

# Perceptual grounding of concepts

Gregor Schöner  
INI RUB

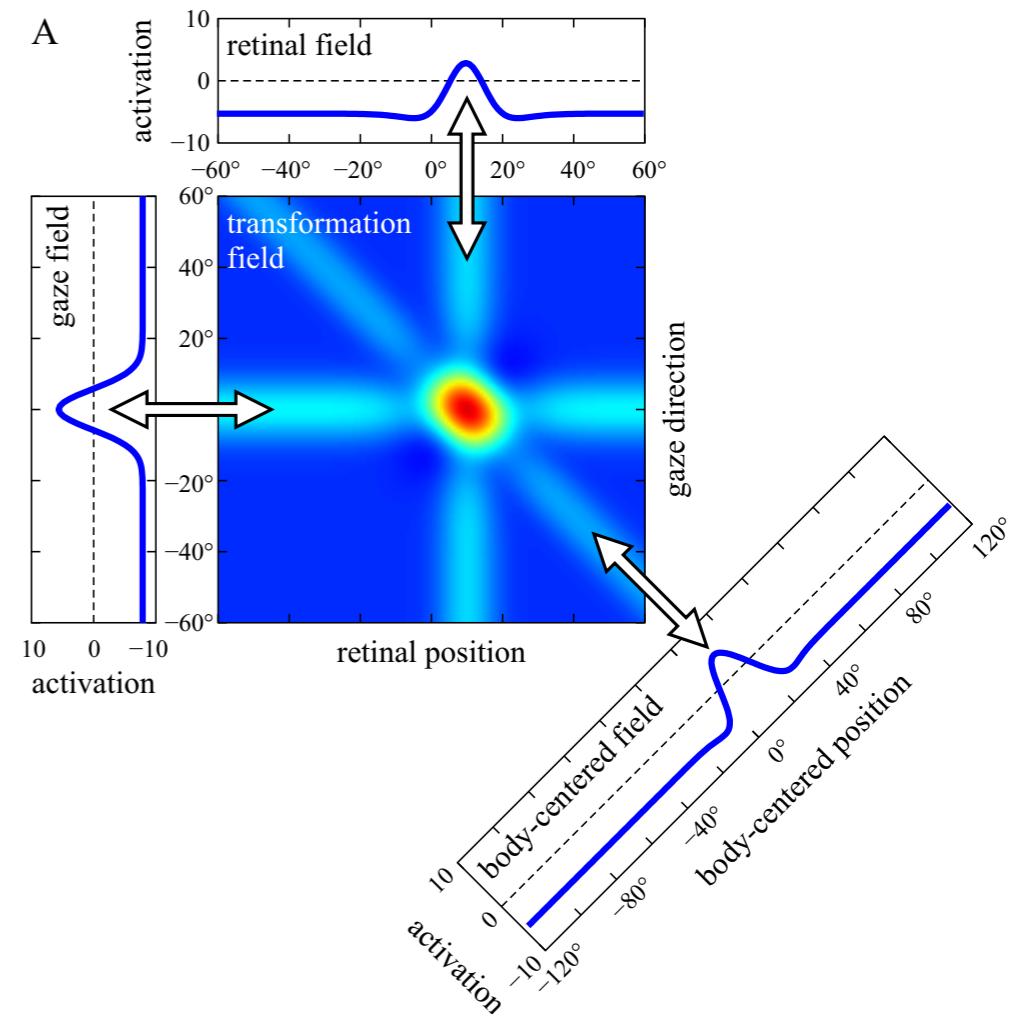
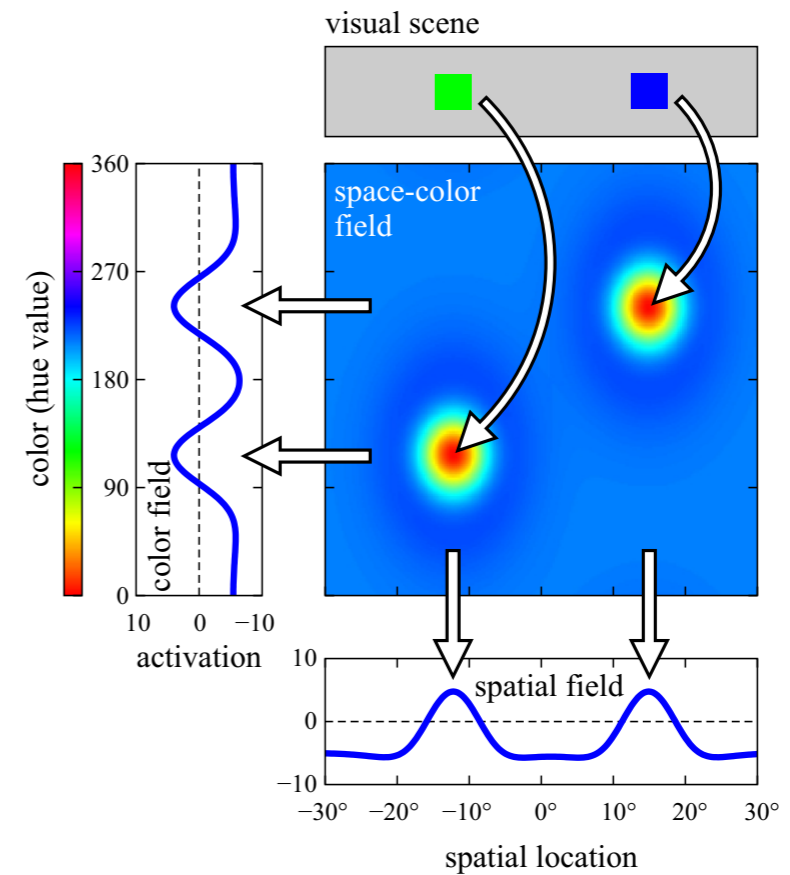
[gregor.schoener@ini.rub.de](mailto:gregor.schoener@ini.rub.de)

# Previous lecture

■ binding through space

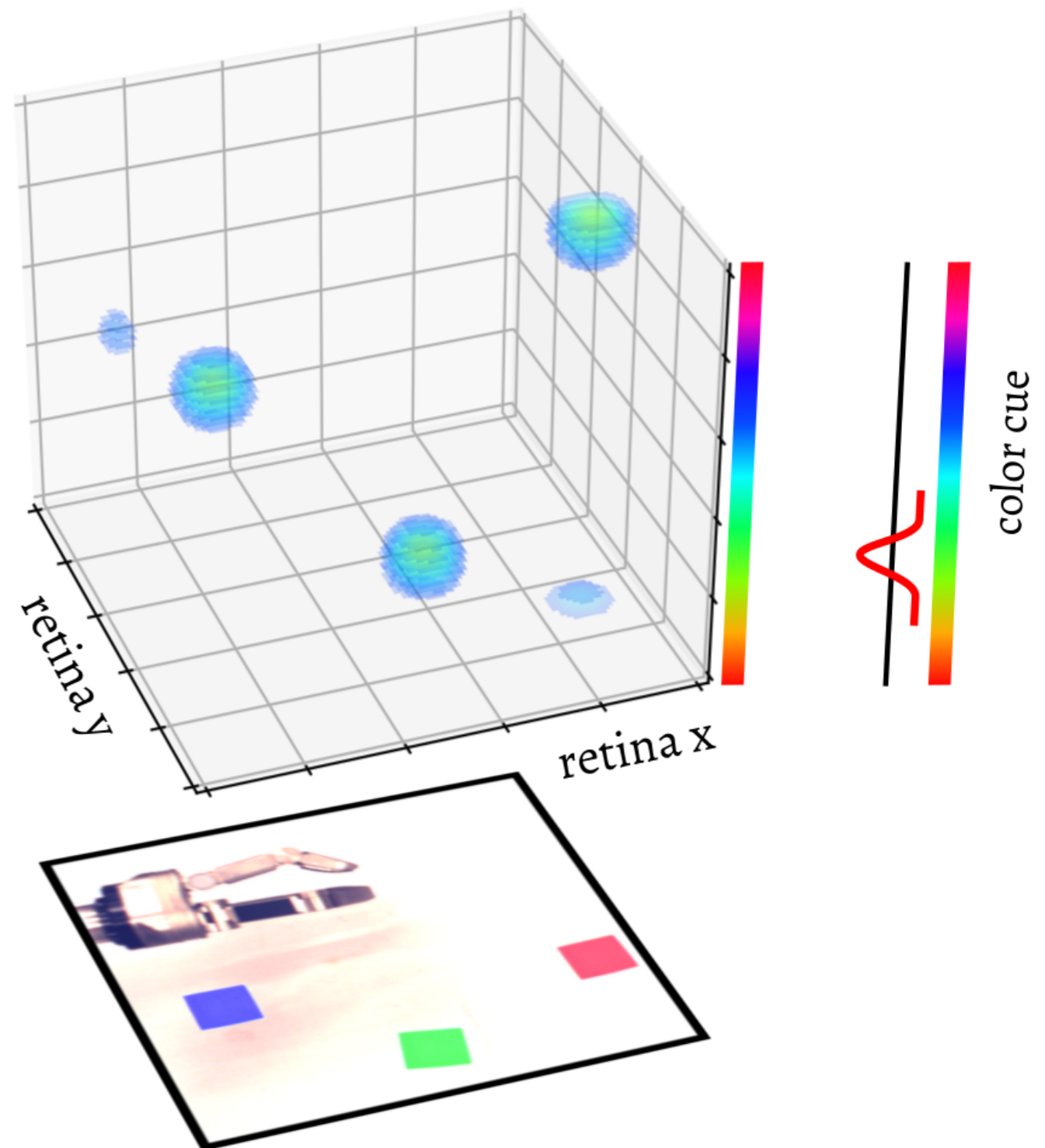
■ visual search... attentional selection

■ coordinate transforms...



# attentional selection

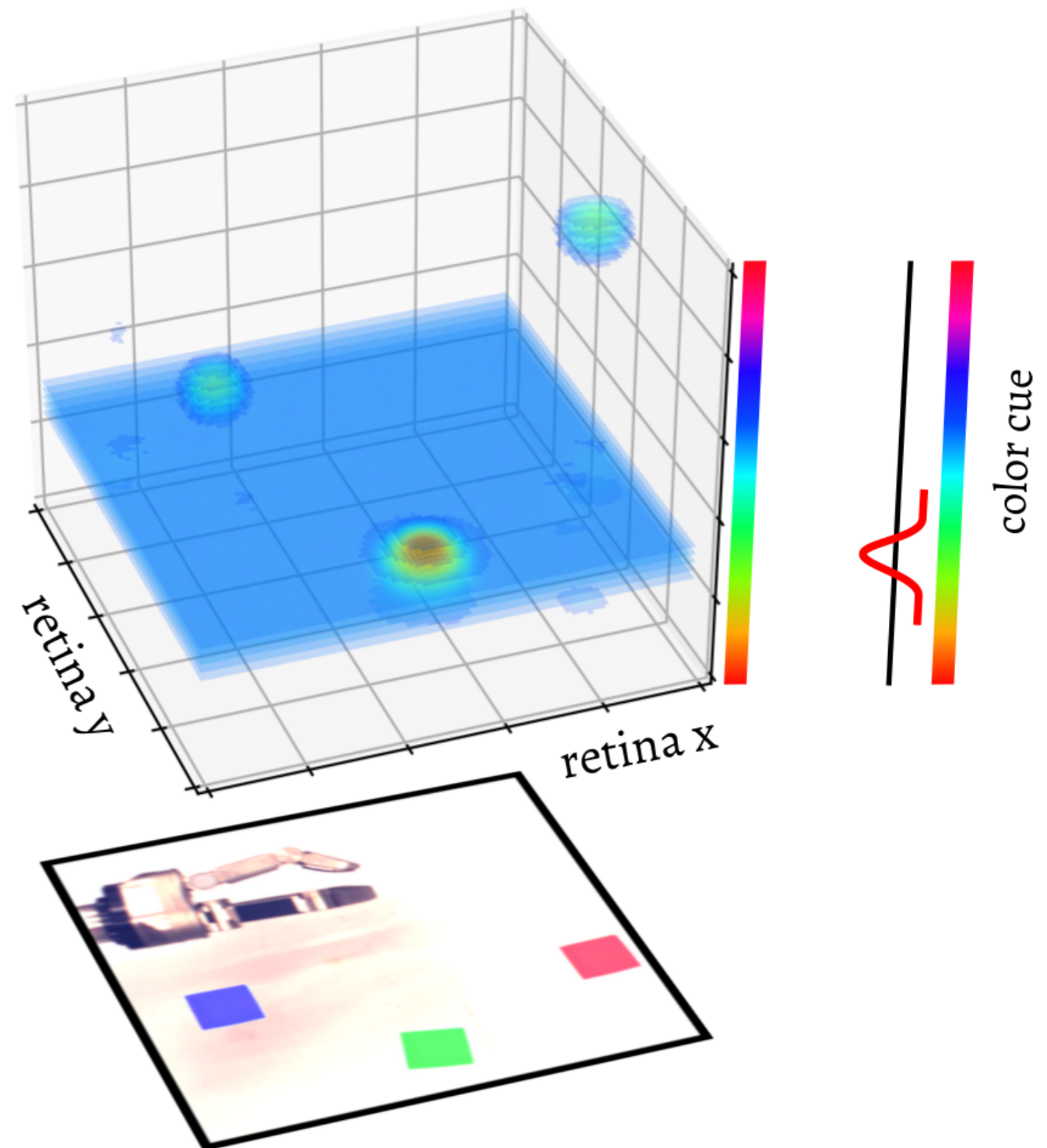
- to bring a cued object into the attentional foreground



[Tekülve, 2020]

# attentional selection

- to bring a cued object into the attentional foreground

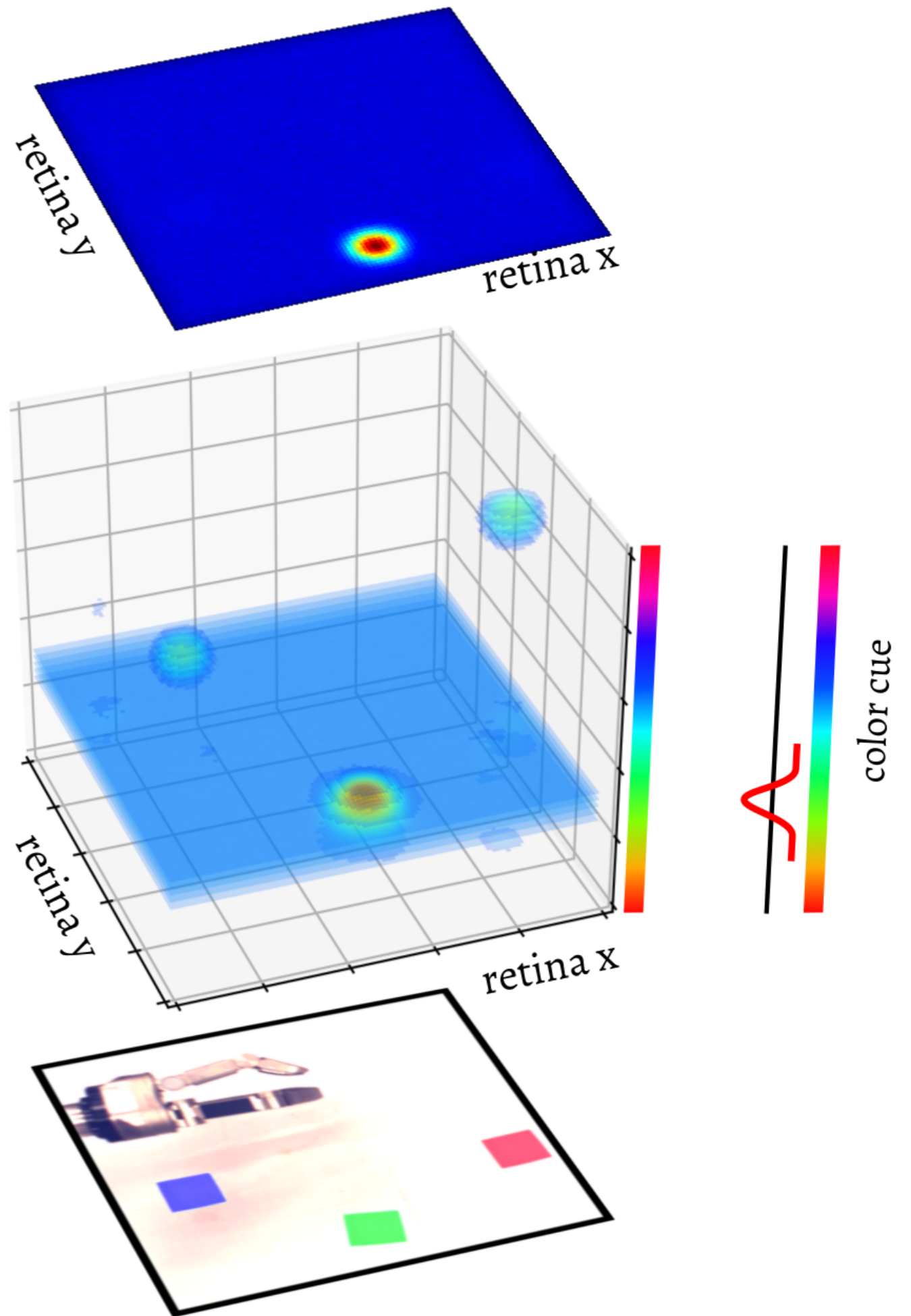


[Tekülve, 2020]



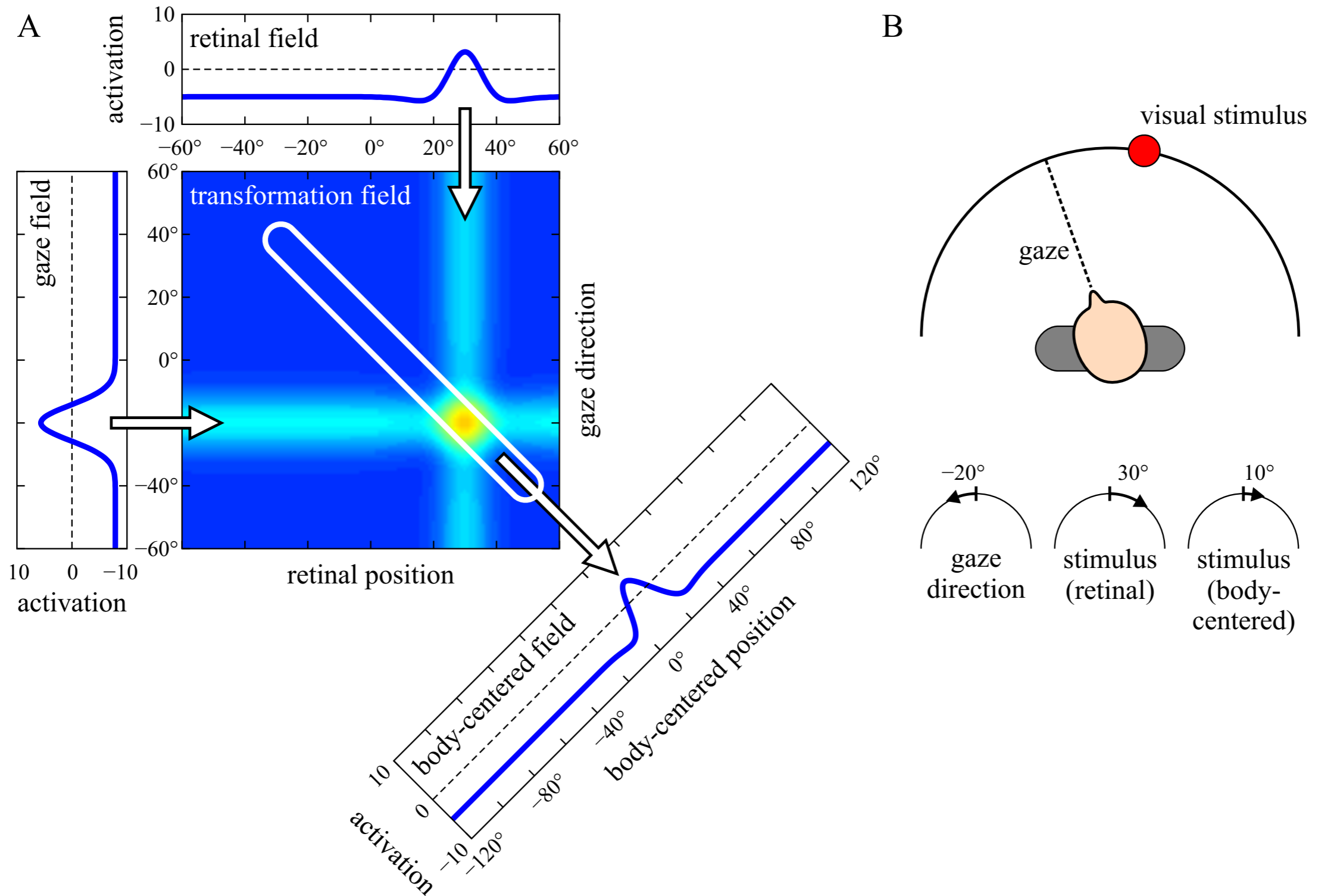
# attentional selection

- to bring a cued object into the attentional foreground



[Tekülve, 2020]

# Coordinate transforms



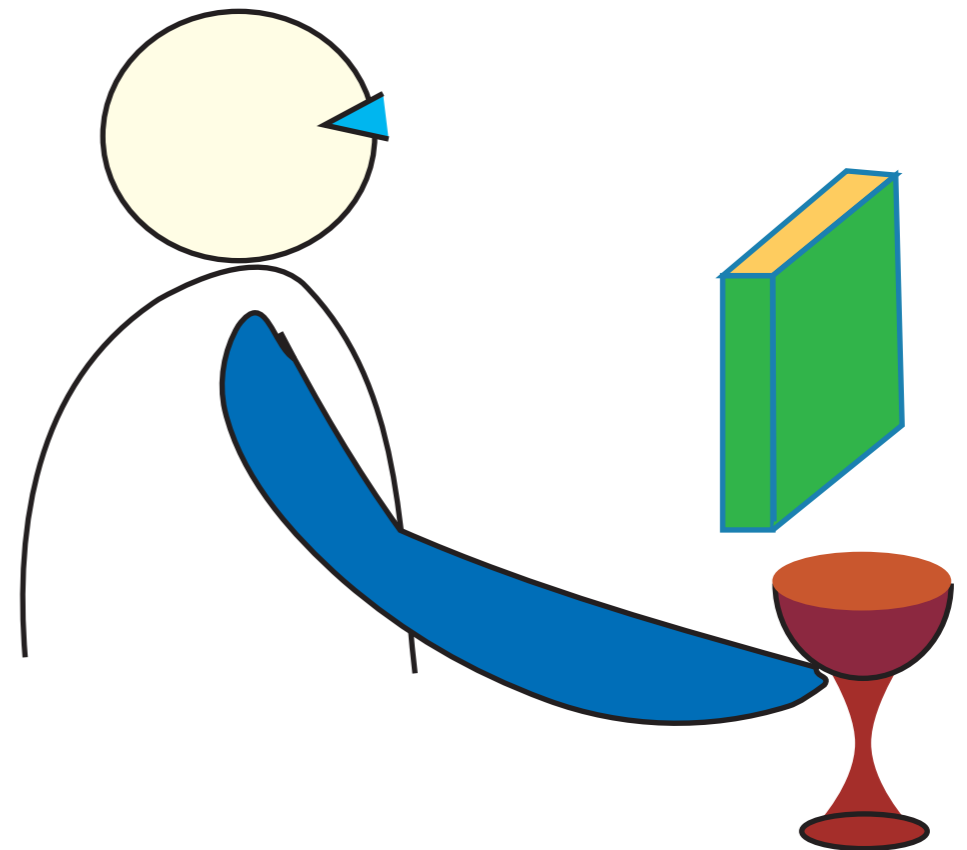
[Schneegans Ch 7 of *DFT Primer*, 2016]

# Today

- these functions play an important role in lifting neural dynamics to higher cognition
- perceptual grounding of concepts/language/relations

# Perceptually grounding language

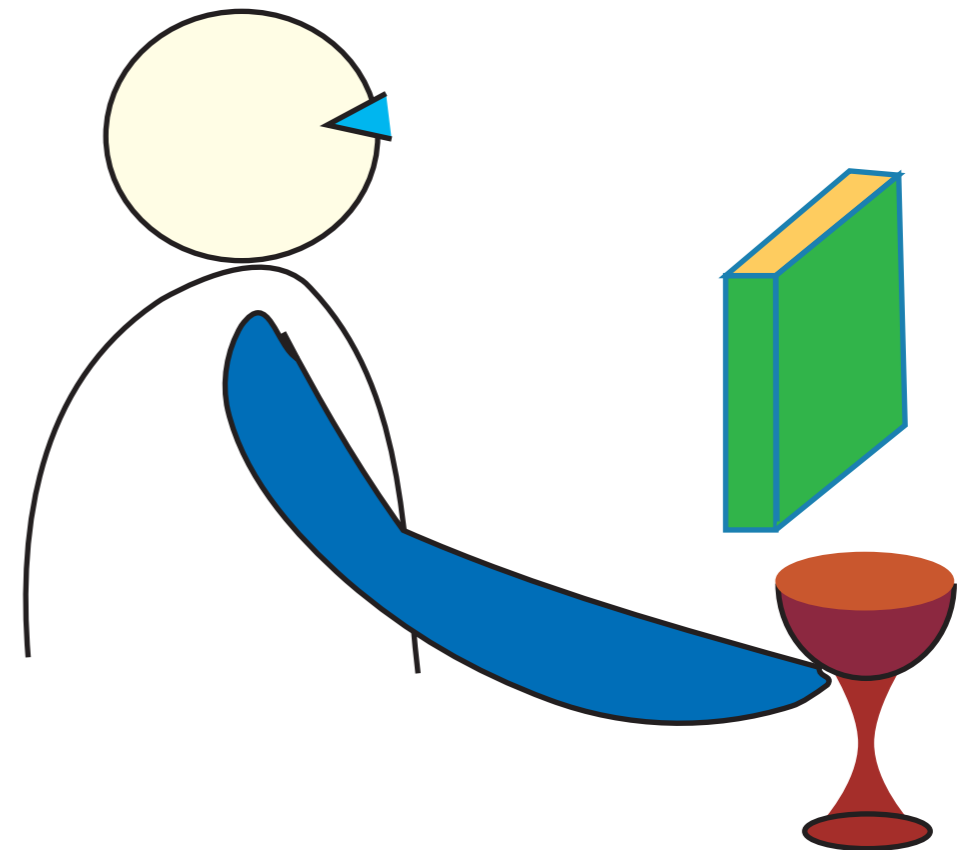
- human communication, in its simplest form, is about objects or events that are perceivable or reachable by action in a shared environment
- e.g., this cup is brown



# Perceptually grounding language

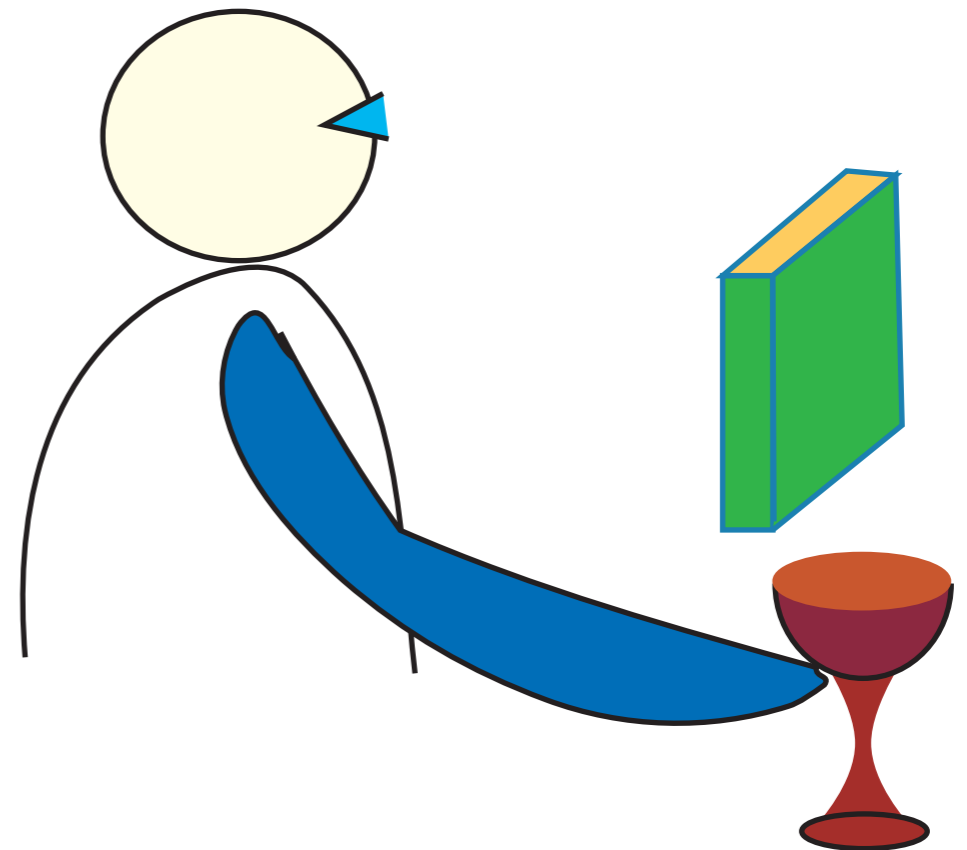
- language has been perceptually grounded when the listener directs attention at the same object that the word “cup” refers to...
- this typically will be “joint attention”, in which the speaker also attends to the object

- the term “perceptual grounding” is not universally used...
- also called “targetting” (Talmy)
- or “referring”



# Perceptually grounding language

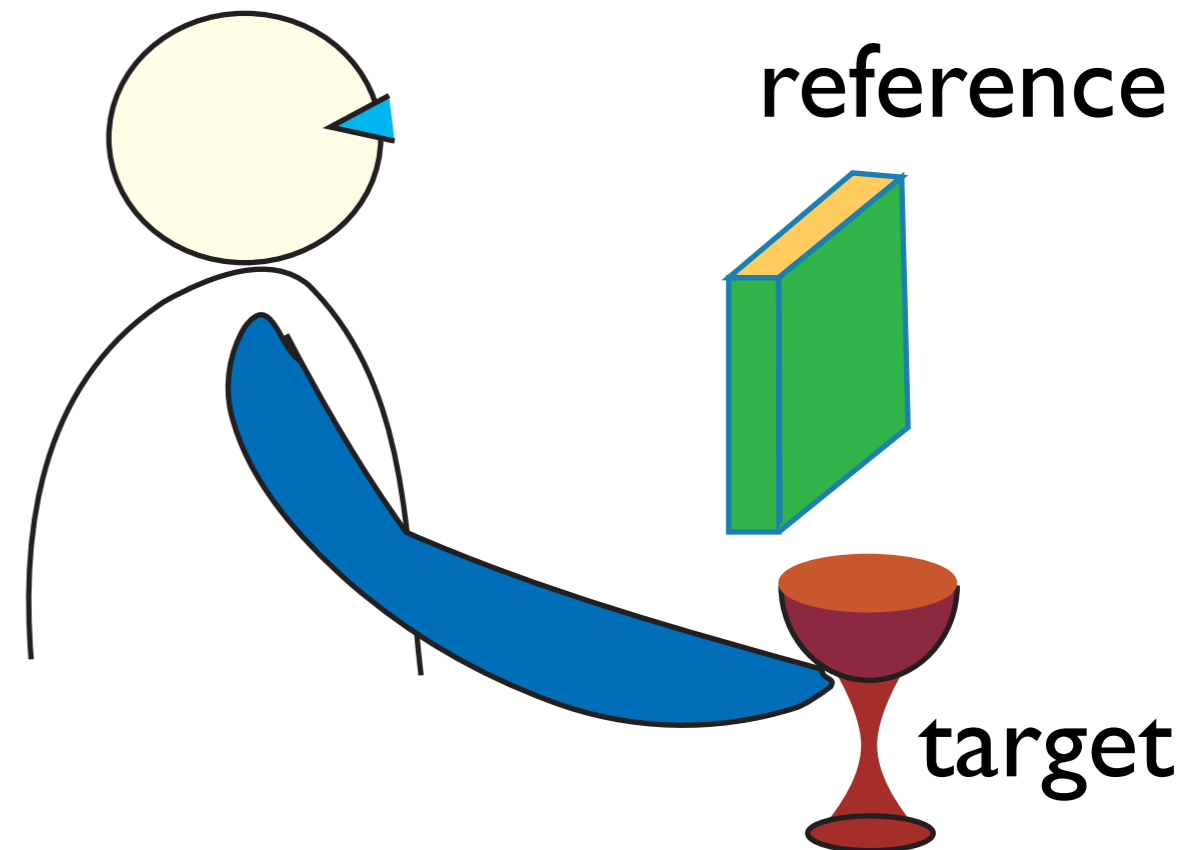
- Grounding process could be and often is mediated by forms of communication other than language
- e.g., pointing (deictic code)
- e.g. context



