

Grounding Spatial Language:

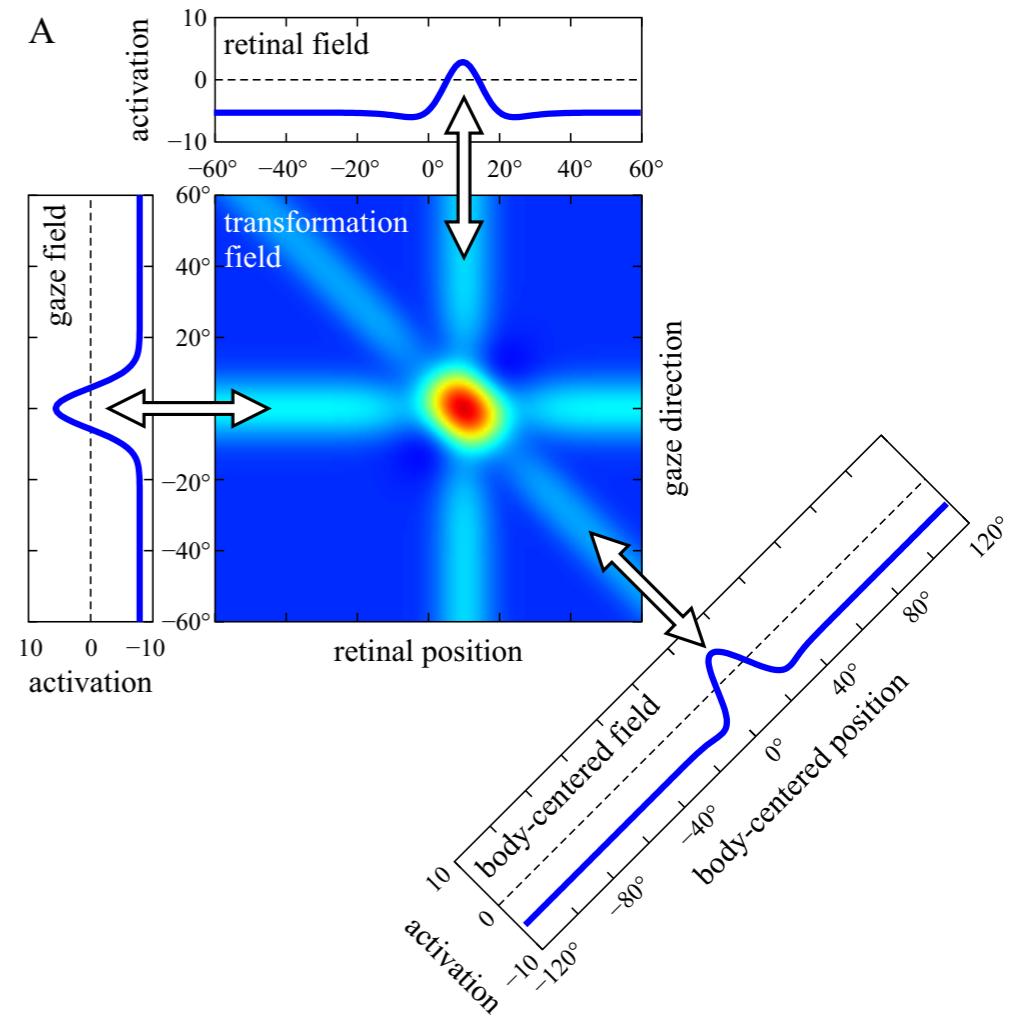
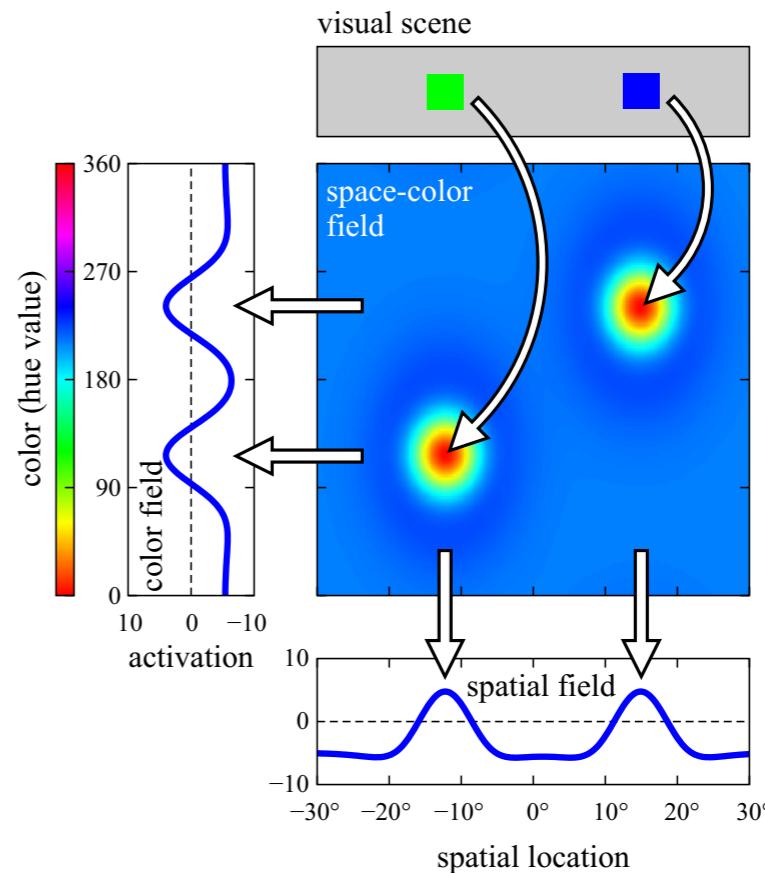
A case study in Dynamic Field Theory as a
framework for neurally grounded architectures for
higher cognition

Gregor Schöner
Institut für Neuroinformatik
Ruhr-Universität Bochum, Germany
gregor.schoener@rub.de

Previous lecture

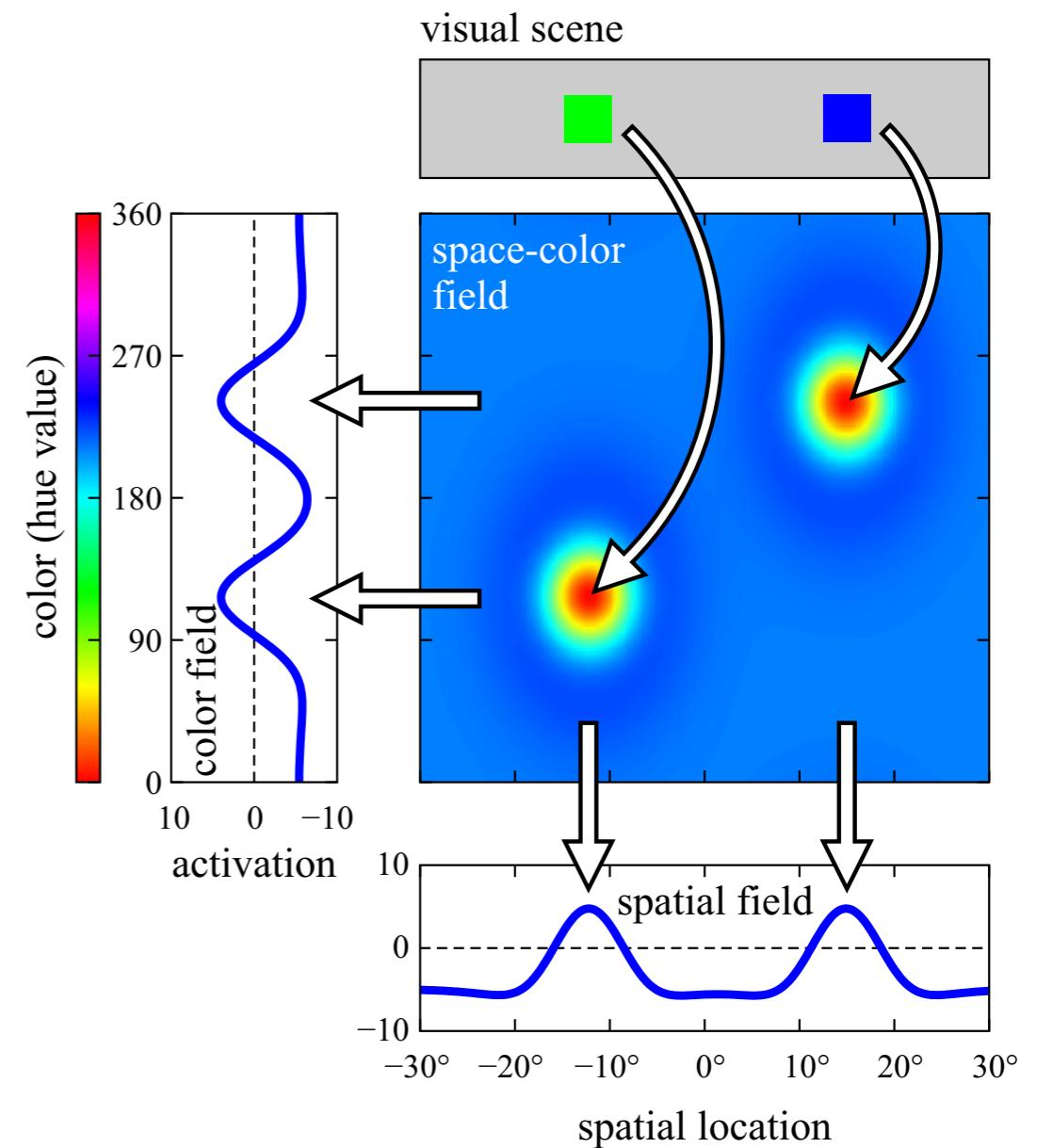
- higher-dimensional neural fields enable new functions

- binding different feature dimensions through space
- visual search... attentional selection
- coordinate transforms...



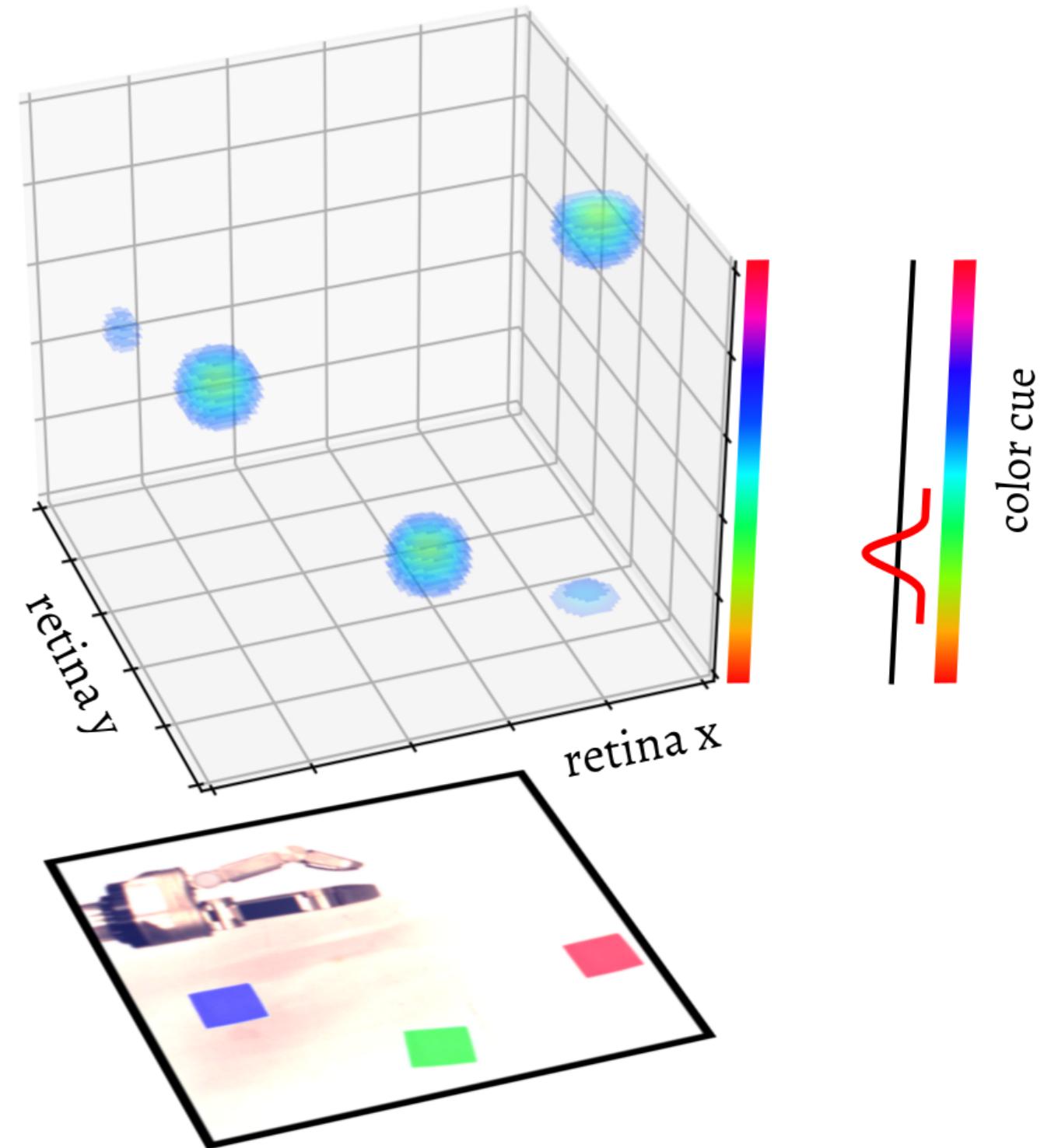
Visual search/scene representation

- Visual search: to direct action at an object first bring it into the attentional foreground...



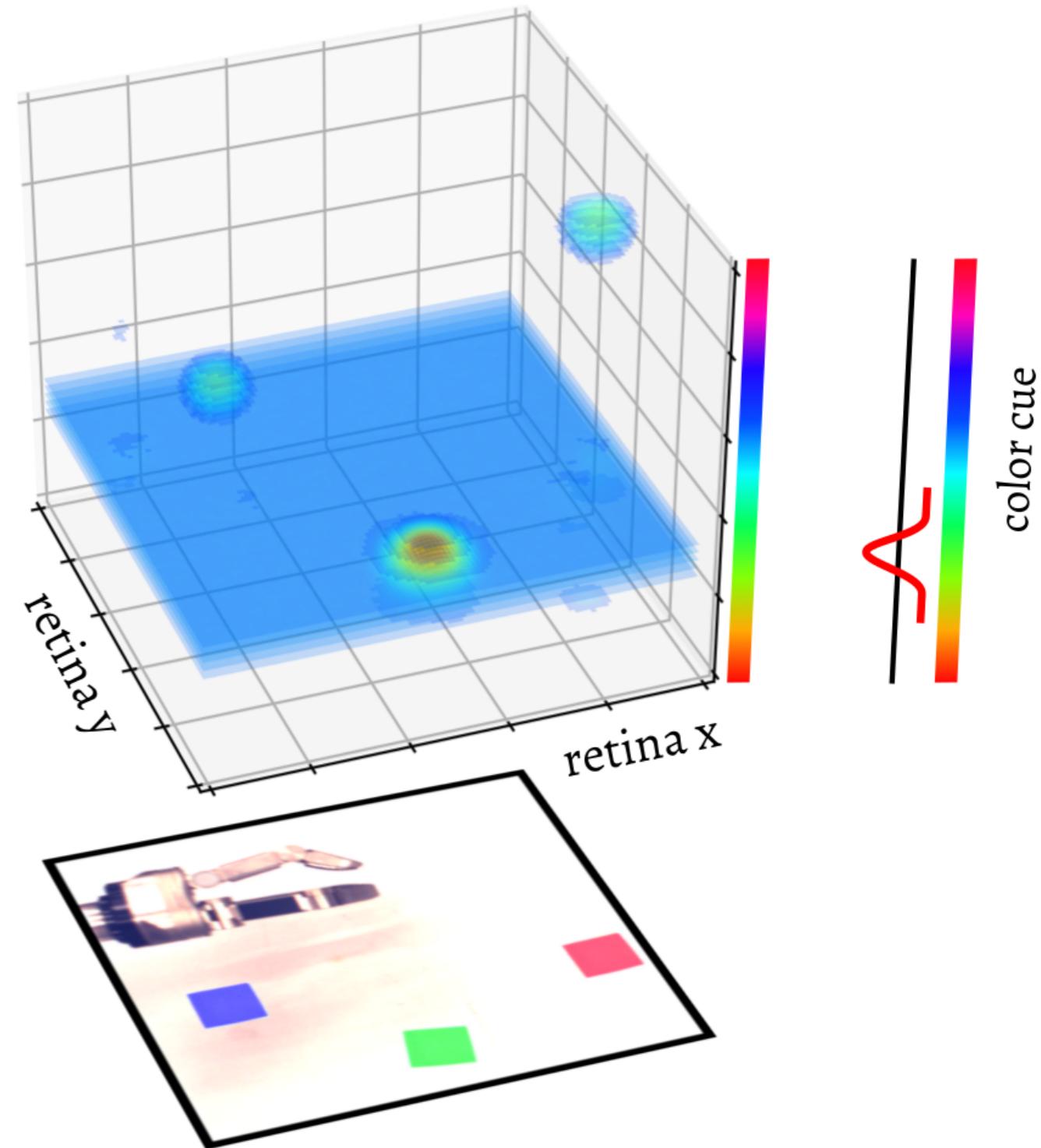
attentional selection

- to bring a cued object into the attentional foreground



attentional selection

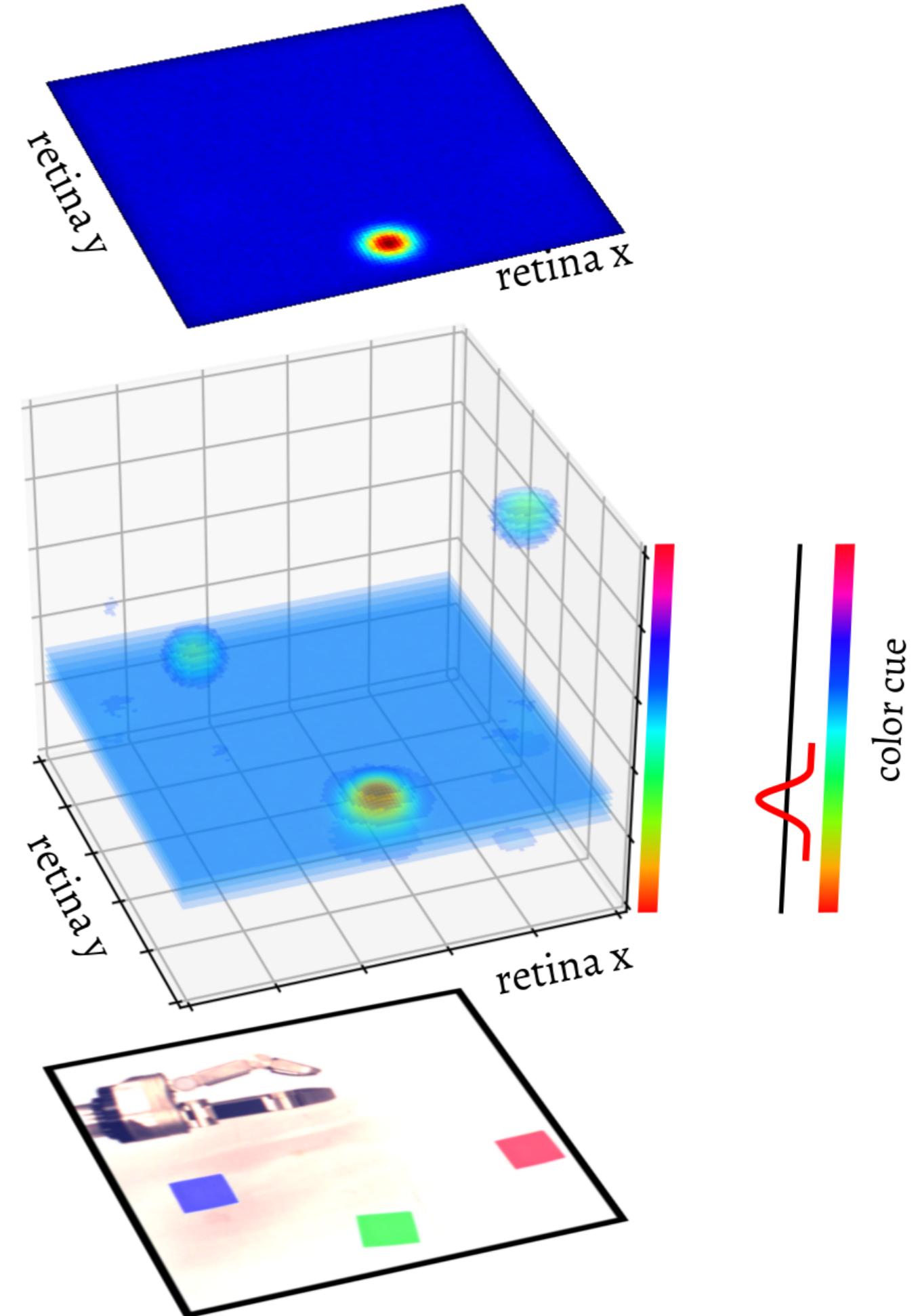
- to bring a cued object into the attentional foreground



