

# Grounding spatial language

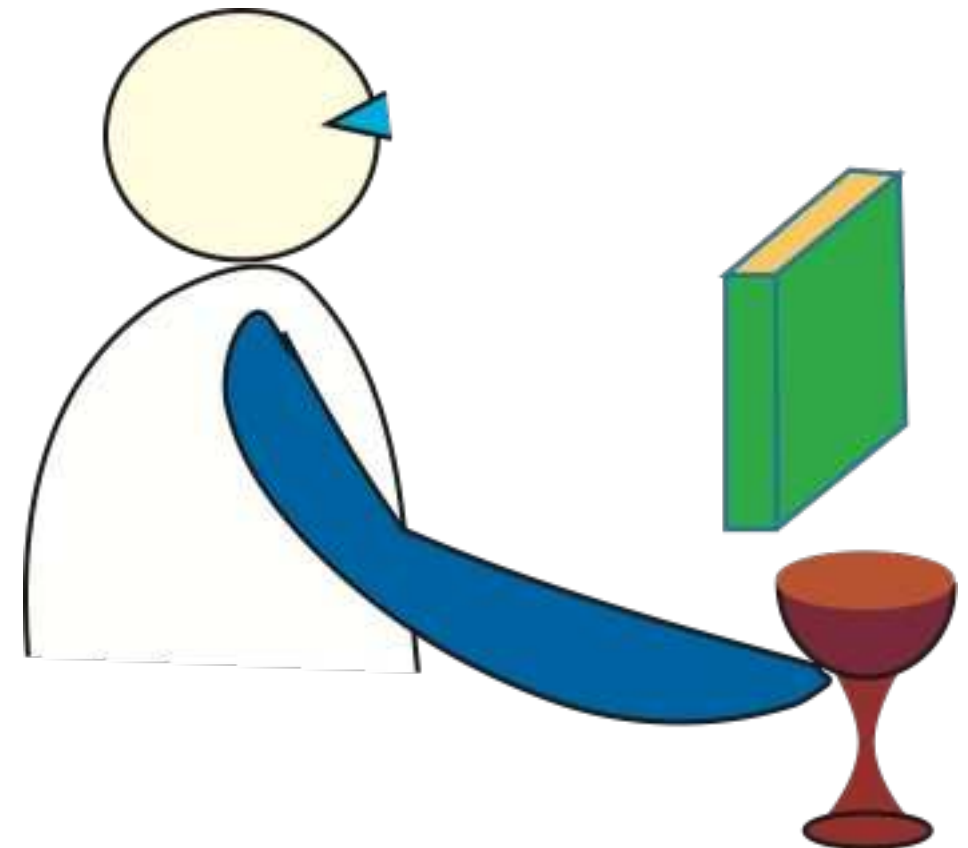
A case study in dynamic field theory as a  
framework for neurally grounded  
architectures for higher cognition

Mathis Richter

Institut für Neuroinformatik

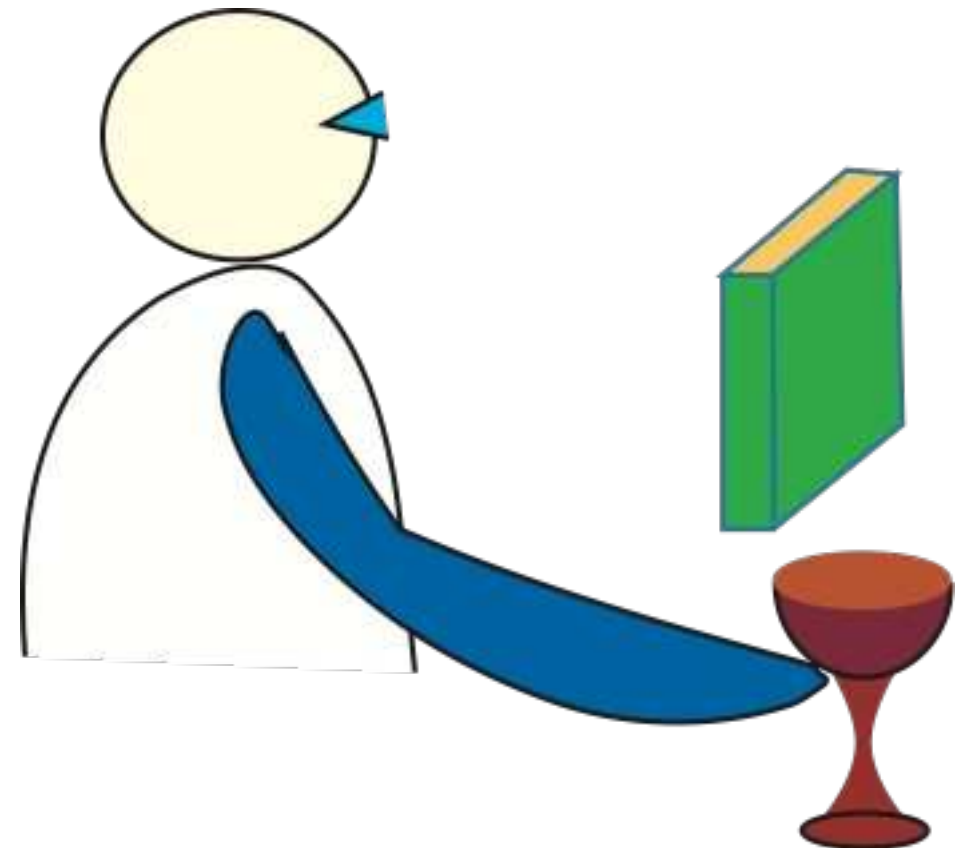
# Perceptually grounding language

- 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



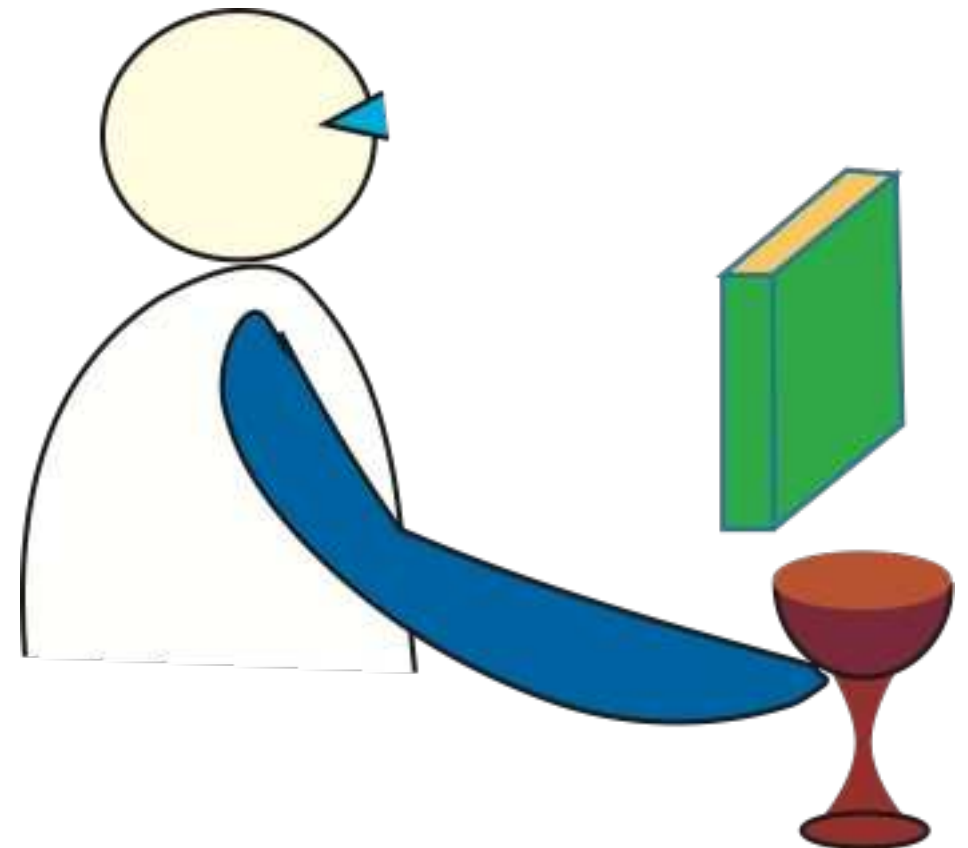
# Perceptually grounding language

- 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



# Perceptually grounding language

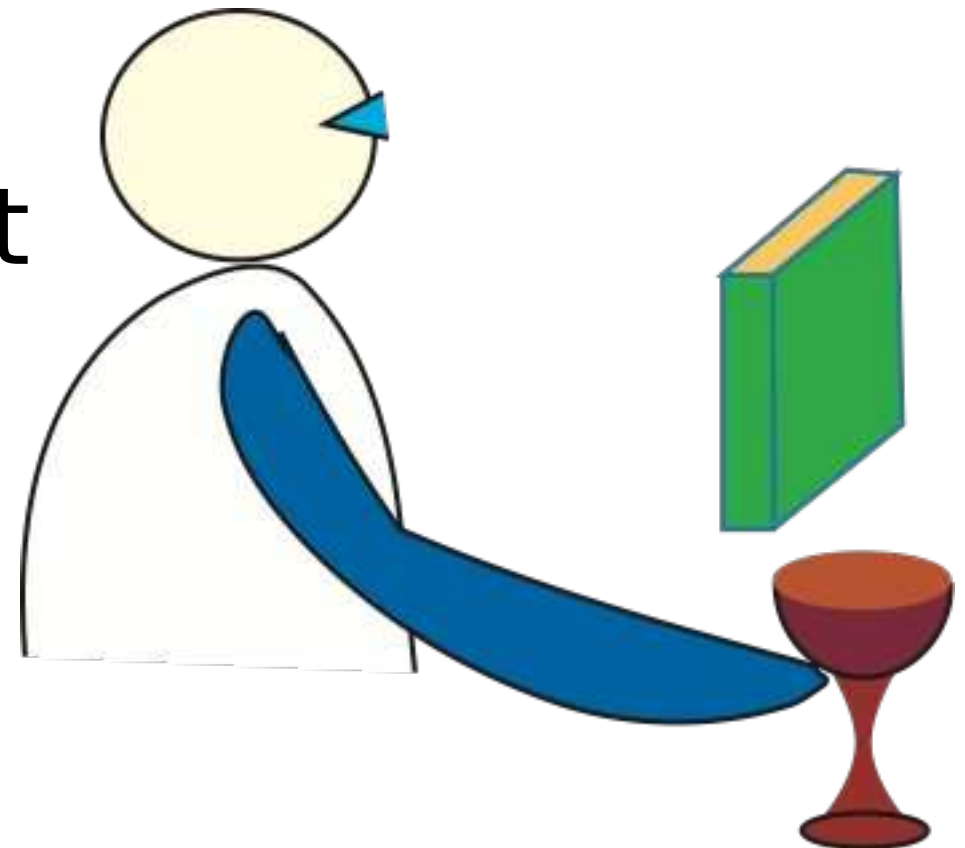
- that process could be mediated by other forms of communication, e.g., pointing (deictic code)



# 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 the name of  
that fruit to the right of  
the lime?”

- 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”, etc.
- a part of language that is 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”

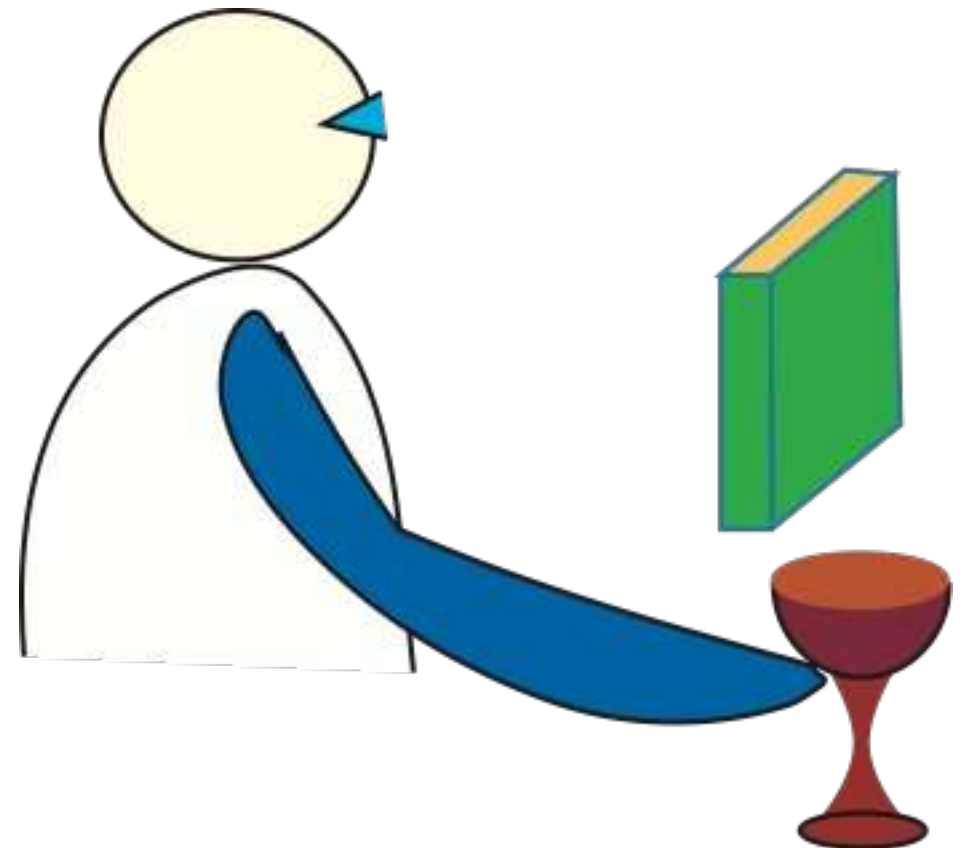
# Grounding spatial language

## ■ 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)

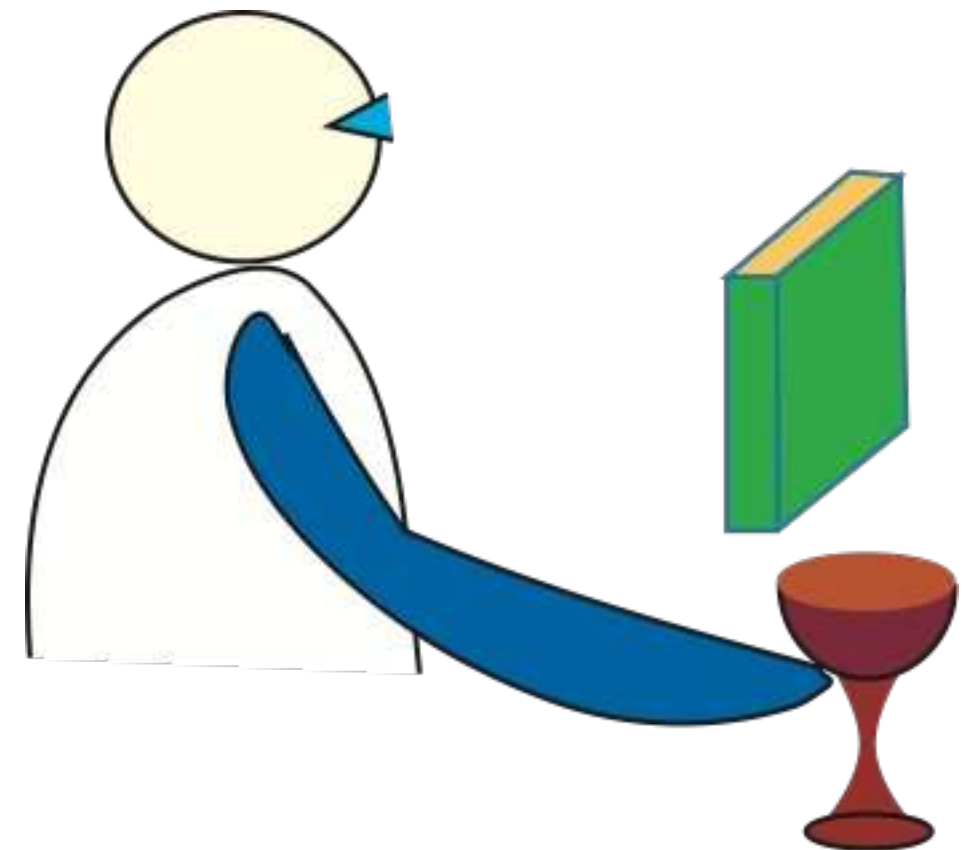
# Grounding spatial language

- 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



# Operations involved in grounding spatial language

- bring objects (target and reference) into the perceptual foreground (visually find them)
- make coordinate transformation
- apply comparison operators

