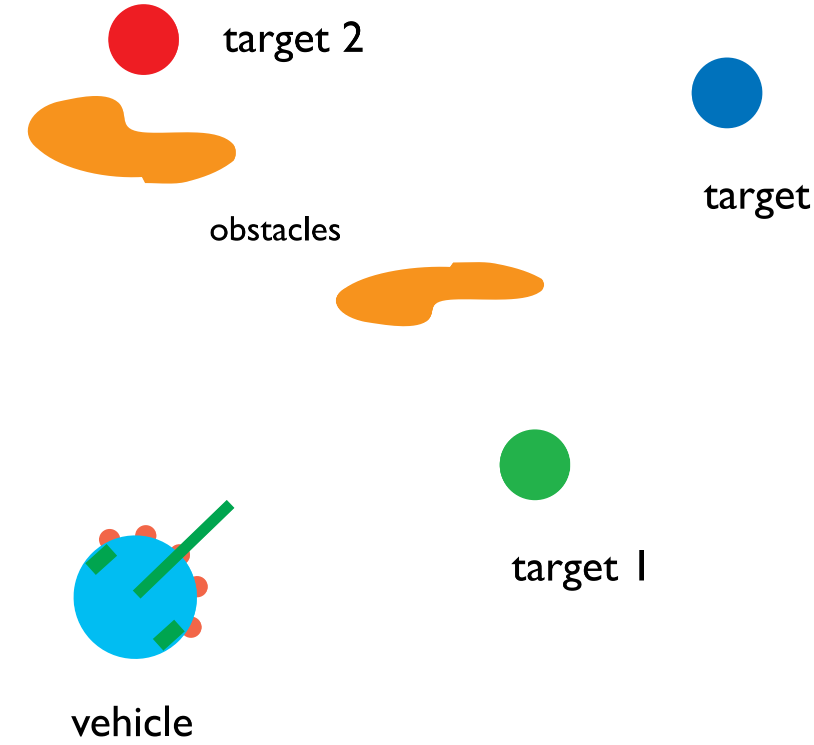


Dynamic Field Theory: autonomy

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

Roadmap

- sequences
- why do DFT architectures work?
- embedding DFT in the theoretical landscape

Sequences

- all behavior and thinking consist of sequences of physical or mental acts
- sometimes in a fixed order as in action routines, or highly trained action patterns
- but potentially highly flexible … as in language, thinking, problem solving …

DFT challenge for sequences

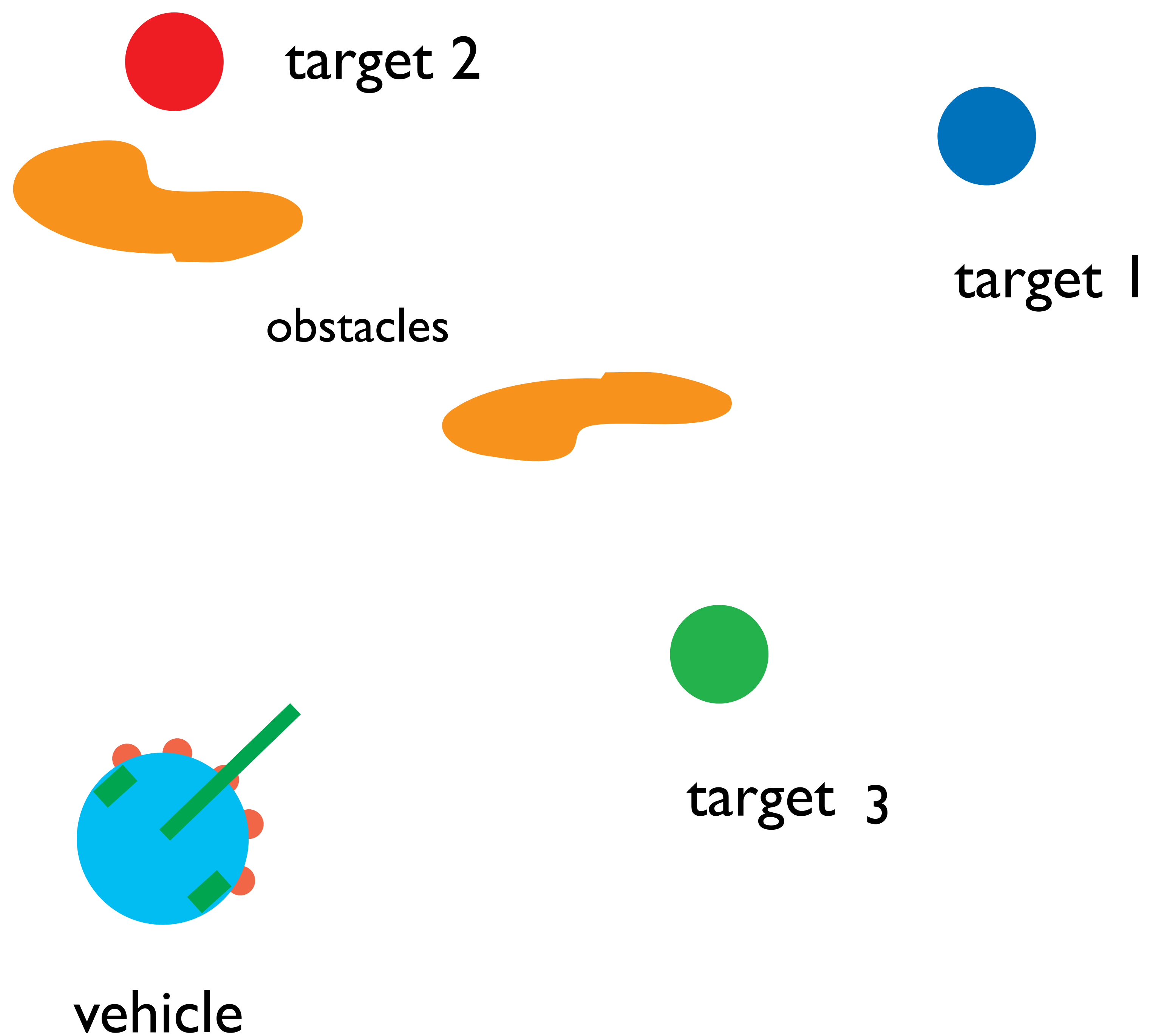
- DFT postulates that all neural states underlying behavior/mental process are attractors that resist change…
- but generating sequences of such states require change of state! Conflicting constraints!
- answer: instabilities are induced systematically to enable switching to a next/new attractor

Sequence generation

- an illustrative example
- the neural/mathematical mechanism

Sequence of physical acts

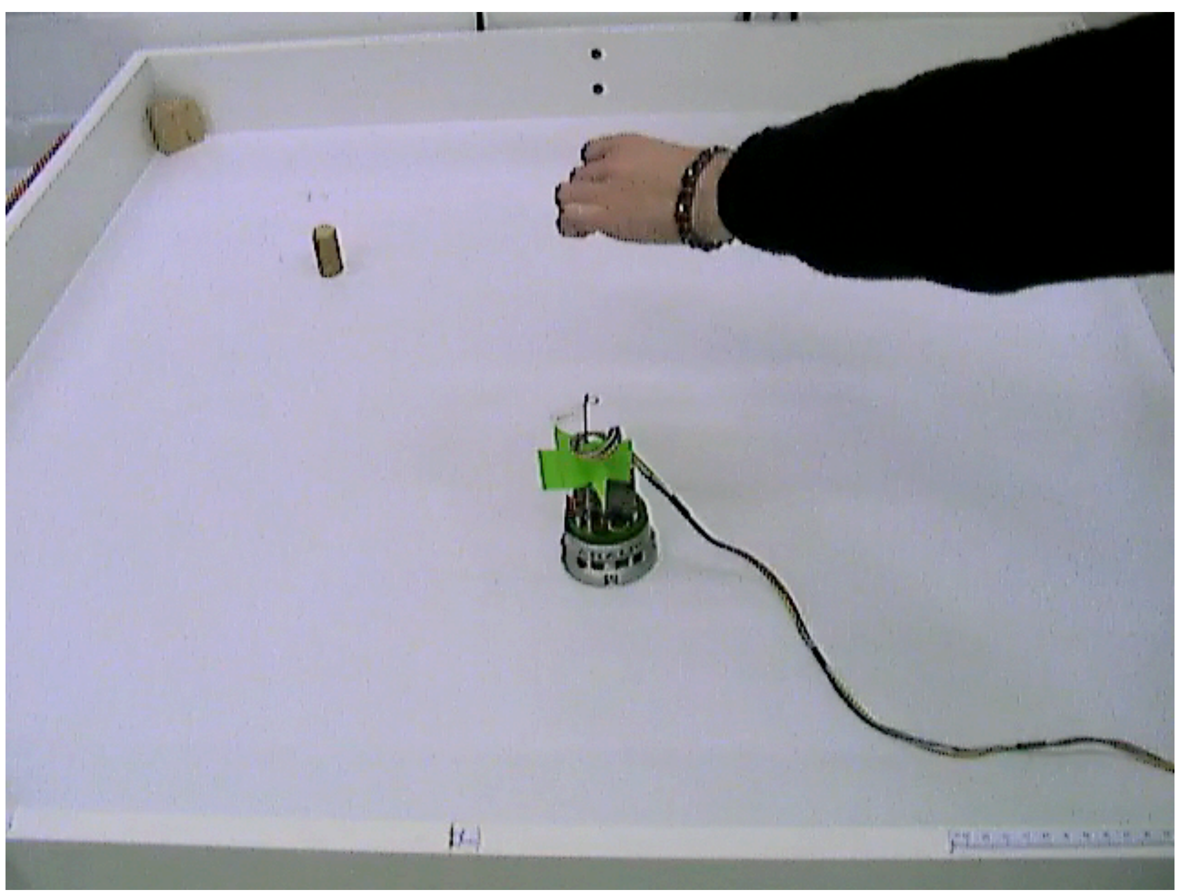
- task: search for objects of a given color in a given order
  - 1 blue
  - 2 red
  - green
- stably couple to objects once they are detected
- ignore objects when their turn has not yet come (distractors)



Implementation as an imitation task

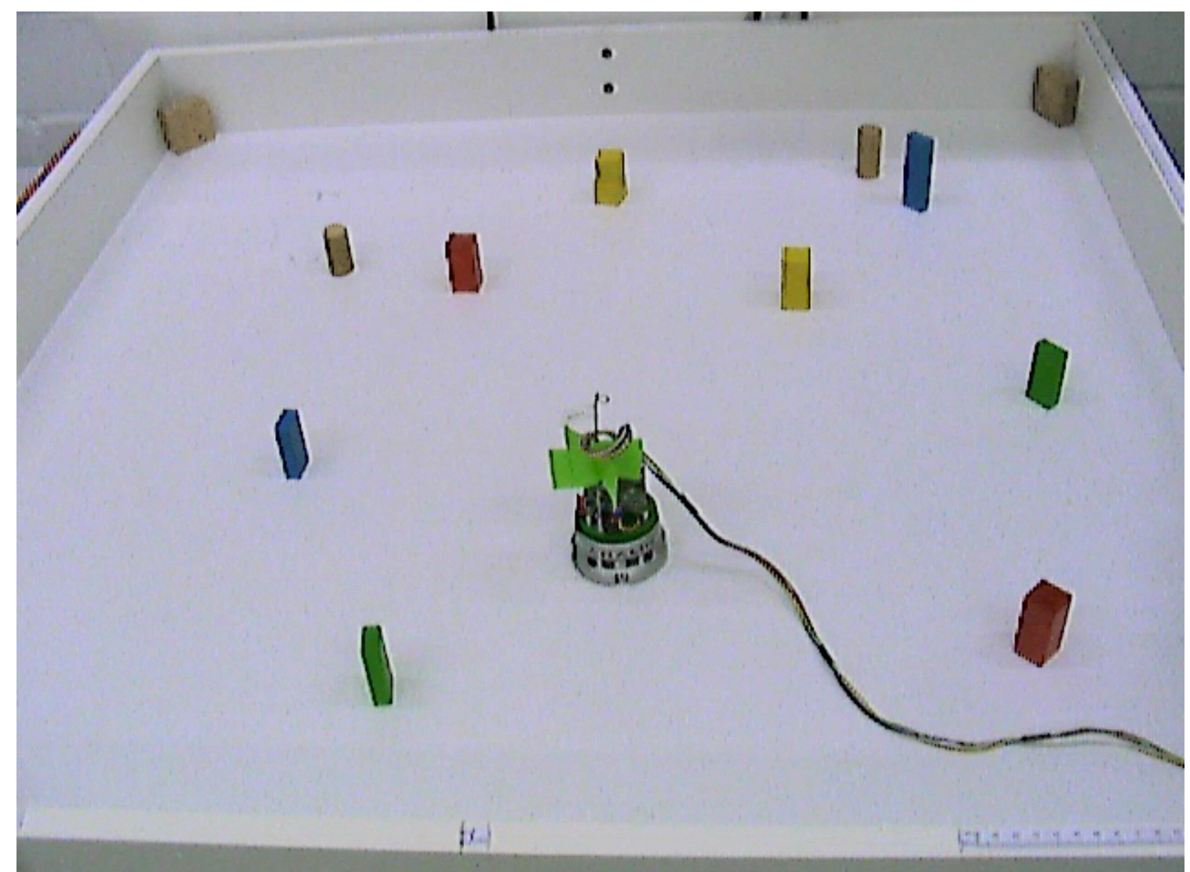
- learn a serially ordered sequence from a single demonstration

yellow-red-green-blue-red



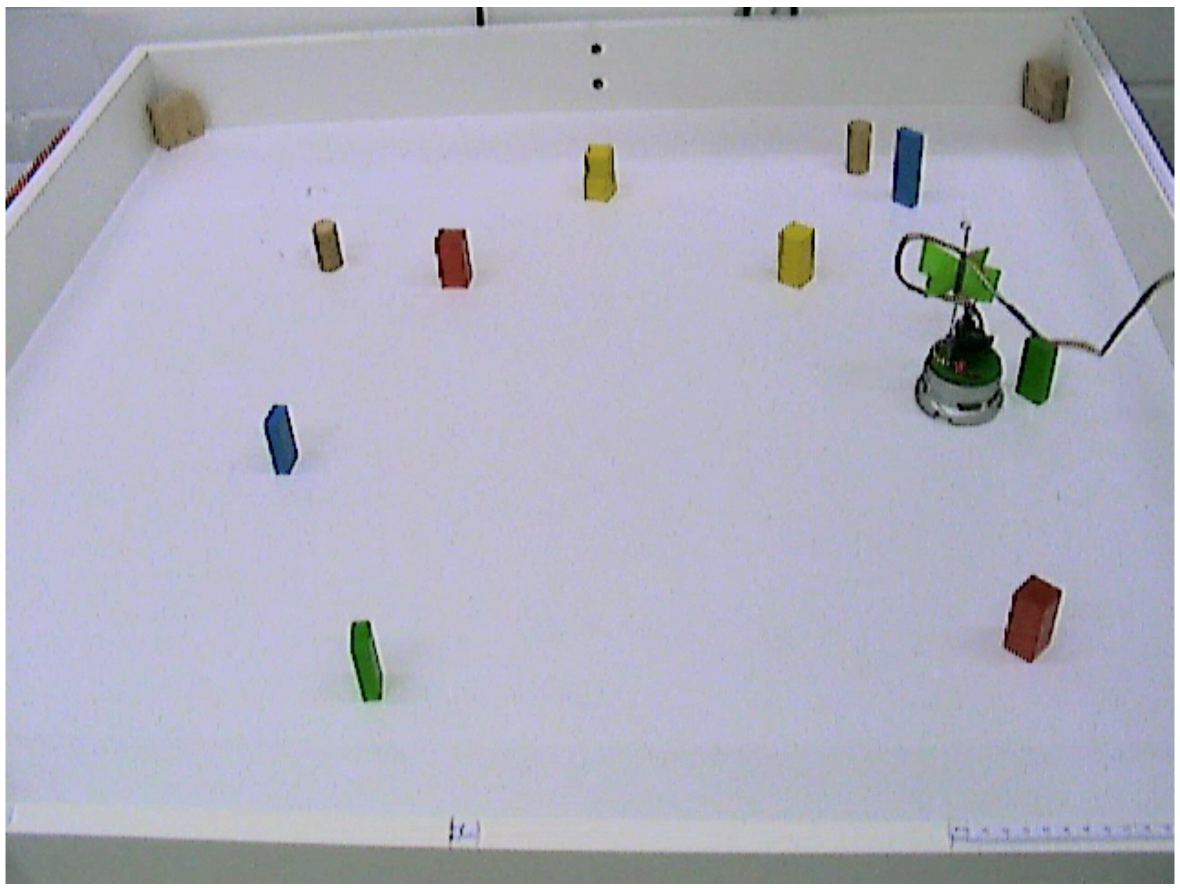
- perform the serially ordered sequence with new timing

