Neural Dynamics For Embodied Cognition

Gregor Schöner, Jan Tekülve
Institute for Neural Computation (INI)
Faculty of Computer Science
Ruhr-University Bochum

Mathis Richter
Research group
Neuromorphics
Intel Corp.

Roadmap

- Neuro-physics
- Neural dynamics
- Recurrent neural dynamics
- Neural fields: dynamics
- Neural fields: dimensions
- Binding
- Sequences
- Coordinate transforms
- Relational concepts, grounding, mental mapping
- Conclusions

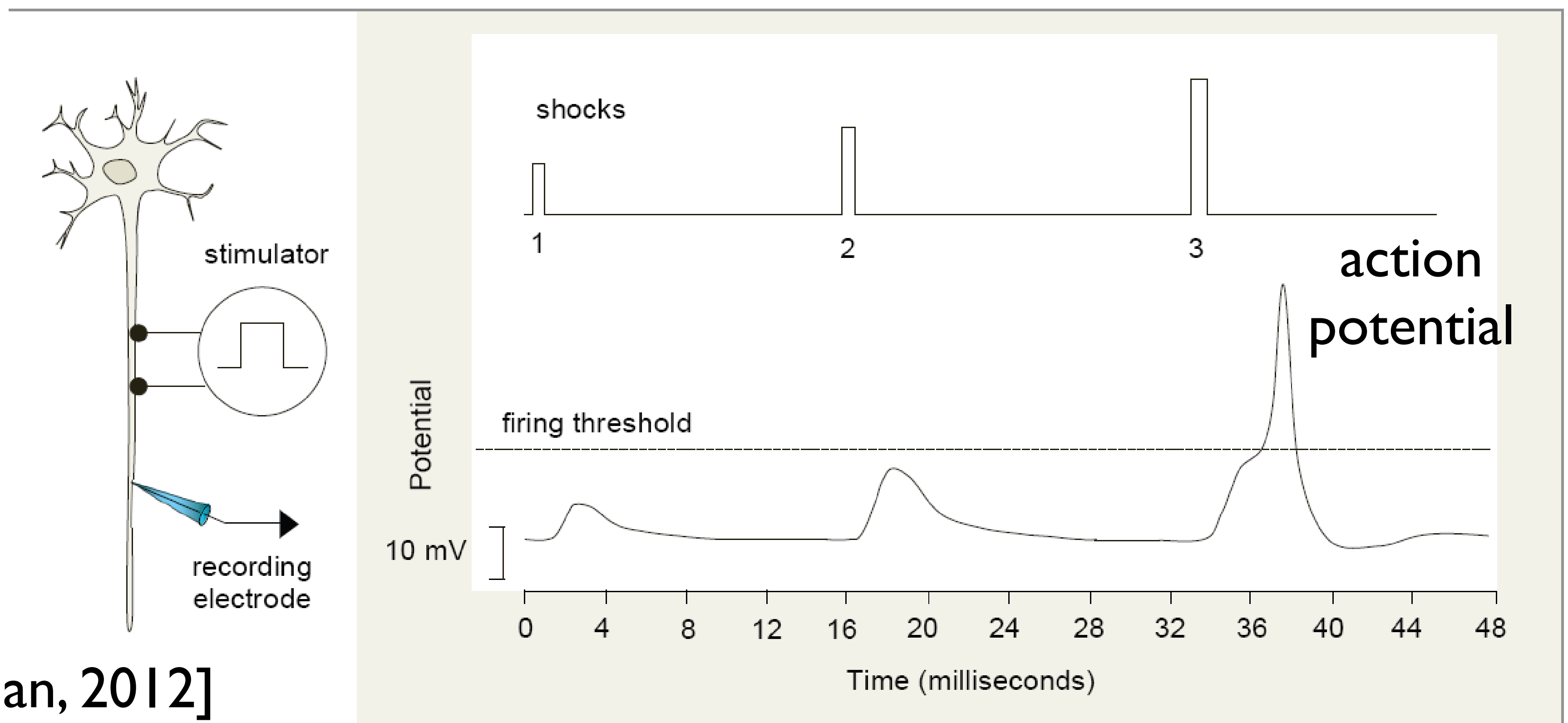
Neuro-physics

- membrane potential, $u(t)$, evolves as a dynamical system

$$\tau \dot{u}(t) = -u(t) + h + \text{input}(t)$$

$$\tau \approx 10 \text{ ms time scale}$$

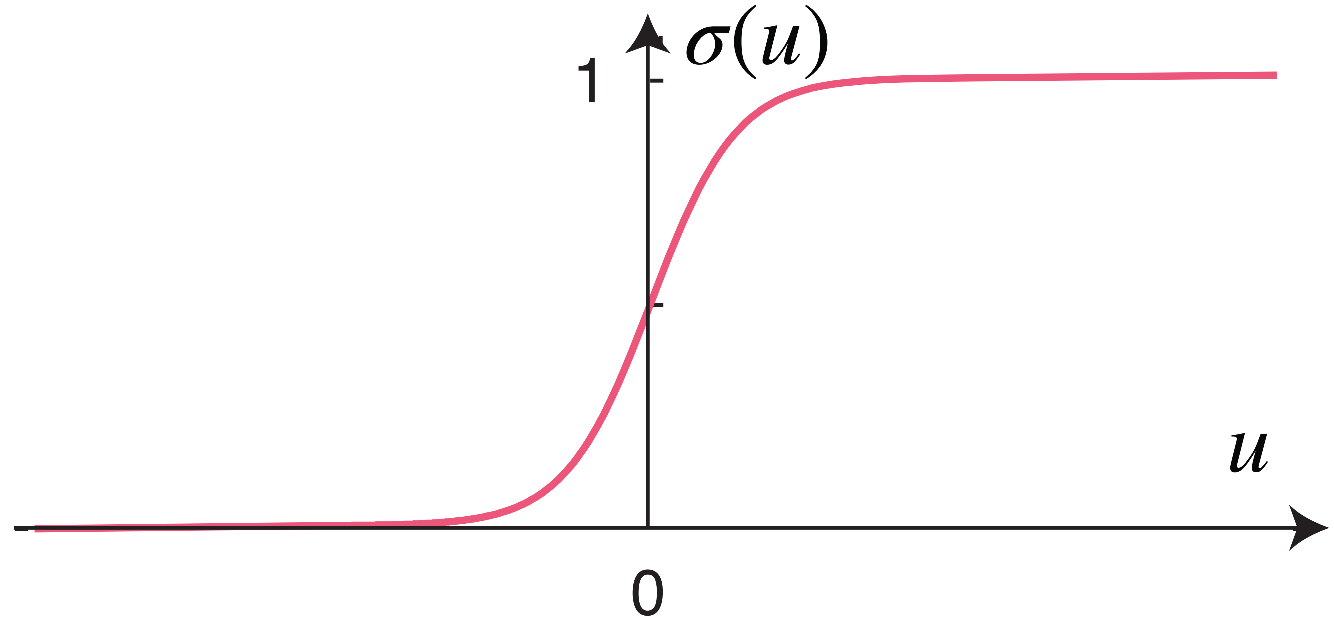
- only when membrane potential exceeds a threshold is activation transmitted to downstream neurons



[from: Tresilian, 2012]

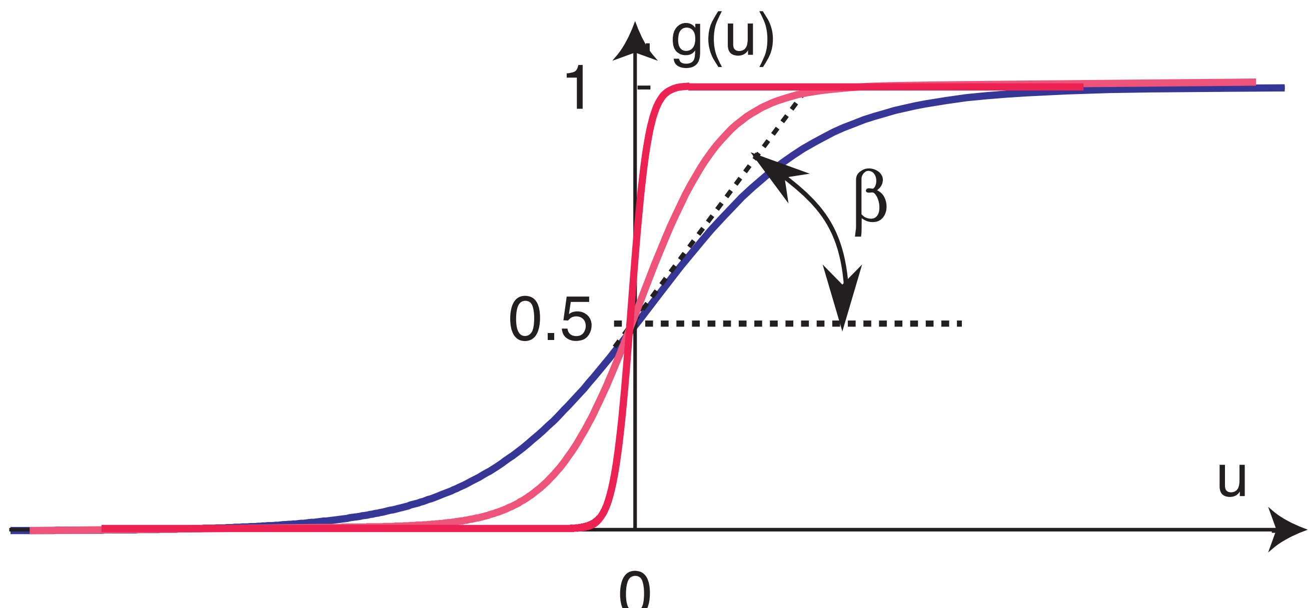
Neural dynamics

- spiking mechanism replaced by a threshold function
- that captures the effective transmission of spikes in populations



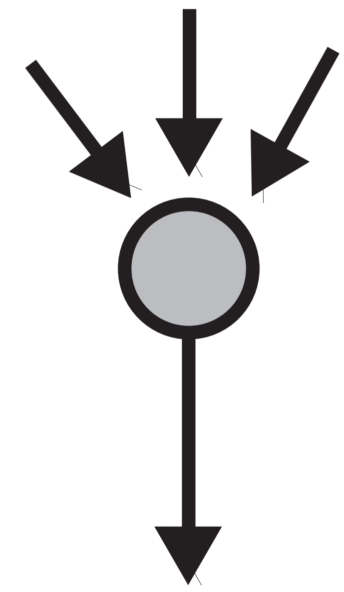
Neural dynamics

- replace spiking mechanism by sigmoid:
  - low levels of activation: not transmitted to downstream systems
  - high levels of activation: transmitted to downstream systems
- abstracting from biophysical details ~ population level membrane potential

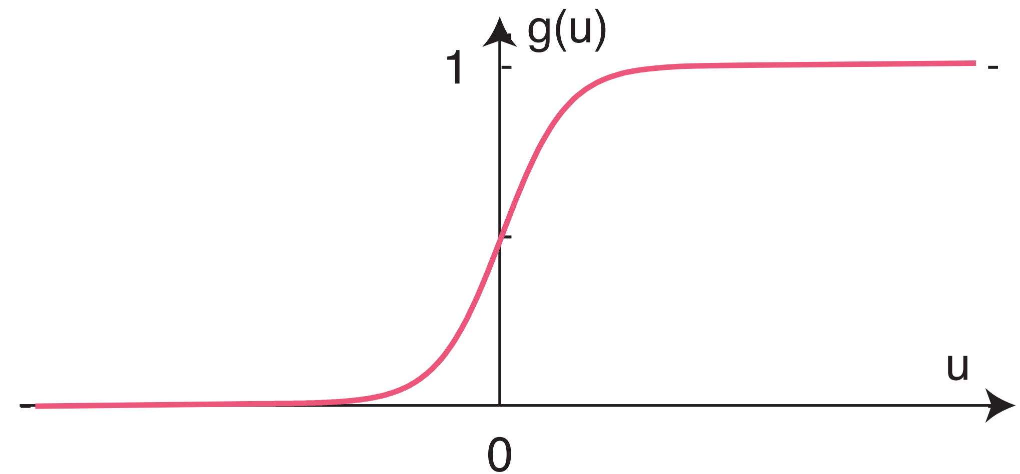


Connectionism

- employs the same abstraction: "neurons" sum input activations and pass them through a sigmoidal threshold function



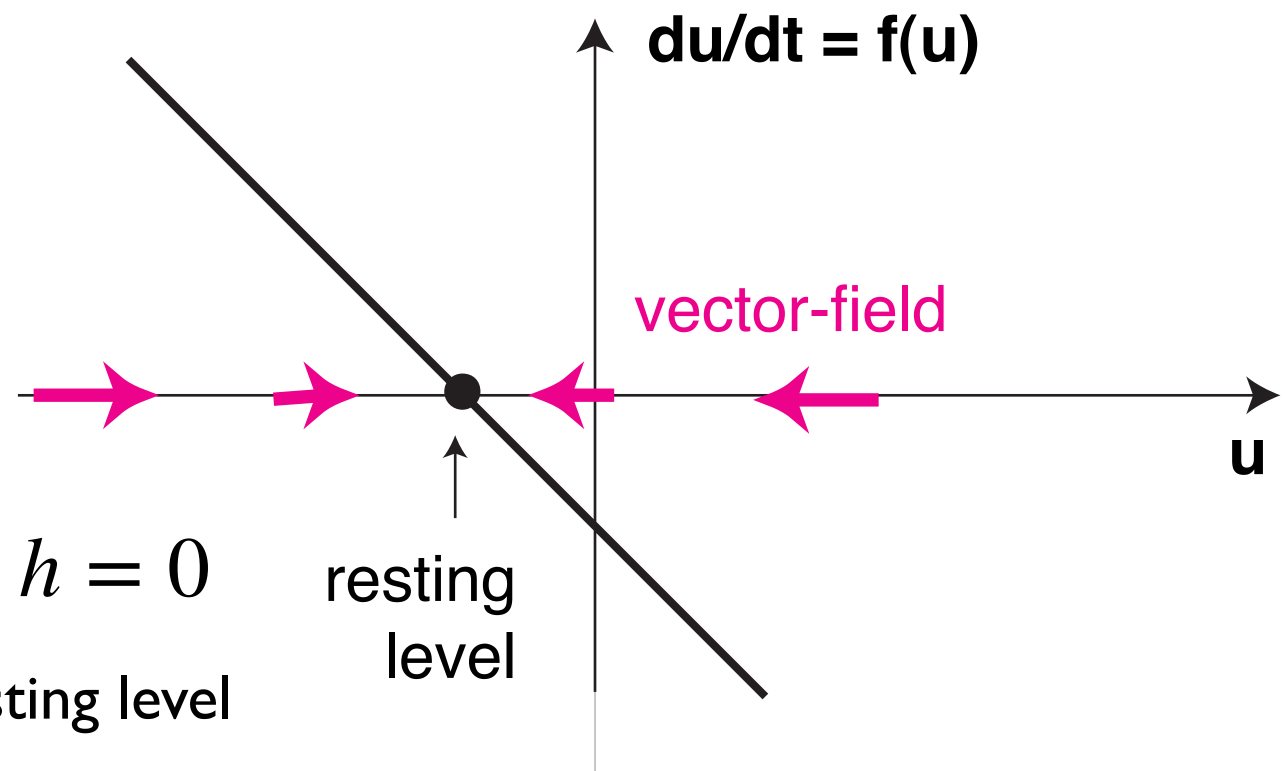
$$\text{output} = g\left(\sum(\text{inputs})\right)$$



Neural dynamics

- dynamical system: the present determines the future
- fixed point = constant solution = stationary state
- stable fixed point = attractor: nearby solutions converge to the fixed point

$$\tau \dot{u}(t) = -u(t) + h$$



$$\tau \dot{u}_{\mathrm{fp}} = -u_{\mathrm{fp}} + h = 0$$

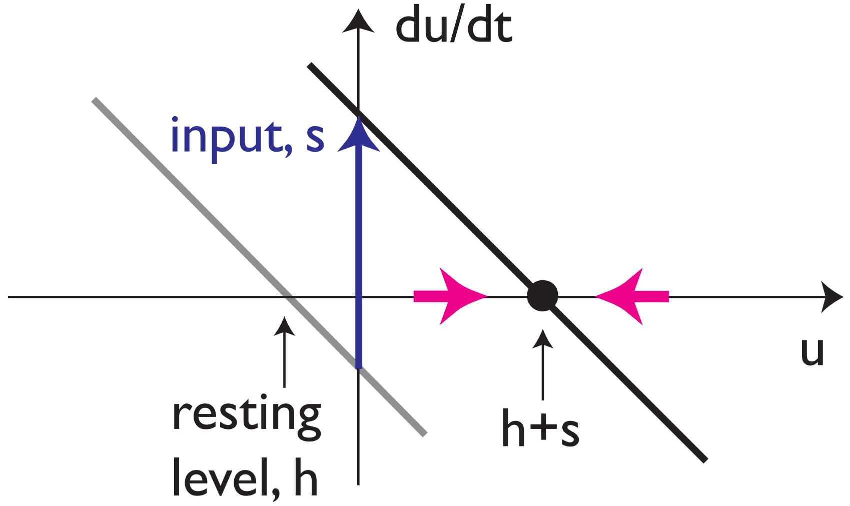
$$\Rightarrow u_{\mathrm{fp}} = h$$ resting level

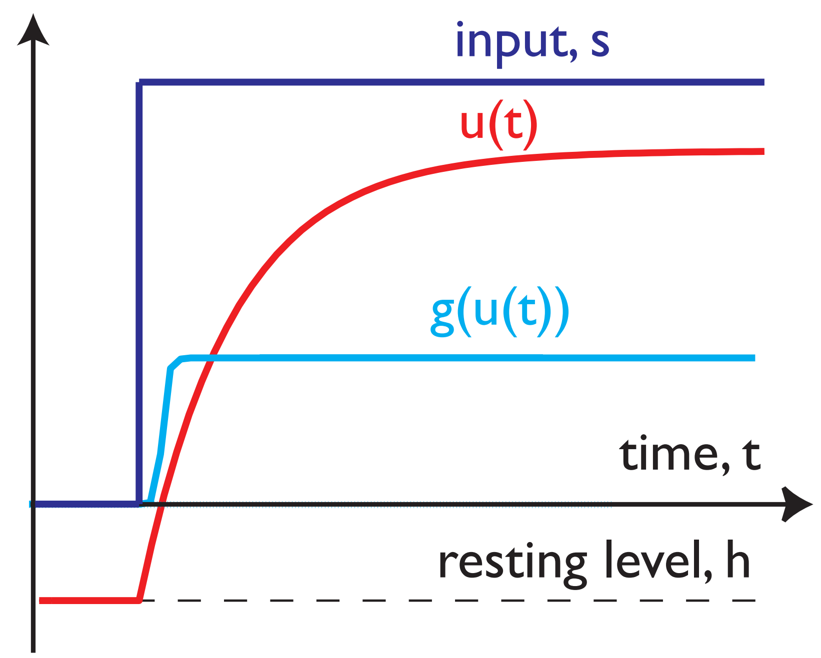
Neural dynamics

- inputs add to the rate of change of activation
  - positive: excitatory
  - negative: inhibitory

$$\tau\dot{u}(t) = -u(t) + h + s(t)$$

- input shifts the attractor
- activation tracks this shift
- $\sigma(u(t))$ transmitted to down-stream neurons