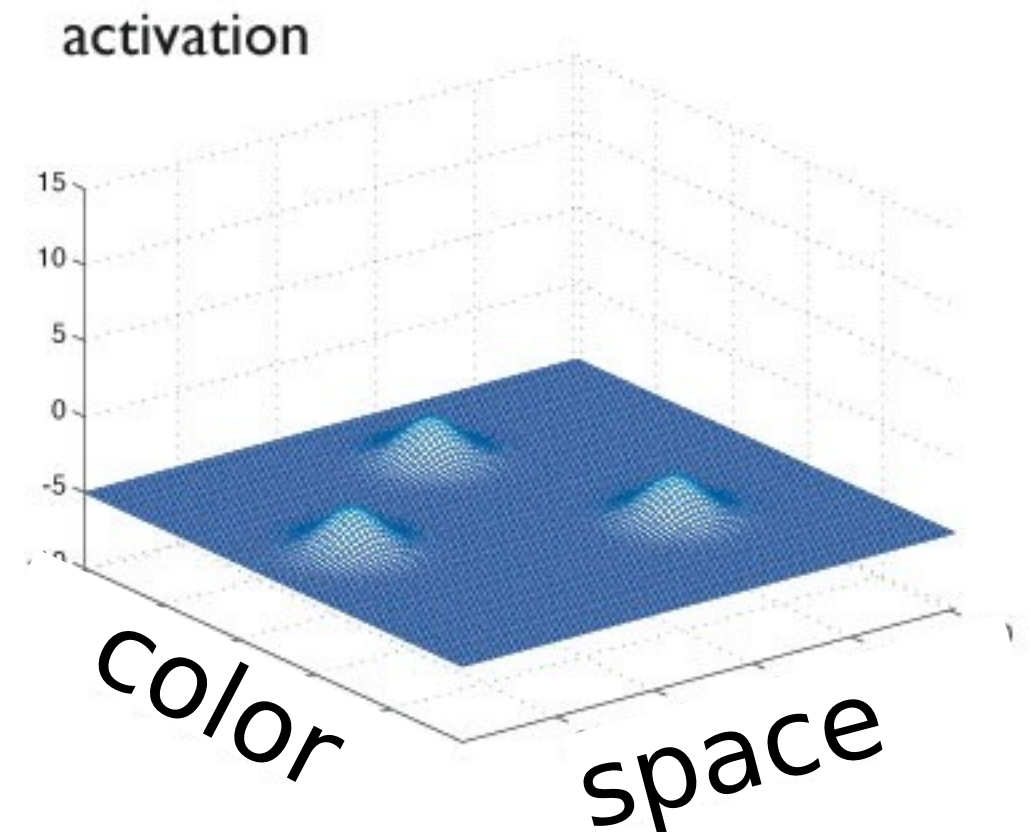
# DFT approach to bringing a perceptual object into the foreground

■ => lecture on higher-dimensional fields

# Bringing an object to the foreground



■ visual  
search:  
“where is the  
red object”?



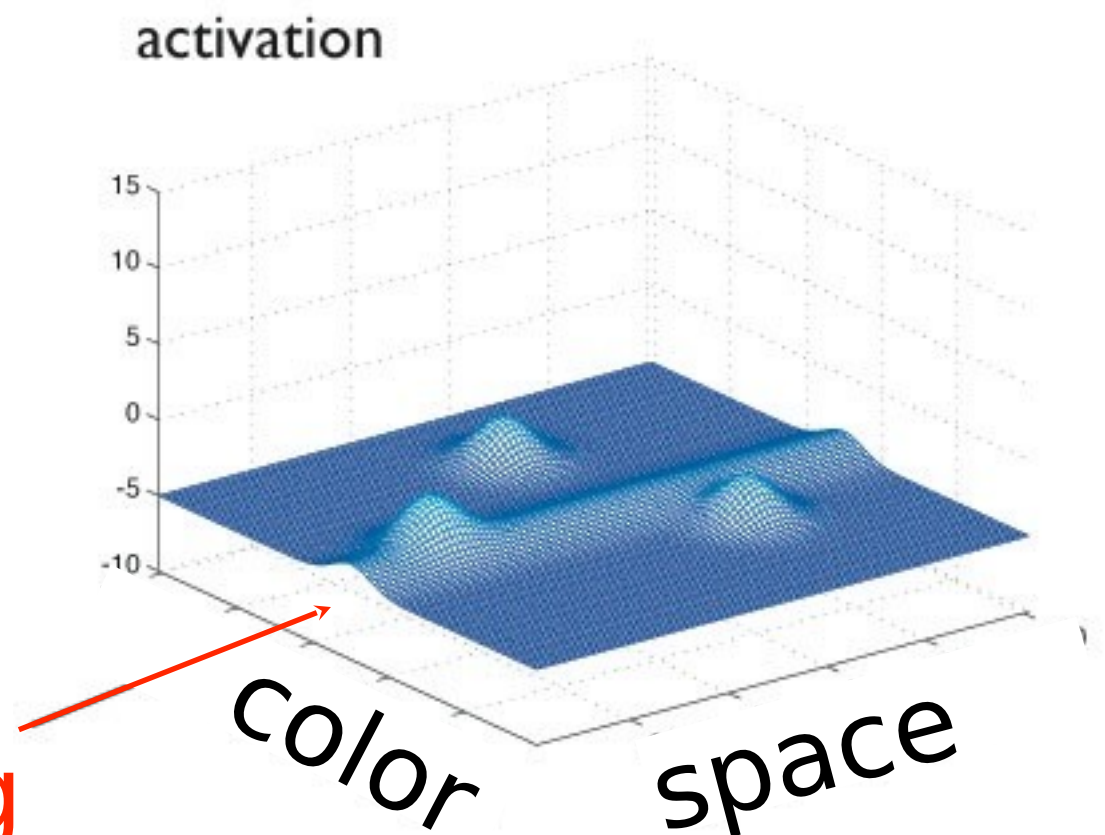


# Bringing an object to the foreground



■ visual  
search:  
“where is the  
red object”?

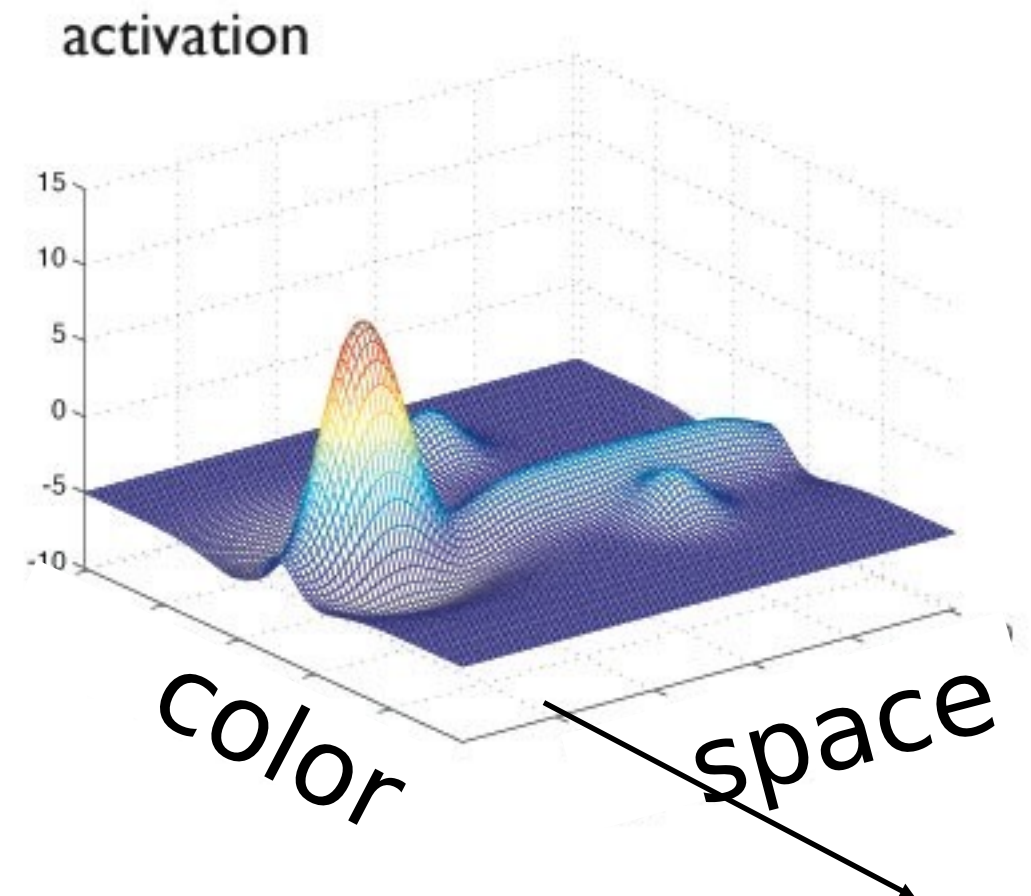
ridge  
specifying  
red



# Bringing an object to the foreground



■ visual  
search:  
“where is the  
red object”?