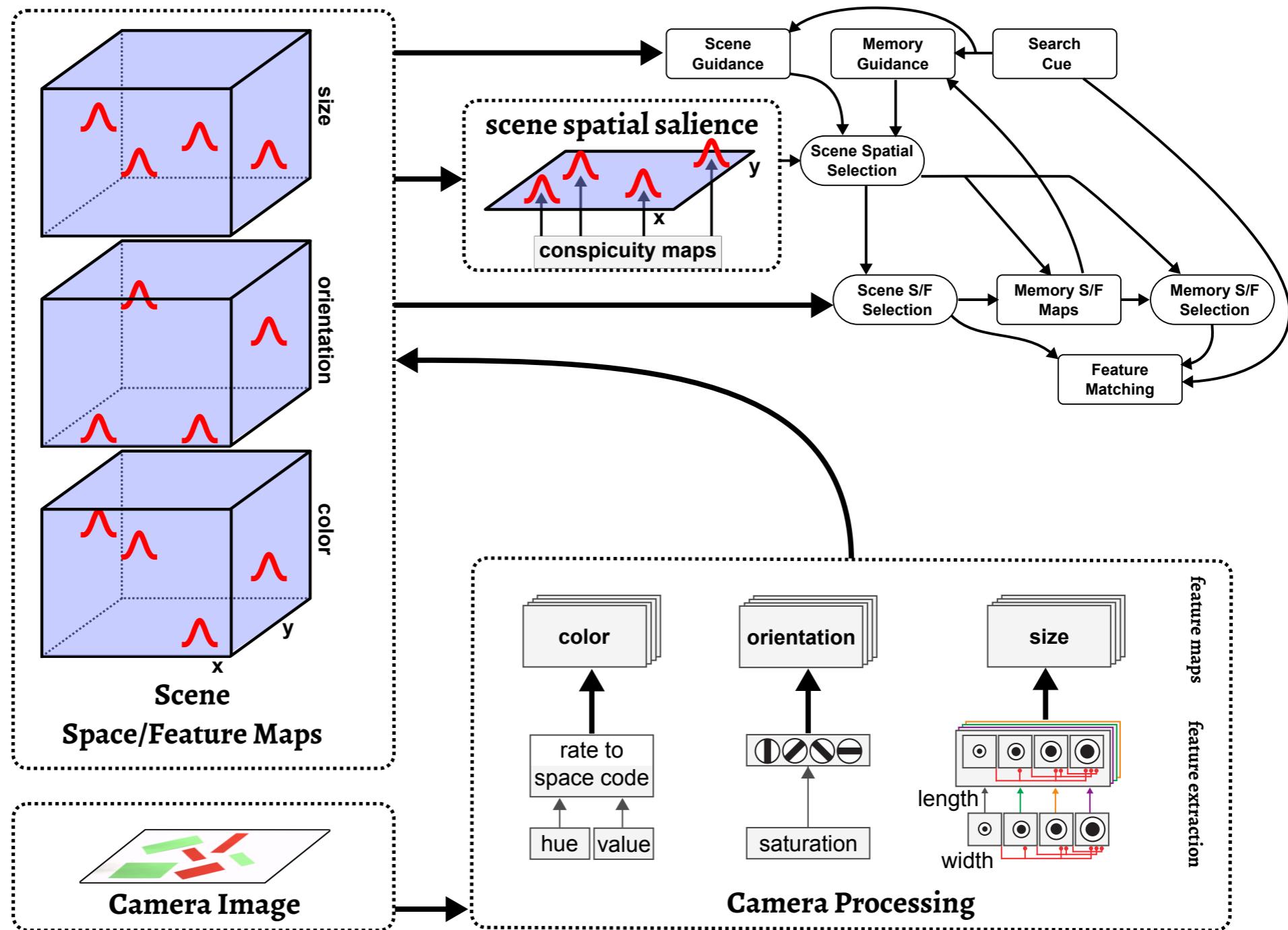
[Tekülve, 2020]

attentional selection

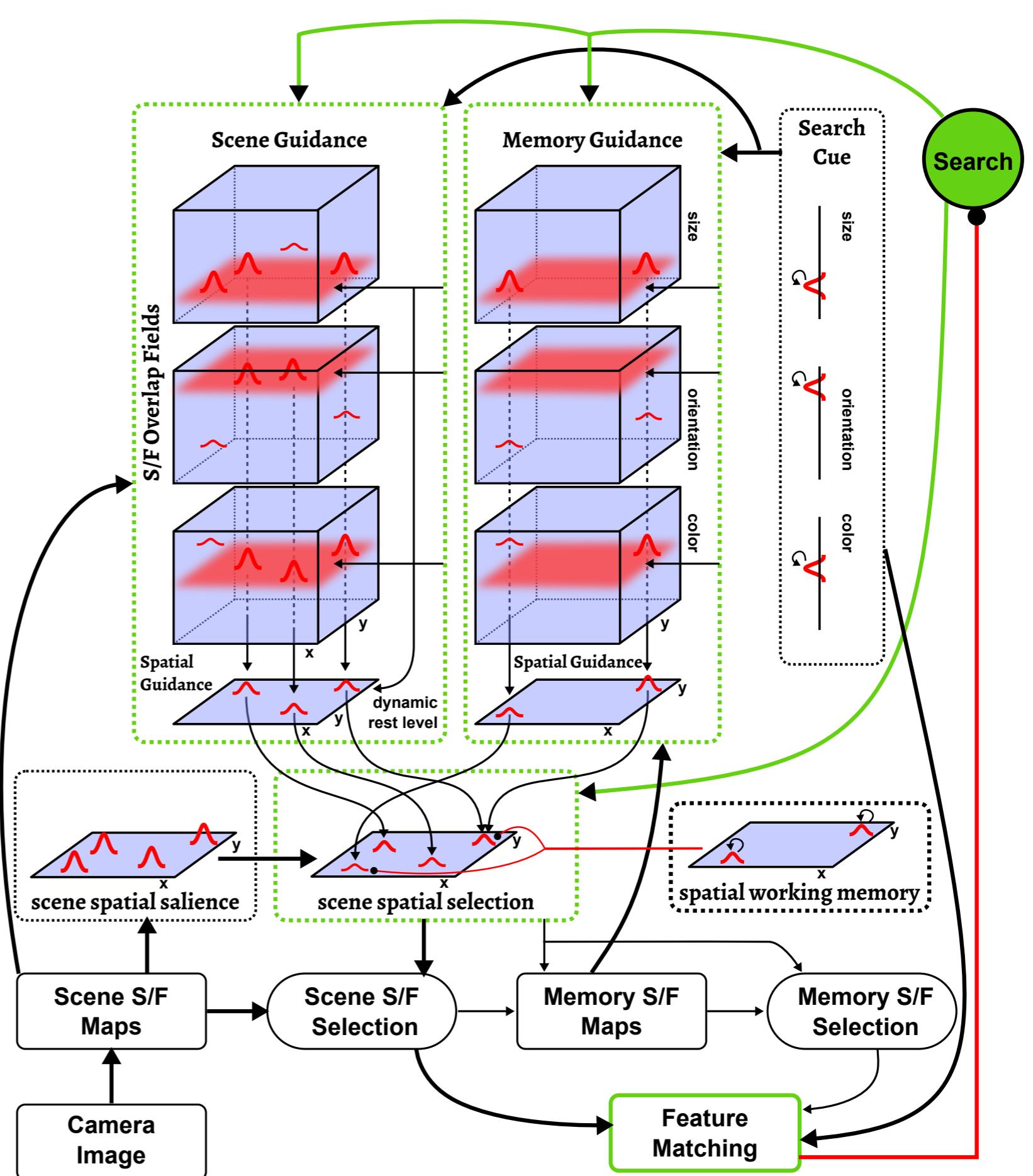
- to bring a cued object into the attentional foreground



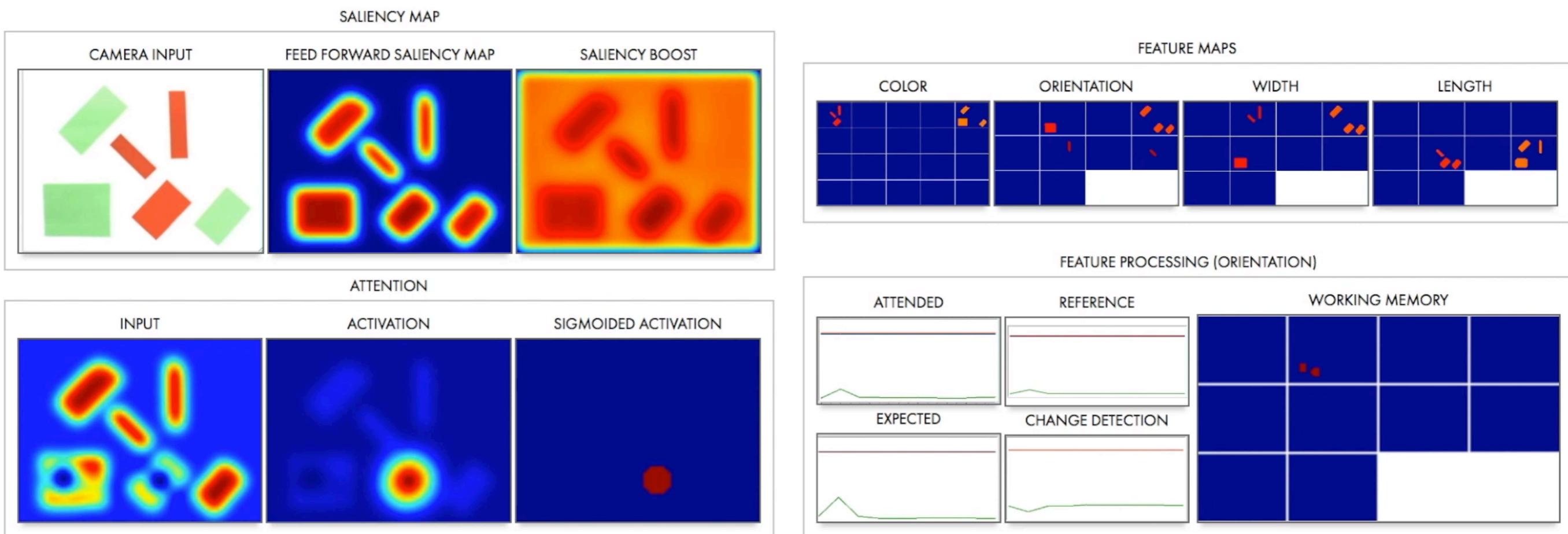
Neural dynamic architecture for visual search and scene memory



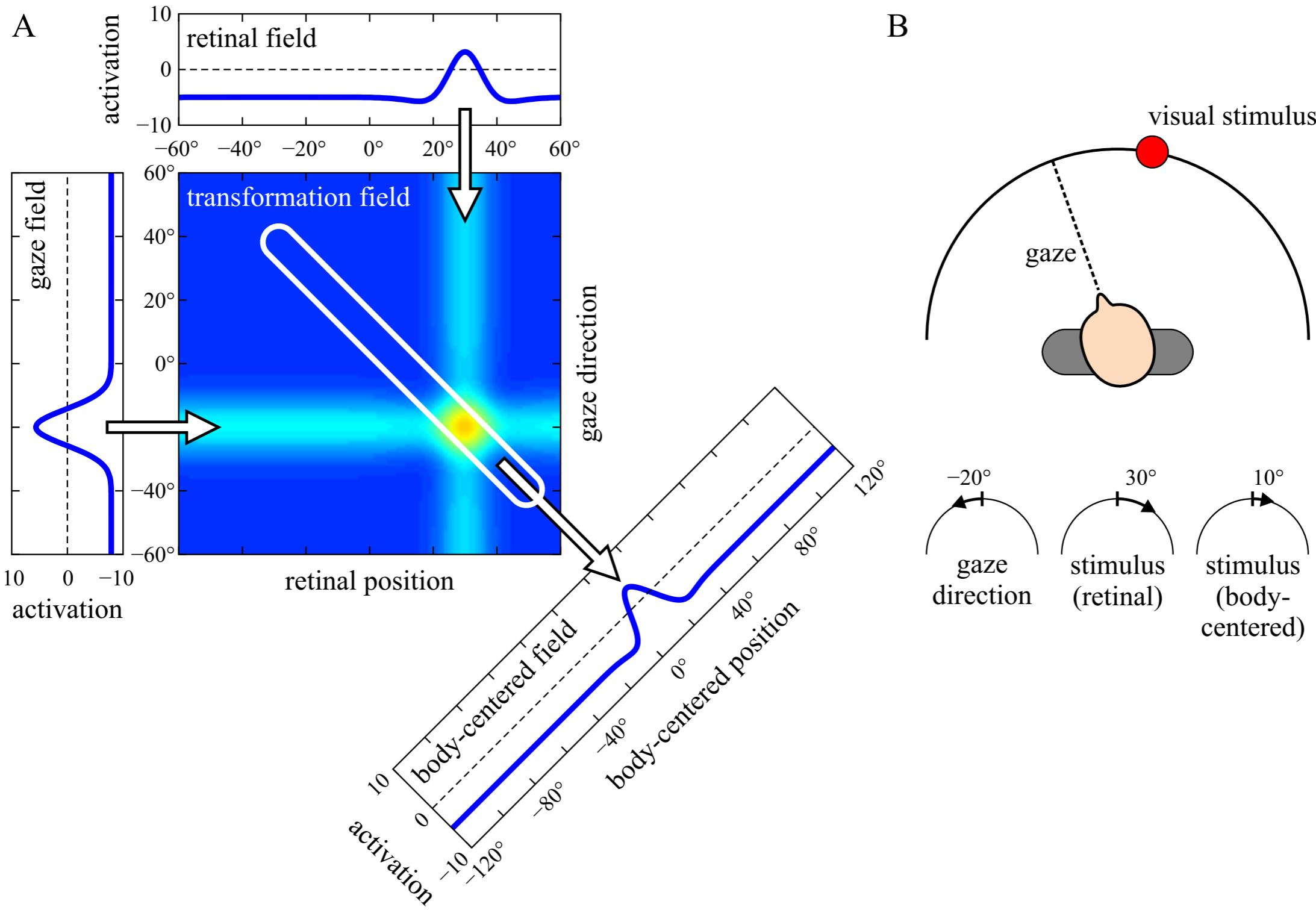
Attentional selection



Autonomous sequences of visual exploration and cued visual search



Coordinate transforms



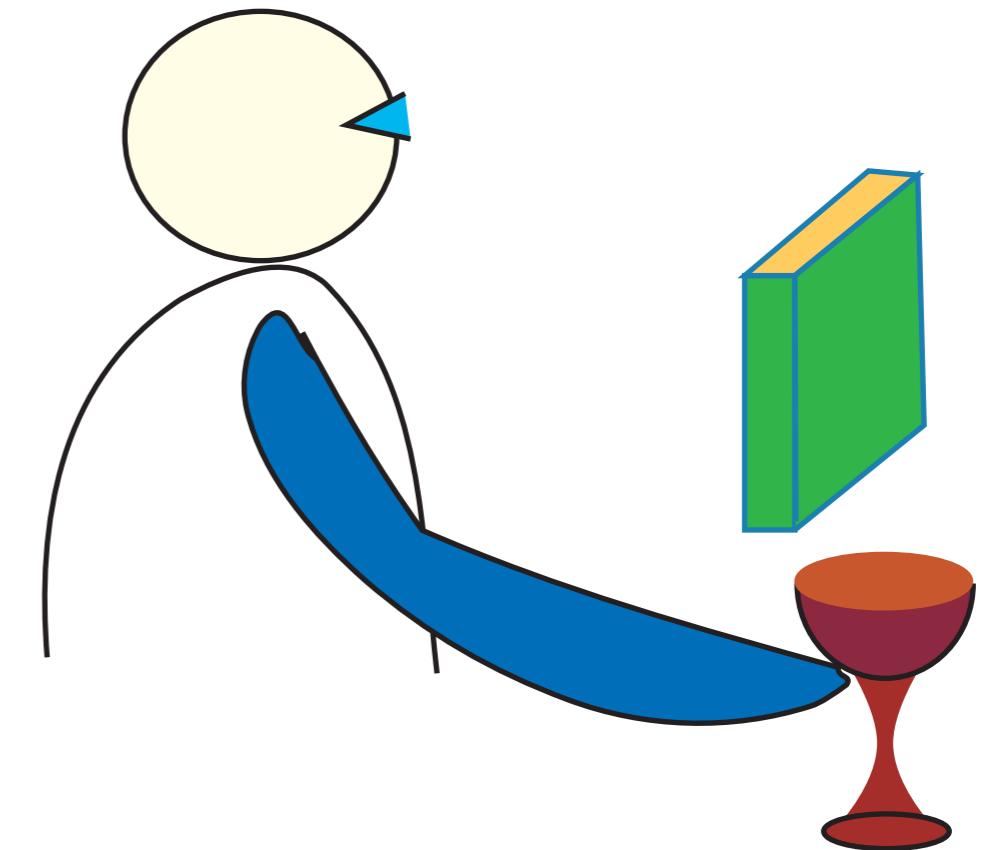
[Schneegans Ch 7 of DFT Primer, 2016]

Today

- illustrate how these functions support lifting neural dynamics to higher cognition
- in the context of the perceptual grounding of concepts/language/relations

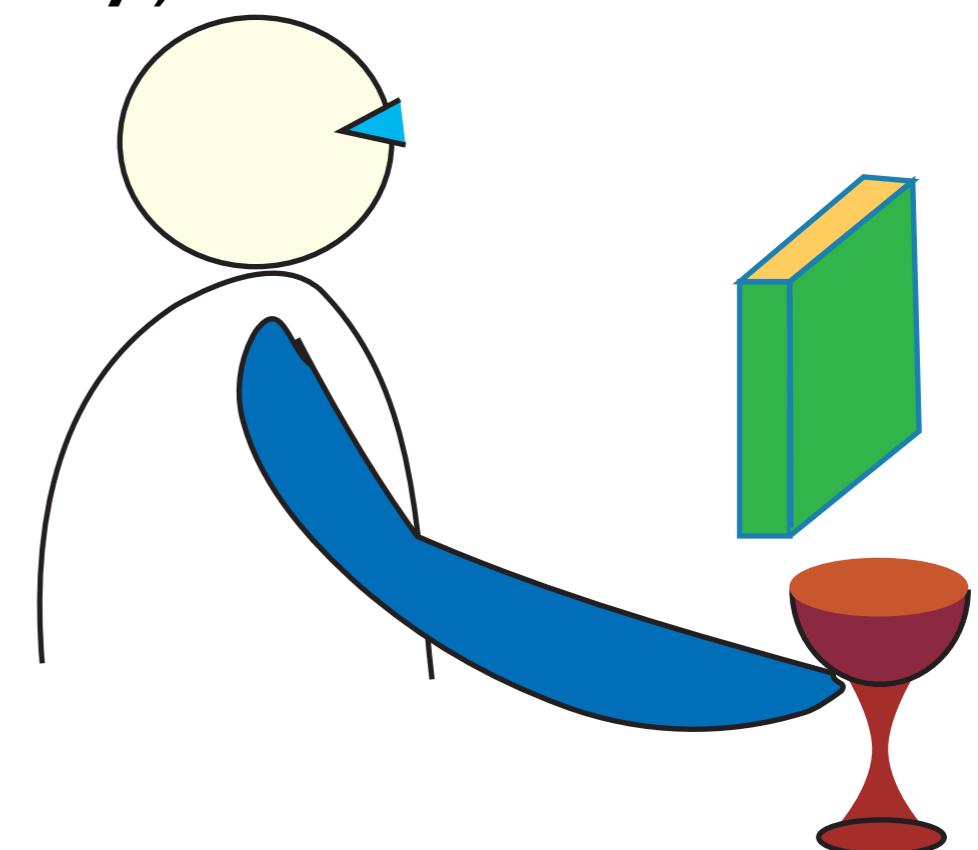
Perceptually grounding language

- human communication in its simplest form is about things that are out there in our environment, perceivable, reachable by action
- e.g., this cup is brown



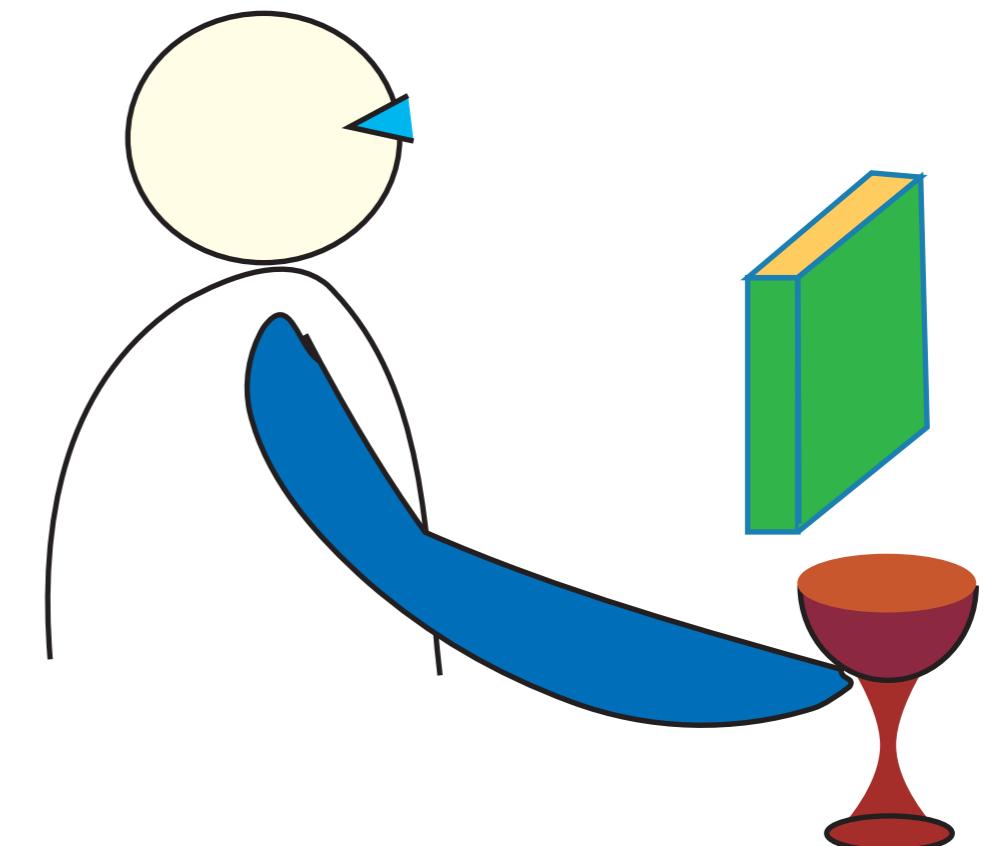
Perceptually grounding language

- this could be based by both the speaker and the listener looking at the scene and grounding the word “cup” by bringing an object of that category into the foreground
- also called “targetting” (Talmy)