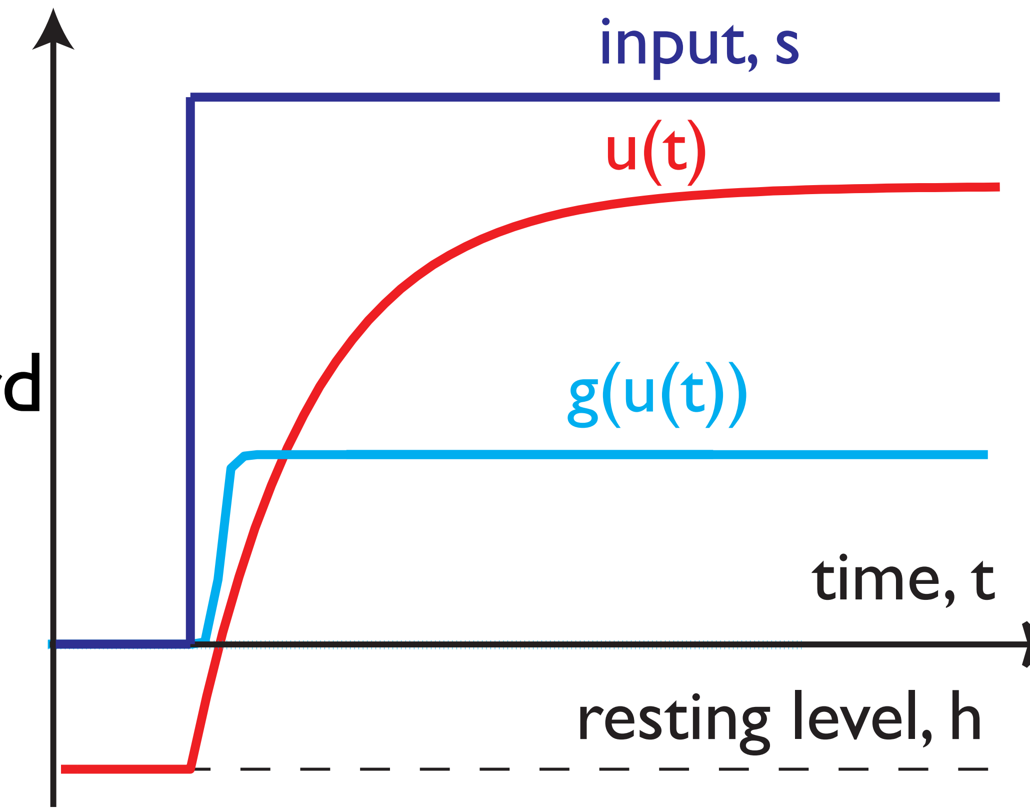


($\sigma(u)$ and $g(u)$ used interchangeably)

Neural dynamics

- so far, the dynamics just does low-pass filtering… (smoothing the time course)
- that would change as a step-function in a forward neural network
- when does neural dynamics make a real difference?

$$\text{output} = g\left(\sum(\text{inputs})\right)$$

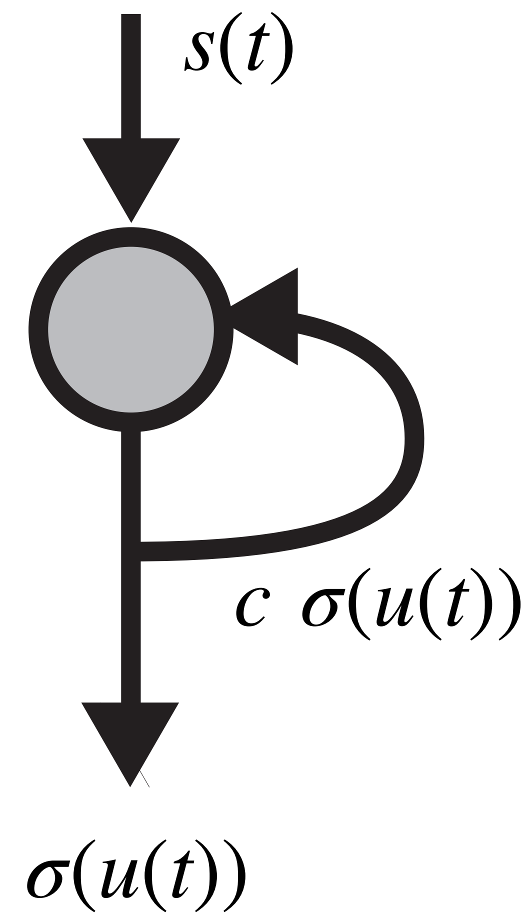


Roadmap

- Neuro-physics
- Neural dynamics
- Recurrent neural dynamics
- Neural fields: dynamics
- Neural fields: dimensions
- Binding
- Sequences
- Coordinate transforms
- Relational concepts, grounding, mental mapping
- Conclusions

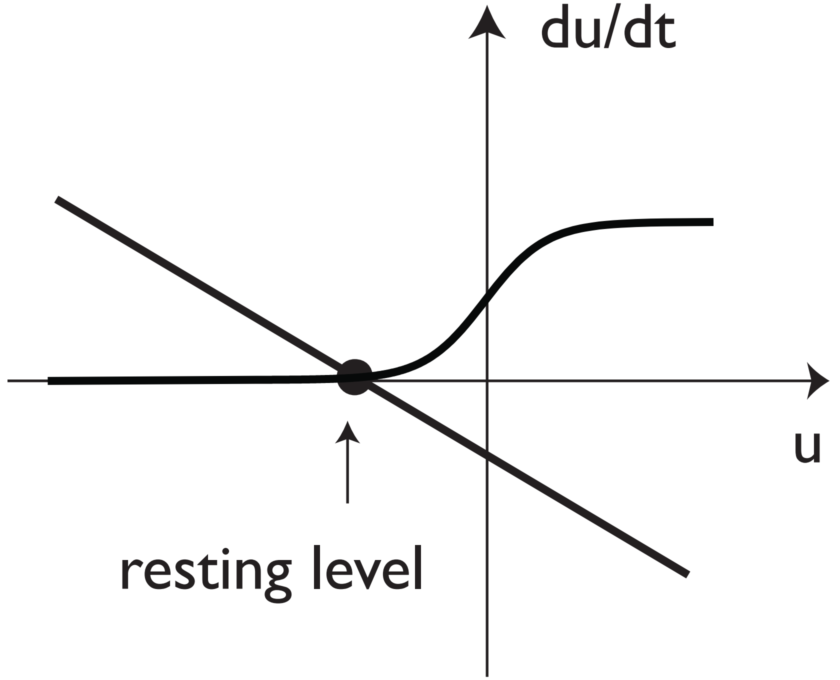
Neuronal dynamics with excitatory recurrent connection = interaction

- in recurrent networks, time is conceptually necessary as some inputs are outputs from the same neuron/population …
- “past outputs are new input”
- => dynamics

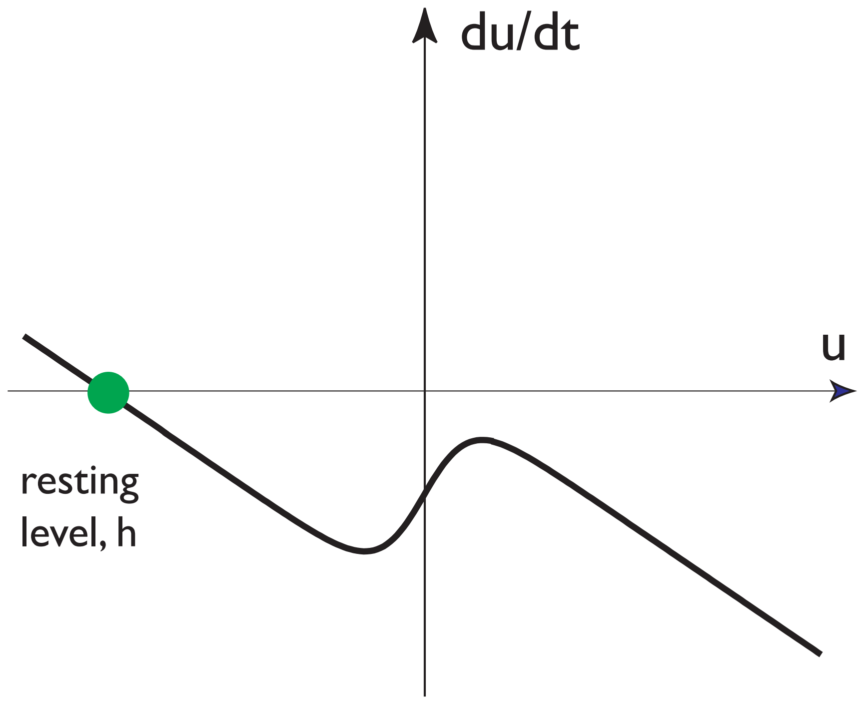


$$\tau \dot{u}(t) = -u(t) + h + s(t) + c\ \sigma(u(t))$$

Neuronal dynamics with self-excitation



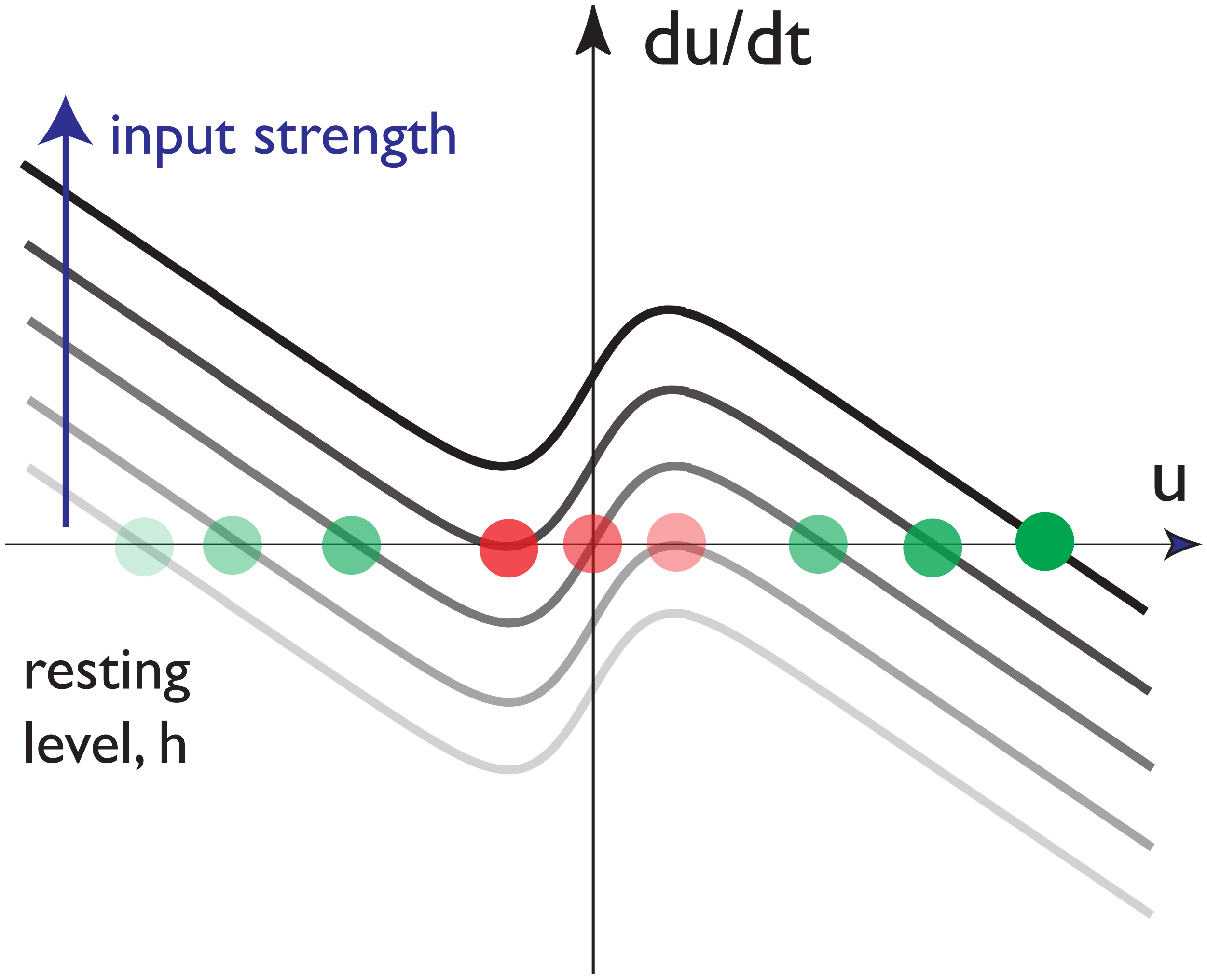
- nonlinear dynamics!



$$\tau \dot{u}(t) = -u(t) + h + s(t) + c\ \sigma(u(t))$$

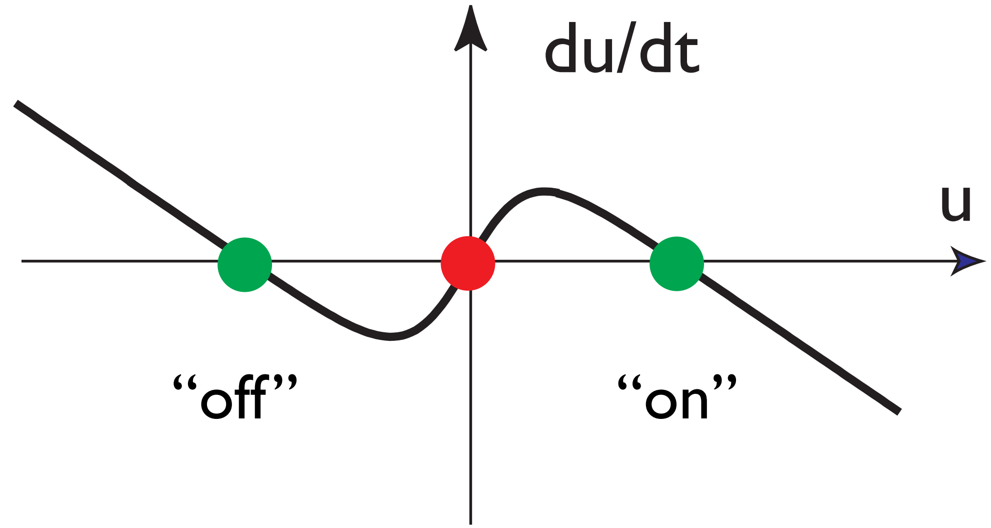
Neuronal dynamics with self-excitation

- varying input
- => number of attractors changes

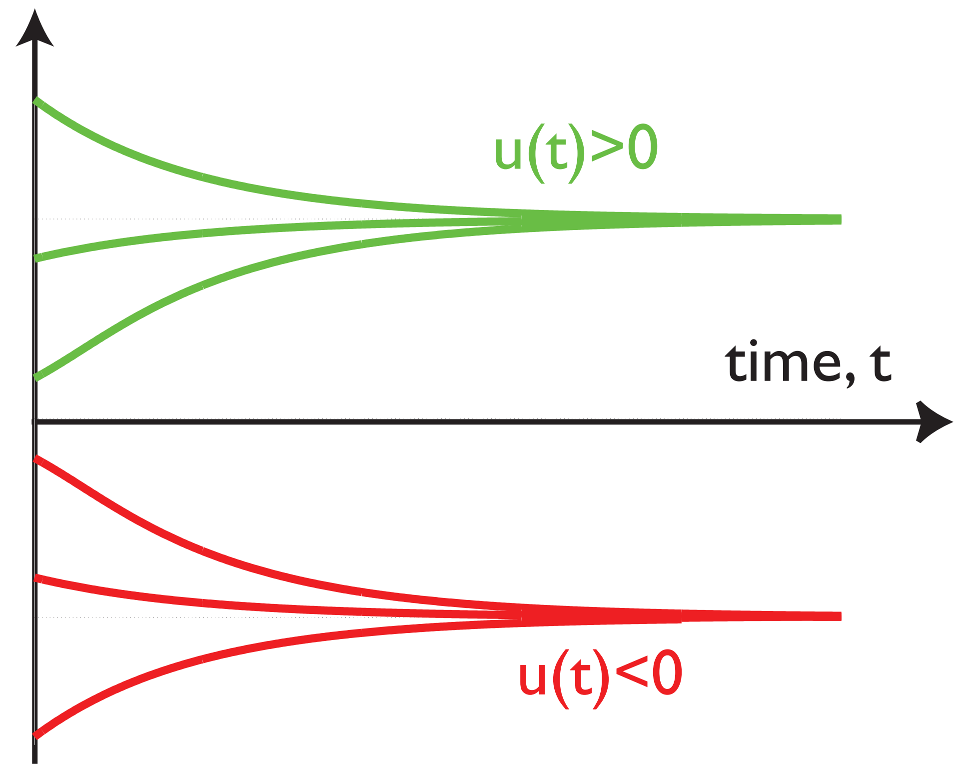


$$\tau \dot{u}(t) = -u(t) + h + s(t) + c\ \sigma(u(t))$$

Neuronal dynamics with self-excitation



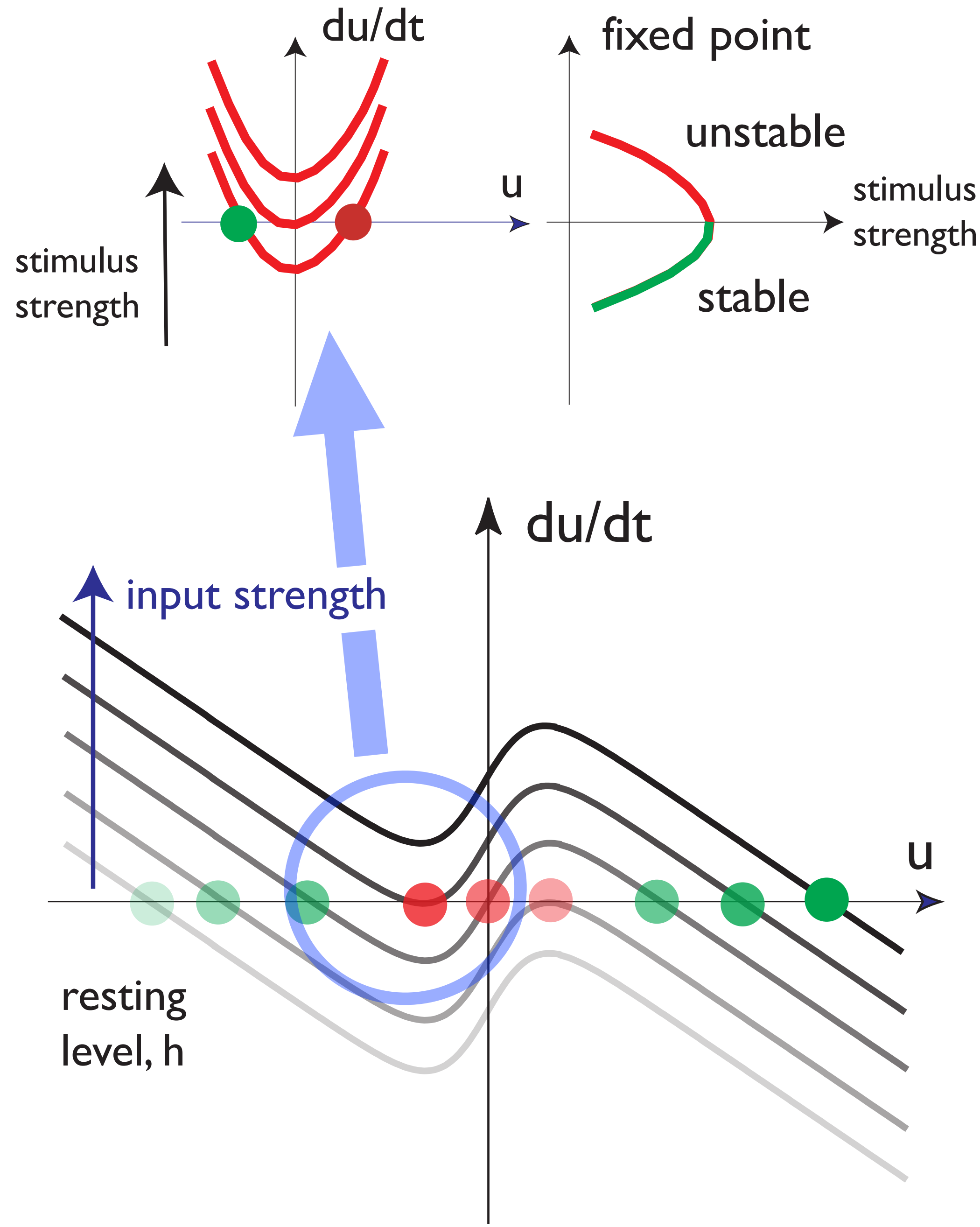
- at intermediate input levels: bistable dynamics
- “on” vs “off” state