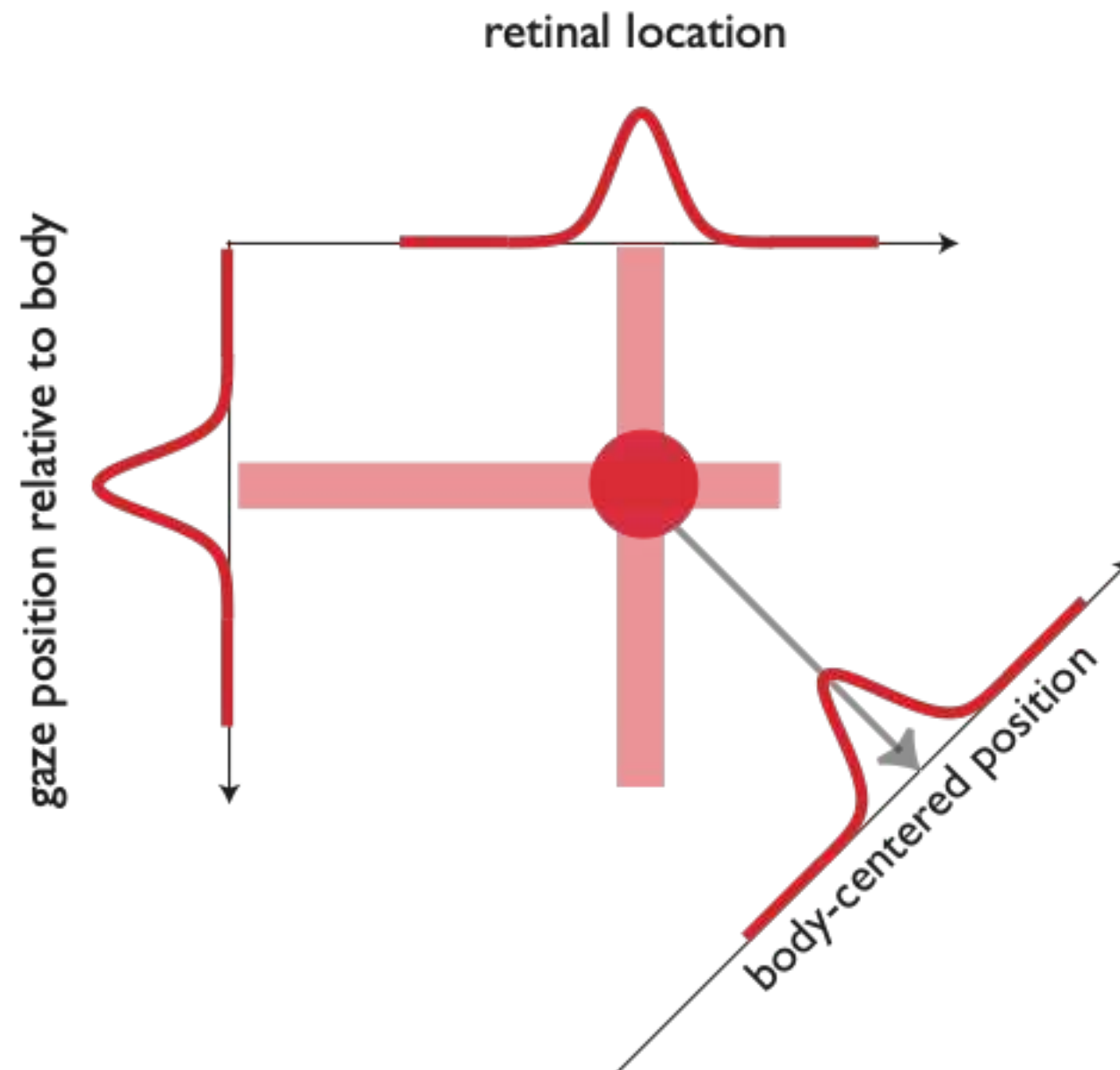


read out spatial  
location  
of red object

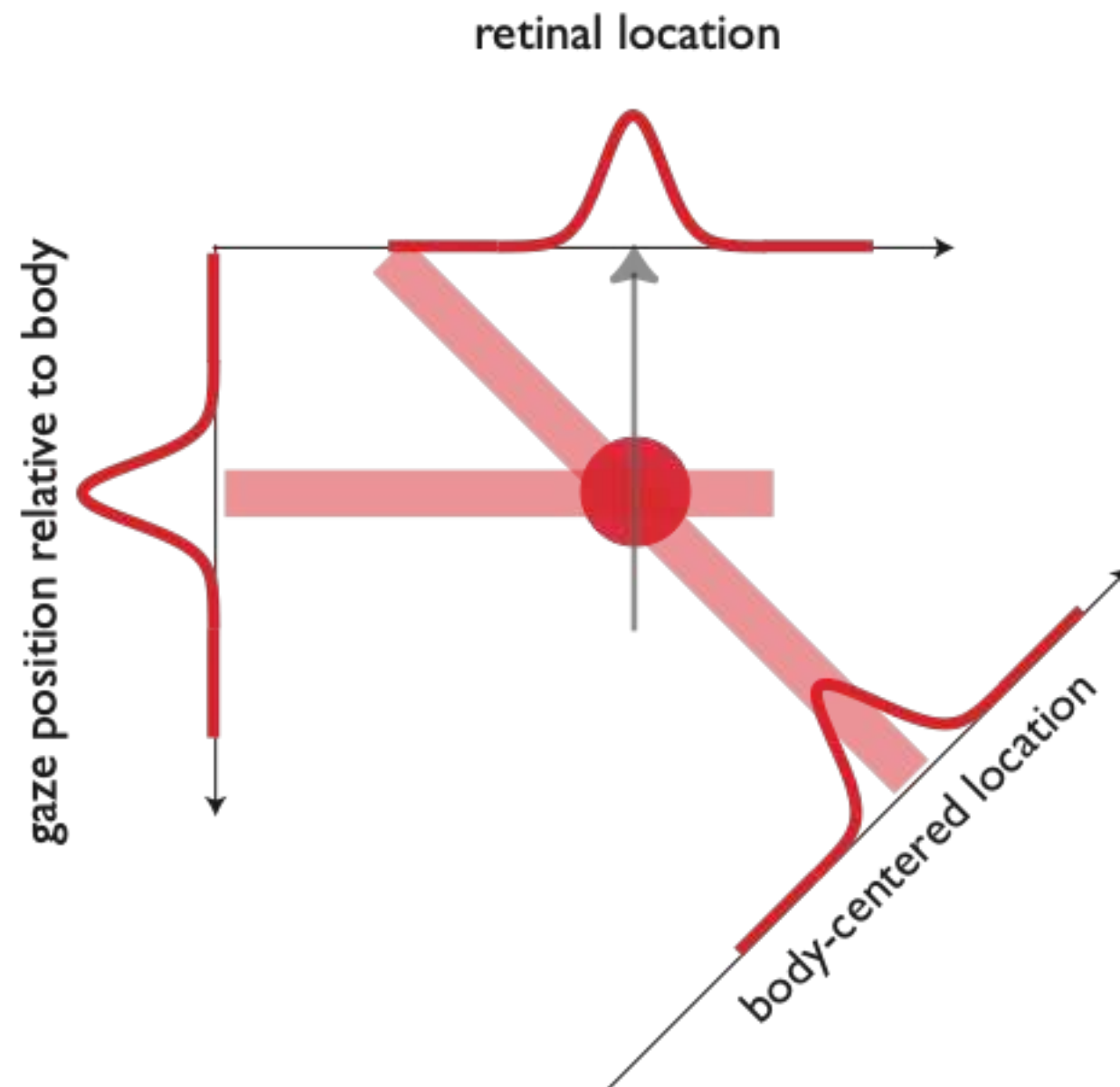
# DFT approach to coordinate transforms

■ => lecture on higher-dimensional  
fields

# Coordinate transformations

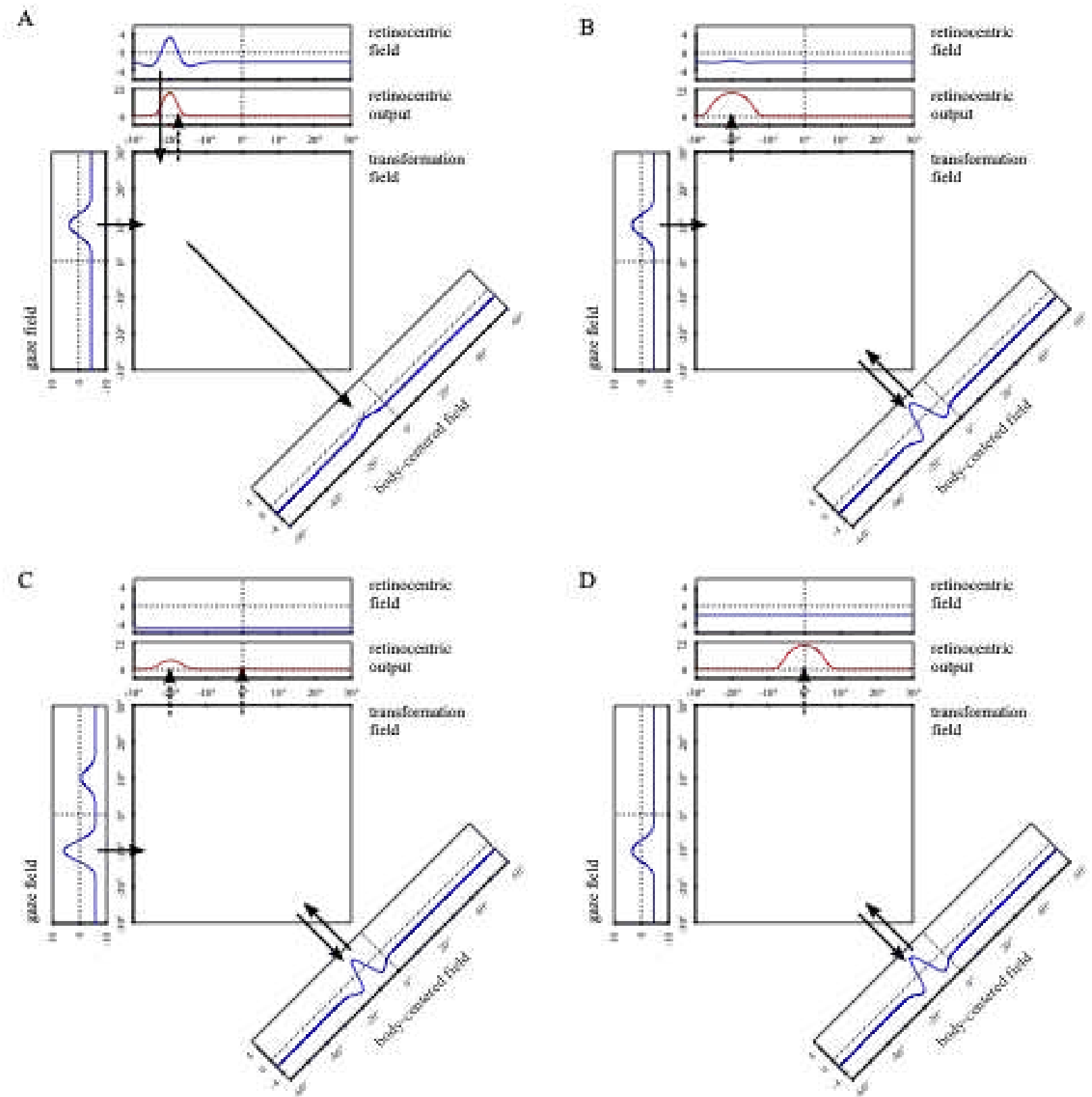


# Coordinate transformations



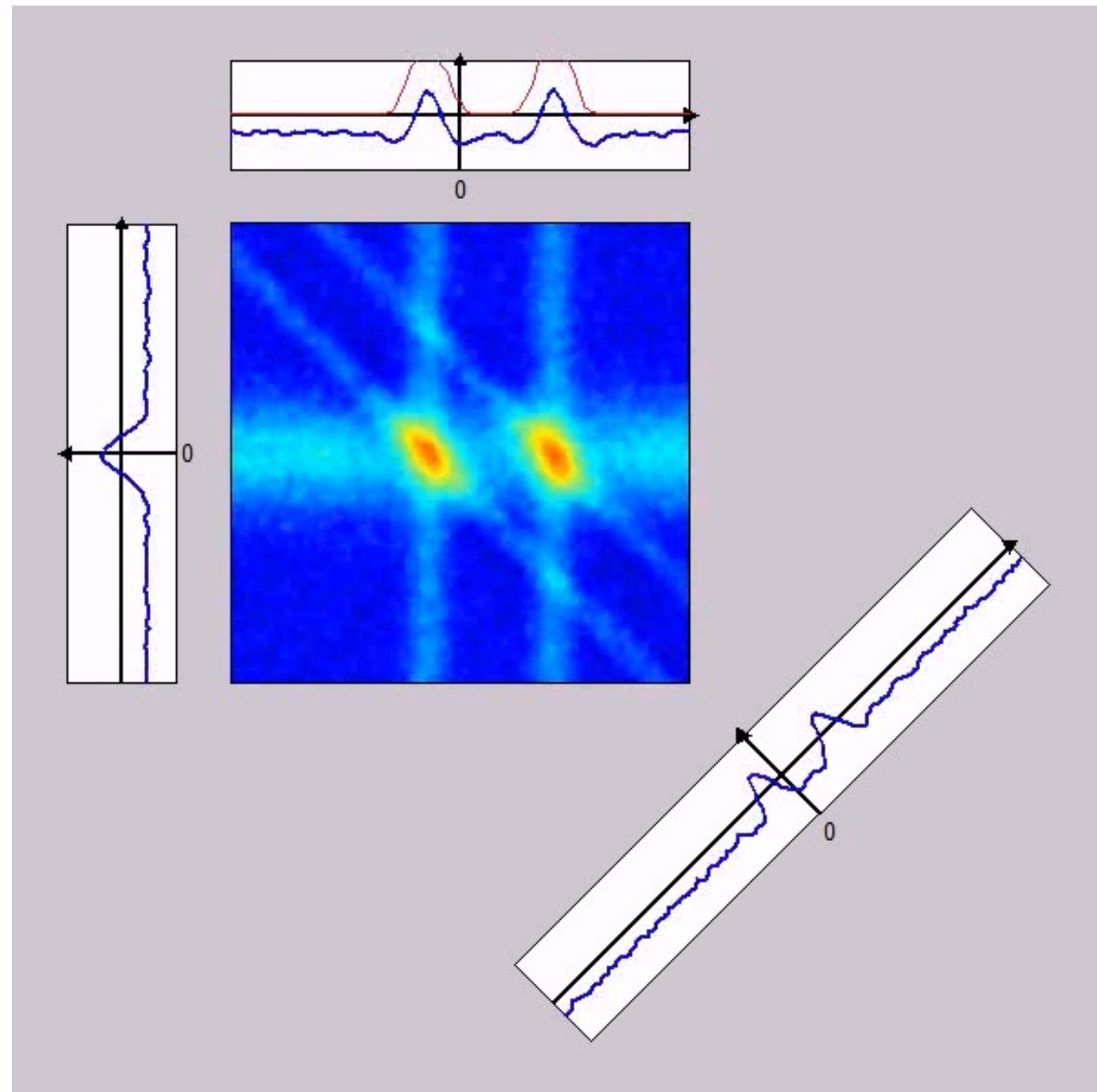
# Coordinate transformations

■ predict  