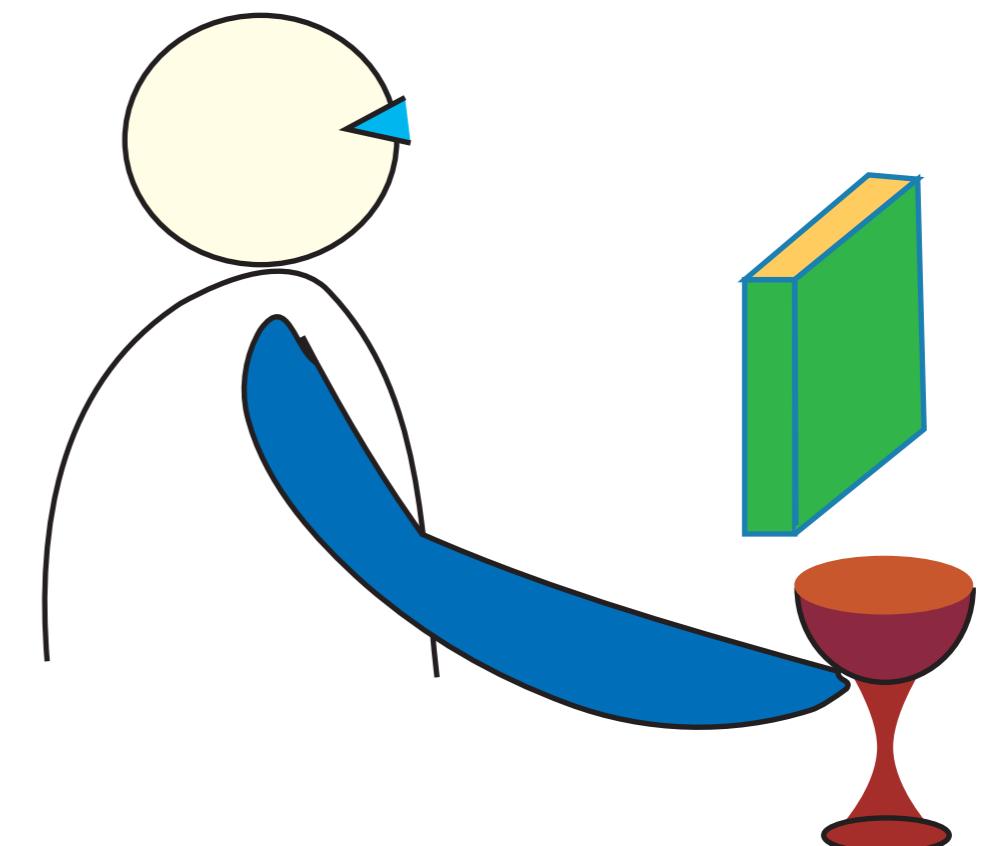
Perceptually grounding language

- that process could be mediated by other forms of communication, e.g., pointing (deictic code)



Perceptually grounding language

- that process could also be mediated by spatial language, e.g., “the cup to the right of the green book is brown” (**spatial language**)
- (which presupposes that the reference object “green book” is grounded for speaker and observer)



Perceptually grounding language vs. describing

“what is to the right of the green object”

- Perceptual grounding: understanding phrases by finding in the visual array the objects to which the phrase refers



- Describing: producing phrases that describe an observed scene or event



Spatial language

- such utterances as “to the left of”, “on top of”, “in”, “in front of”, “toward the south”, “in front of” etc.
- a part of language that deep: evolves slowly in languages, with profound differences between languages and cultures, that is particularly challenging for “grounding”

Spatial language

■ Examples:

- some cultures use absolute directions “north”, “south” etc. even on a local scale (e.g., “the car north of the house” rather than “the car in front of the house”).
- others have special spatial language referring to geographical landmarks (e.g., islanders who have a word for “toward the beach” vs. “away from the beach, toward the inland”)
- “in front of” is used differently even in different indo-european languages

Grounding spatial language

- involves necessarily reference frames... there are 4 basic and commonly used reference frames

Grounding spatial language

- orientation relative to speaker, position centered in speaker
 - “on my left”
- orientation relative to world/object, position centered in speaker:
 - “north”, “south...” or “leeward”, “windward” ...
- orientation relative to speaker, position centered in object
 - “the cup to the right of the bottle”
- orientation relative to object, position centered in object
 - “leave the train on the right hand side”

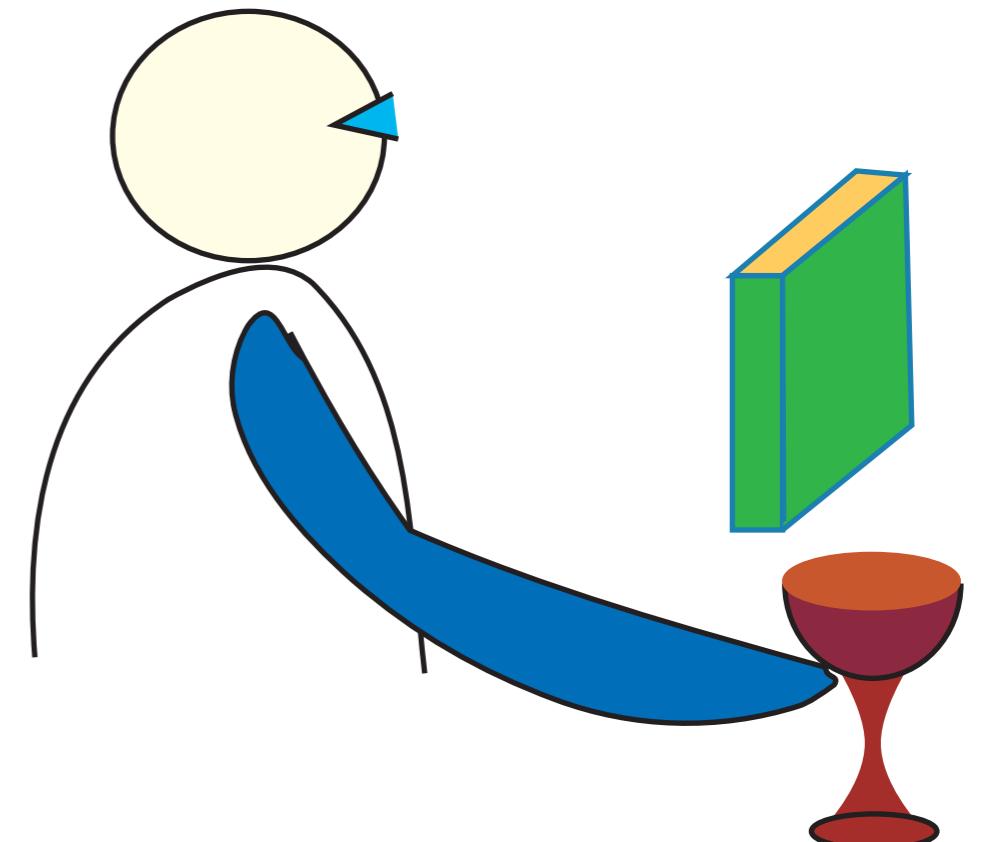
Grounding spatial language

■ reference frames are subtle

- Example: “in front of” can be in an ego-centric frame if the object has no special long axis and front end (e.g., “in front of the tree” meaning “between me and the tree”)
- but can be in an object centered frame if the object has a long axis and front end (e.g. “in front of the car” meaning “on the side of the car in the direction in which its front end points”)
- (and on this count different languages differ)

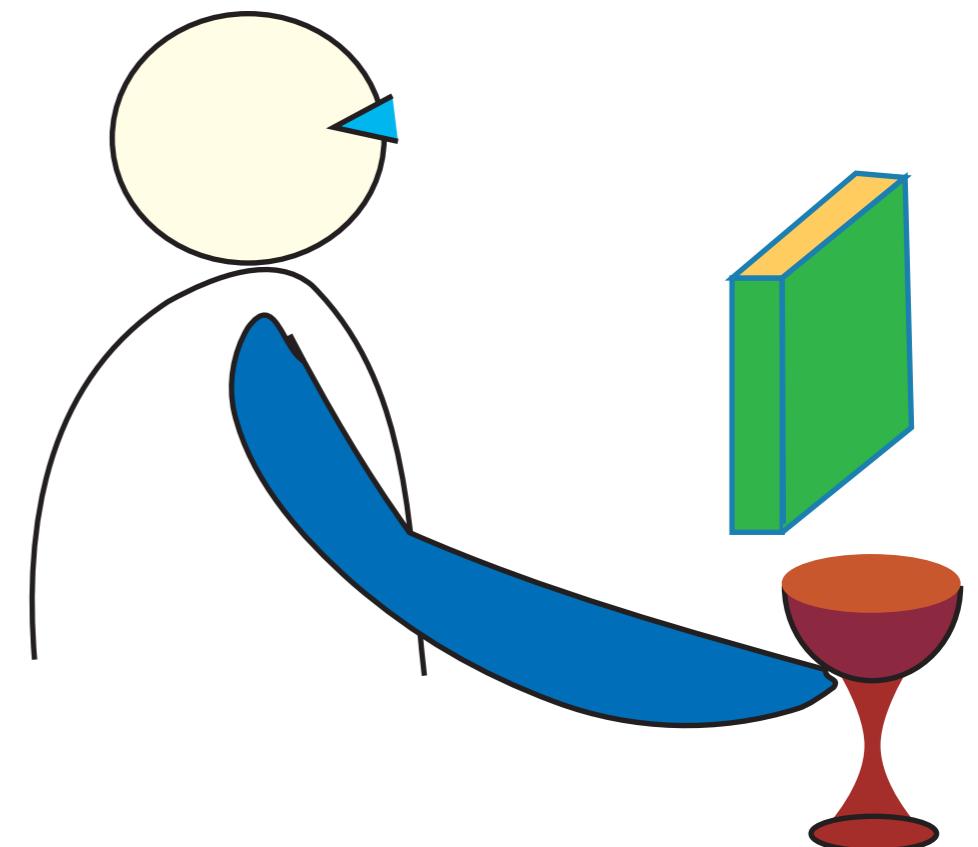
Grounding spatial language

- spatial language often involves reference objects
 - Example: “to the right of the green book”: this is a statement in an ego-centric reference frame for direction but that is spatially centered in an object

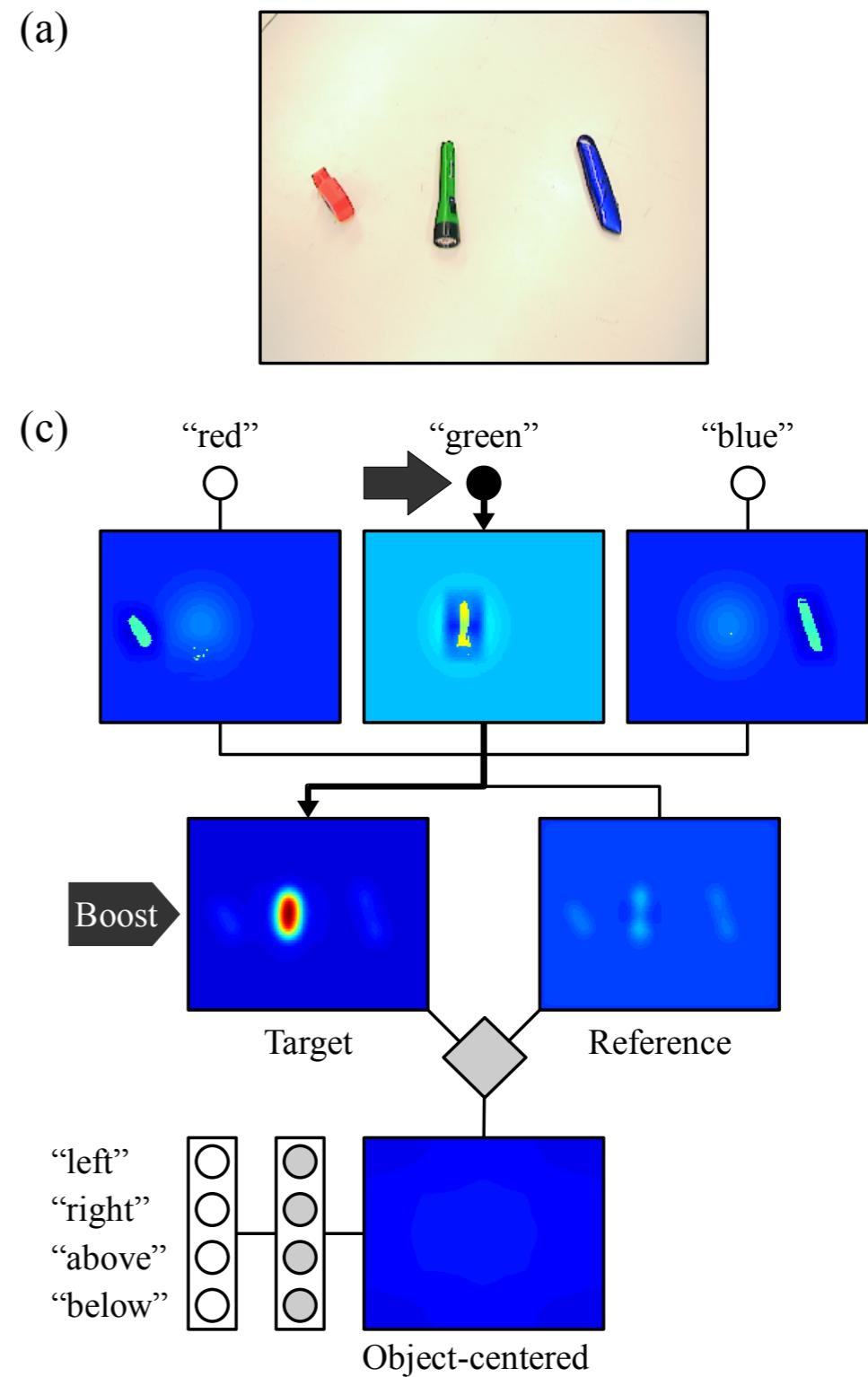


Grounding spatial language

- spatial language often involves coordinate transforms
 - e.g., “to the right of the green book”: coordinate transformation: from the speaker/observer centered reference frame into a frame centered in the reference object
 - e.g., “to my right” requires the listener to transform the reference frame from his or her own view to the directional and positional frame of the speaker

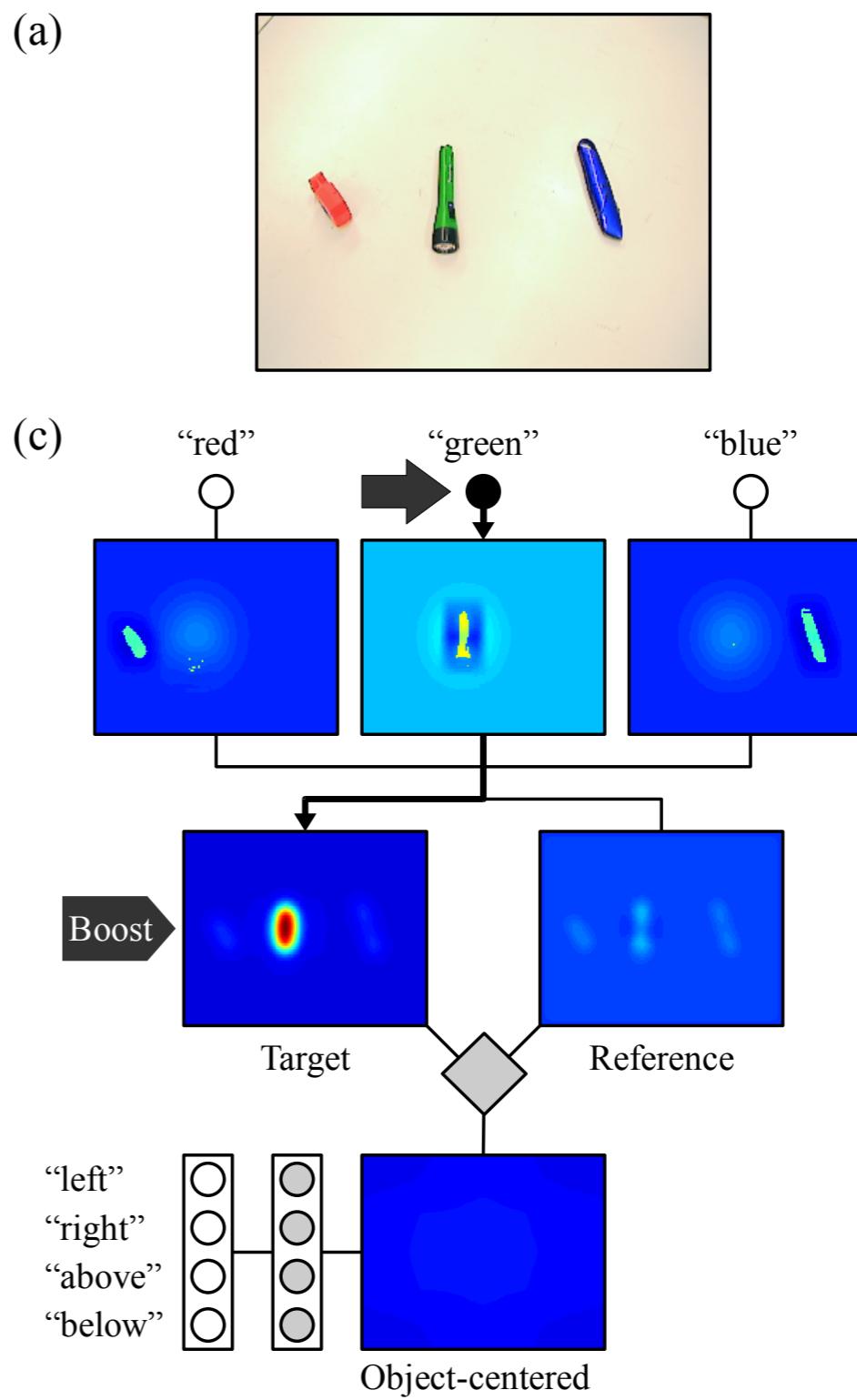


A cognitive architecture for grounded spatial language in DFT



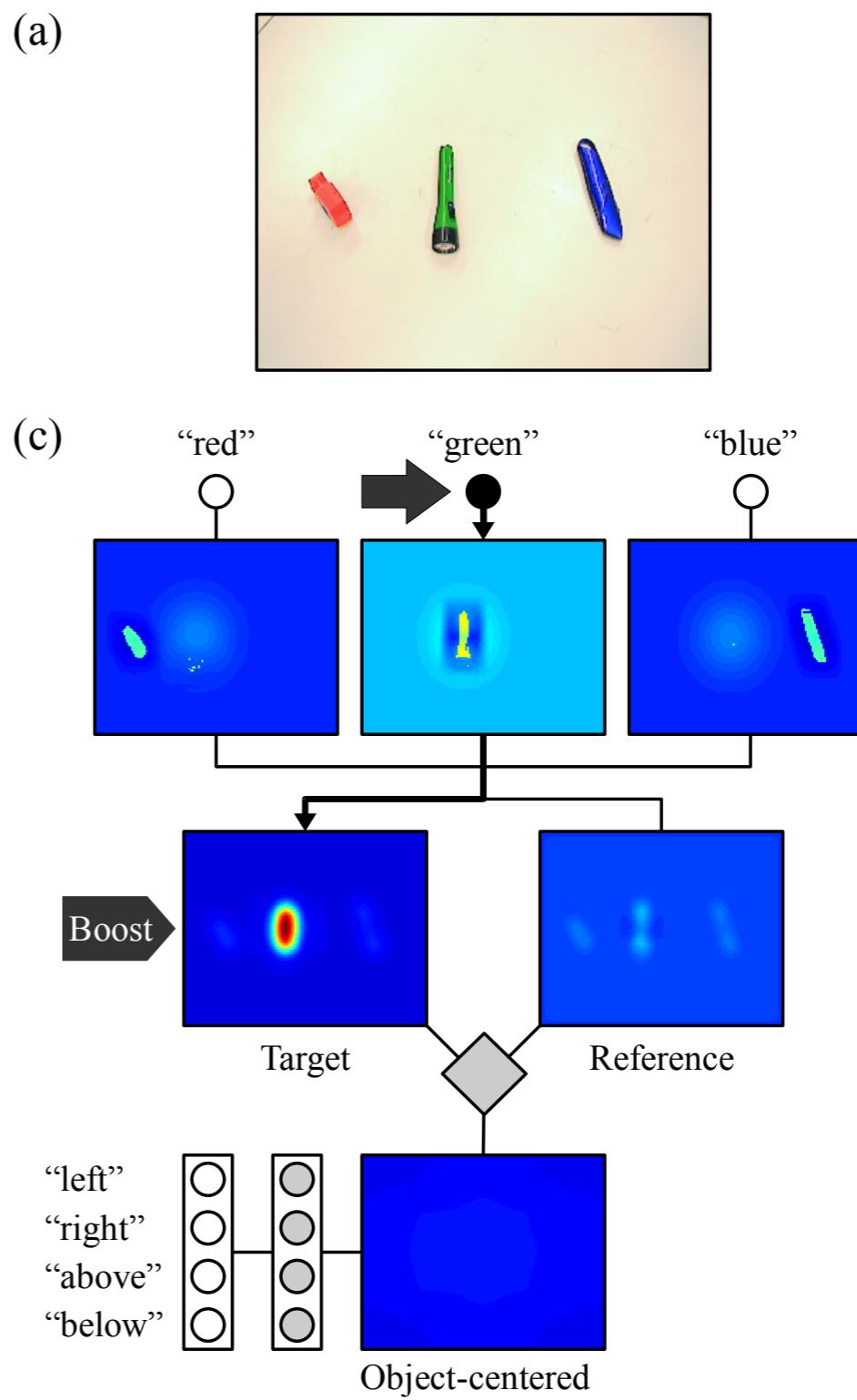
Spatial comparison in DFT

- bring objects into foreground
- make coordinate transformation
- apply comparison operators



