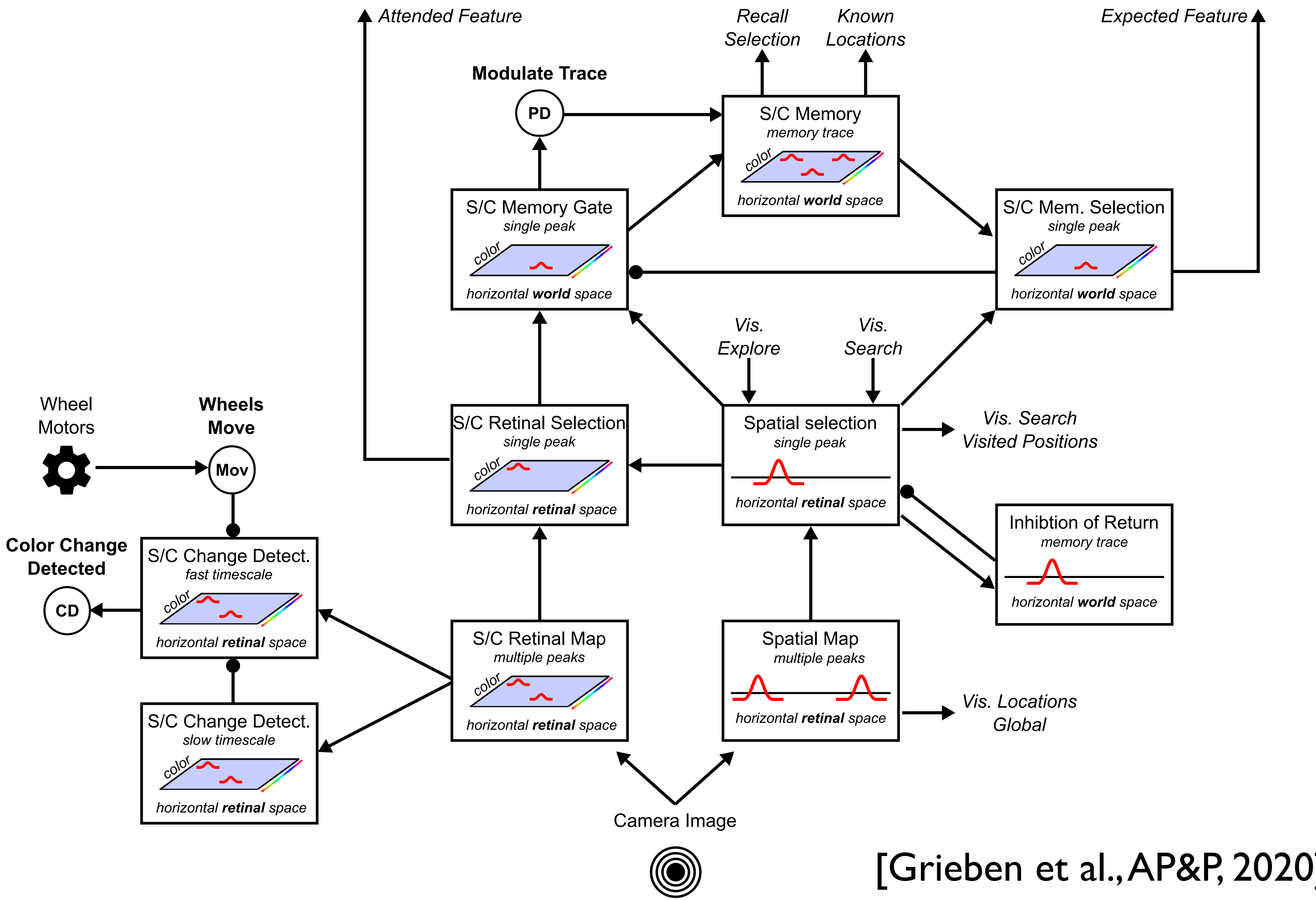
# Intention in action: reach



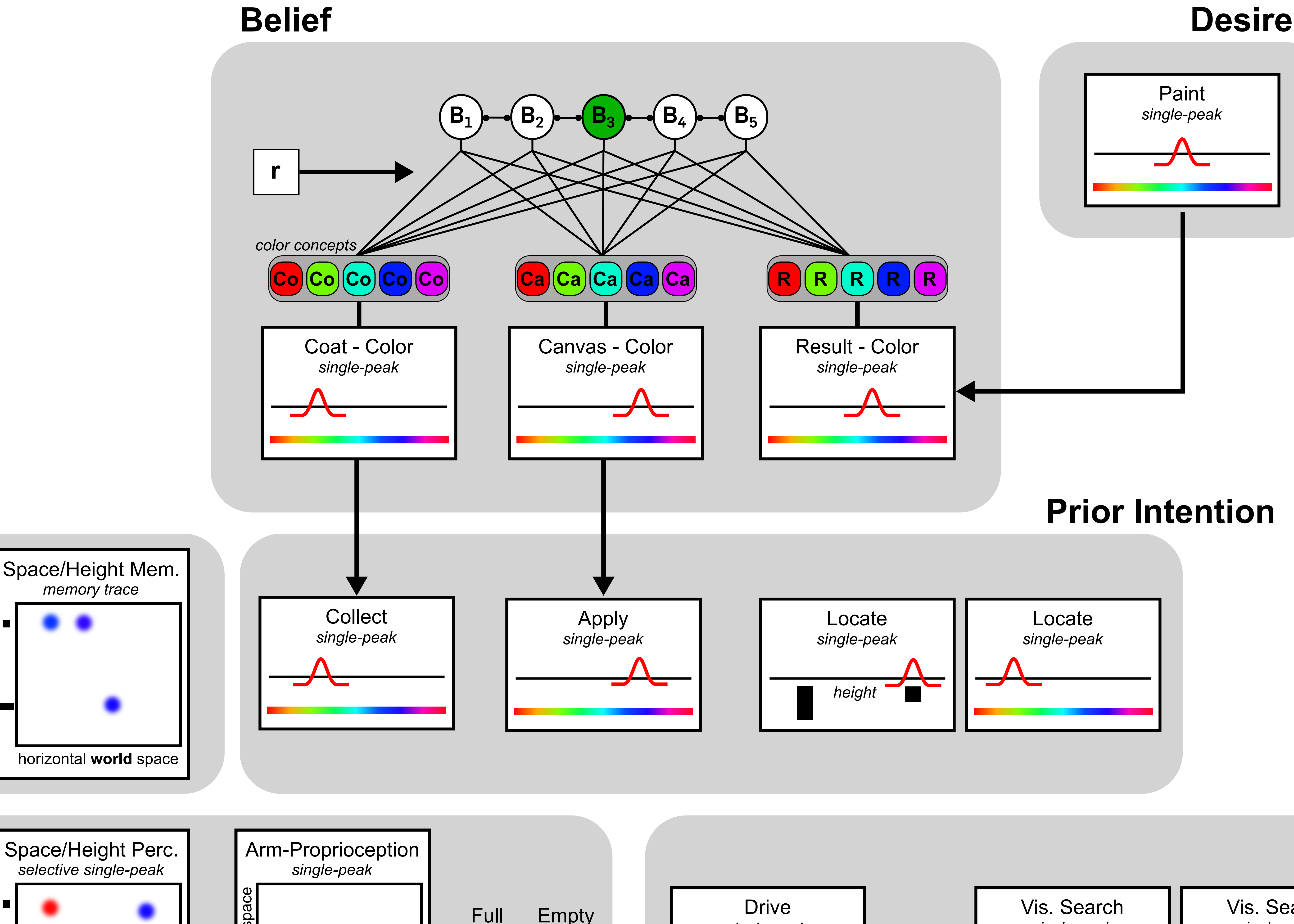



Memory
Space/Color Mem.
memory trace
horizontal world space
Space/Height Mem.
memory trace
horizontal world space
Collect
single-peak
Apply
single-peak
Perception
Current Position
single-peak
horizontal world space
Color Change
Detection
Perc.
Space/Color Perc.
selective single-peak
horizontal retinal space
Space/Height Perc.
selective single-peak
horizontal retinal space