yellow-red-green-blue-red

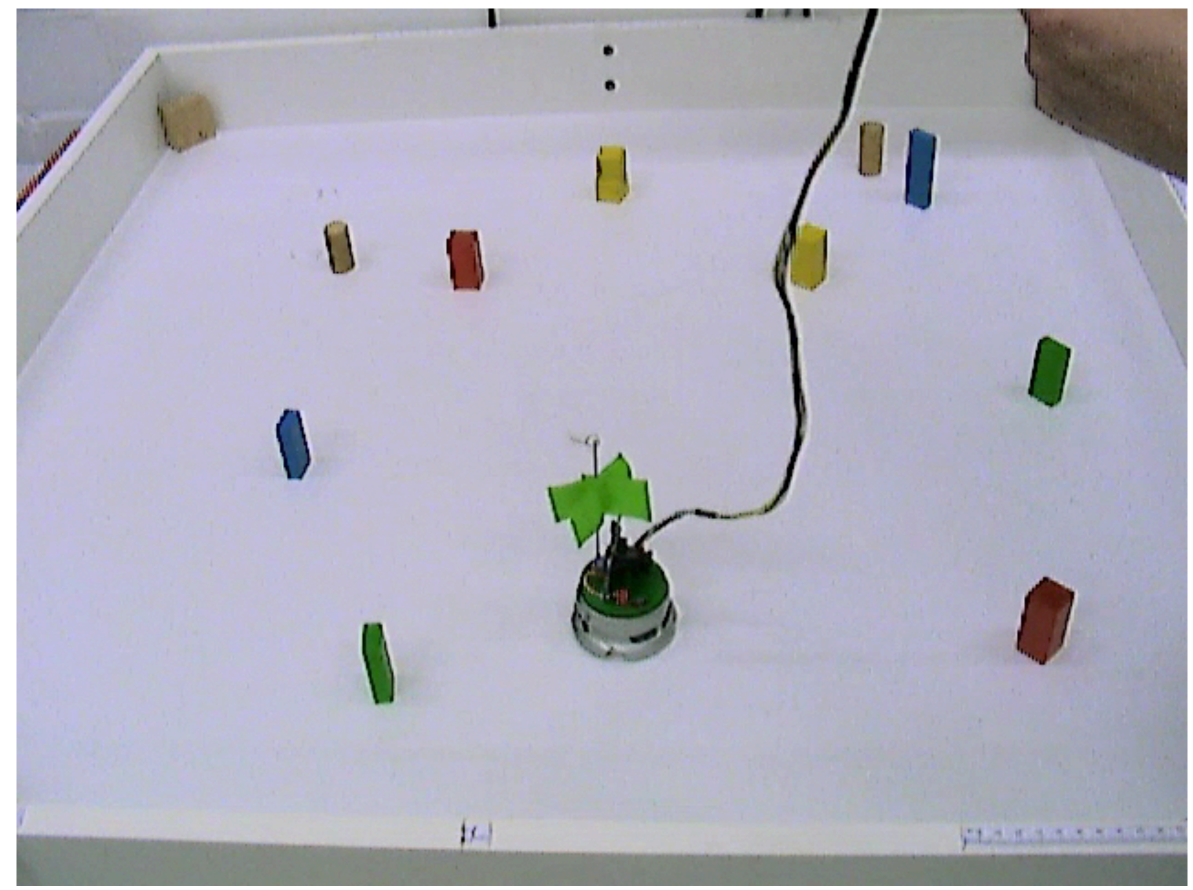


[Sandamirskaya, Schöner: *Neural Networks* 23:1163 (2010)]

red a distractor

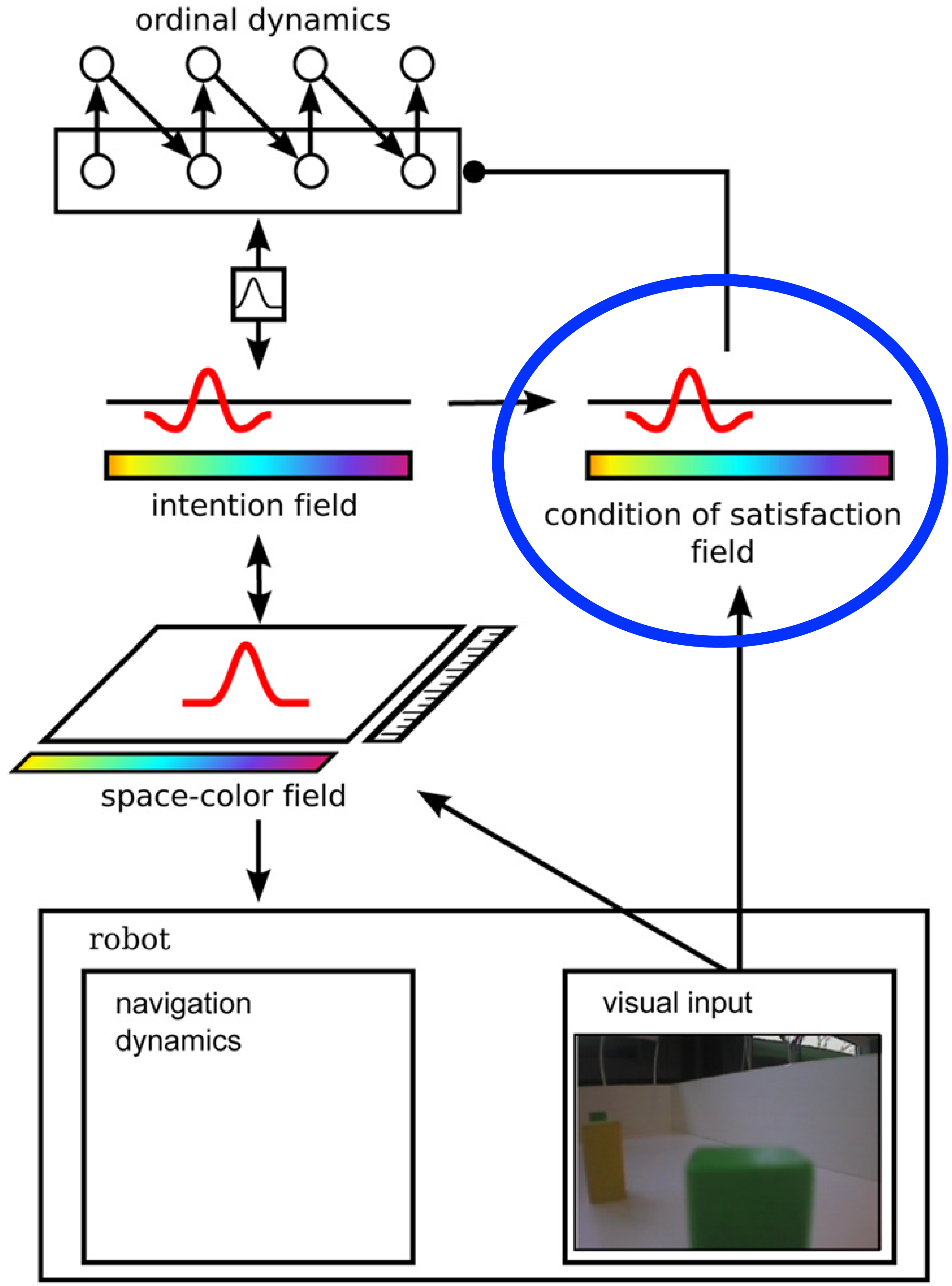


red a target



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

Condition of Satisfaction (CoS)

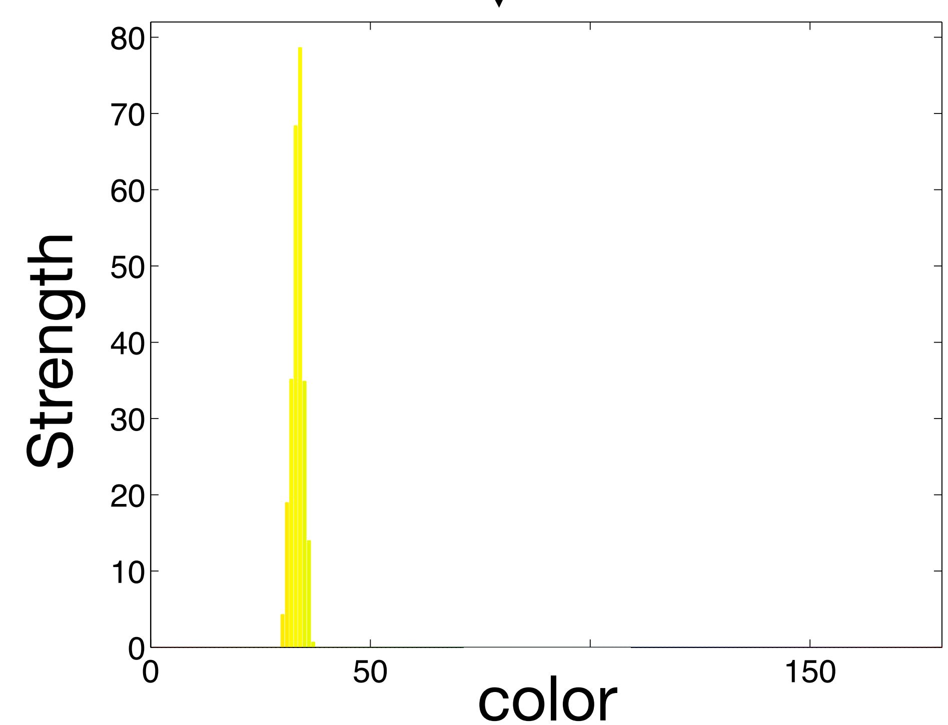


[Sandamirskaya, Schöner: *Neural Networks* 23:1163 (2010)]

Visual search

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

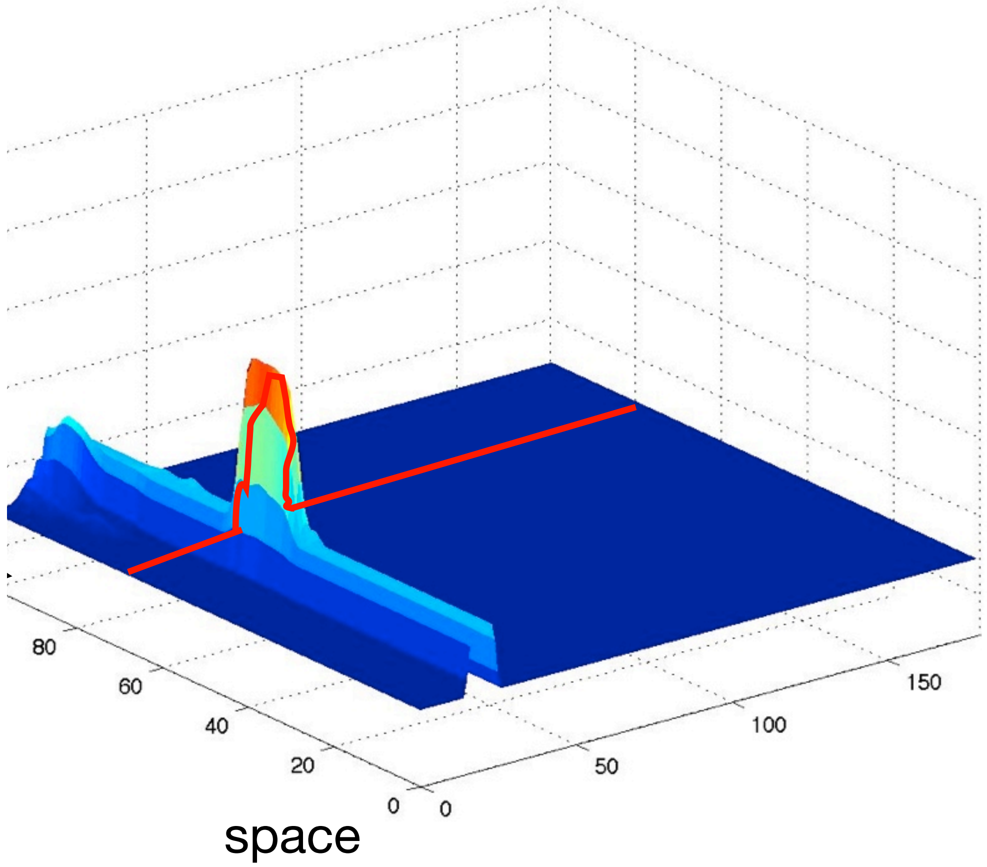
Camera image

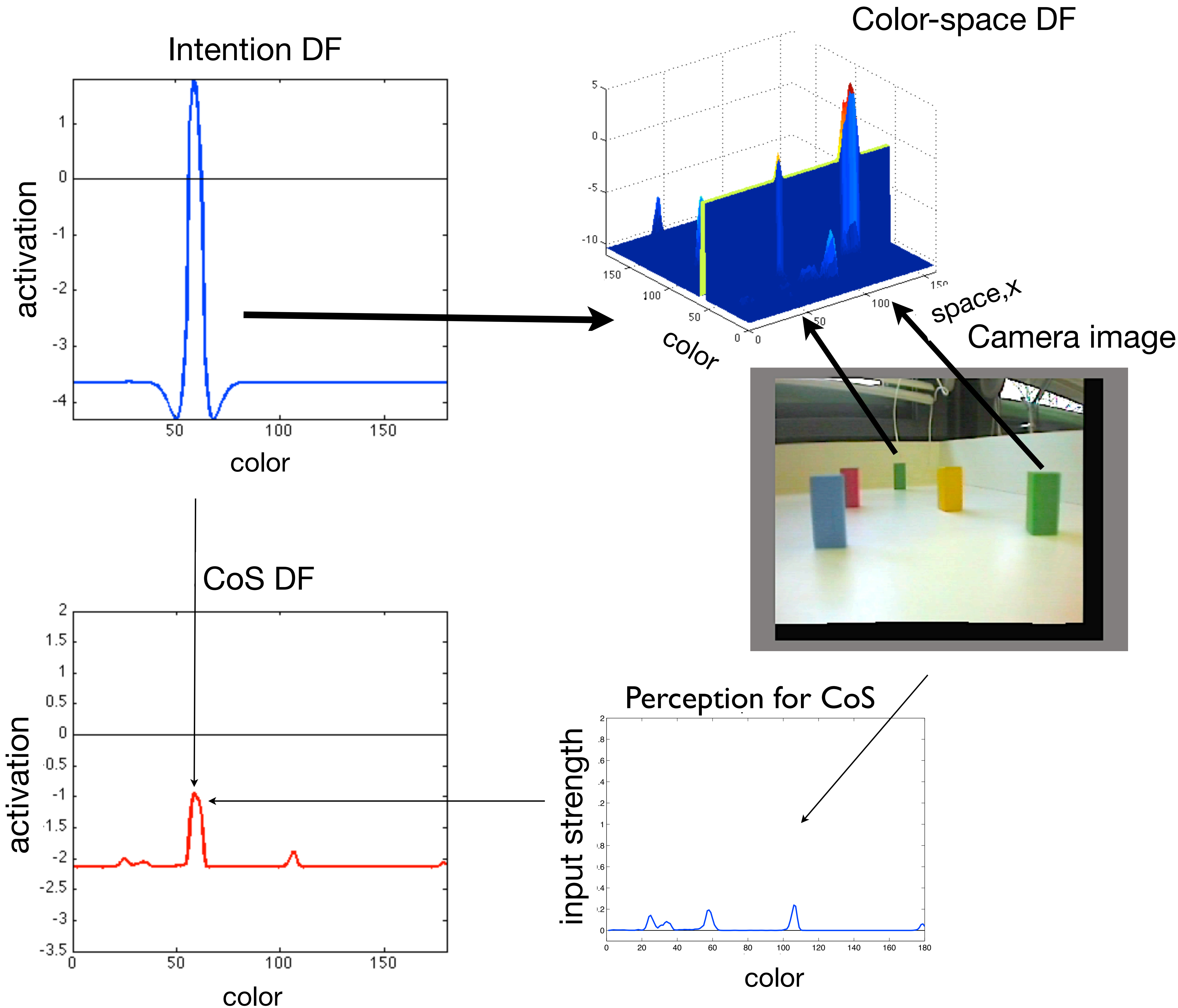


Visual search

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

search cue

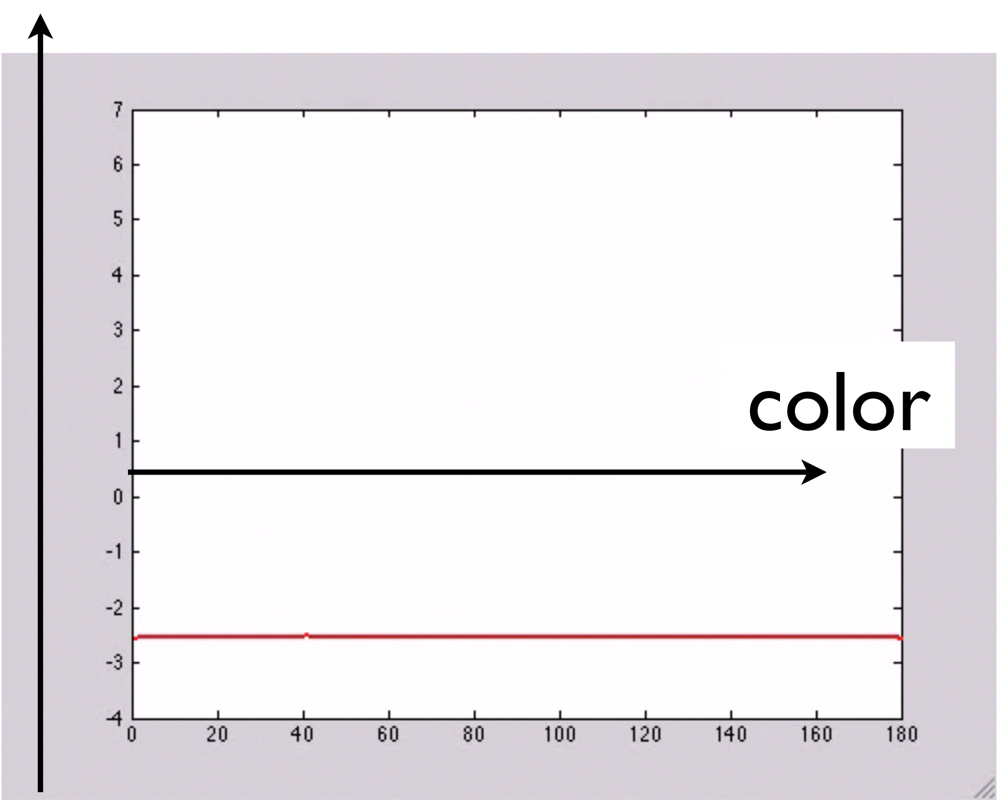



Intention DF
activation
color
Color-space DF
space,x
color
Camera image
CoS DF
activation
color
Perception for CoS
input strength
color