# Perceptually grounding language

- grounding may be mediated by **spatial language**
- in which a **target** object is related to a **reference** object
- presupposes that the reference object is also perceptually grounded for speaker and listener

the cup to the right of  
the green book



# Perceptually grounding language vs. describing

- 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

“what is to the right of the green object”





# Spatial relations

- 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 ... I’ll come back to that

# Grounding spatial language

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

# Grounding spatial language

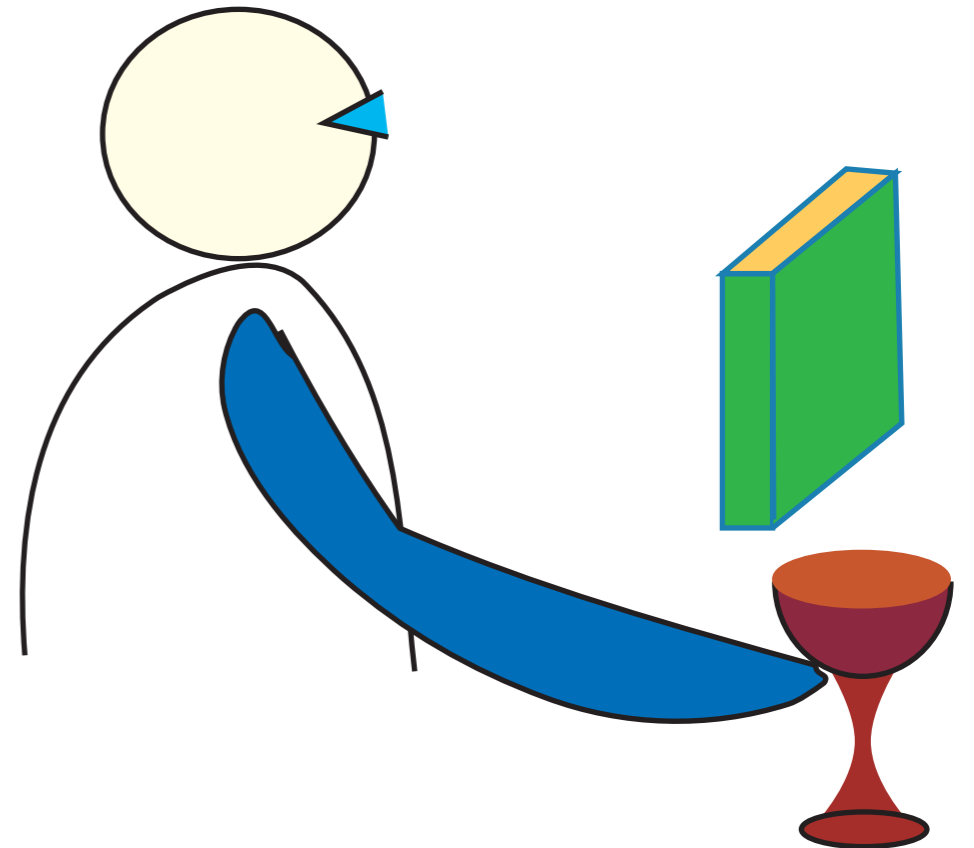
- orientation relative to speaker, position centered in speaker
  - “on my left”
- orientation relative to world/object, position centered in speaker:
  - “north”, “south...” or “leeward”, “windward” ...
- orientation relative to speaker, position centered in object
  - “the cup to the right of the bottle”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

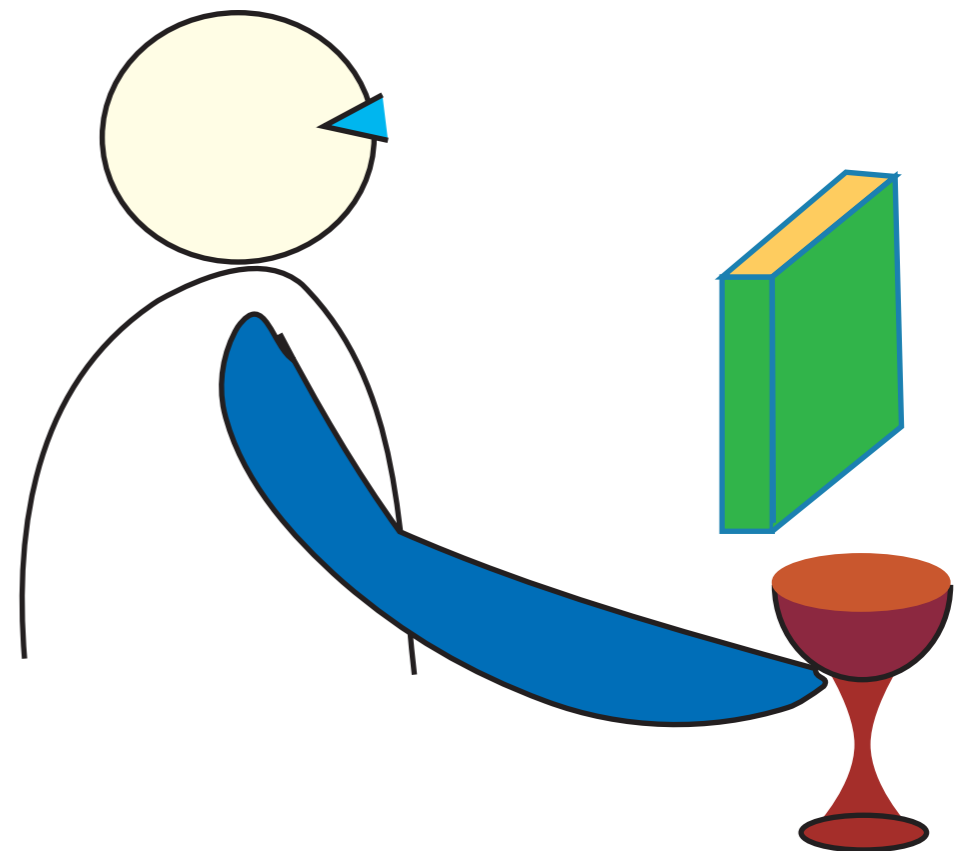
# Grounding spatial language

- 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
- => relational concepts
- reaches a very broad class of mental operations



# Grounding spatial relations

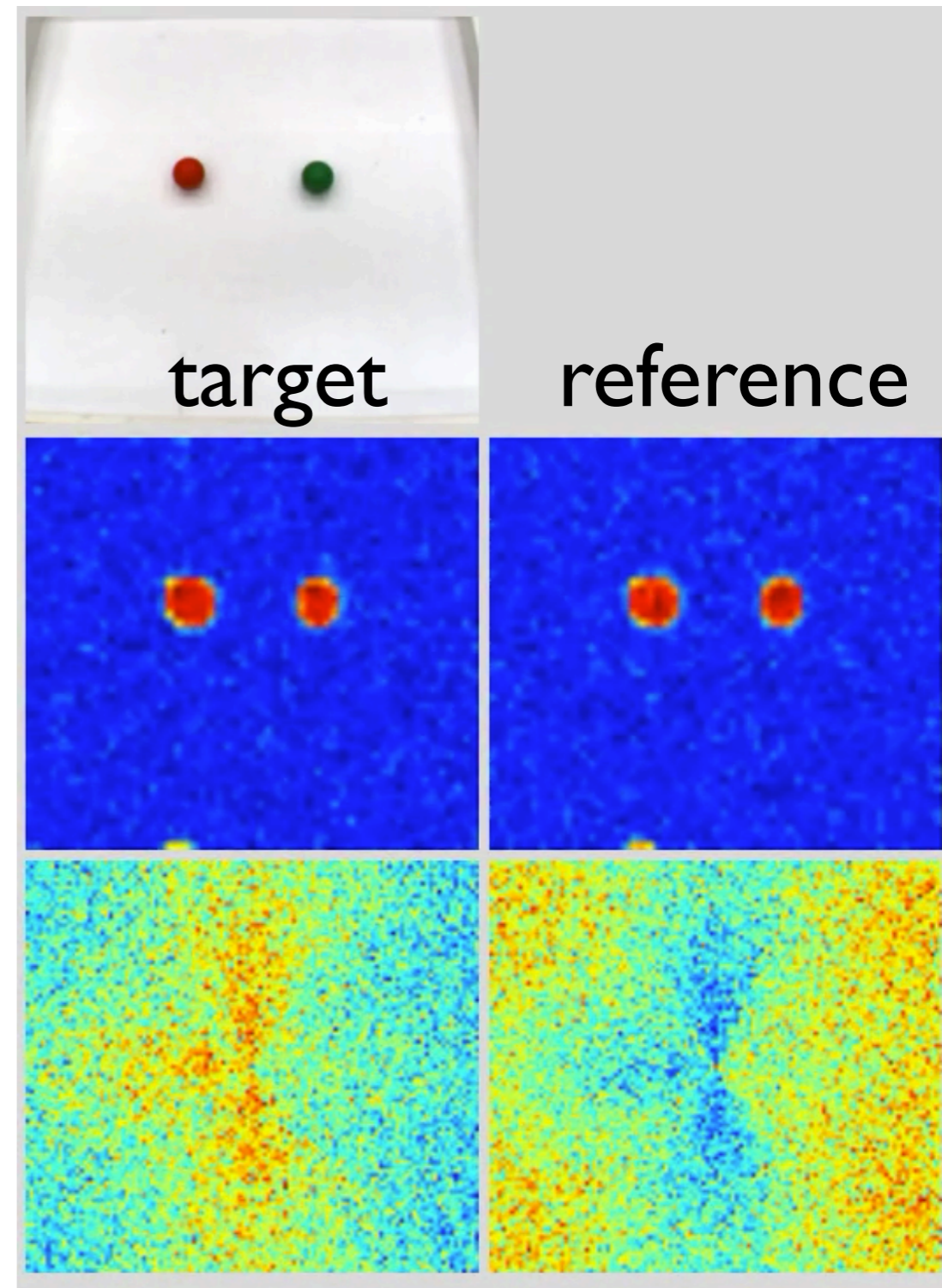
- => spatial language often involves coordinate transforms
- e.g., “to the right of the green book”: coordinate transformation: from the speaker/observer centered reference frame into a frame centered in the reference object
- e.g., “to my right” requires the listener to transform the reference frame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



# Grounding in DFT

- bringing the target object of a relational phrase into the attentional foreground

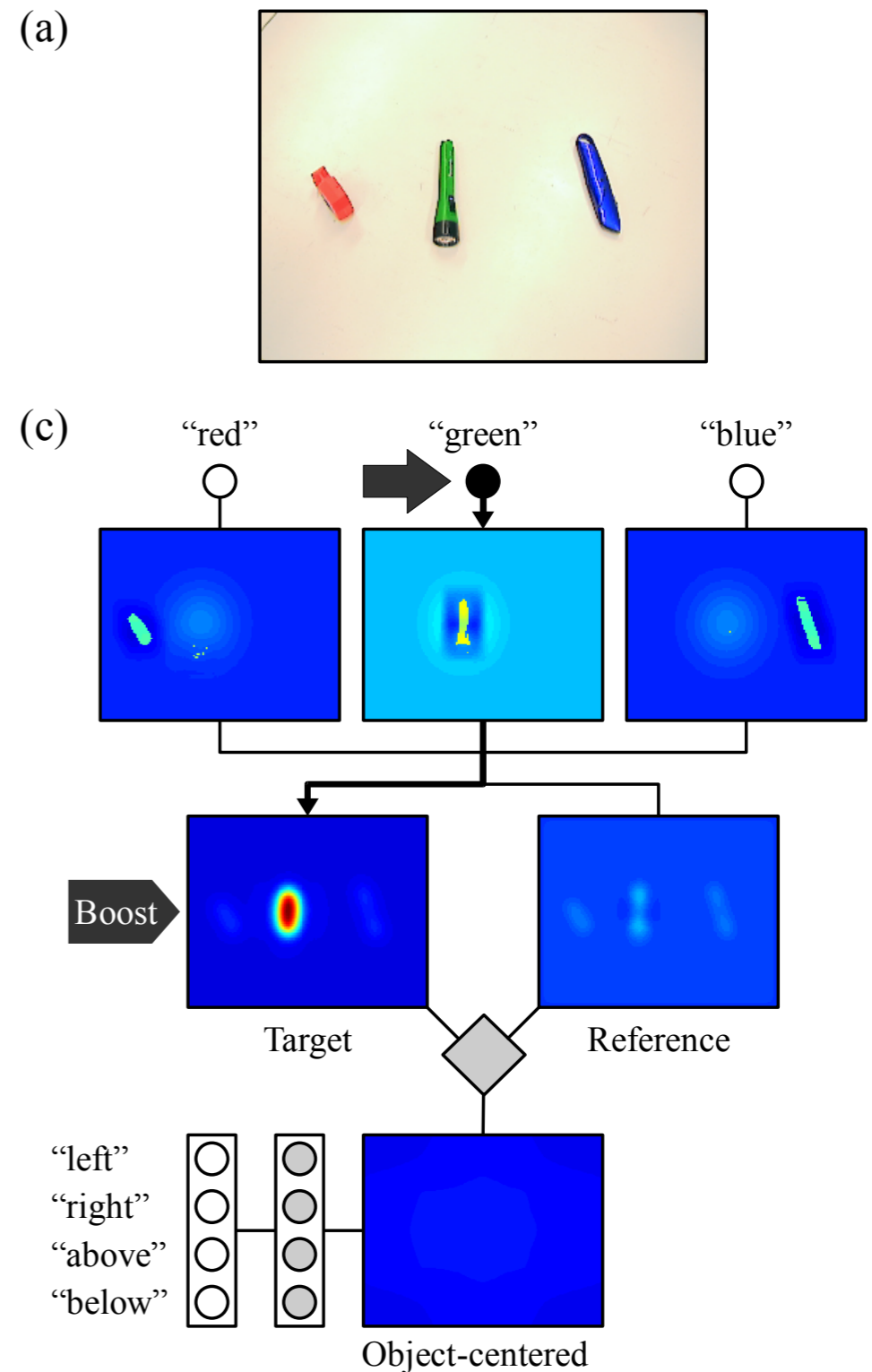
“red to the left of green”



[Lipinski, Sandamirskaya, Schöner 2009  
... Richter, Lins, Schöner, *Topics* 2017]

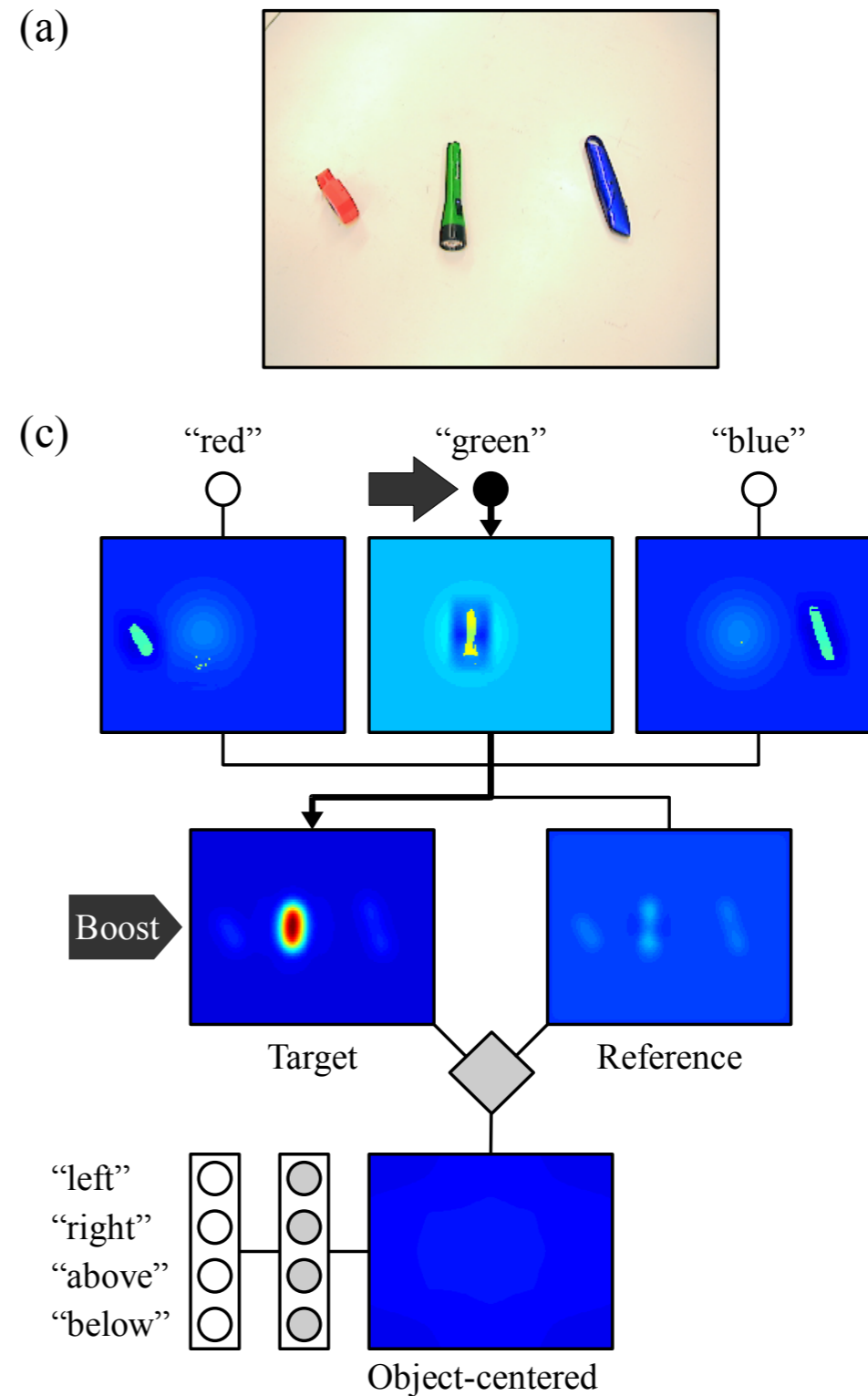


# Core idea for grounding spatial relations in DFT



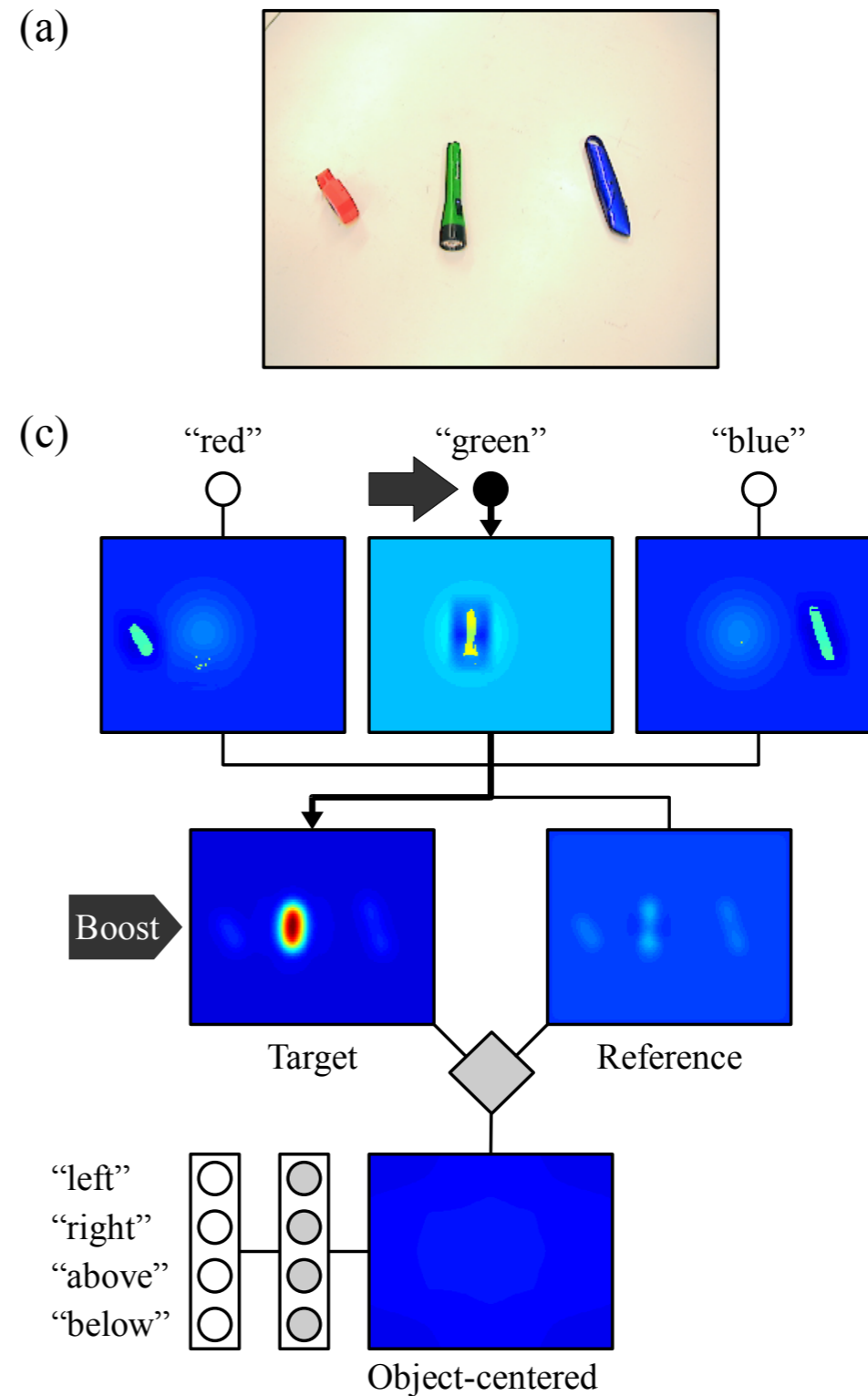
# Core idea for grounding spatial relations in DFT

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



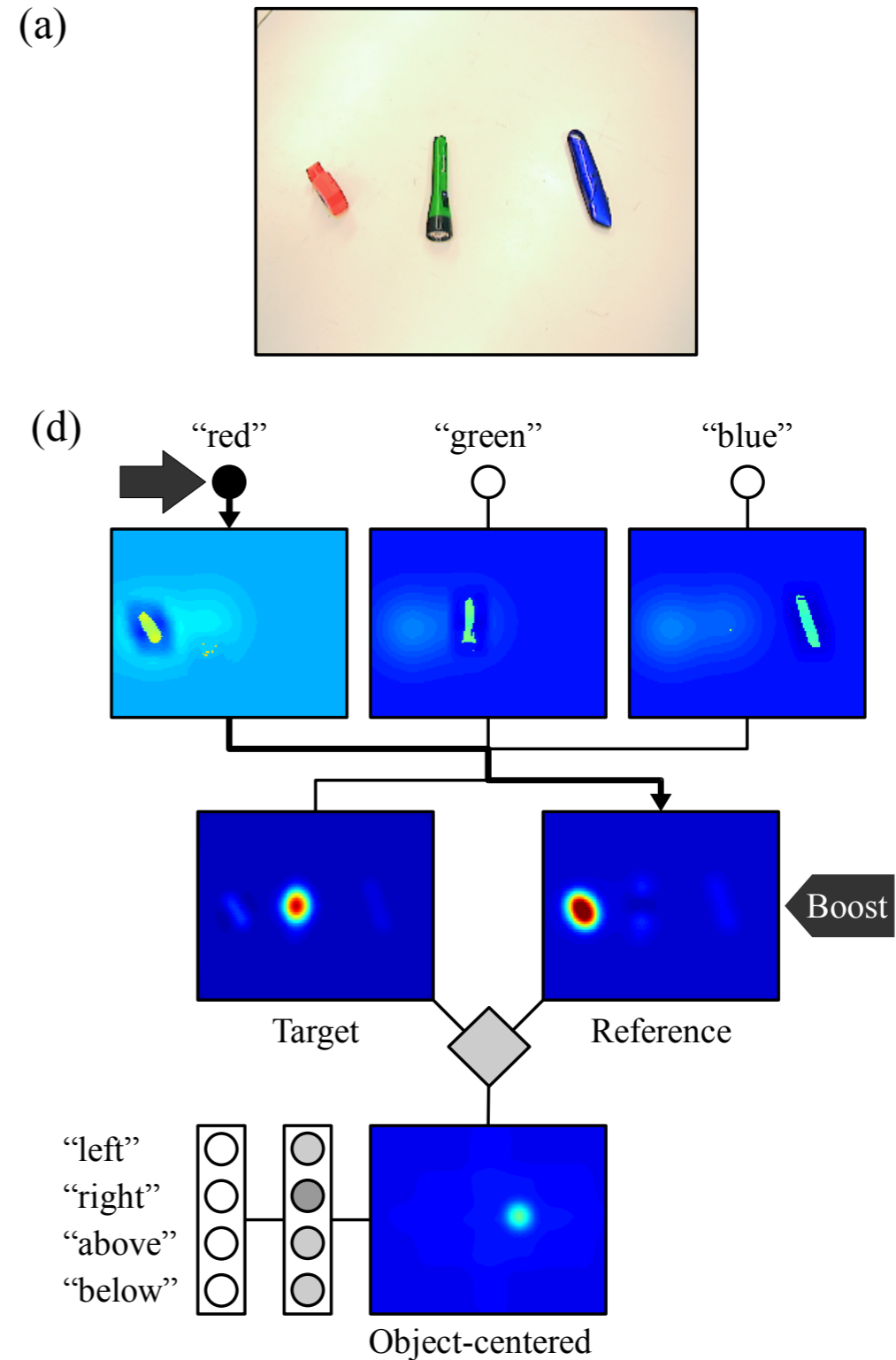
# Core idea for grounding spatial relations in DFT

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



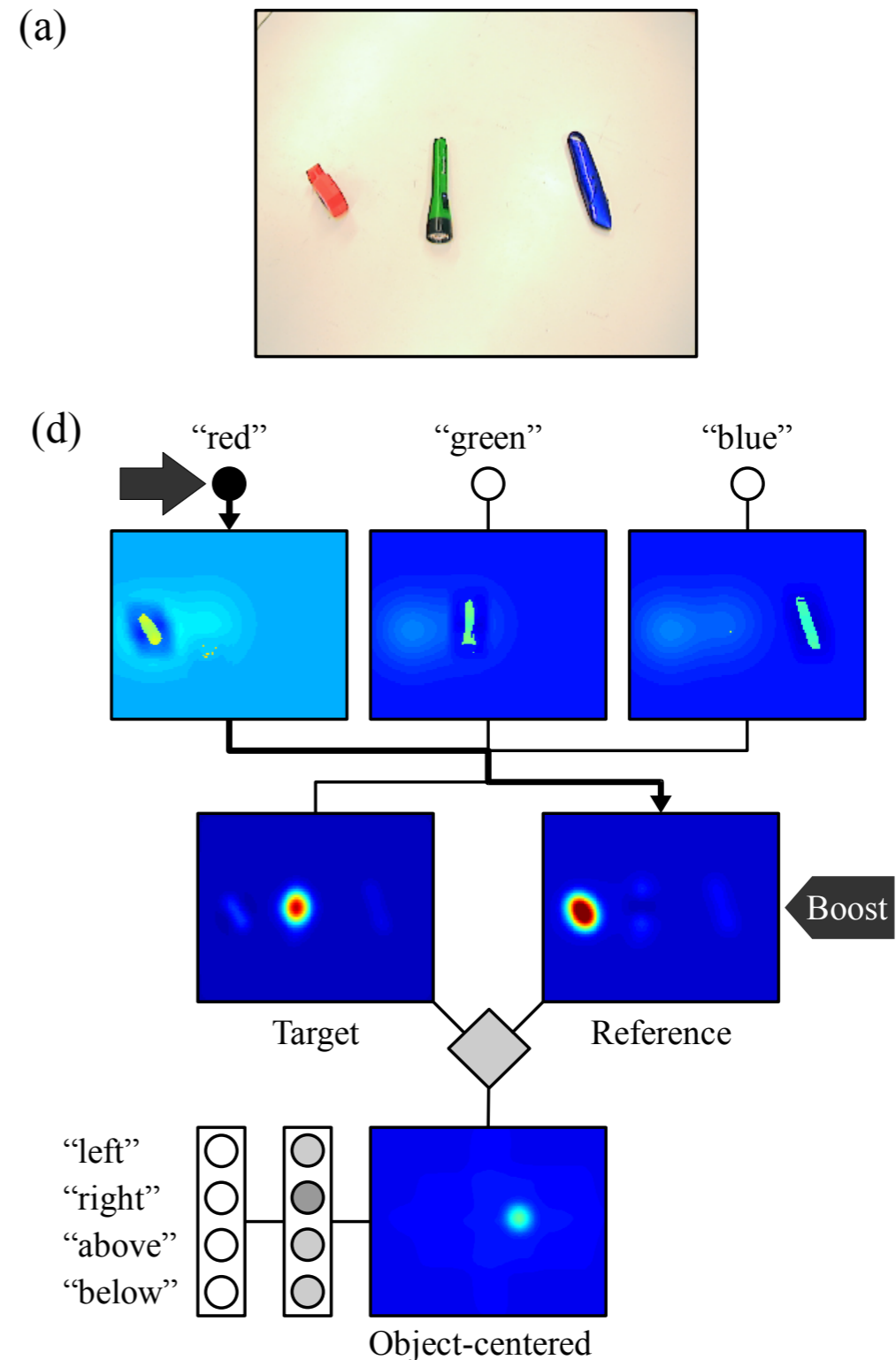
# Core idea for grounding spatial relations in DFT