retinal  
location  
following  
gaze  
shift



# Coordinate transformations

- predict retinal location following gaze shift

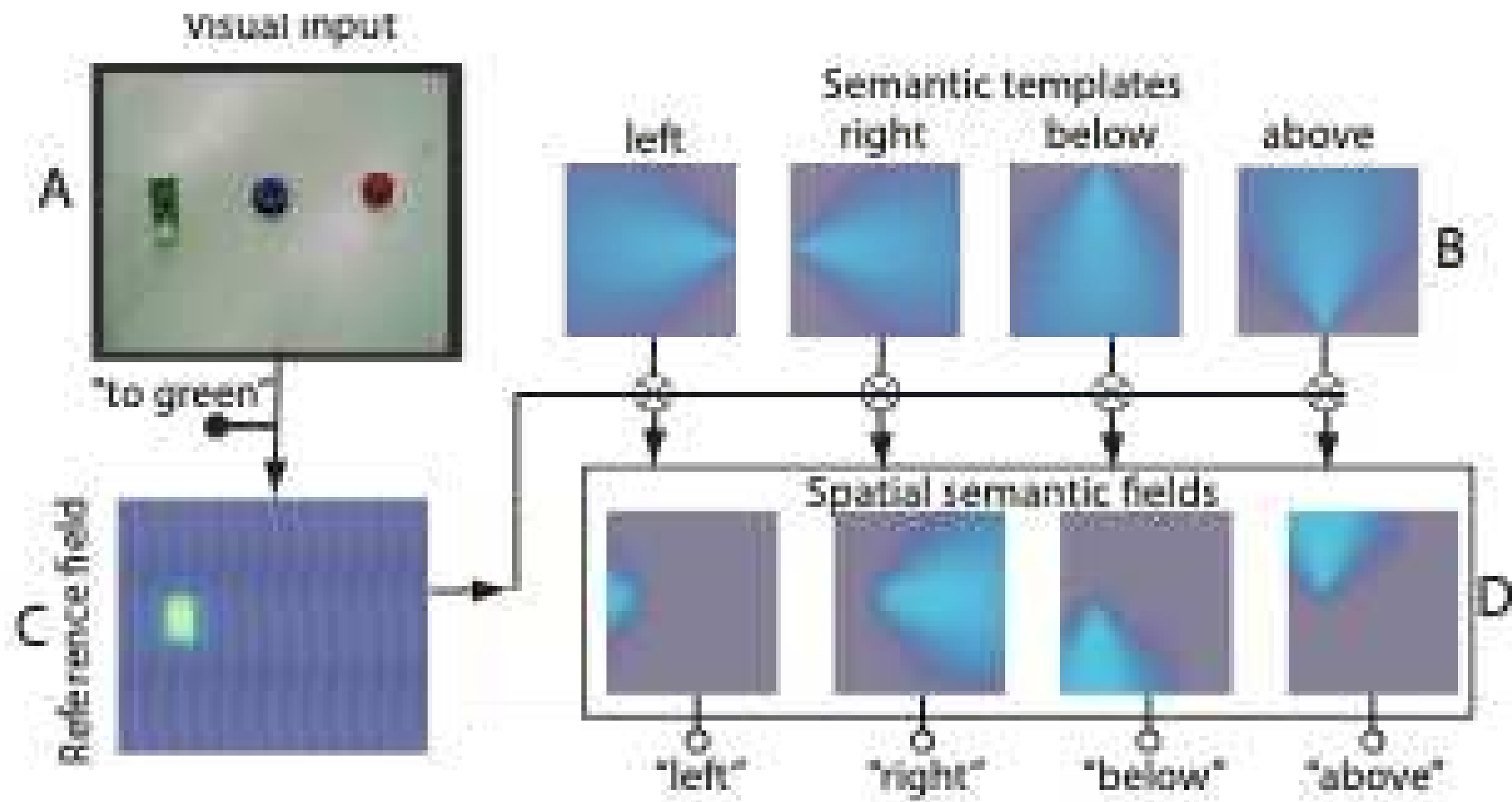


# DFT approach to applying operators



# DFT approach to applying operators

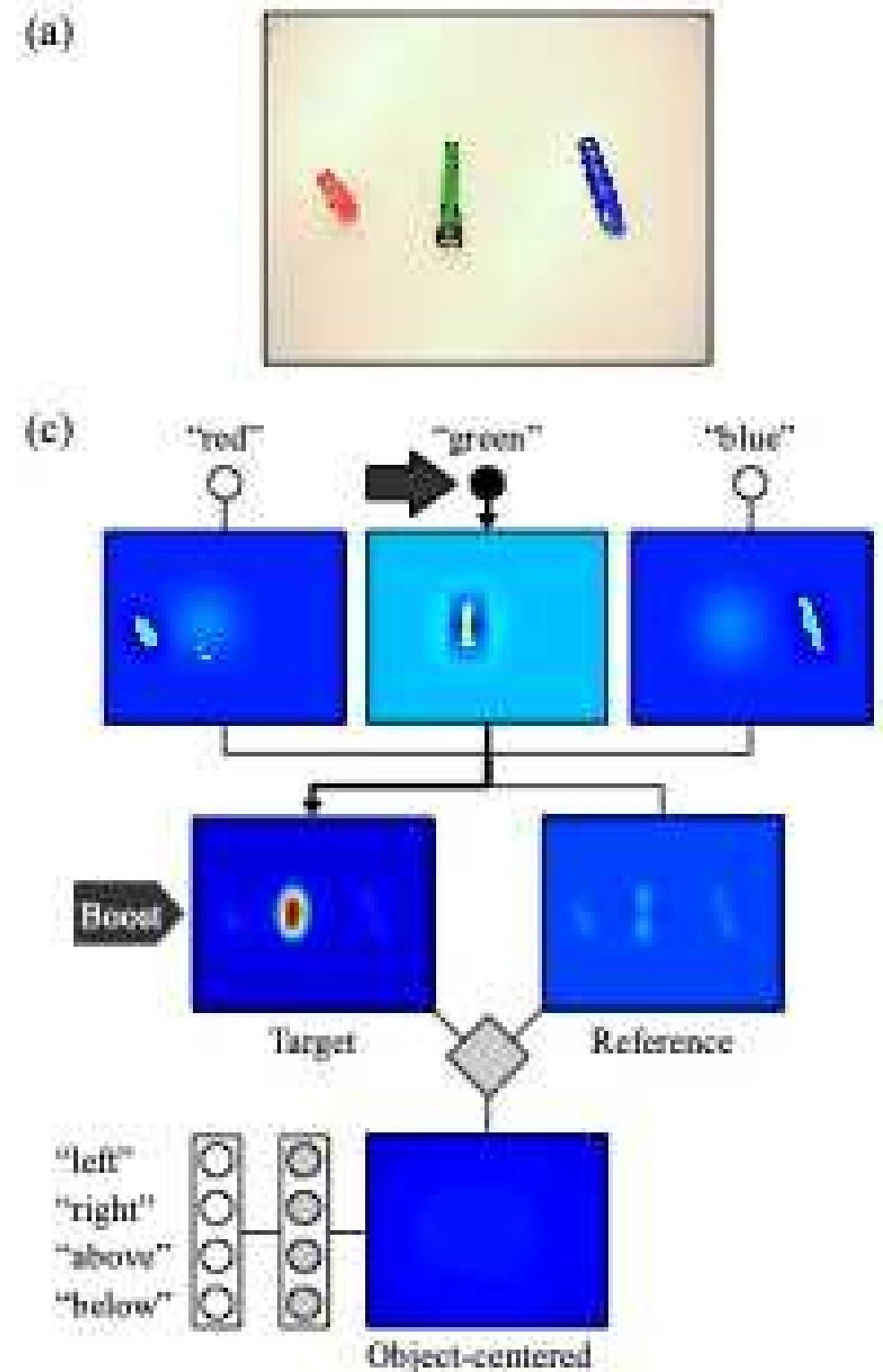
- based on convolution of fields with kernels



[from: Lipinski, Sandamirskaya, Schöner, 2009]