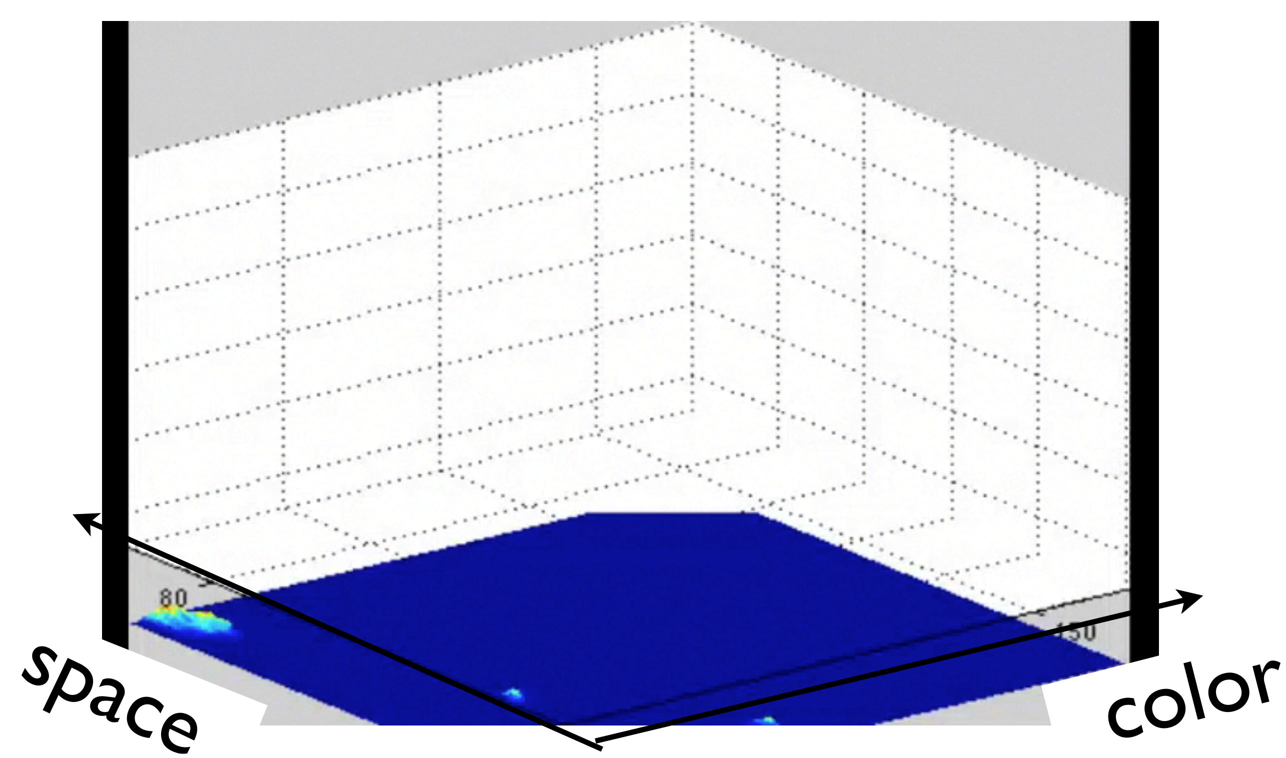
ordinal stack

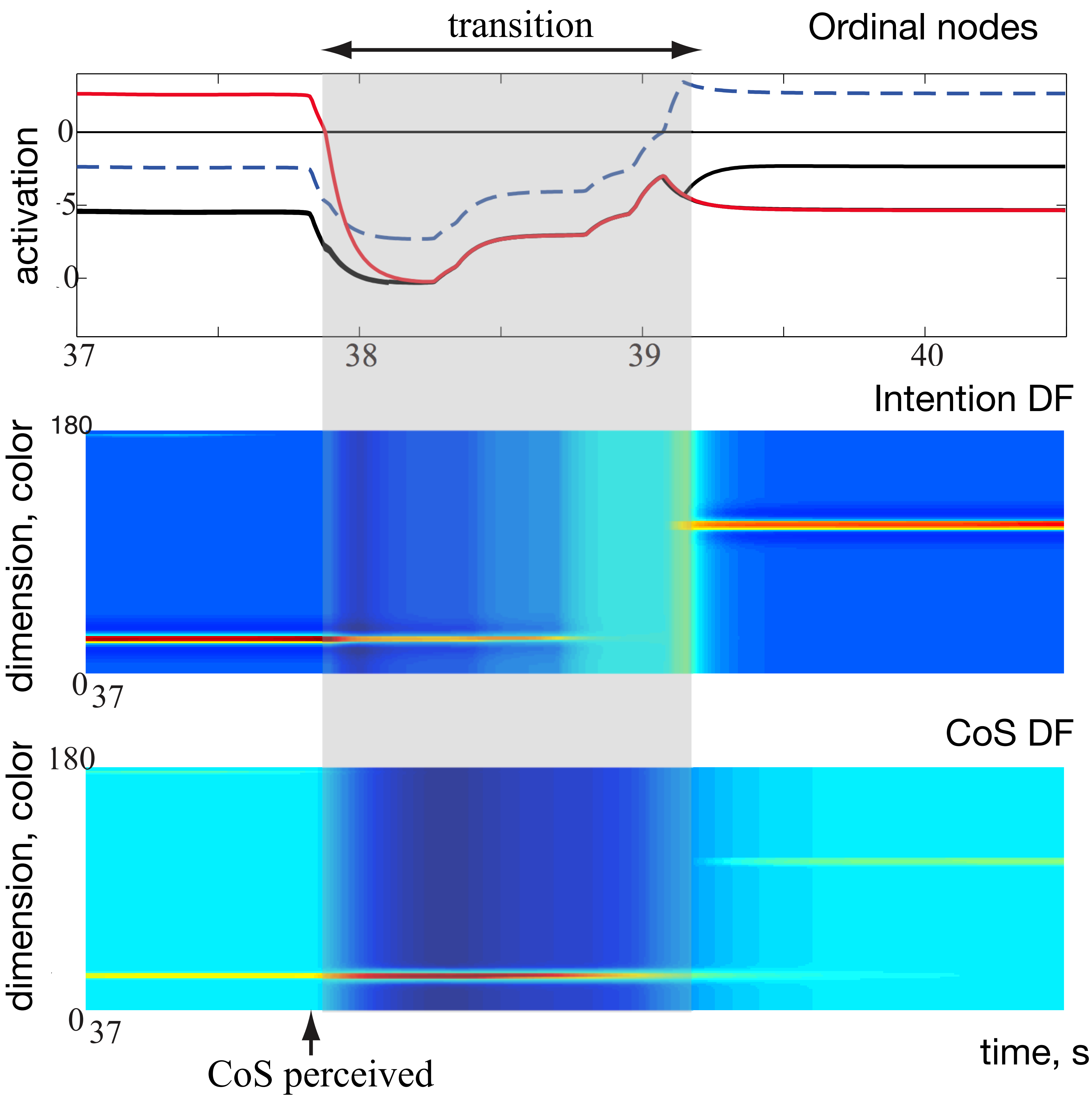
condition of satisfaction (CoS)

intentional state

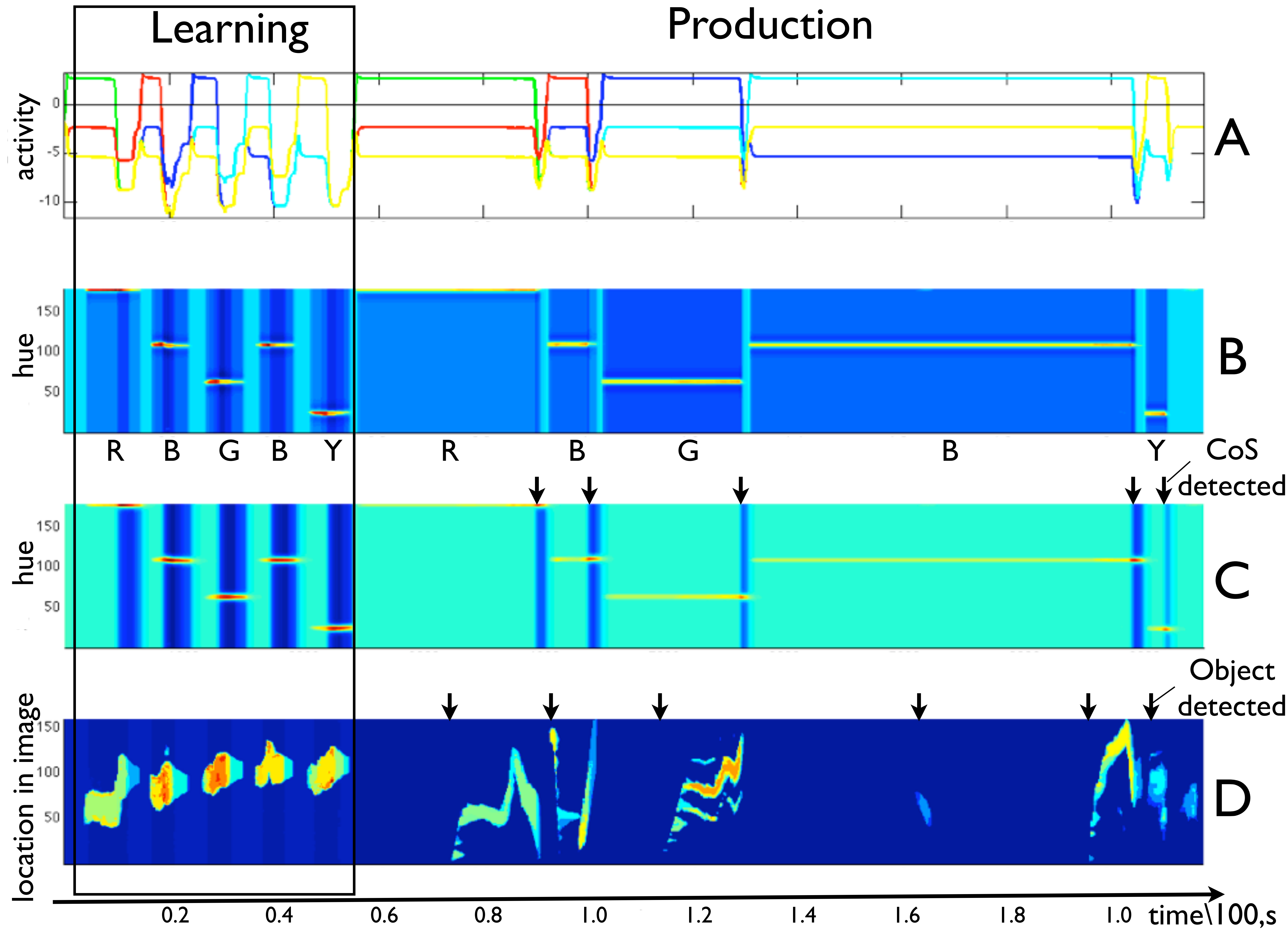


2D color-space field



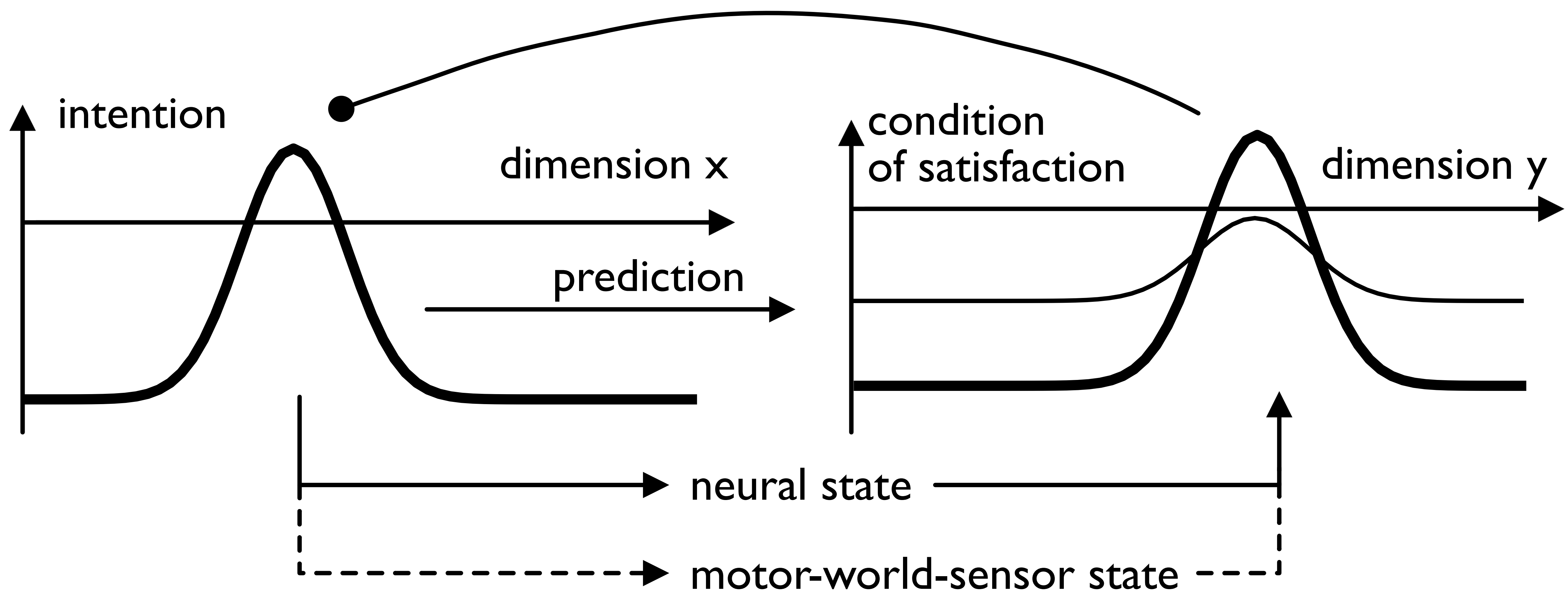


transition
Ordinal nodes
activation
0
-5
-10
37
38
39
40
Intention DF
dimension, color
180
0
37
CoS DF
dimension, color
180
0
37
CoS perceived
time, s



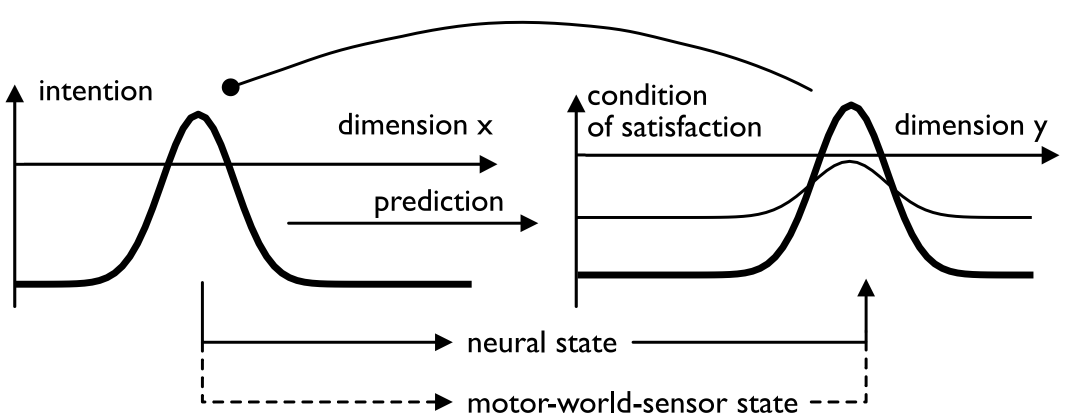
Learning
Production
activity
0
-5
-10
A
hue
150
100
50
B
R B G B Y
R B G B Y
CoS detected
hue
150
100
50
C
Object detected
location in image
150
100
50
D
0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 1.0
time\100,s