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



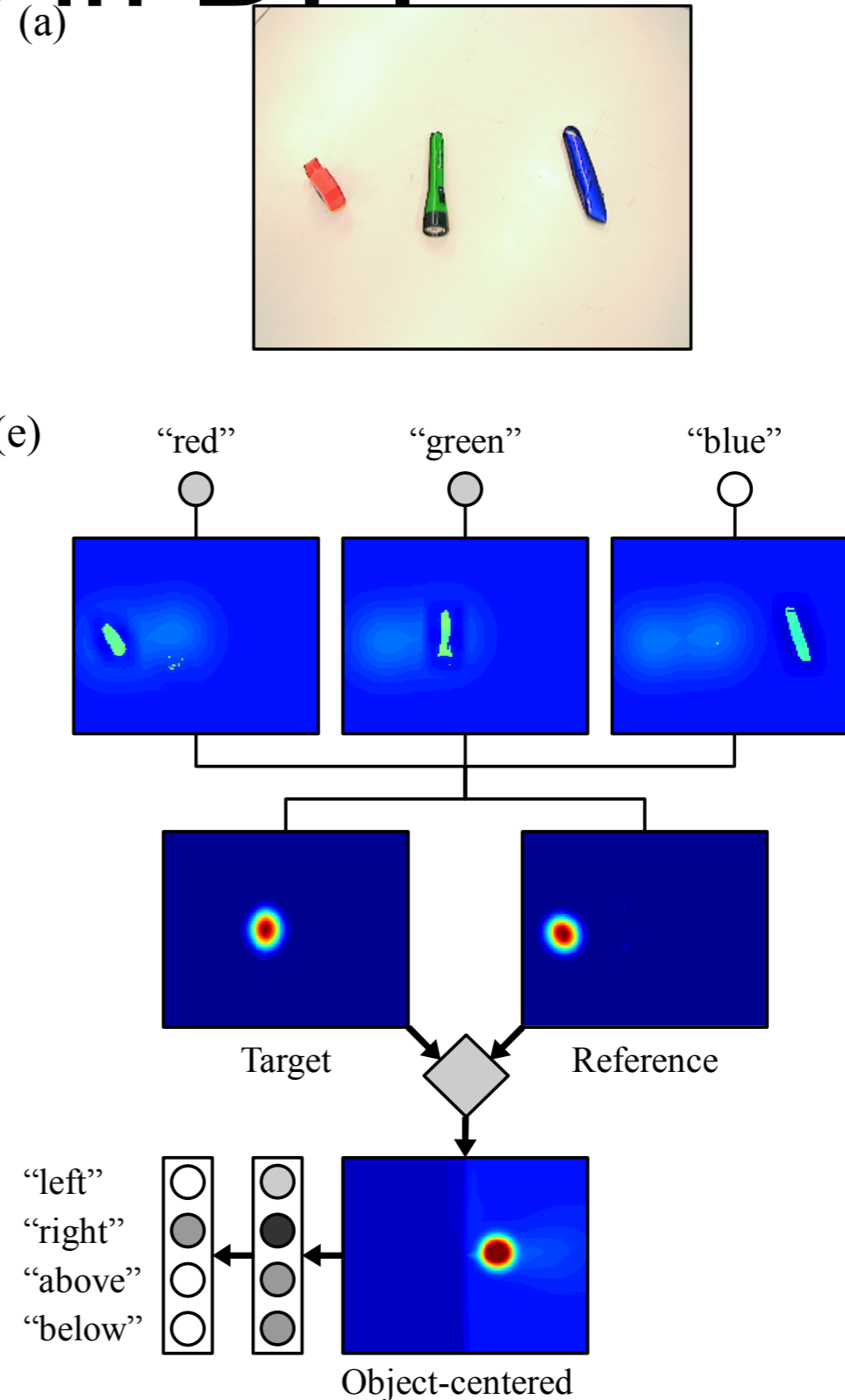
# Core idea for grounding spatial relations in DFT

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

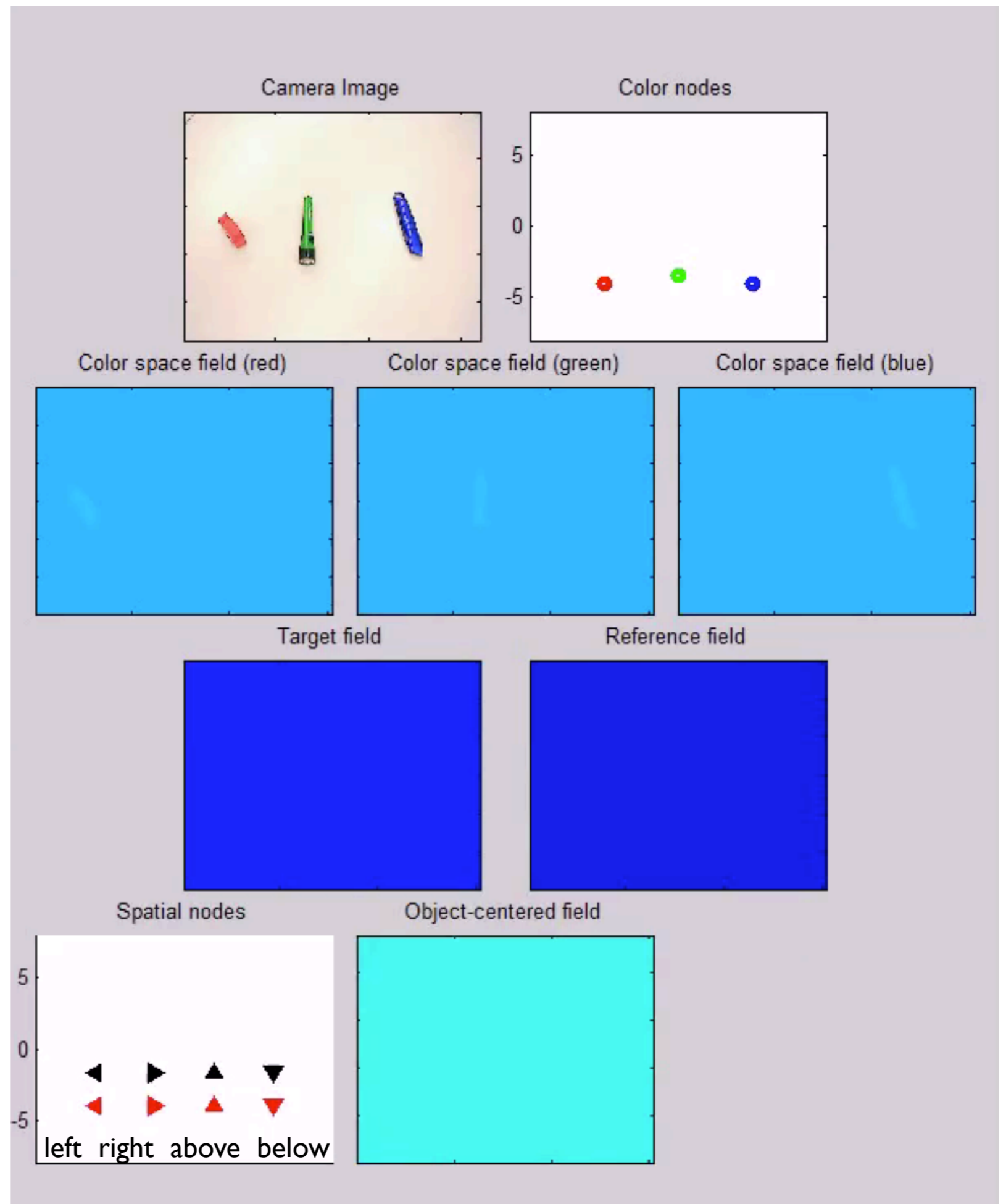


# Core idea for grounding spatial relations in DFT

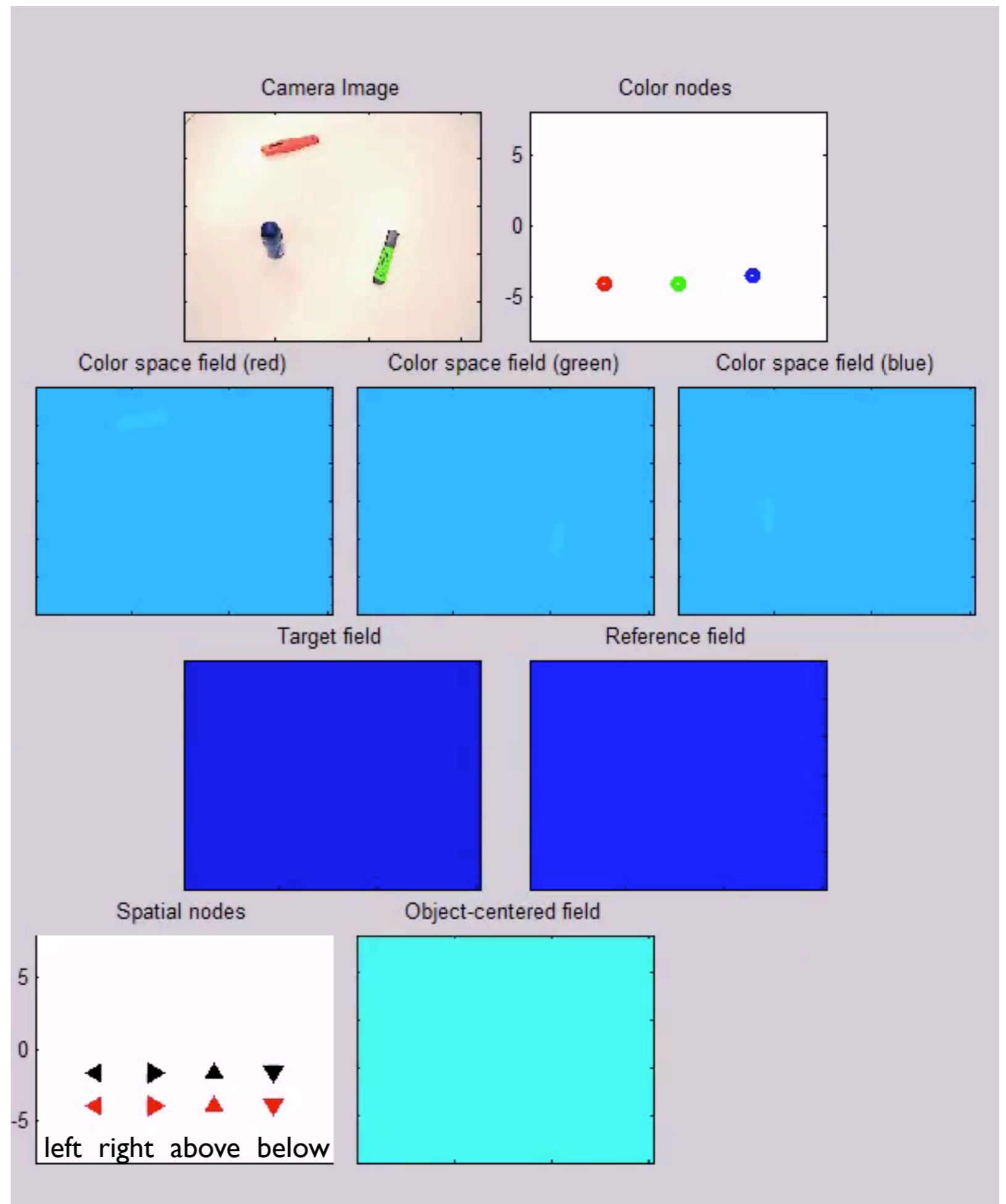
- bring objects into foreground
- make coordinate transformation
- apply comparison operators



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

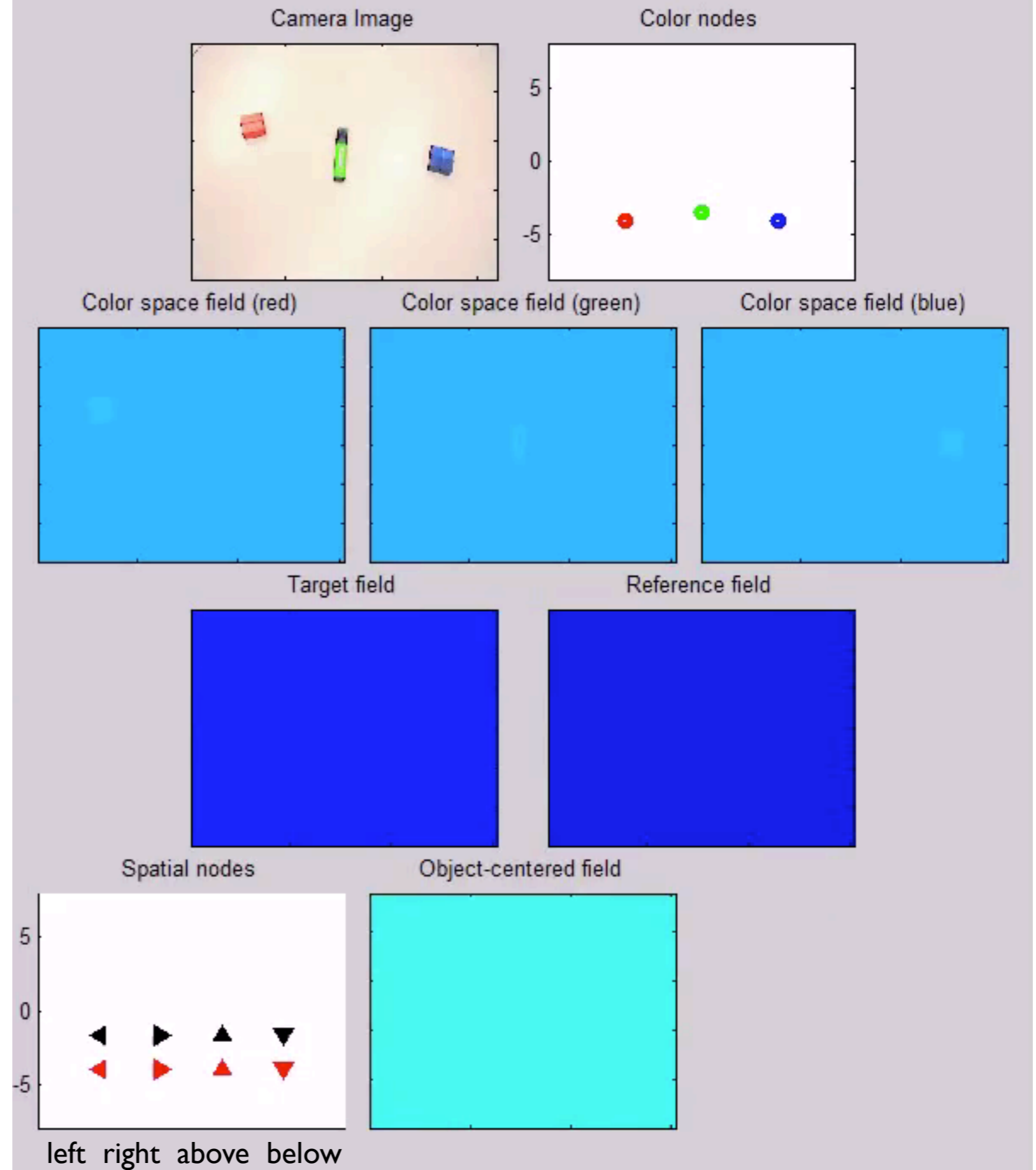


■ “which object is above the blue object?”



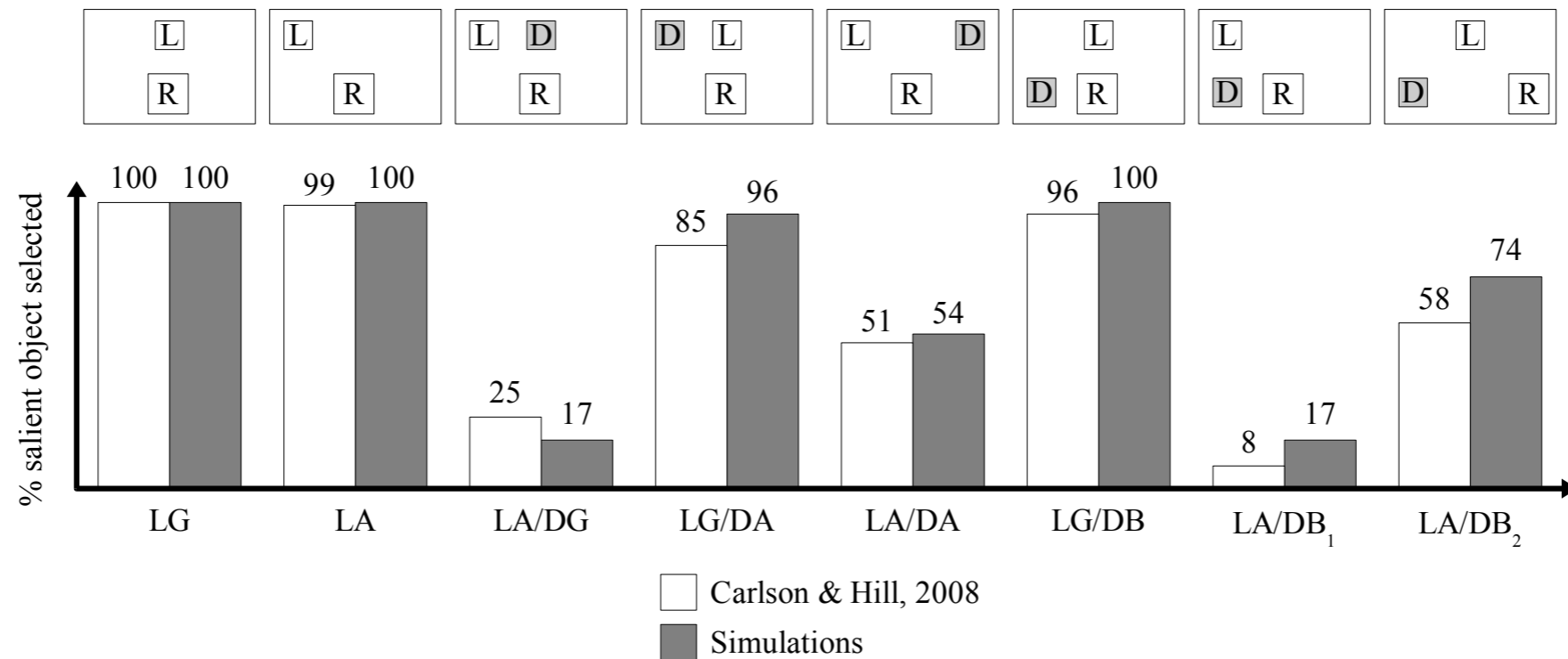


■ “where is the green object?”



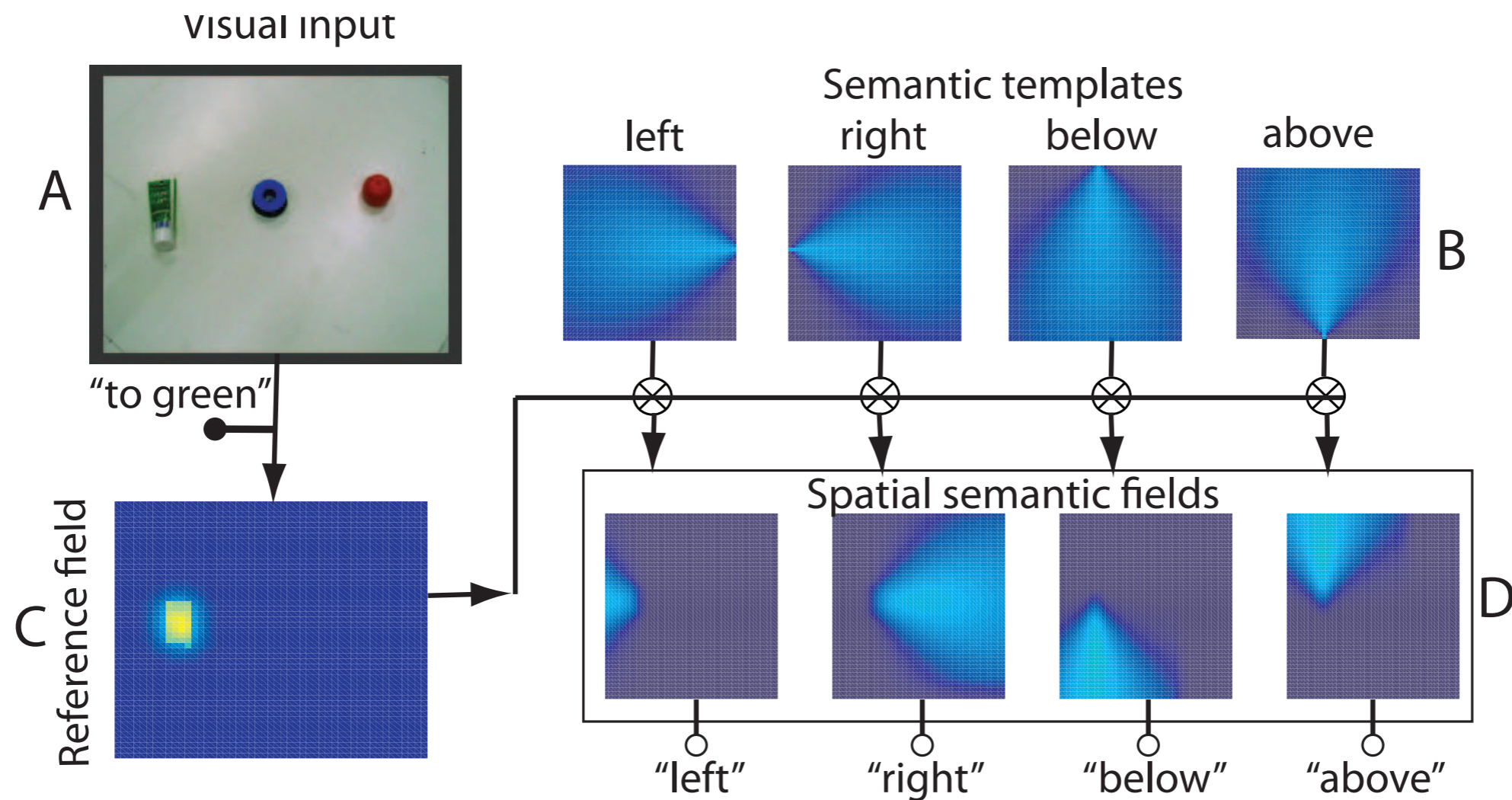
# Spatial comparison in DFT

■ accounts for human data



# Alternative: coordinate transform applied to the neural operator

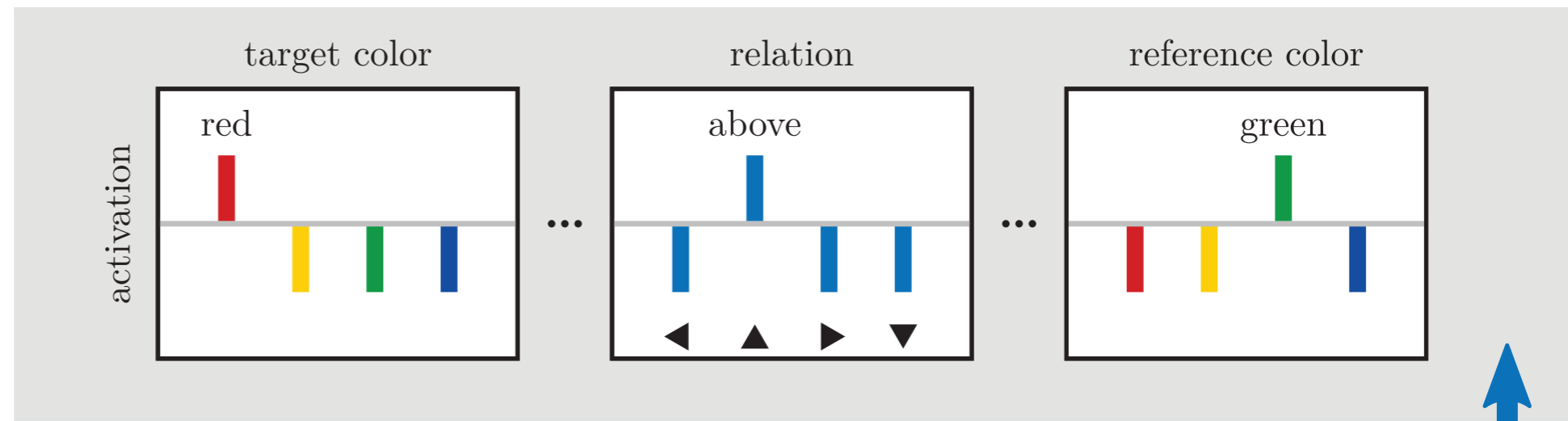
- based on convolution of fields with kernels



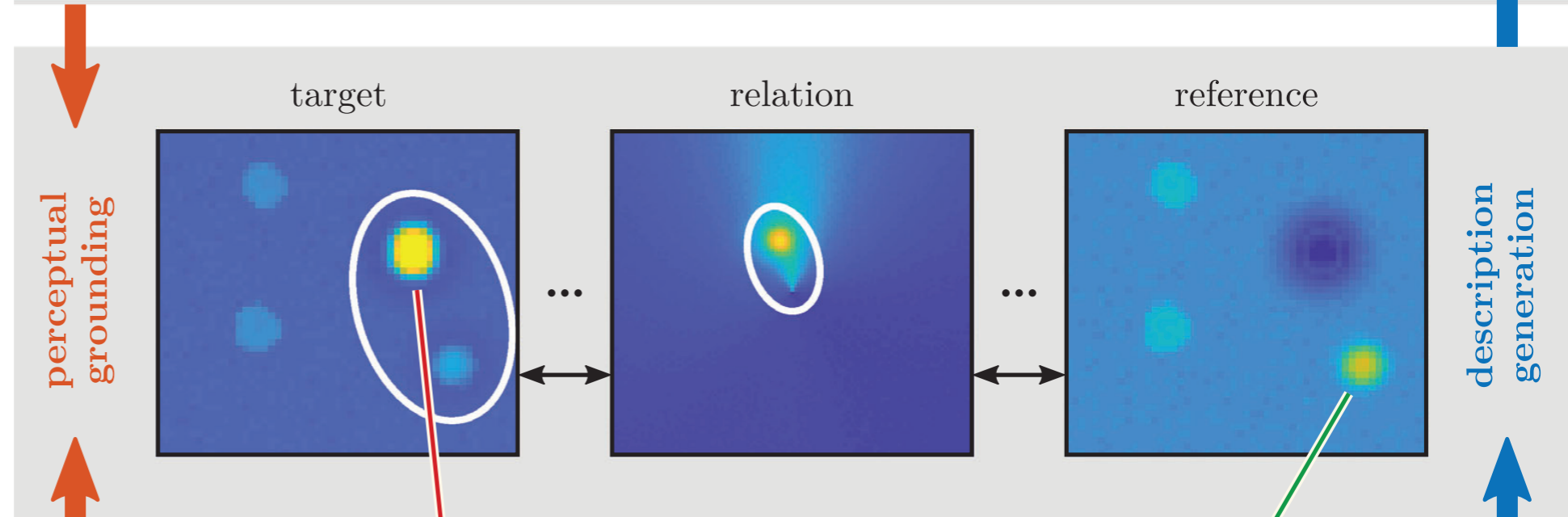
[from: Lipinski, Sandamirskaya, Schöner, 2009]