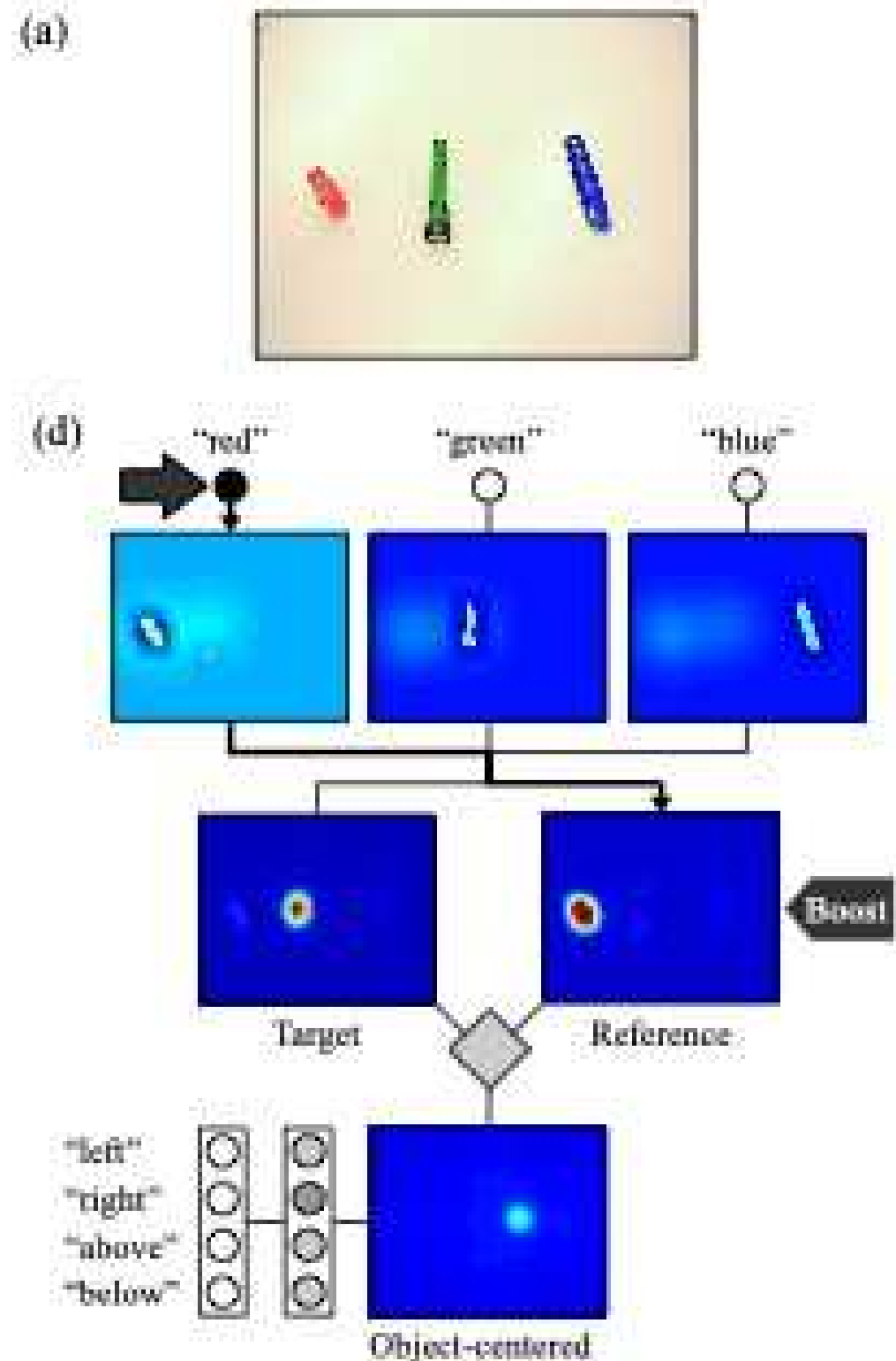
# Spatial comparison in DFT

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



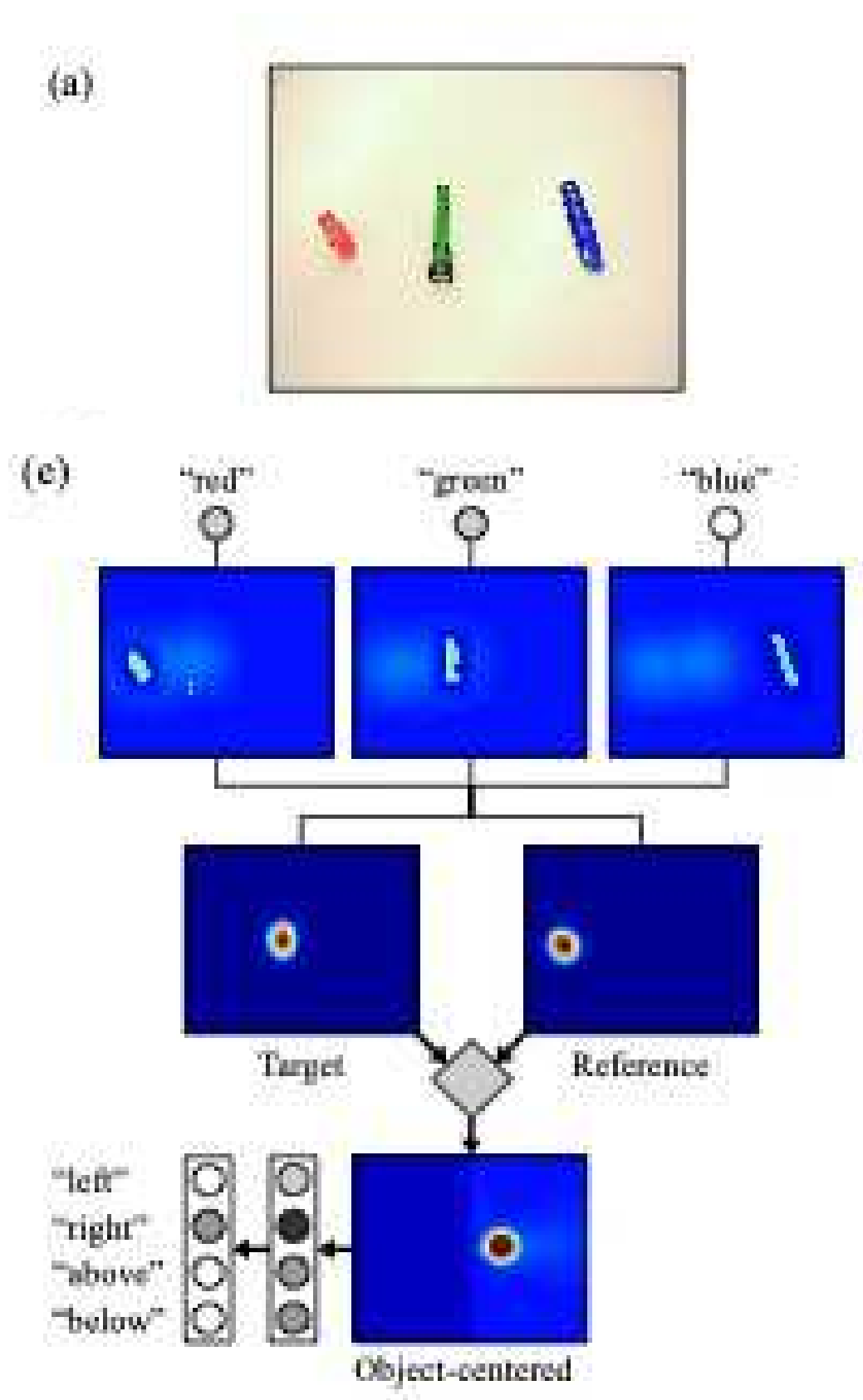
# Spatial comparison in DFT

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

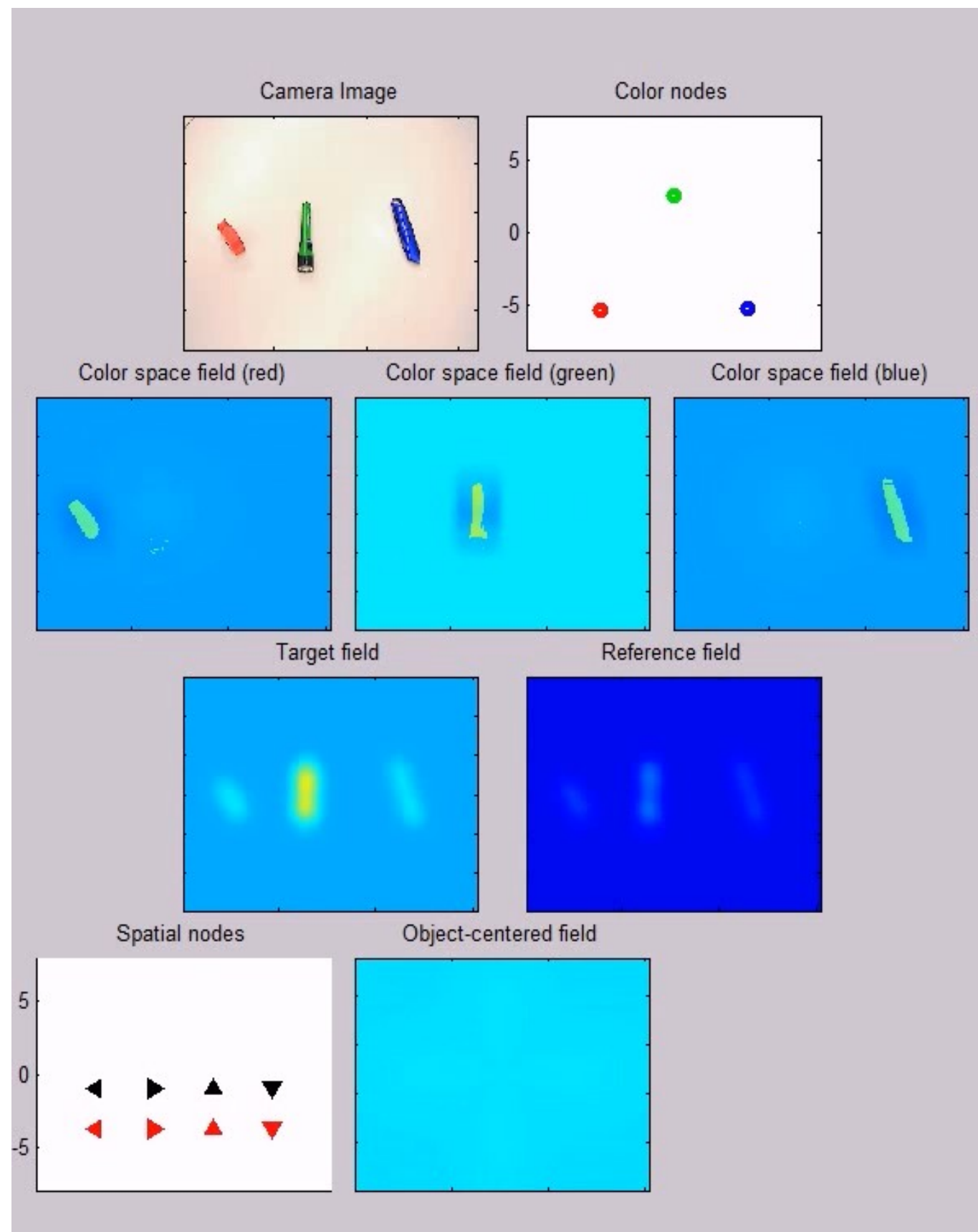


# Spatial comparison in DFT

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

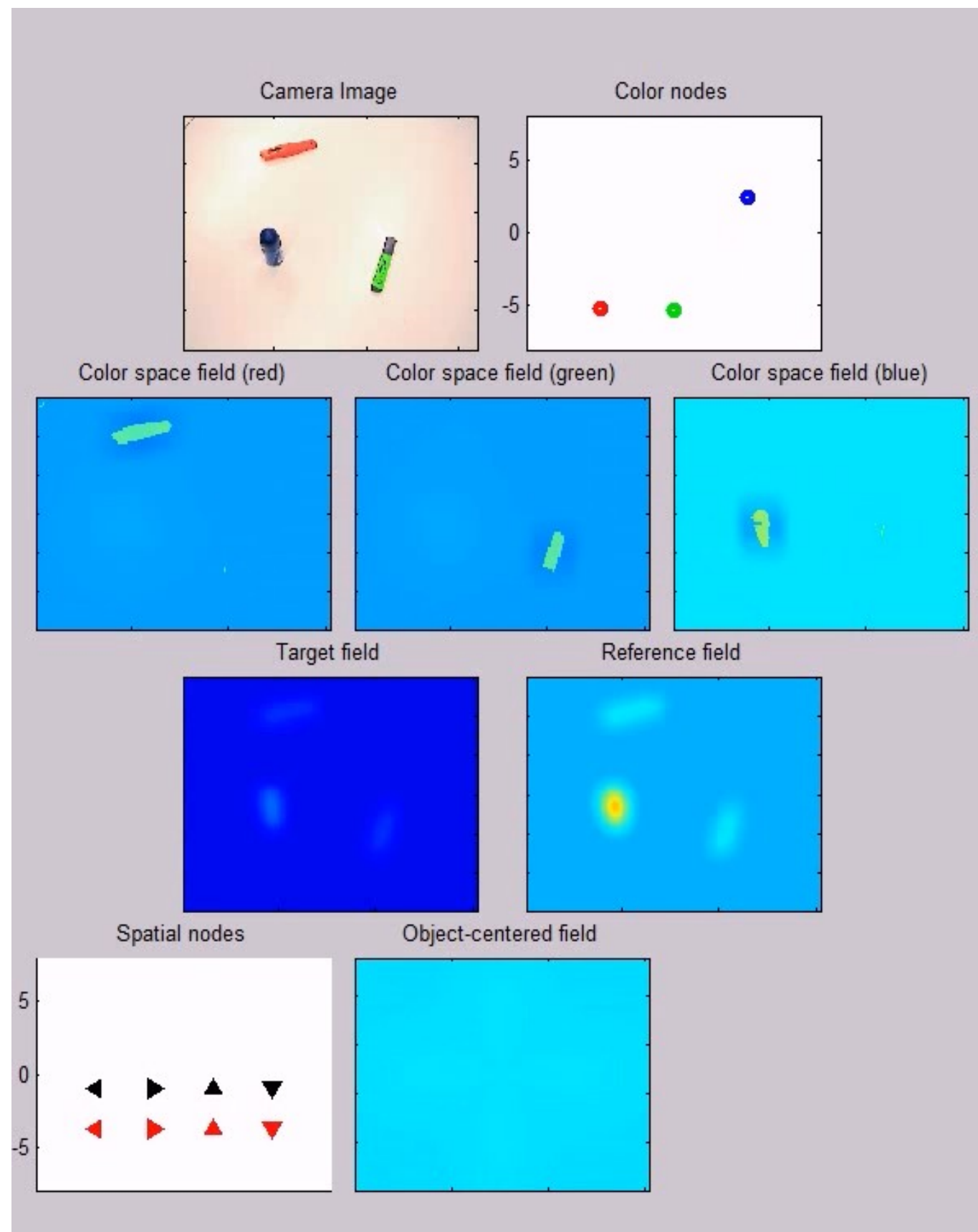


■ “where is the green object relative to the red object?”



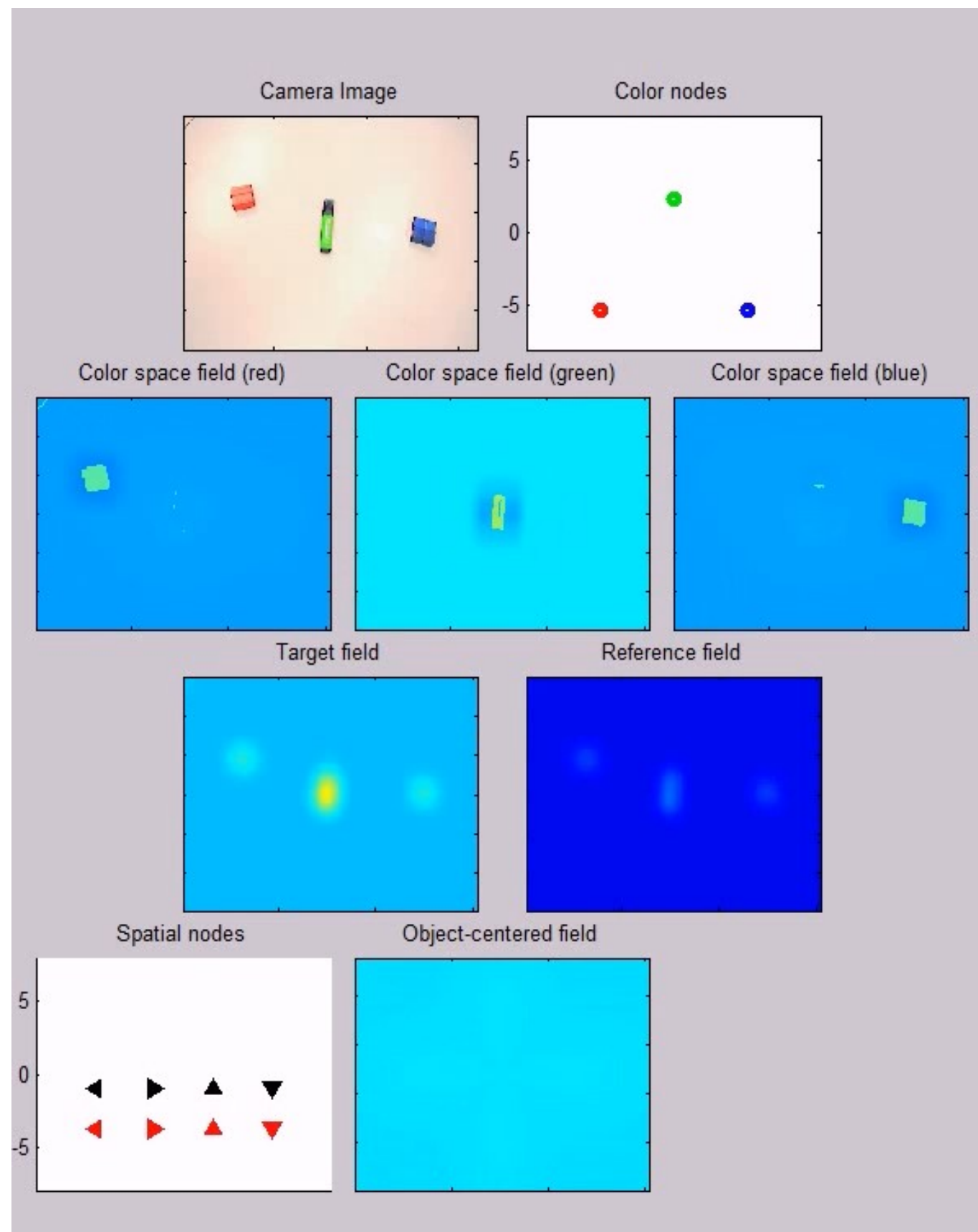
left right above below

■ “which object  
is above the  
blue object?”



left right above below

■ “where is the green object?”

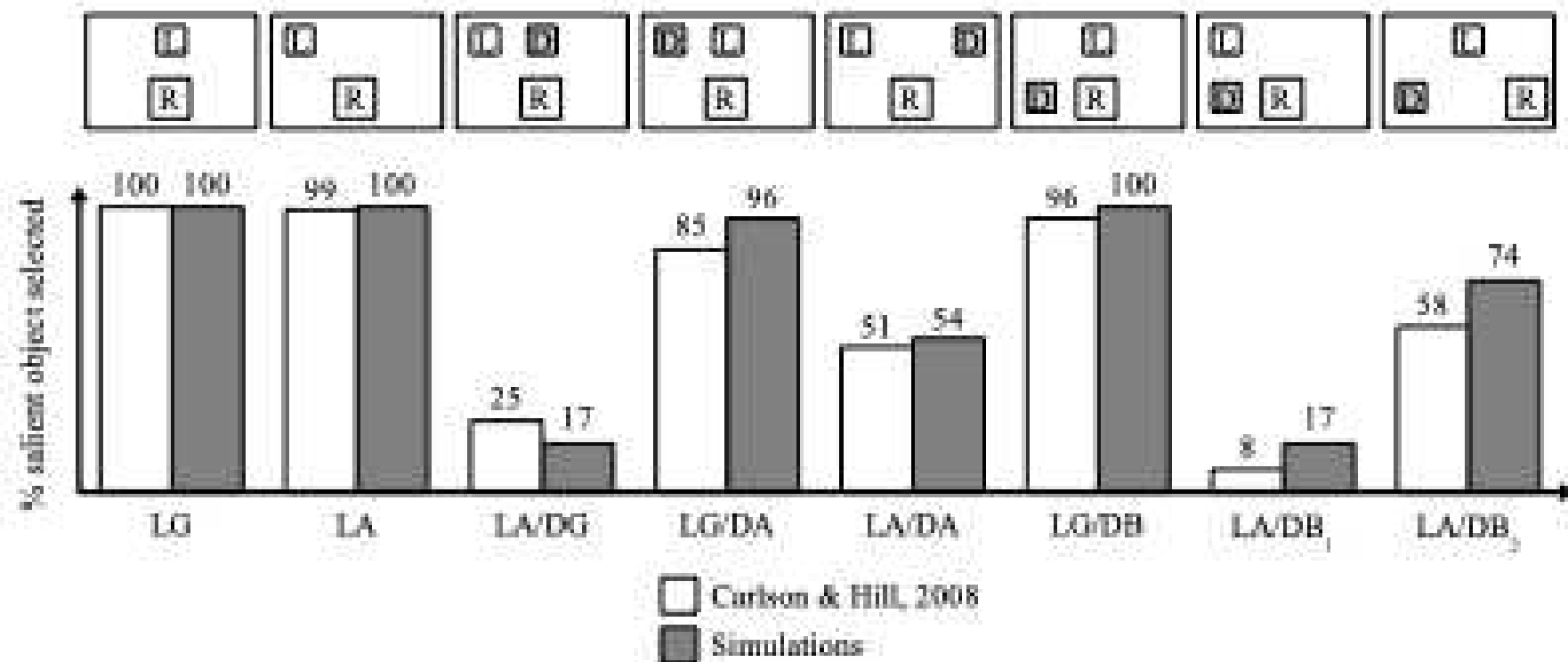


left right above below

[Lipinski et al: JEP:LMC (2011)]