$$\tau \dot{u}(t) = -u(t) + h + s(t) + c\ \sigma(u(t))$$

Neuronal dynamics with self-excitation

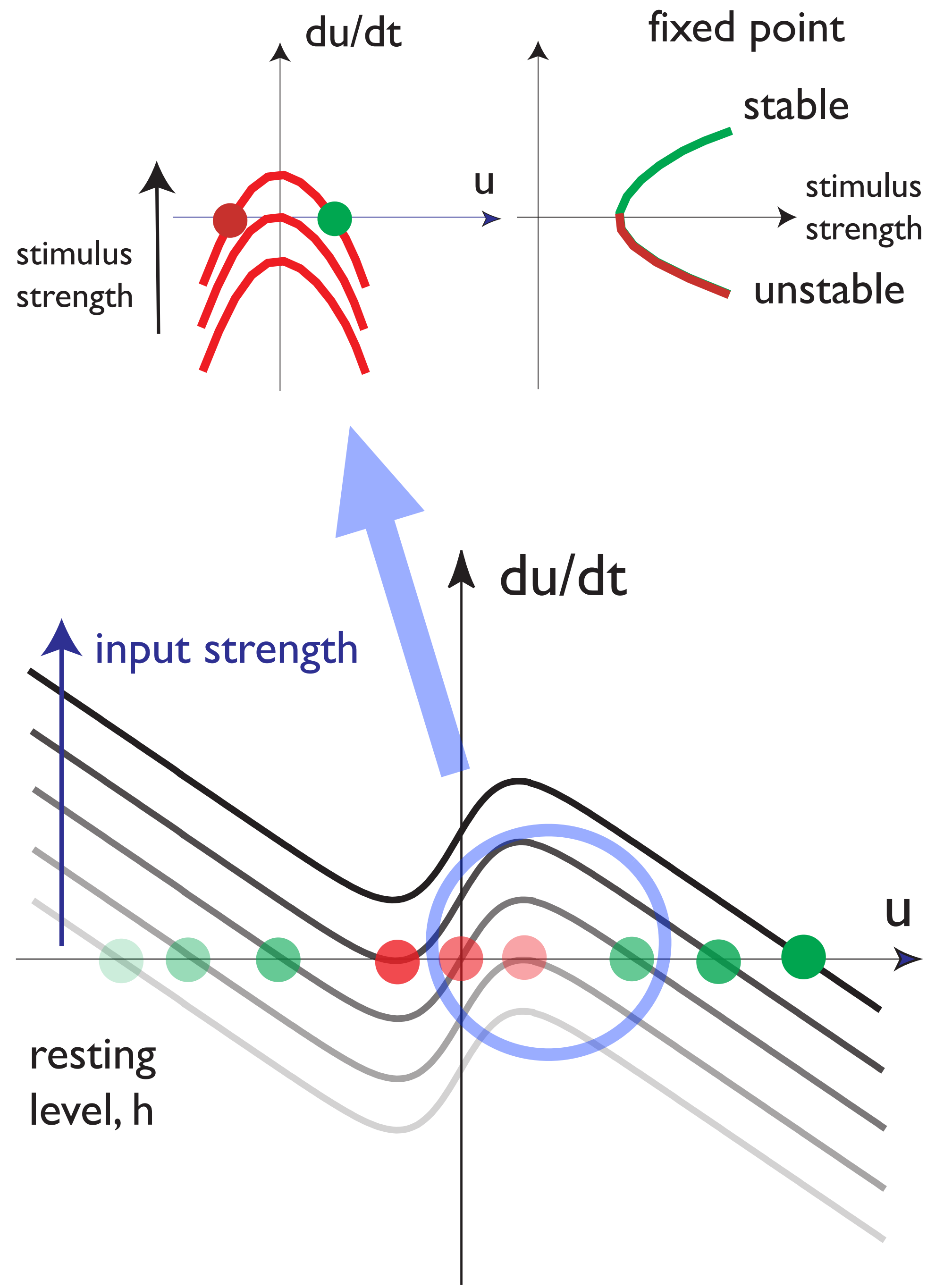
- increasing input strength => detection instability



$$\tau\dot{u}(t) = -u(t) + h + s(t) + c\ \sigma(u(t))$$

Neuronal dynamics with self-excitation

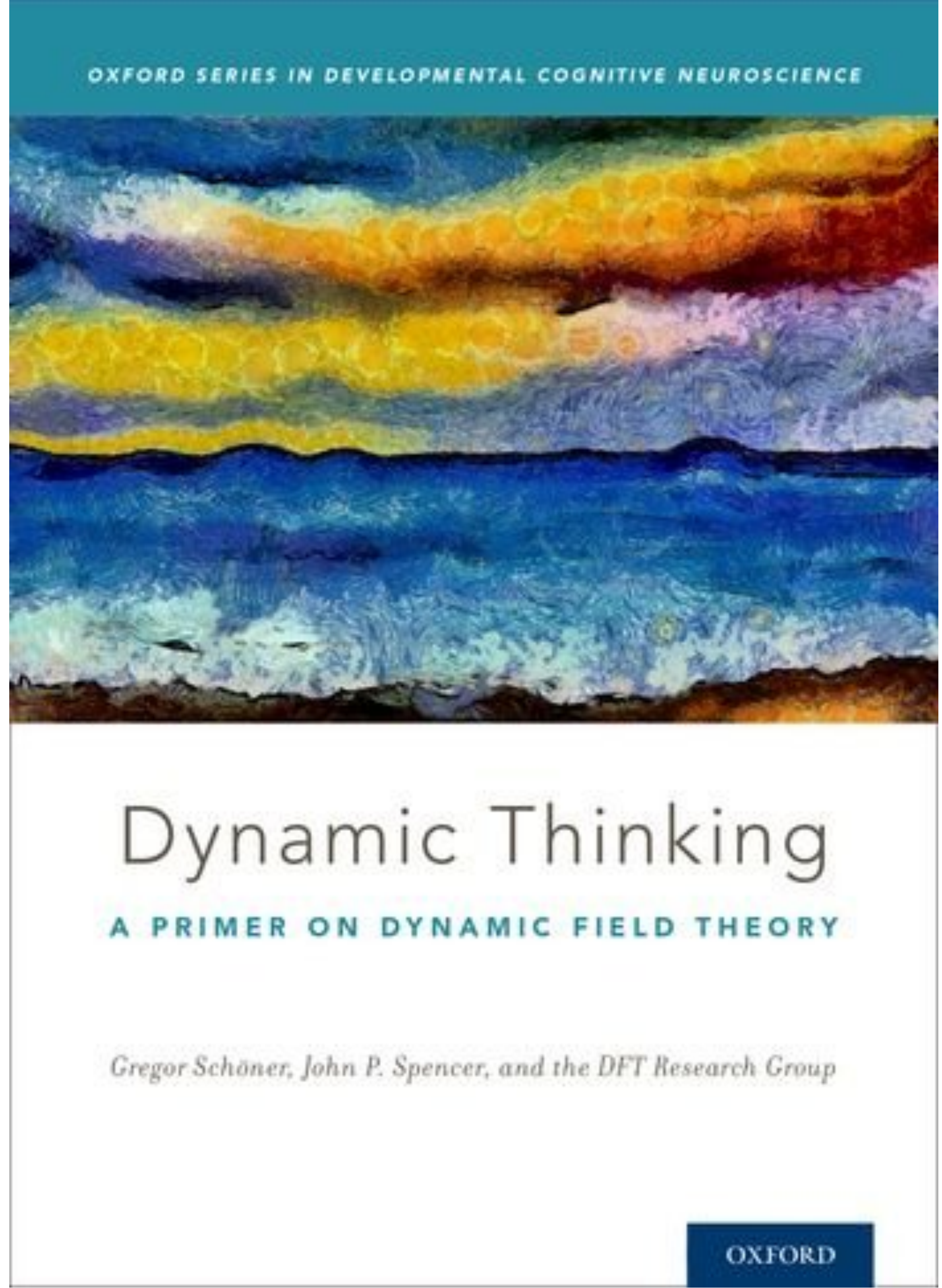
- decreasing input strength => reverse detection instability



$$\tau \dot{u}(t) = -u(t) + h + s(t) + c\ \sigma(u(t))$$

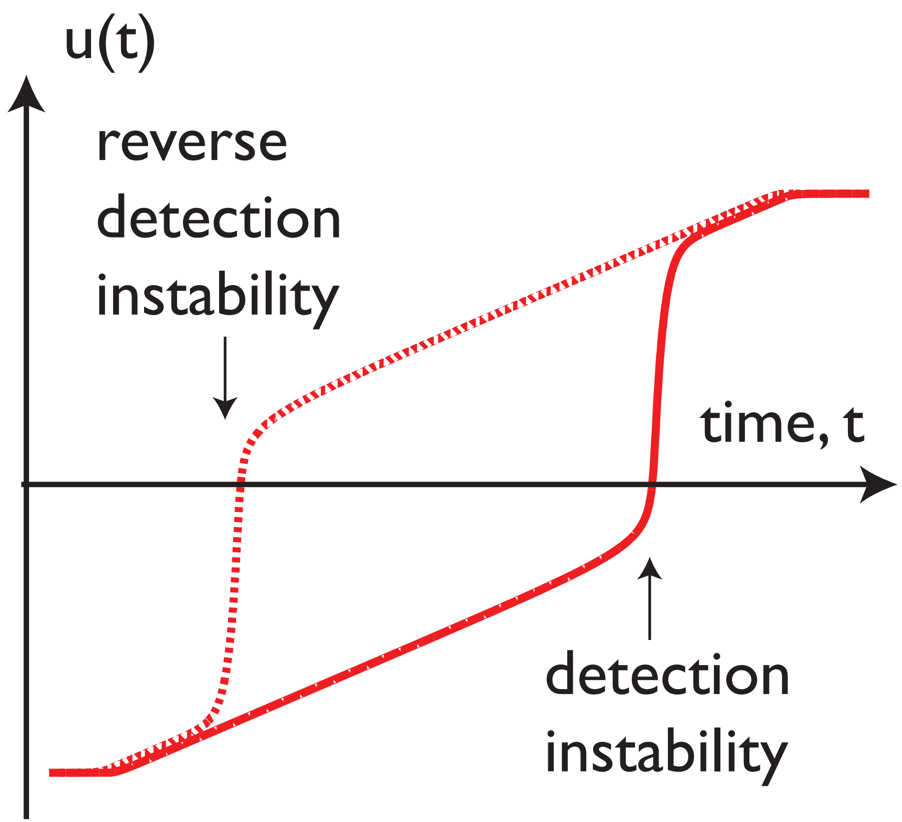
=> simulation

- dynamicfieldtheory.org



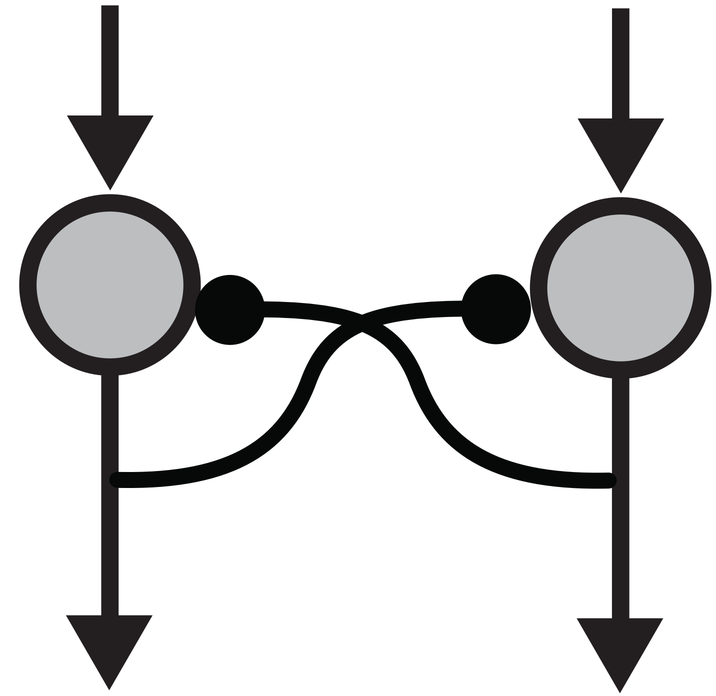
Neuronal dynamics with self-excitation

- the detection and its reverse create events at discrete times from time-continuous changes



$$\tau\dot{u}(t) = -u(t) + h + s(t) + c\ \sigma(u(t))$$

Neuronal dynamics with inhibitory recurrent connectivity



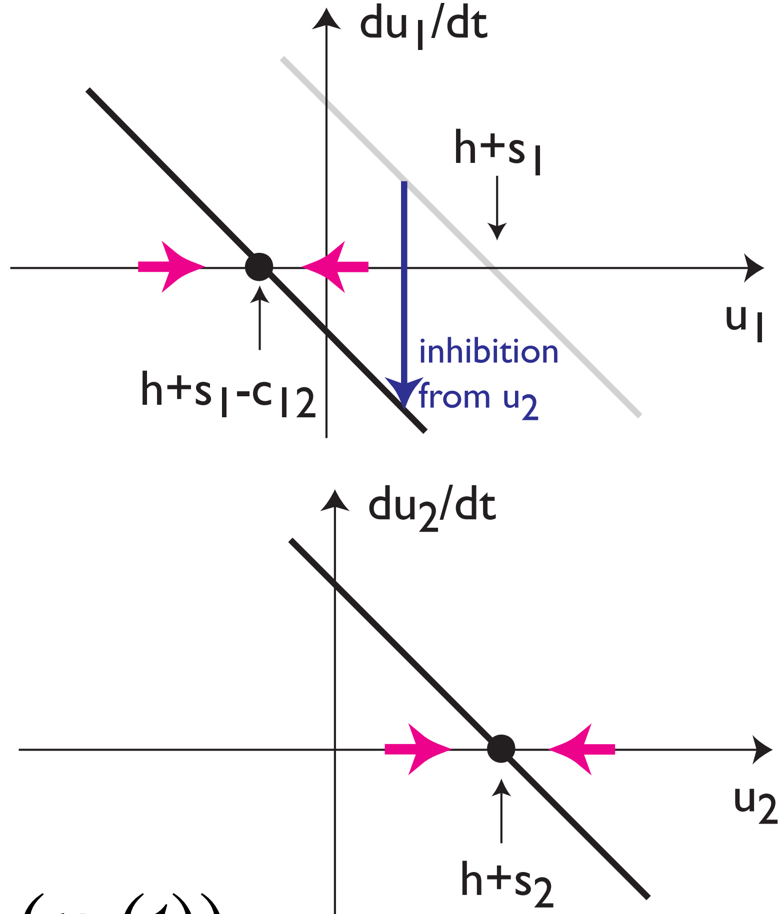
coupling/interaction

$$\tau \dot{u}_1(t) = -u_1(t) + h + s_1(t) - c_{12}\sigma(u_2(t))$$

$$\tau \dot{u}_2(t) = -u_2(t) + h + s_2(t) - c_{21}\sigma(u_1(t))$$

Neuronal dynamics with competition

- assume $u_2 > 0$ => u_2 inhibits u_1
- => attractor for $u_1 < 0$
- => u_1 does not inhibit u_2

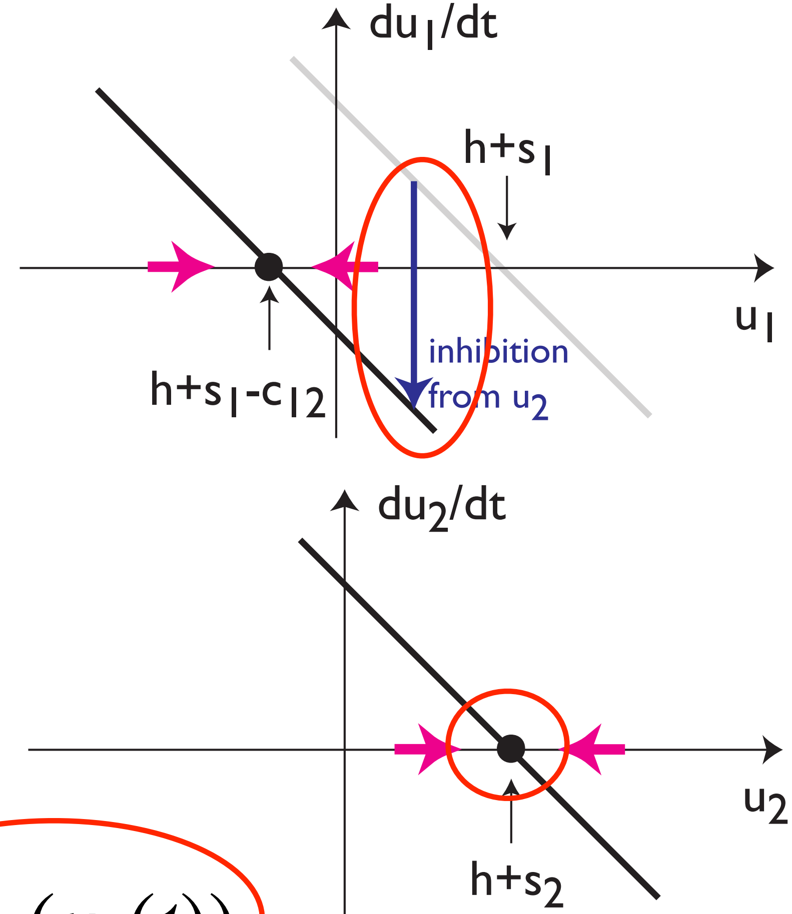


$$\tau \dot{u}_1(t) = -u_1(t) + h + s_1(t) - c_{12}\sigma(u_2(t))$$

$$\tau \dot{u}_2(t) = -u_2(t) + h + s_2(t) - c_{21}\sigma(u_1(t))$$

Neuronal dynamics with inhibitory recurrent connectivity

- assume $u_2 > 0$ => u_2 inhibits u_1
- => attractor for $u_1 < 0$
- => u_1 does not inhibit u_2