# Perceptual grounding vs describing

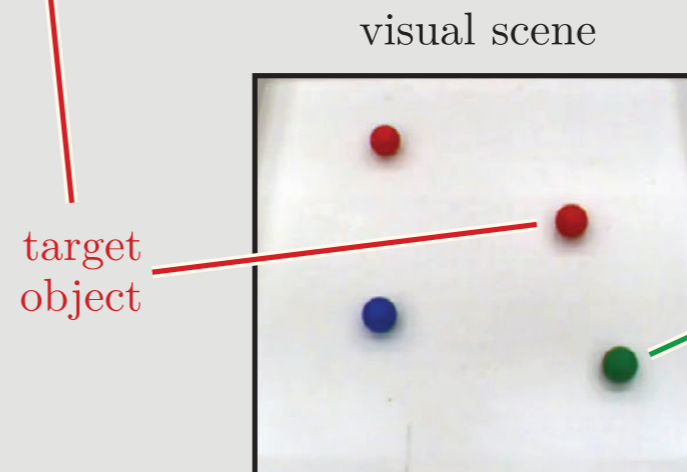
conceptual  
structure

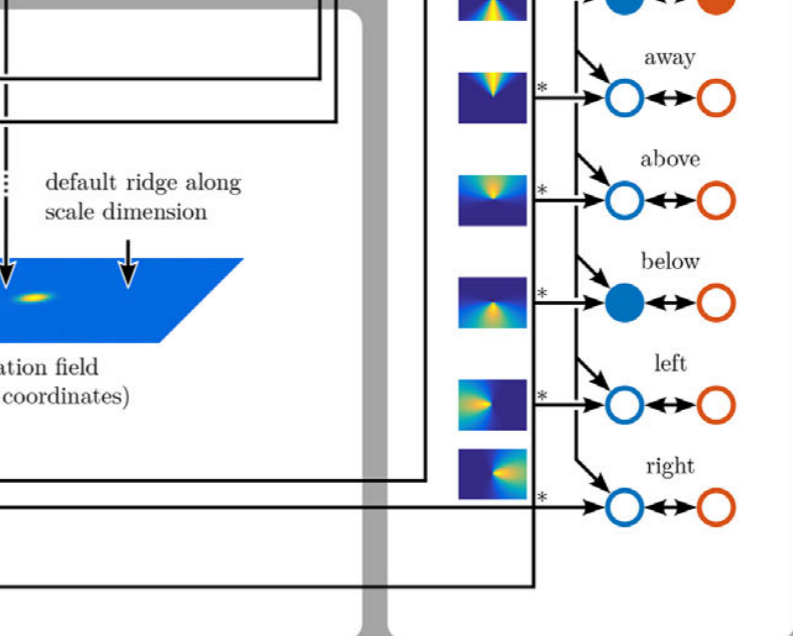
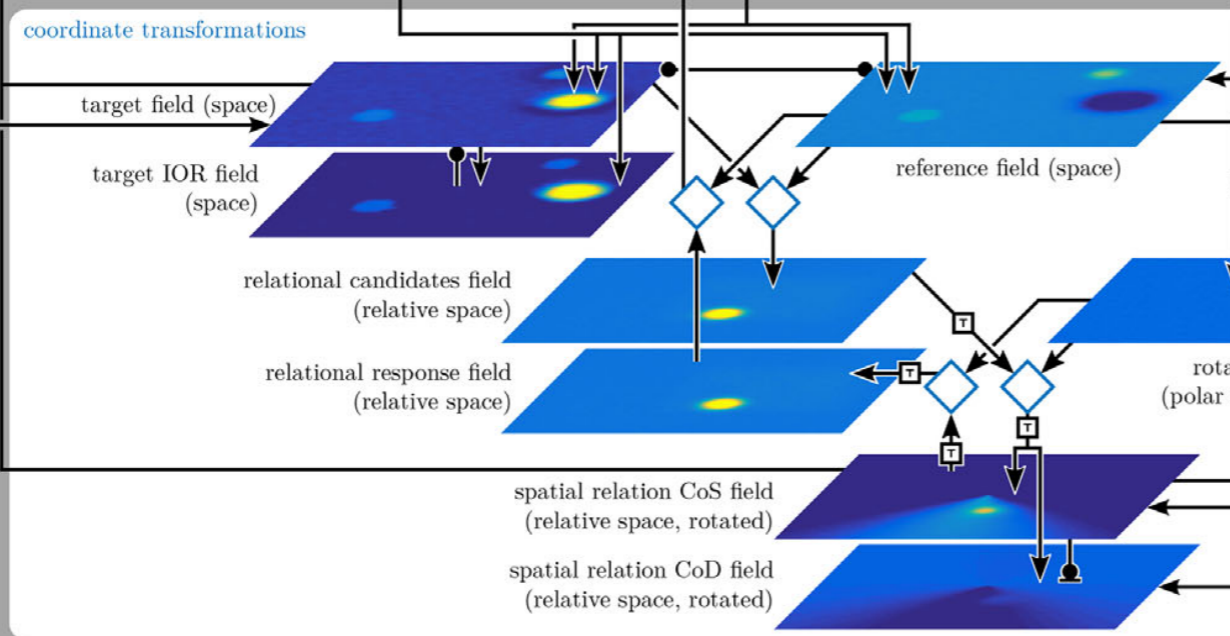
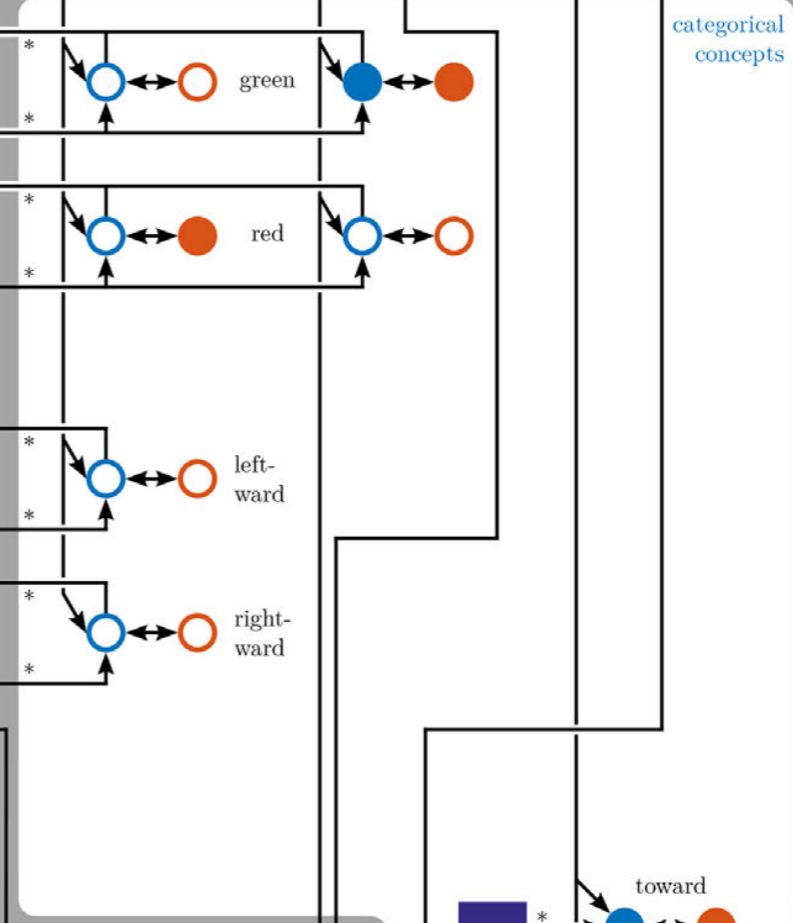
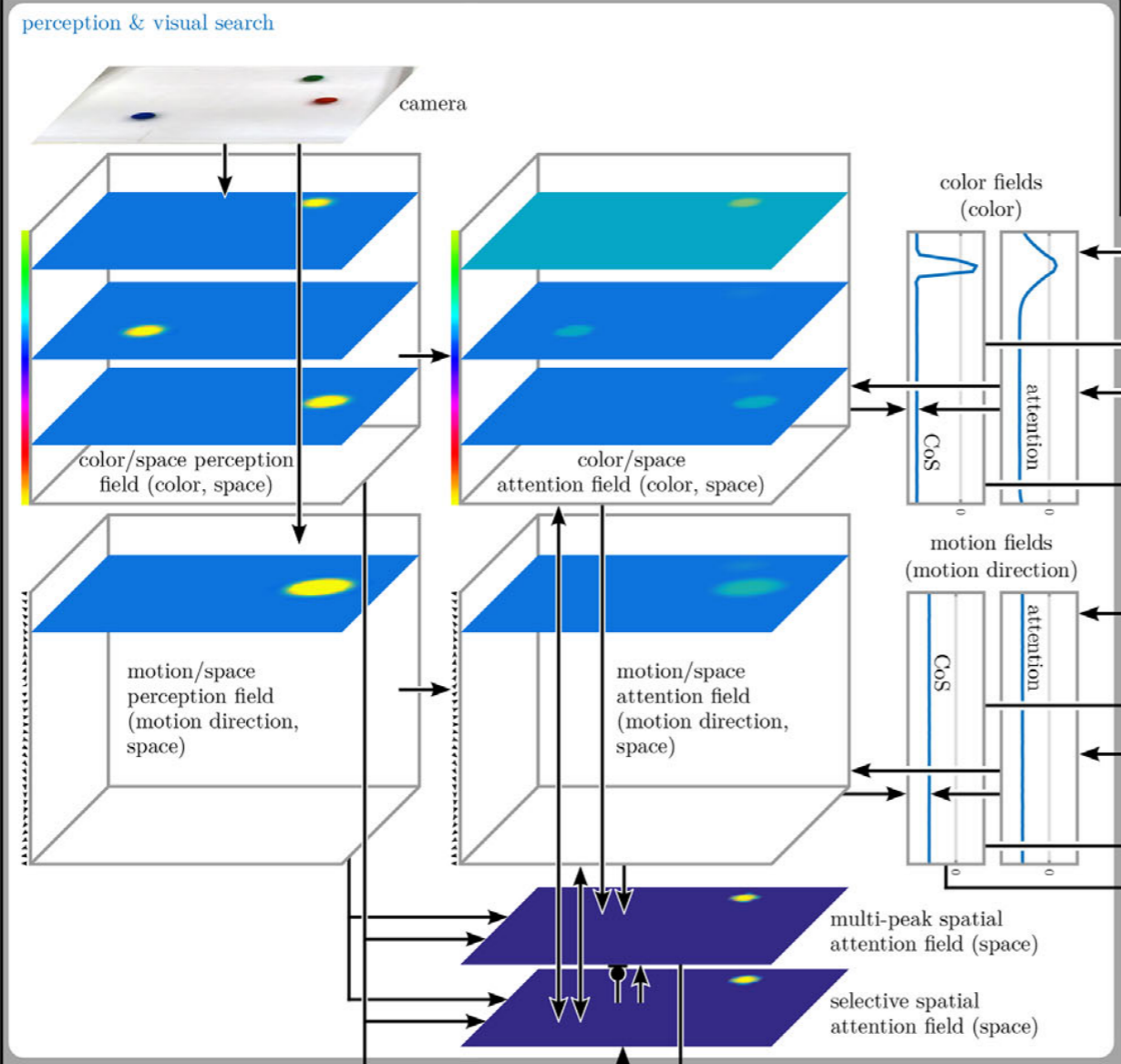
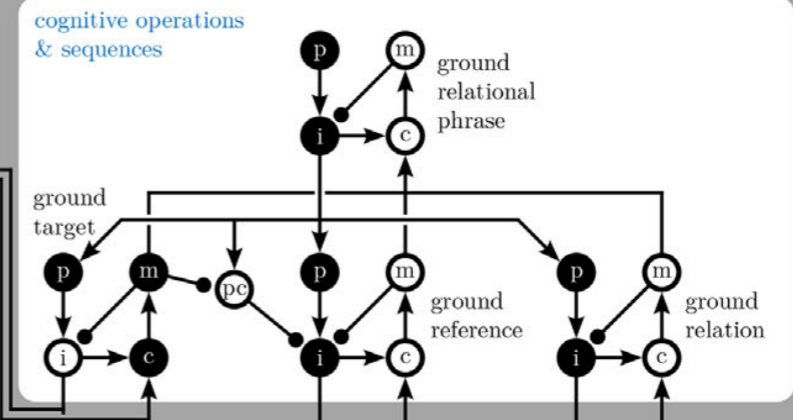
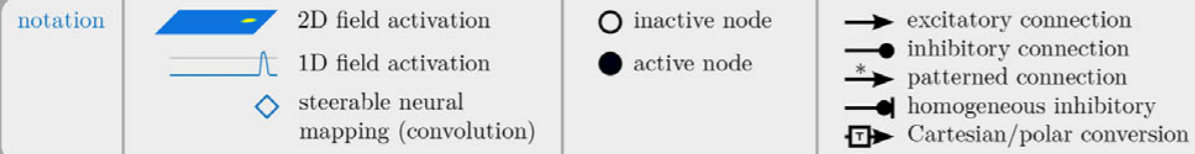


grounded  
representation

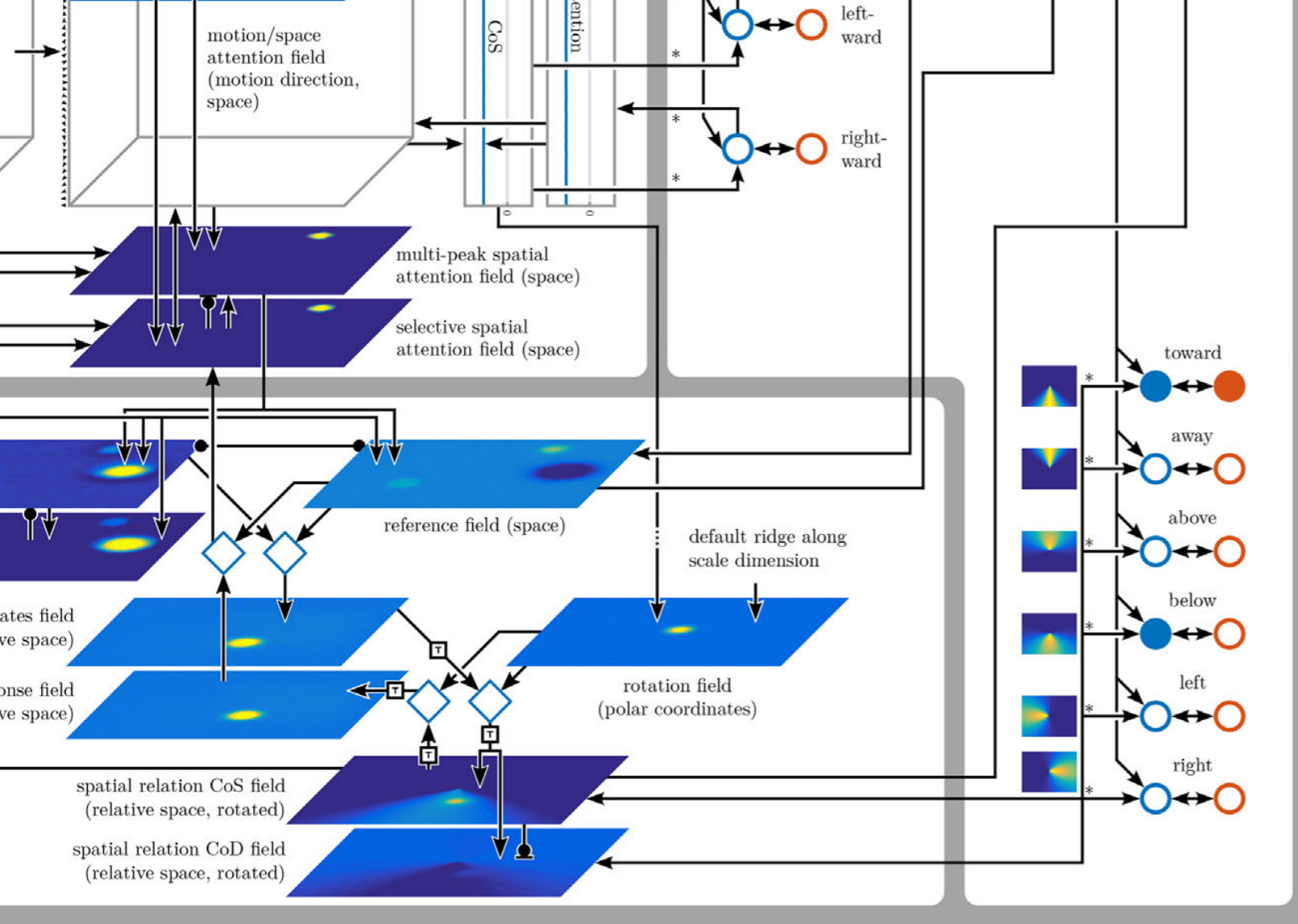


[Richter, Lins,  
Schöner, 2021]









# COGNITIVE SCIENCE

A Multidisciplinary Journal



Cognitive Science 45 (2021) e13045

© 2021 The Authors. *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).

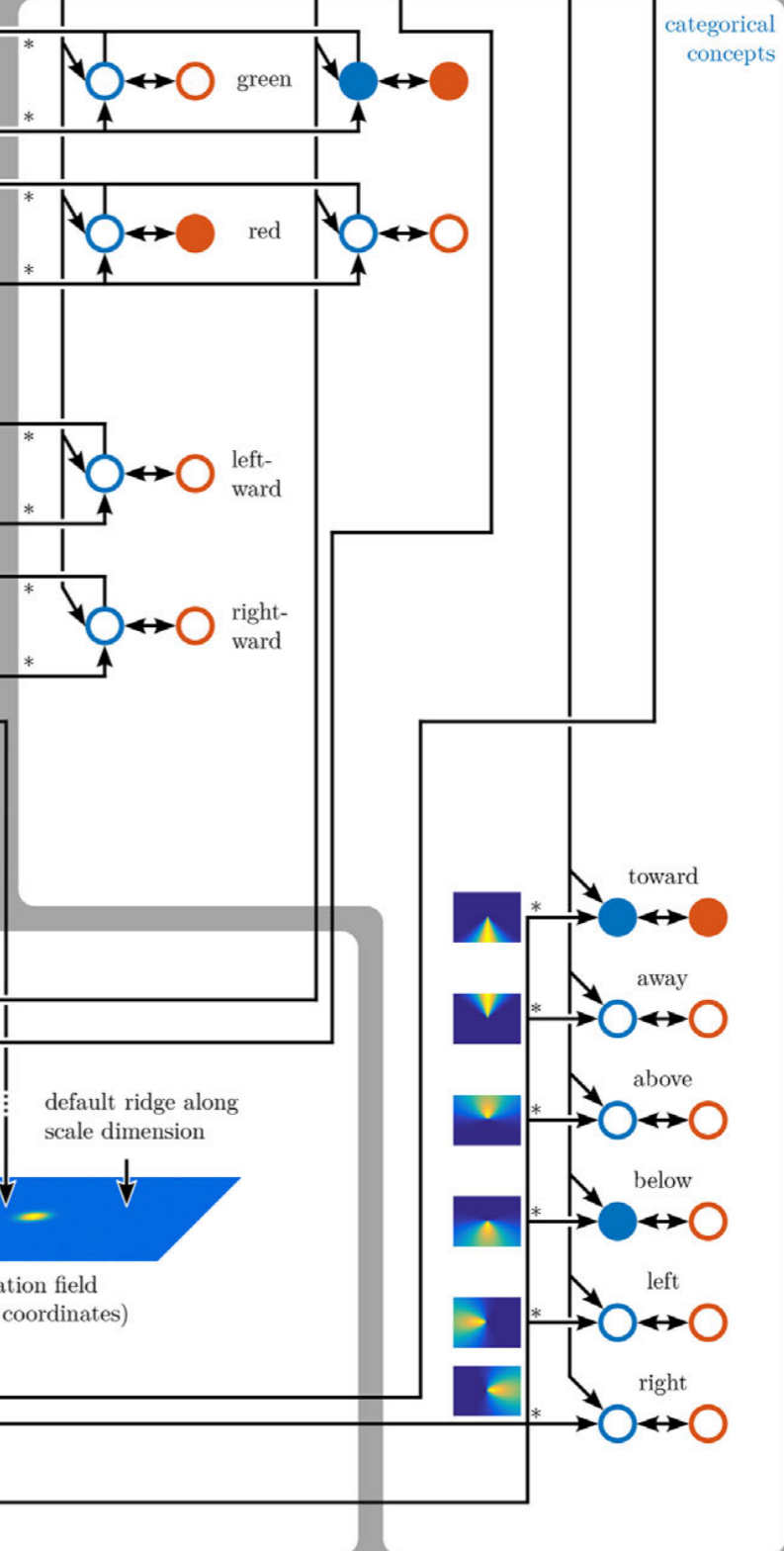
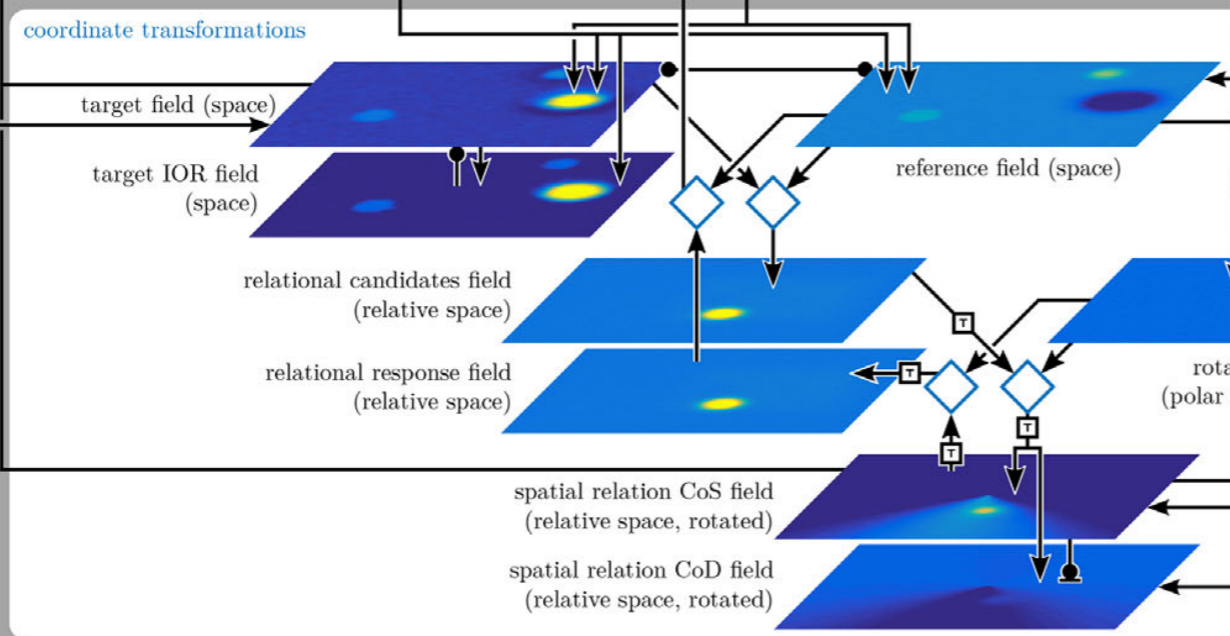
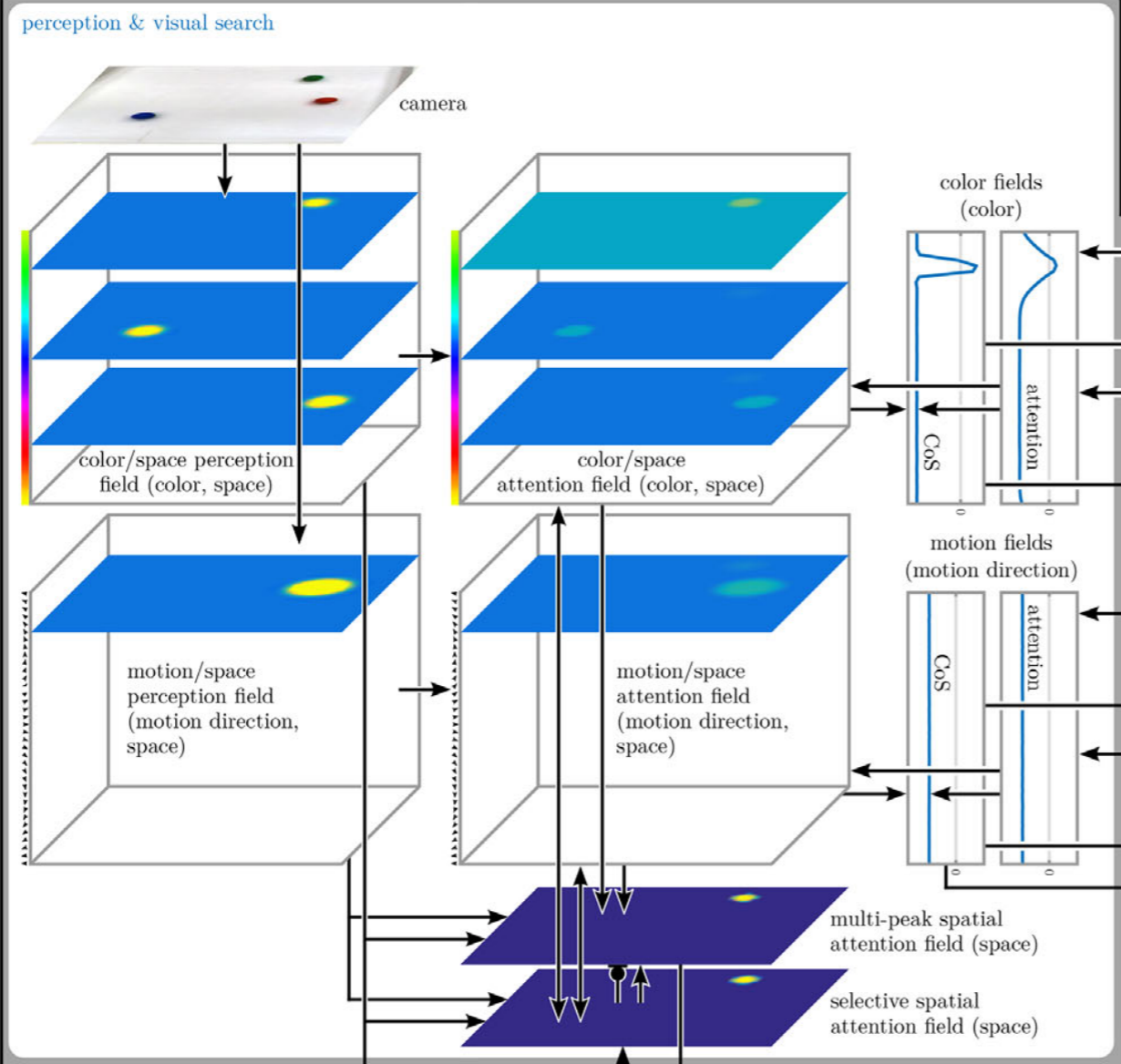
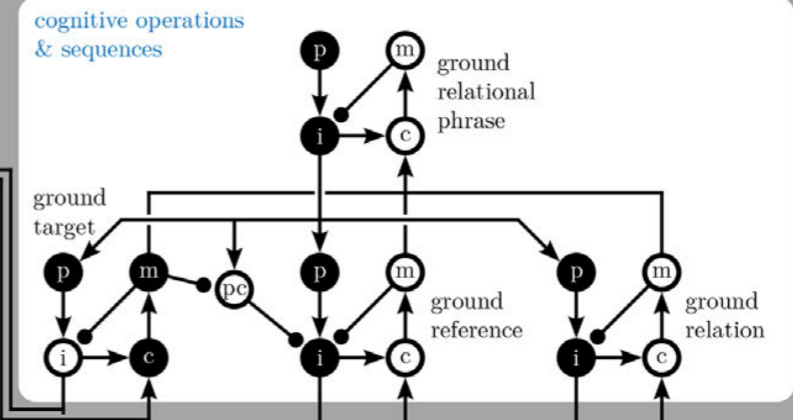
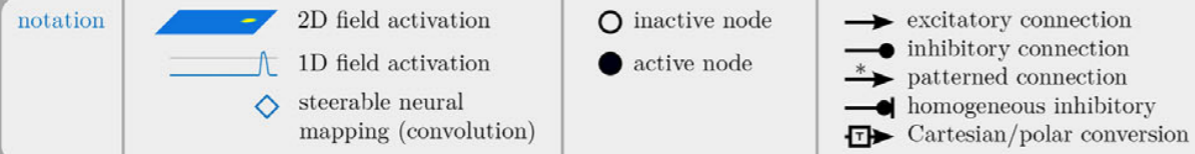
ISSN: 1551-6709 online

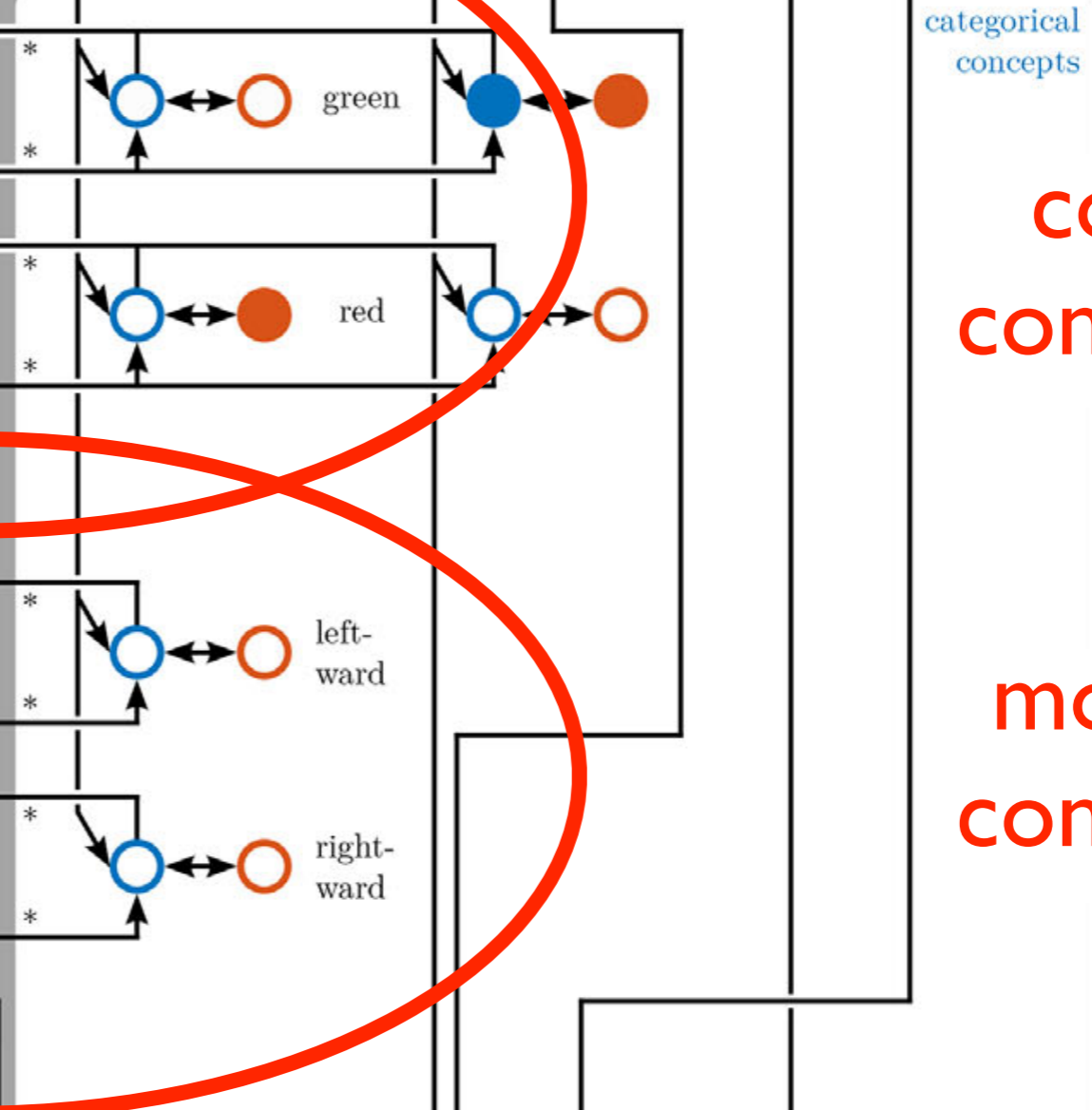
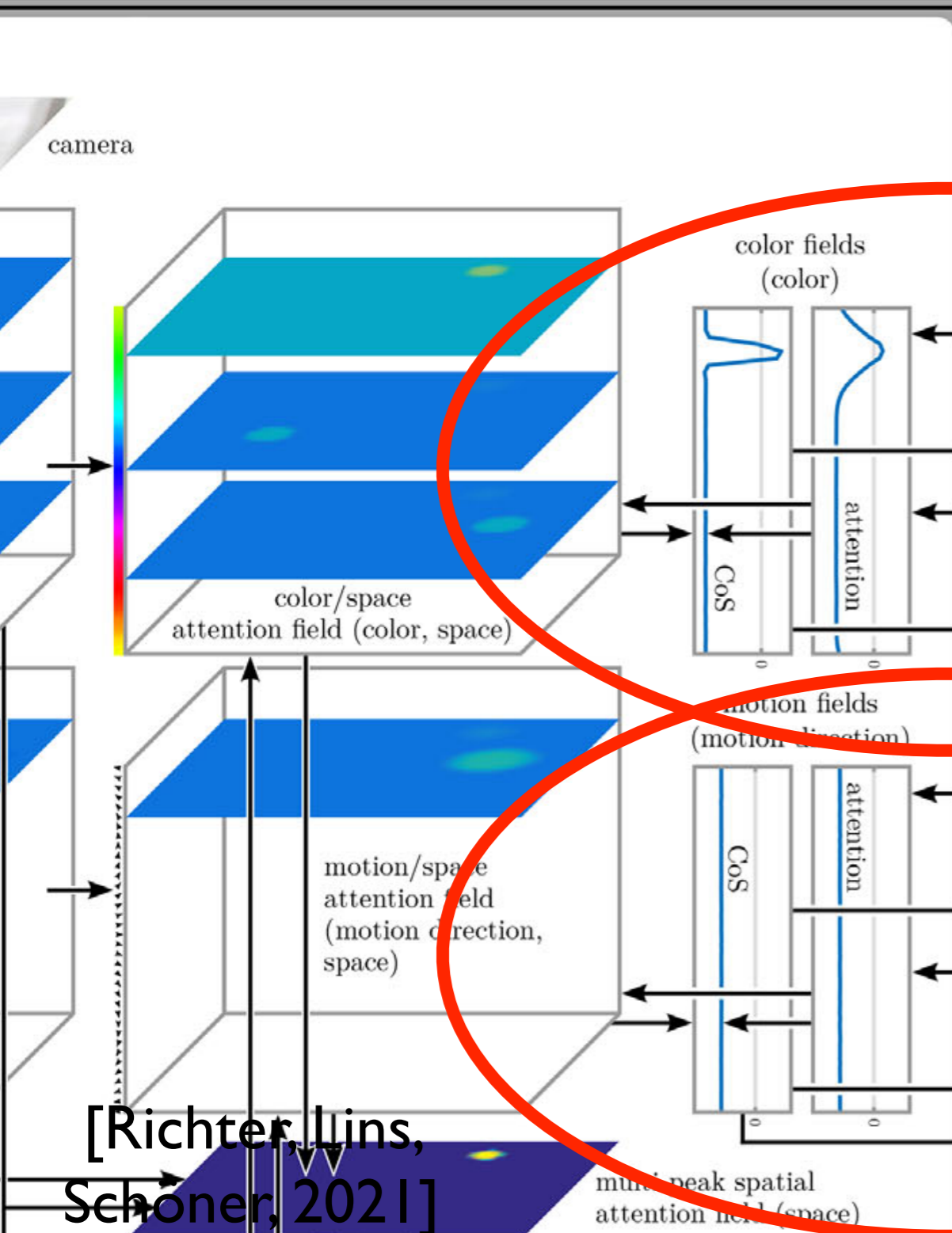
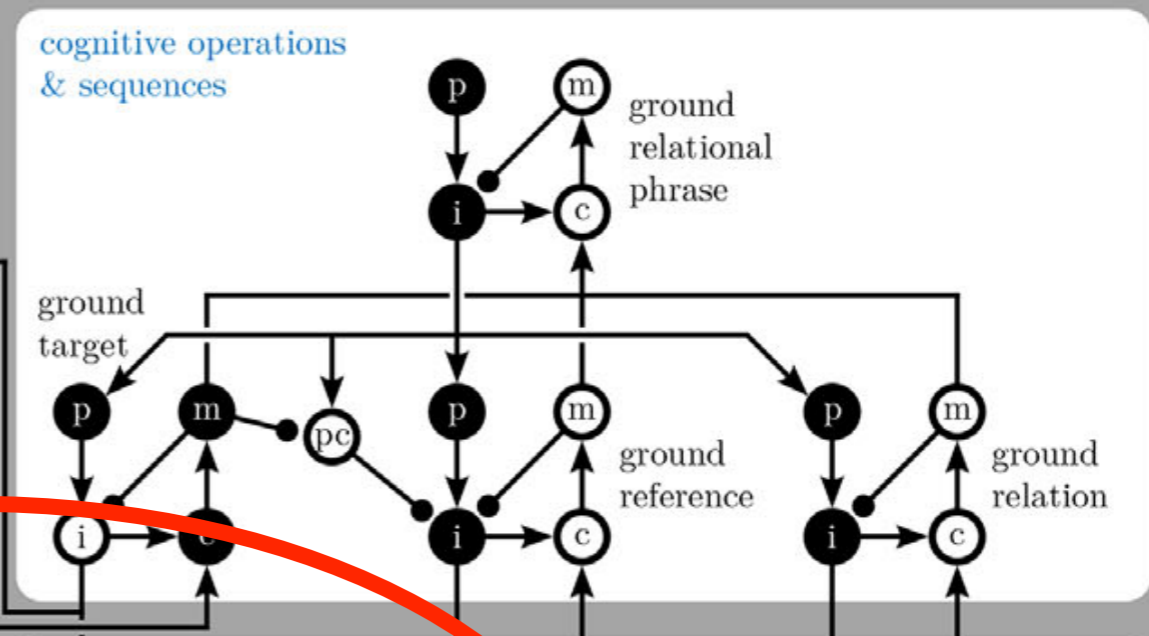
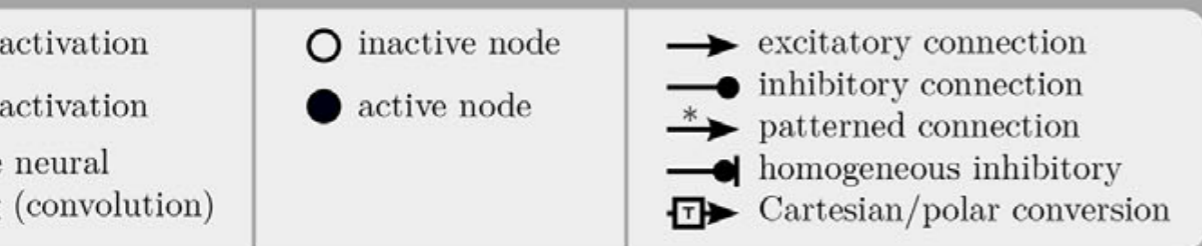
DOI: 10.1111/cogs.13045