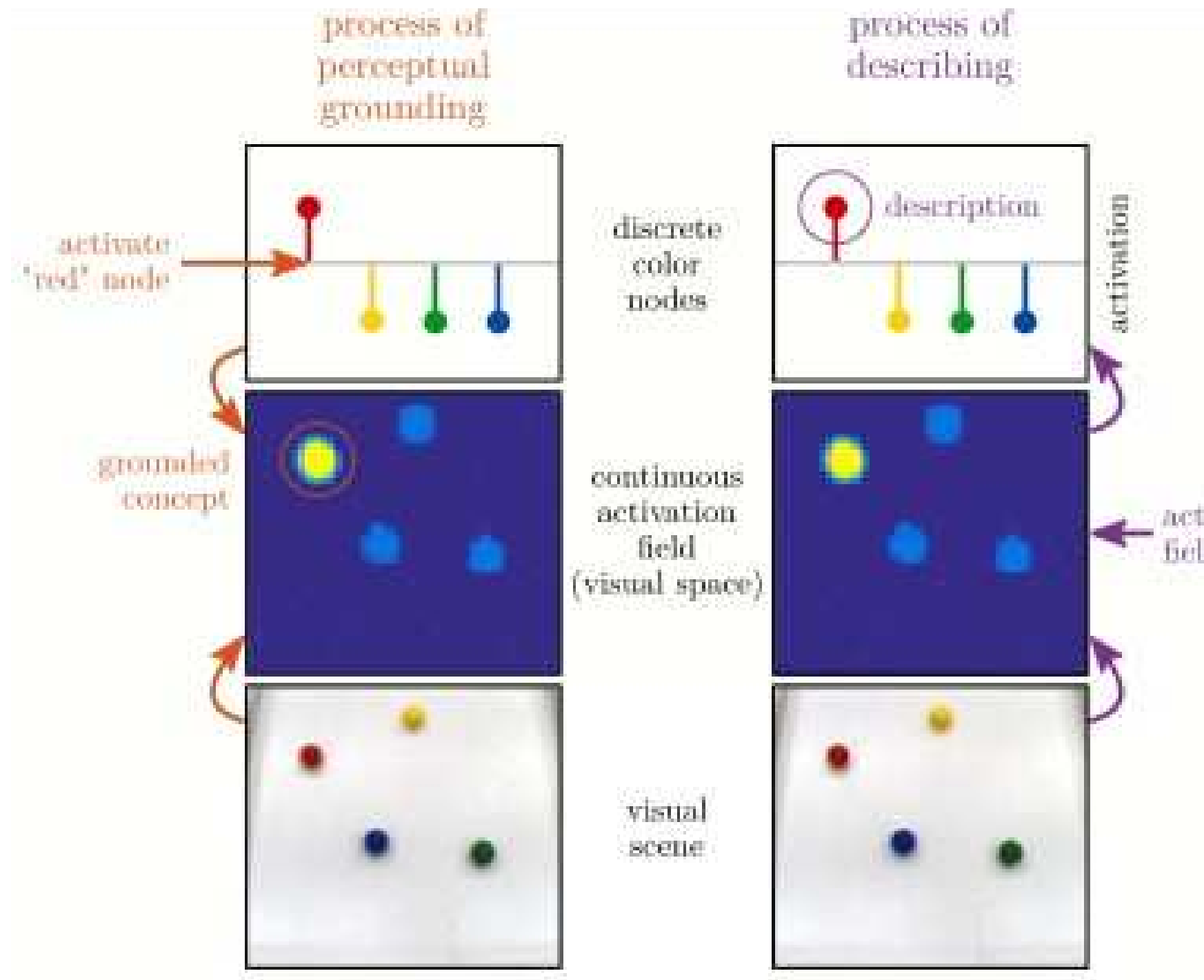
# Spatial comparison in DFT

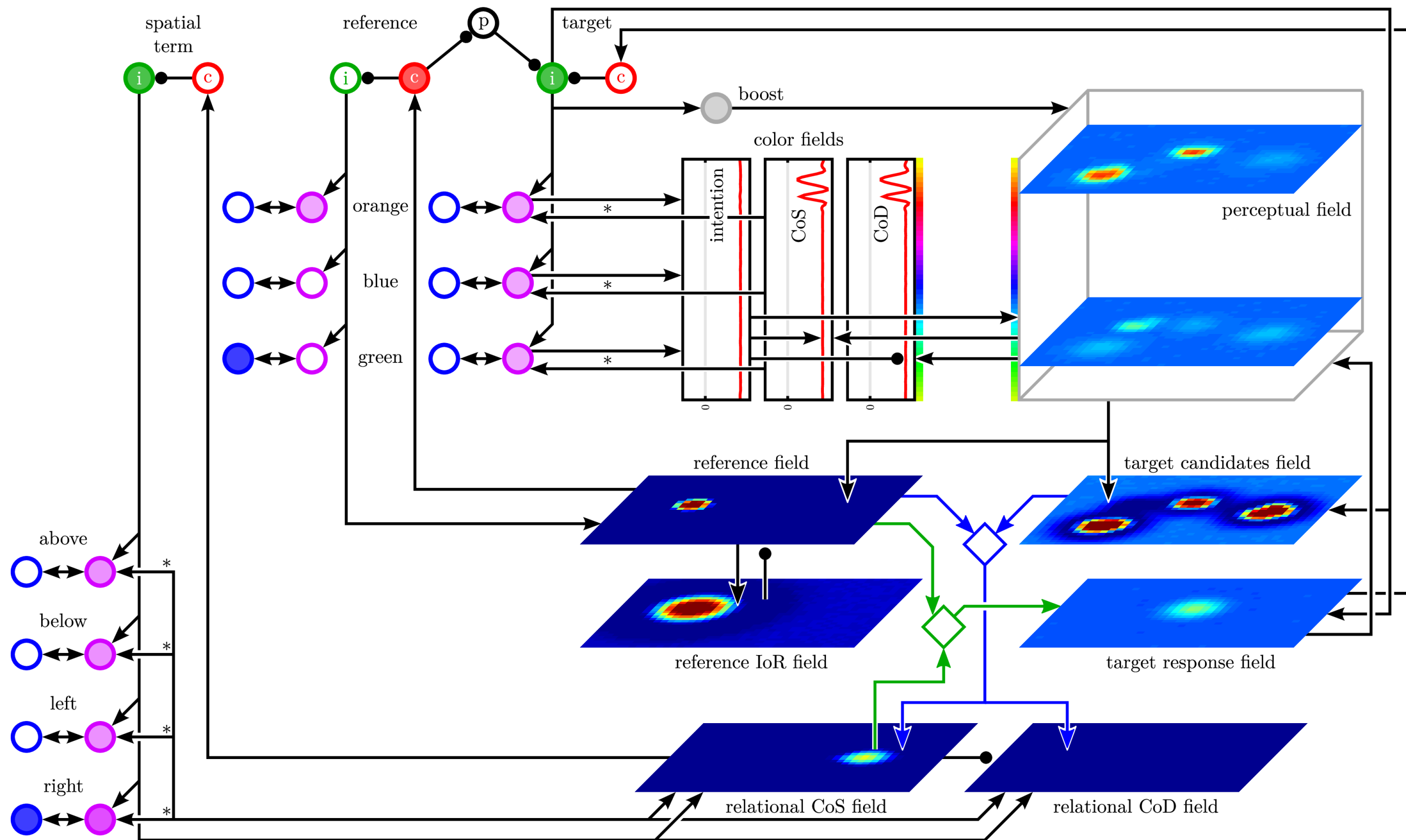
■ accounts for human data



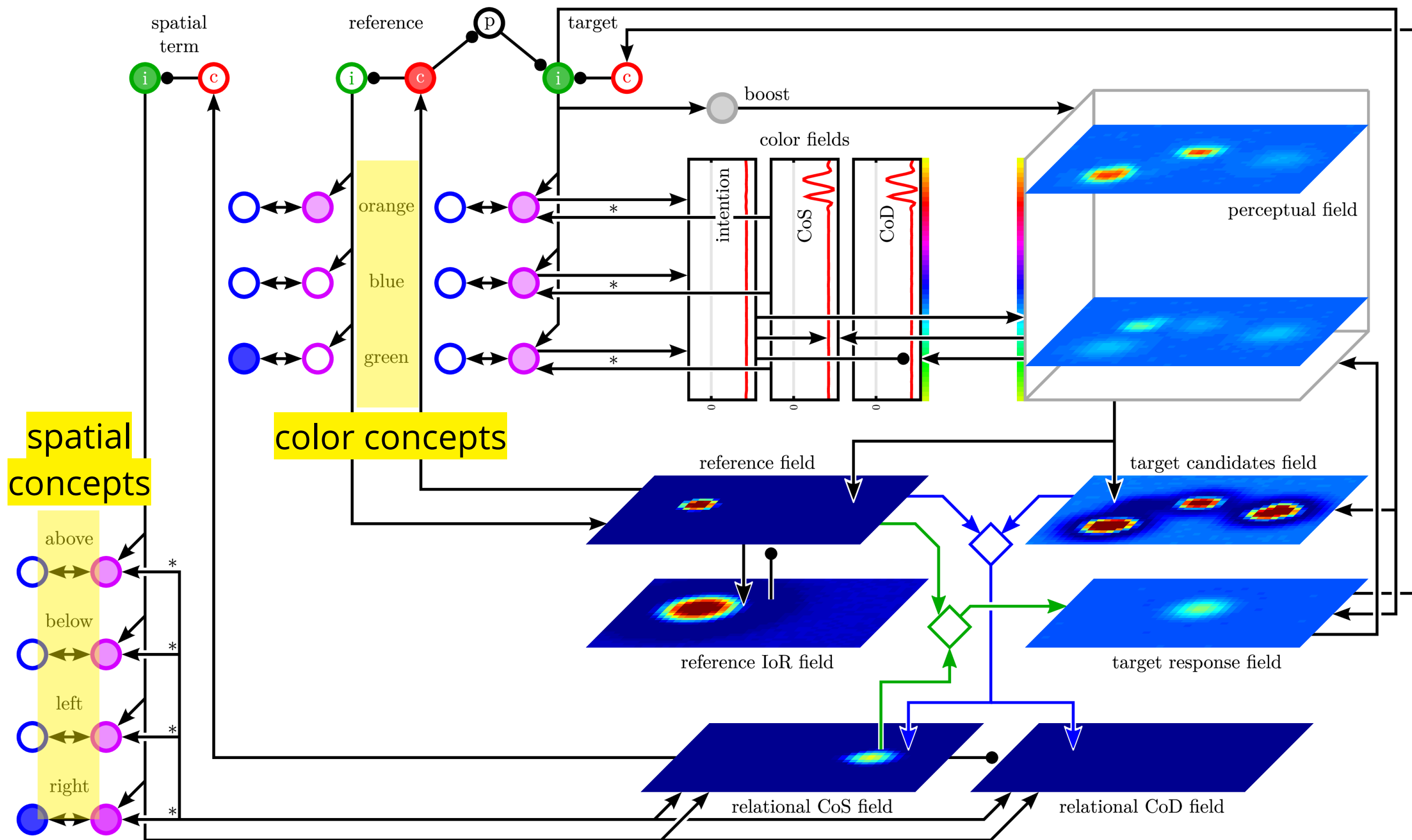


# A DFT architecture that does both grounding and describing

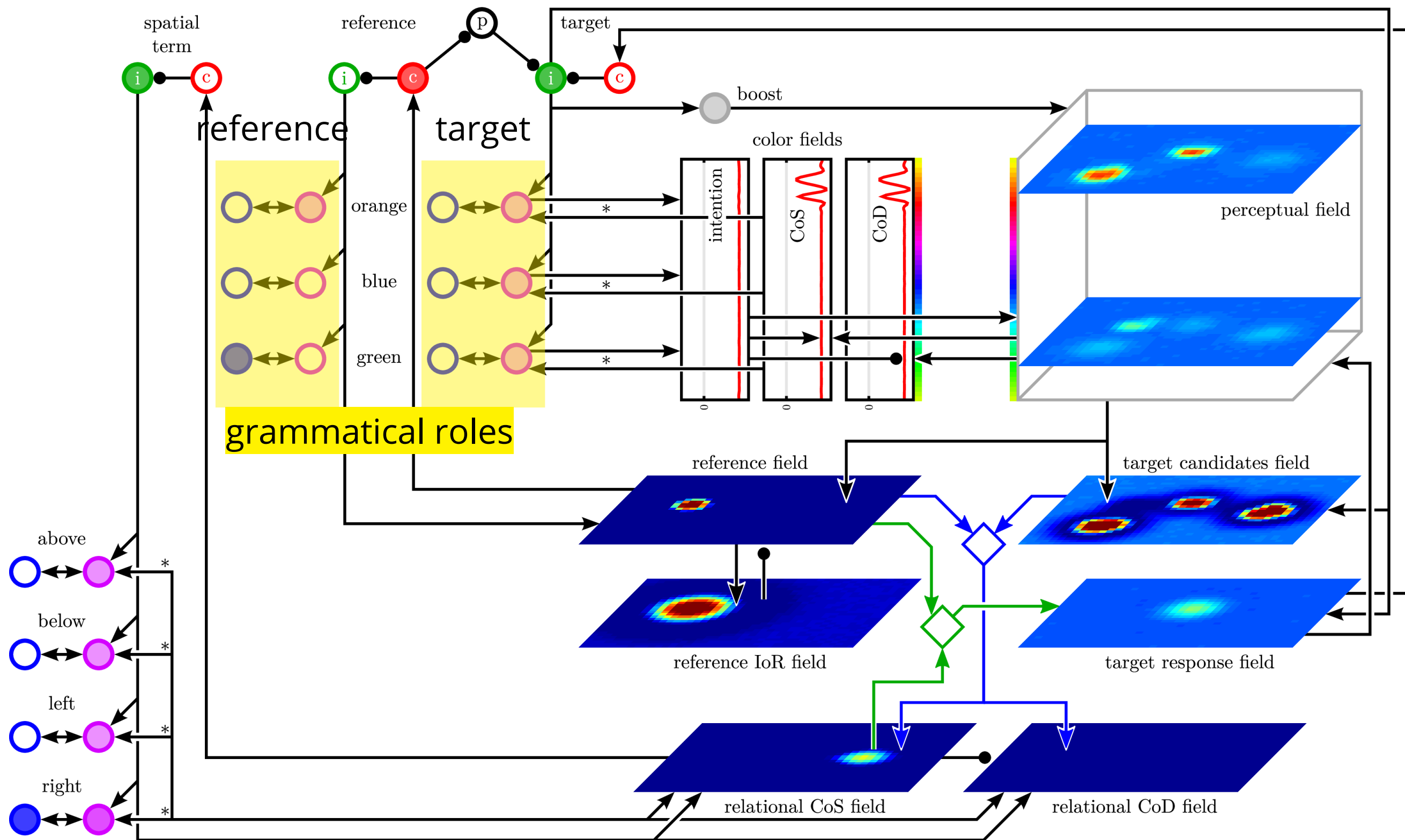




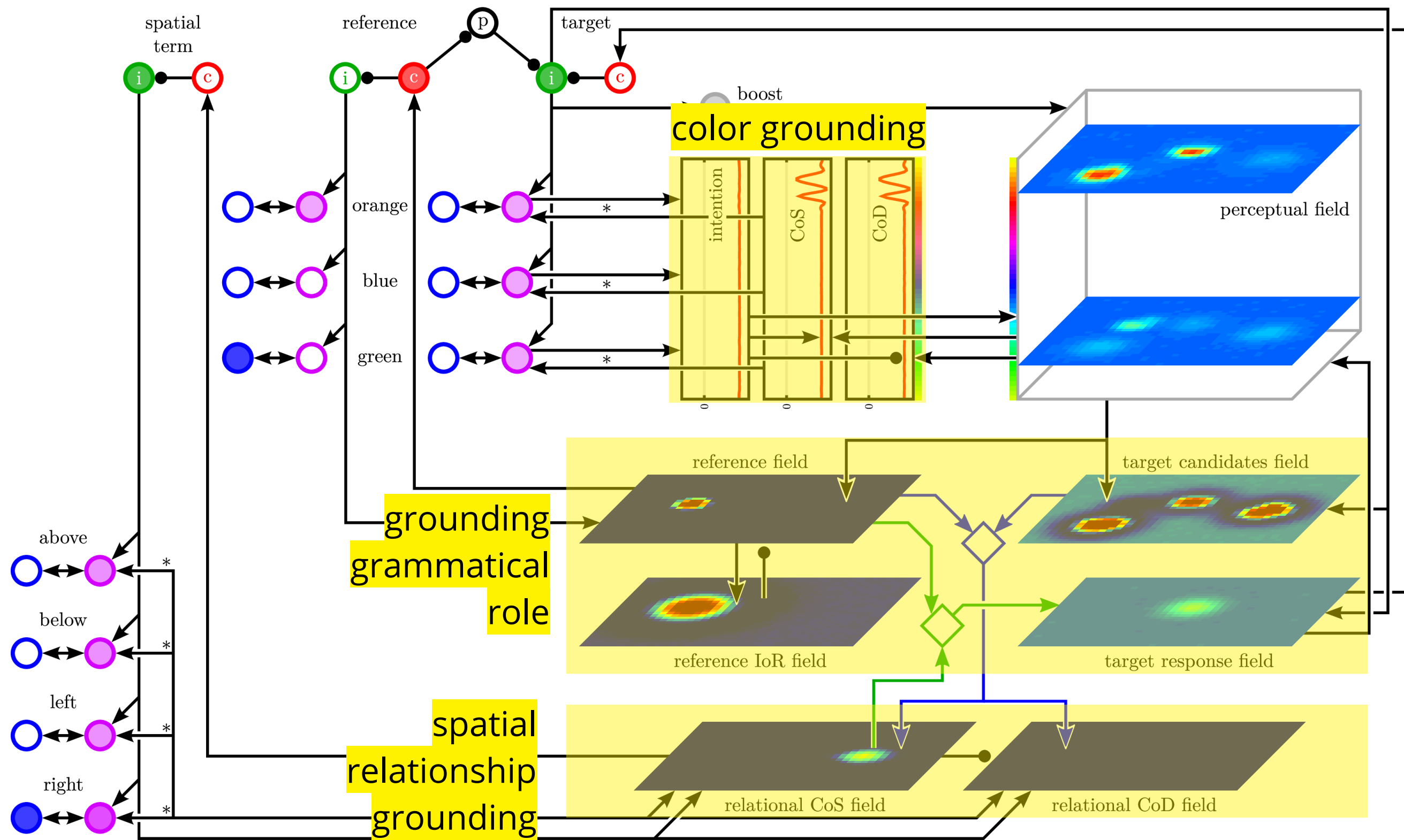
[Richter, Lins et al. ICANN 2014]