Arm-Proprioception
single-peak
retinal depth space
horizontal retinal space
Full
Can
Perc.
Empty
Can
Perc.
Position
Sensor
Transient
Detector
Visual
Sensor
Arm
Can
Sensor
Sensor/Motor S

# Perception and memory




[Grieben et al., AP&P, 2020]



Belief
Desire
Paint
single-peak
B1
B2
B3
B4
B5
r
color concepts
Co
Ca
R
Coat - Color
single-peak
Canvas - Color
single-peak
Result - Color
single-peak
Prior Intention
Space/Height Mem.
memory trace
horizontal world space
Collect
single-peak
Apply
single-peak
Locate
single-peak
height
Locate
single-peak
Space/Height Perc.
selective single-peak
Arm-Proprioception
single-peak
space
Full
Empty
Drive
Vis. Search

# Intentional systems

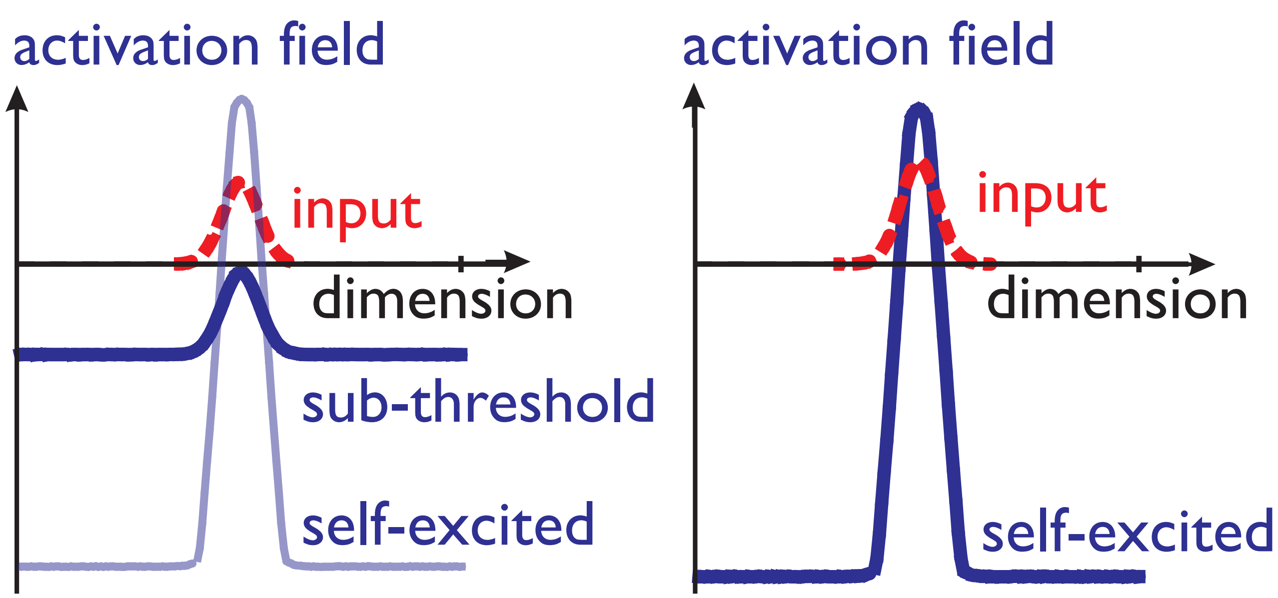
- => special lecture Jan Tekülve on Friday

# What does it all mean…

- why do neural dynamic architectures work?
- how do embodied (neural dynamic) architectures relate to classical cognitive architectures ?
- what does embodiment mean?
- how does DFT relate to deep NN, to VSA?