## A Neural Dynamic Model of the Perceptual Grounding of Spatial and Movement Relations

Mathis Richter,  Jonas Lins, Gregor Schöner

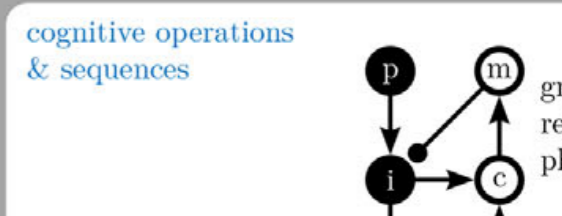
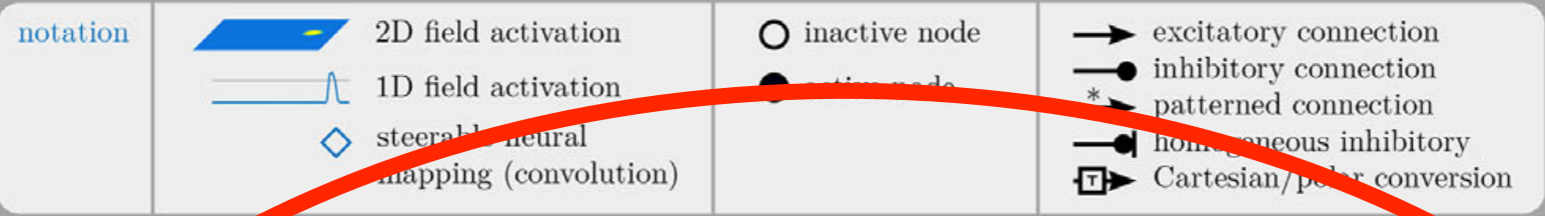
*Institut für Neuroinformatik, Ruhr-Universität Bochum*





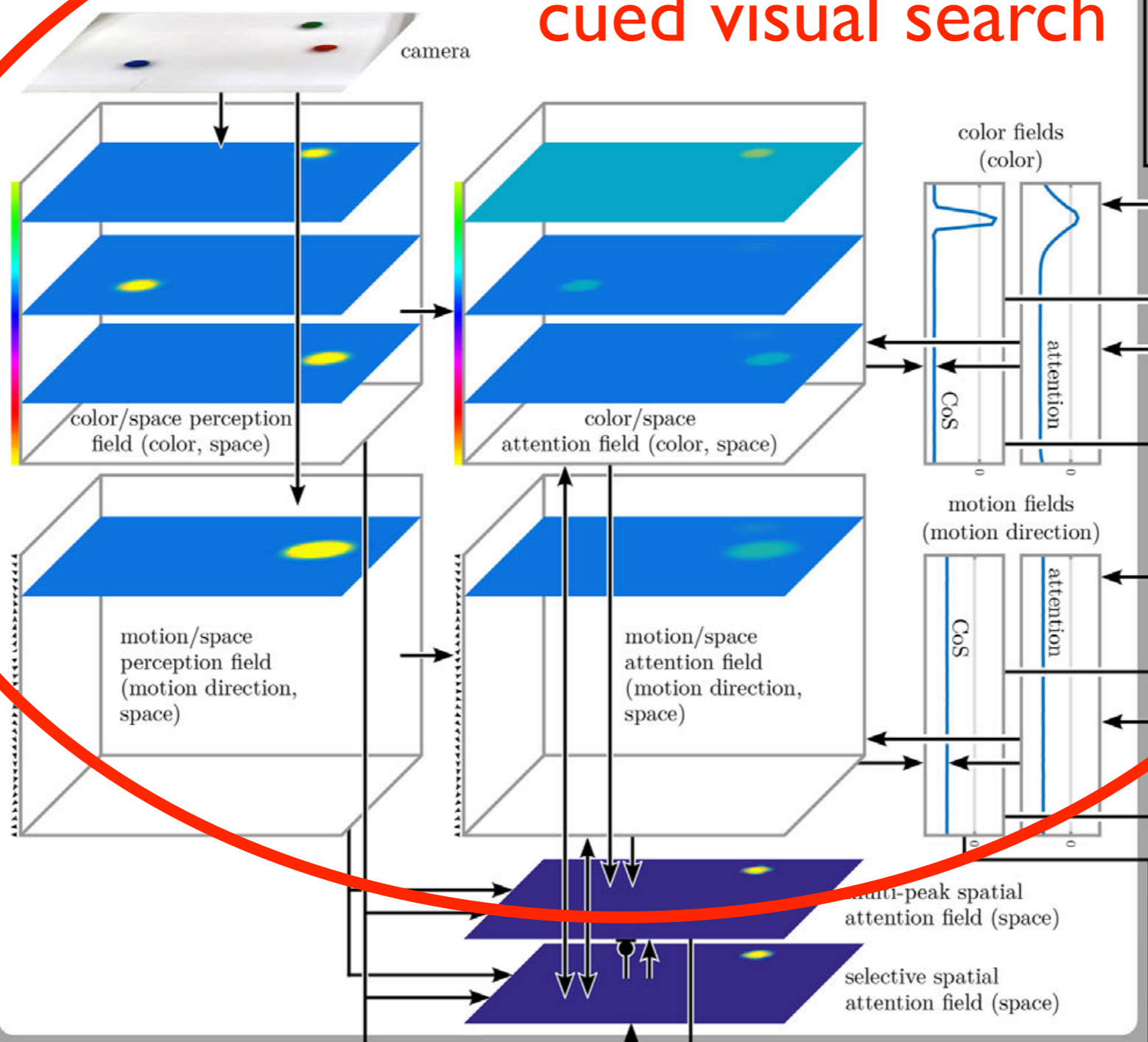
[Richter, Jins, Schöner, 2021]



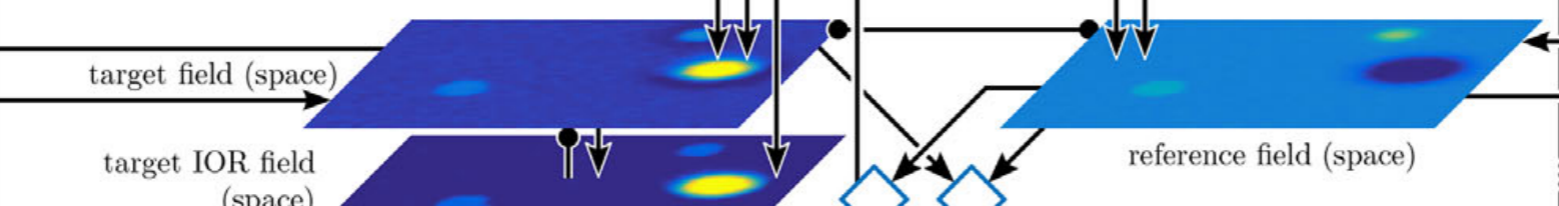


# cued visual search

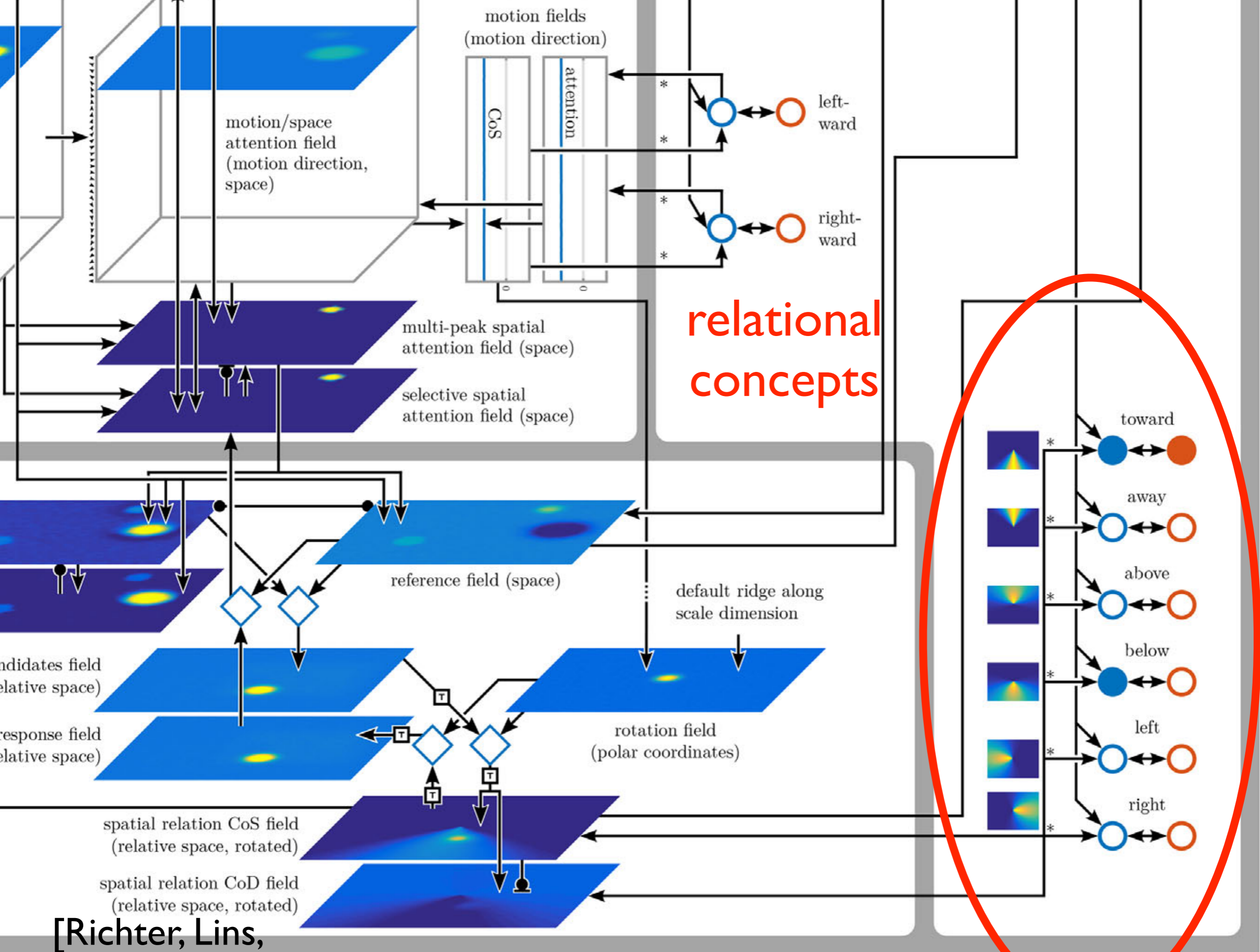
**perception & visual search**



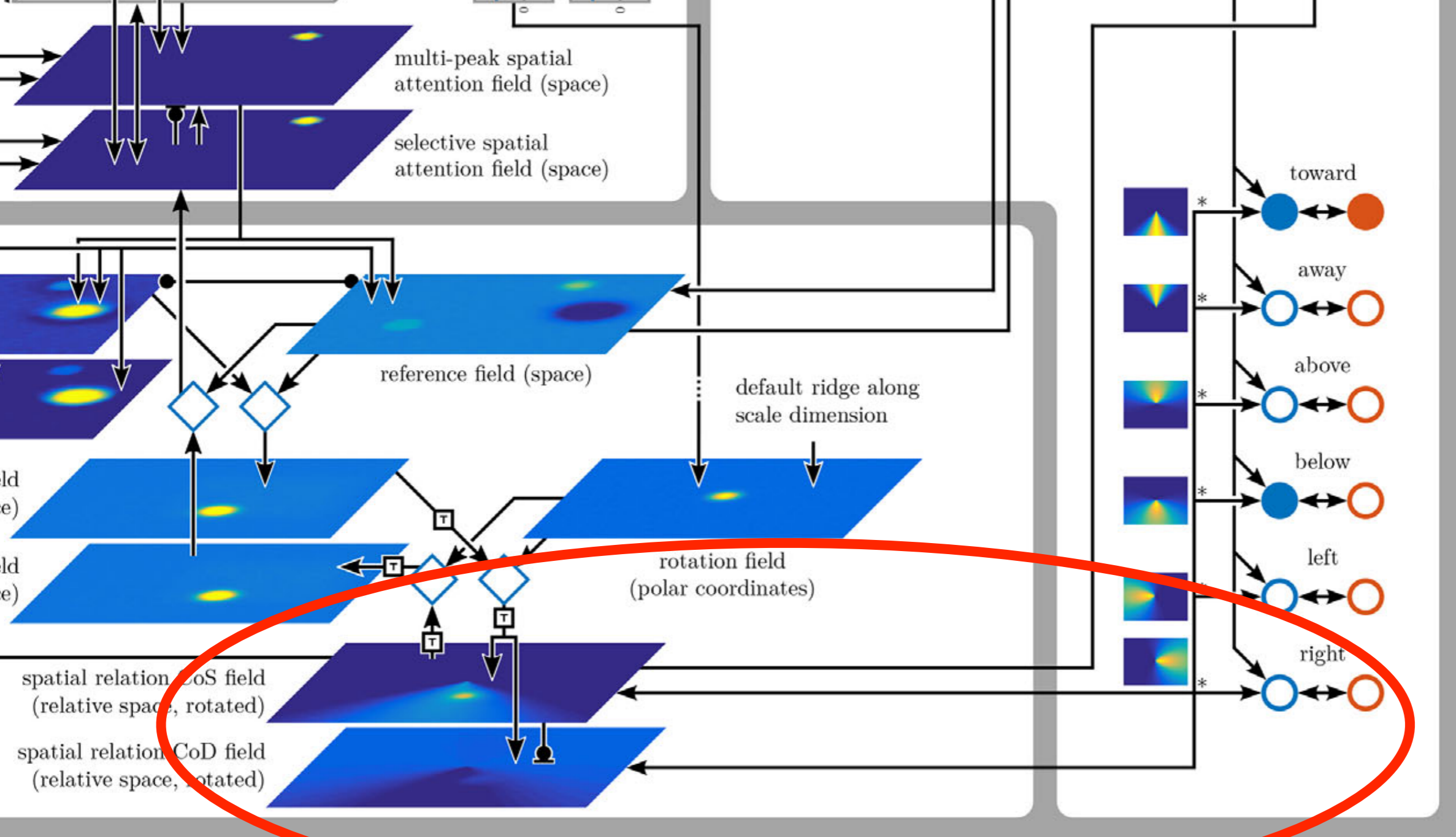
**coordinate transformations**



[Richter, Lins, Schöner, 2021]



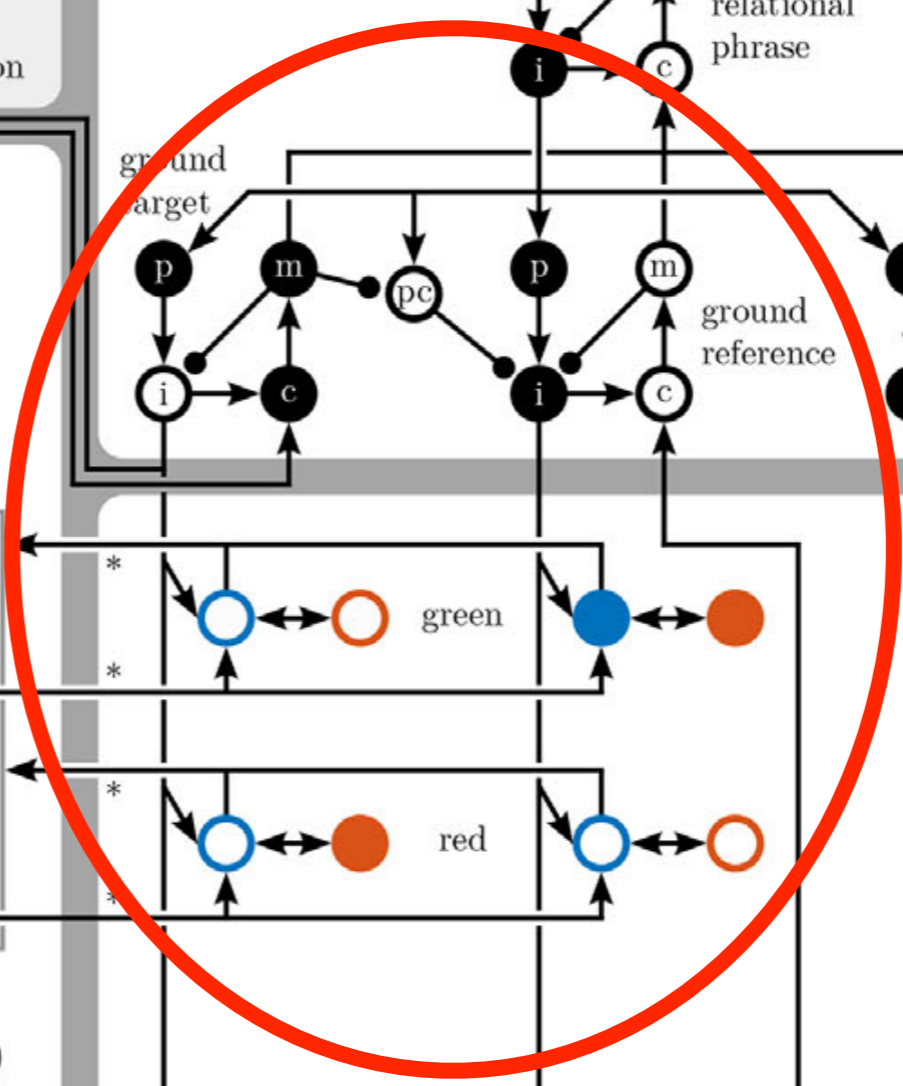
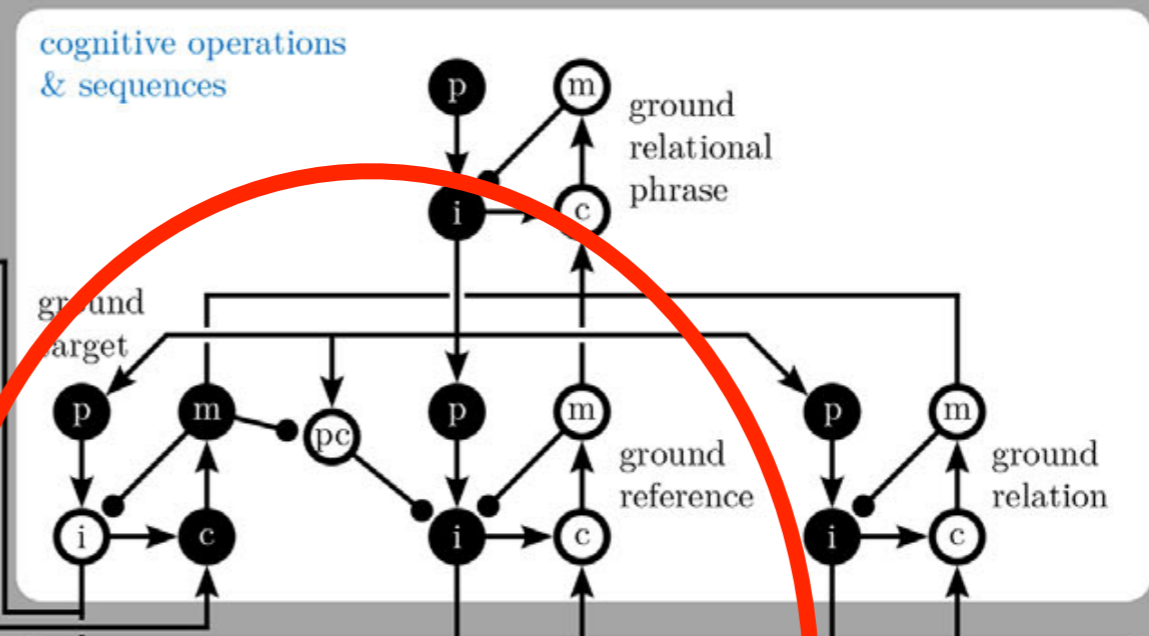
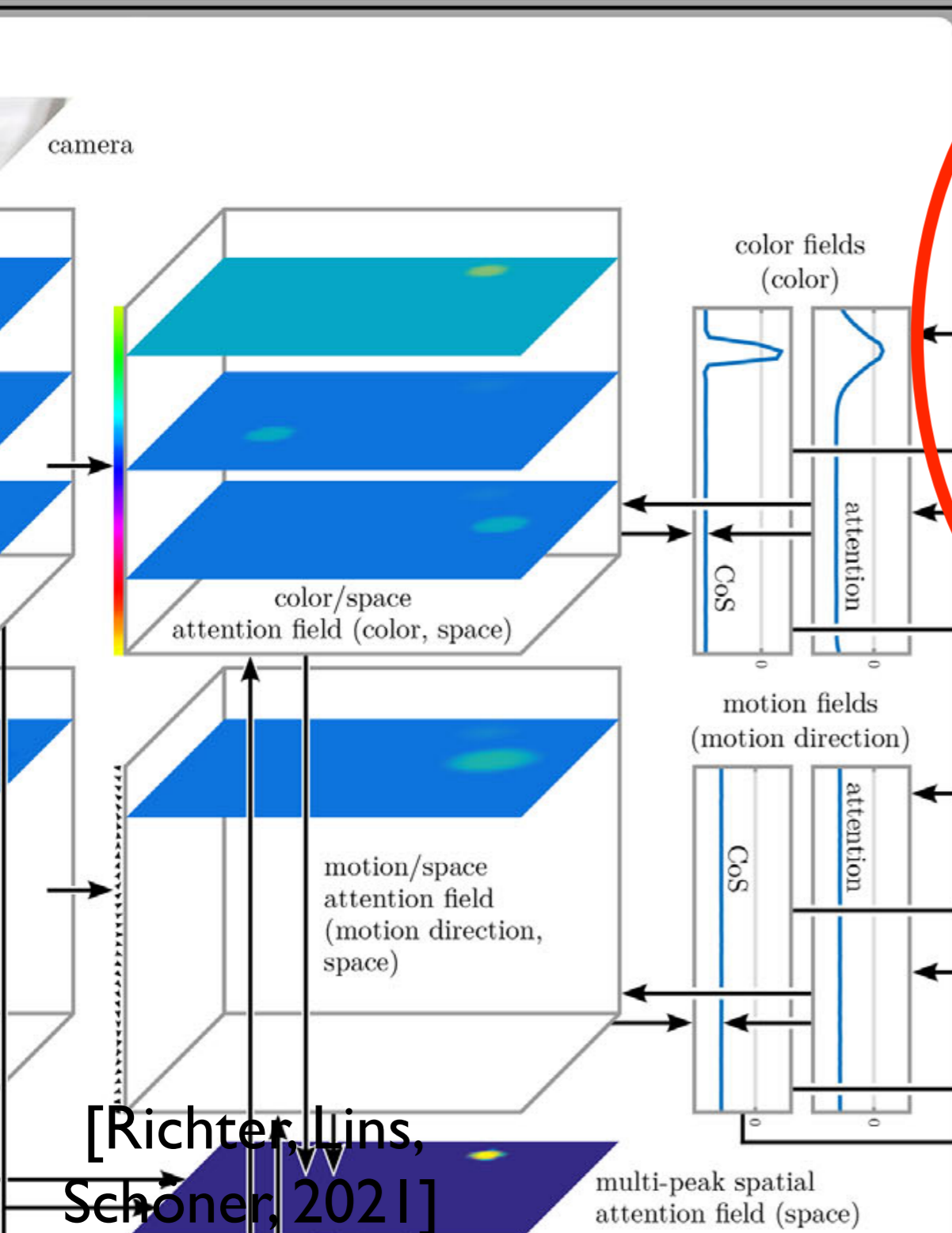
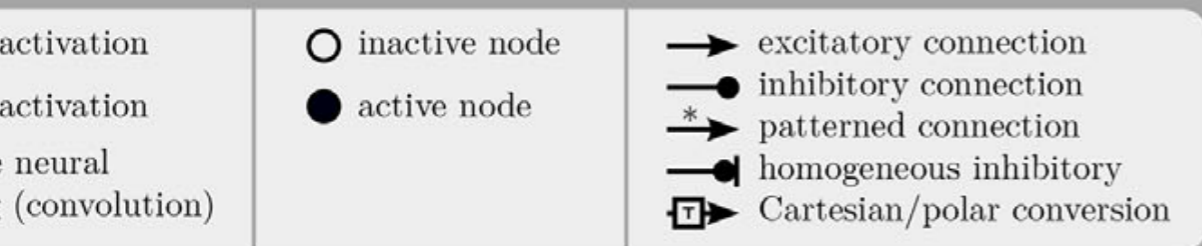
[Richter, Lins,  
Schöner, 2021]



grounding of relational concepts

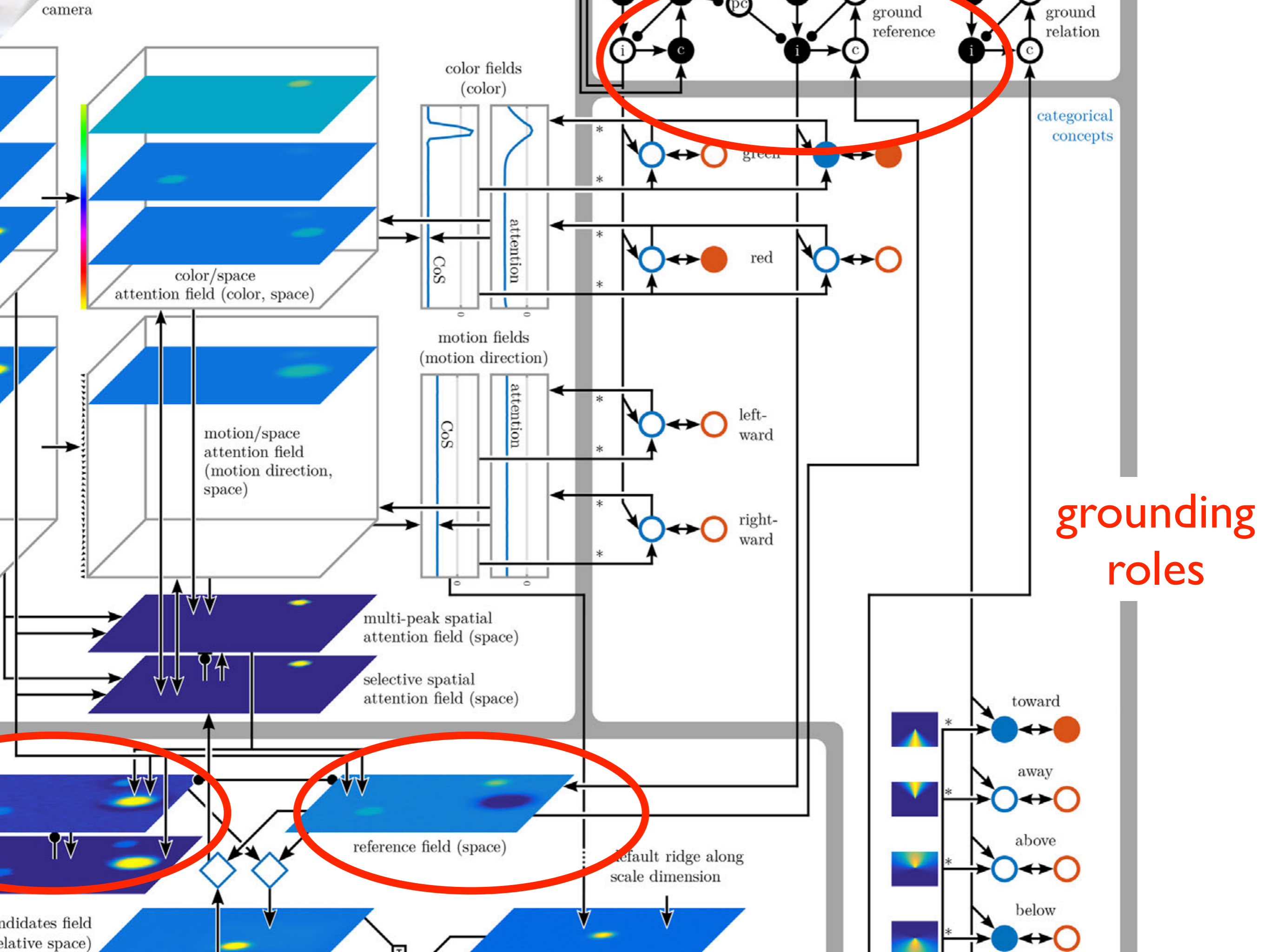
[Richter, Lins, Schöner, 2021]