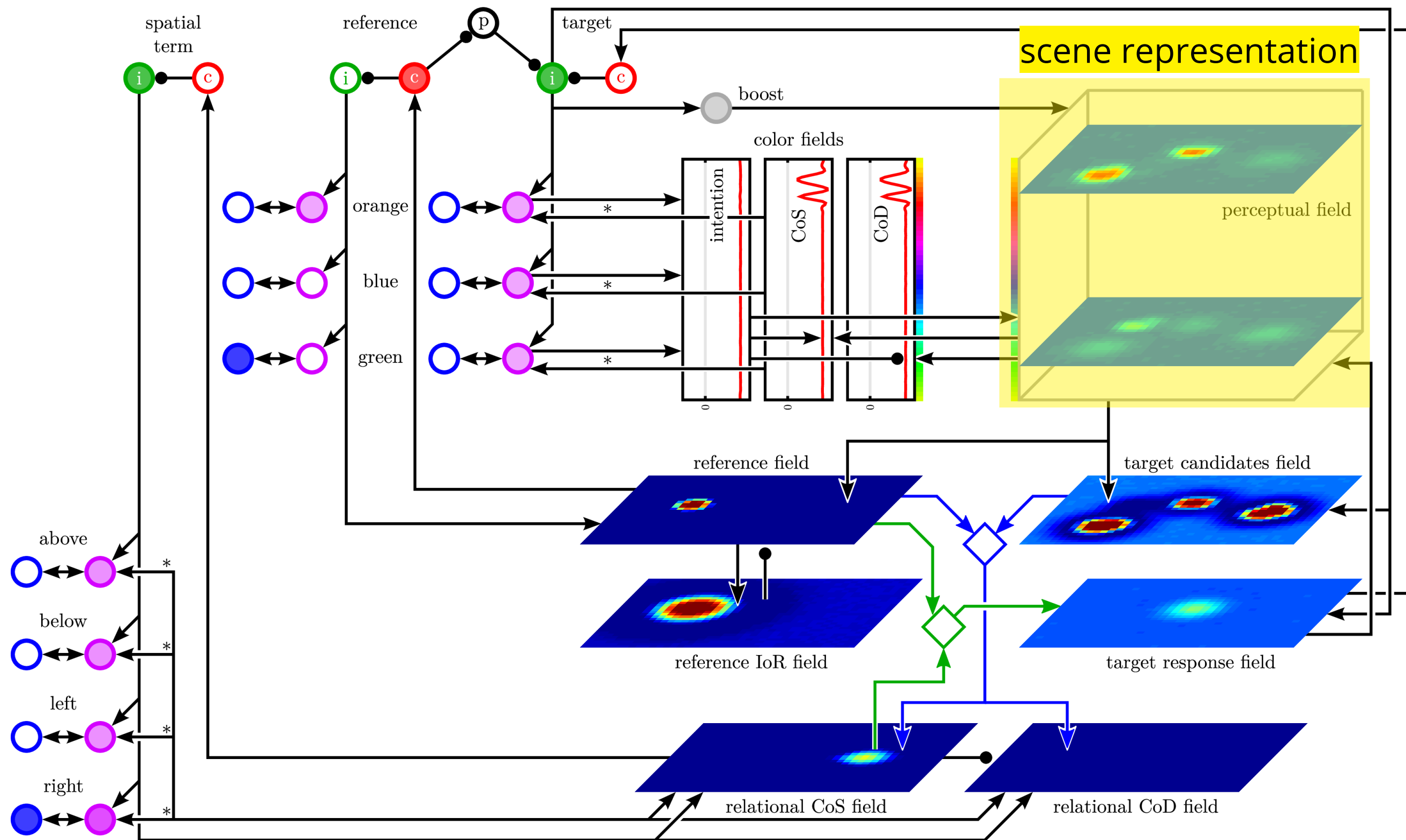
[Richter, Lins et al. ICANN 2014]



[Richter, Lins et al. ICANN 2014]



[Richter, Lins et al. ICANN 2014]

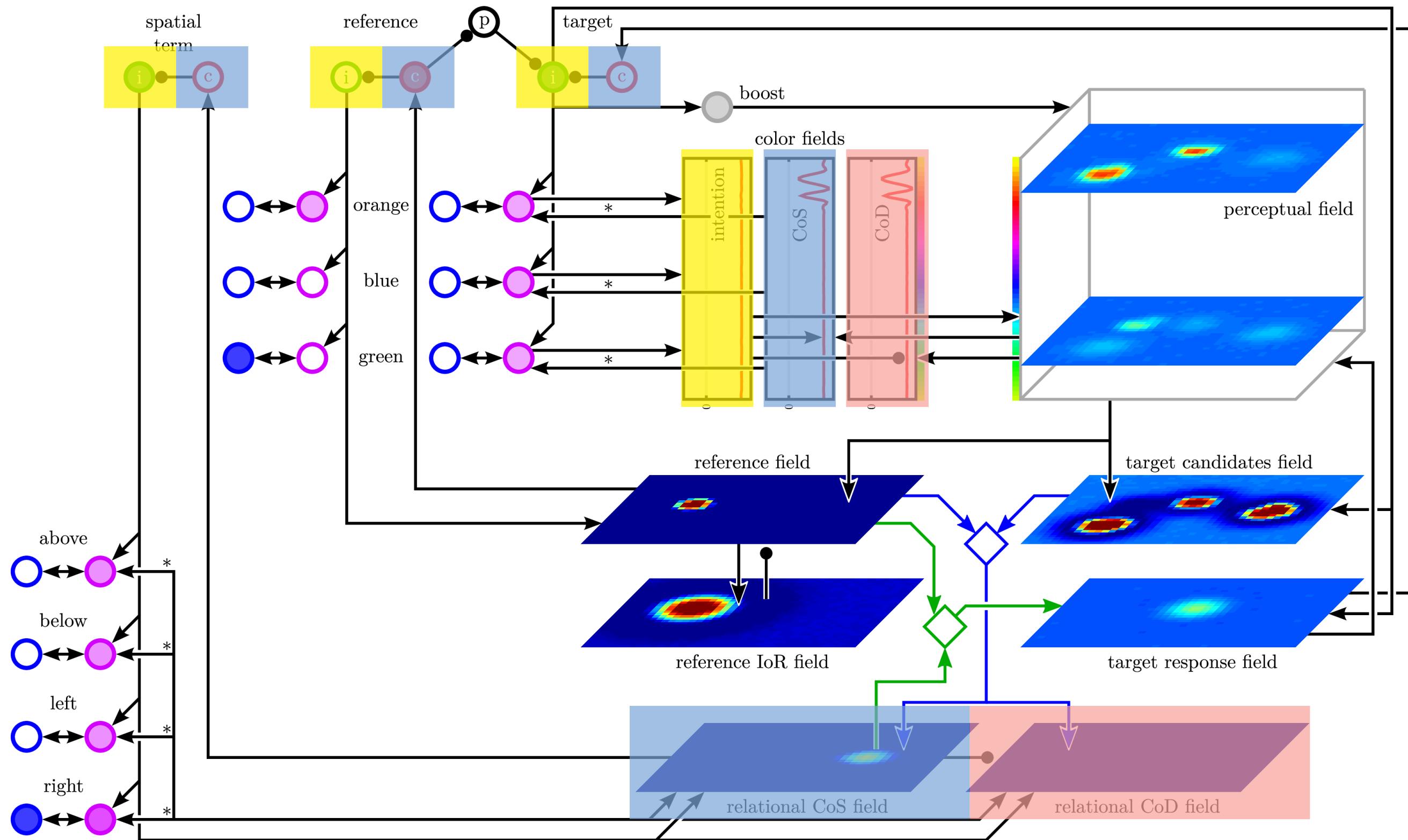


[Richter, Lins et al. ICANN 2014]

intention

condition of satisfaction

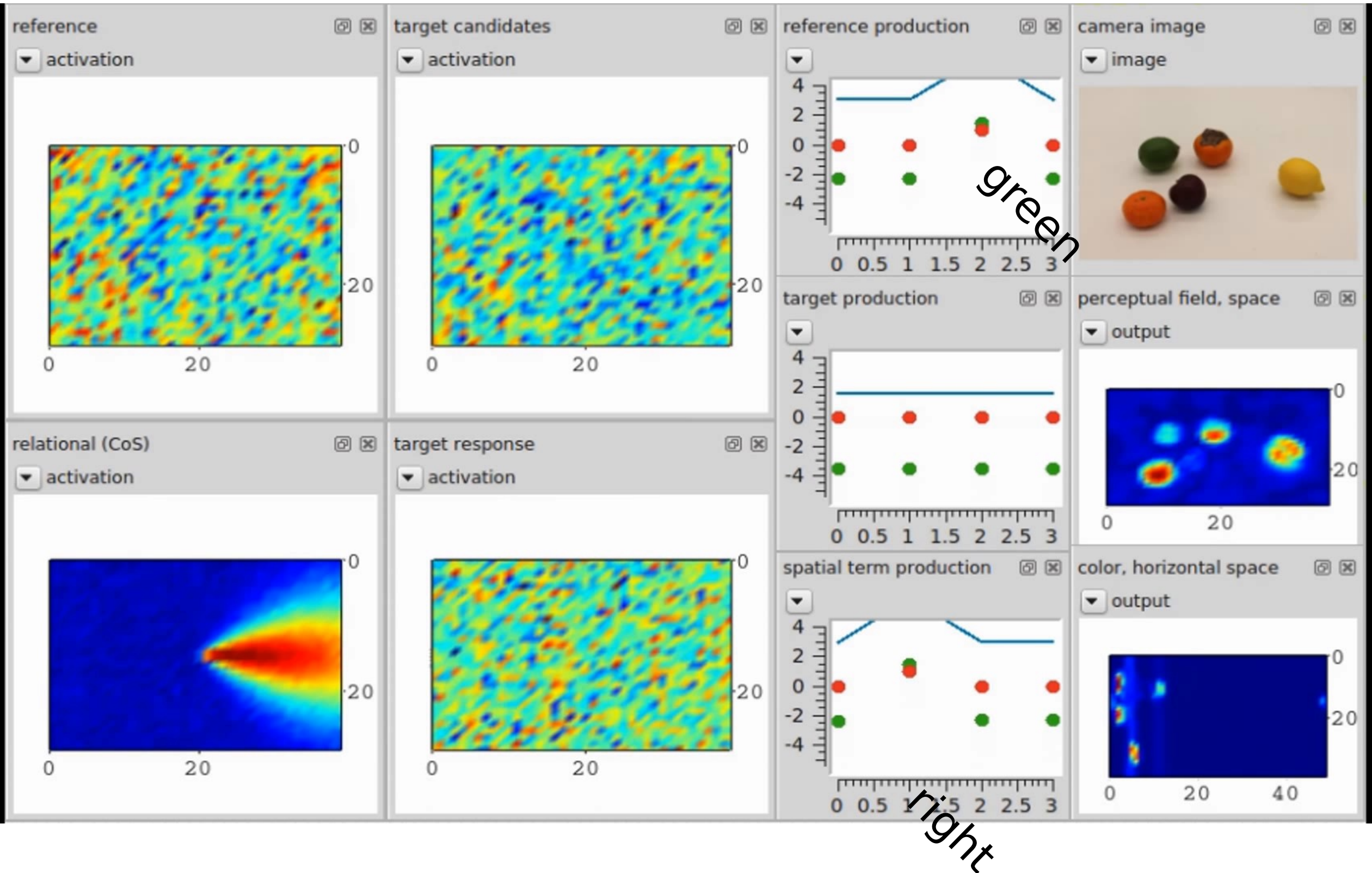
condition of dissatisfaction



[Richter, Lins et al. ICANN 2014]

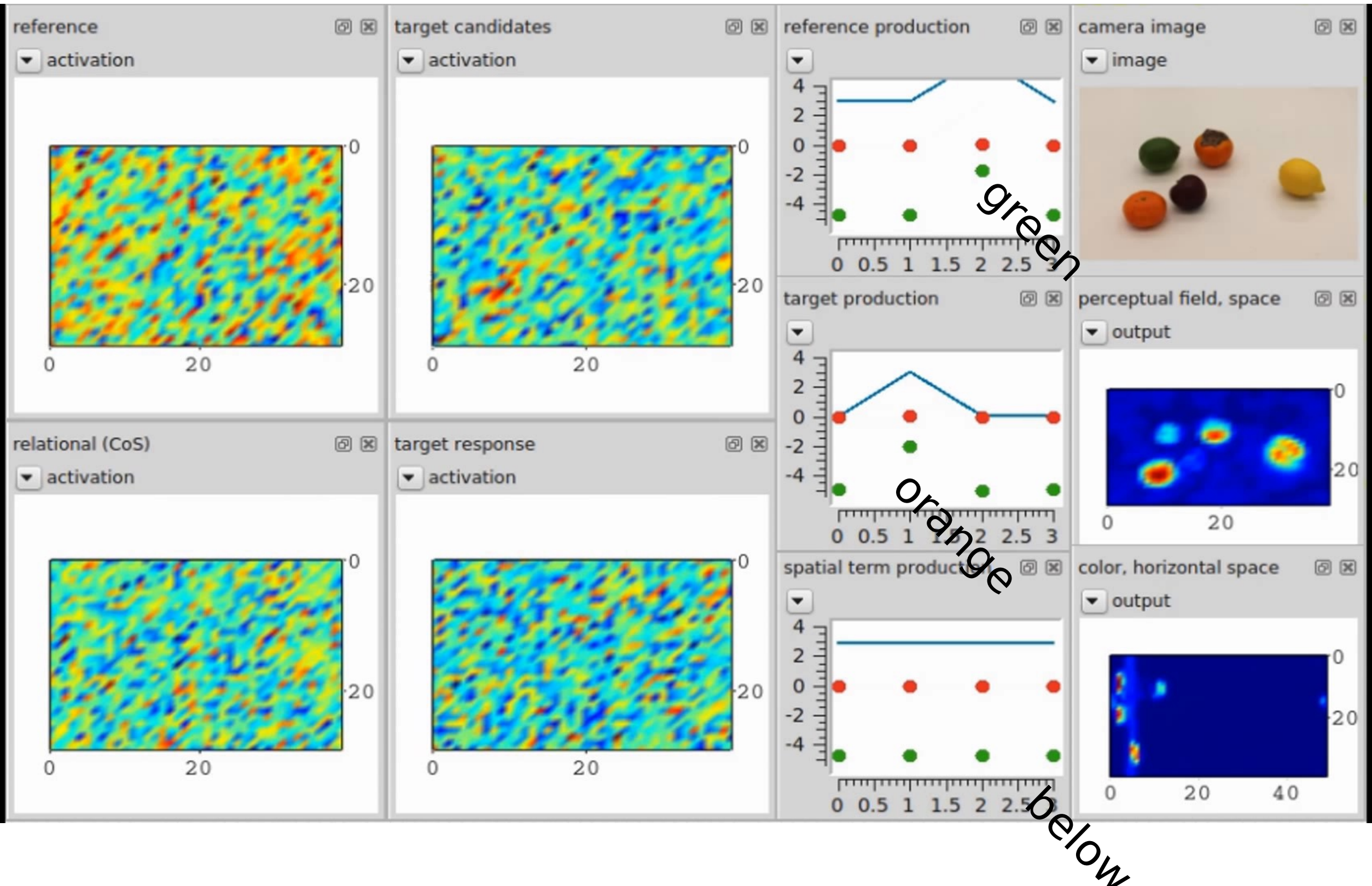


■ what is to the right of the green object?