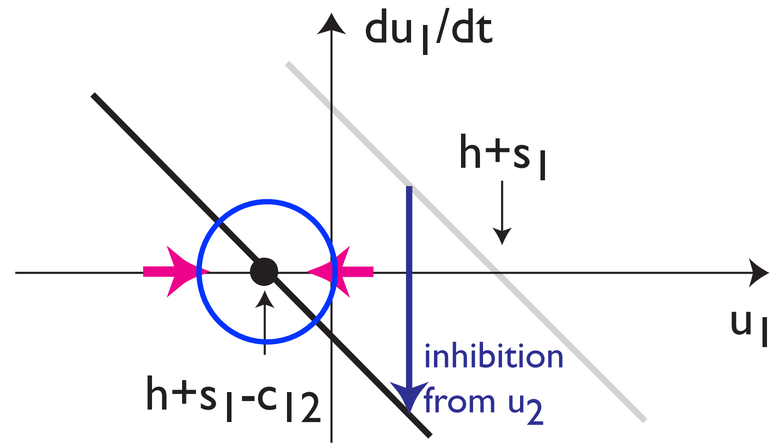


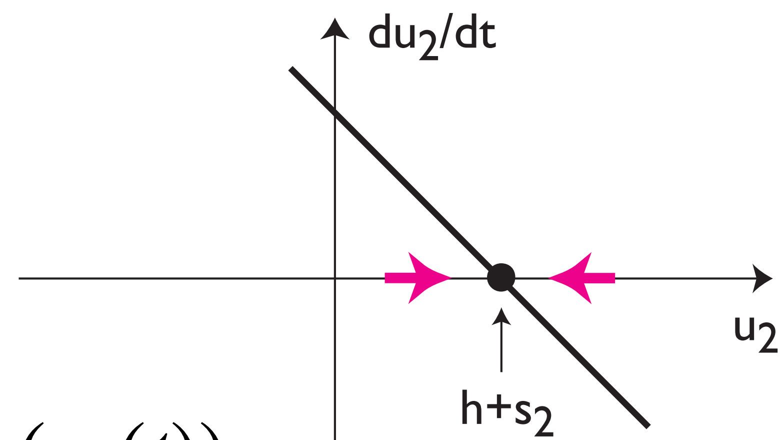
$$\tau \dot{u}_1(t) = -u_1(t) + h + s_1(t) - c_{12}\sigma(u_2(t))$$

$$\tau \dot{u}_2(t) = -u_2(t) + h + s_2(t) - c_{21}\sigma(u_1(t))$$

Neuronal dynamics with inhibitory recurrent connectivity

- assume $u_2 > 0$ => u_2 inhibits u_1
- => attractor for $u_1 < 0$
- => u_1 does not inhibit u_2



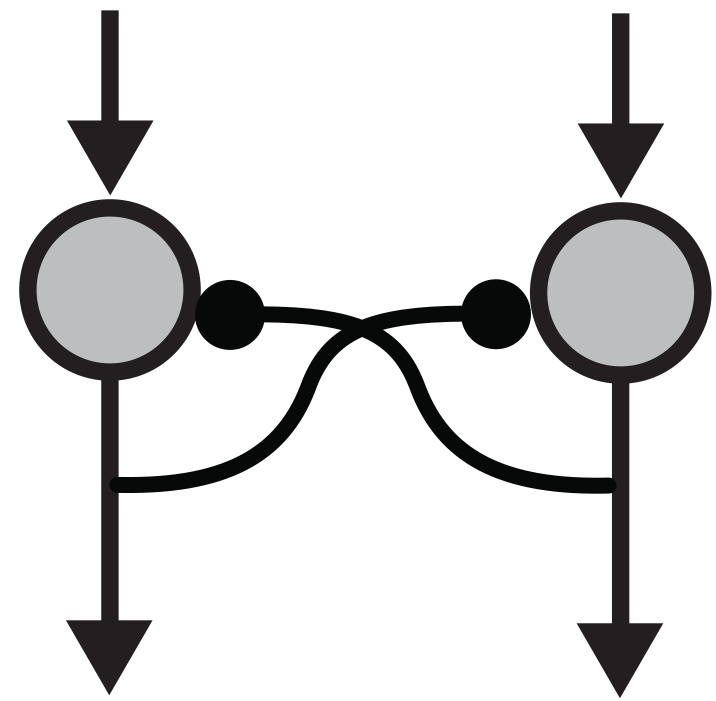


$$\tau \dot{u}_1(t) = -u_1(t) + h + s_1(t) - c_{12}\sigma(u_2(t))$$

$$\tau \dot{u}_2(t) = -u_2(t) + h + s_2(t) - c_{21}\sigma(u_1(t))$$

Neuronal dynamics with inhibitory recurrent connectivity

- by symmetry, to possible attractor stats
  - $u_2 > 0$ and $u_1 < 0$
  - $u_2 < 0$ and $u_1 > 0$
- => competition/selection

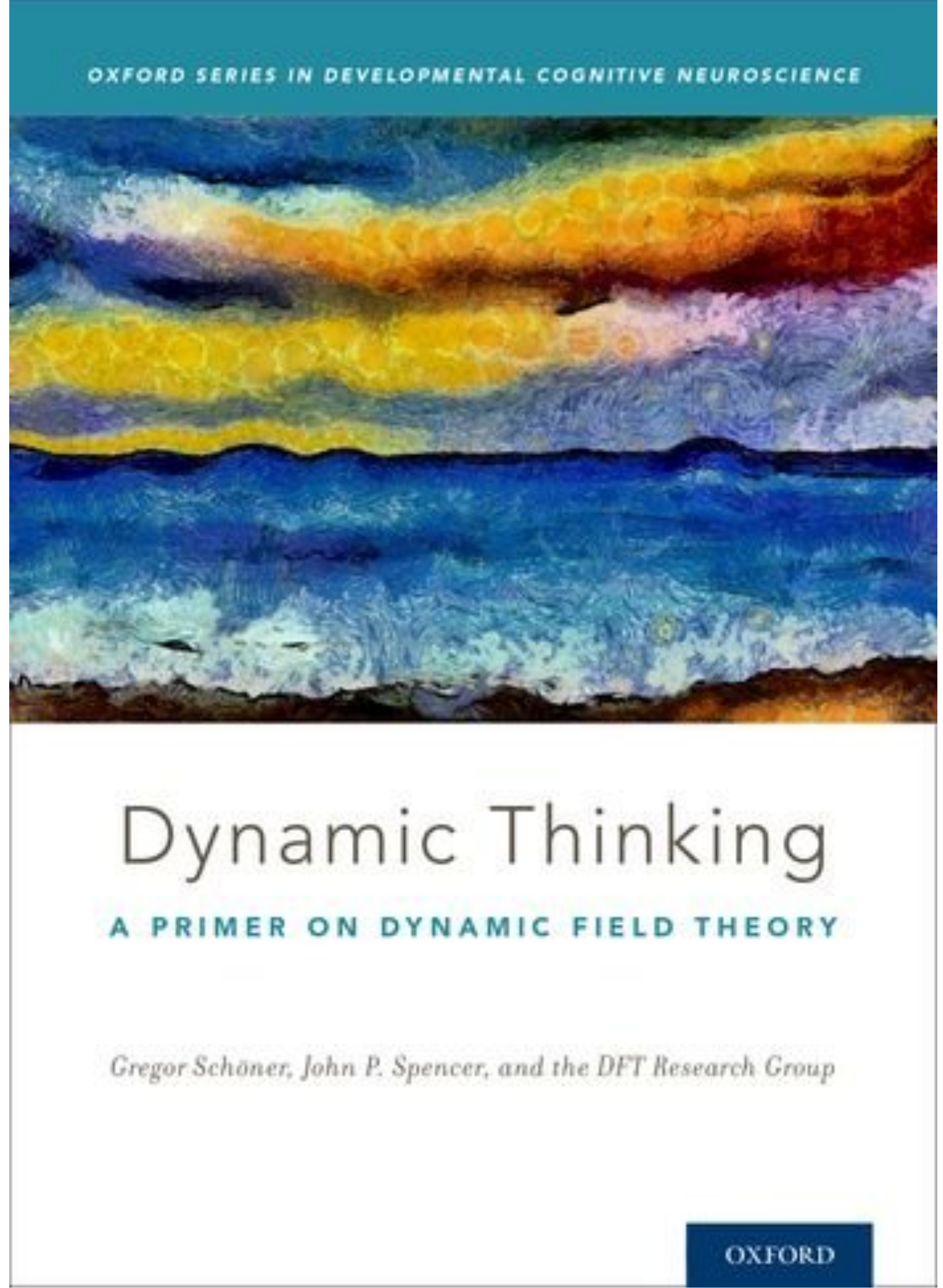


$$\tau \dot{u}_1(t) = -u_1(t) + h + s_1(t) - c_{12}\sigma(u_2(t))$$

$$\tau \dot{u}_2(t) = -u_2(t) + h + s_2(t) - c_{21}\sigma(u_1(t))$$

=> simulation

- dynamicfieldtheory.org

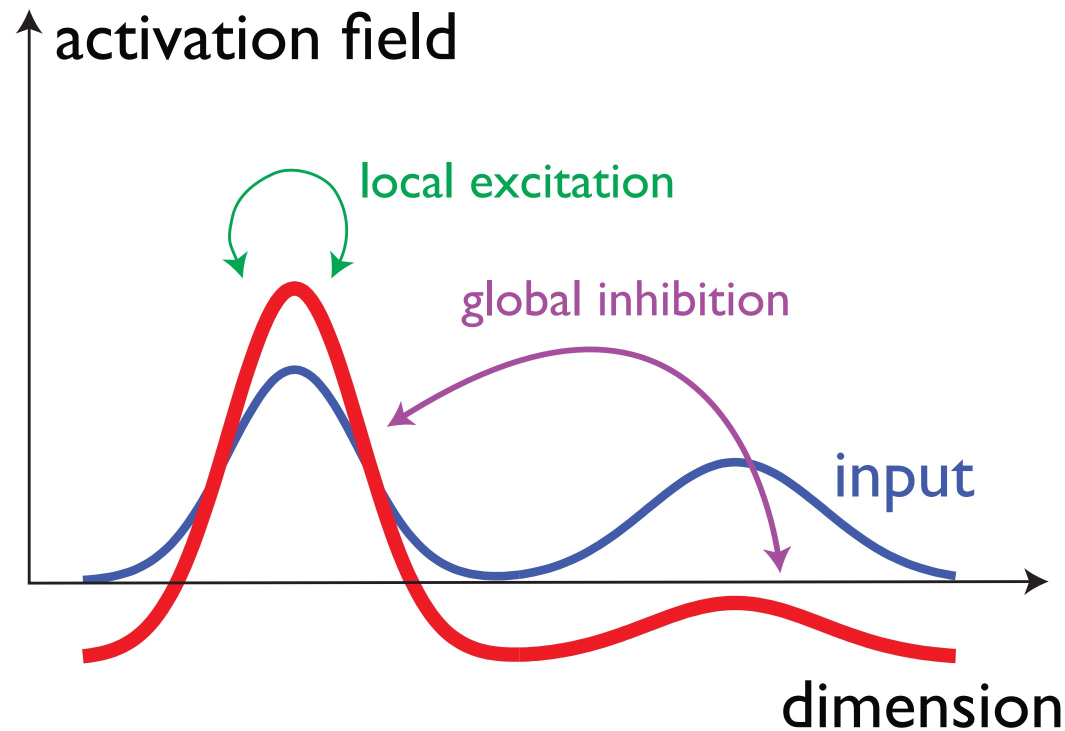


Roadmap

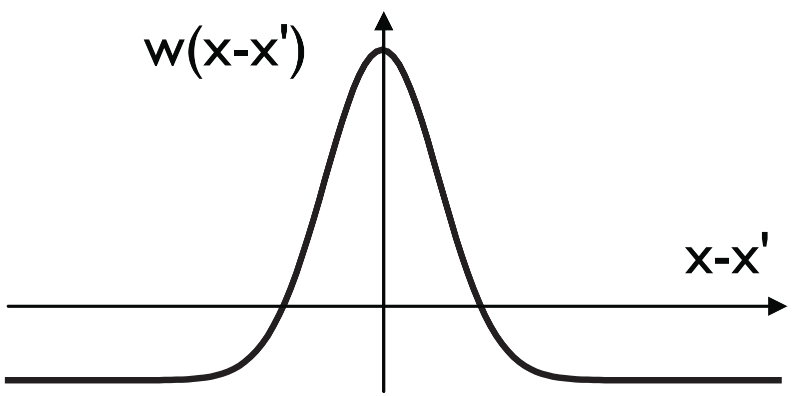
- Neuro-physics
- Neural dynamics
- Recurrent neural dynamics
- Neural fields: dynamics
- Neural fields: dimensions
- Binding
- Sequences
- Coordinate transforms
- Relational concepts, grounding, mental mapping
- Conclusions

Neural dynamics of fields

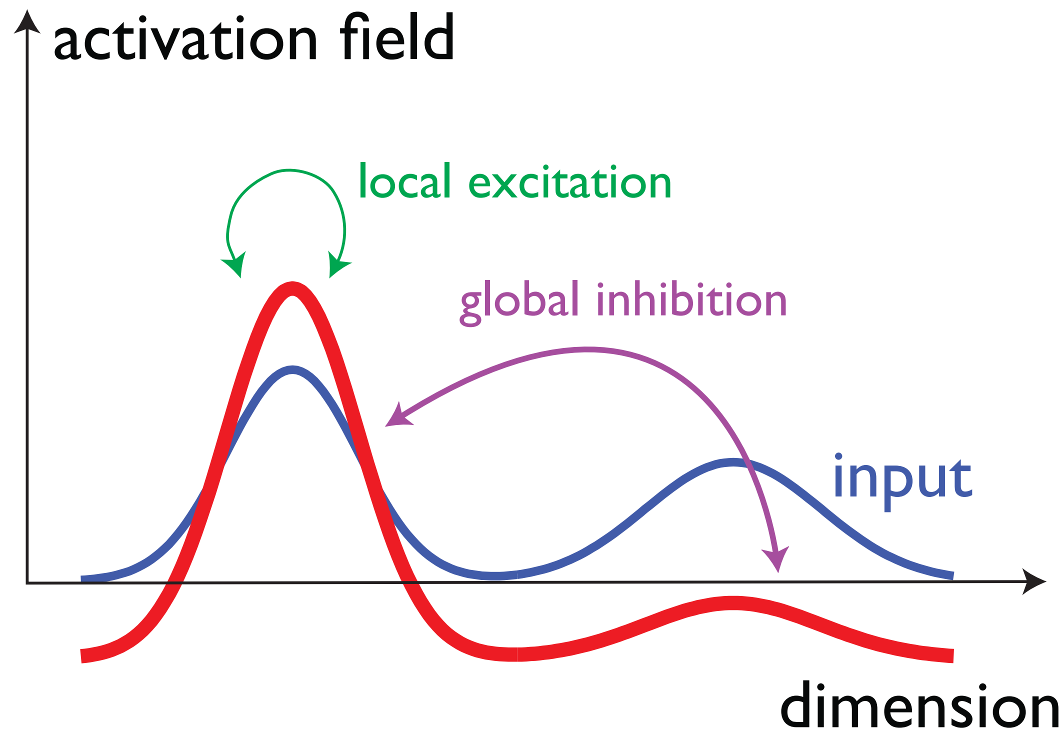
- continuously many activation variables
  - spanned by a dimension
- combine detection with selection
- => local excitation/ global inhibition



Neural dynamics of fields

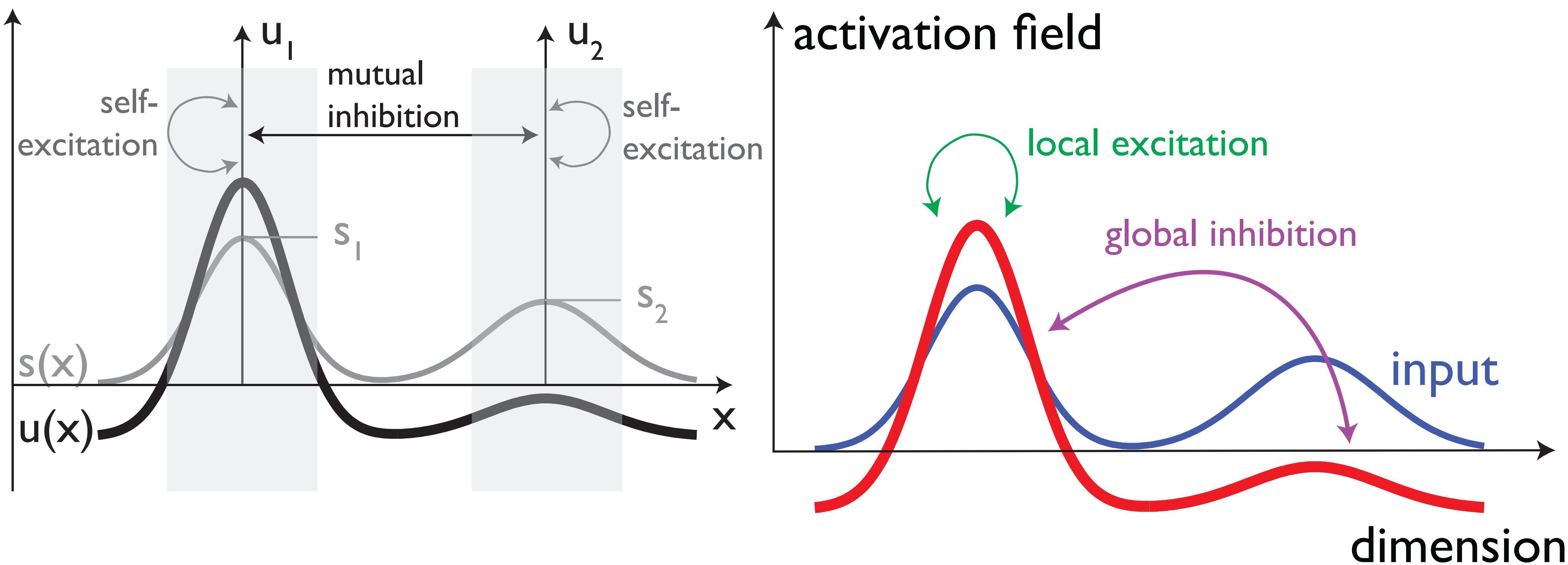


$$w(x - x') = w_{\text{exc}} e^{-\frac{(x - x')^2}{2\sigma^2}} - w_{\text{inh}}$$



$$\tau \dot{u}(x, t) = -u(x, t) + h + s(x, t) + \int dx' \; w(x - x') \; \sigma(u(x'))$$