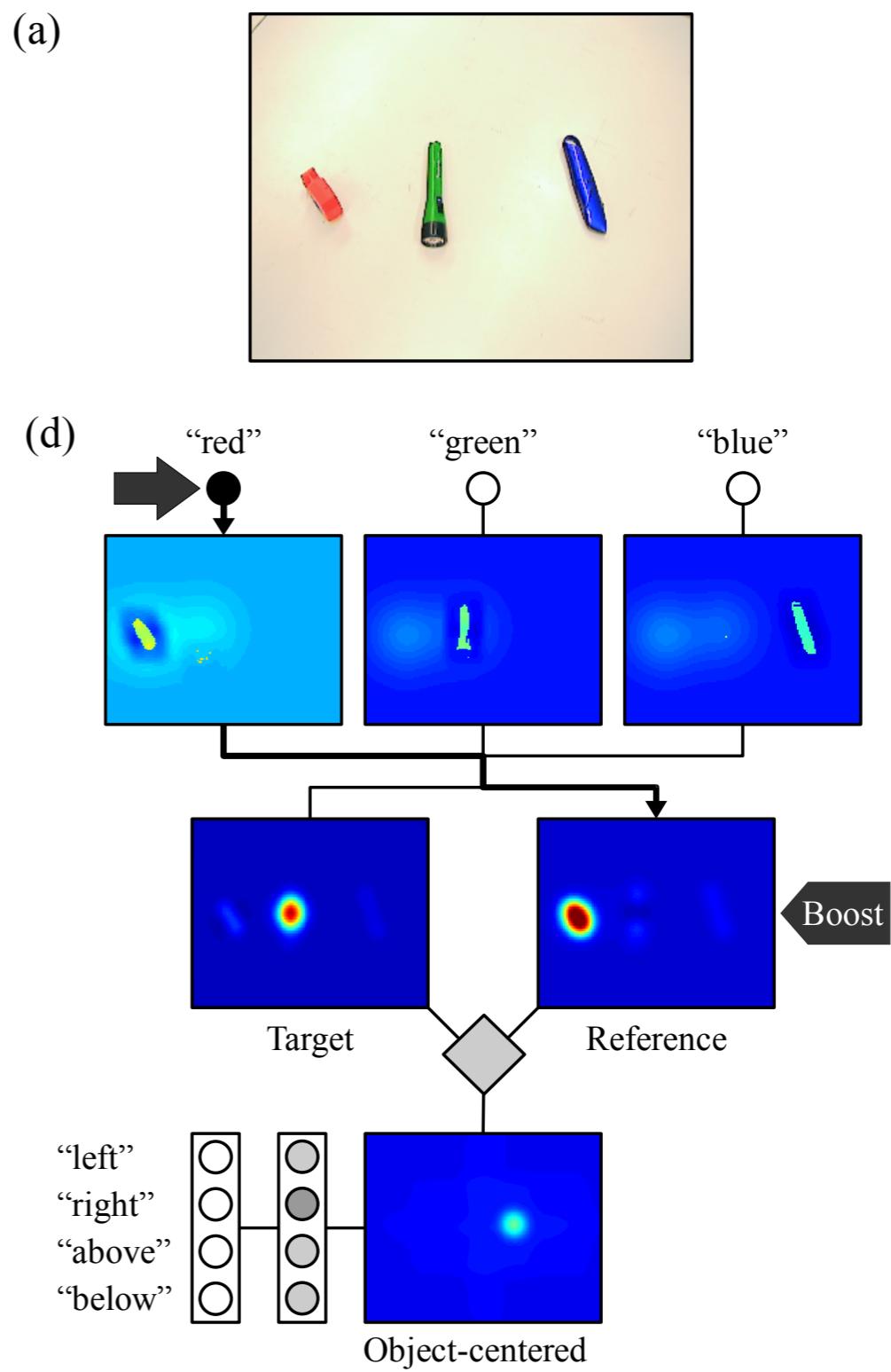
Spatial comparison in DFT

- bring objects into foreground
- make coordinate transformation
- apply comparison operators



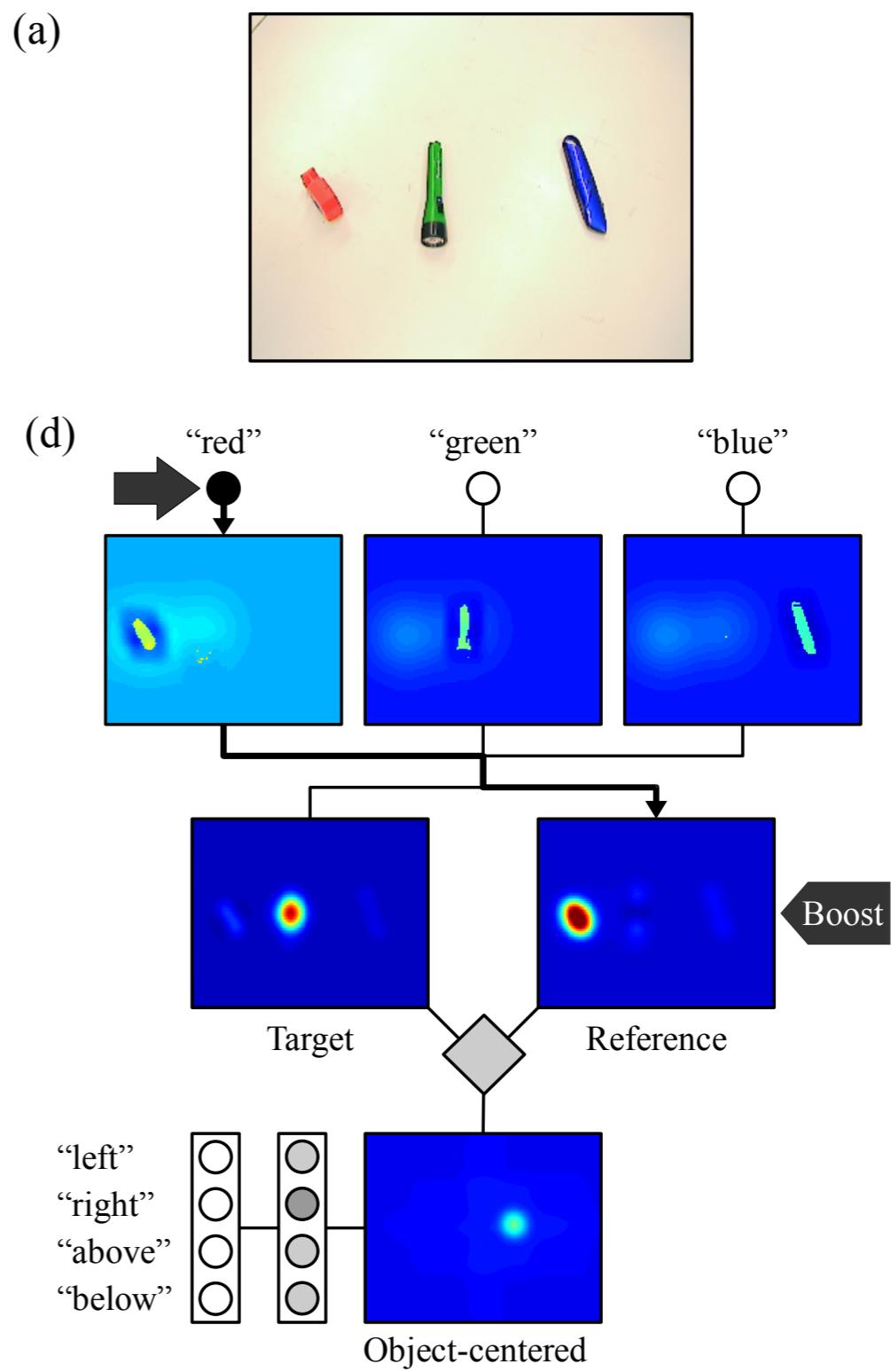
Spatial comparison in DFT

- bring objects into foreground
- make coordinate transformation
- apply comparison operators



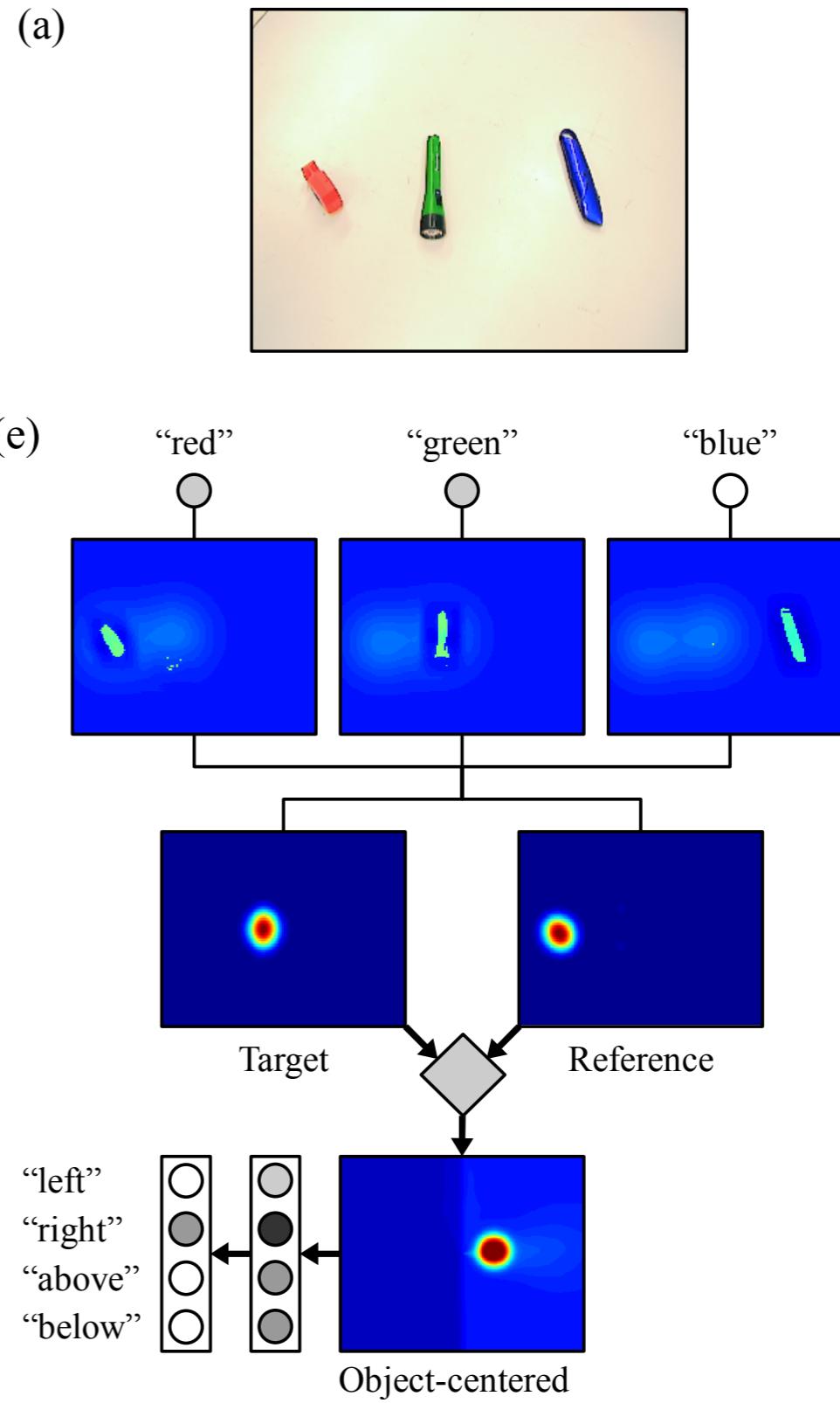
Spatial comparison in DFT

- bring objects into foreground
- make coordinate transformation
- apply comparison operators

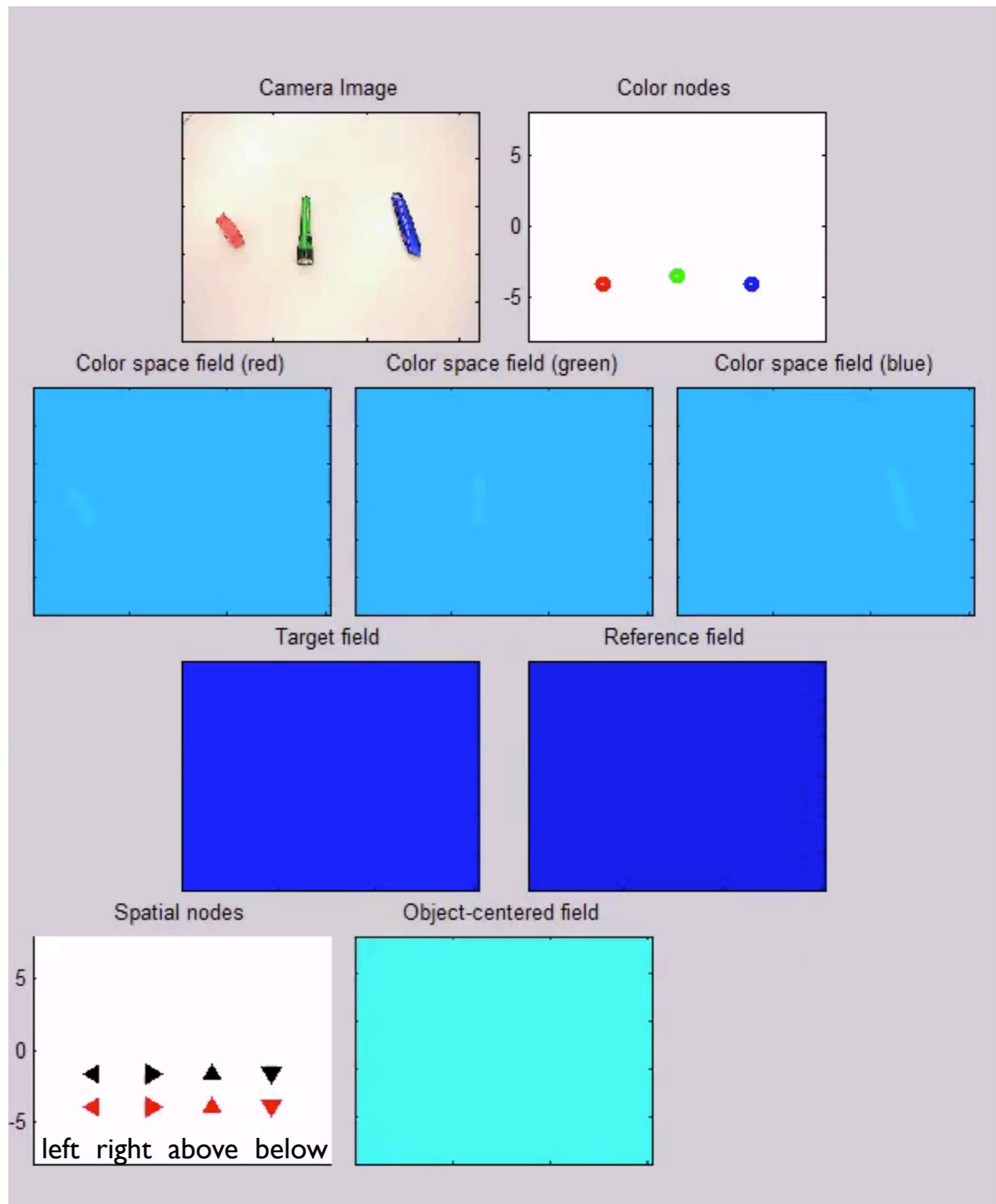


Spatial comparison in DFT

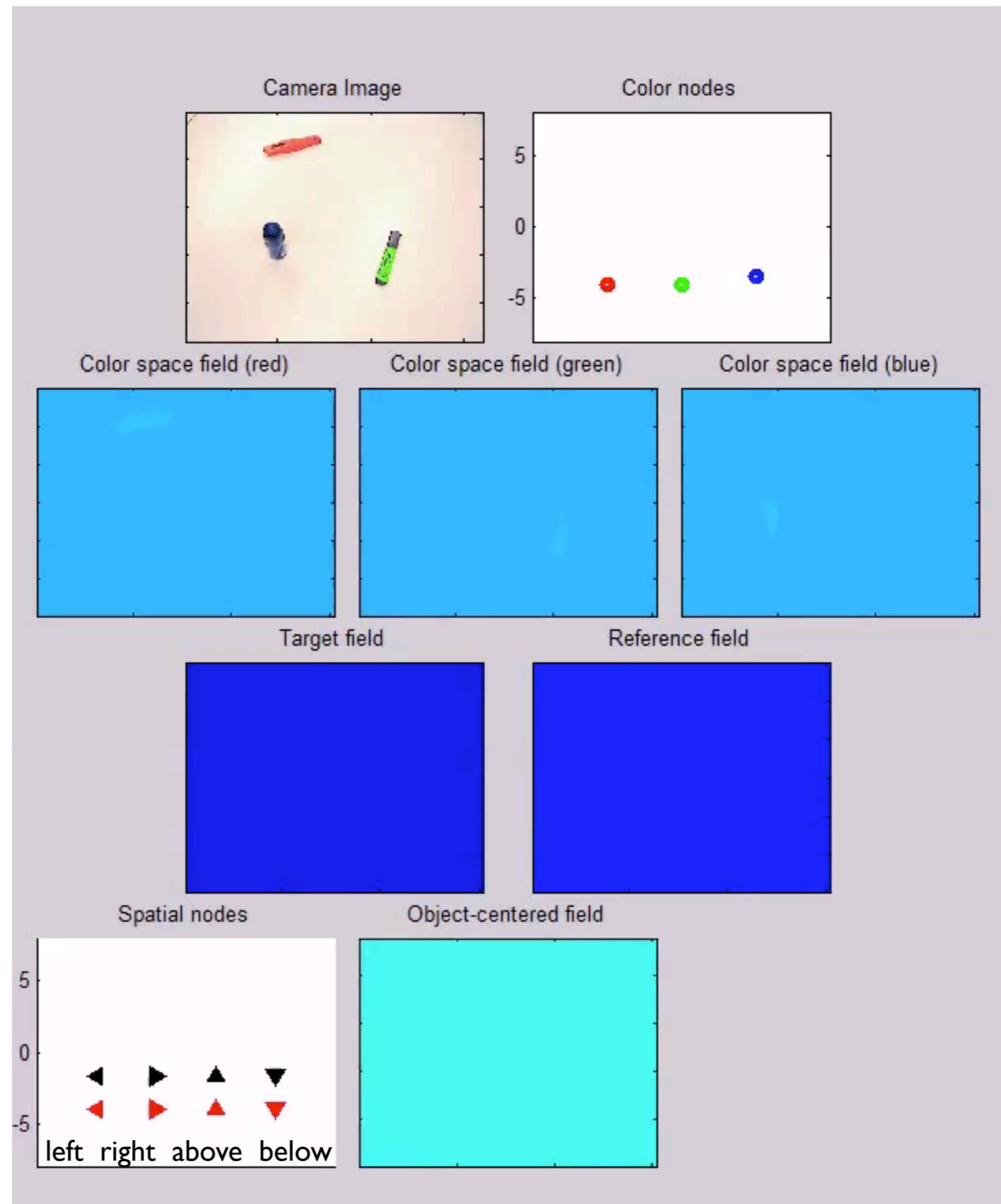
- bring objects into foreground
- make coordinate transformation
- apply comparison operators



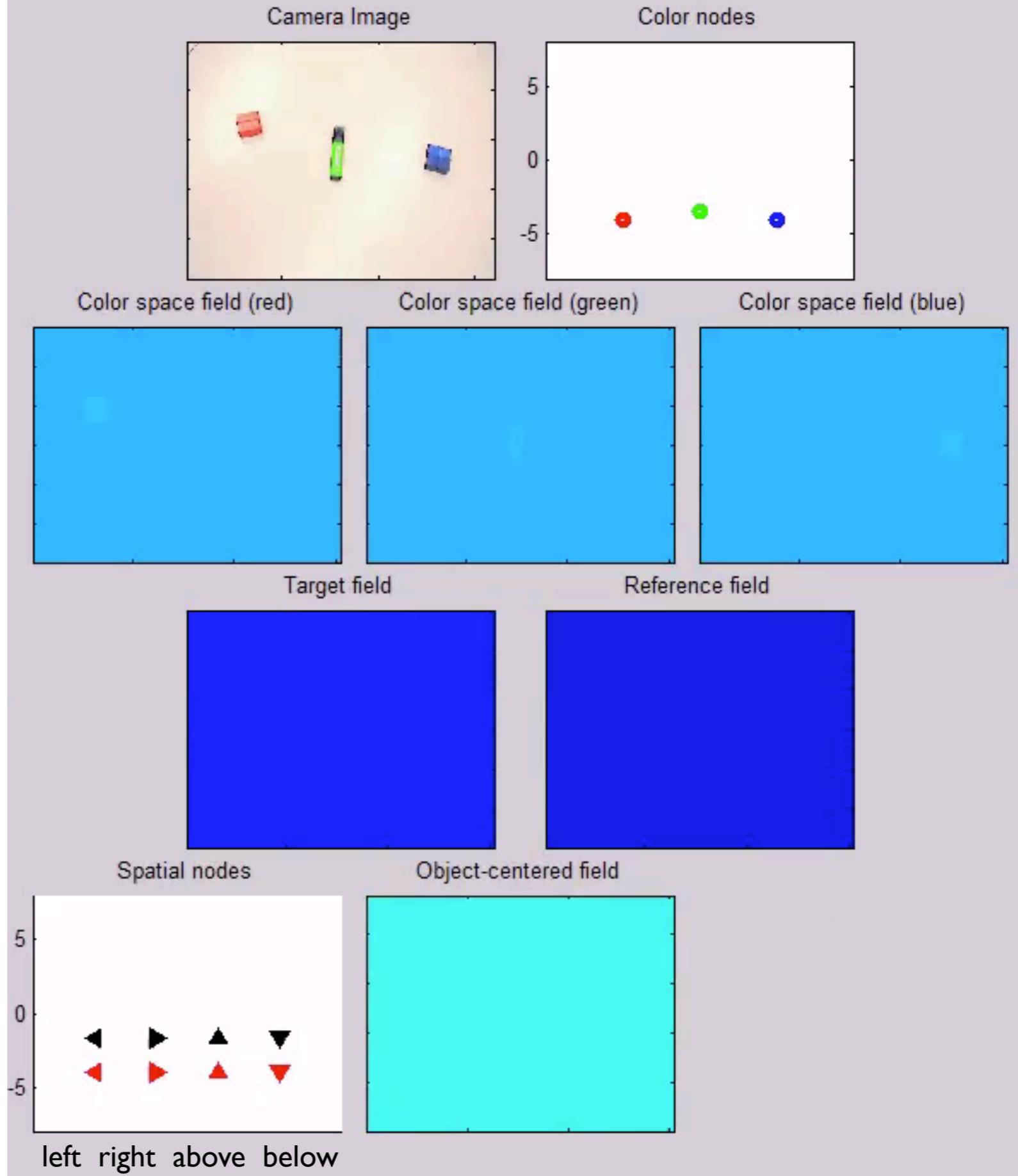
“where is
the green
object
relative
to the
red
object?”