Mathematical mechanism



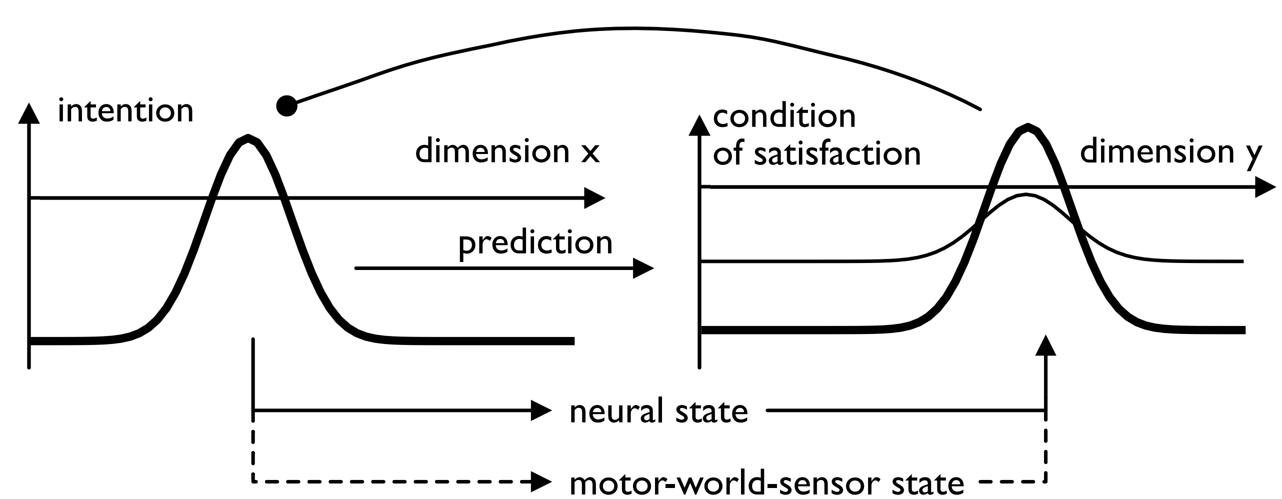
Sequence of instabilities

- 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



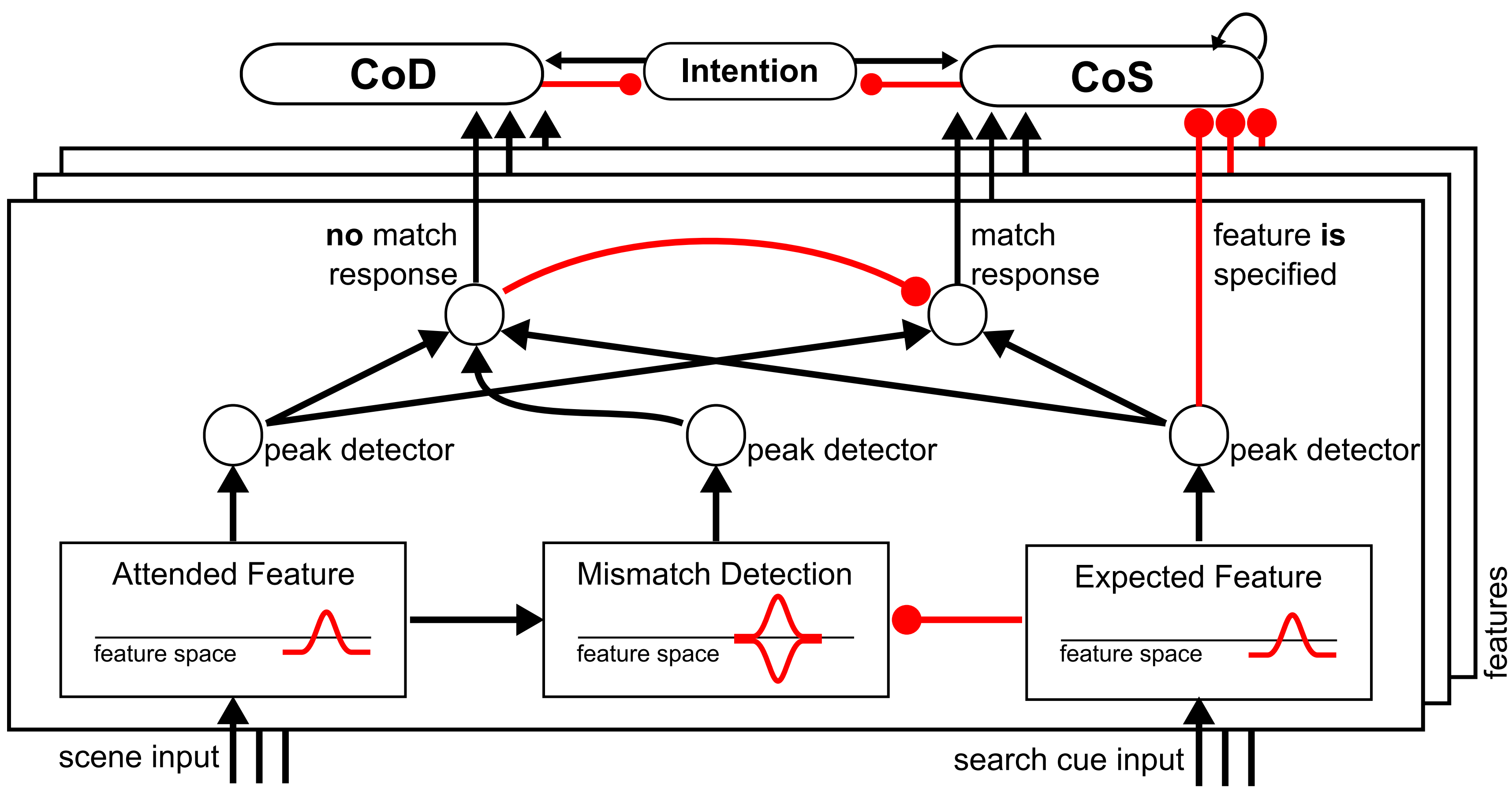
CoS and efference copy

- one could think of the "prediction" implied in the CoS as being a form of efference copy
- that does act inhibitorily…
- but it does so on the (motor)intention, not on the perception of the outcome that is predicted!



Generalization

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



[Grieben, Schöner, *CogSci* 2021]

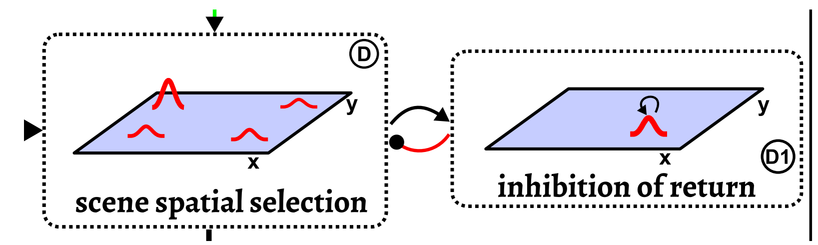
Roadmap

How is the next state selected?

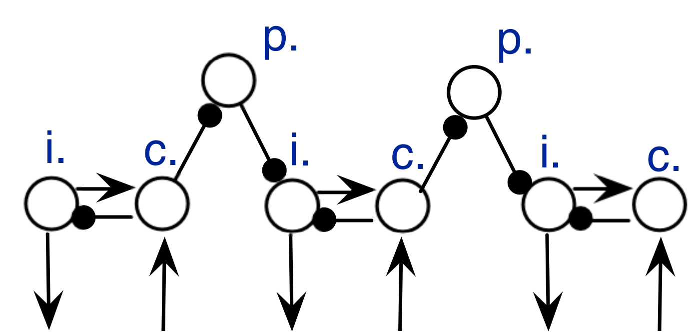
- once the current state has been de-activated…
- three notions
  - gradient-based selection
  - chaining
  - positional representation
- an illustration

How is the next state selected?

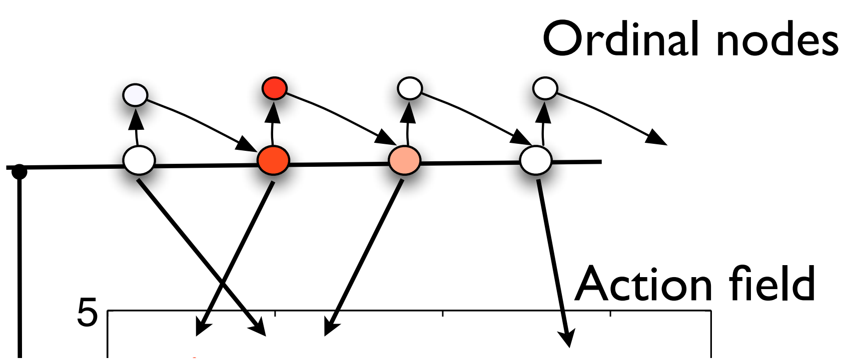
- once the current state has been deactivated…
- 3 notions (~Henson Burgess 1997)



- 1 gradient-based selection

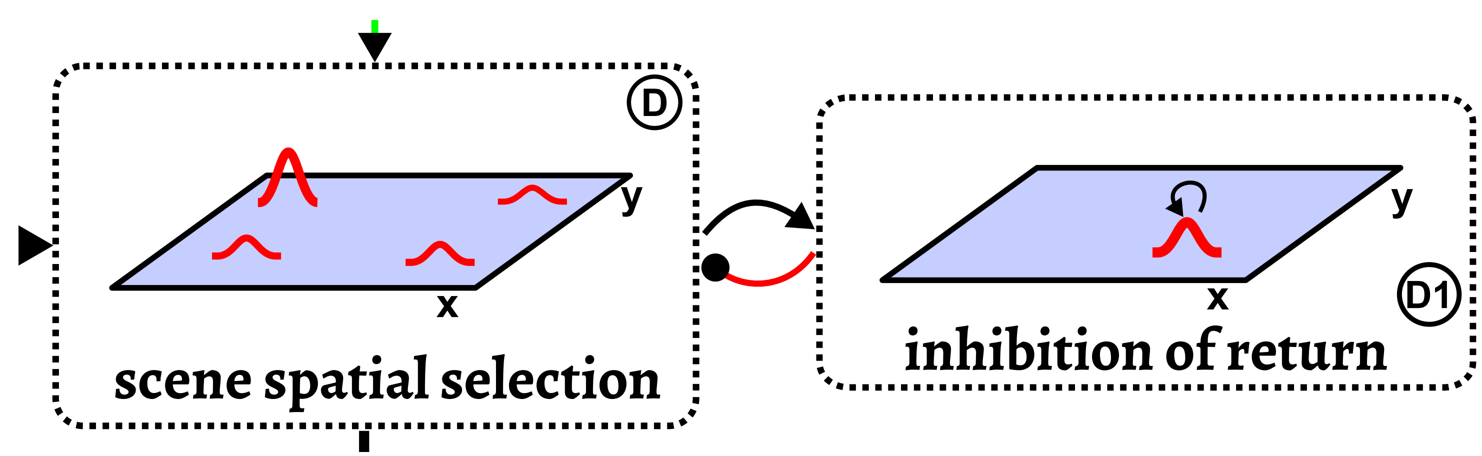


- 2 chaining
- 3 positional representation



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

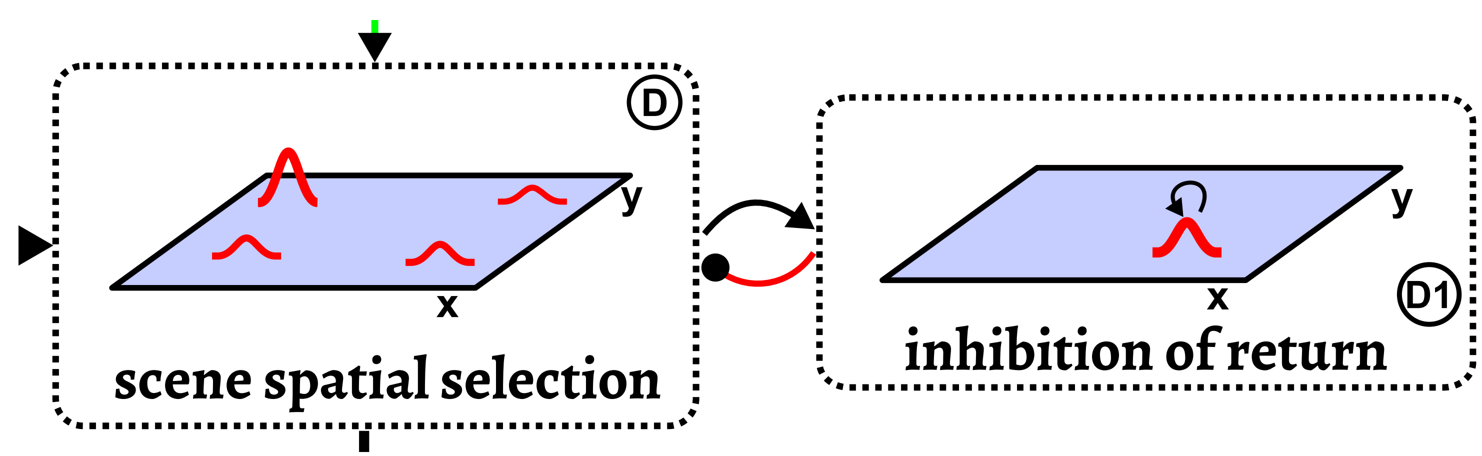


[Grieben, Schöner, *CogSci* 2021]

Gradient-based

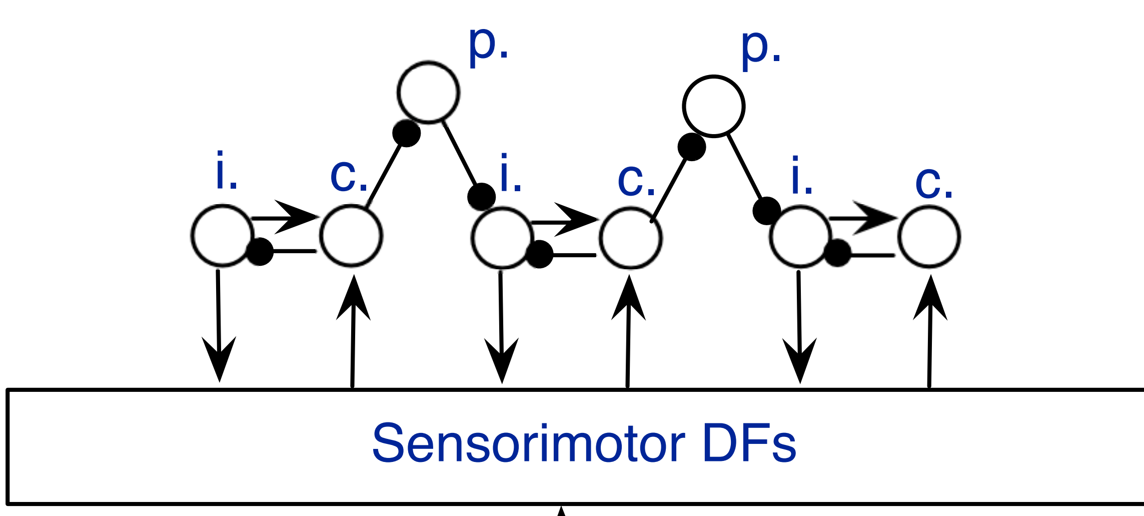
- this is used in many of the DFT architectures
  - visual search
  - relational grounding
  - mental mapping

[Grieben, Schöner, *CogSci* 2021]



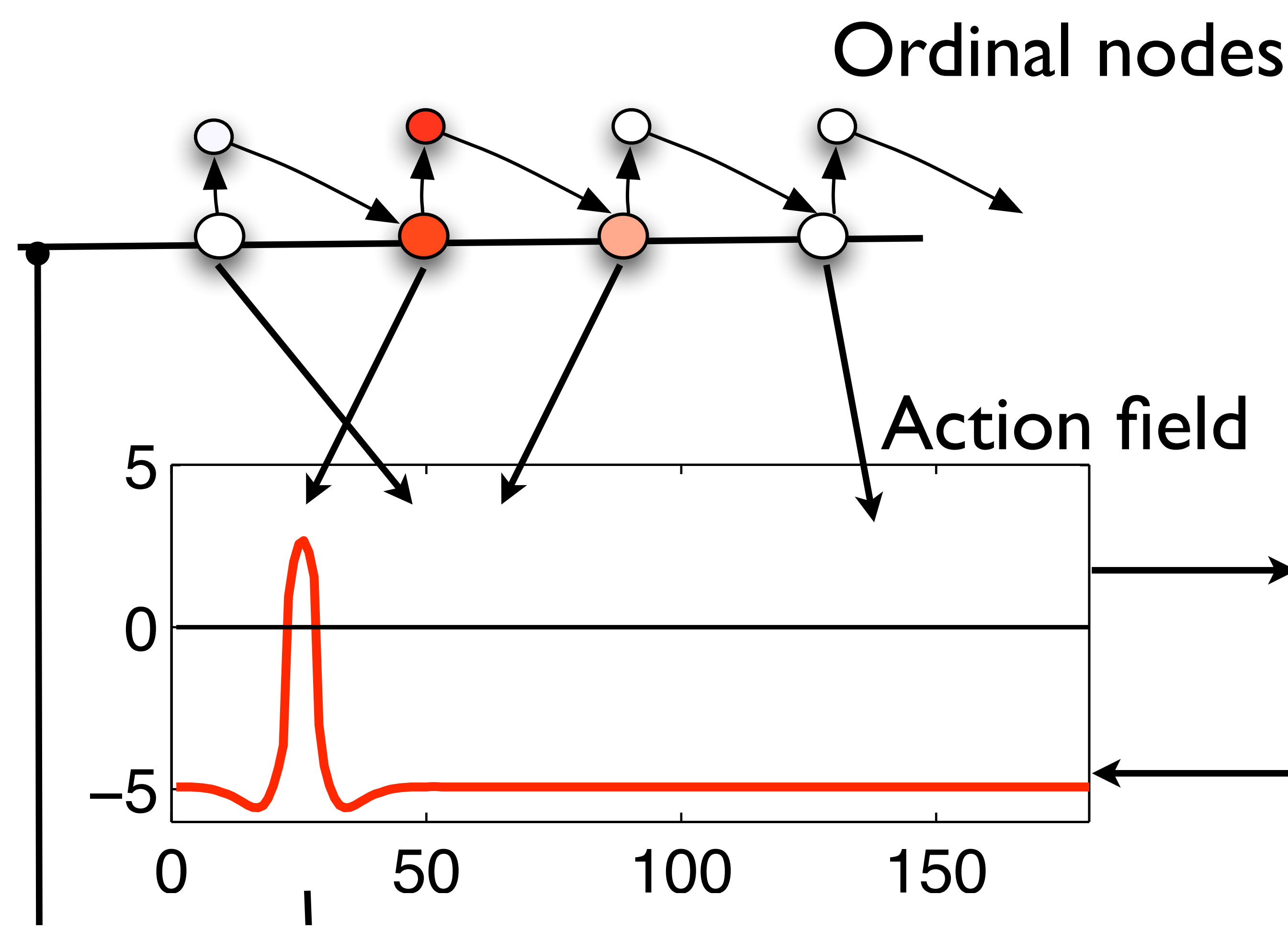
Chaining

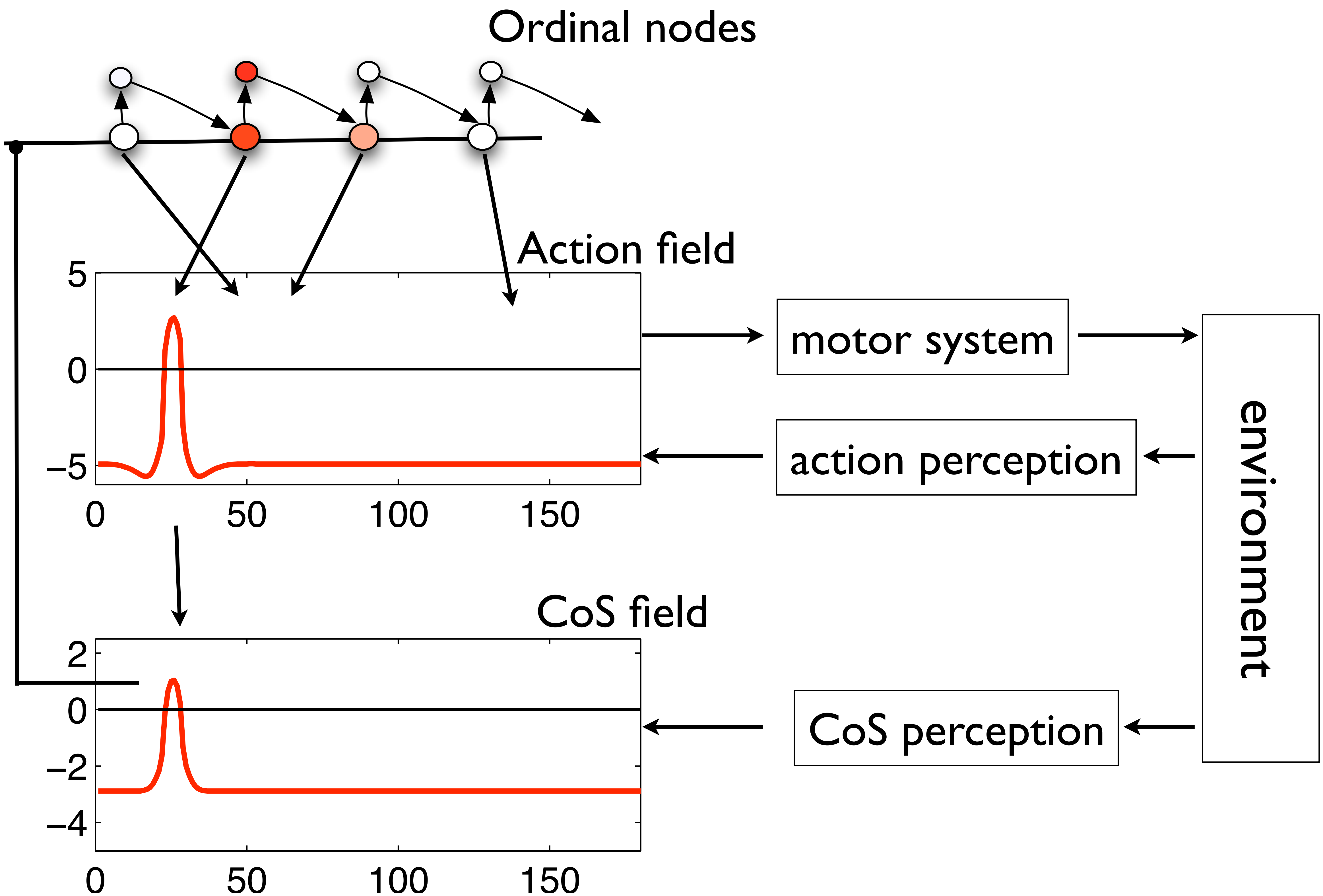
- for fixed sequences…
  - e.g. reach-grasp
  - fixed order of mental operations… e.g. ground reference object first, then target object
- less flexible (e.g.. when going through the same state with different futures)
- could be thought to emerge with practice/habit from the positional system