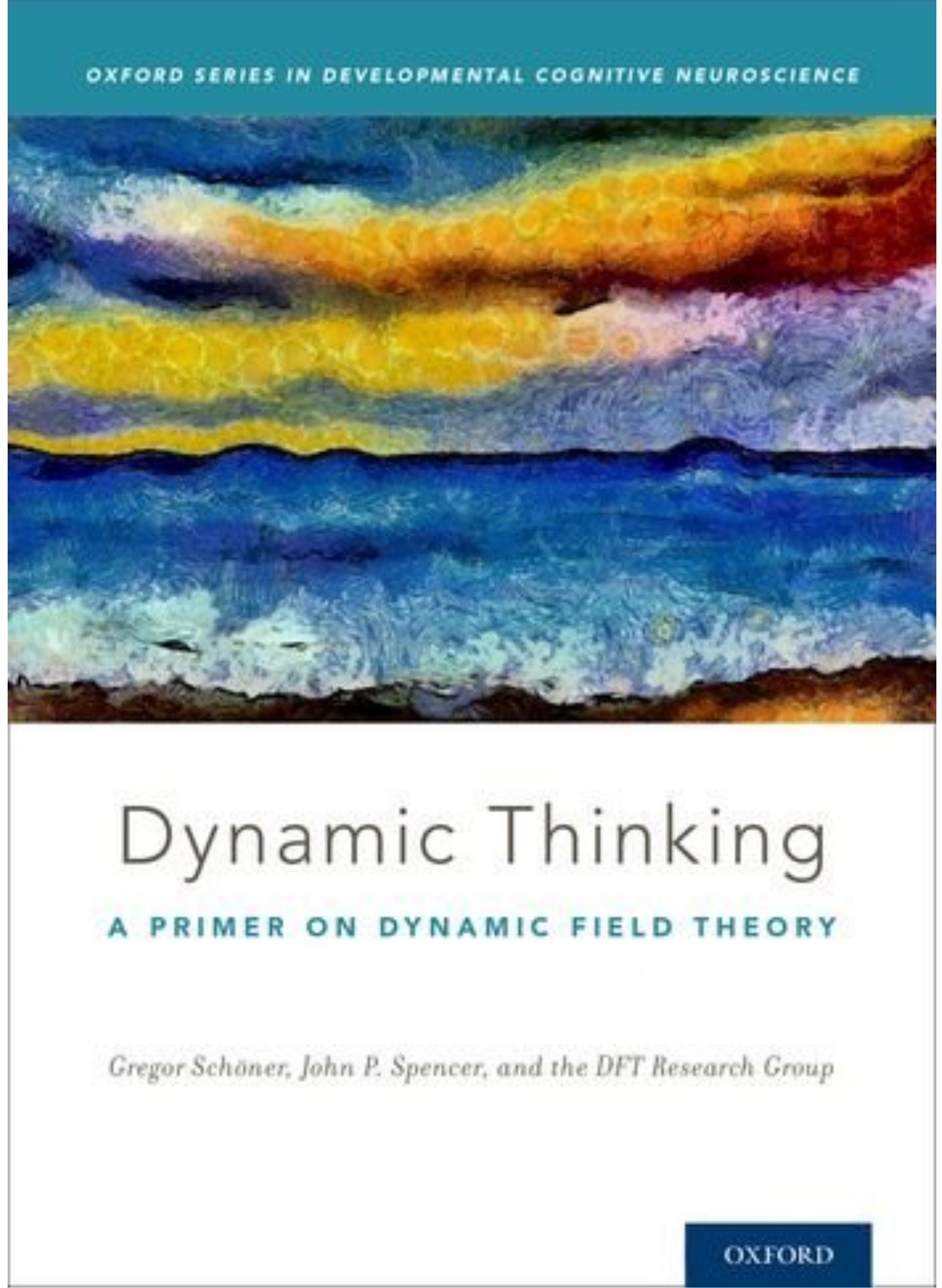
Neural dynamics of fields



$$\tau \dot{u}(x,t) = -u(x,t) + h + s(x,t) + \int dx'\ w(x-x')\ \sigma(u(x'))$$

=> simulation

- dynamicfieldtheory.org



Attractors and their instabilities

- input driven solution (sub-threshold)
- self-stabilized solution (peak, supra-threshold)
- selection / selection instability
- working memory / memory instability
- boost-driven detection instability

detection instability

reverse detection instability

Noise is critical near instabilities

Dynamic regimes

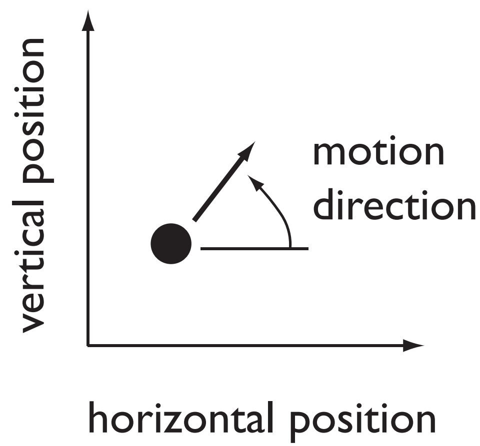
- which attractors and instabilities arise as input patterns are varied
- examples
  - "perceptual regime": mono-stable sub-threshold => bistable sub-threshold/peak => mono-table peak..
  - "working memory regime" bistable sub-threshold/peak => mono-table peak.. without mono-stable sub-threshold
  - single ("selective") vs. multi-peak regime

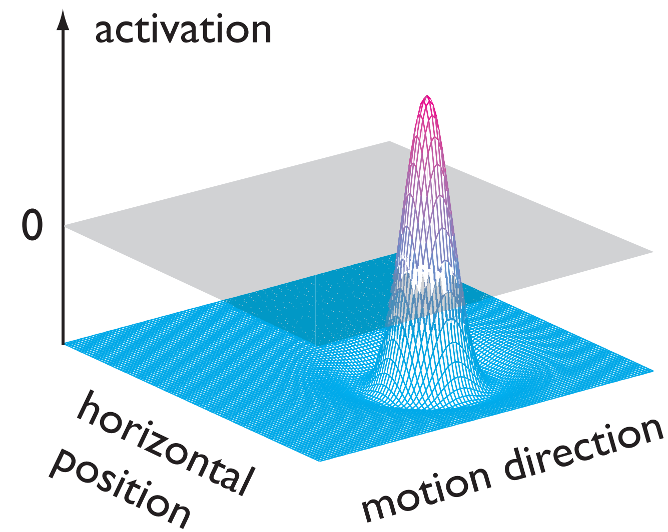
Roadmap

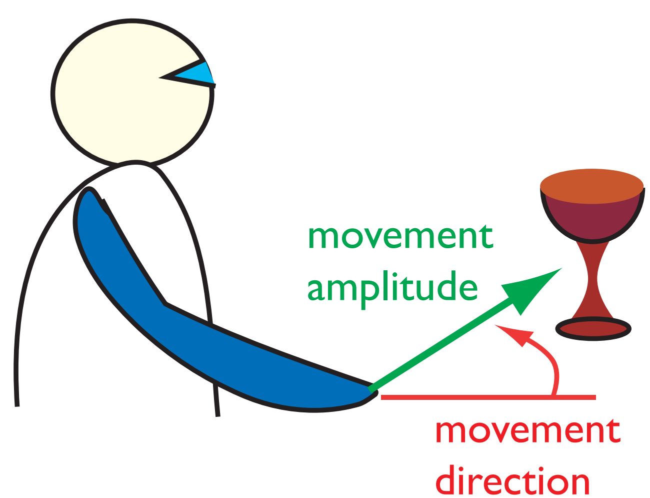
- Neuro-physics
- Neural dynamics
- Recurrent neural dynamics
- Neural fields: dynamics
- Neural fields: dimensions
- Binding
- Sequences
- Coordinate transforms
- Relational concepts, grounding, mental mapping
- Conclusions

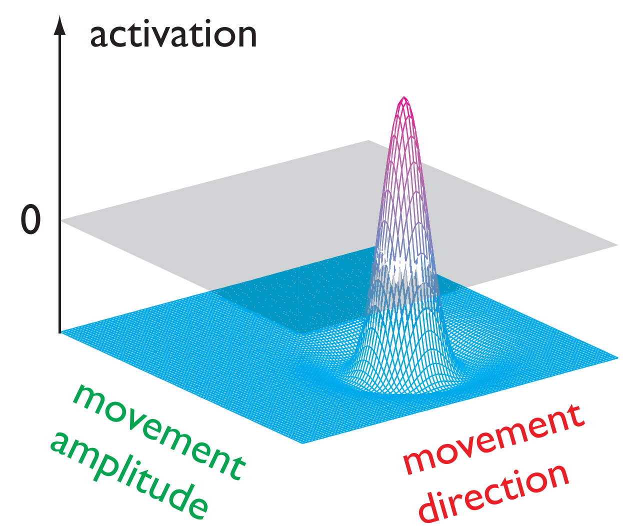
What do activation patterns mean?

- how do neural fields come to “represent” feature spaces?

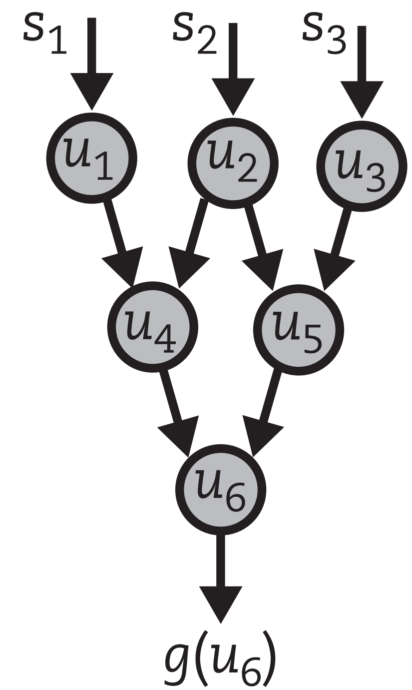




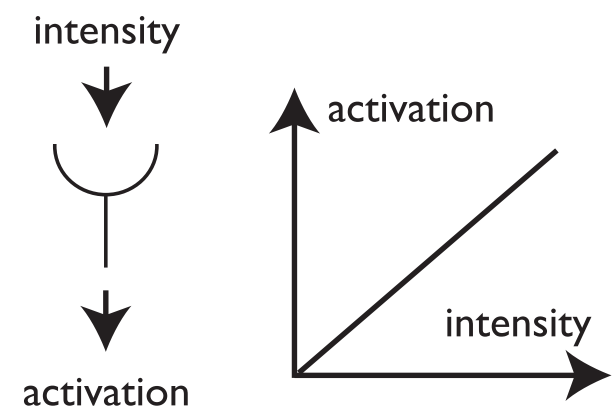


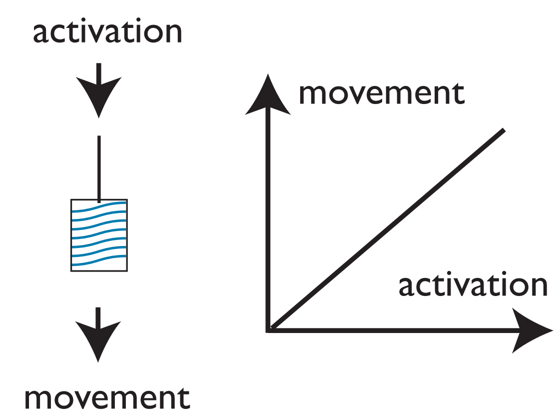


Neural networks



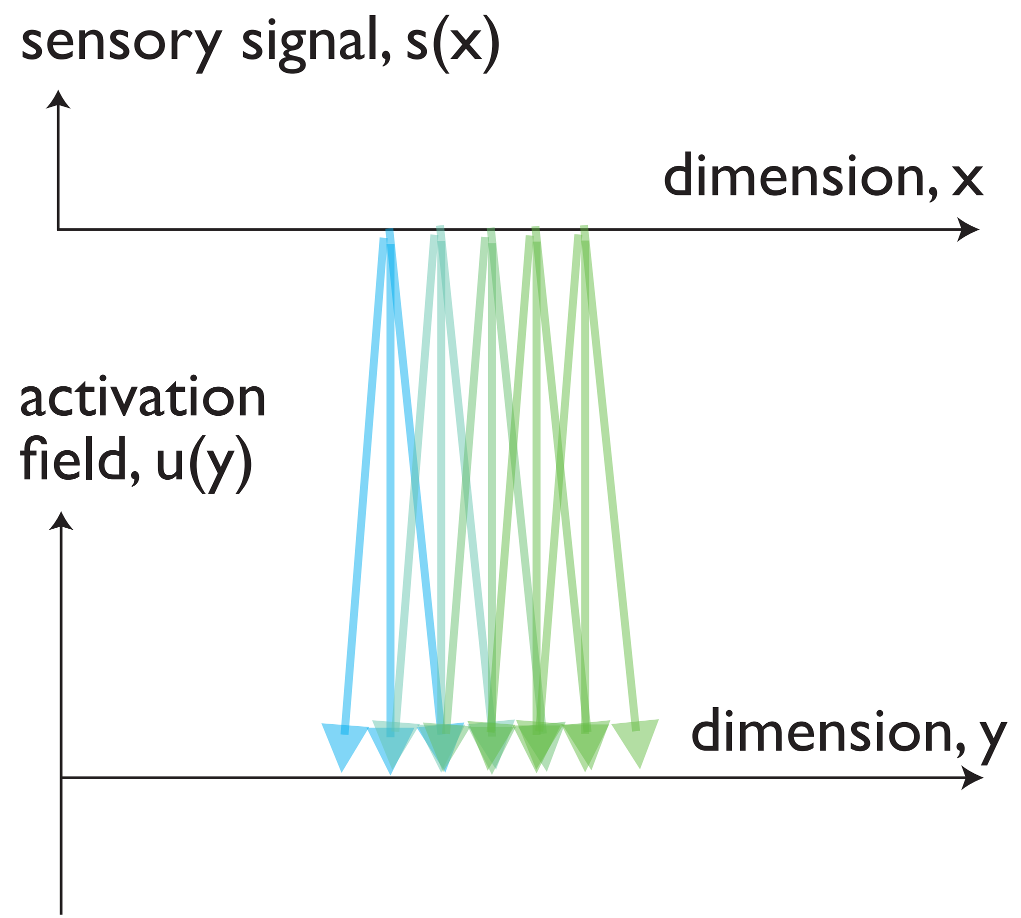
- forward connectivity determines "what a neuron stands for"= space code (or labelled line code)
- while the activation level may "stand for" intensities =rate code
- generic neural networks combine both codes





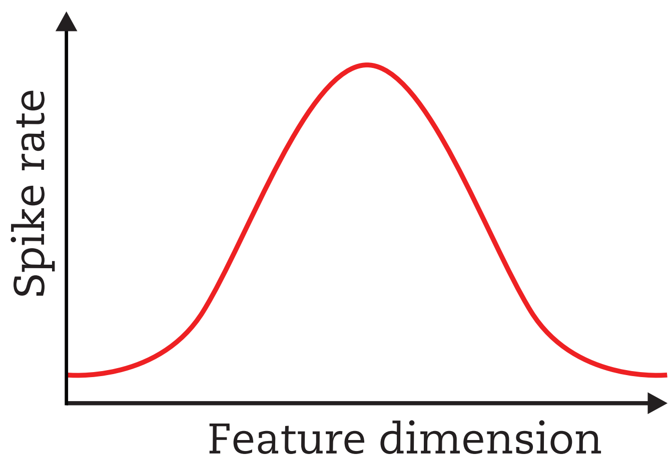
Neural fields

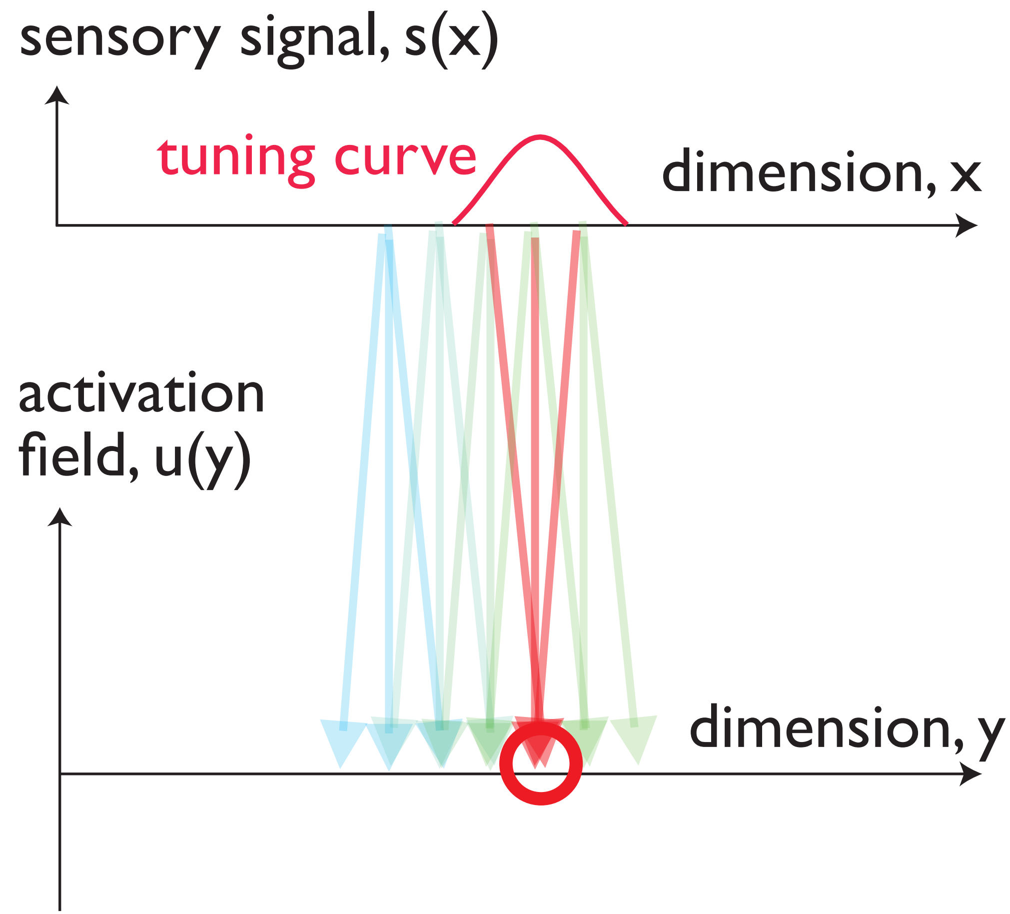
- forward connectivity from the sensory surface extracts perceptual feature dimensions