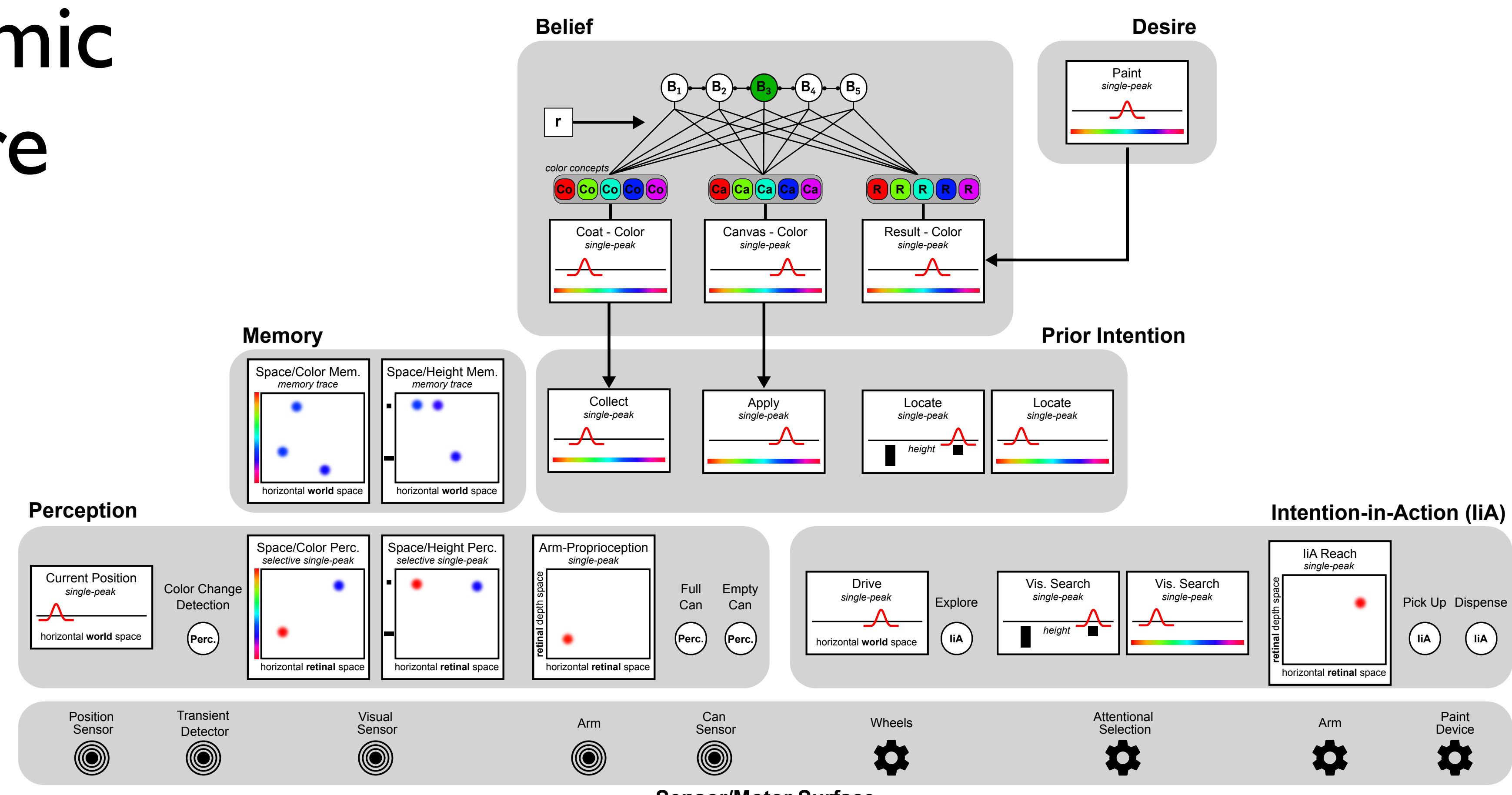
# DFT architectures

- why are attractors and their instabilities preserved as fields are coupled into architectures?
- stability => structural stability = invariance of solutions under change of the dynamics
- => dynamic modularity: fields retain their dynamic regime as activation elsewhere varies