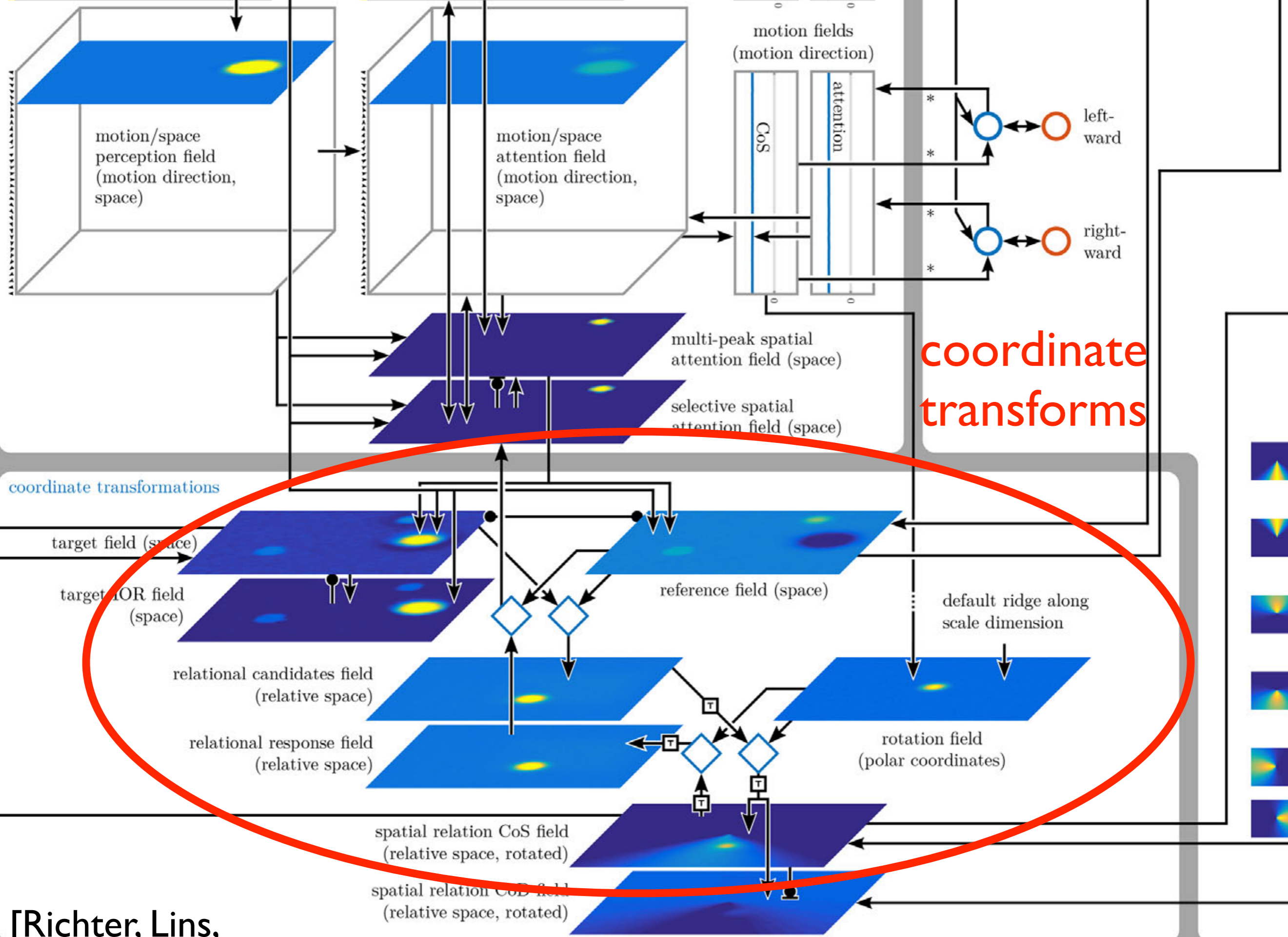


categorical concepts

role filler binding

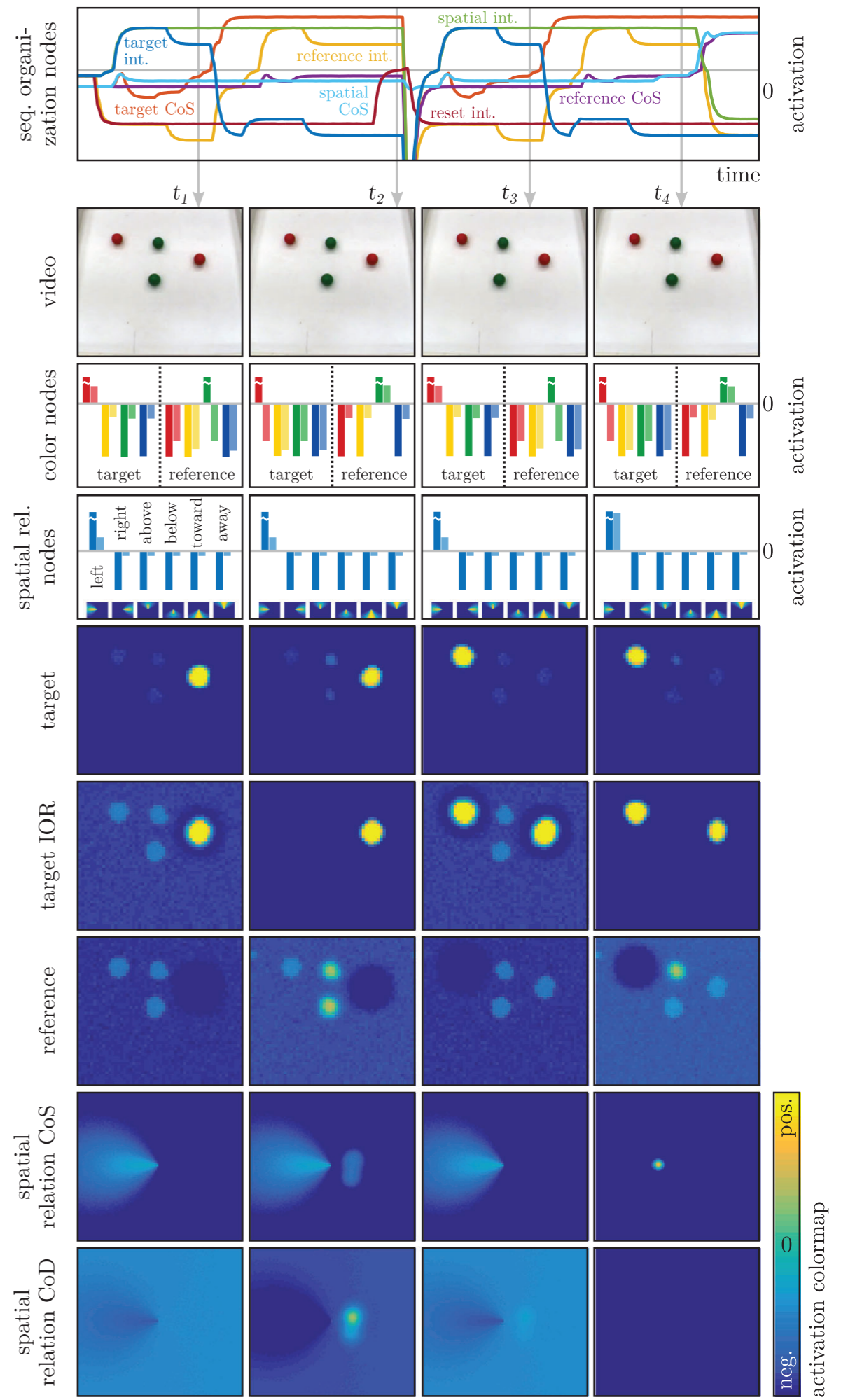






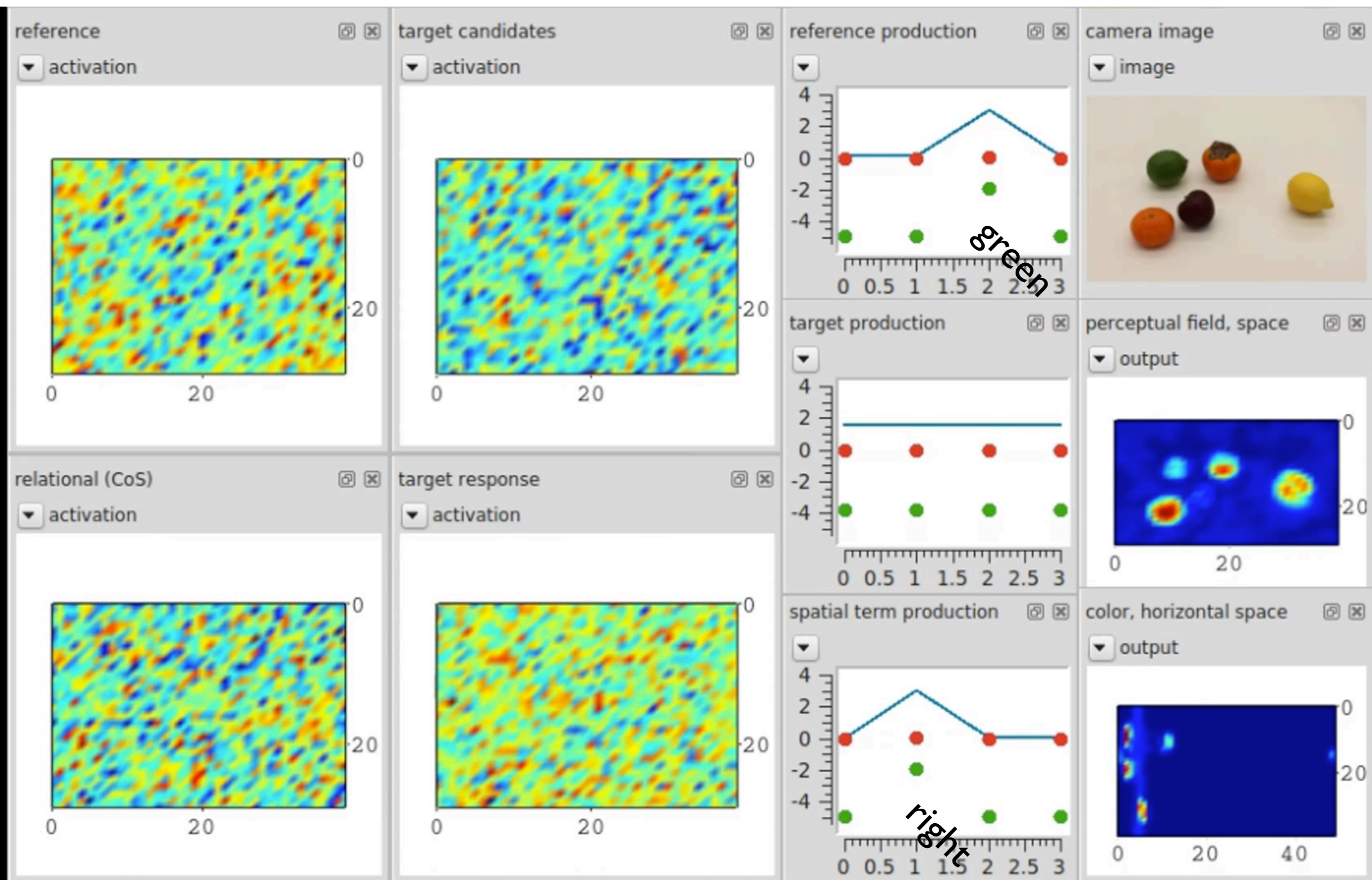
[Richter, Lins, Schöner, 2021]

“the red object to the left of the green object”



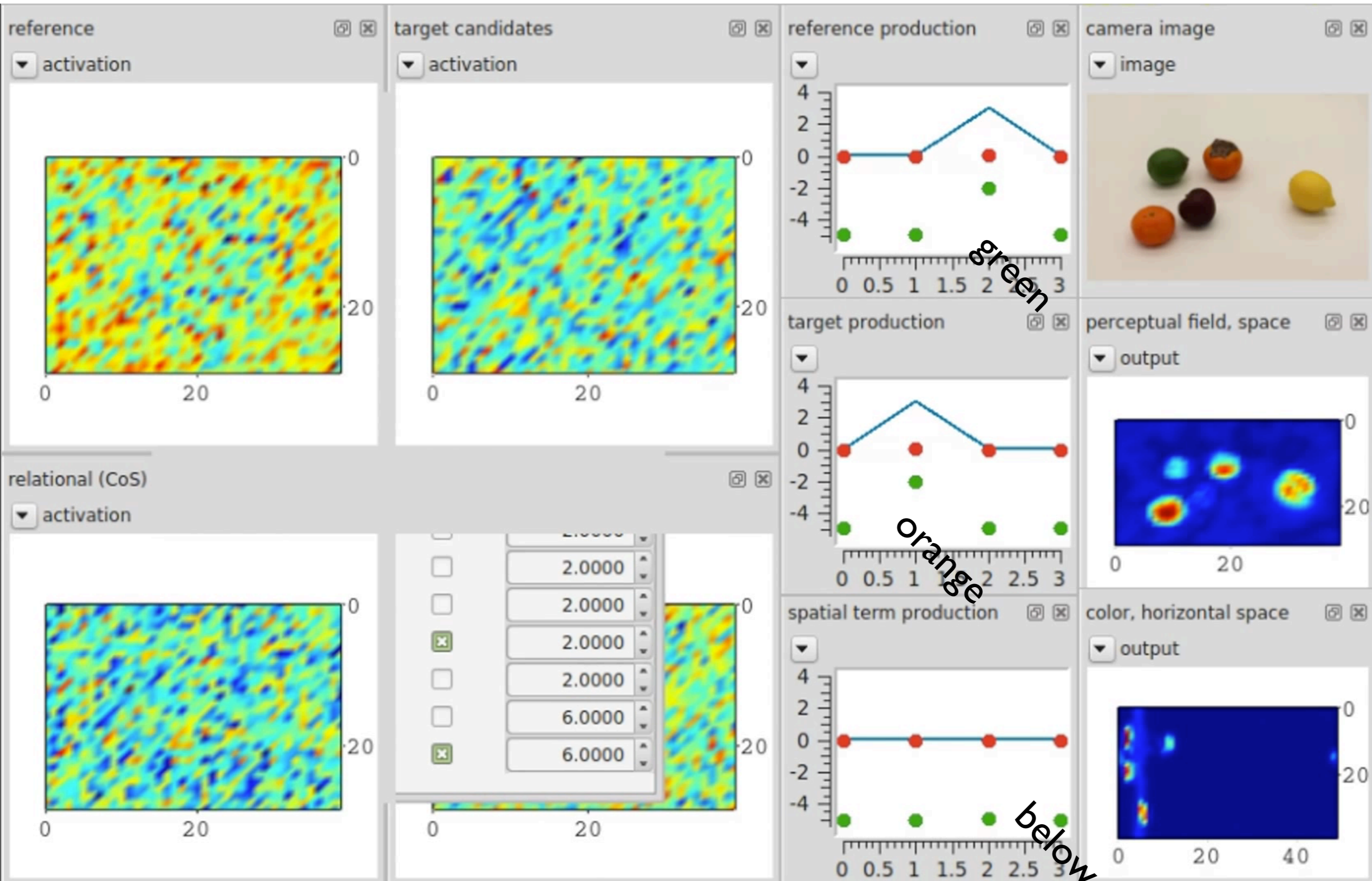


■ what is to the right of the green object?

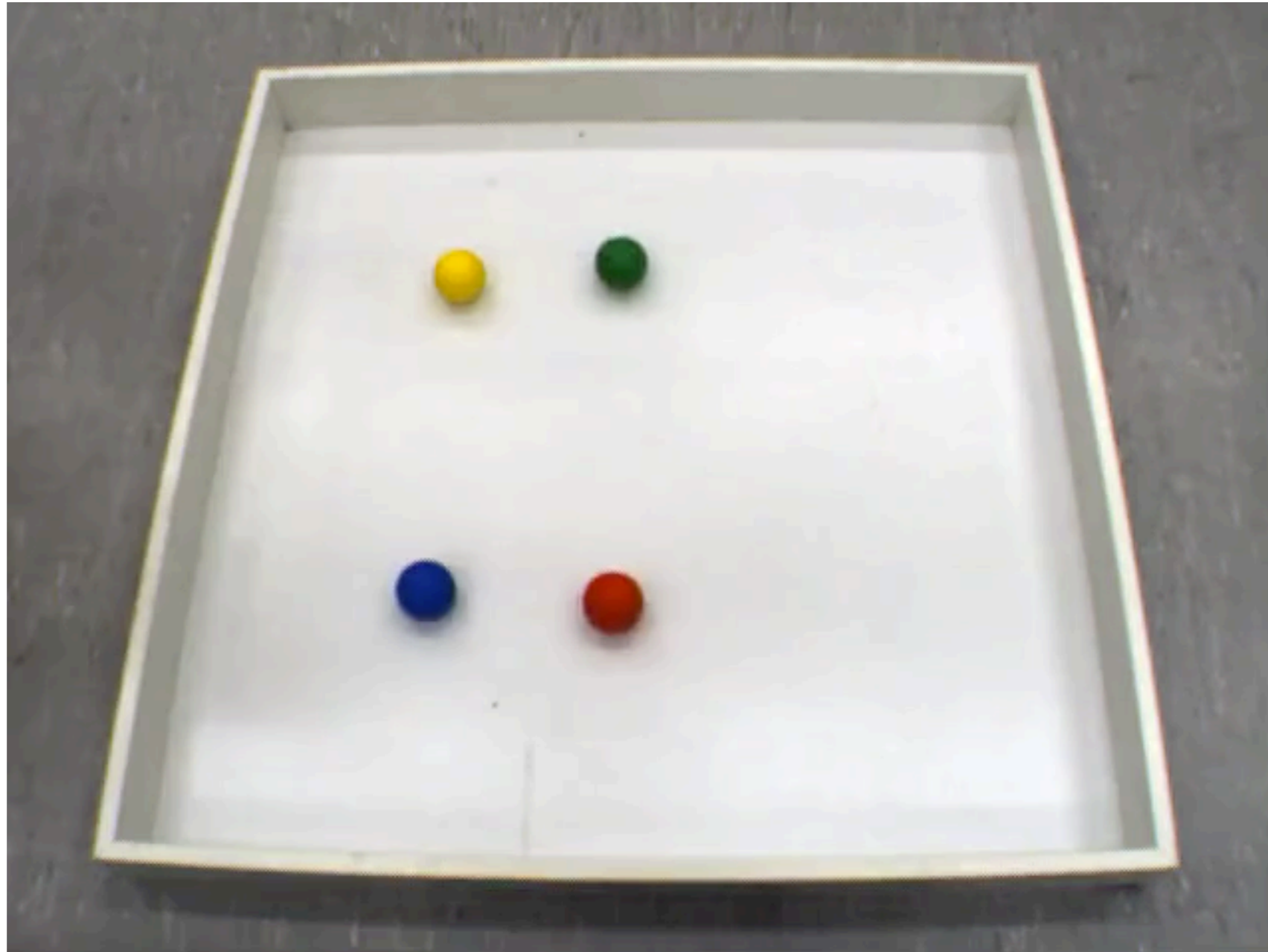




■ where is the orange relative to the green object



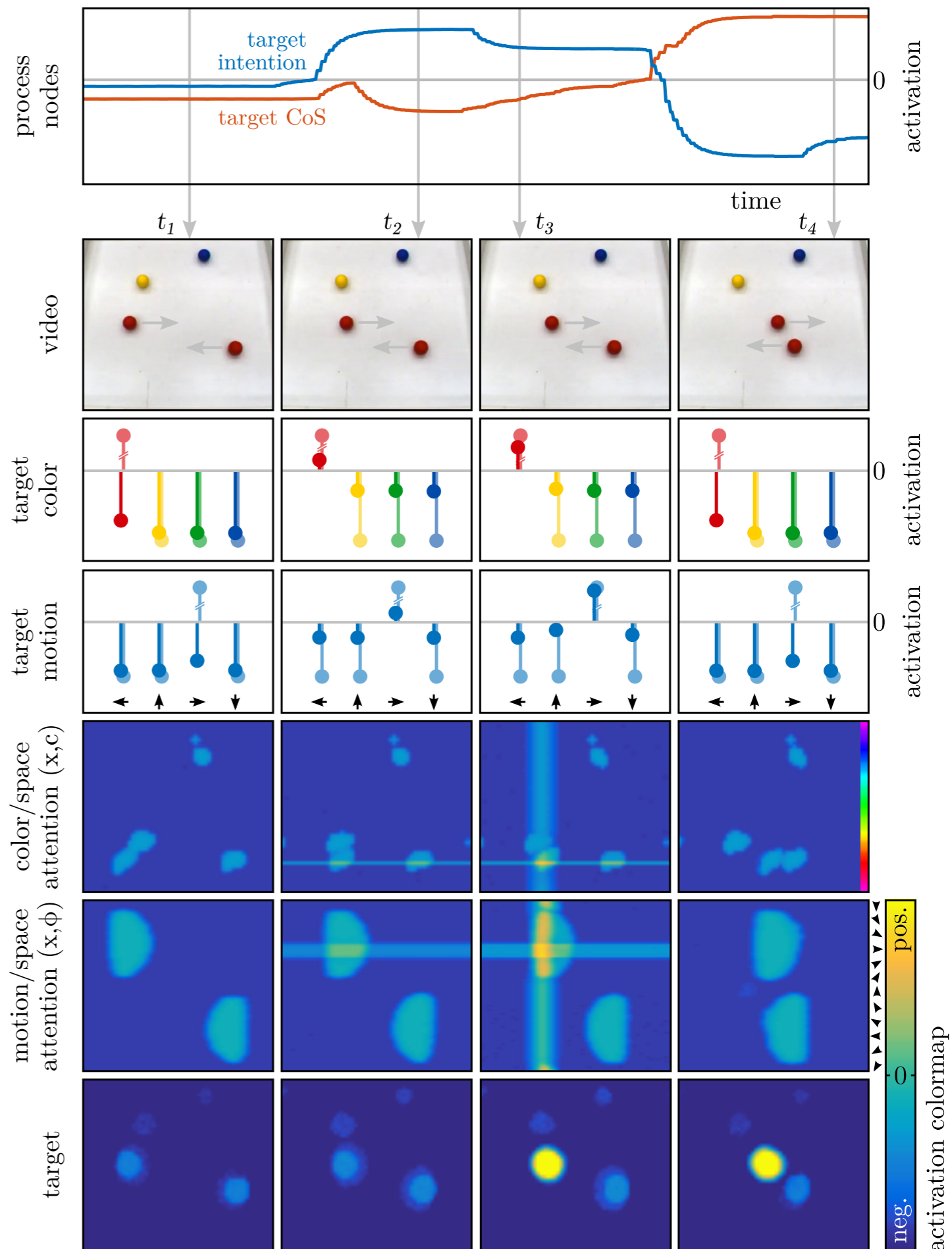
# Grounding movement relations



“the red  
moving to the  
right”

perceptual  
grounding

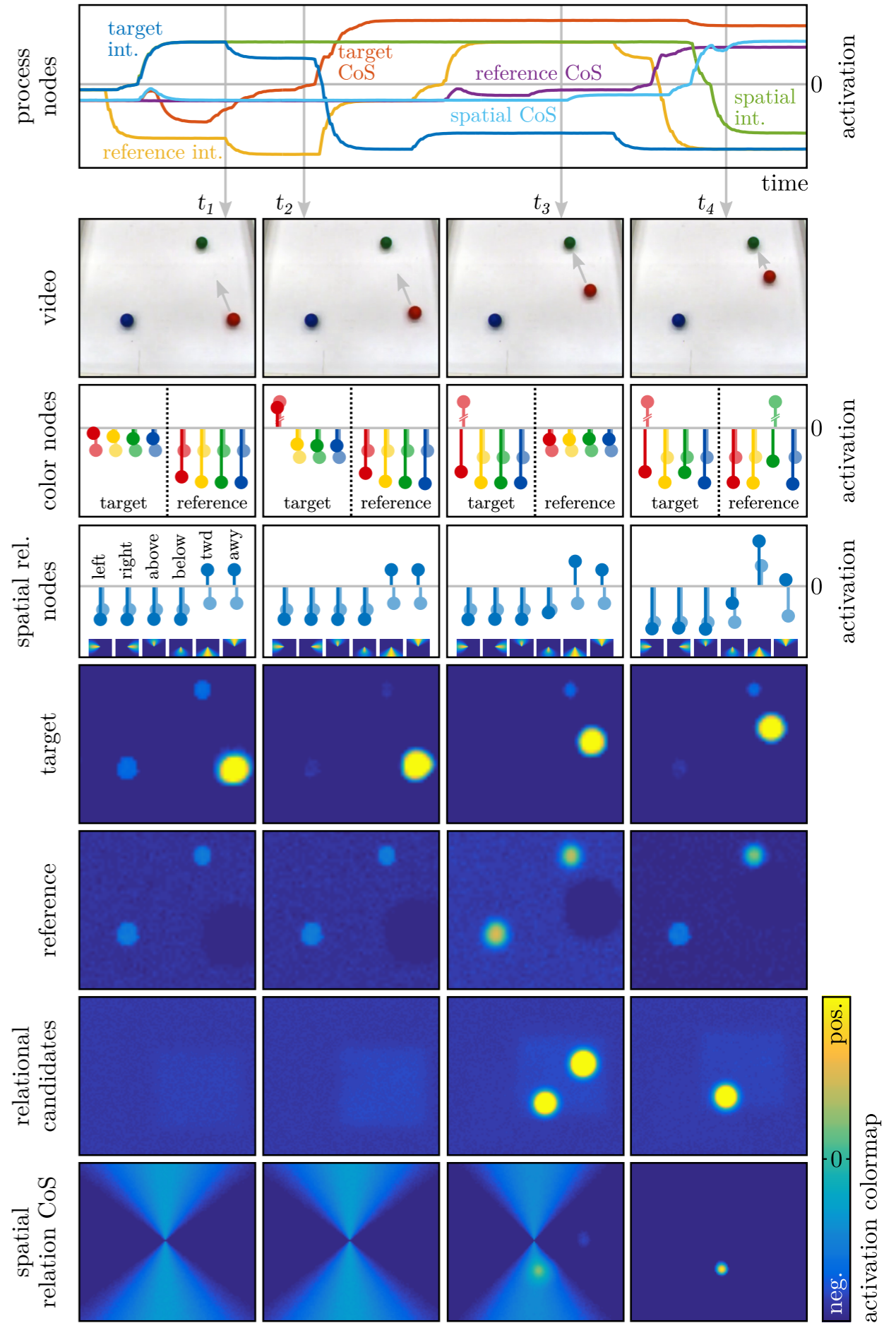
[Richter et al]





# description

[Richter et al]

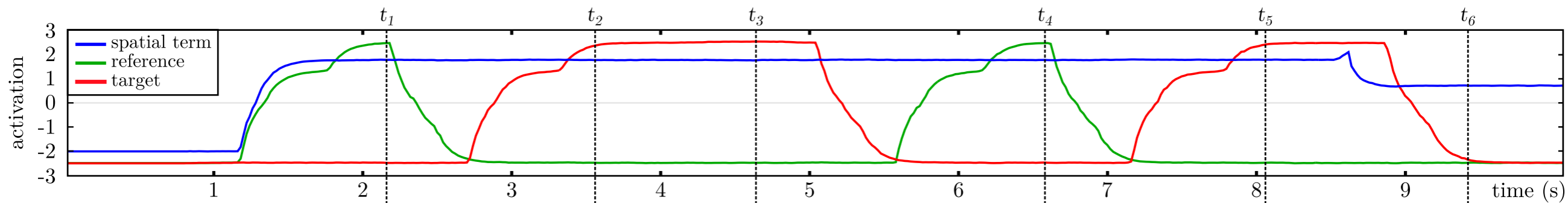


# Autonomous hypothesis testing

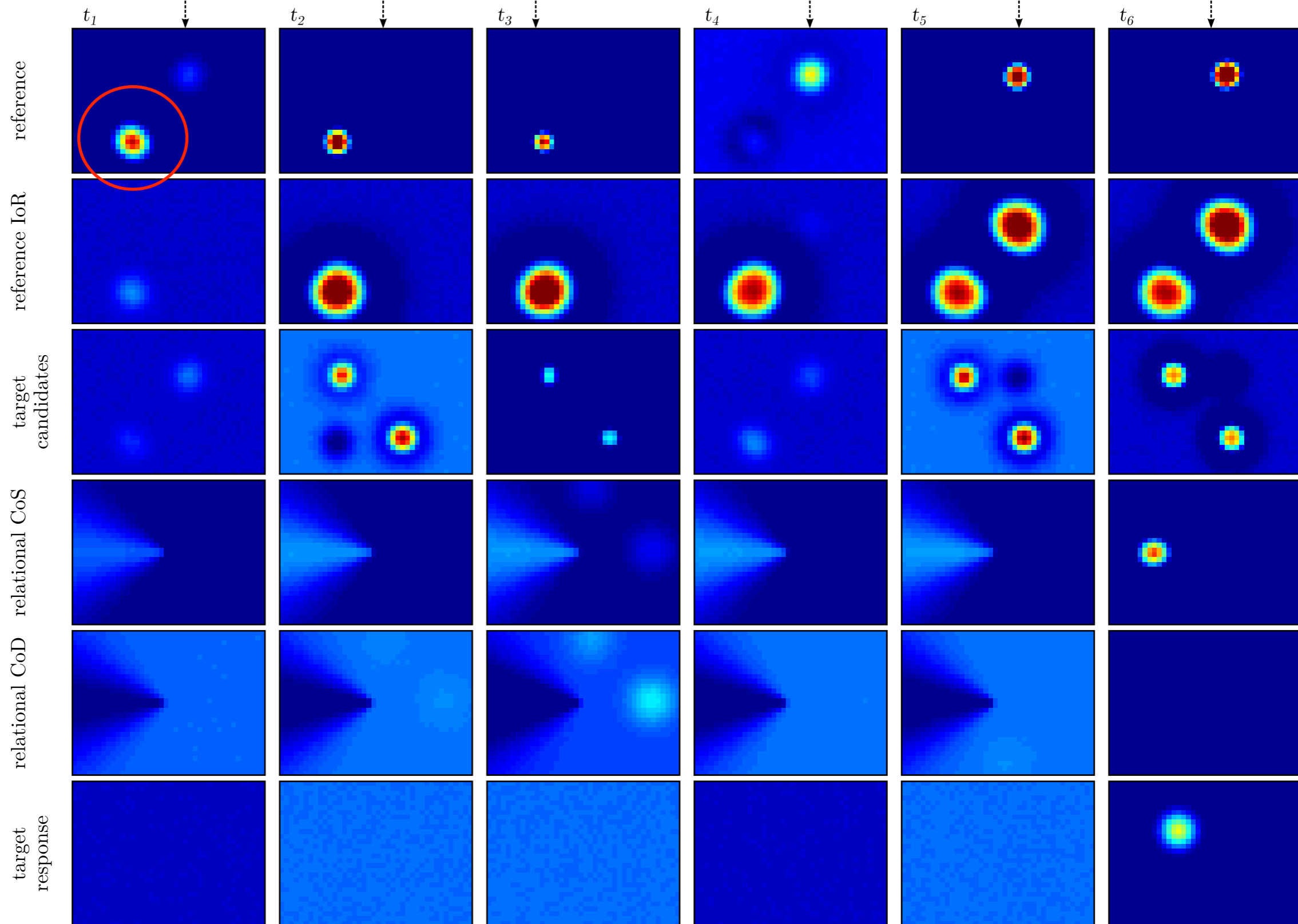
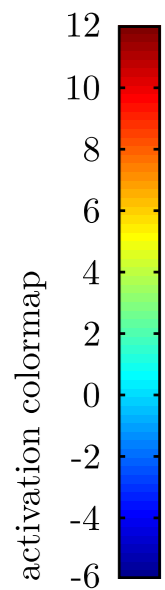