Neural fields

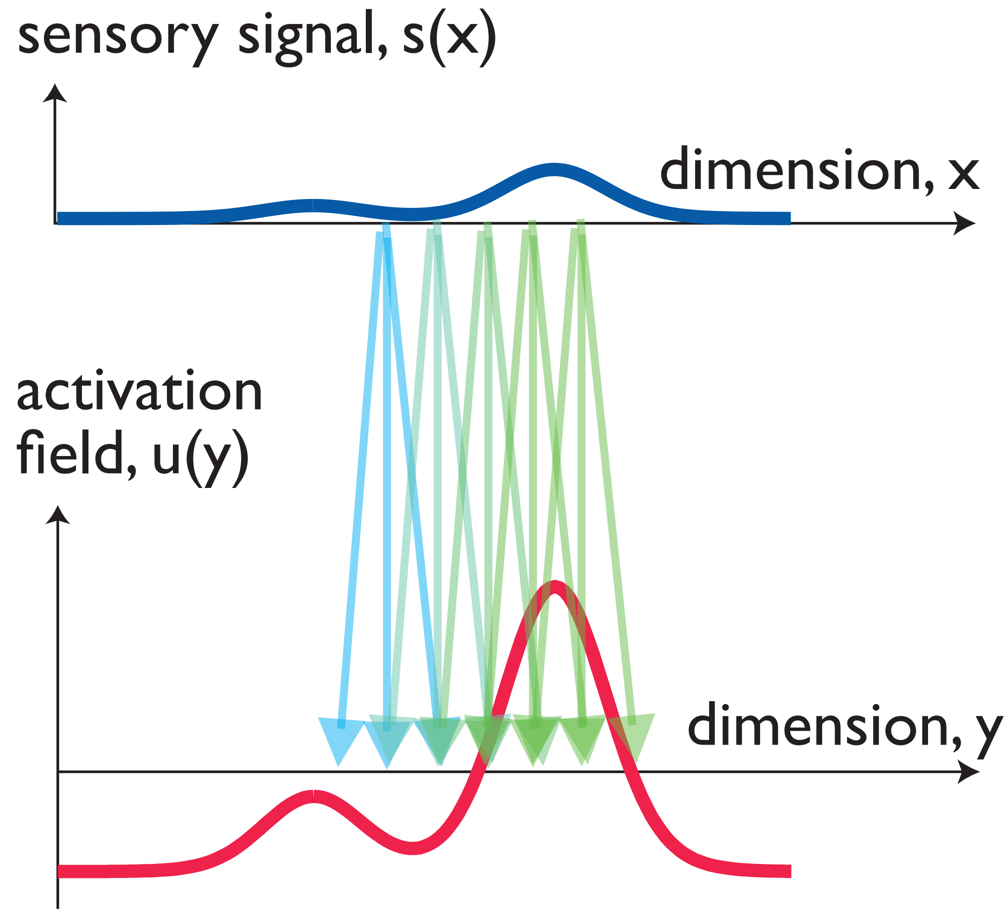
- as described by tuning curves or receptive fields





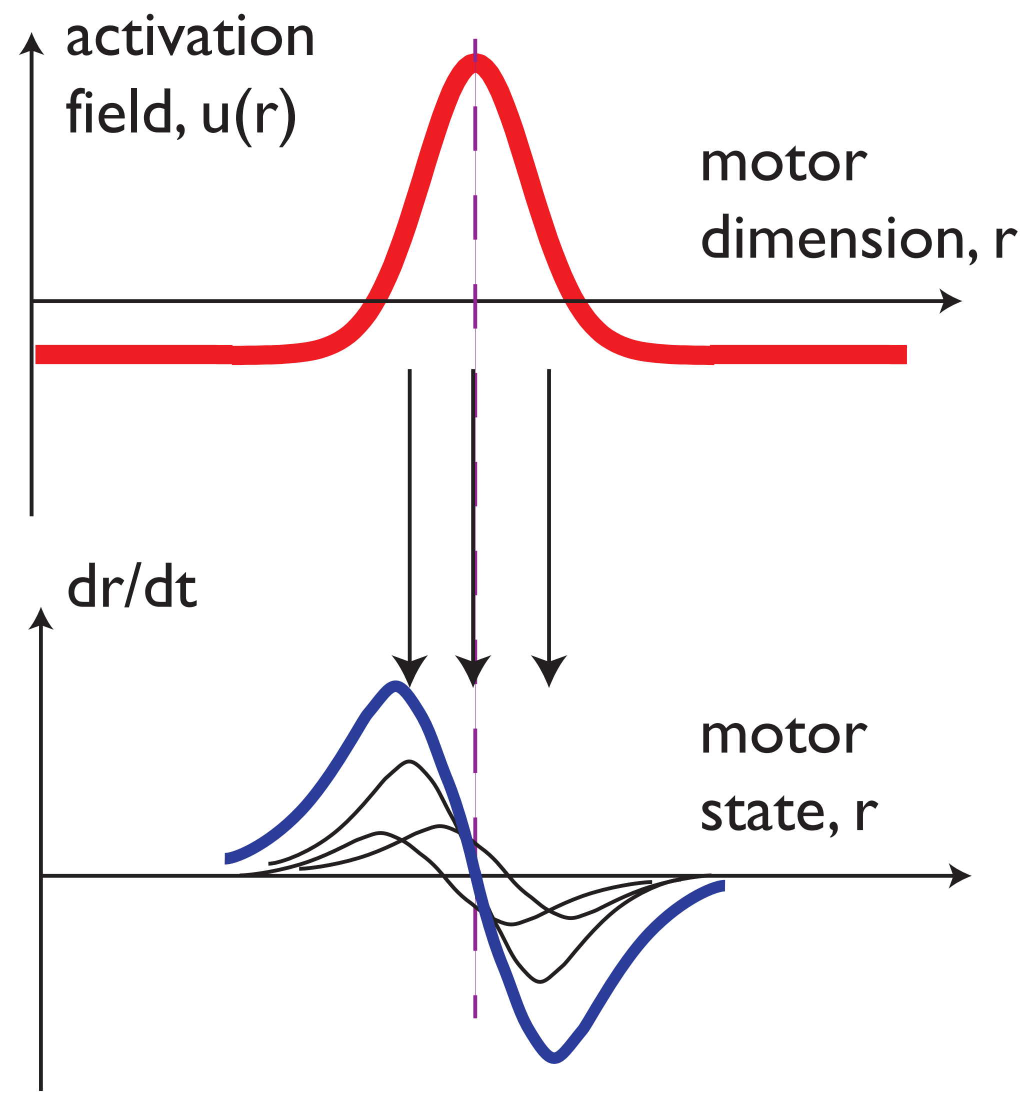
Neural fields

- => neural map from sensory surface to feature dimension
- neglect the sampling by individual neurons => activation field

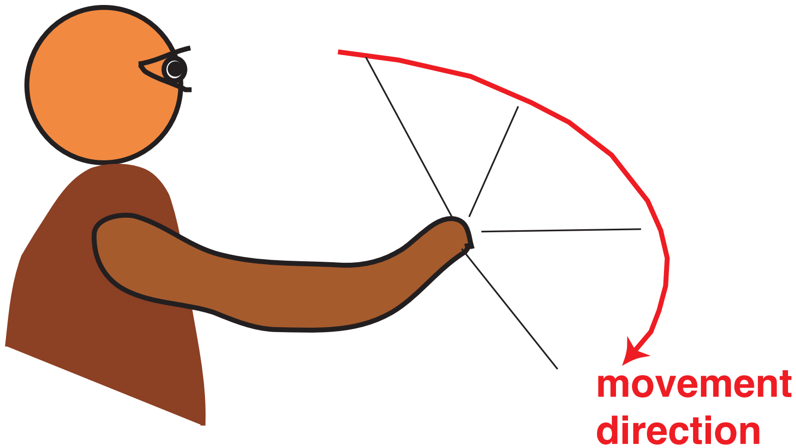


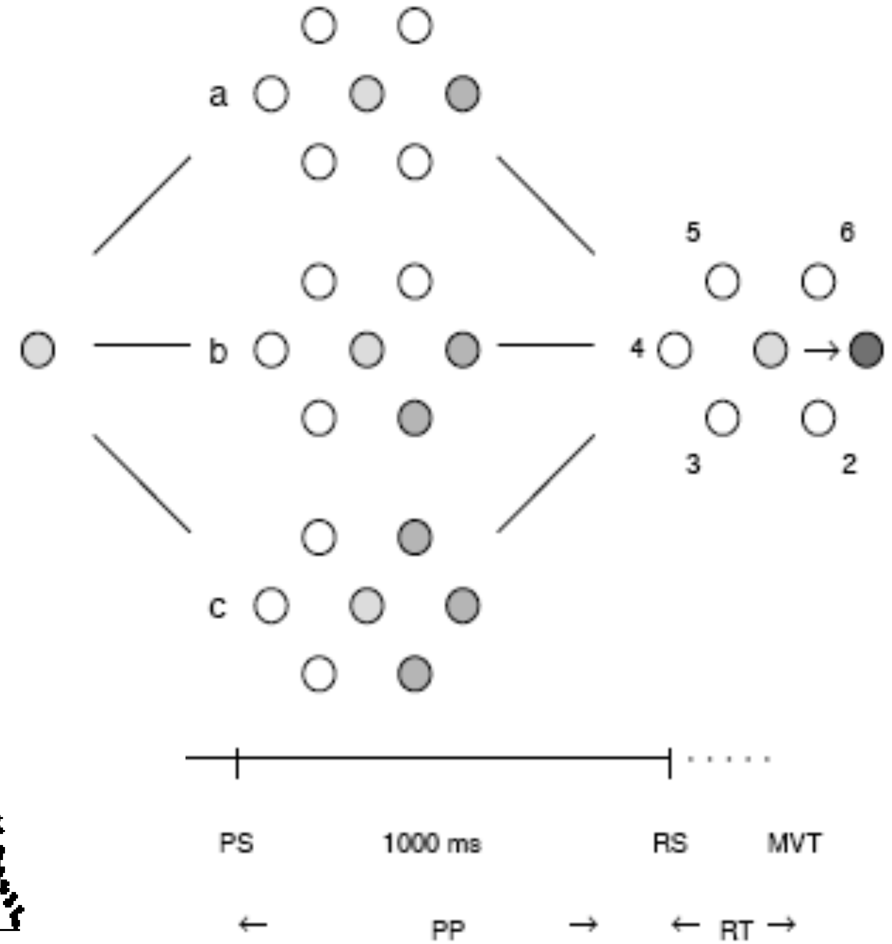
Neural fields

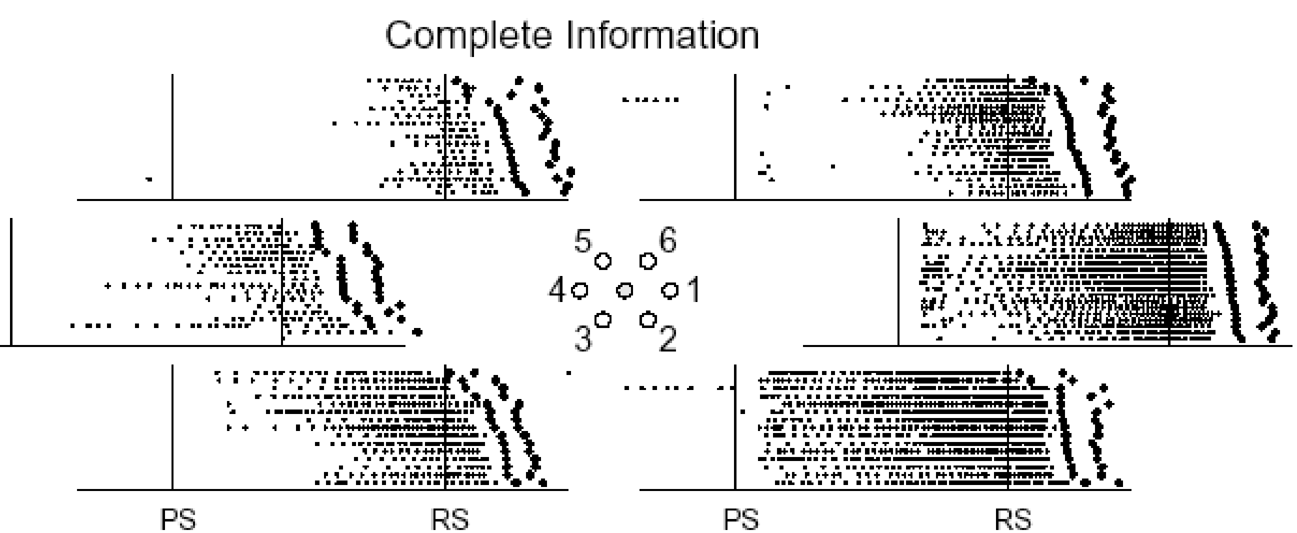
- analogous for projection onto to motor surfaces…
- which actually involves behavioral dynamics (e.g., through neural oscillators and peripheral reflex loops)



Neural estimation of fields



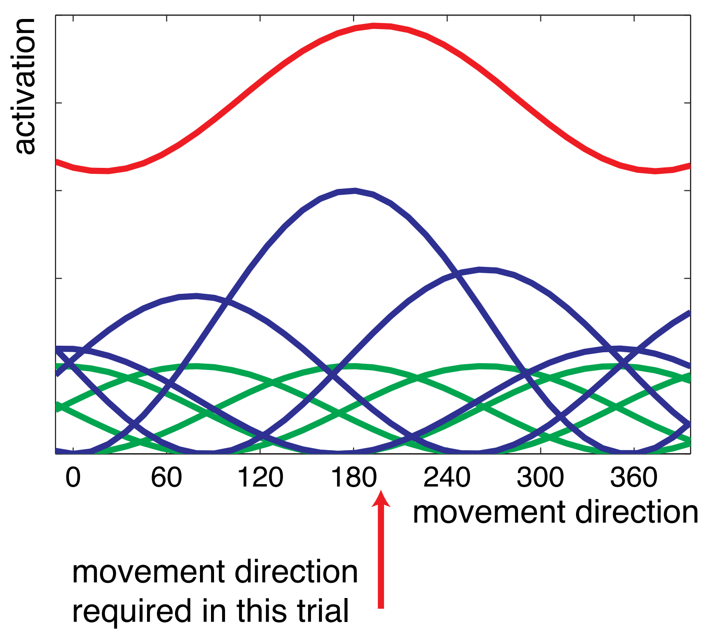


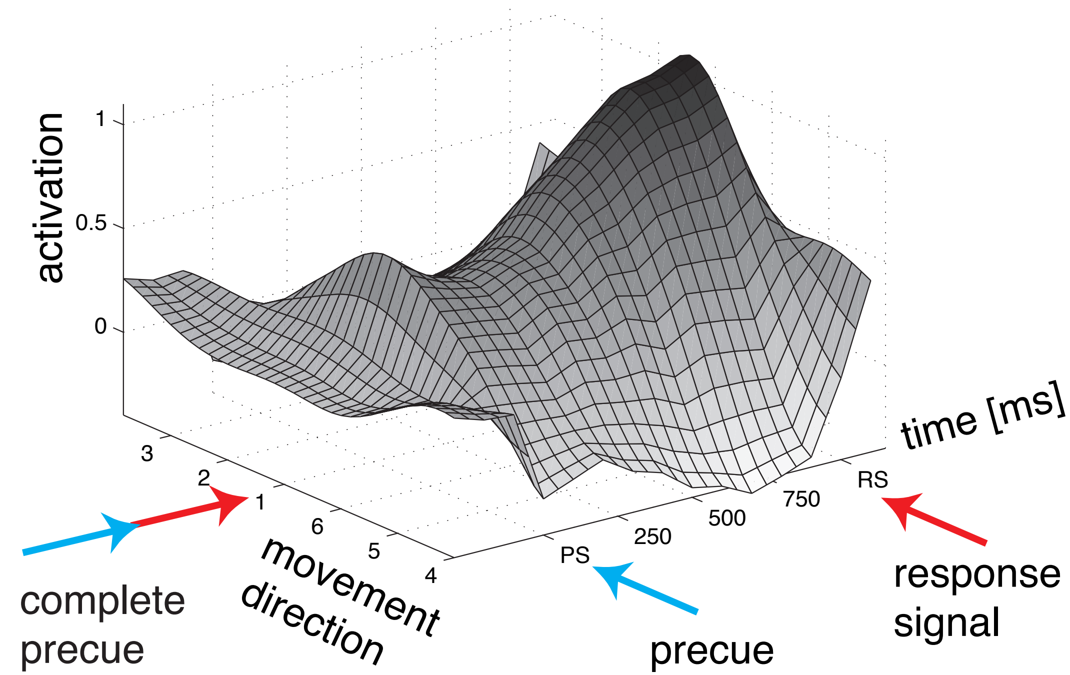


Bastian, Riehle, Schöner, 2003

Distribution of Population Activation (DPA) <=> neural field

Distribution of population activation = $\sum_{\text{neurons}}$ tuning curve * current firing rate

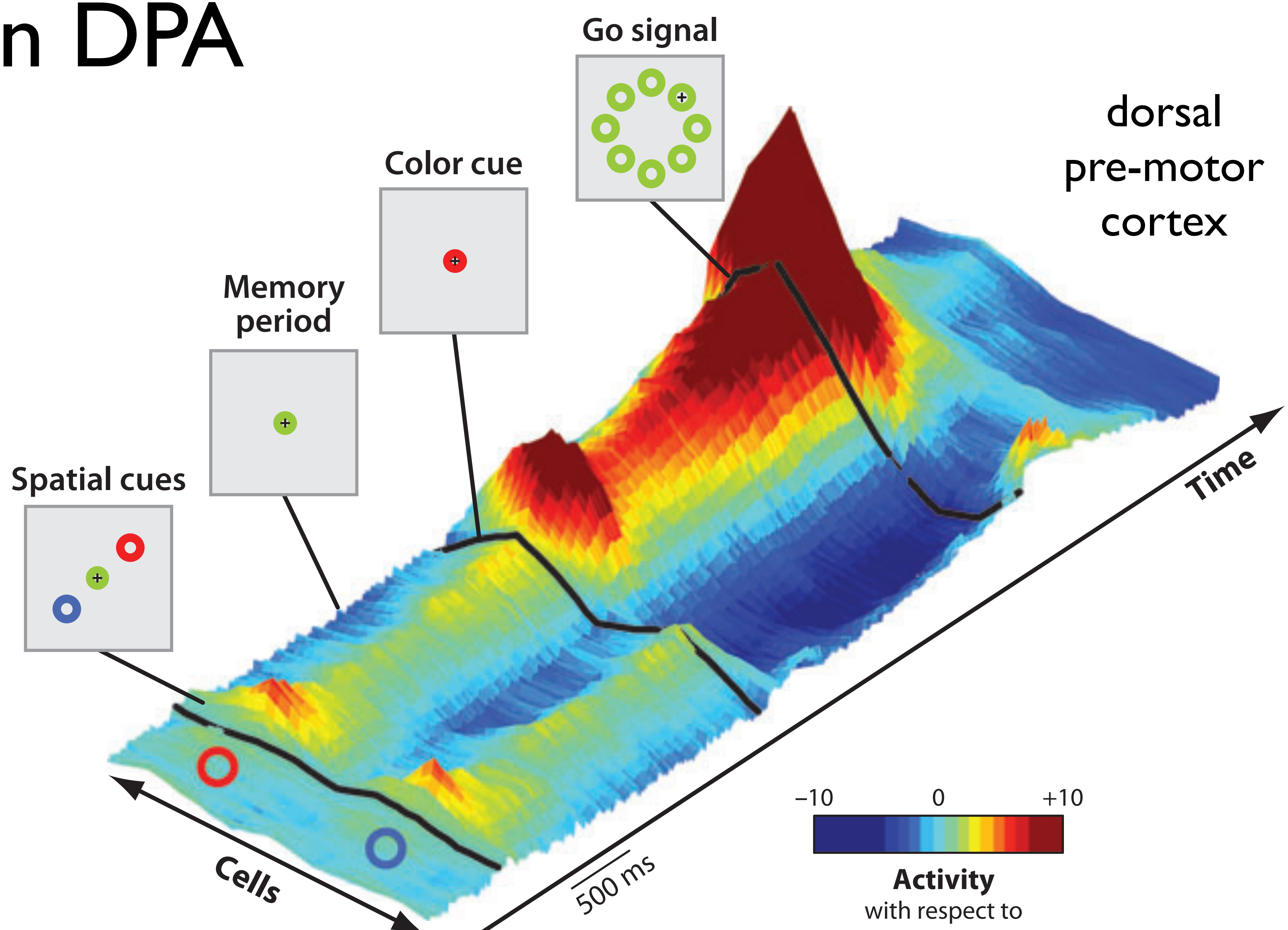




- note: neurons are not localized within DPA!