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…





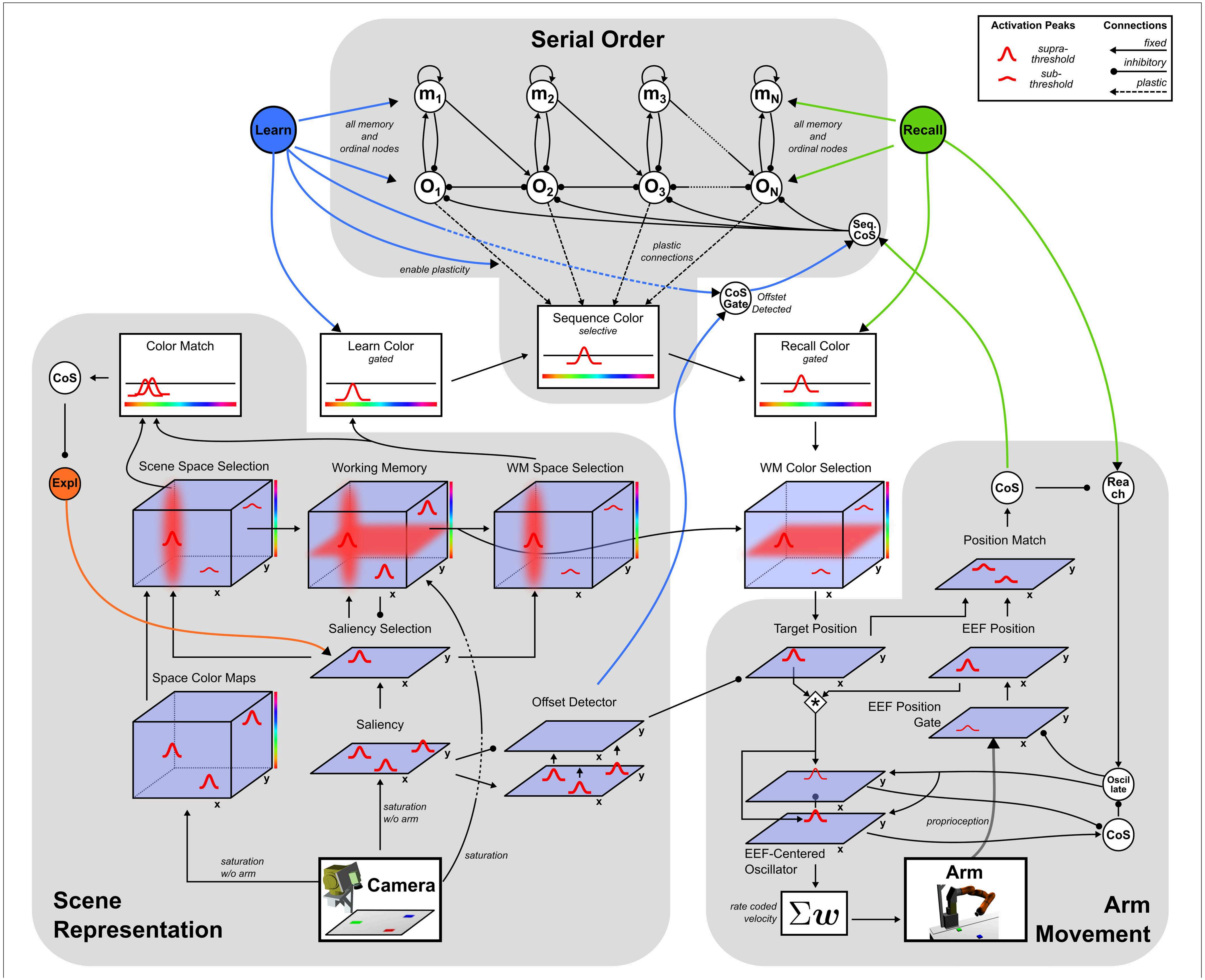
[Sandamirskaya, Schöner: Neural Networks 23:1163 (2010)]

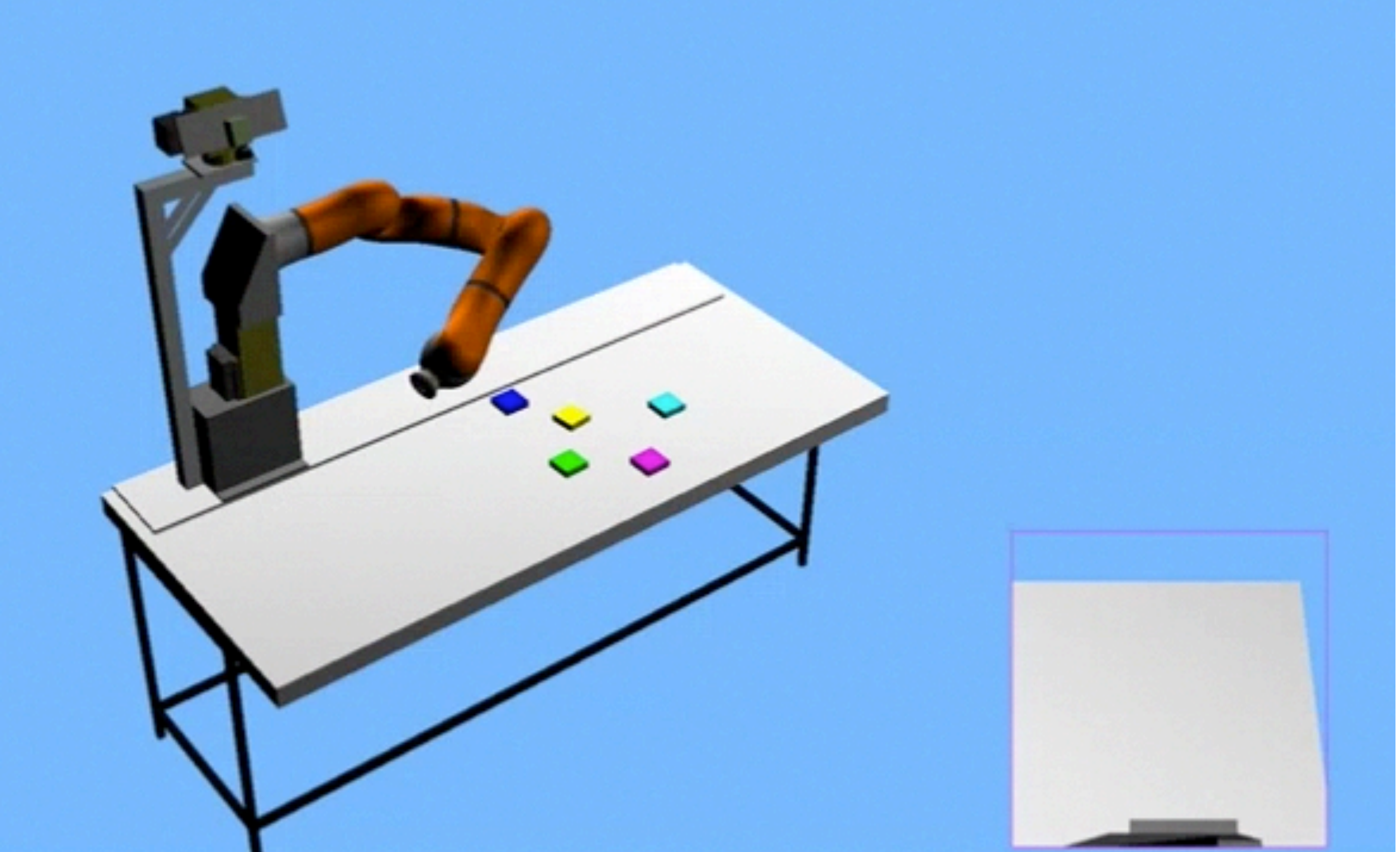
Positional representation

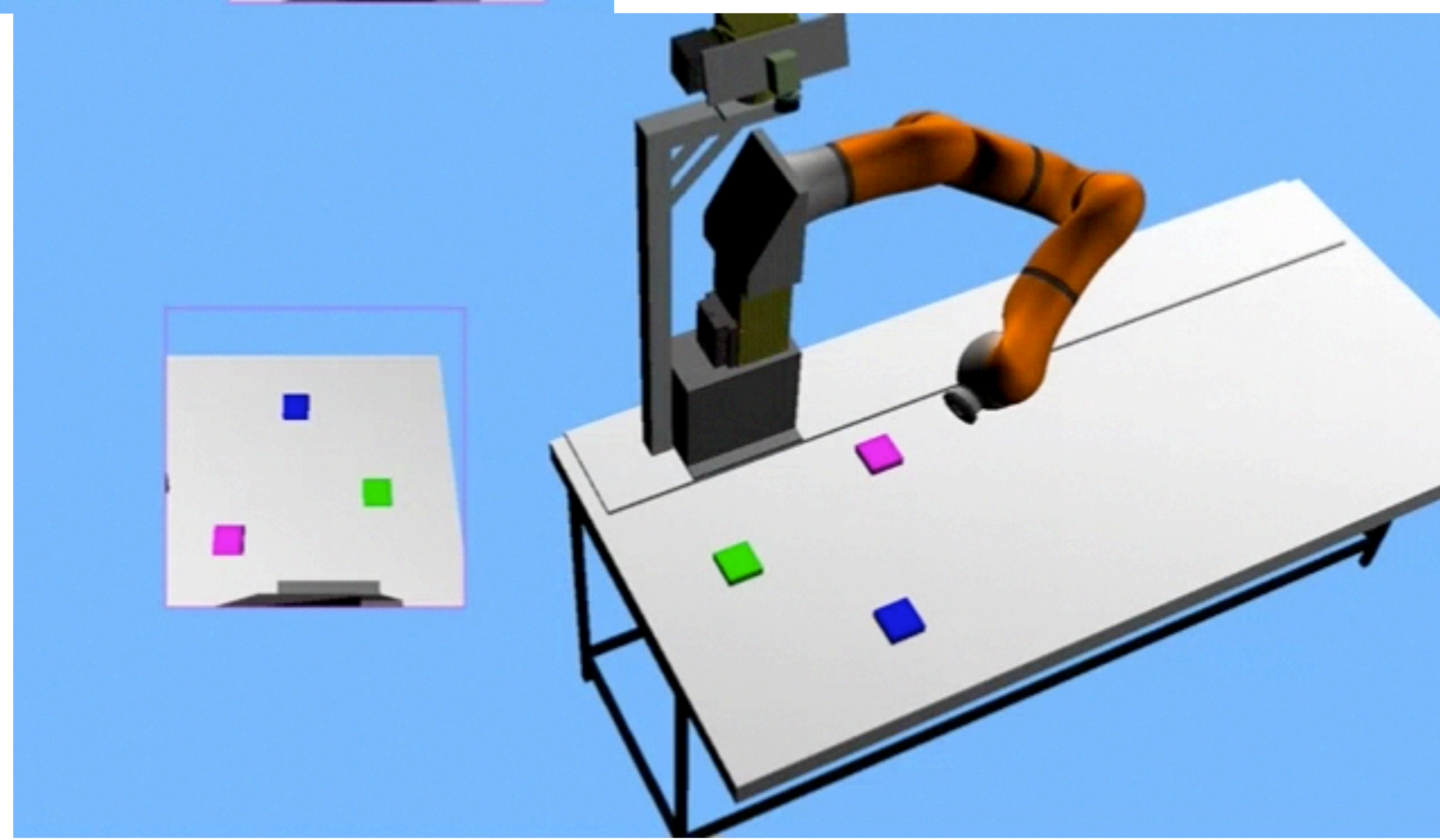
- essentially chaining with flexible contents
- good for fast learning of sequences…
  - e.g. imitation
  - a Hippocampus function?
- but: must have potential synaptic links to many representations…
- => such ordinal systems must exist for sub-representations… embodiment effects…

Serial order demonstrated/enacted

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







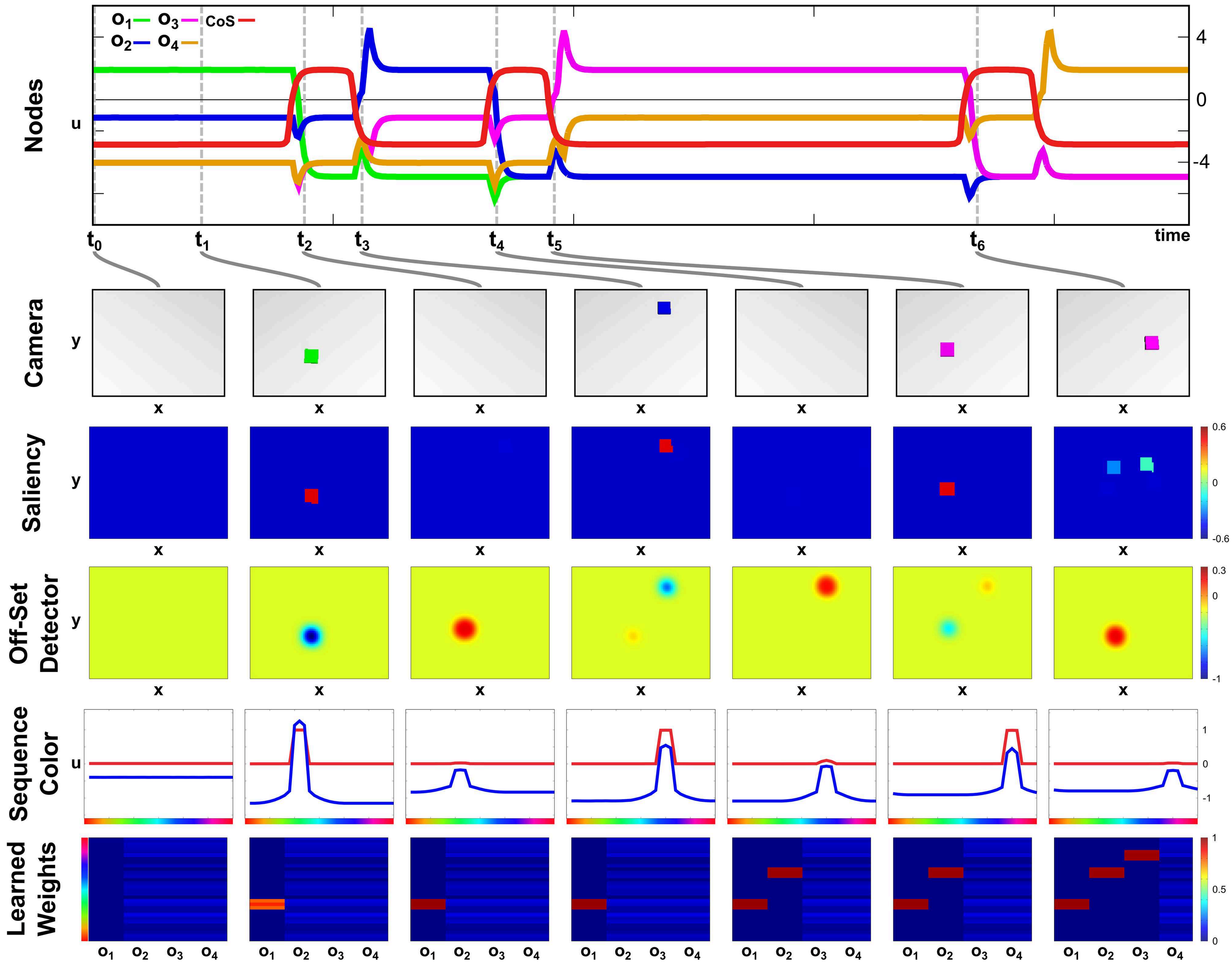


FIGURE 5 | Time course of learning a three element sequence with varying presentation time.