■ “which object is above the blue object?”

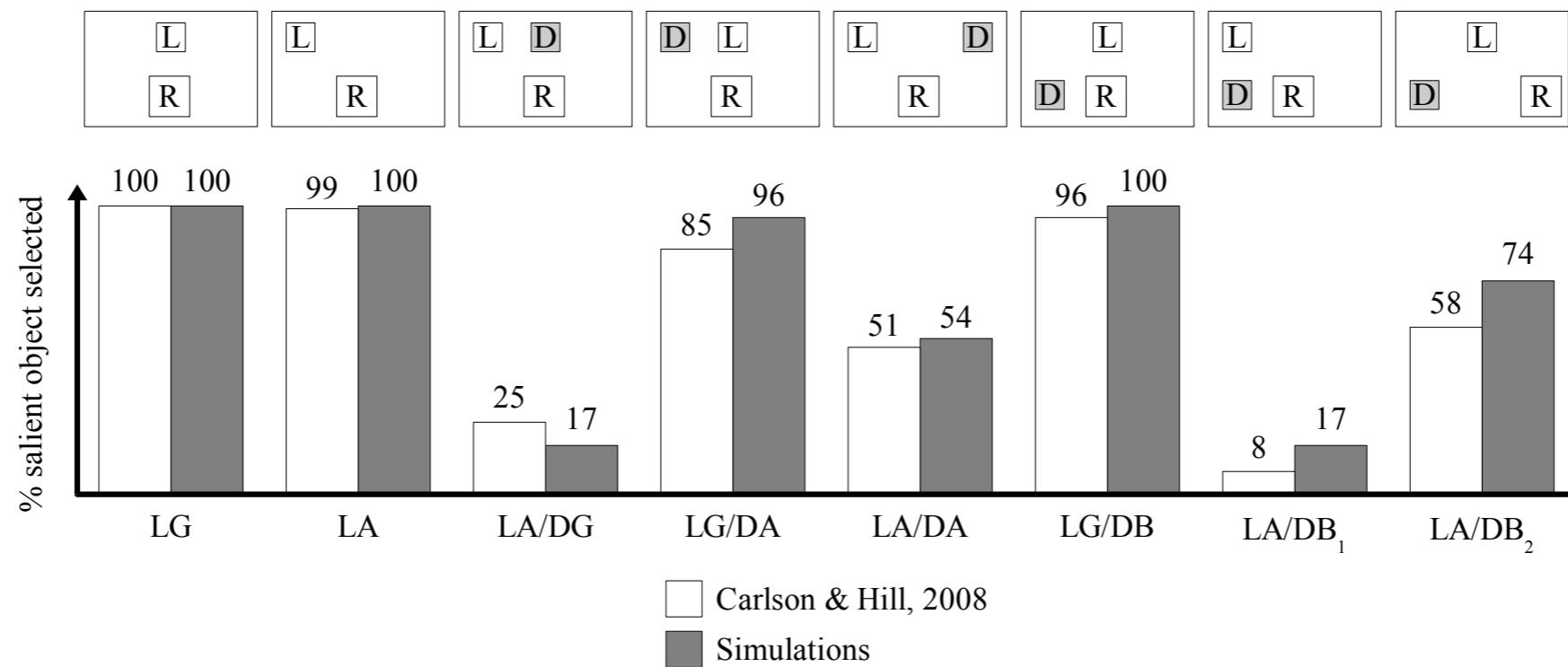


“where is
the green
object?”



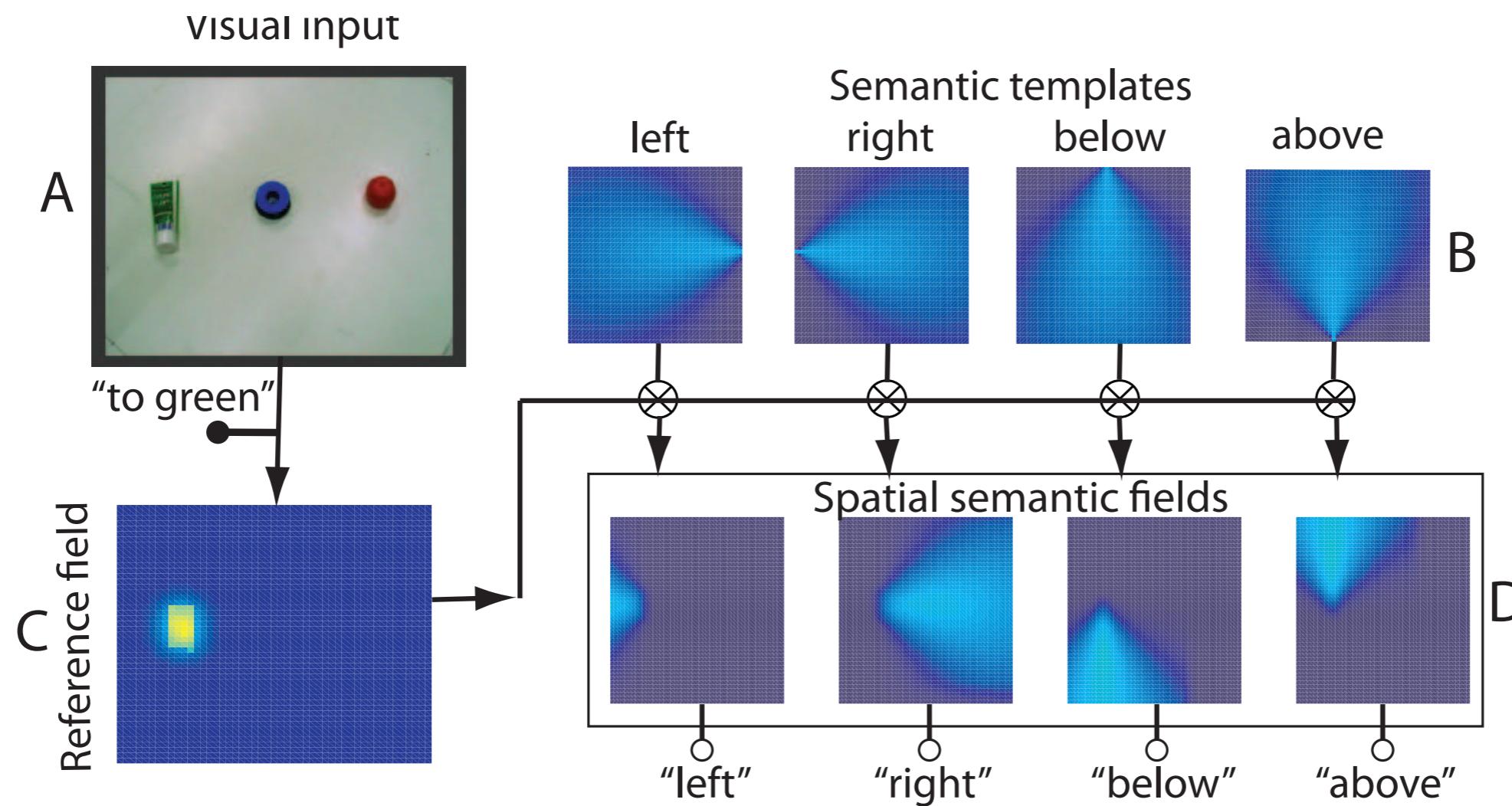
Spatial comparison in DFT

■ accounts for human data



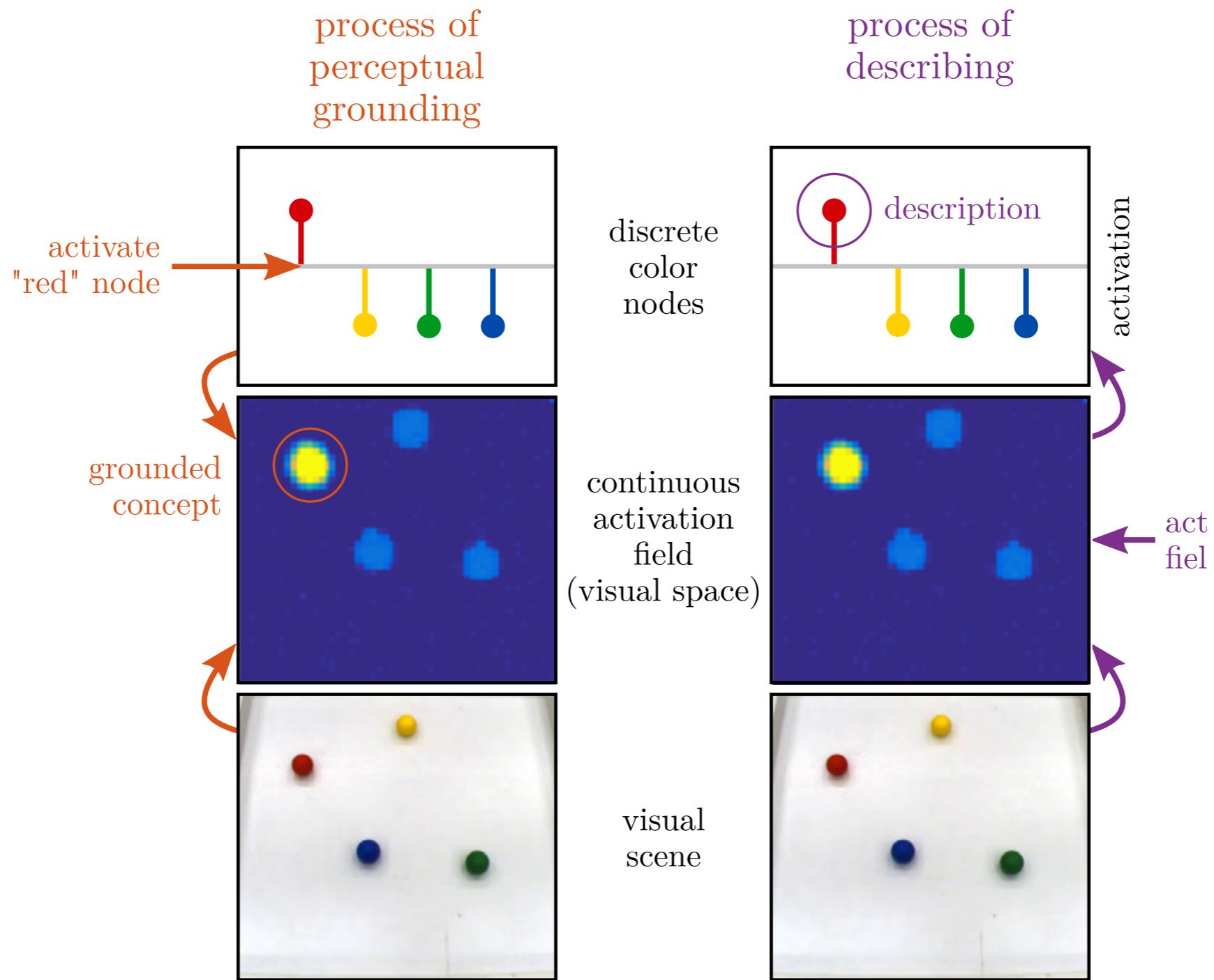
Alternative: coordinate transform applied to the neural operator

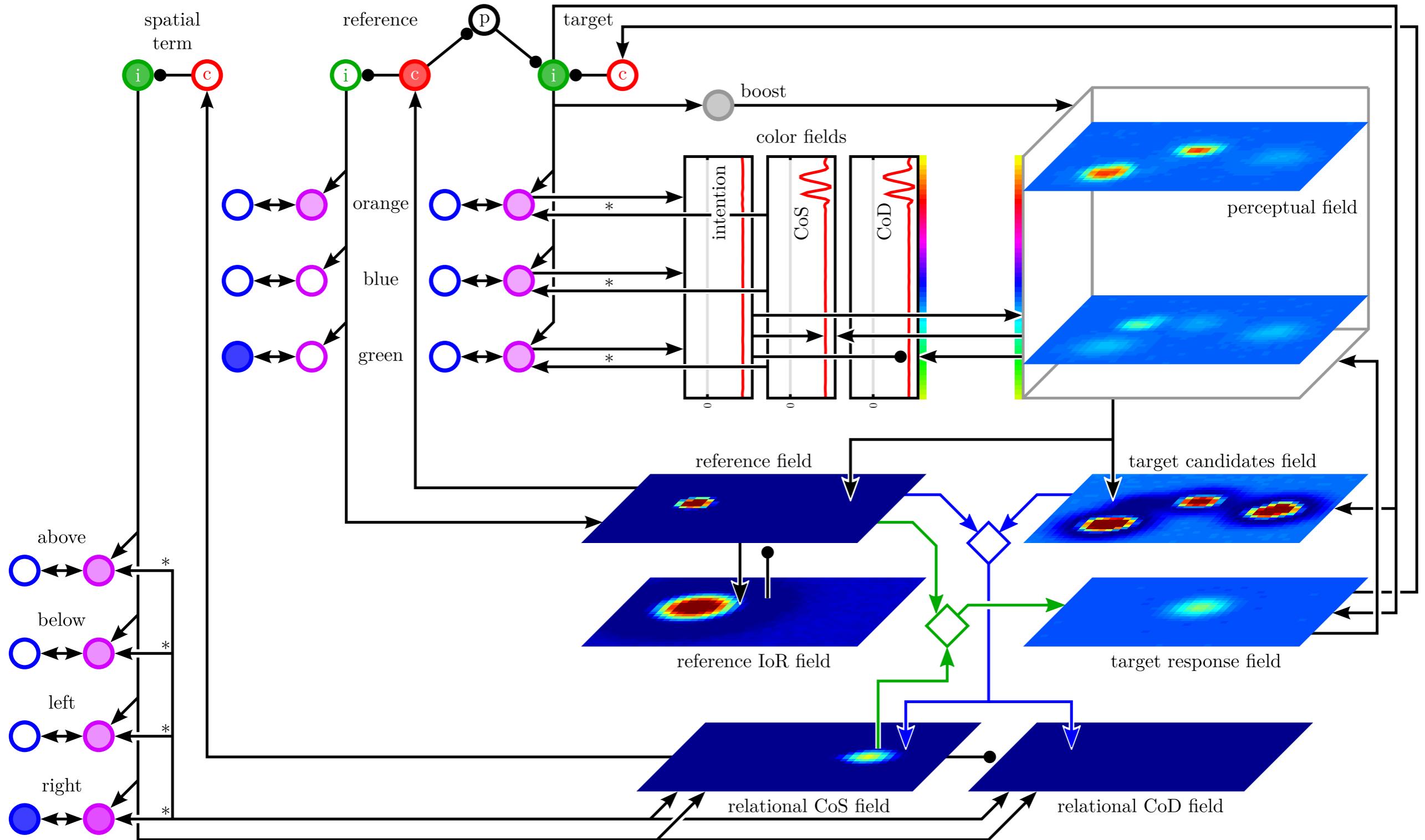
- based on convolution of fields with kernels



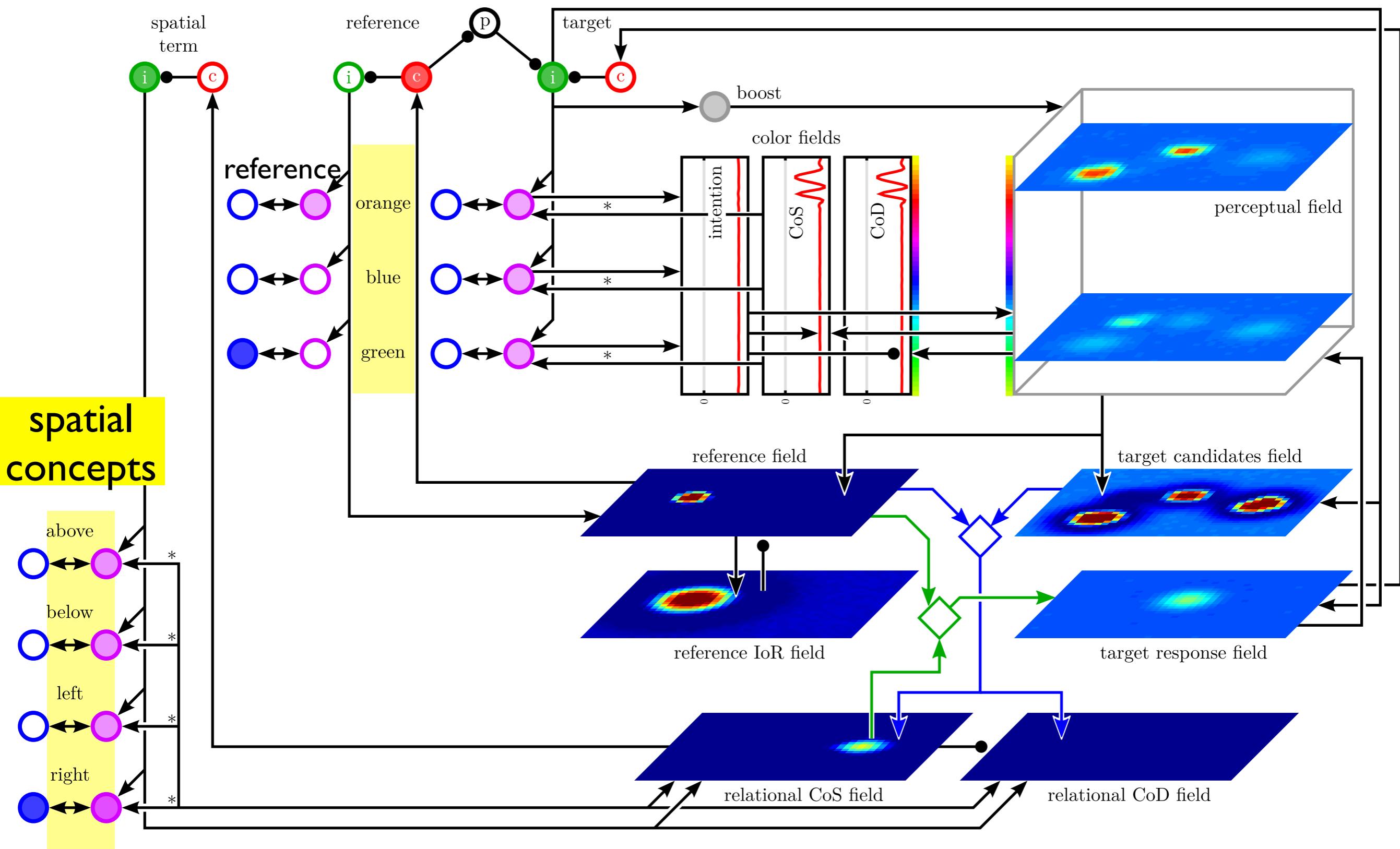
[from: Lipinski, Sandamirskaya, Schöner, 2009]