[Bastian, Riehle, Schöner, 2003]

Decision making in DPA



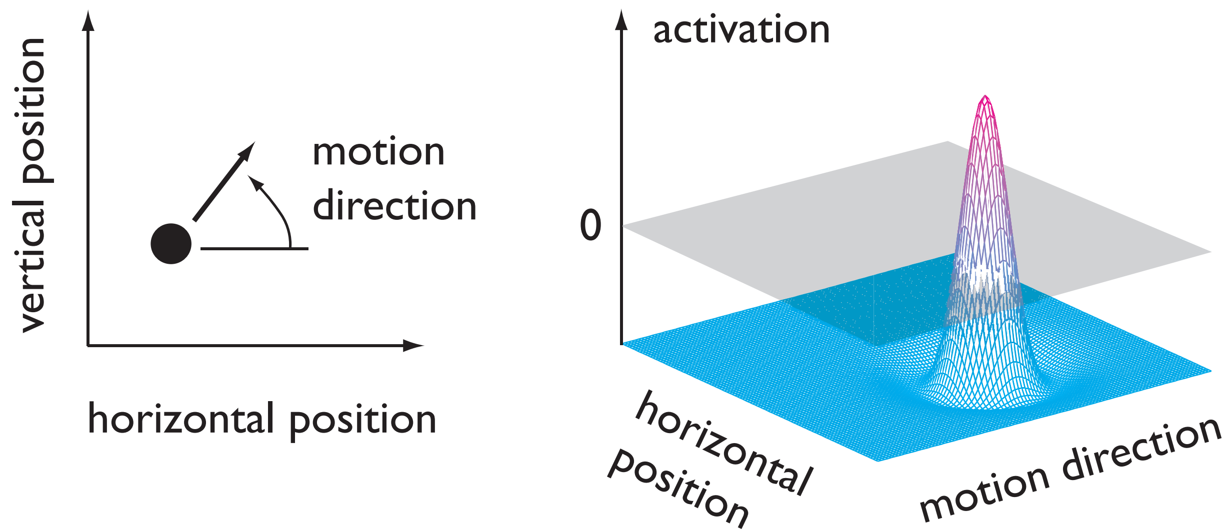
dorsal pre-motor cortex

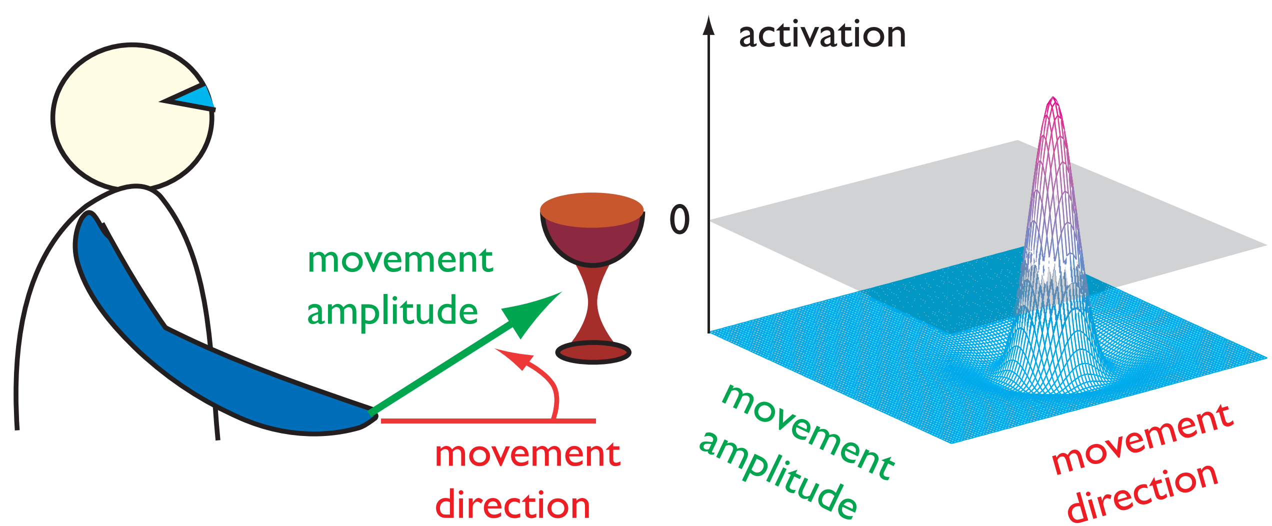
[Cisek, Kalaska 2005]

Roadmap

- Neuro-physics
- Neural dynamics
- Recurrent neural dynamics
- Neural fields: dynamics
- Neural fields: dimensions
- Binding
- Sequences
- Coordinate transforms
- Relational concepts, grounding, mental mapping
- Conclusions

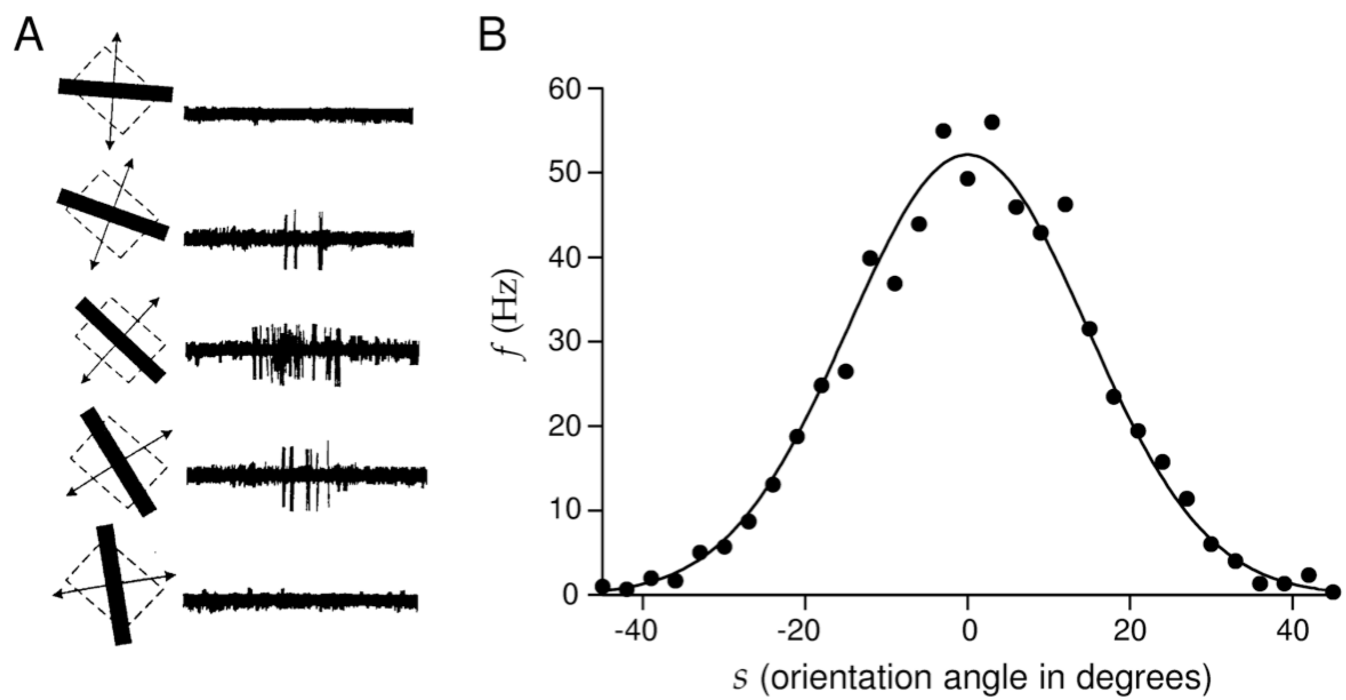
What does it mean when different dimensions are combined?





Combining different feature dimensions

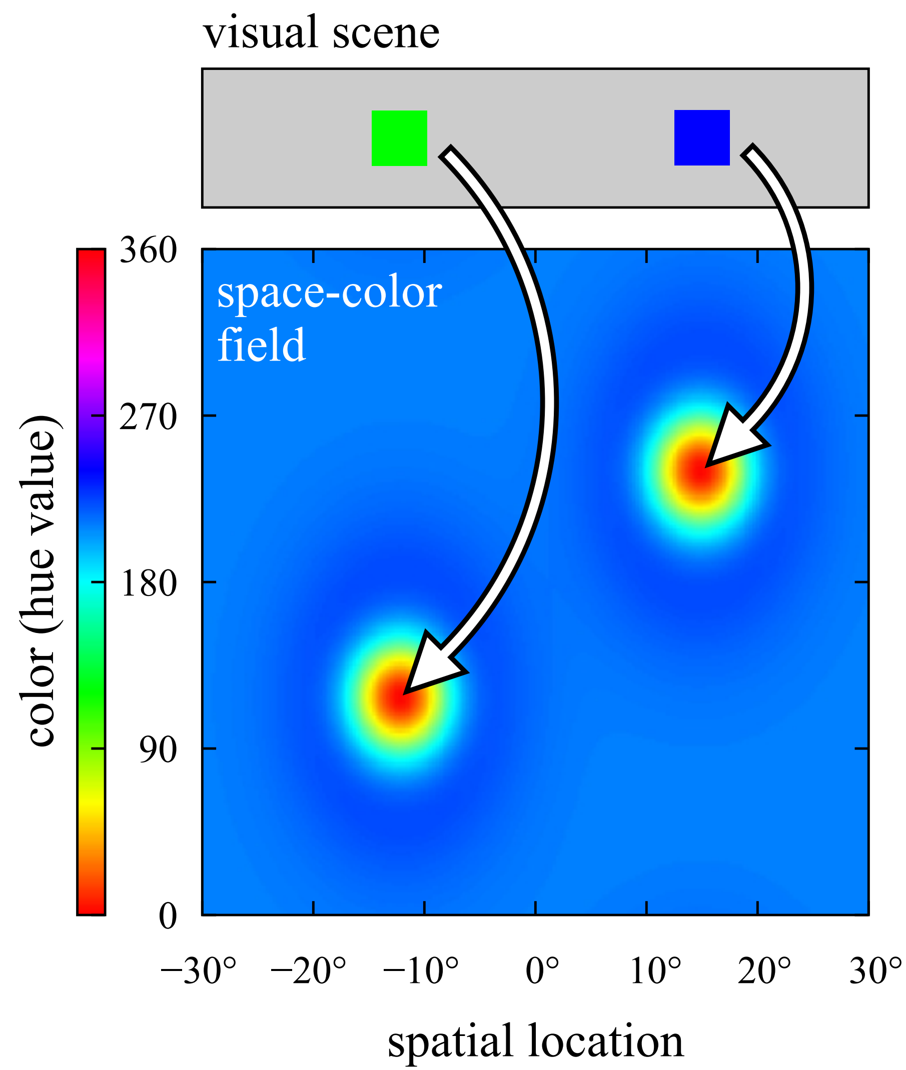
- neurons tuned to multiple dimensions
  - e.g. receptive field + direction tuning
  - => combines visual space and orientation
- "anatomical" binding



[Hubel, Wiesel, 1962]

Combining different feature dimensions

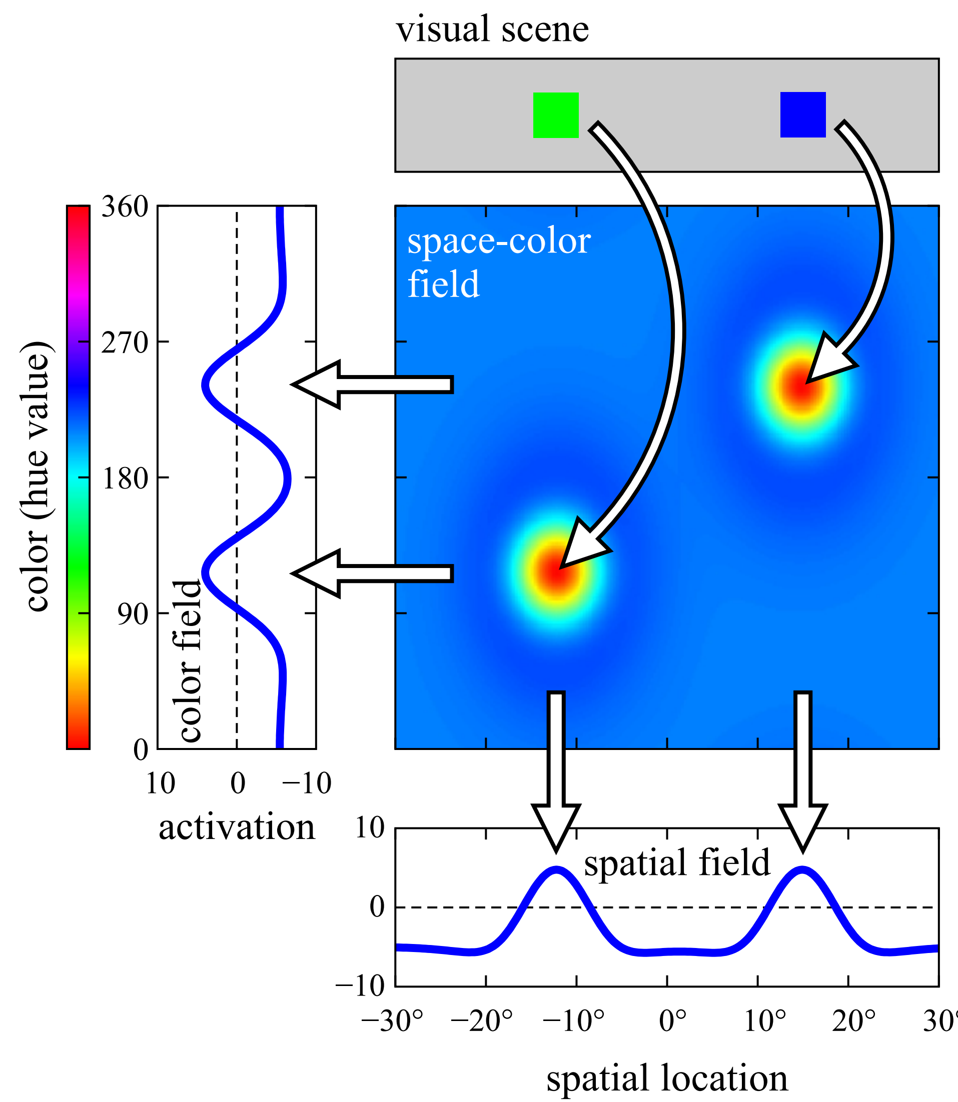
- example: a joint representation of color and visual space "binds" these two dimensions



[Schneegans et al., Ch 5 of *DFT Primer*, 2016]

Extract the bound features

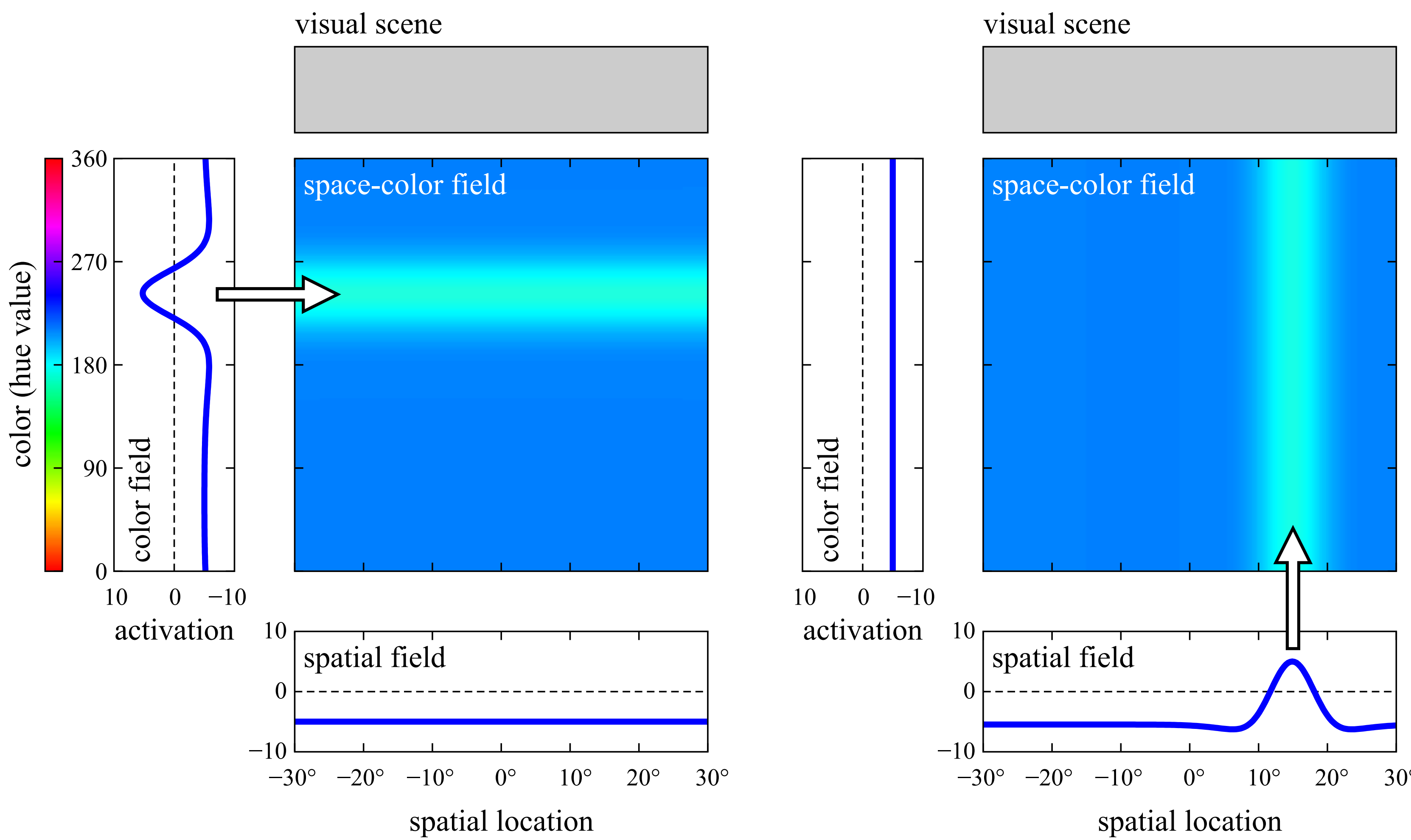
- project to lower-dimensional fields
- by summing along the marginalized dimensions
- (or by taking the soft-max)