■ where is the orange relative to the green object



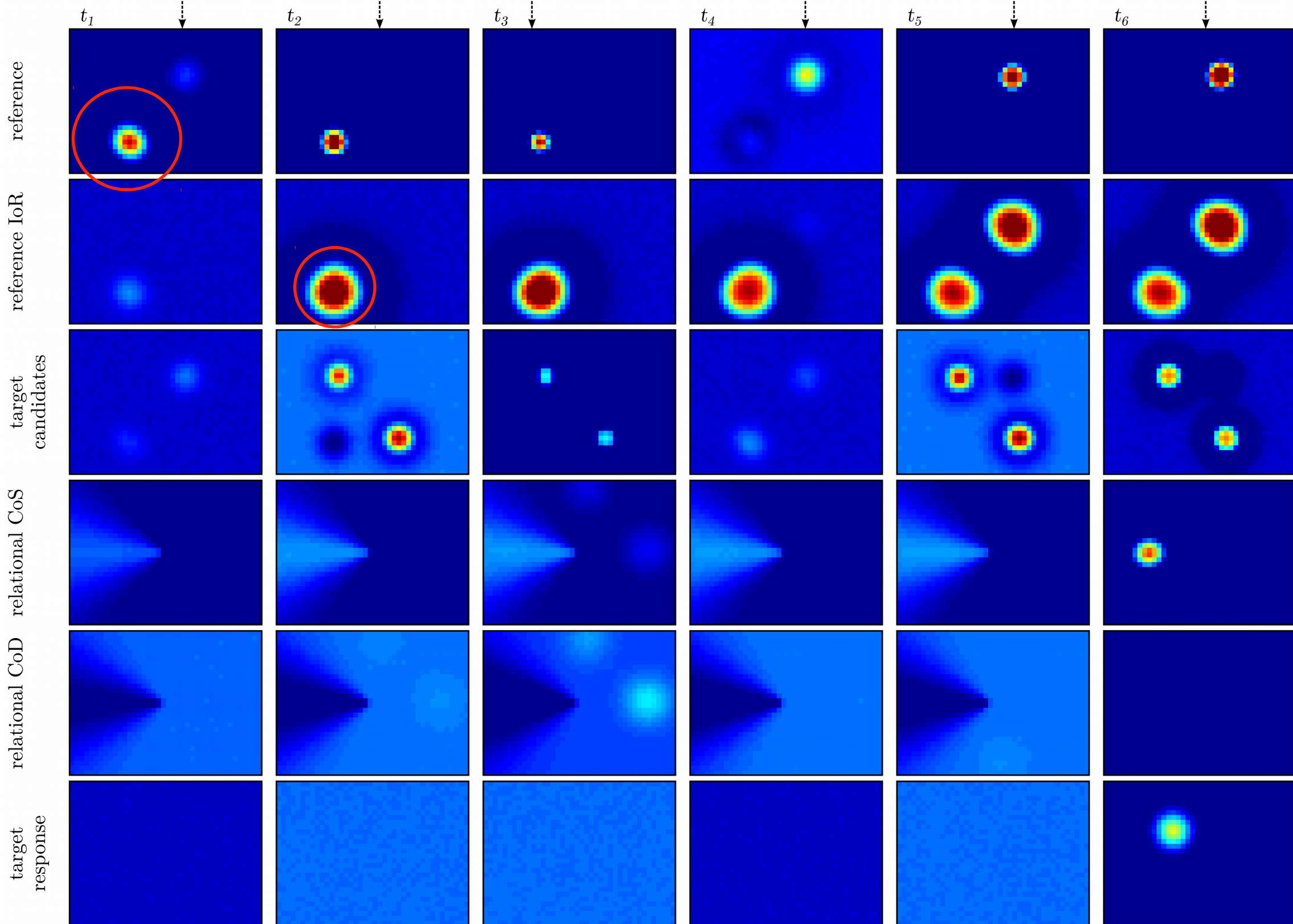
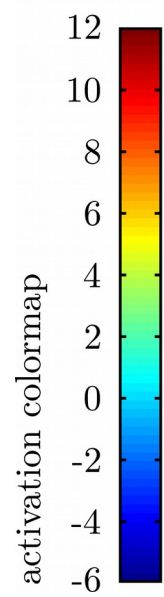
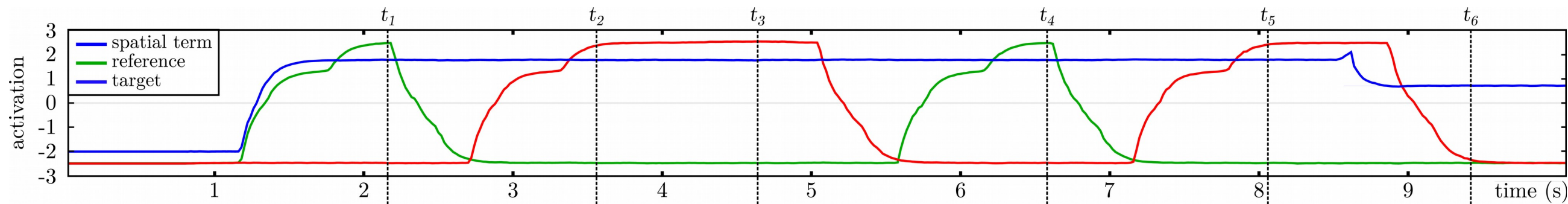
# Autonomous hypothesis testing



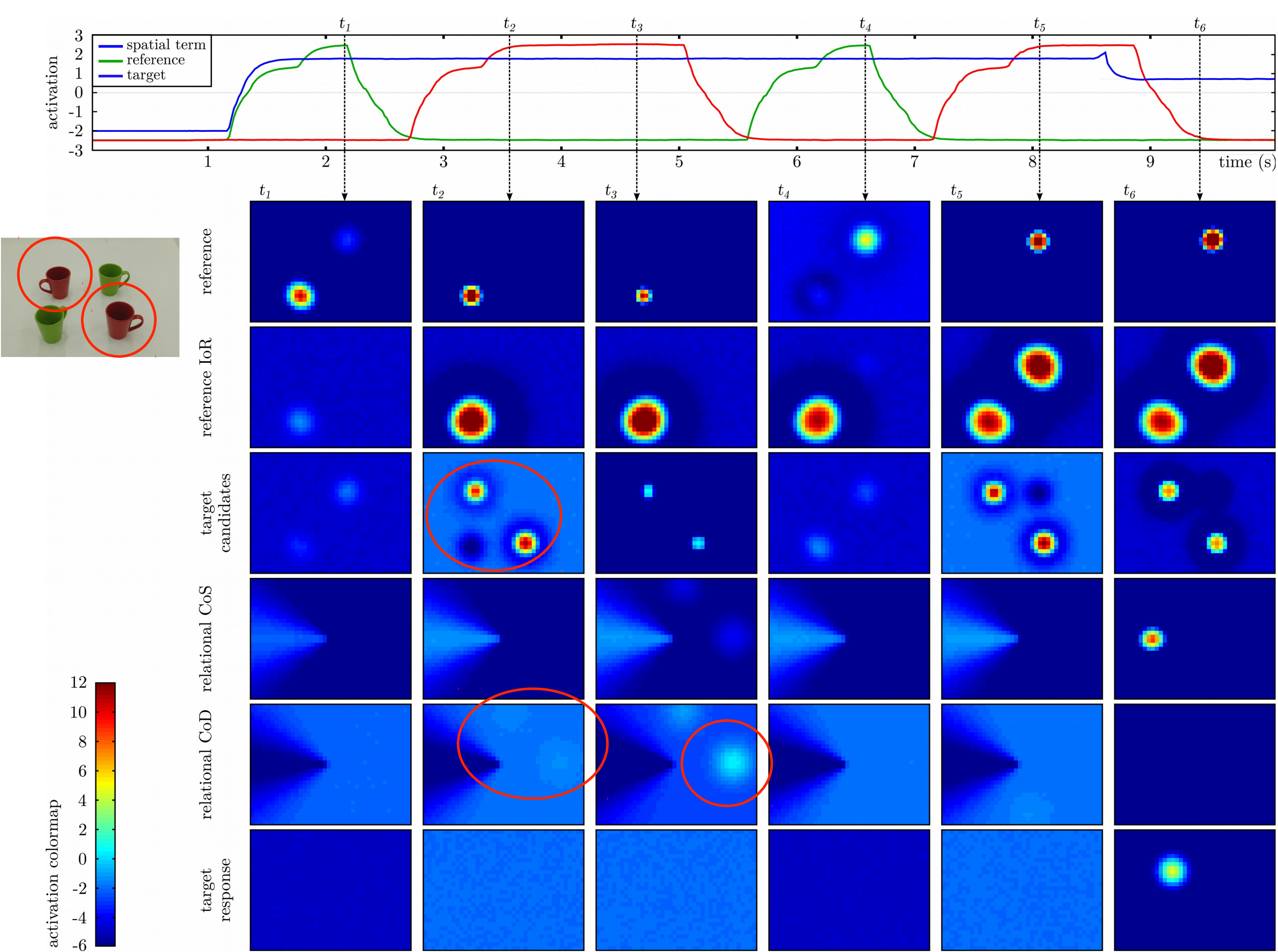
“the cup that is to the left of the green cup”

[Richter, Lins et al, CogSci 2014]

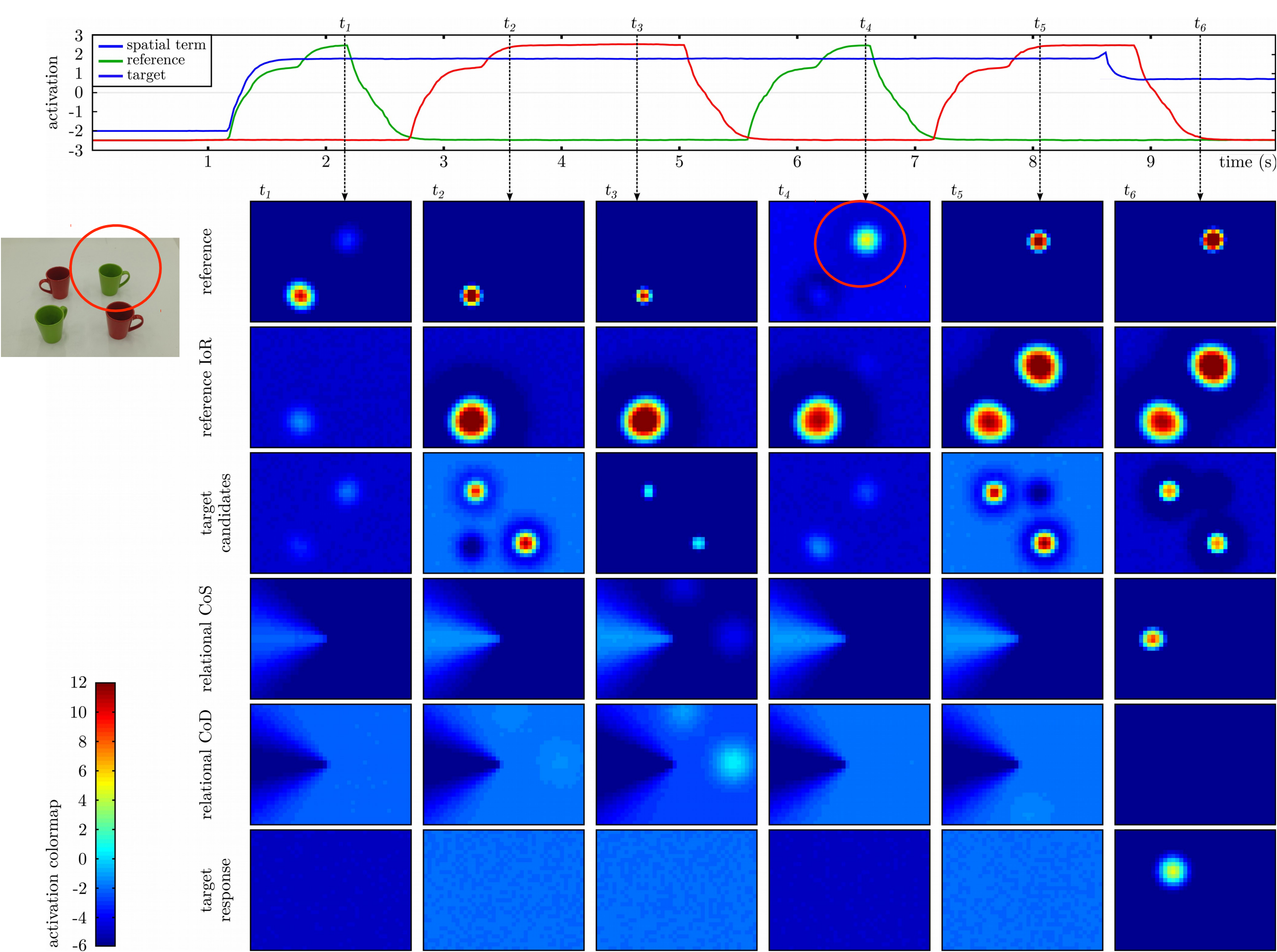




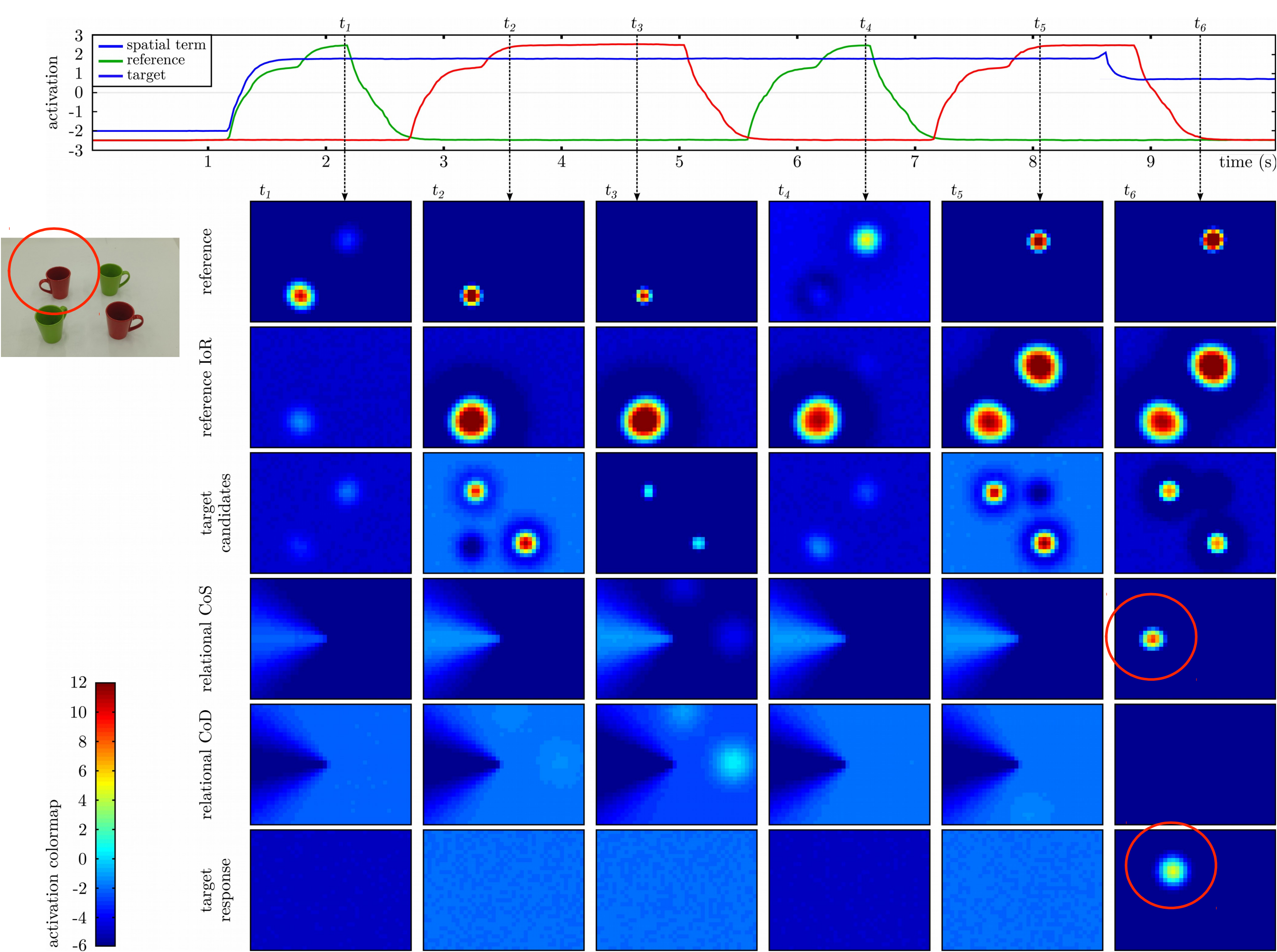














■ “find red to the left of green”