Time course of attention selection and building of scene memory

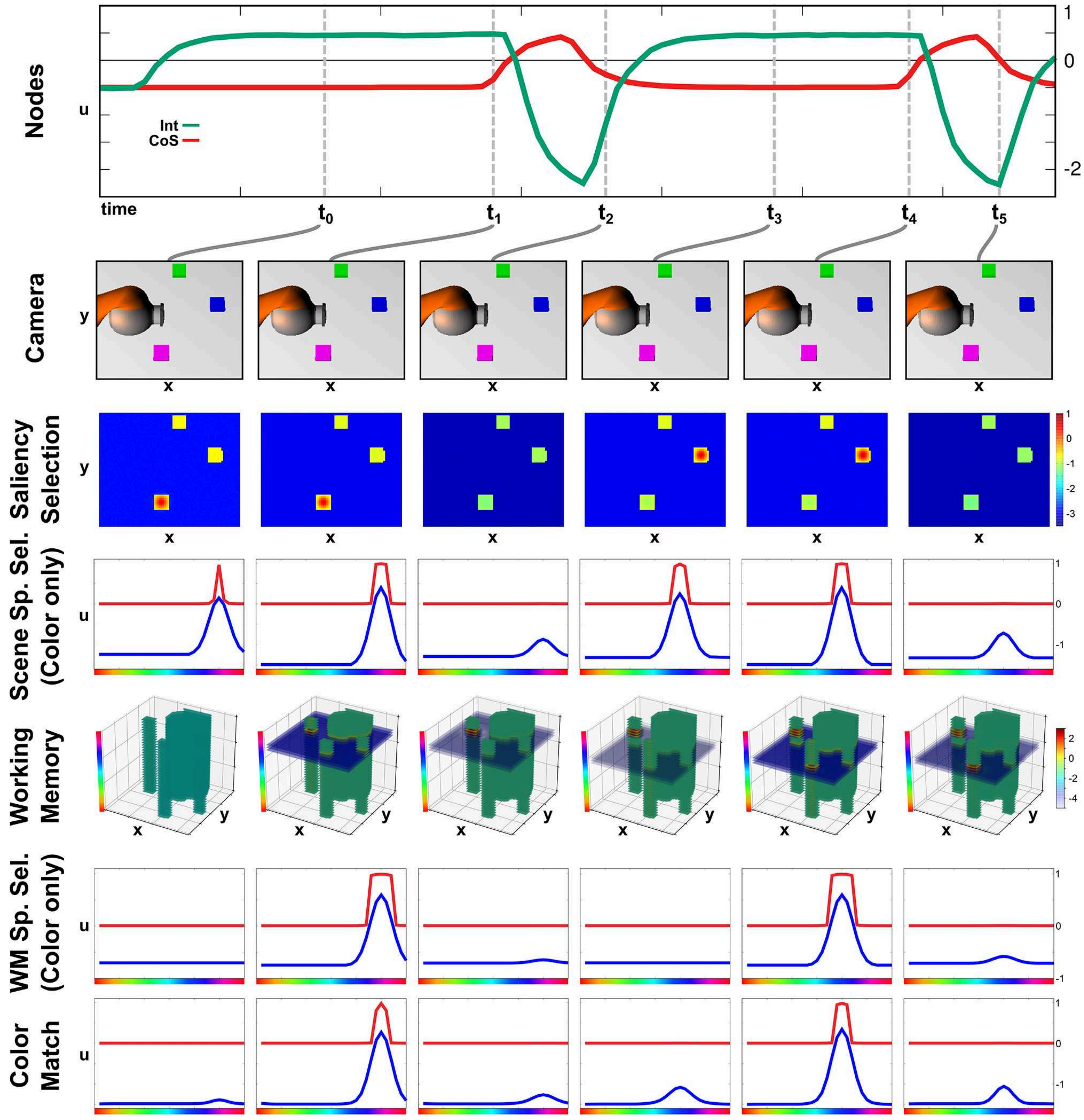


FIGURE 4 | Time course of building a scene memory.

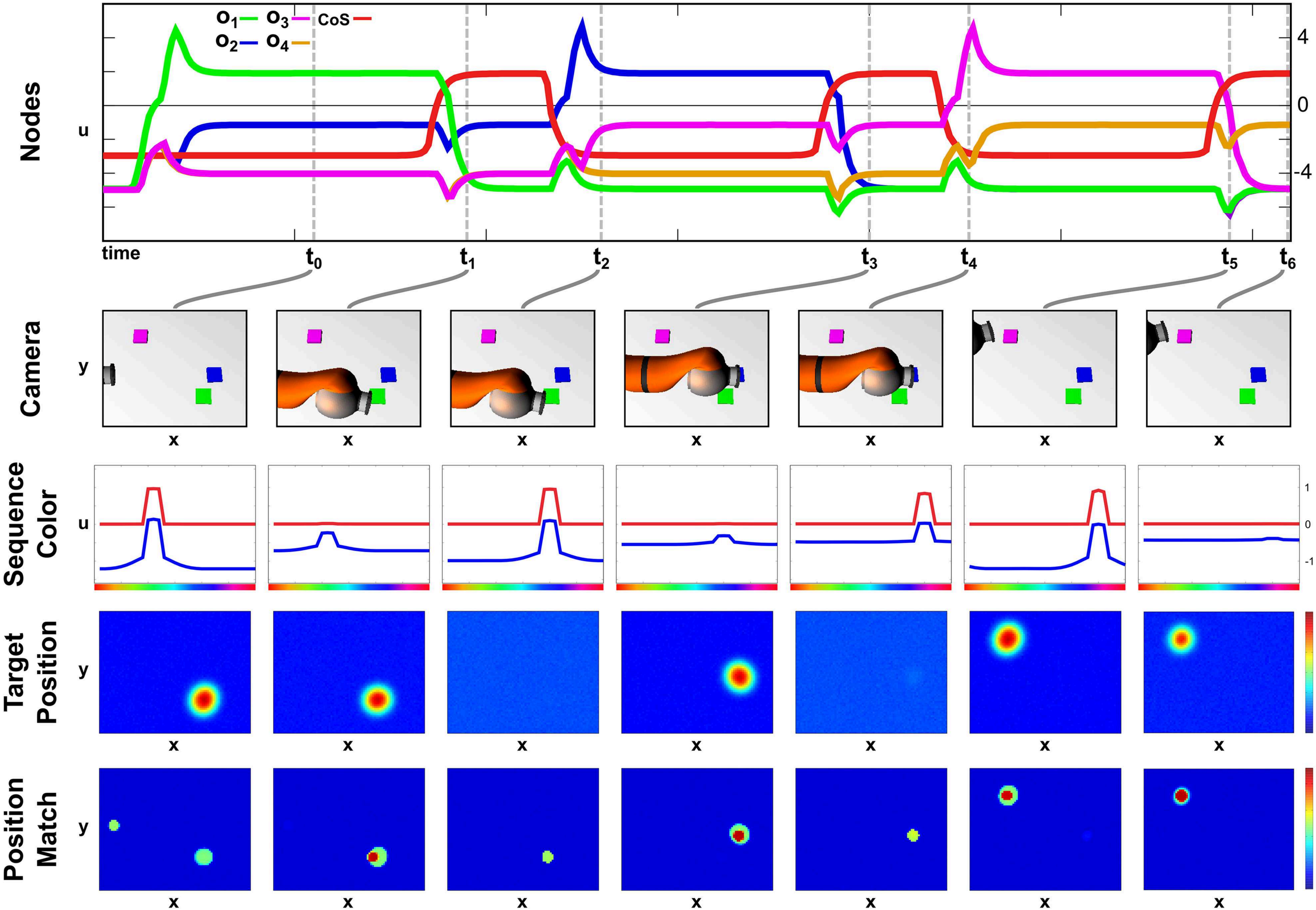
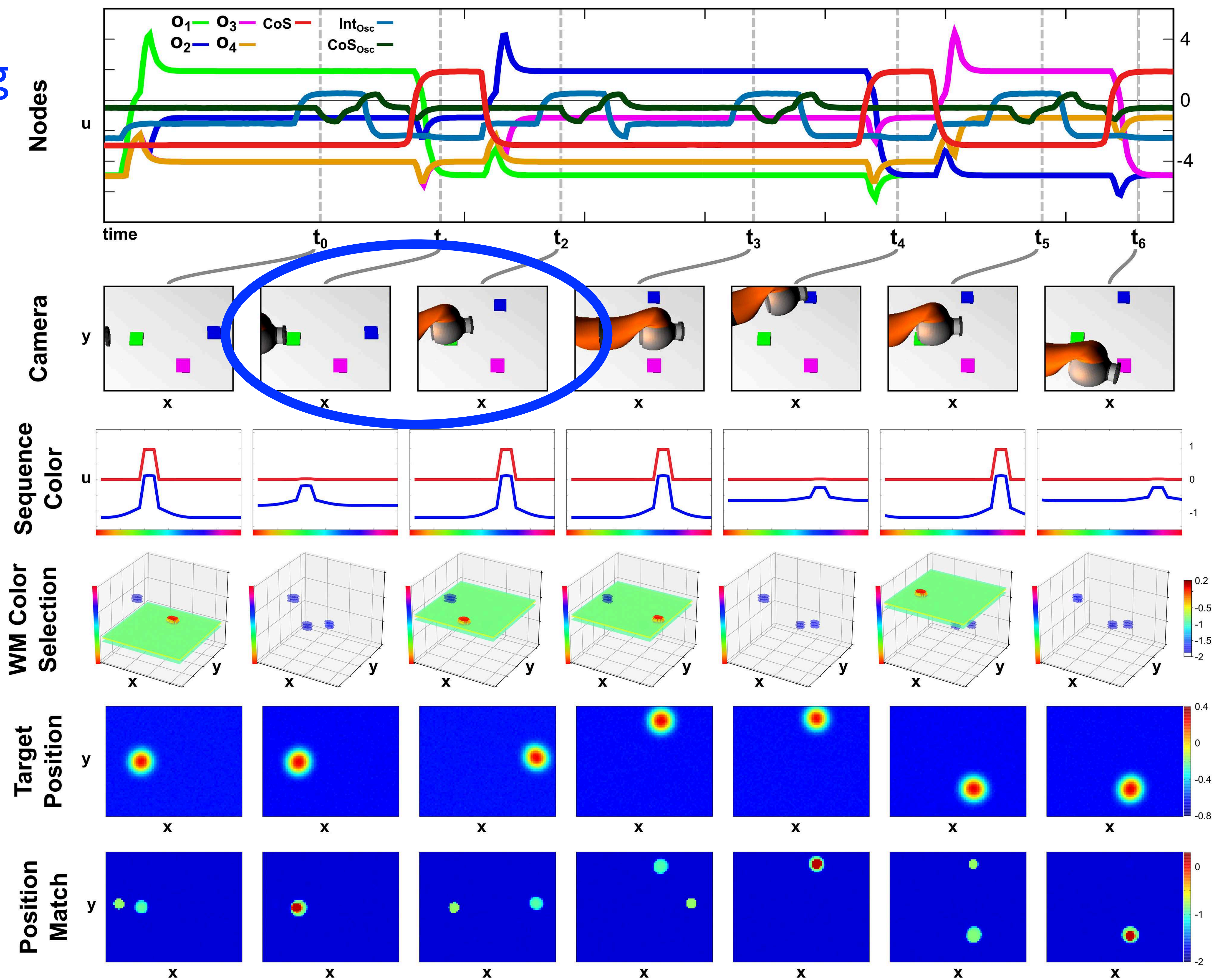


FIGURE 6 | Time course of recalling a three element sequence through pointing at colored objects.

online
updating

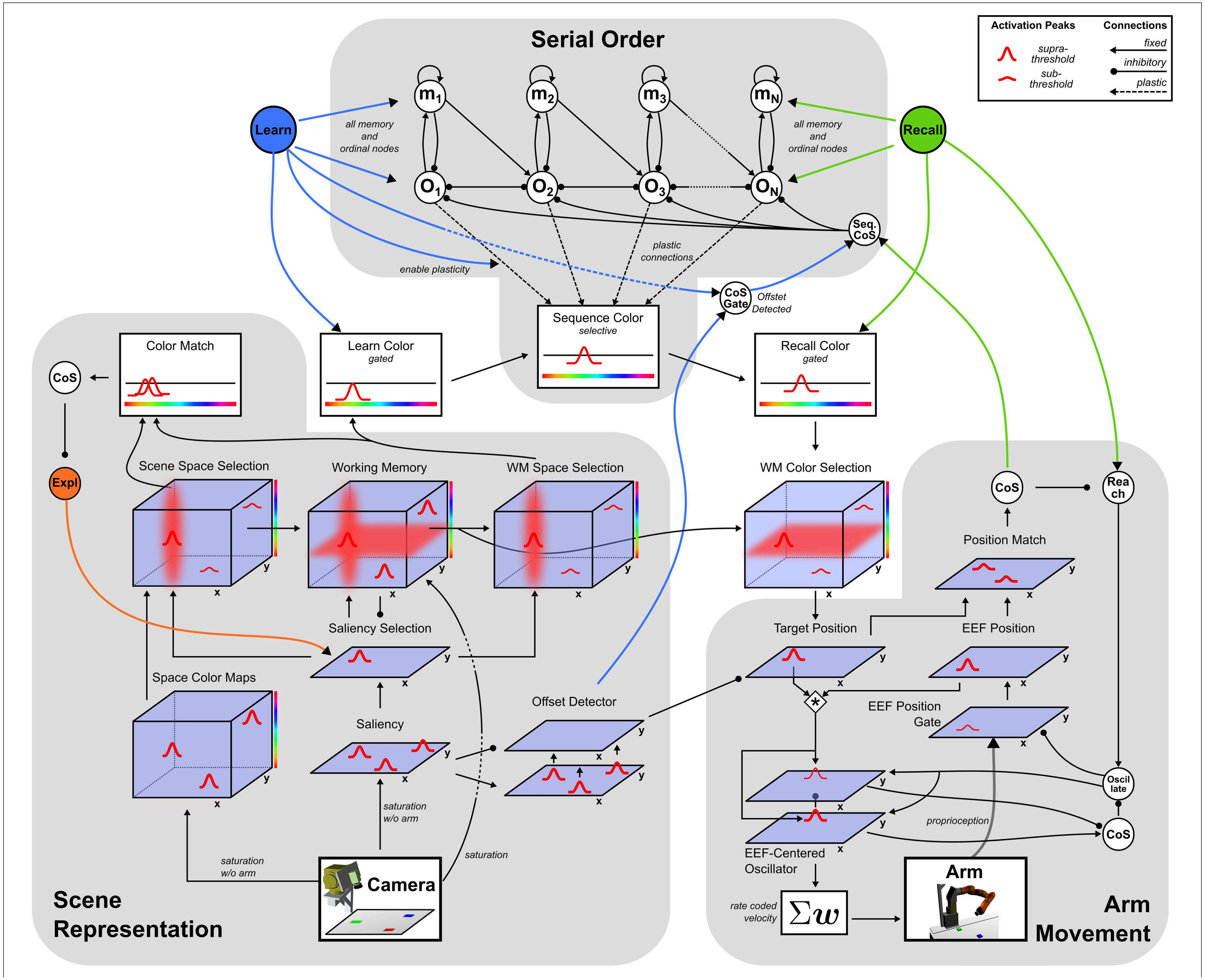


Roadmap

Why do neural dynamic architectures work?

- dynamic structural stability
- the “non-synesthesia” principle

Neural dynamic architecture



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

Neural dynamic architectures

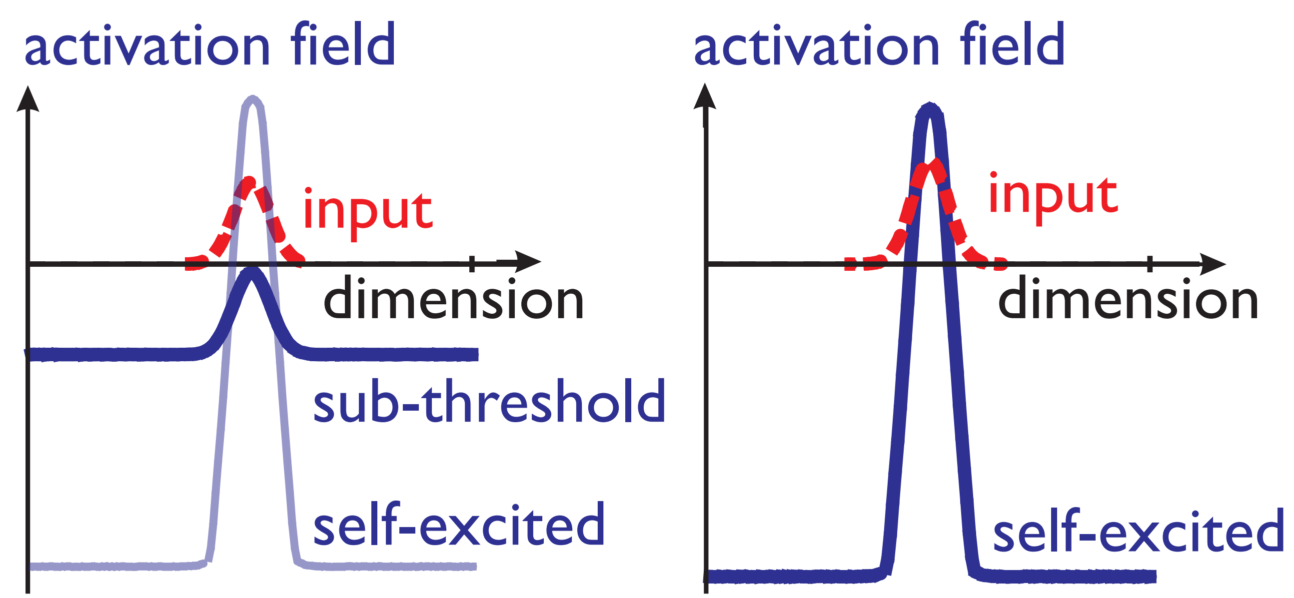
- when we label each field/set of fields with a "function", we presuppose that activation in that subpopulation has a fixed functional significance…
- [which may misleadingly give the impression that DFT architectures are information processing architectures]
- why is it possible to do that even though the DFT architecture really is just one big dynamical system?