“the red cup that is to the left of the green cup”

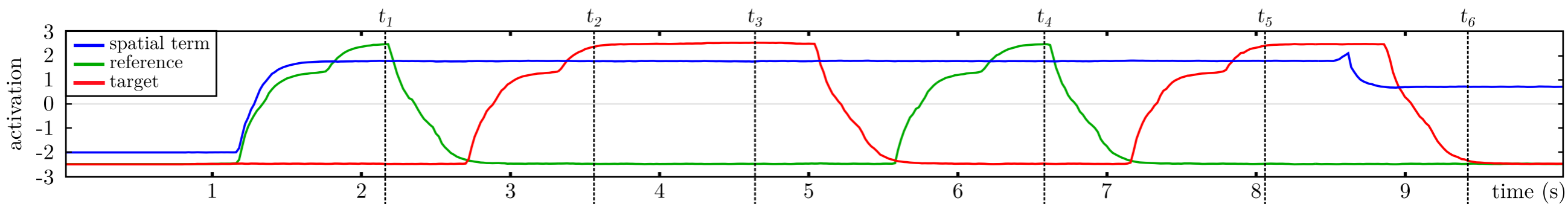
[Richter, Lins et al, CogSci 2014]



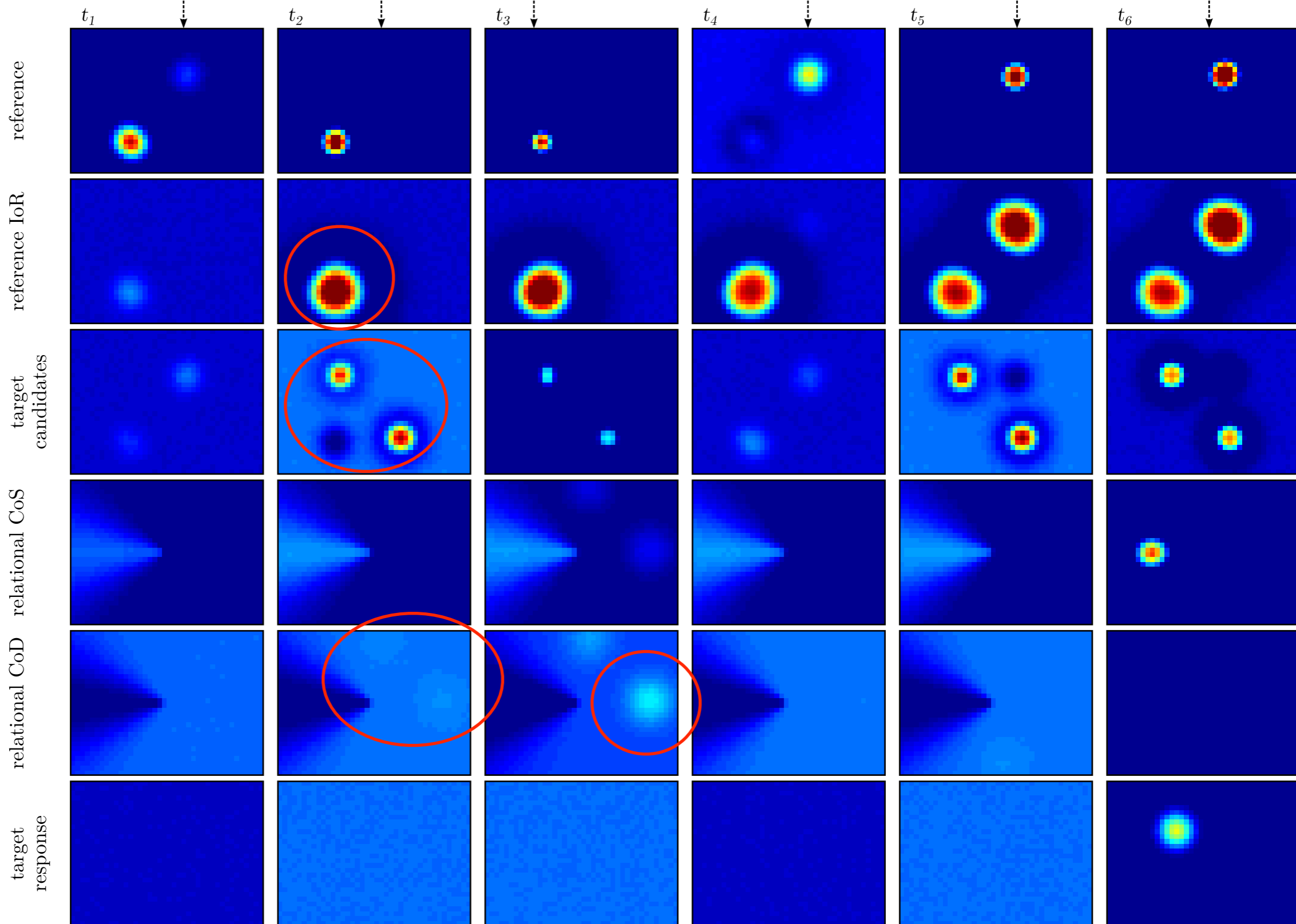
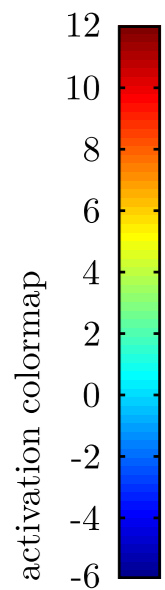
“the red cup  
that is to the  
left of the  
green cup”

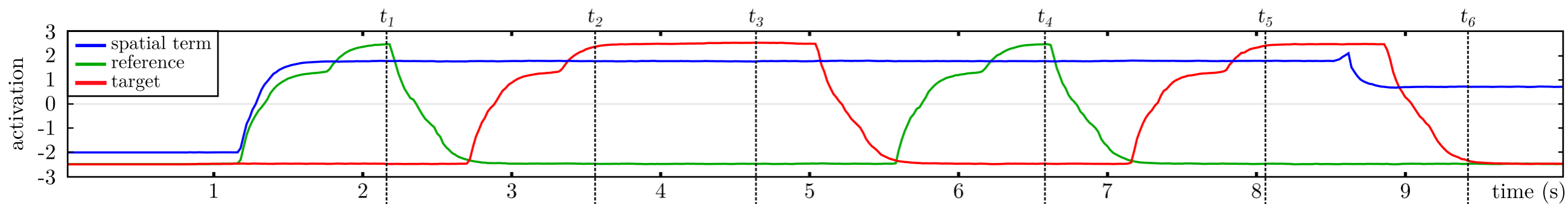




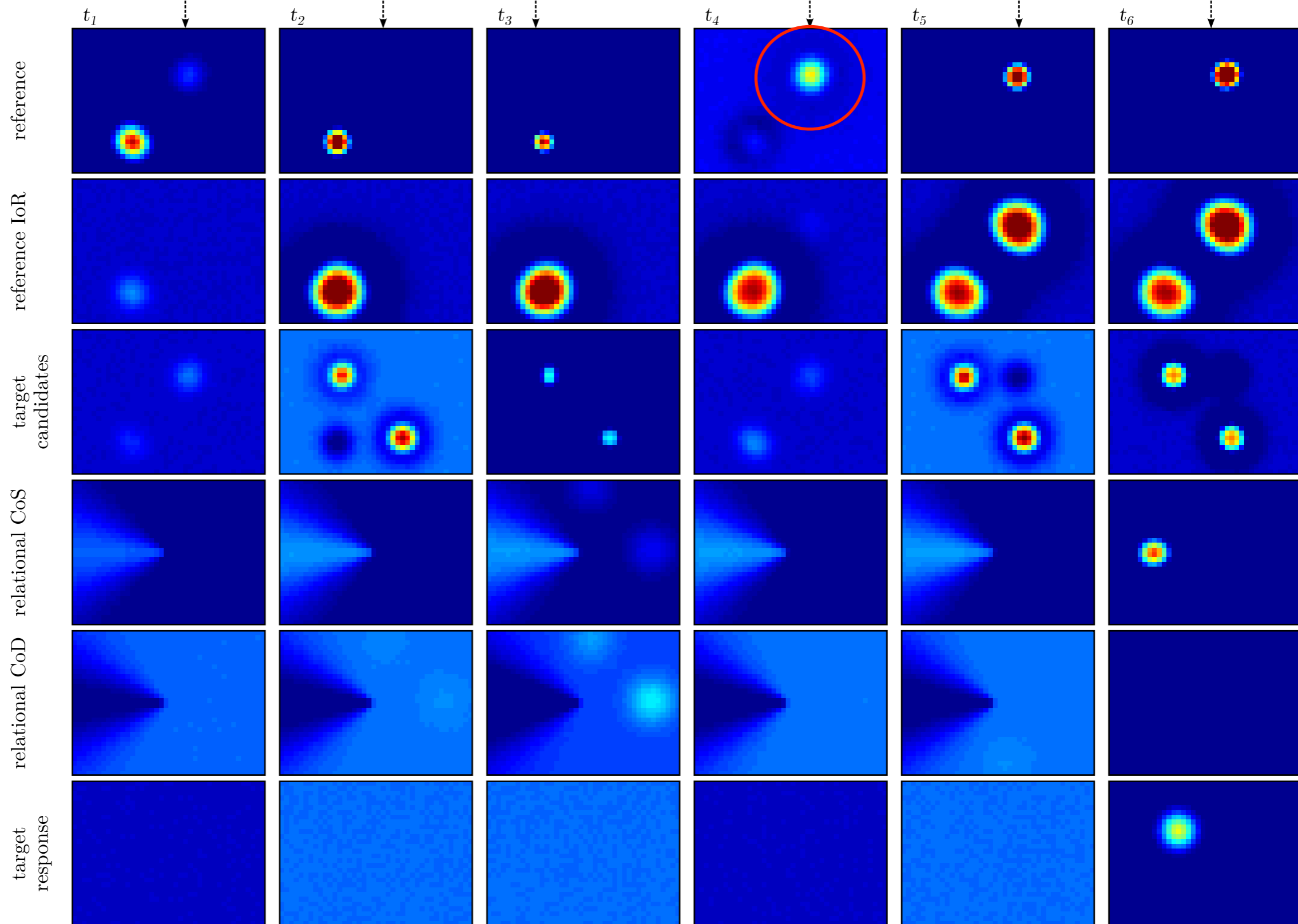
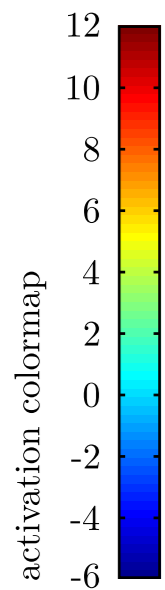


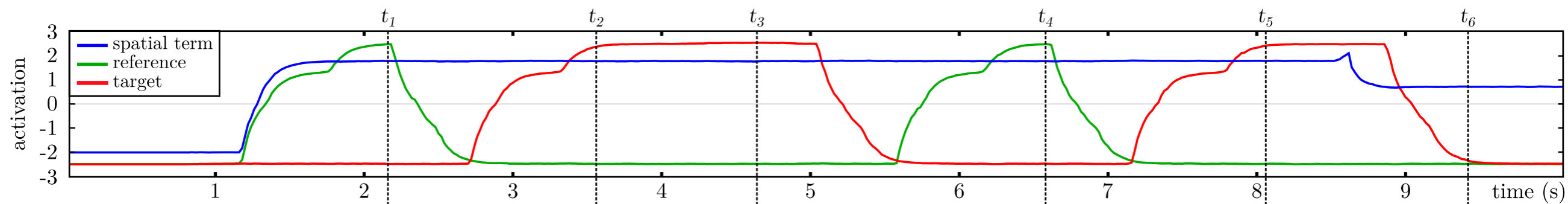
“the red cup  
that is to the  
left of the  
green cup”



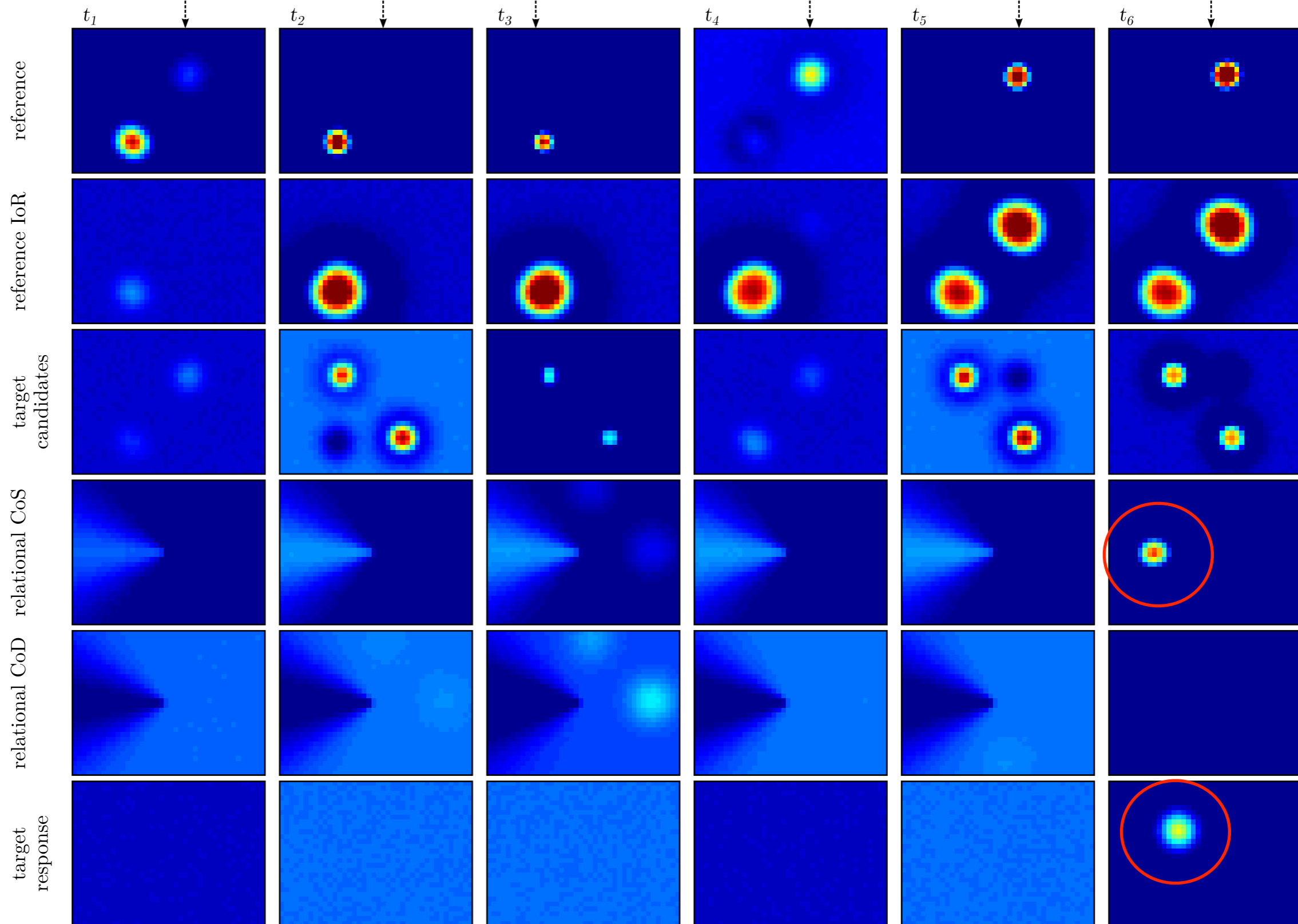
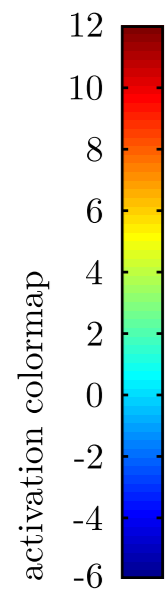


“the red cup  
that is to the  
left of the  
green cup”



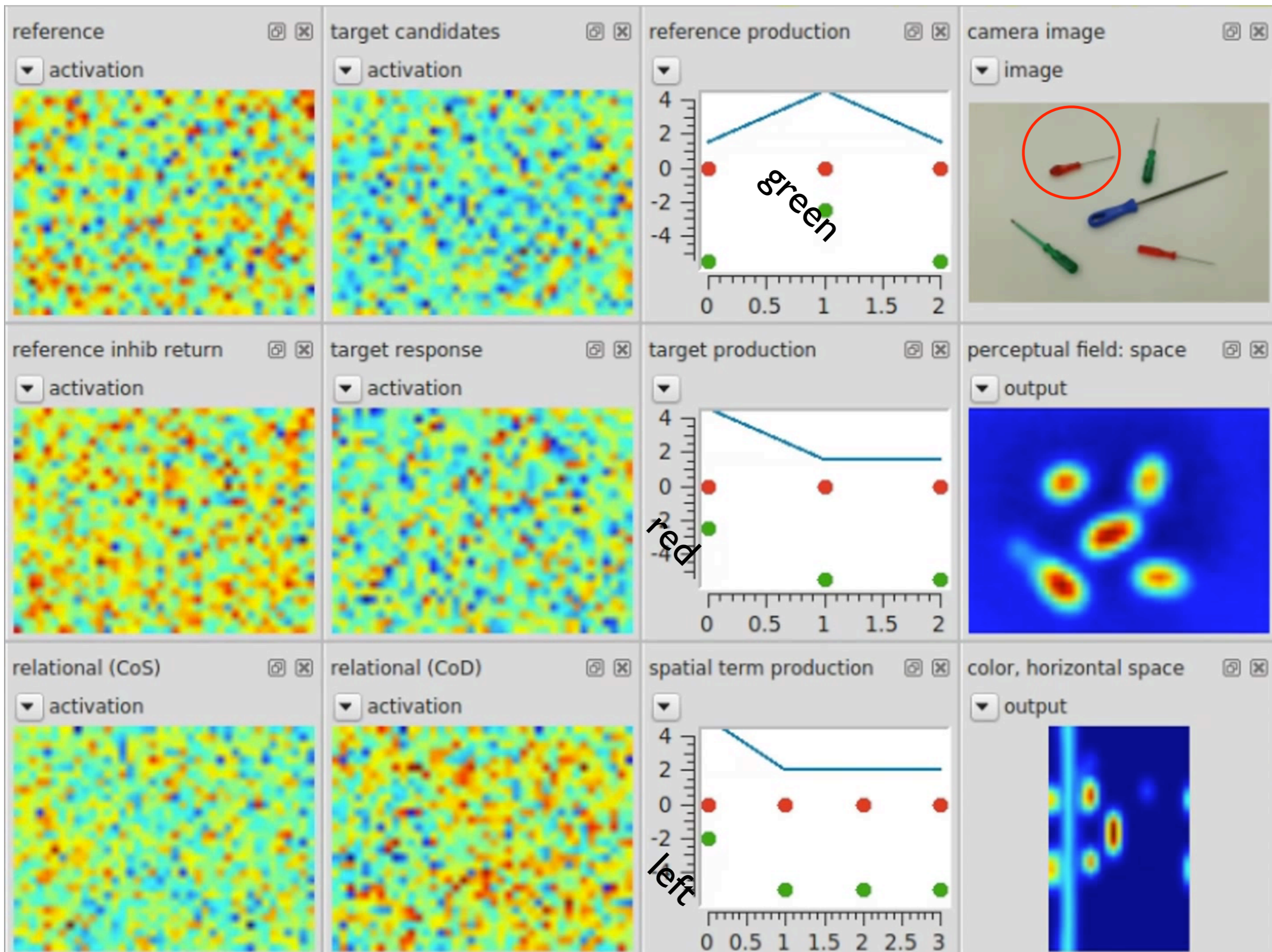


“the red cup  
that is to the  
left of the  
green cup”





“the red to the left of the green”



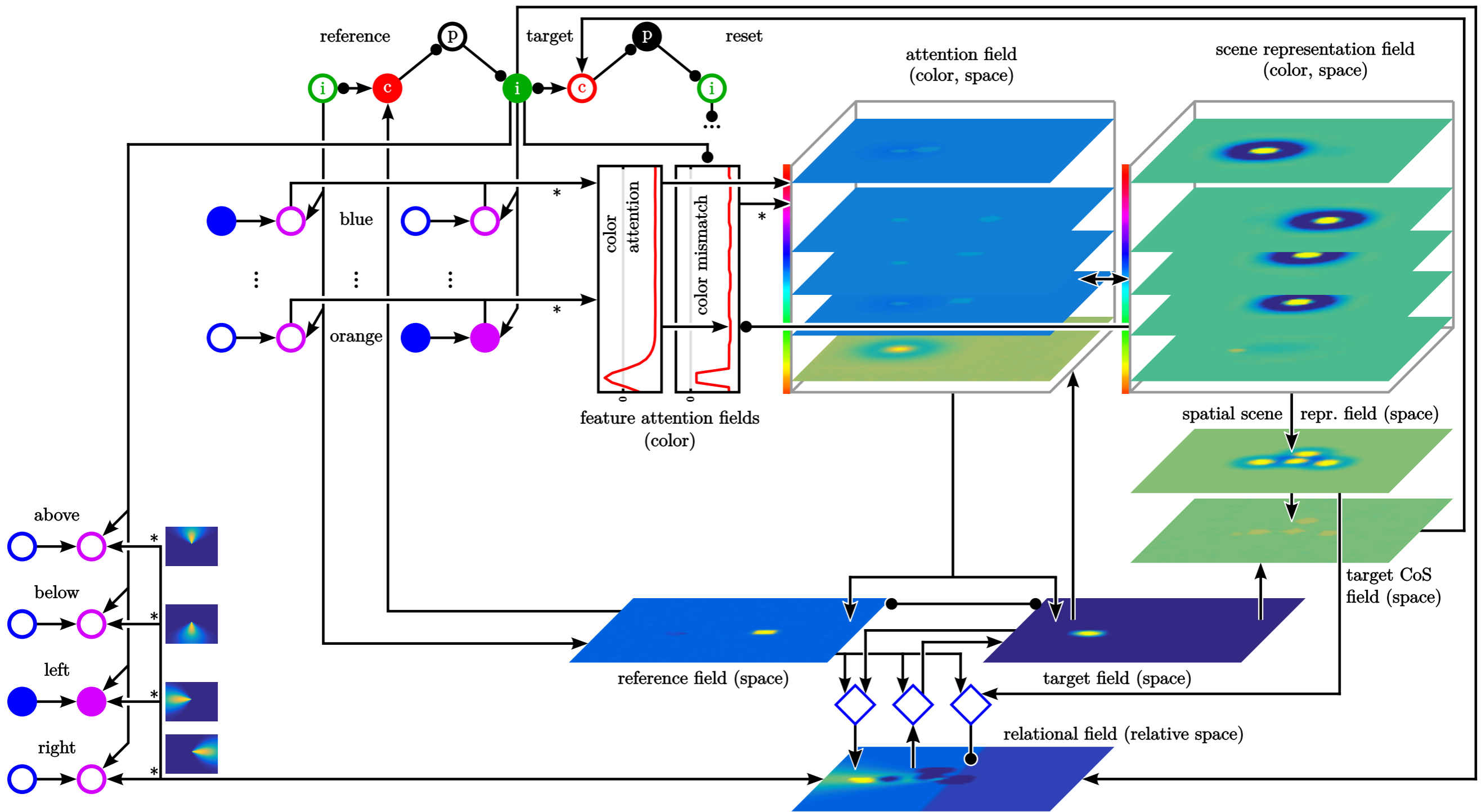
# Mental mapping and inference

## ■ propositions

- “There is a cyan object above a green object.”
- “There is a red object to the left of the green object.”
- “There is a blue object to the right of the red object.”
- “There is an orange object to the left of the blue object.”

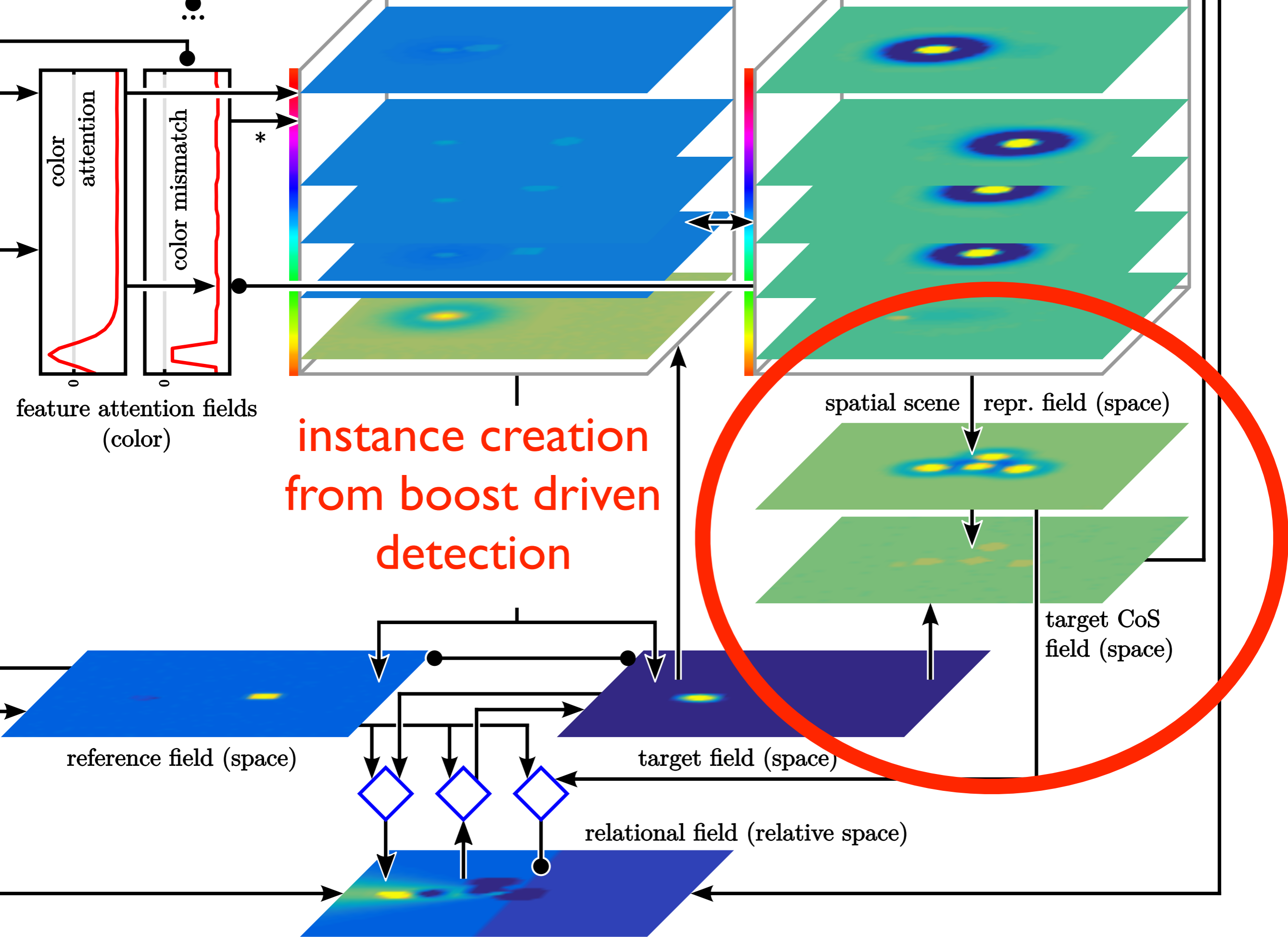
## ■ inference

- “Where is the blue object relative to the red object?”