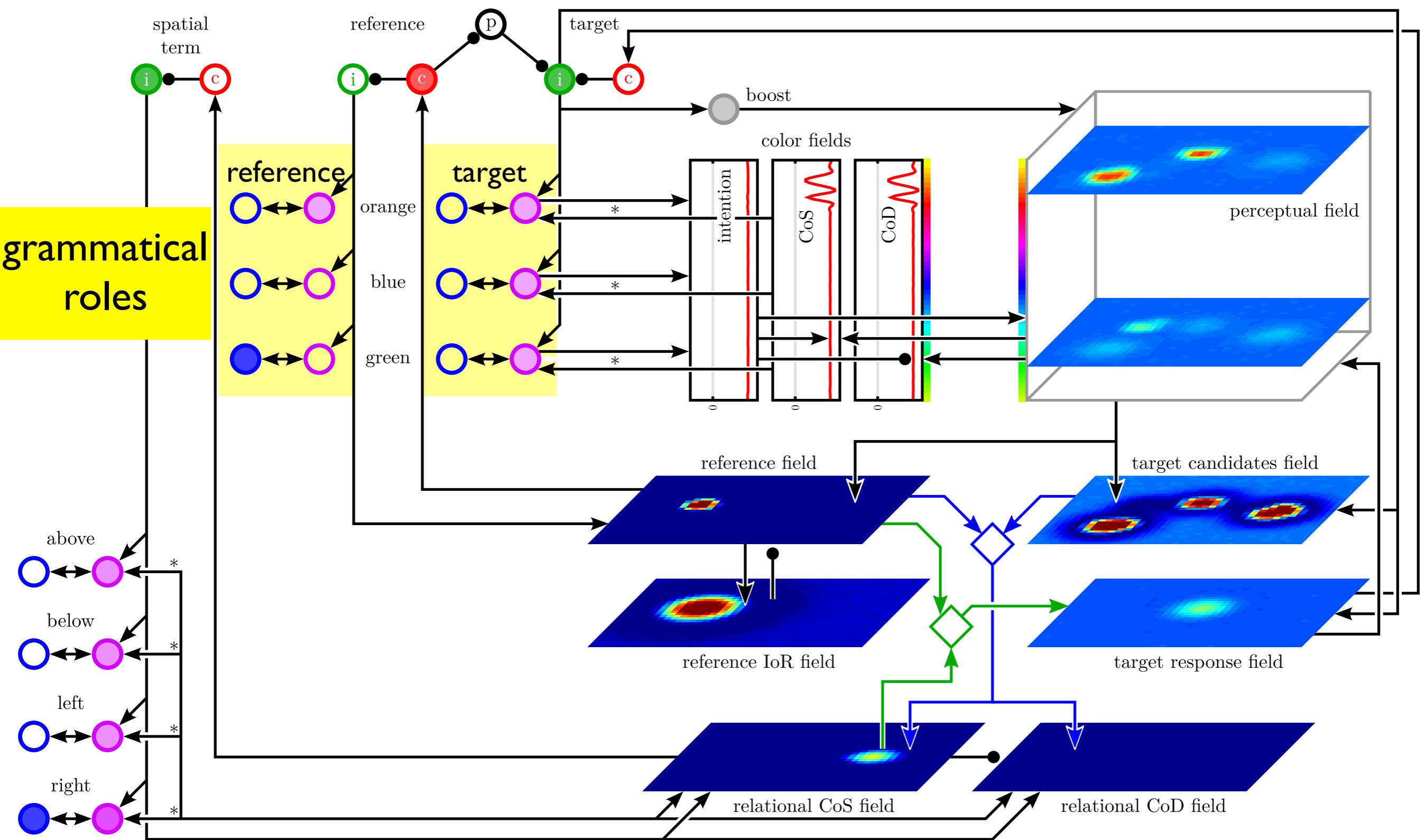
A DFT architecture that does both grounding and describing

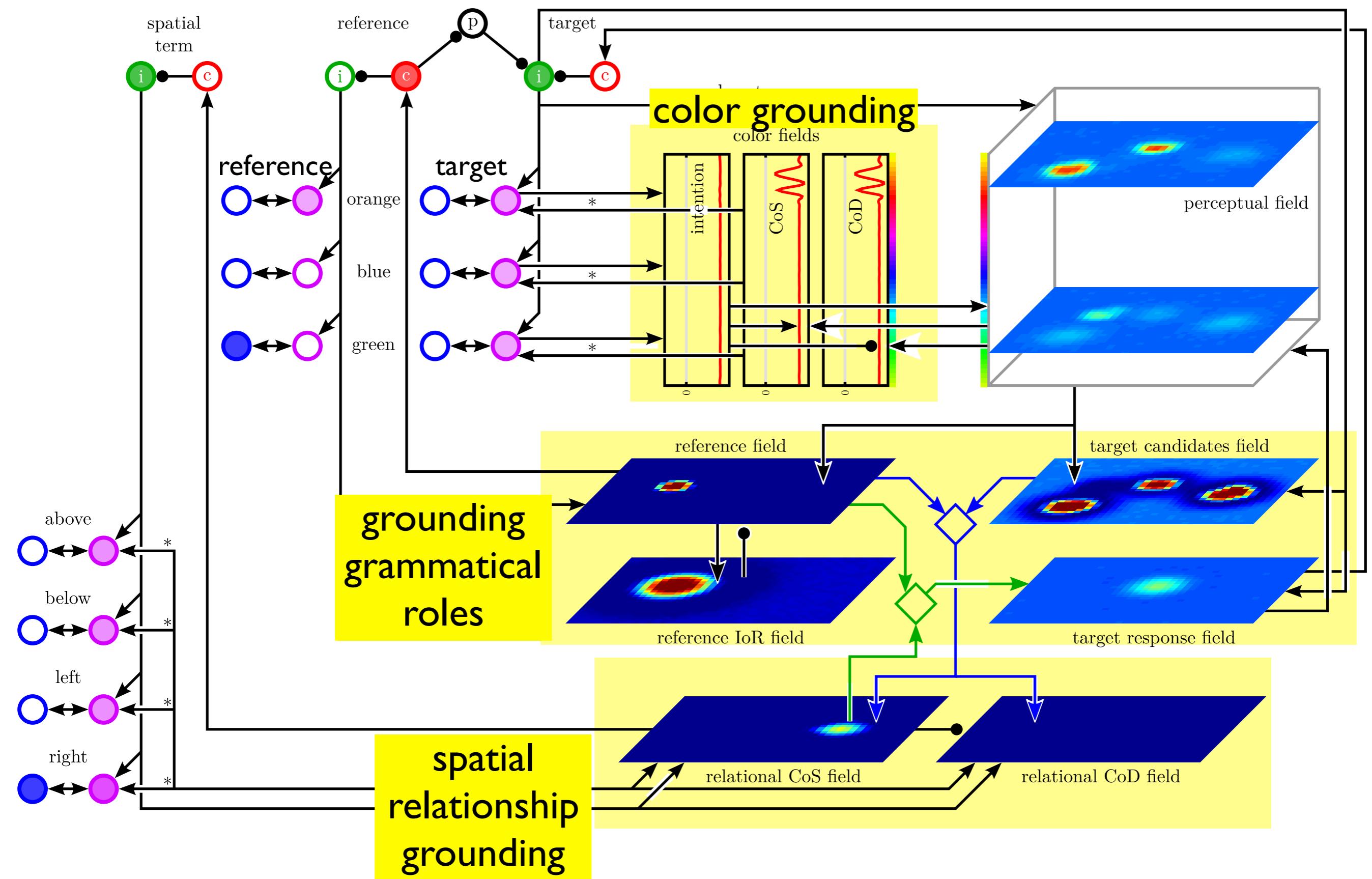




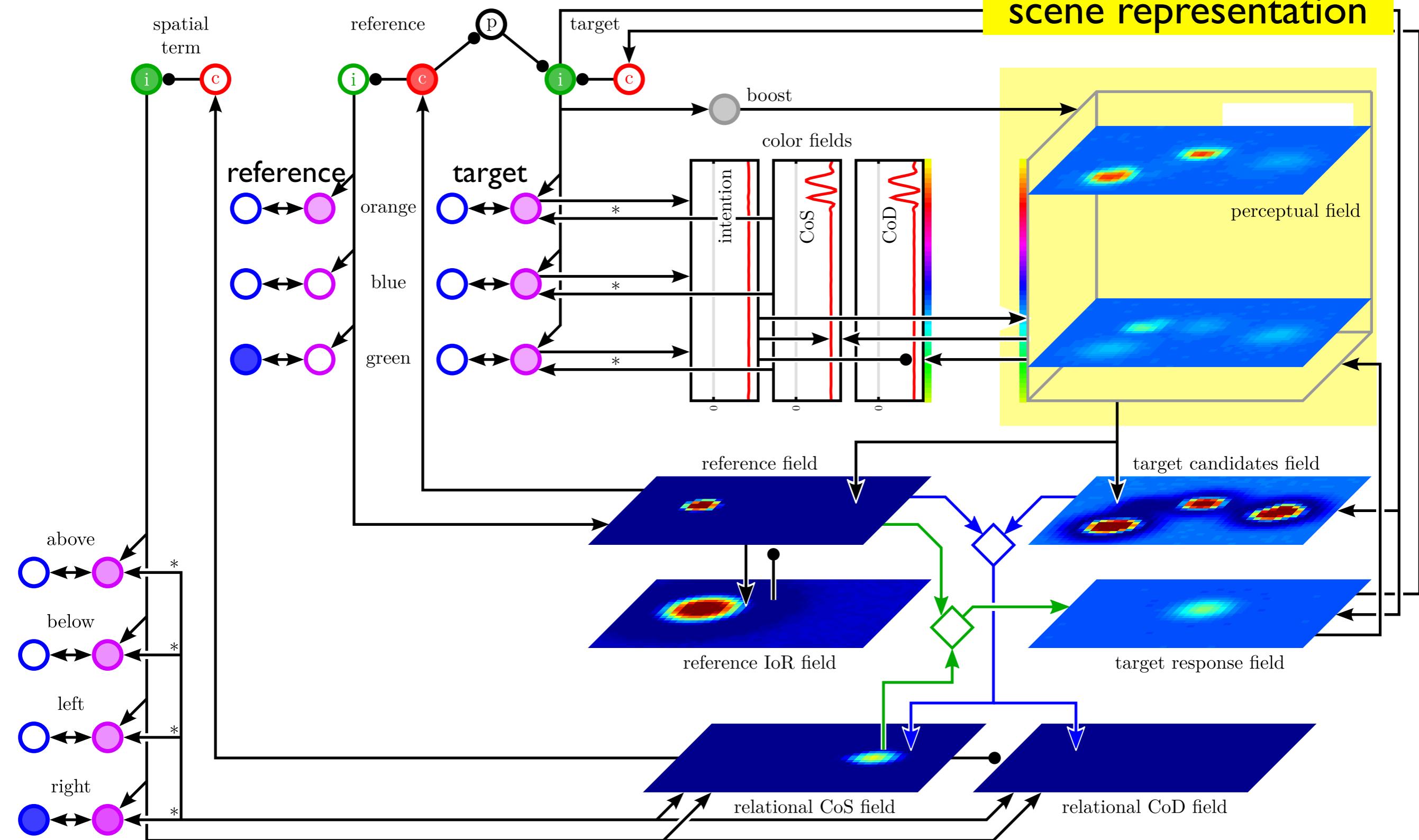
color concepts







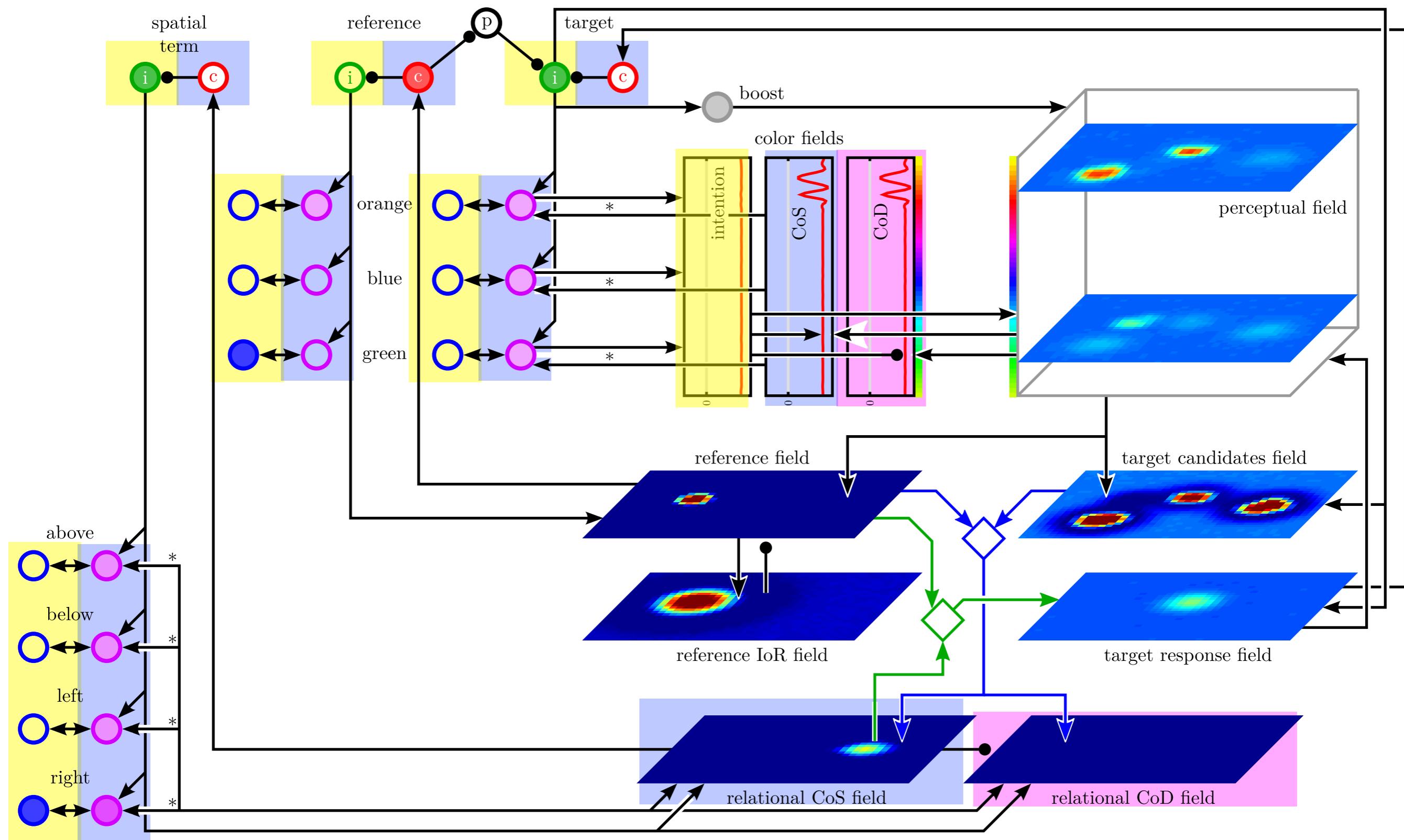
scene representation



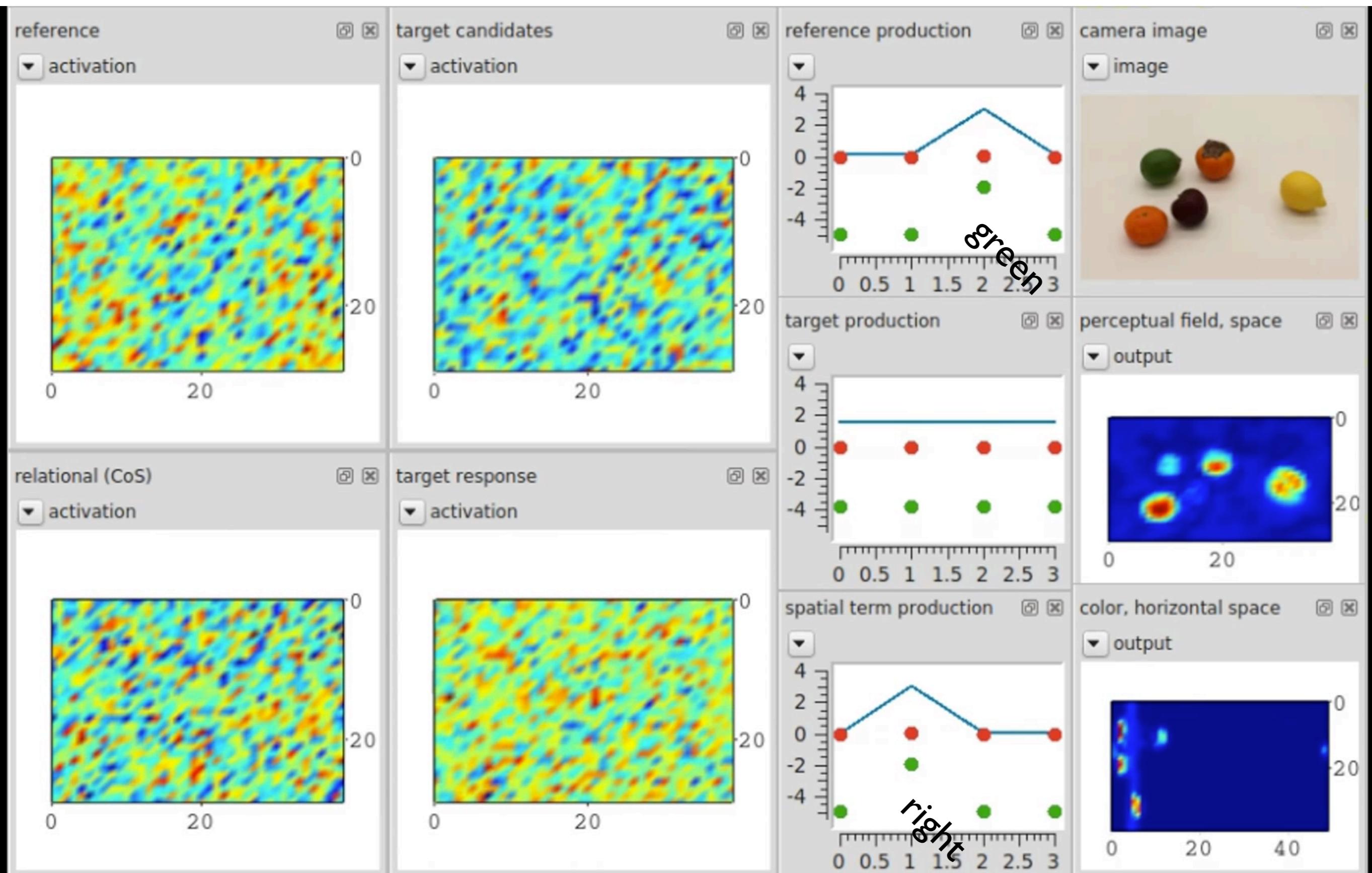
Intention

Condition of Satisfaction

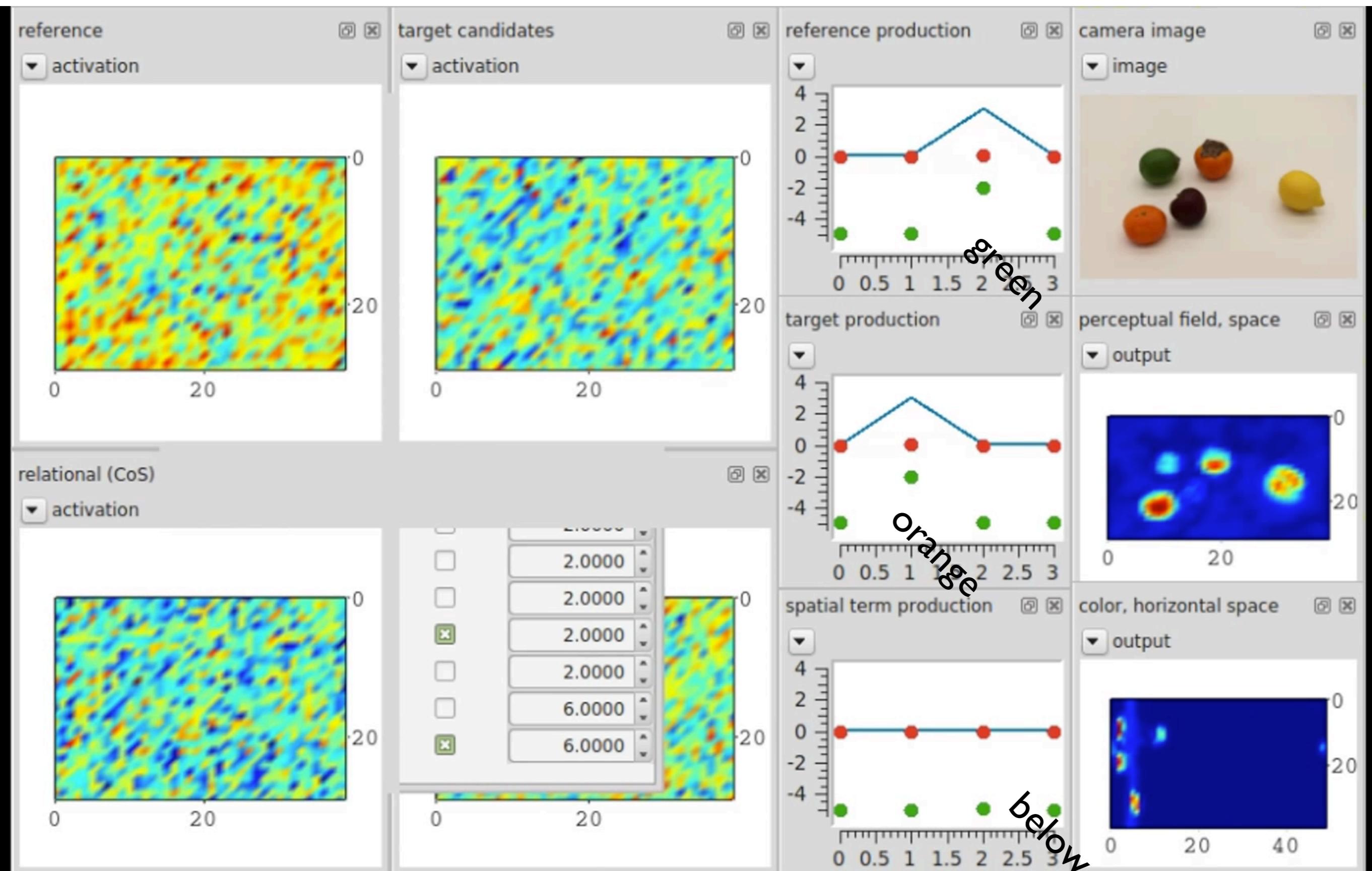
Condition of Dissatisfaction



what is to the right of the green object?