Two invariances

- Two questions are contained here
- 1) why is the dynamic regime ("selection", "working memory", "detection", "match" etc.) of a component field invariant as we couple it into a larger architecture?
- 2) why is the content (the feature space over which fields are defined, the content of a concept node) of a component field invariant as we couple it into a larger architecture?

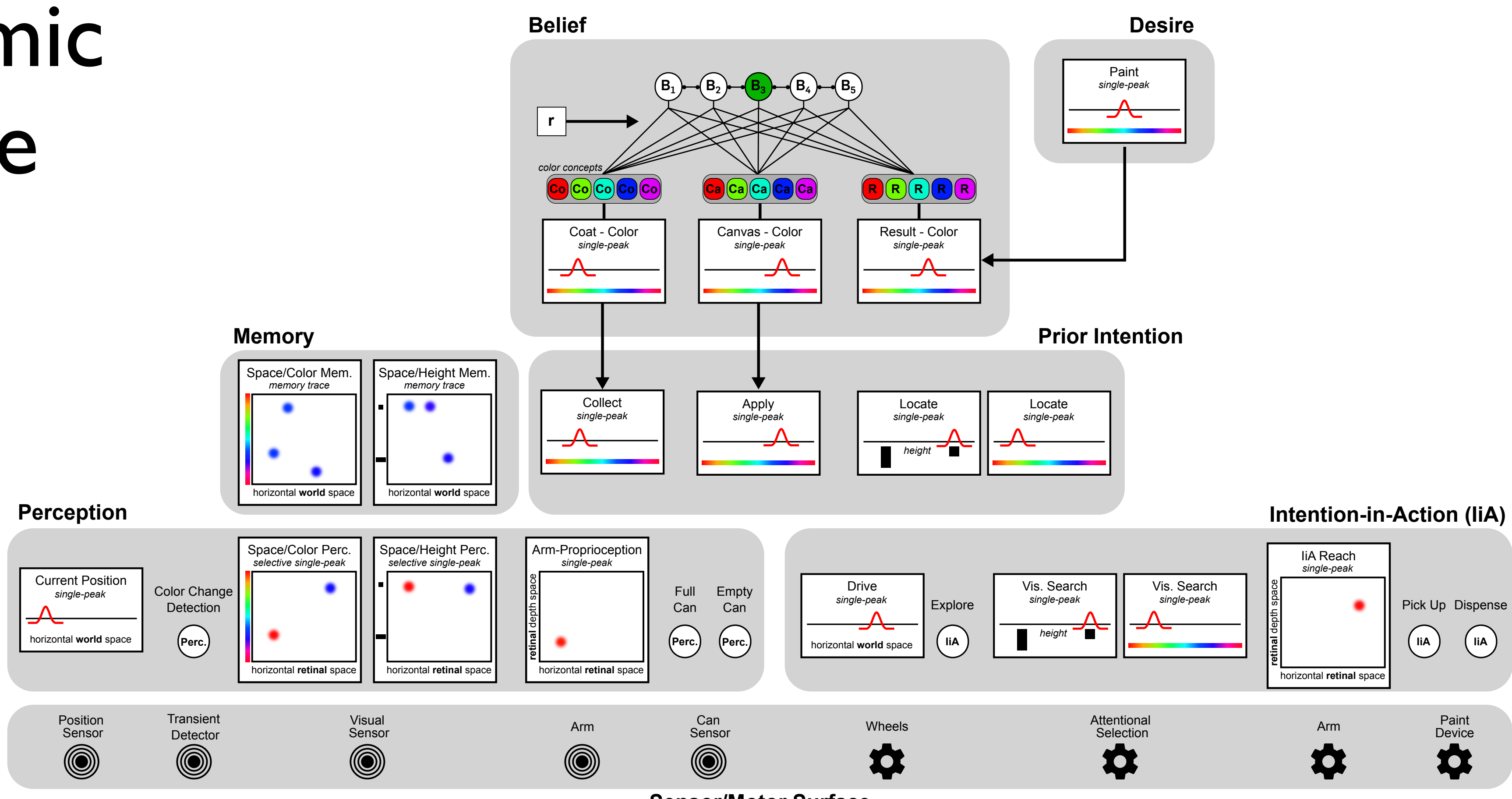
DFT architectures

- 1) 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

- 2) 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



Roadmap

Positioning DFT in the theoretical landscape

- in which sense is cognition emerging in DFT architectures embodied?
- DFT vs connectionism/DNN
- DFT vs. cognitive architectures/symbol manipulation
- DFT vs. VSA/SPAUN

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!

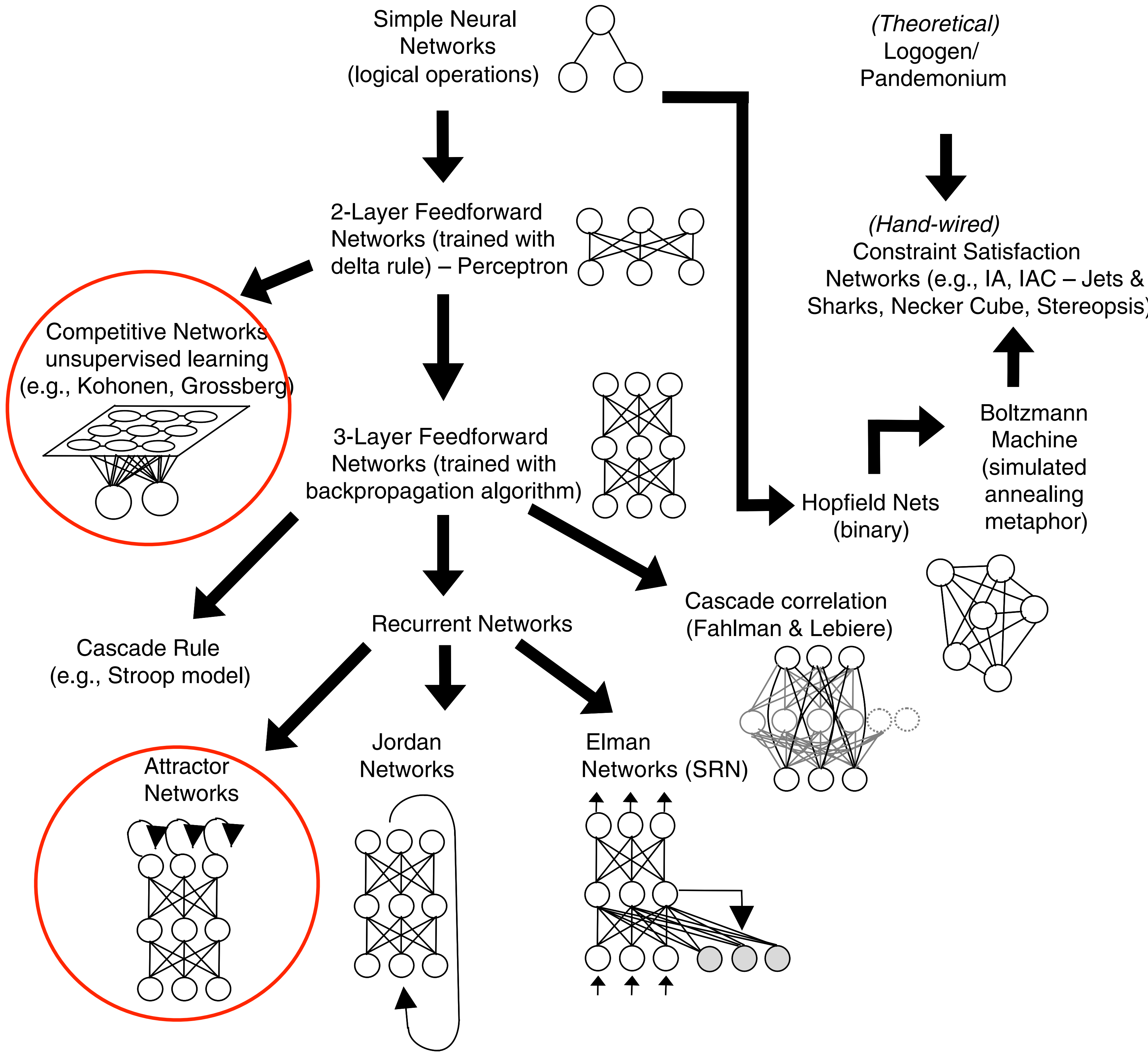
Embodiment hypothesis

- 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

DFT vs connectionism/NN

- DFT models are neural network models in the most general sense…
- sharing level of description (activation, sigmoid)



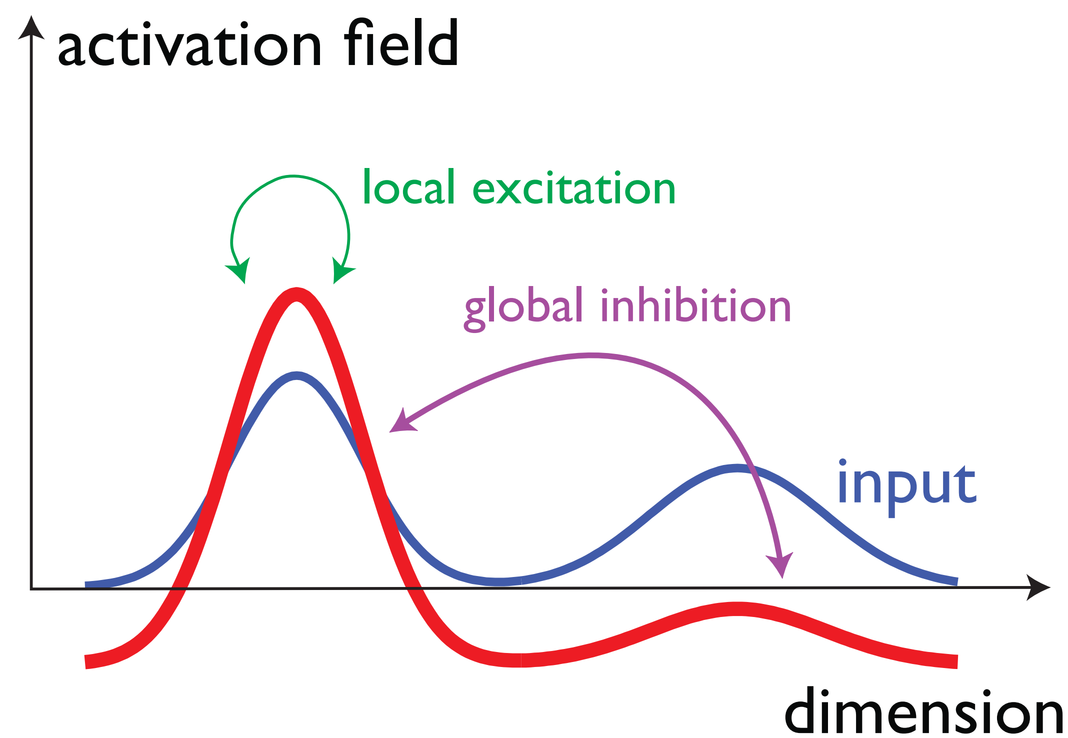
[Thomas, McClelland, 2008]

DFT makes more specific commitments

- stability of functionally significant states
- populations as the level of description at which regularities of behavior/thinking can be understood
- instabilities as key elements of neural processing

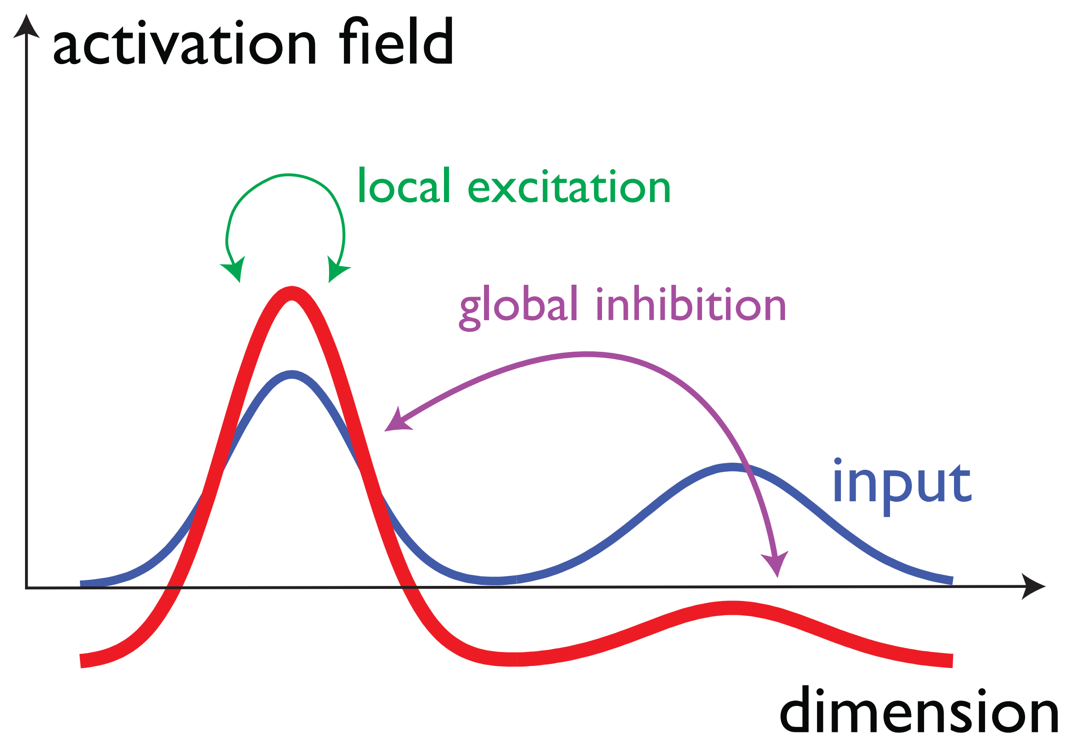
DFT’s commitments differ from connectionist commitments

- DFT: all autonomous cognition is based on localist representations
- => which are necessarily low-dimensional
- to enable the homogeneous form of neural interaction
- to enable stable representations of new patterns
- to enables instabilities => sequences



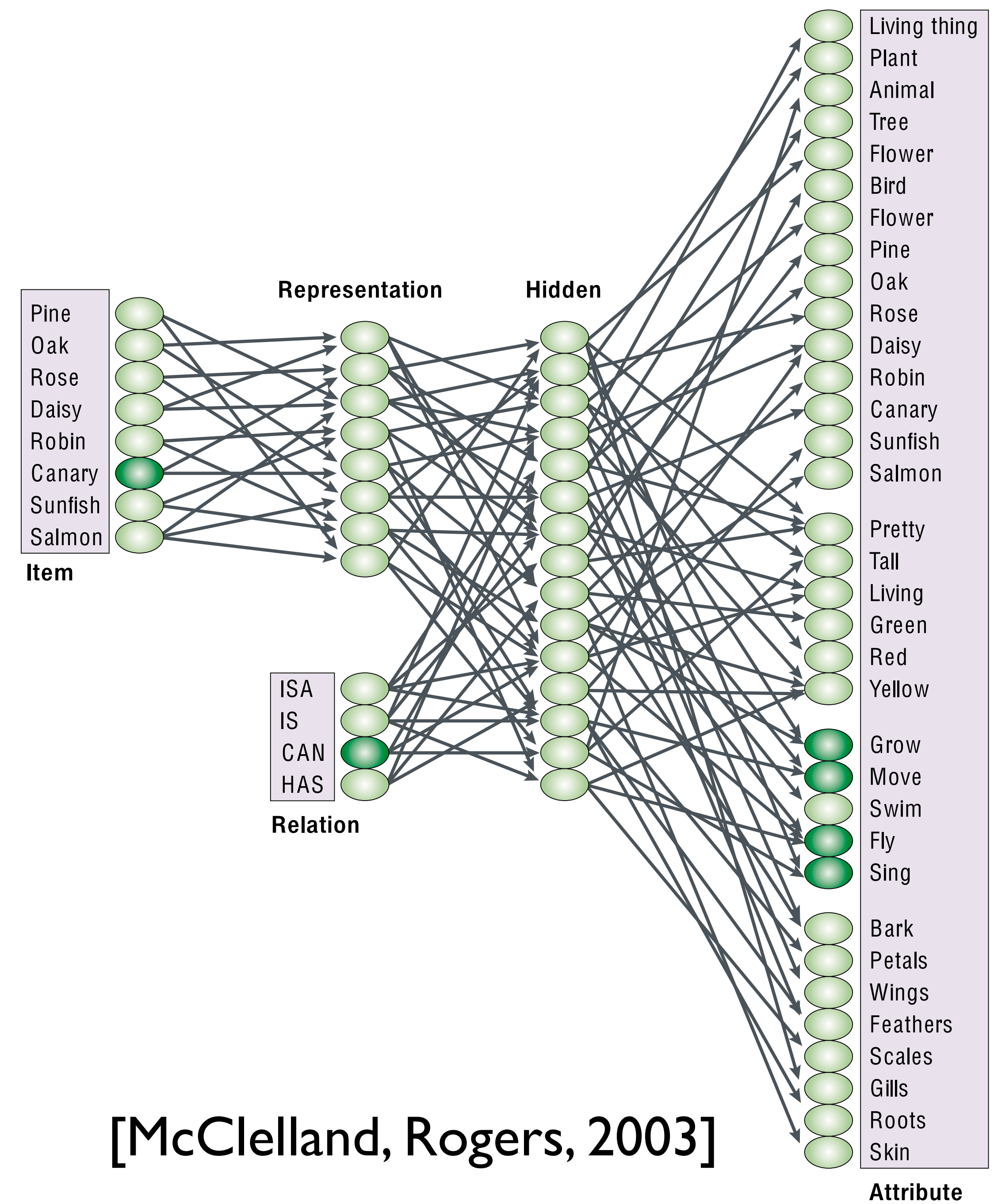
DFT's commitments differ from connectionist commitments