# DFT architectures

- why do fields retain their meaning…
- coupling among fields must preserve the fields' dimensions: "non-synesthesia principle"
- informational modularity (encapsulation)
- => neural dynamic architectures are specific = constrained by evolution and development



# What does "embodiment" mean?

- cognition activates motor systems?
- cognition is based on sensor systems?
- not necessarily!

# What does "embodiment" mean?

- continuous state, continuous time
- continuous/intermittent link to the sensory and motor surfaces is *possible*
- closed loop => stability!

# Embodiment hypothesis

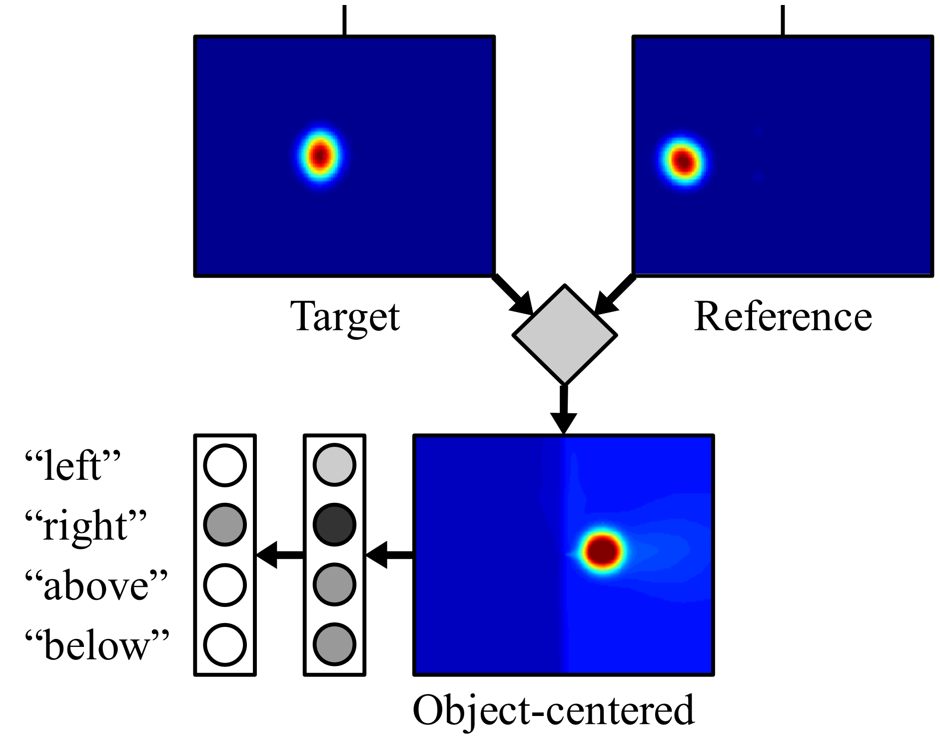
- all cognitive processes inherit the dynamic properties of sensory-motor cognition: stability, instabilities…

- cognition is embedded in the specific embodied cognitive architectures that emerged in evolution/development