■ where is the orange relative to the green object

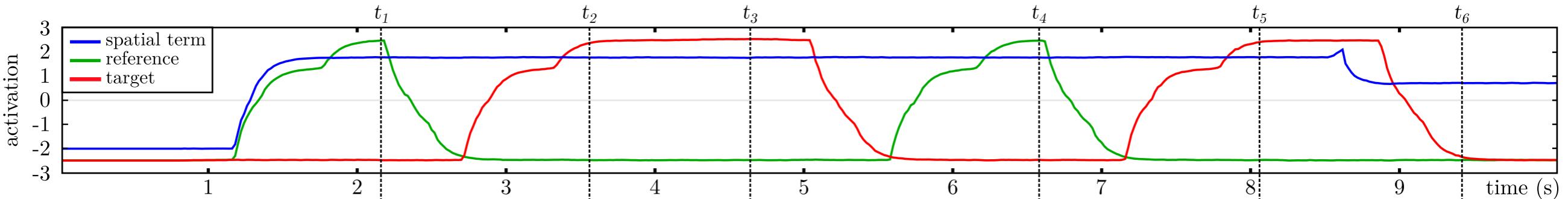


Autonomous hypothesis testing

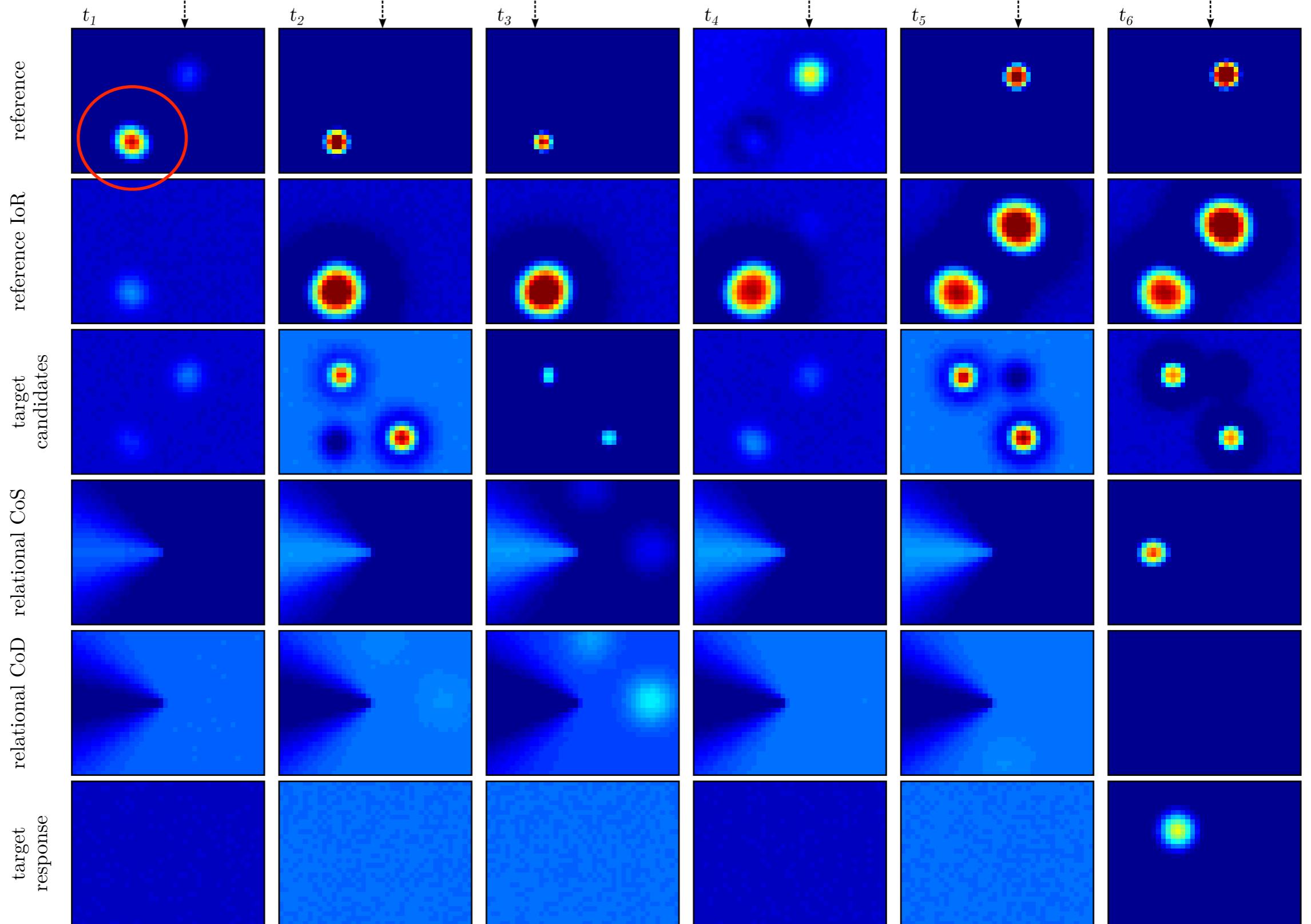


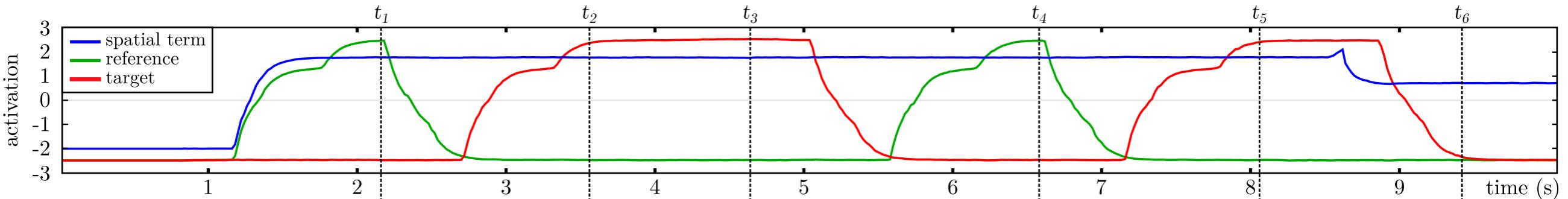
“the red cup that is to the
left of the green cup”

[Richter, Lins et al, CogSci 2014]

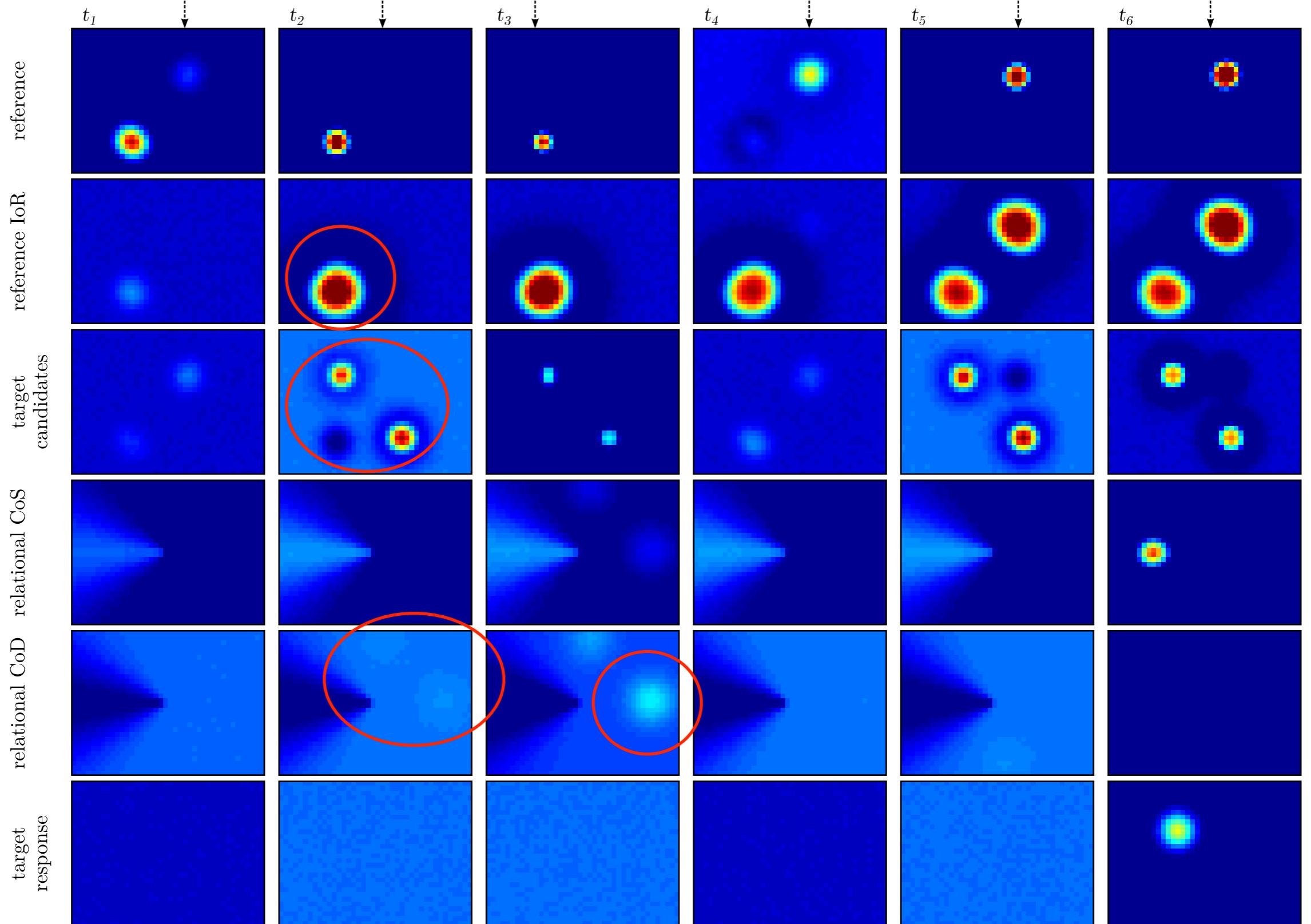


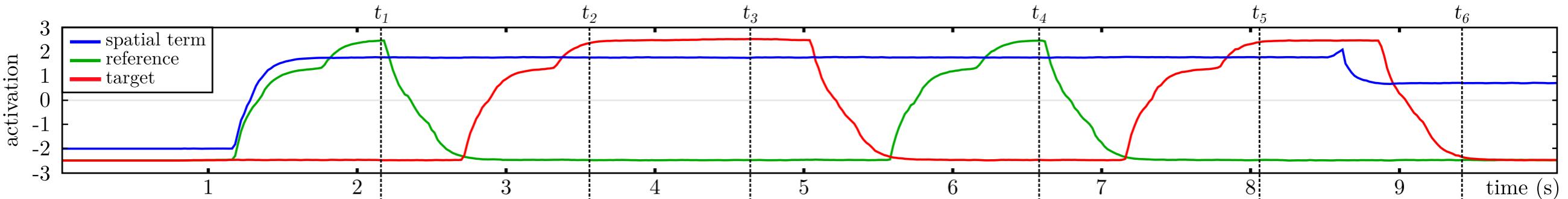
“the red cup
that is to the
left of the
green cup”



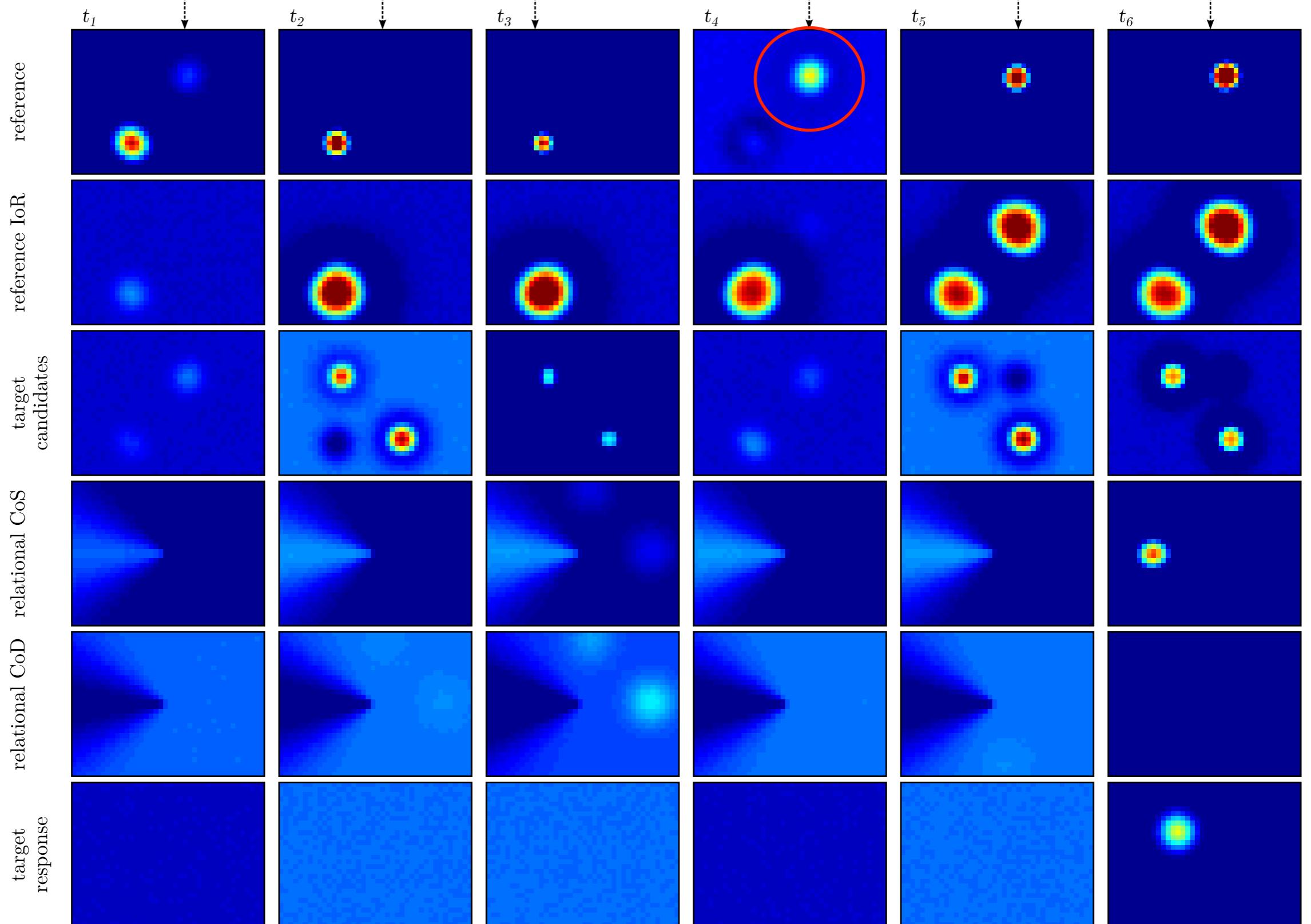


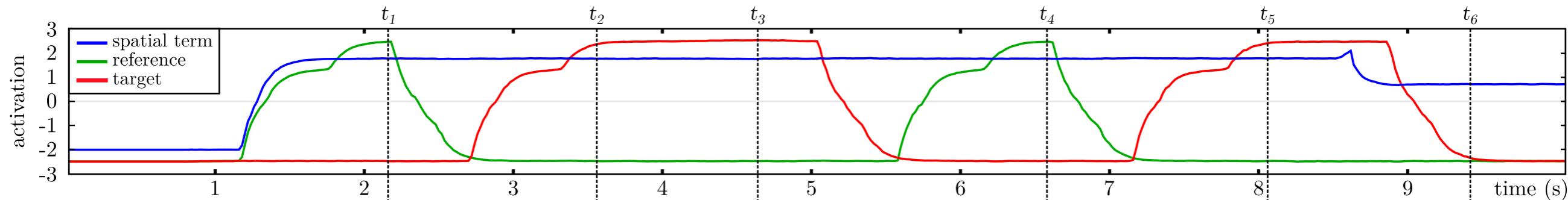
“the red cup
that is to the
left of the
green cup”



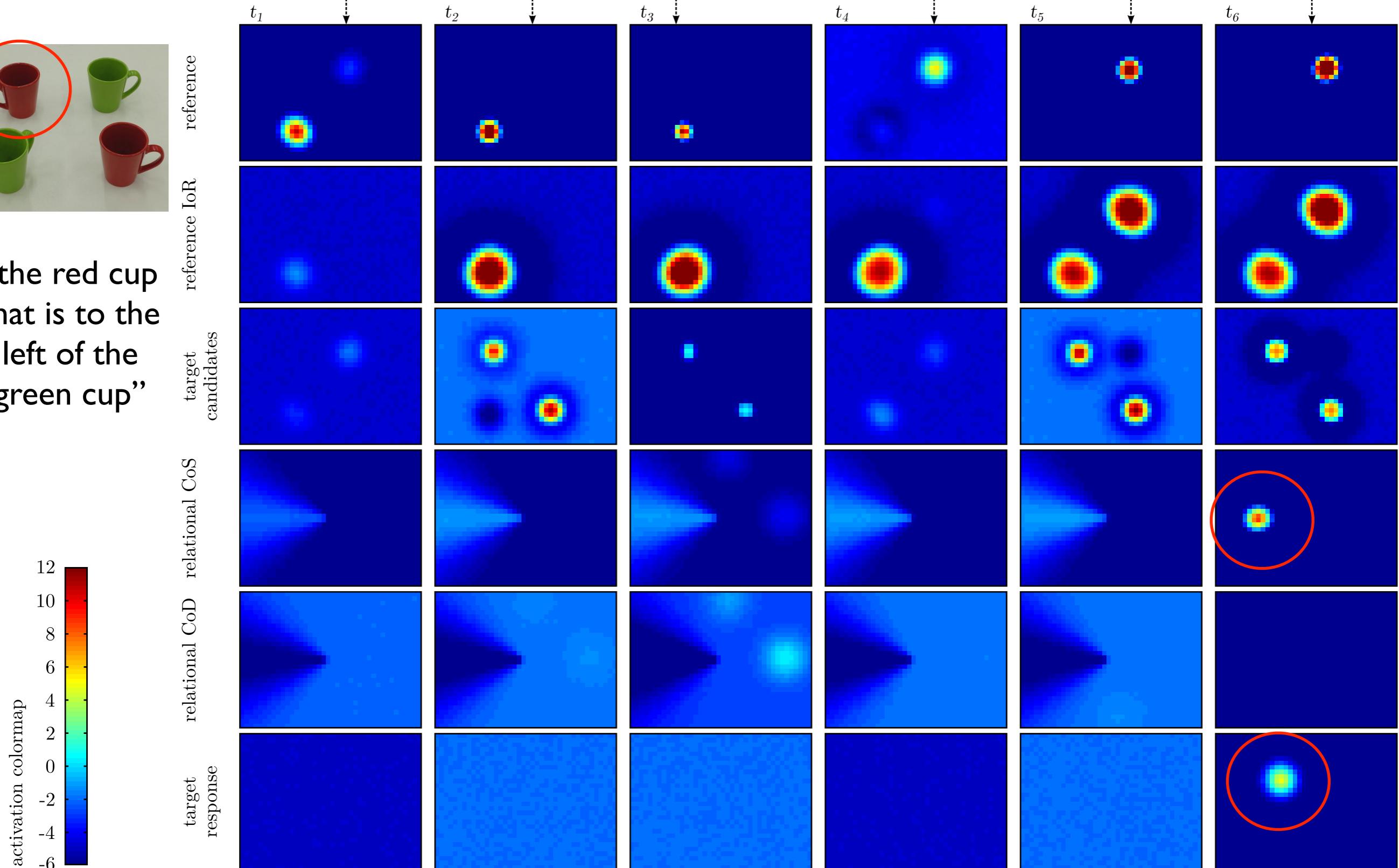


“the red cup
that is to the
left of the
green cup”