- => this leads to the special role of the memory trace.. a possible theory of memory



DFT's commitments differ from connectionist commitments

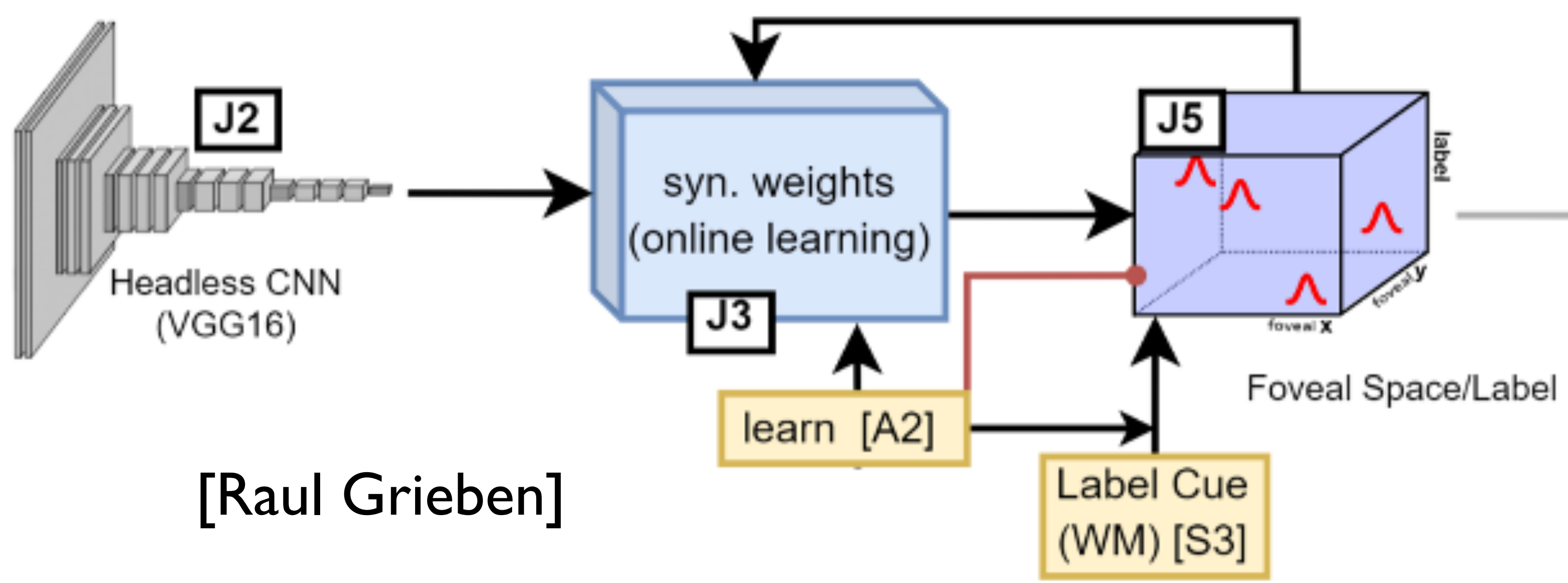
- eliminates role of distributed representations in association !
- e.g., in DFT Rumelhart/ McClelland's account for concepts as feature associations is acually a form of binding among localist representations



[McClelland, Rogers, 2003]

DFT's commitments differ from connectionist commitments

- high-dimensional neural representations that resemble distributed representations play a special role in discrimination/classification… that is effective only when these processes are driven by sensory inputs



[Raul Grieben]

DFT vs symbol manipulation

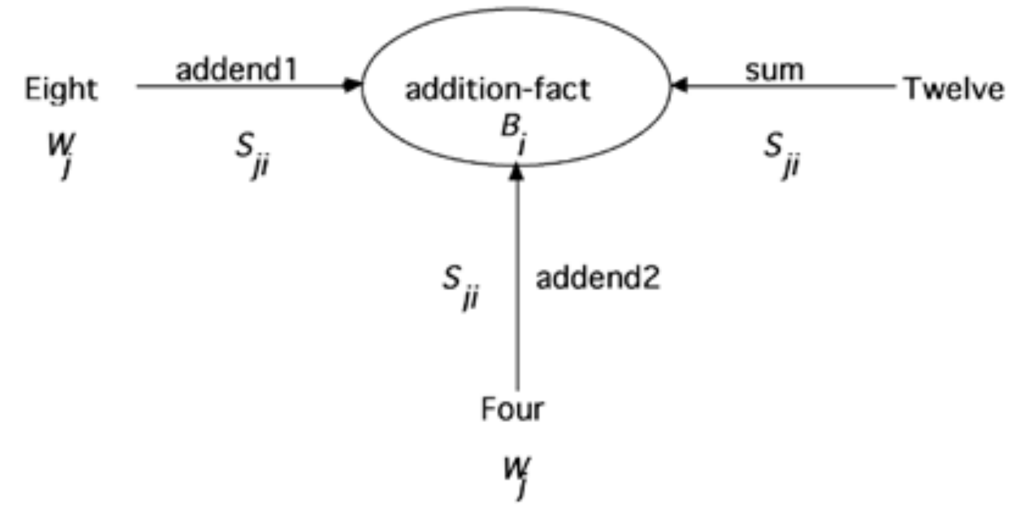
- the "information processing" perspective of cognition is based on "function calls" that hand on "arguments"… <=> symbol manipulation

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

- 1) this is at the core of classical cognitive architectures

DFT vs symbol manipulation

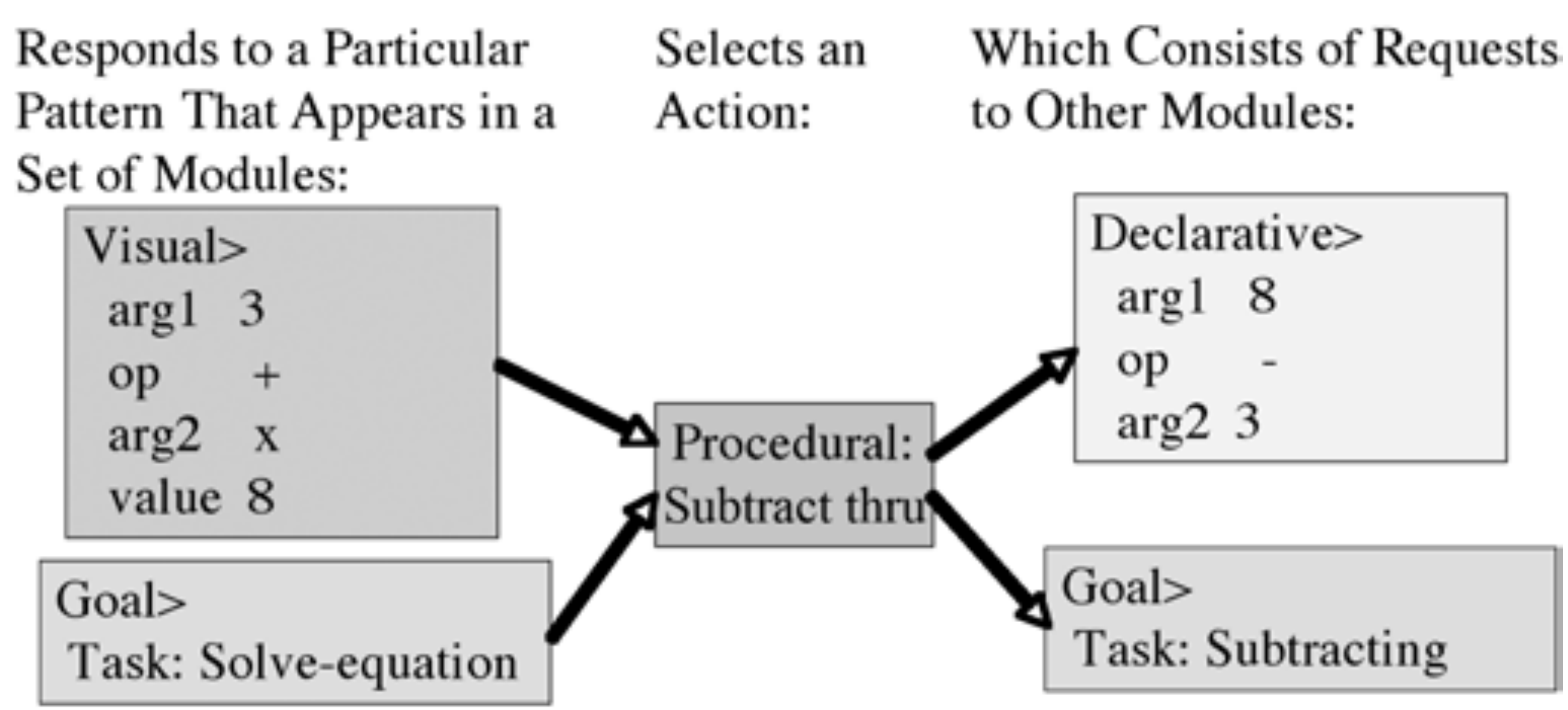
- example: ACT-R for mental arithmetic
  - contents: symbol
  - control: activation/weights



. A representation of a chunk with its subsymbolic quantities.

- 2) sequence generation is driven by an external computational cycle

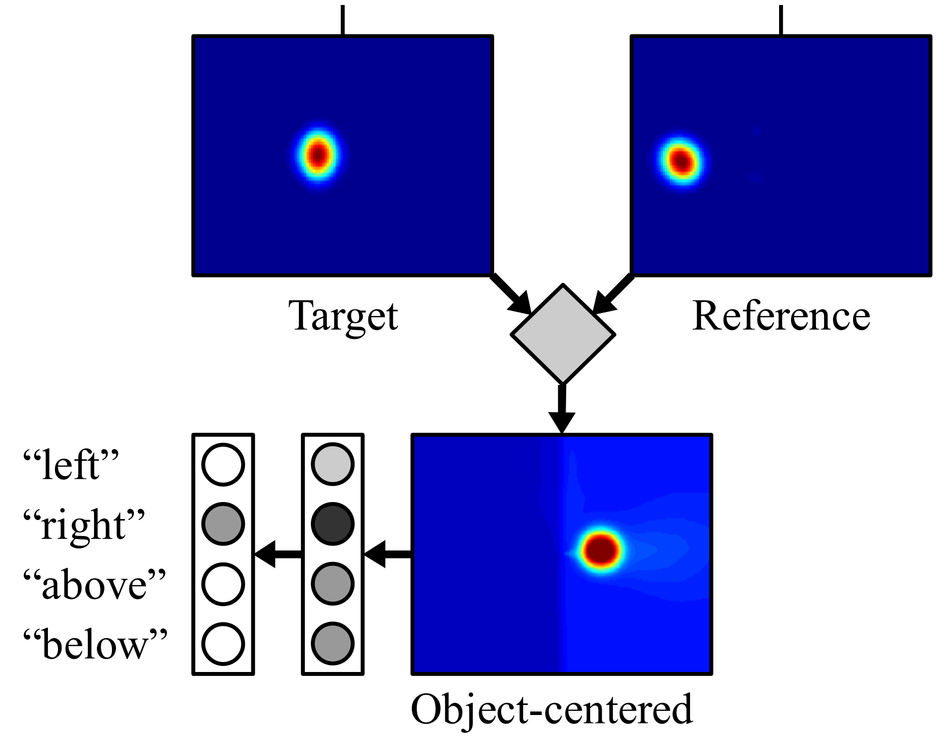
[Anderson, 2007]



DFT: a neural theory for higher cognition

- 1) attentional selection, coordinate transformation, sequential processing … emulate "function calls"

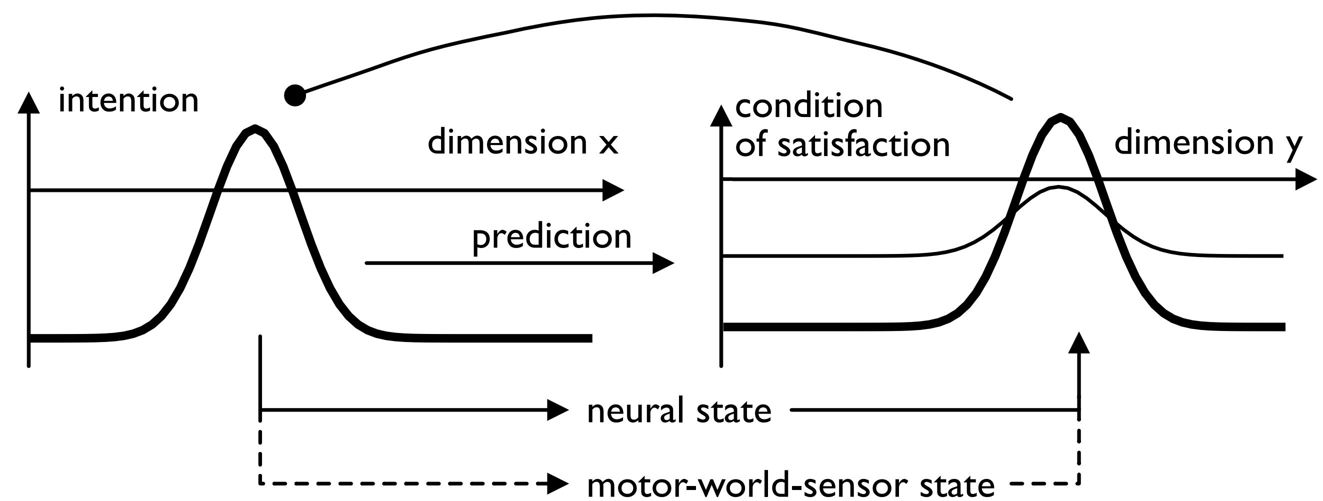
$$\text{to the left of} = f(\text{target, reference})$$



- much more constrained and costly in processing structure … explains signature of human cognition
- all concepts are grounded by their very nature…
- open to learning… and memory

DFT: a neural theory for higher cognition

- 2) the sequences of processing steps emerge from dynamic instabilities.



- robust under embodiment!

DFT vs VSA

- Vector-symbolic architectures (VSA) are a an alternative theoretical proposal for a neural account for higher cognition
- 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

DFT vs VSA

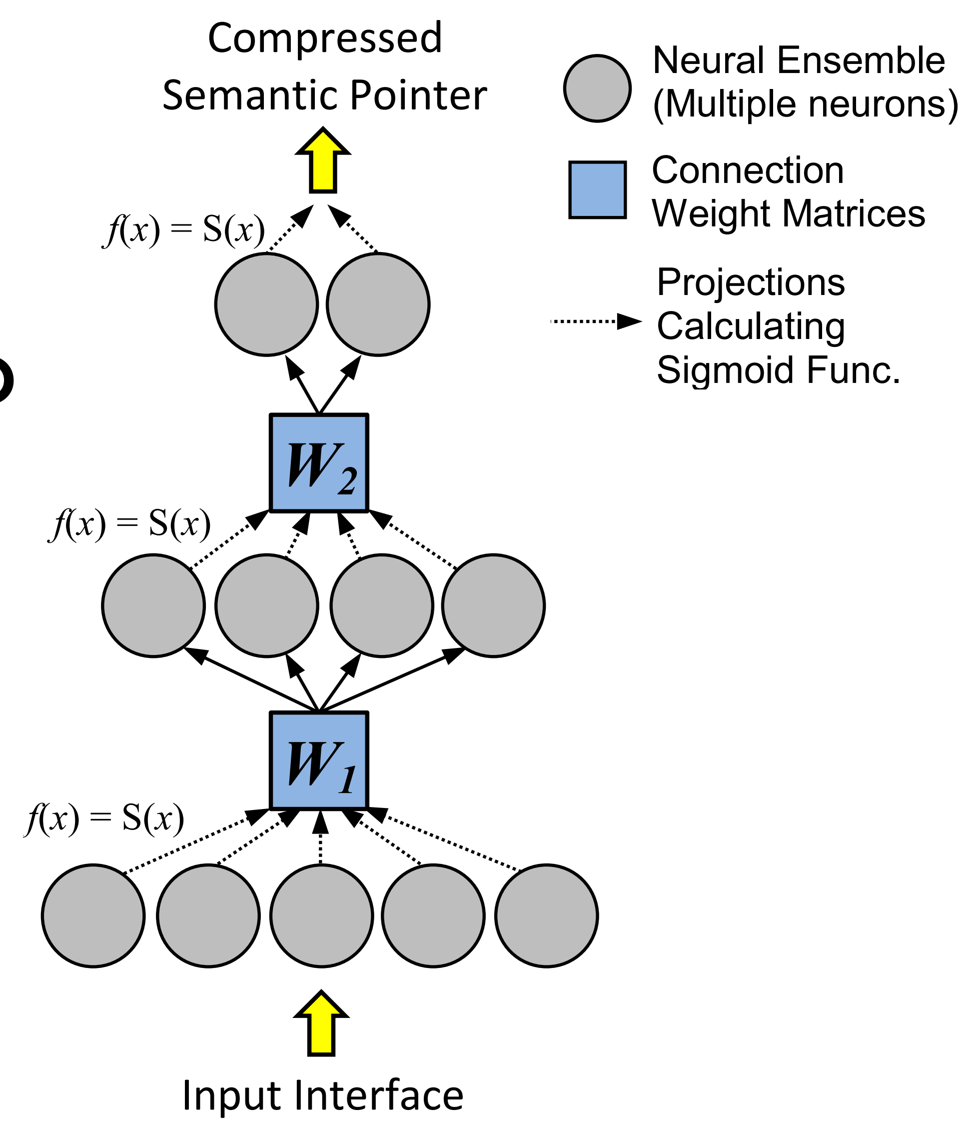
- requires that the symbol grounding problem is solved at encoding/decoding

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

DFT vs VSA

- But: to preserve the original vectors, connectivity in VSA/NEF (SPAUN) architectures is very special: decode and re-encode..
- => SPAUN brains are not robust against learning/development due to non-local inter-dependence of connectivities
- (and other issues)



[Choo Feng Xuan, 2018]

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…