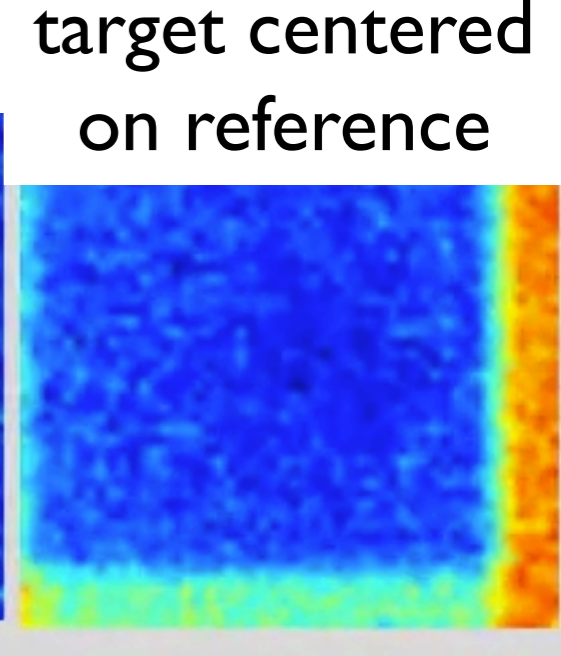
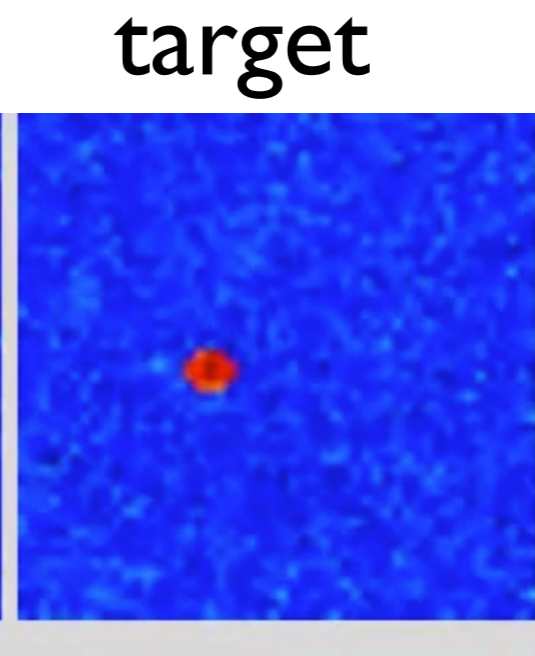
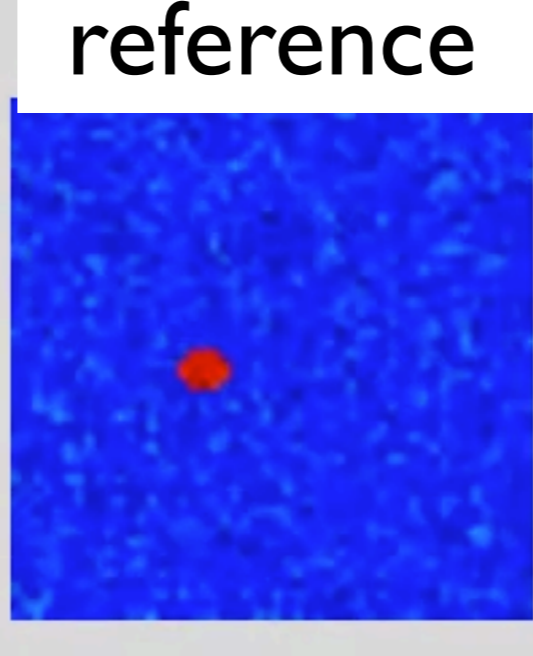
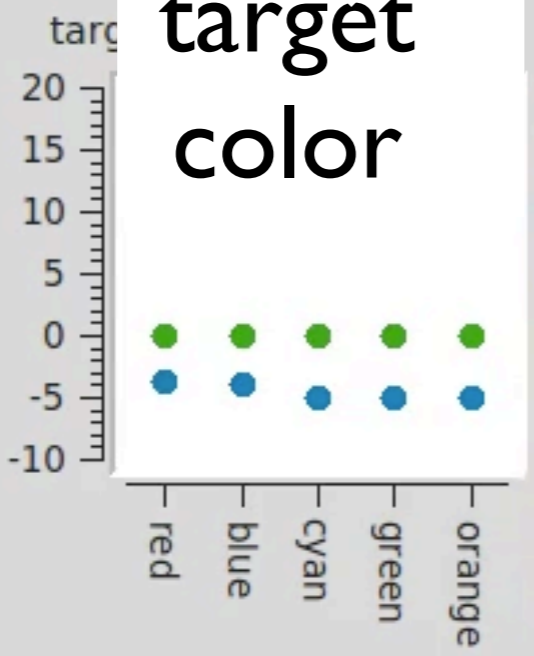
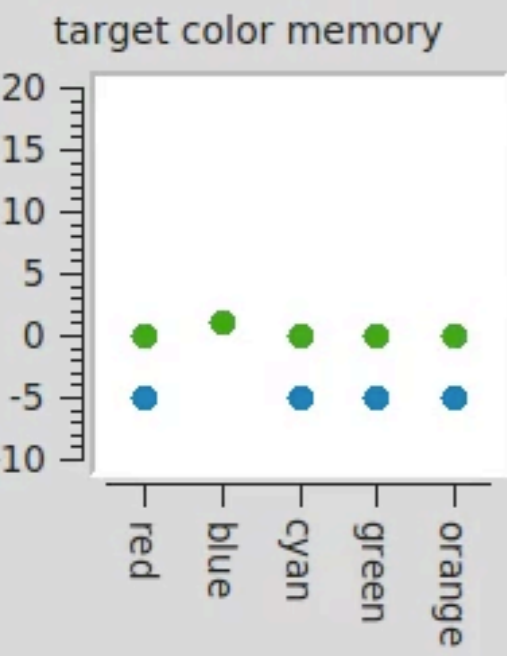
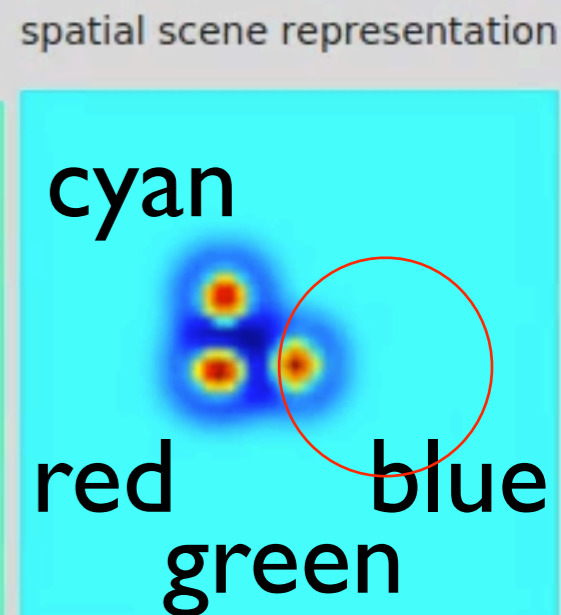
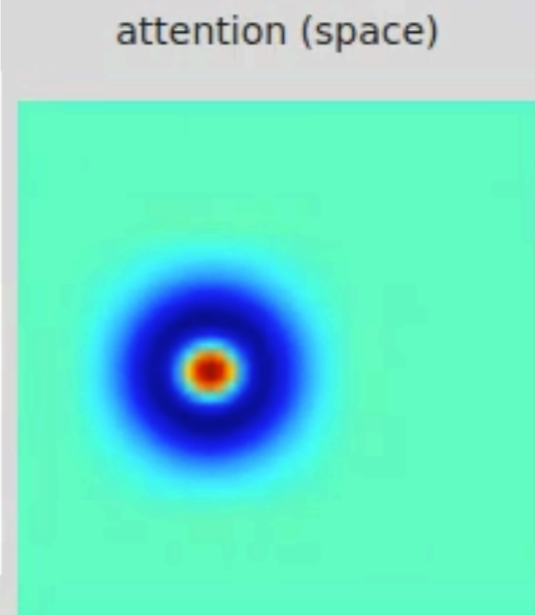
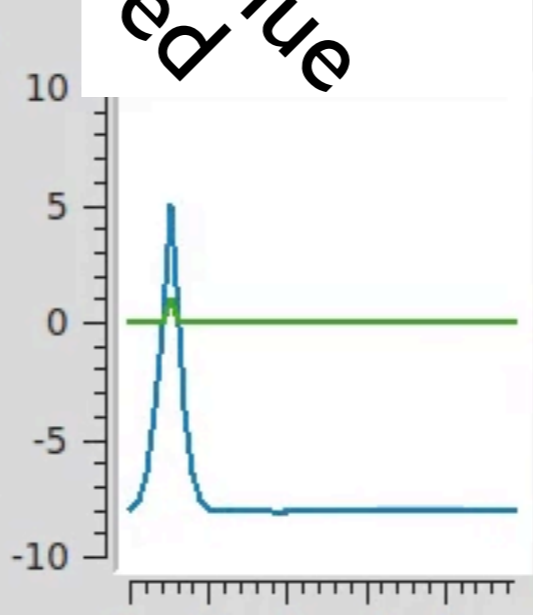
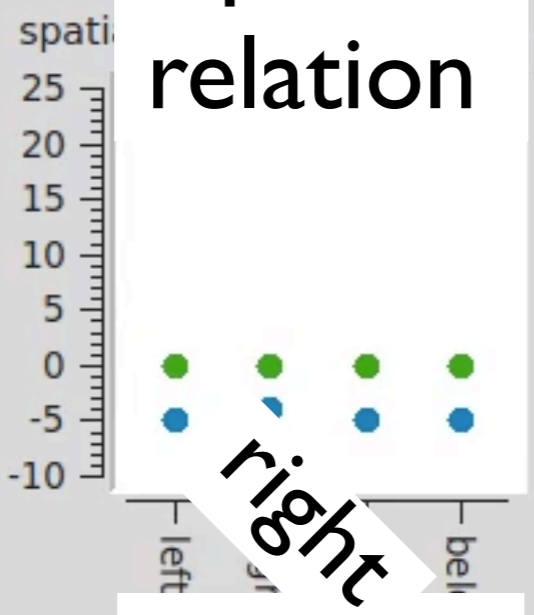
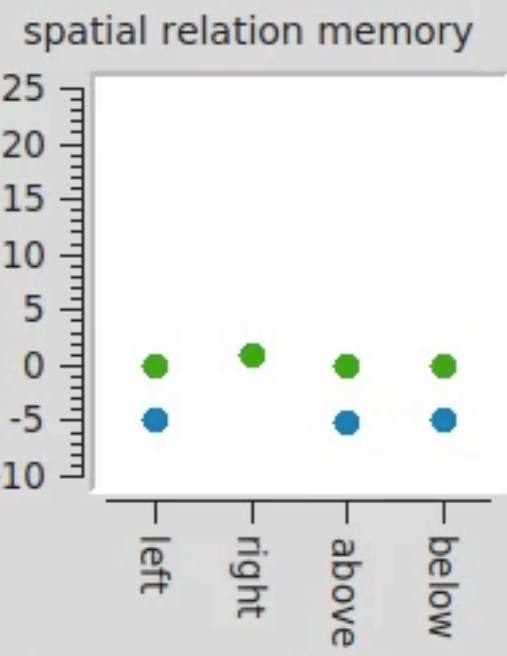
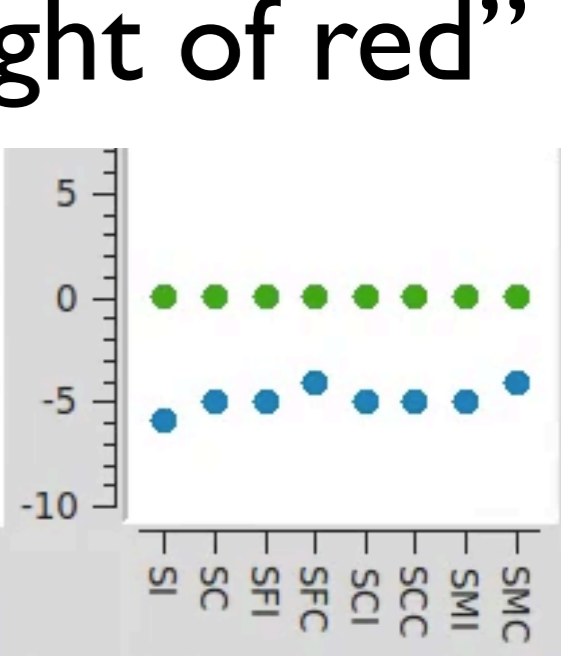
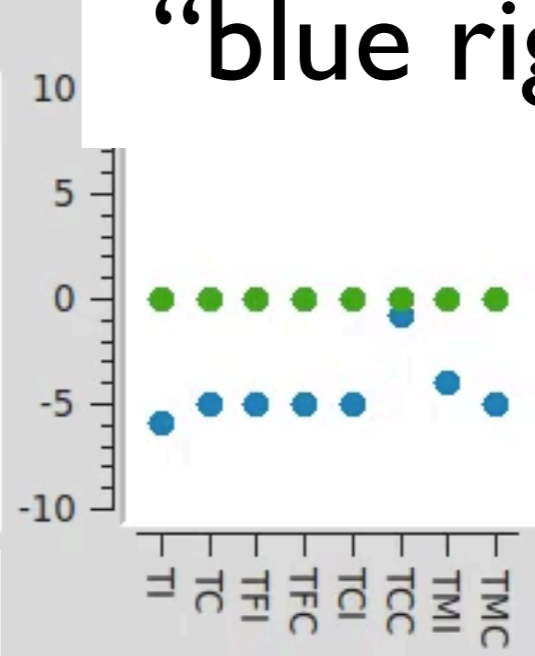
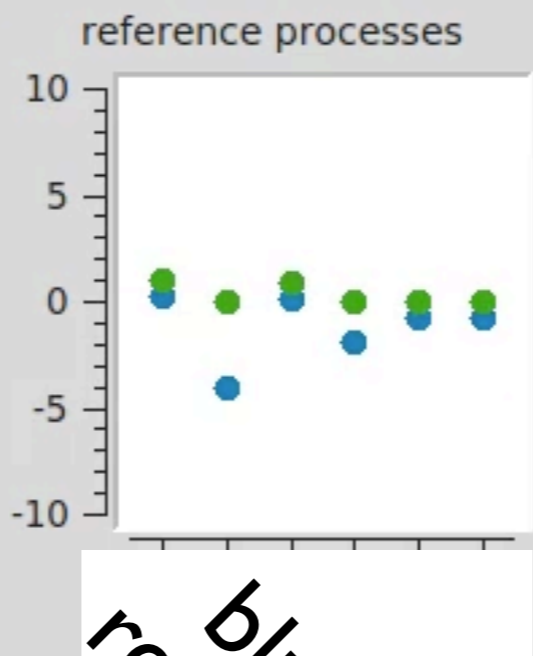
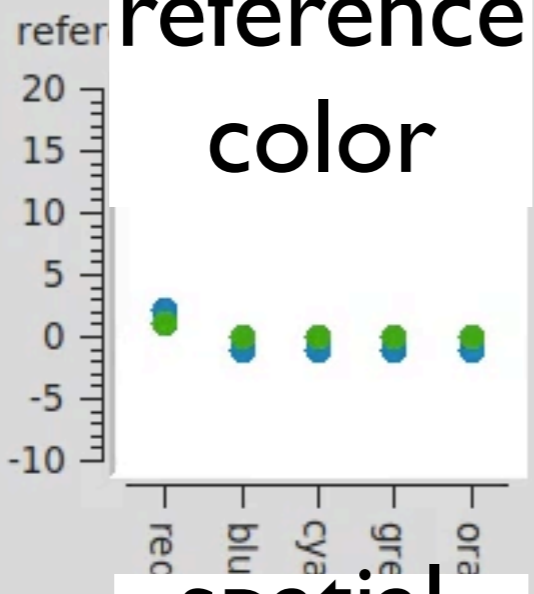
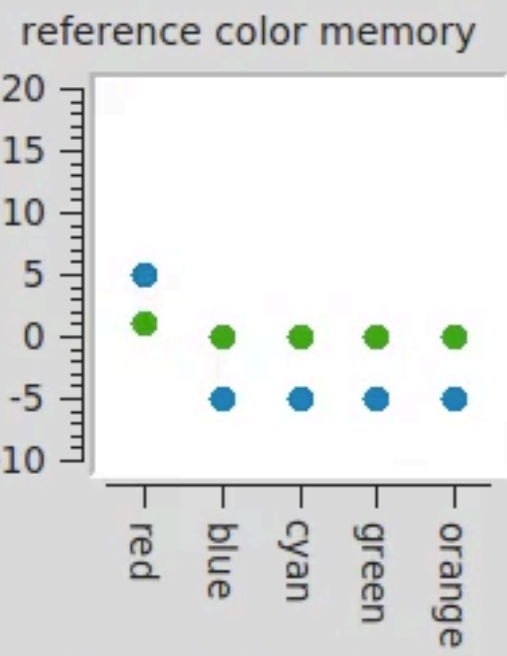
[Kounatidou, Richter, Schöner, CogSci 2018]







“blue right of red”



reference color

spatial relation

target color

right

reference

target

target centered on reference

cyan

red

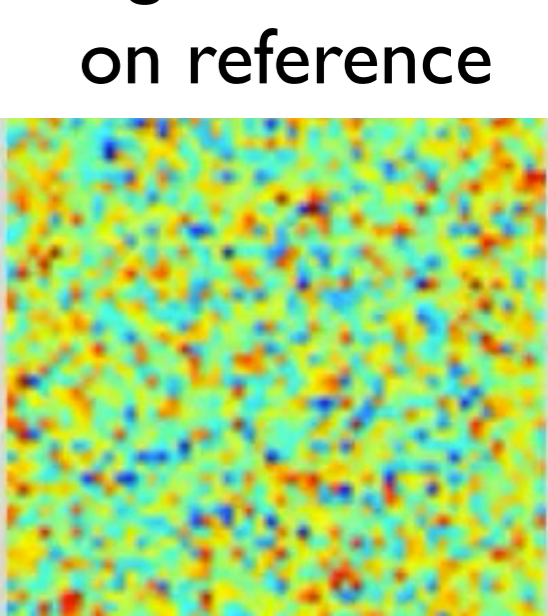
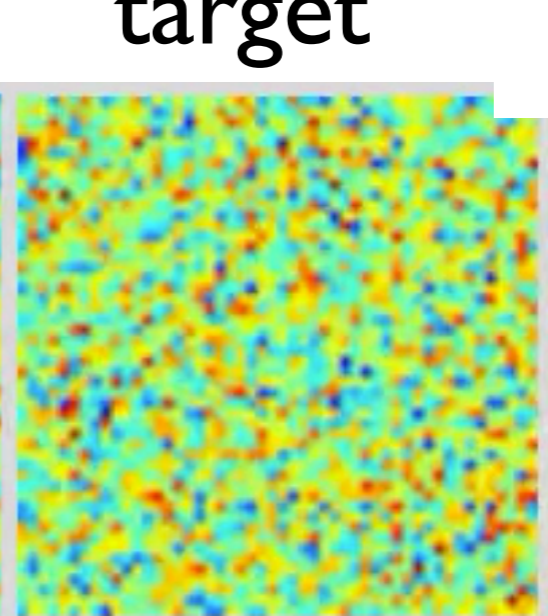
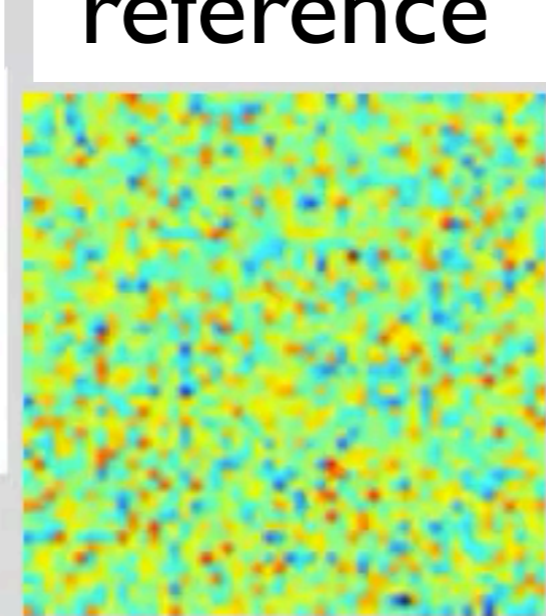
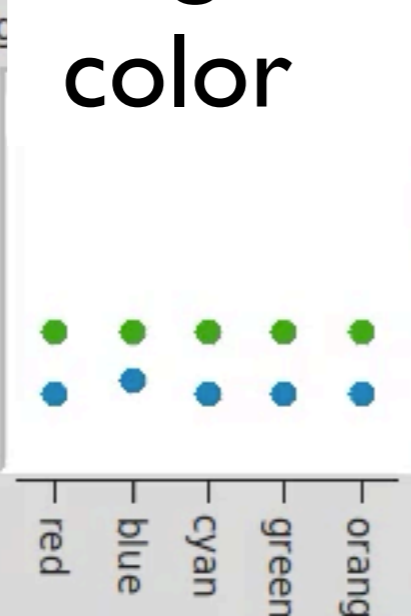
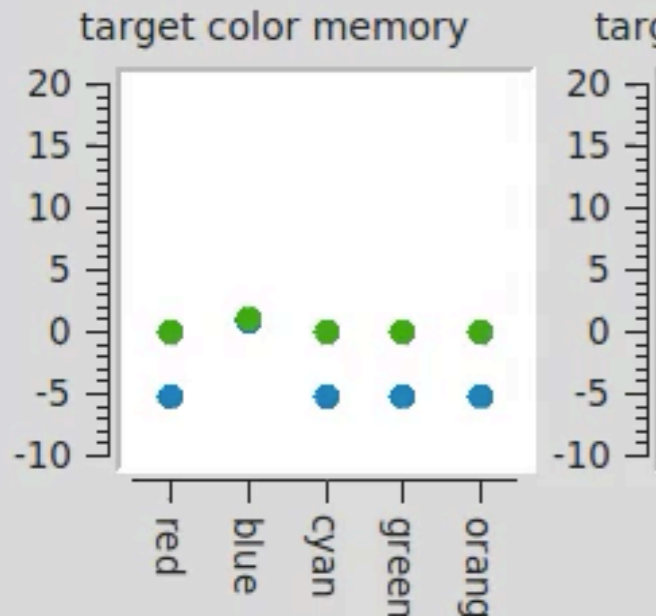
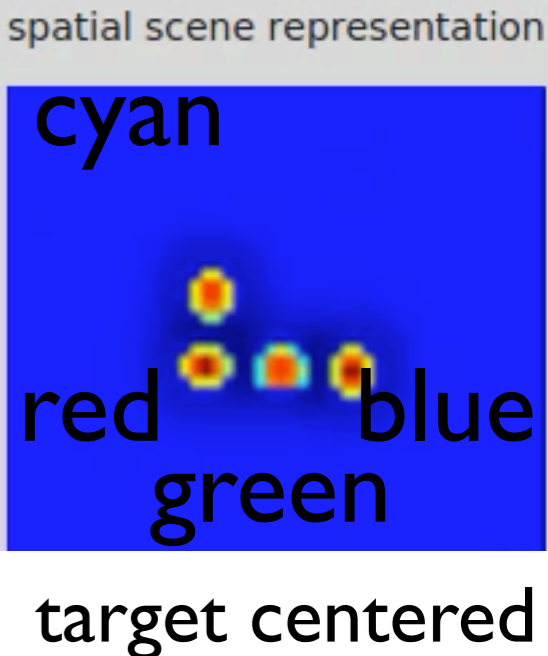
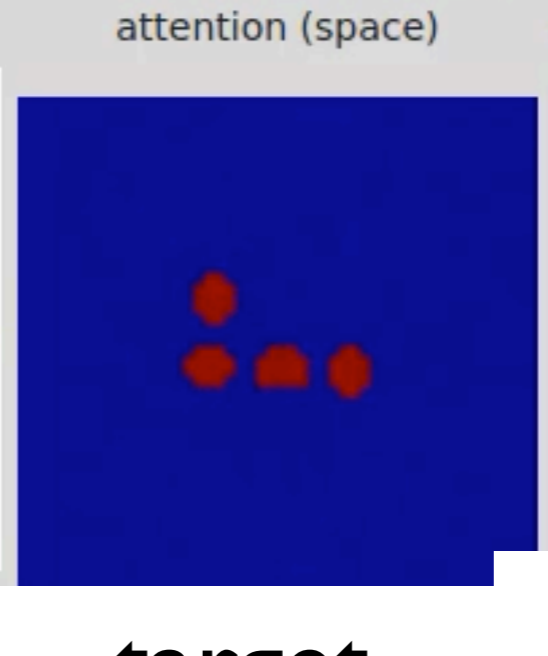
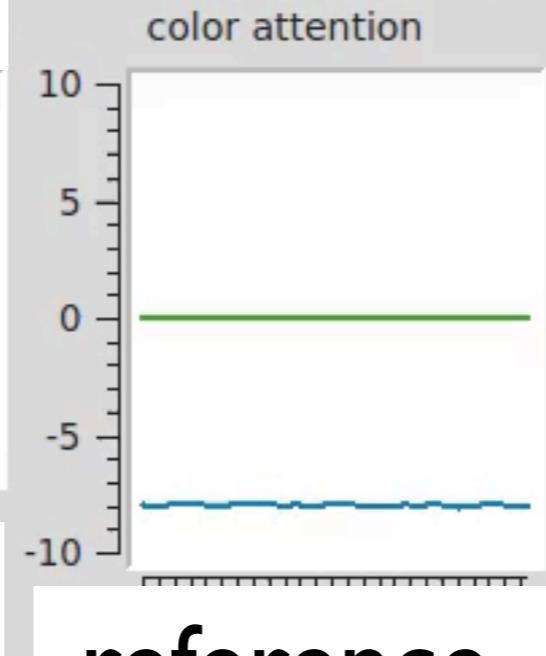
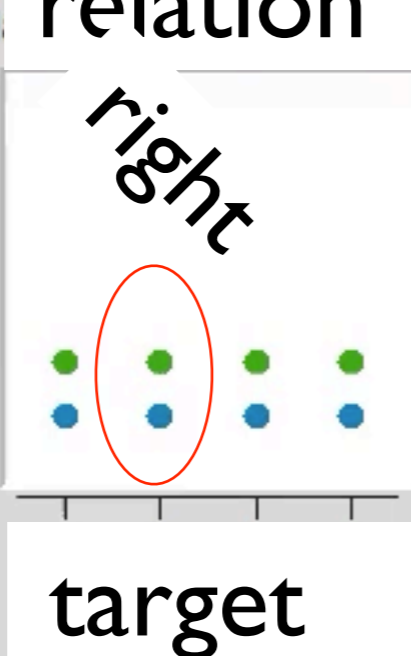
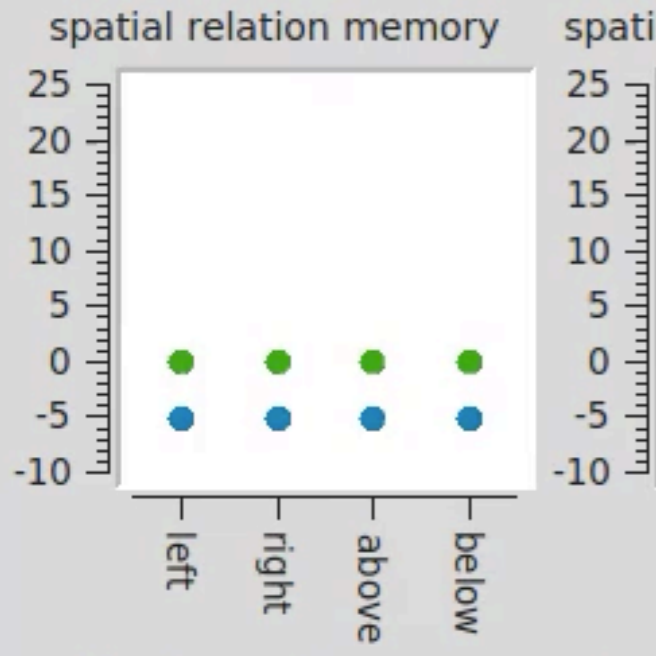
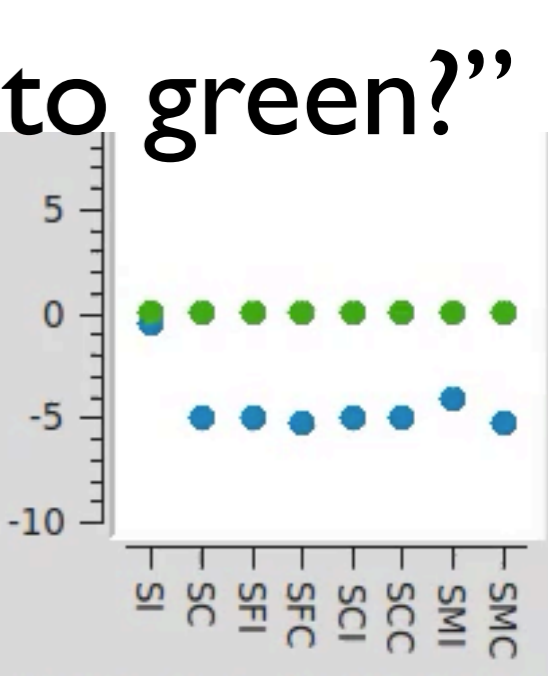
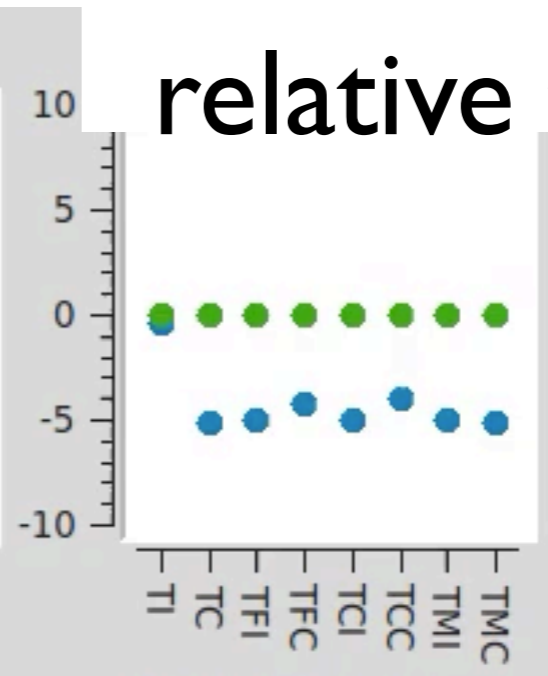
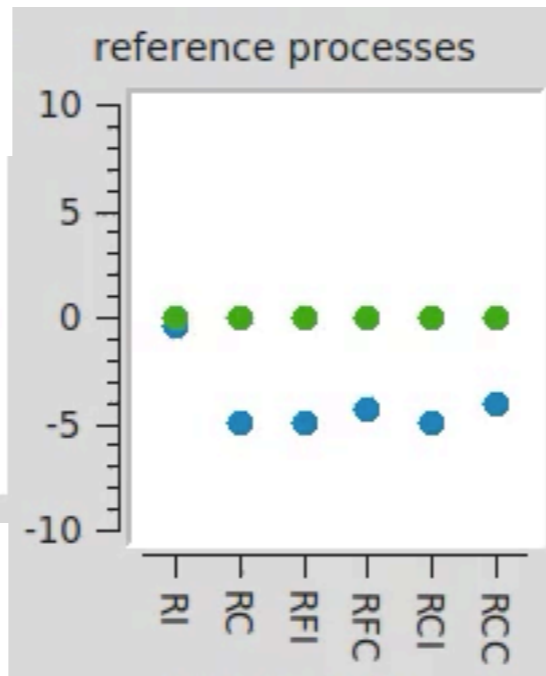
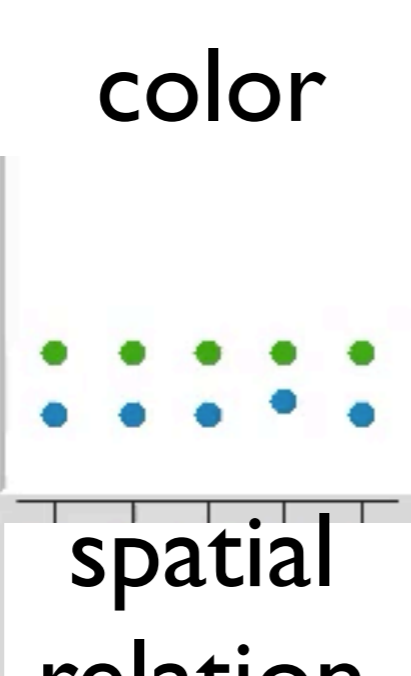
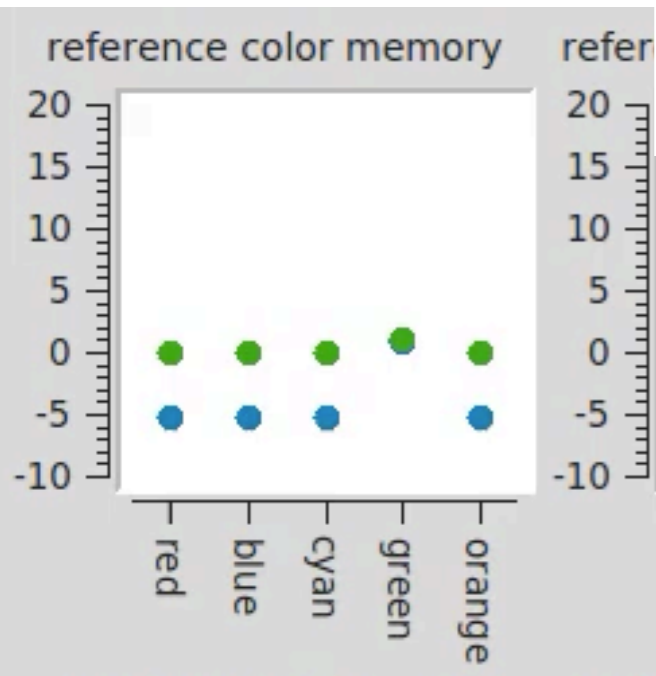
green

blue



reference  
color

“where is blue  
relative to green?”



reference

target

target centered  
on reference

# Conclusion

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