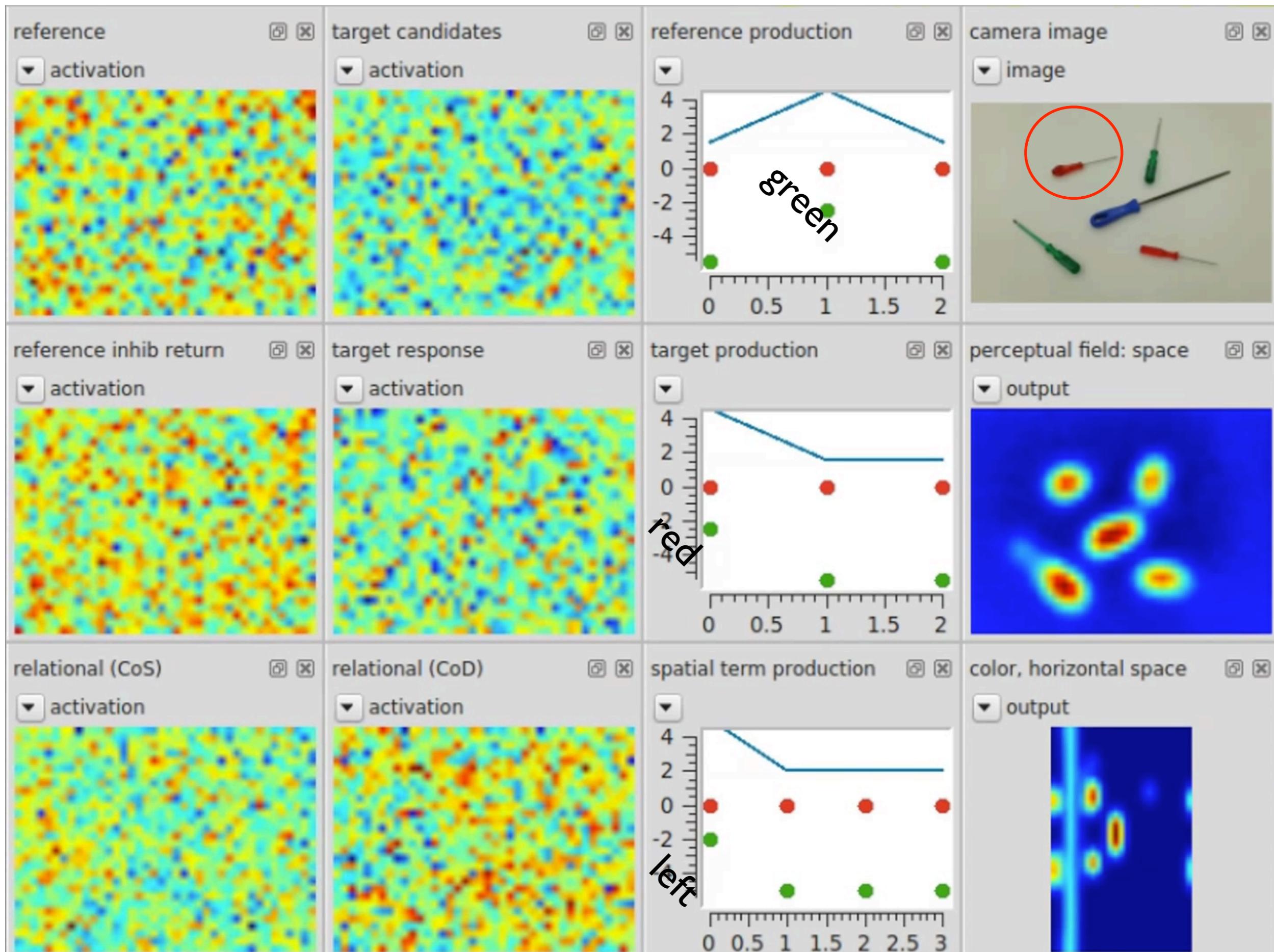




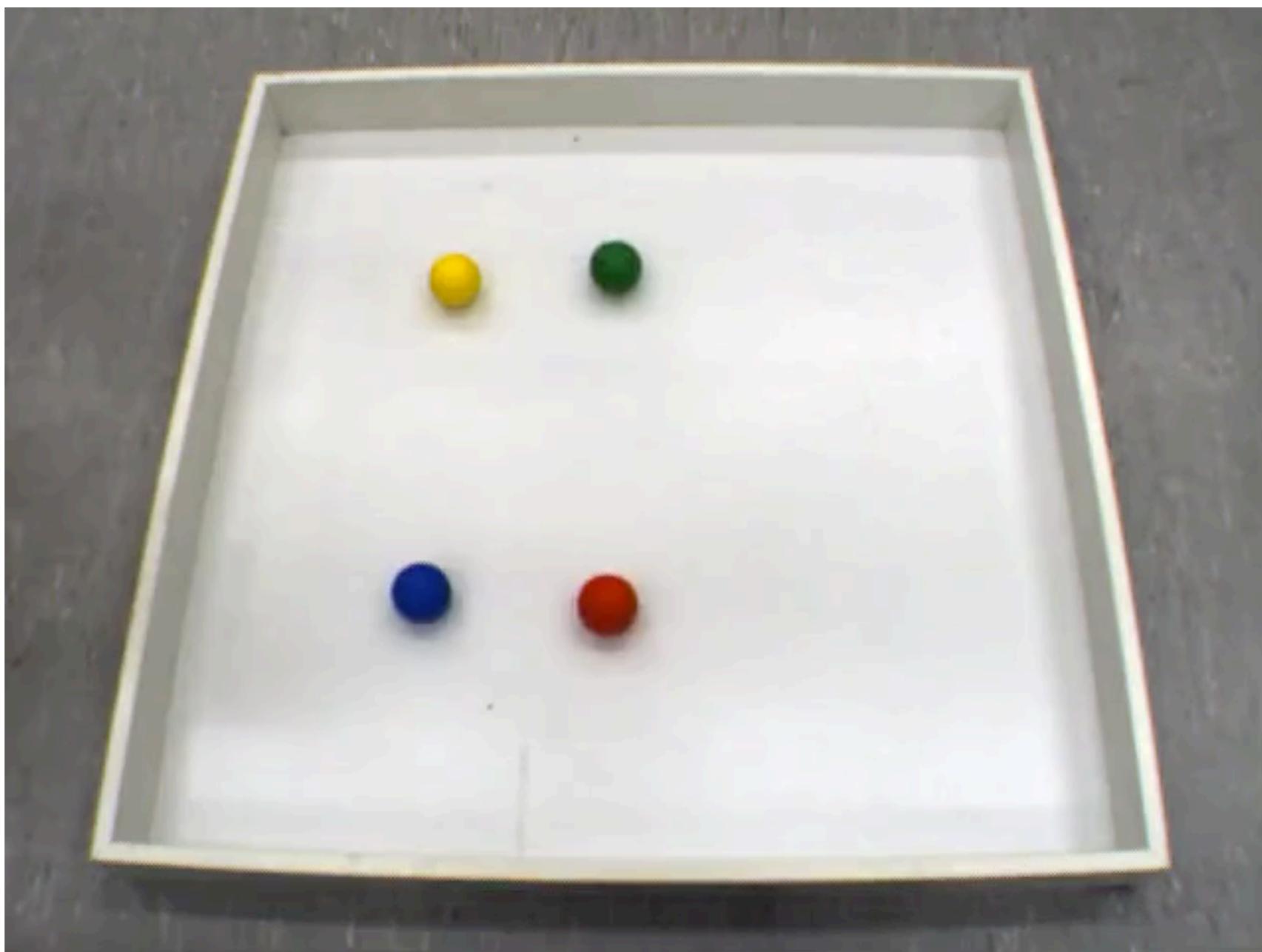
“the red cup
that is to the
left of the
green cup”



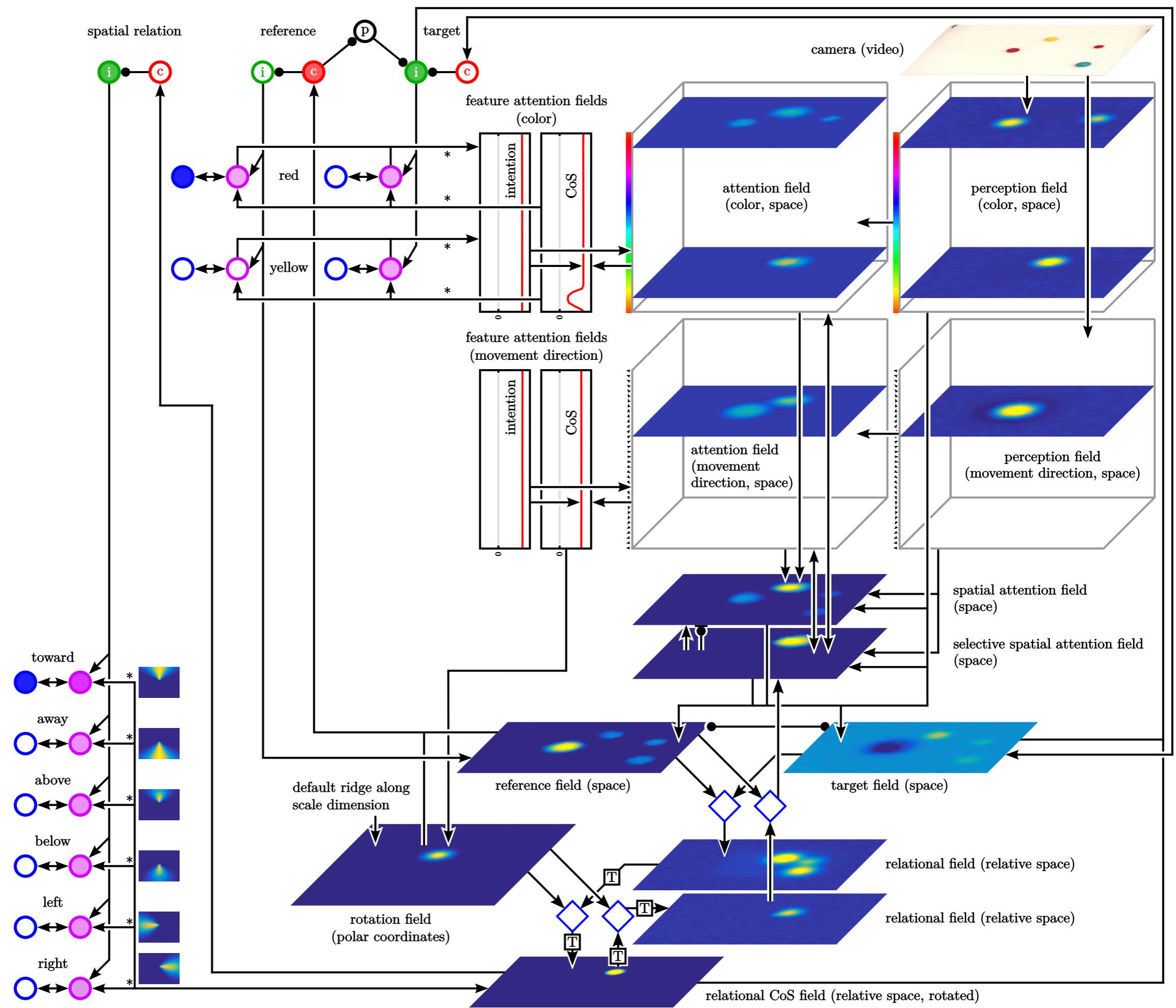
“the red to the left of the green”

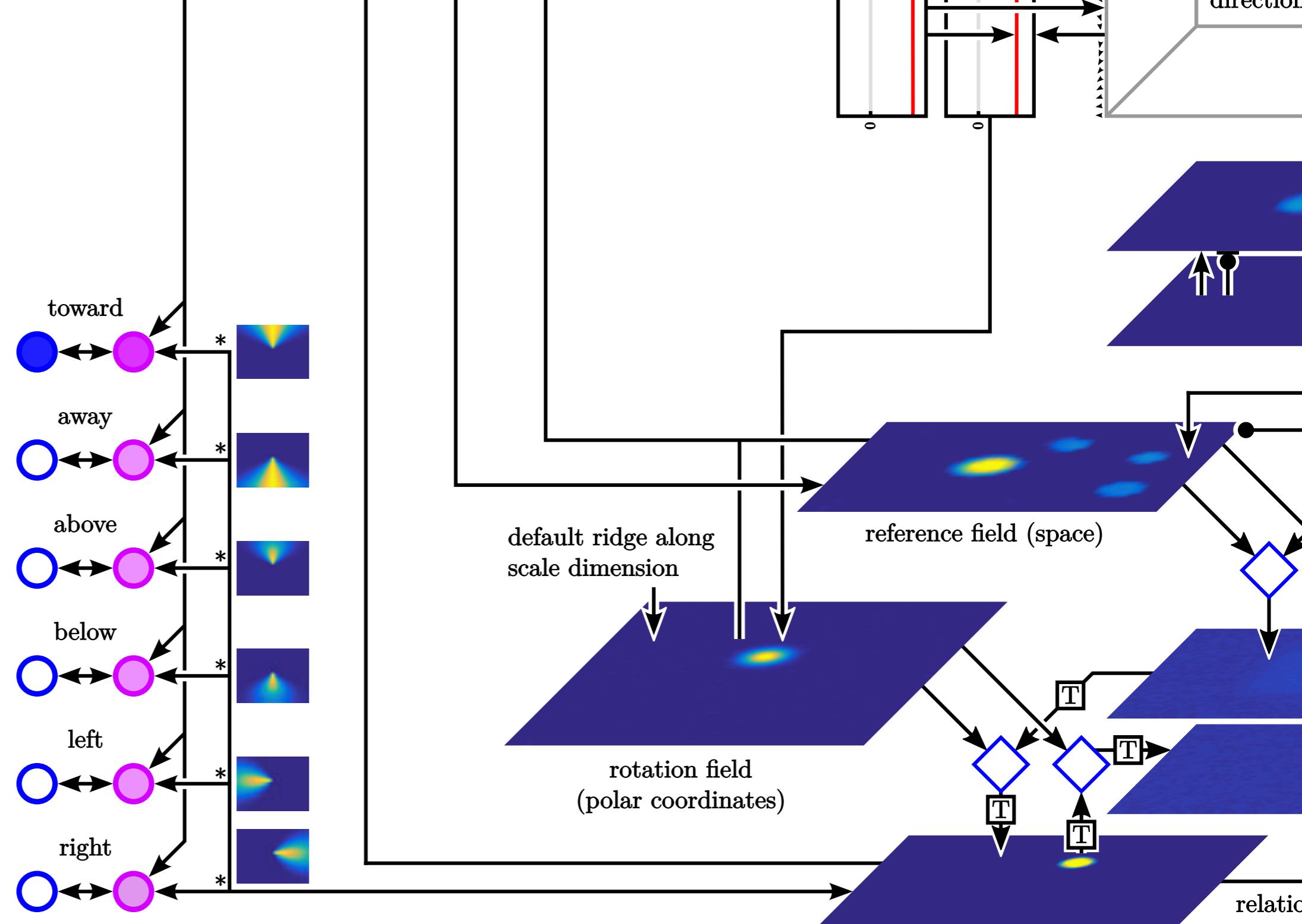


Grounding movement relations



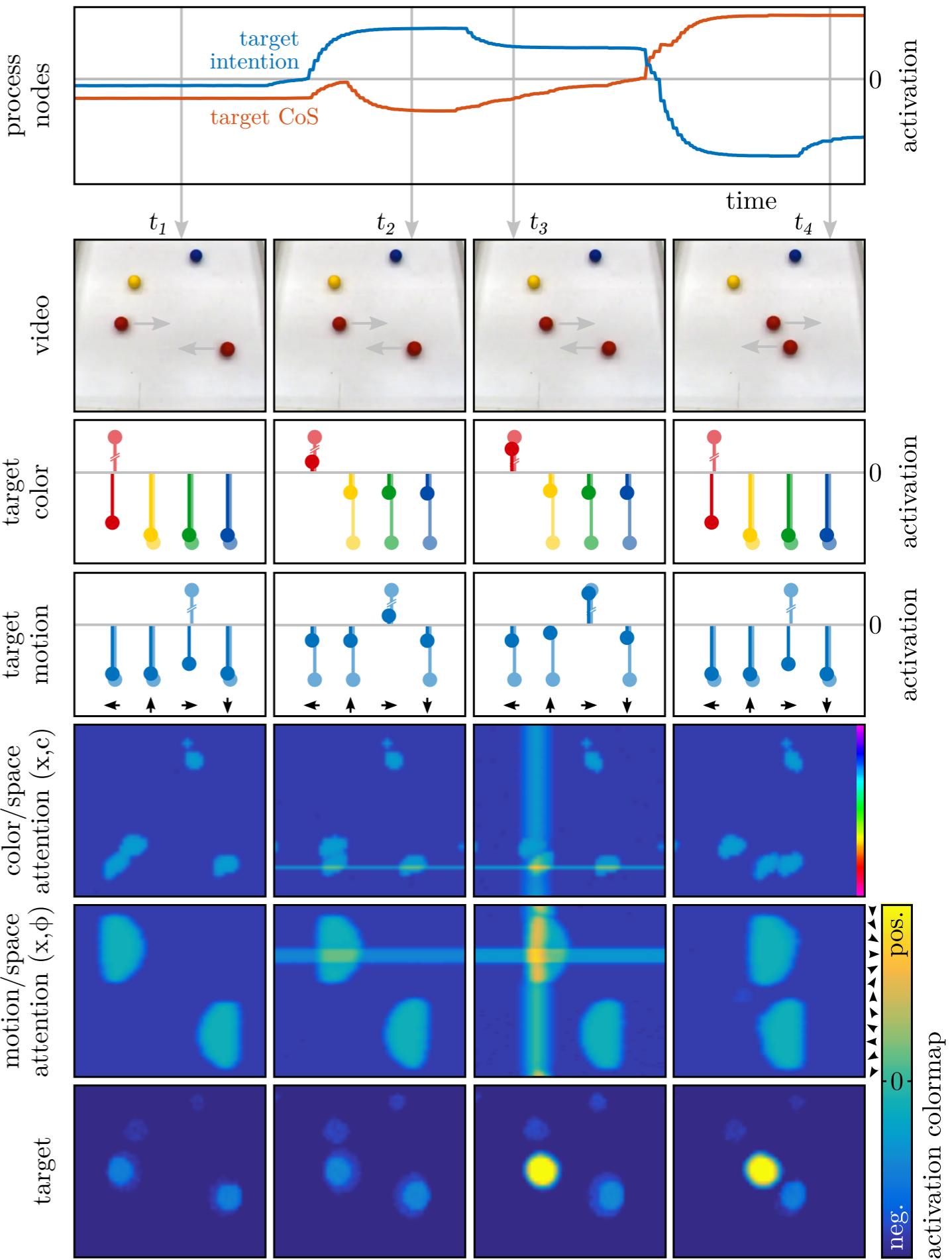
[Richter,
Lins,
Schöner,
ToPiC
(2017)]





“the red
moving to the
right”

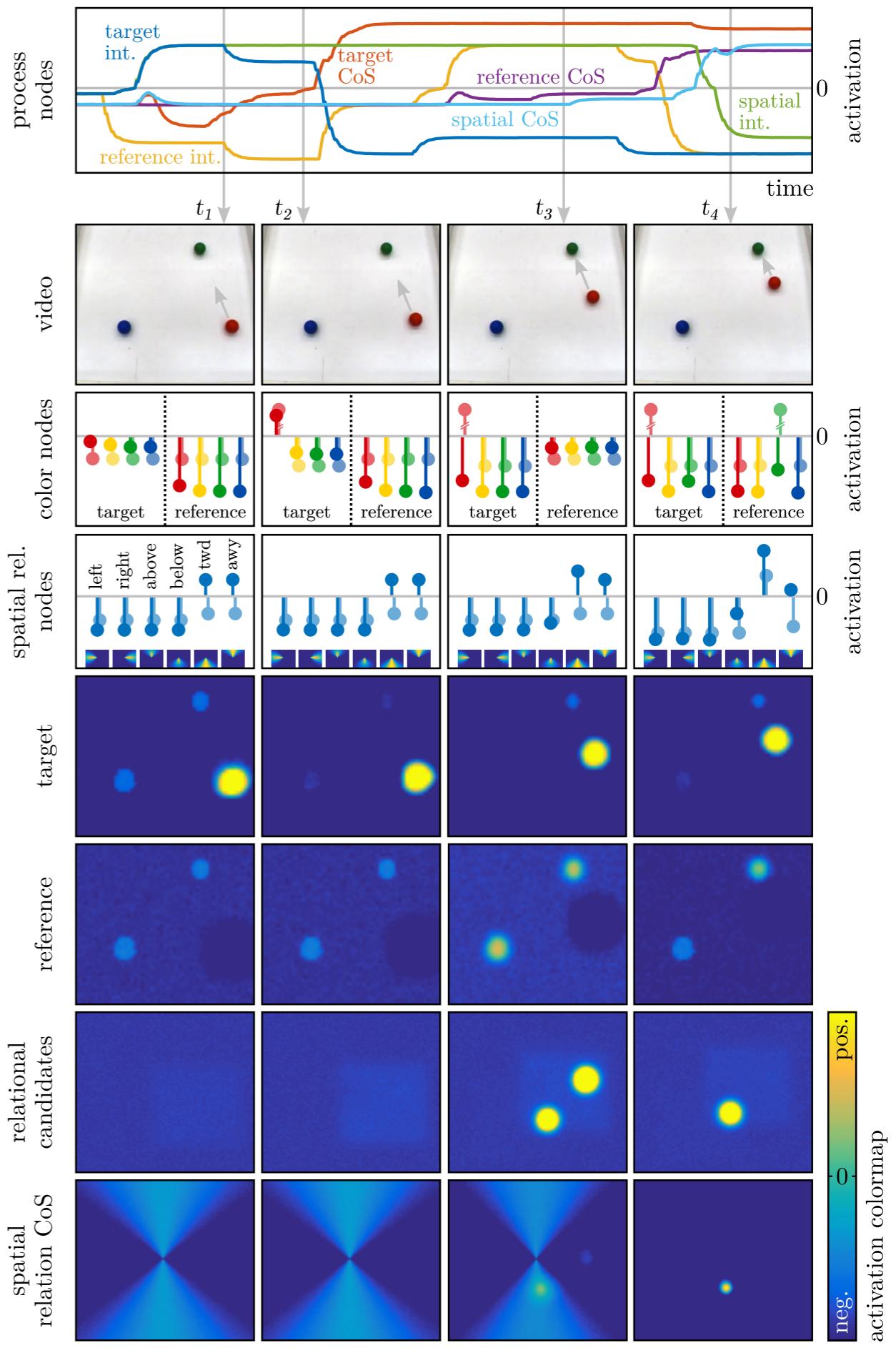
perceptual
grounding



[Richter et al]

description

[Richter et al]



Conclusion

- ... higher dimensional fields
- arranged in architectures...
- deliver higher cognitive functions
- such as perceptual grounding, and describing scenes