# How is higher cognition reached?

- attentional selection, coordinate transformation, sequential processing … emulates "function calls"

$$\text{to the left of} = f(\text{target}, \text{reference})$$

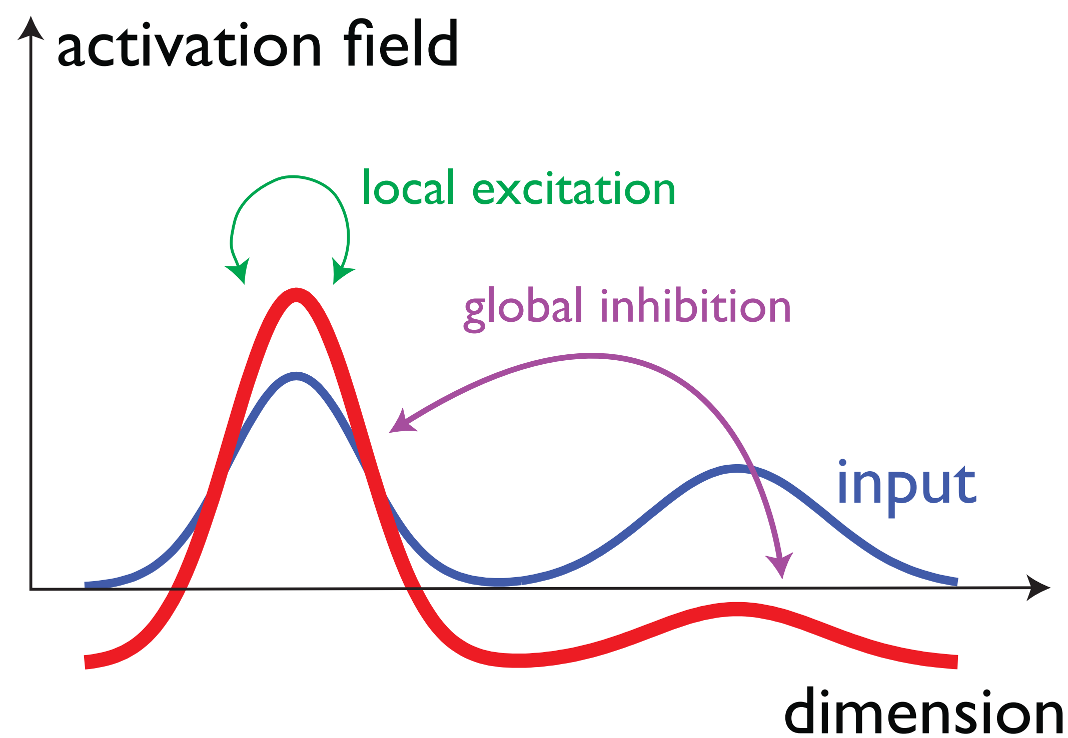



- … not as flexible as symbol manipulation and costly in processing structure …
- but all concepts are grounded by their very nature…

# Localist vs. distributed

- DFT hypothesis: all autonomous cognition happens in localist representations which are necessarily low-dimensional

- they don't have to be easy to grasp and observe
- they could be latent representations
- high-dimensional distributed representations subserve primarily classification, which is embedded in the neural dynamics of competing nodes



# DFT vs VSA

- Vector-symbolic architectures (VSA) are a theoretical alternative
- in the original version (Smolensky): role-filler binding… compatible with DFT
- in the Gayler/Kanerva/Plate version: high-dimensional vectors as symbols that afford binding, and function calling … not neurally feasible: autonomy
- requires that the symbol grounding problem is solved at encoding/decoding

# DFT vs VSA

- Eliasmith's Neural Engineering Framework (NEF) as a possible neural implementation of VSA
  - vectors represented by (small) populations of spiking neural networks
- NEF is "model neutral"… essentially a method to "numerically" implement any neural model
- But: to preserve the original vectors, connectivity in VSA/NEF (SPAUN) architectures is very special => non-local dependence of connectivities on each other…

# Outlook/challenges

- sequences of relational concepts that interrelate, exchange arguments, have hierarchical structure
  - “the box to the right of the bottle that stands under the lamp”
- sequences of actions that are directed at goals, and have hierarchical structure
  - “open the box to get the screwdriver with which you remove the screw to take of the cover of the toaster…”
- goals and their dynamics, motivation…
  - emotions…