[Schneegans et al., Ch 5 of *DFT Primer*, 2016]

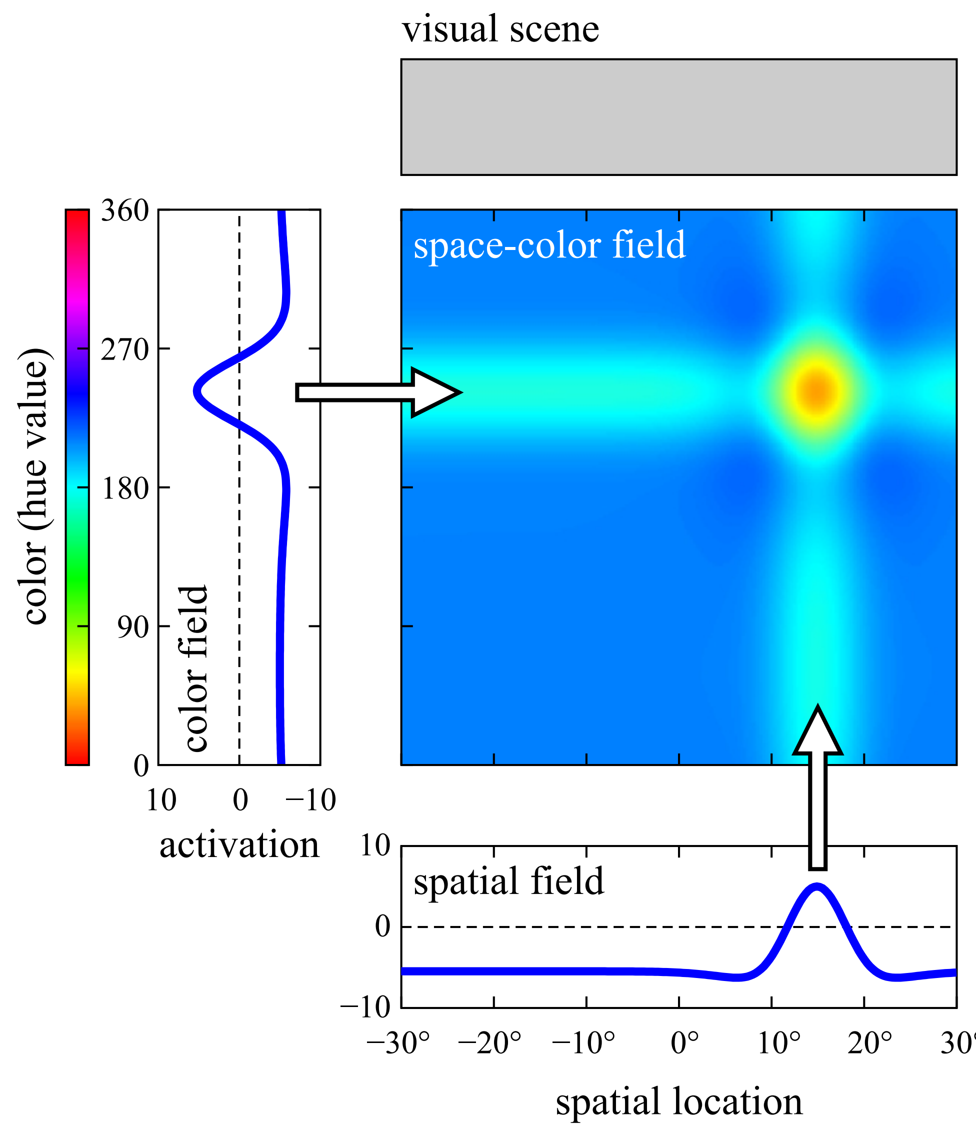
Assemble bound representations

- project lower-dimension field onto higher-dimensional field as “ridge input”



[Schneegans et al., Ch 5 of *DFT Primer*, 2016]

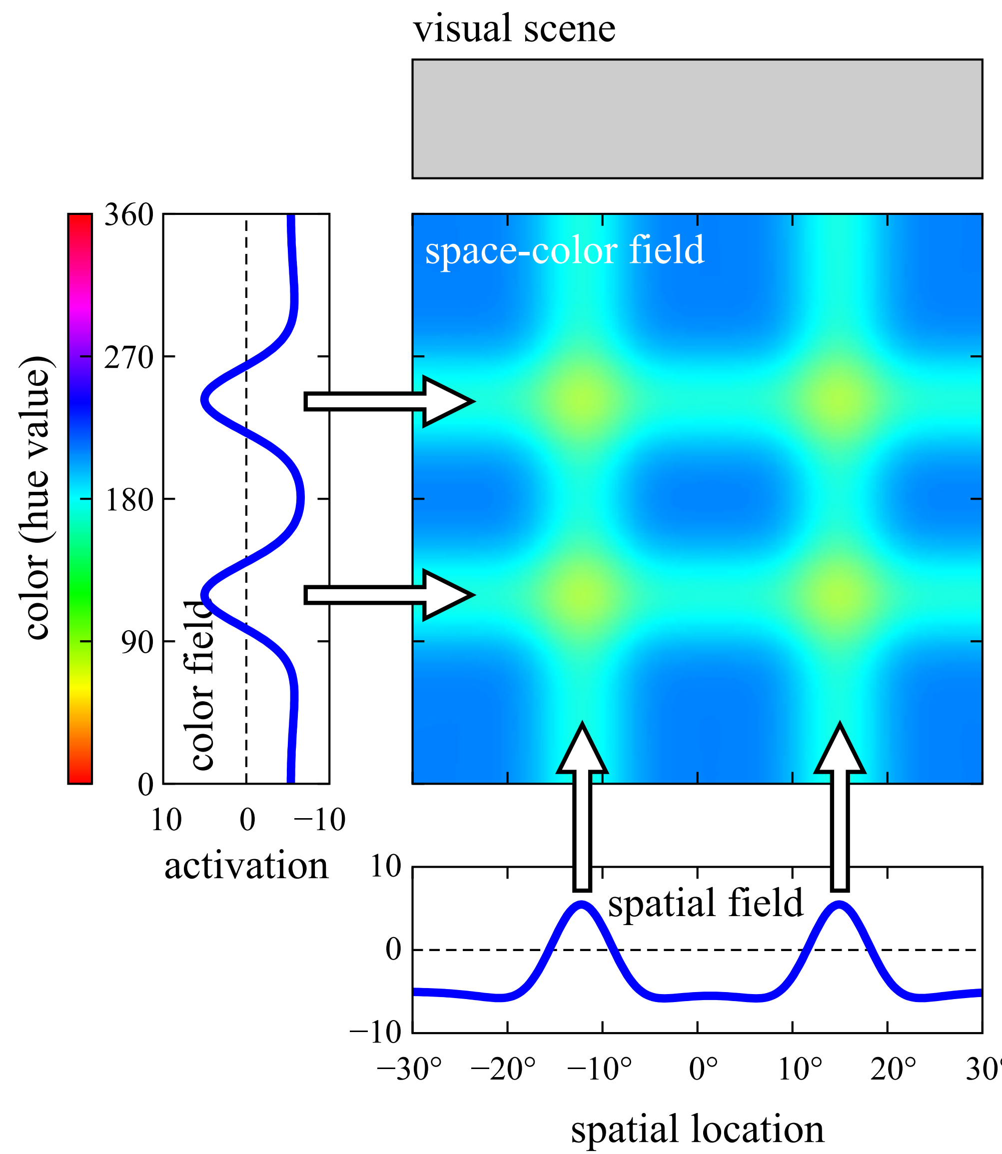
Assemble bound representations



[Schneegans et al.,Ch 5 of *DFT Primer*, 2016]

Assemble bound representations

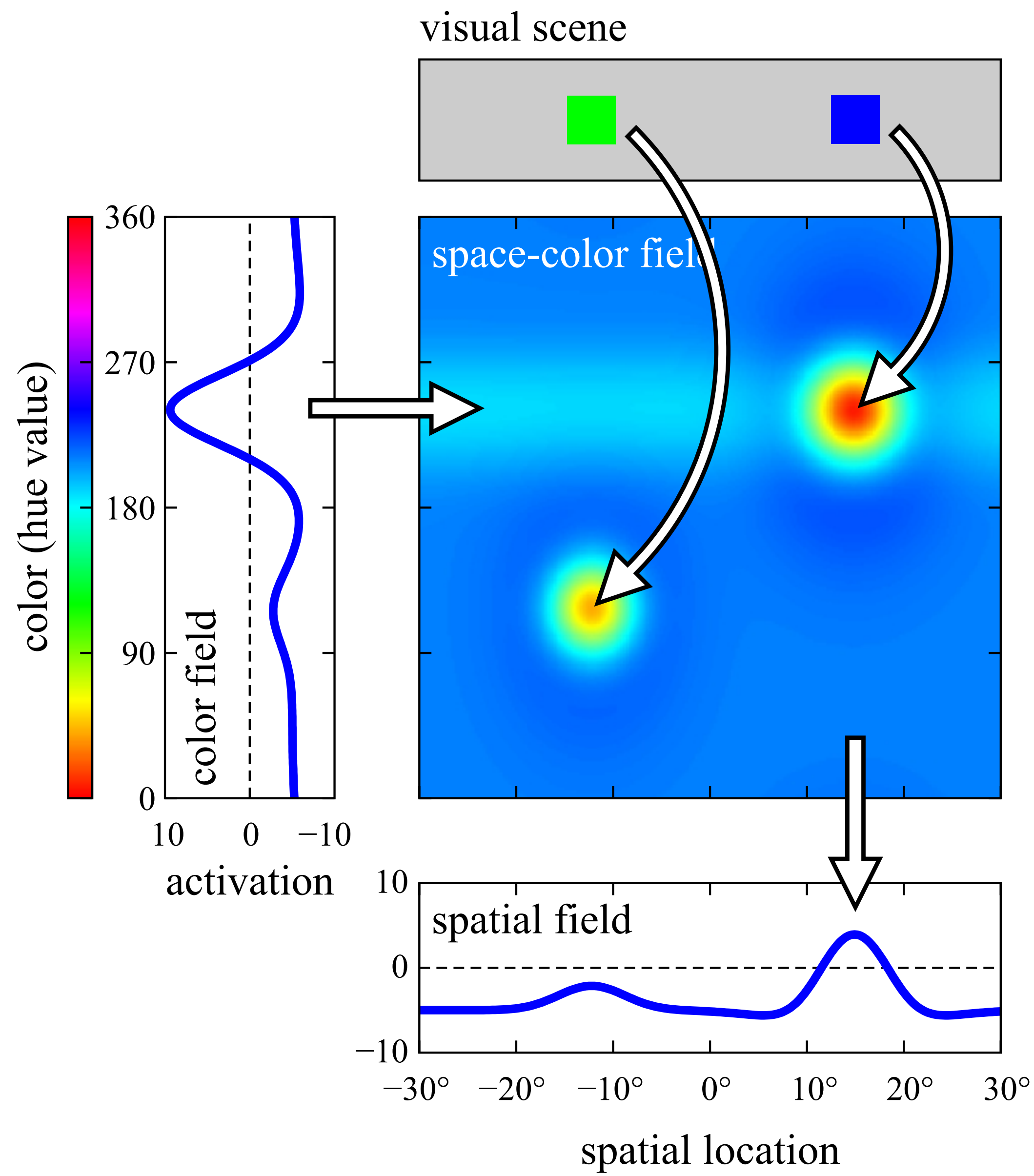
- binding problem: multiple ridges along lower-dimensional space lead to a correspondence problem
- => assemble one object at a time…
- => sequentiality bottle-neck!



[Schneegans et al.,Ch 5 of *DFT Primer*, 2016]

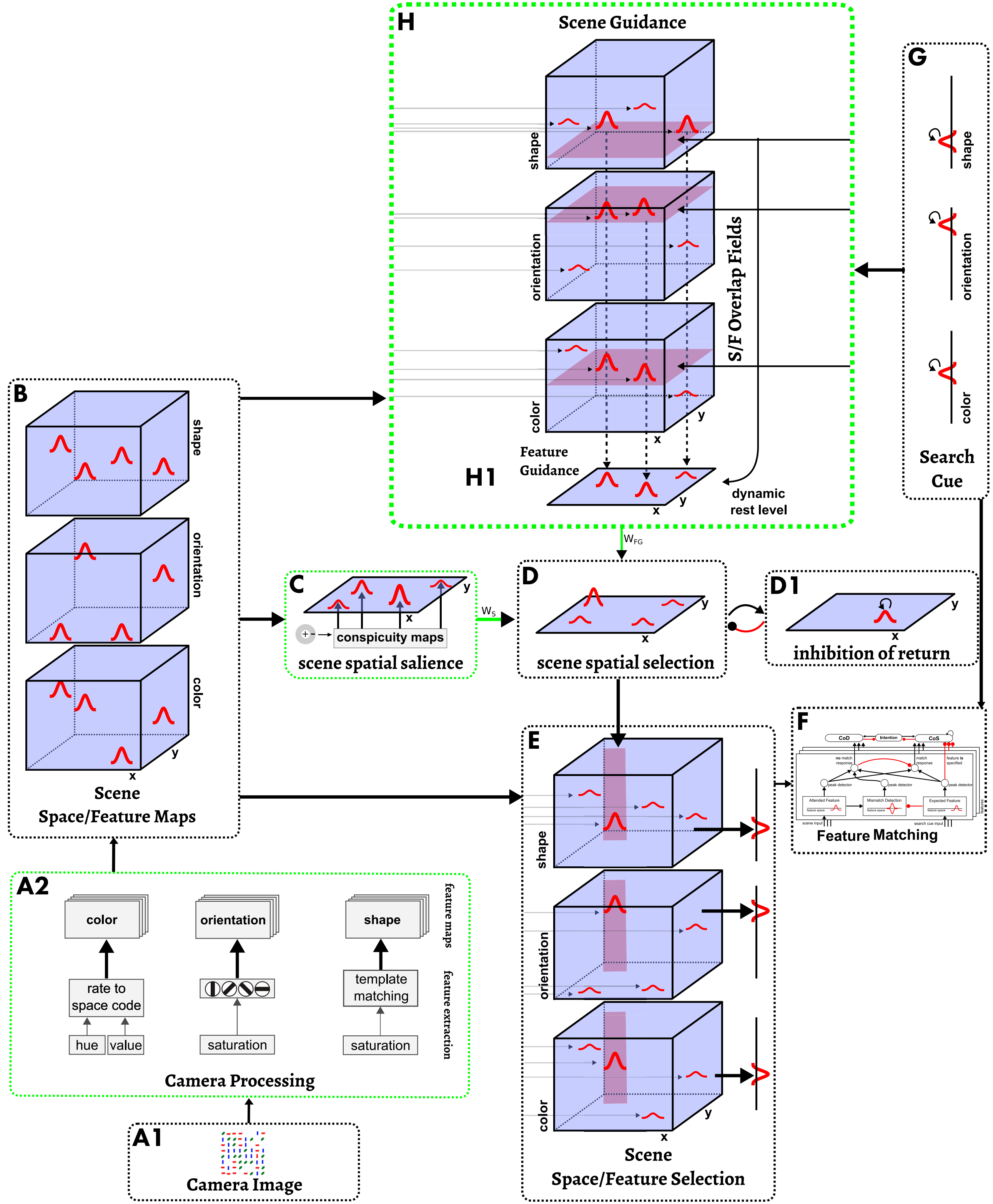
Search

- ridge input along one dimension extracts from bound representation matching objects
- other dimensions of those objects can then be extracted
- e.g. visual search

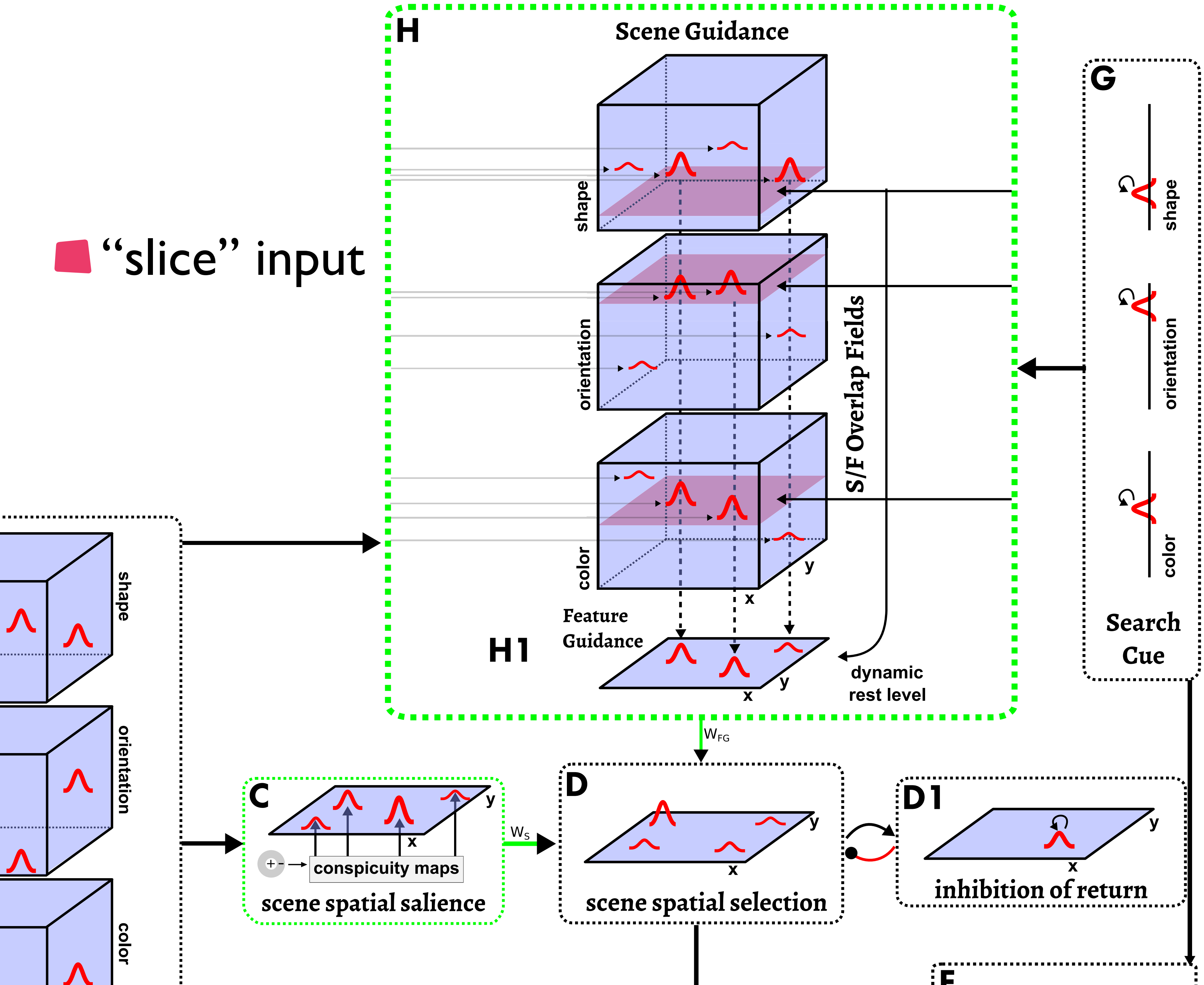


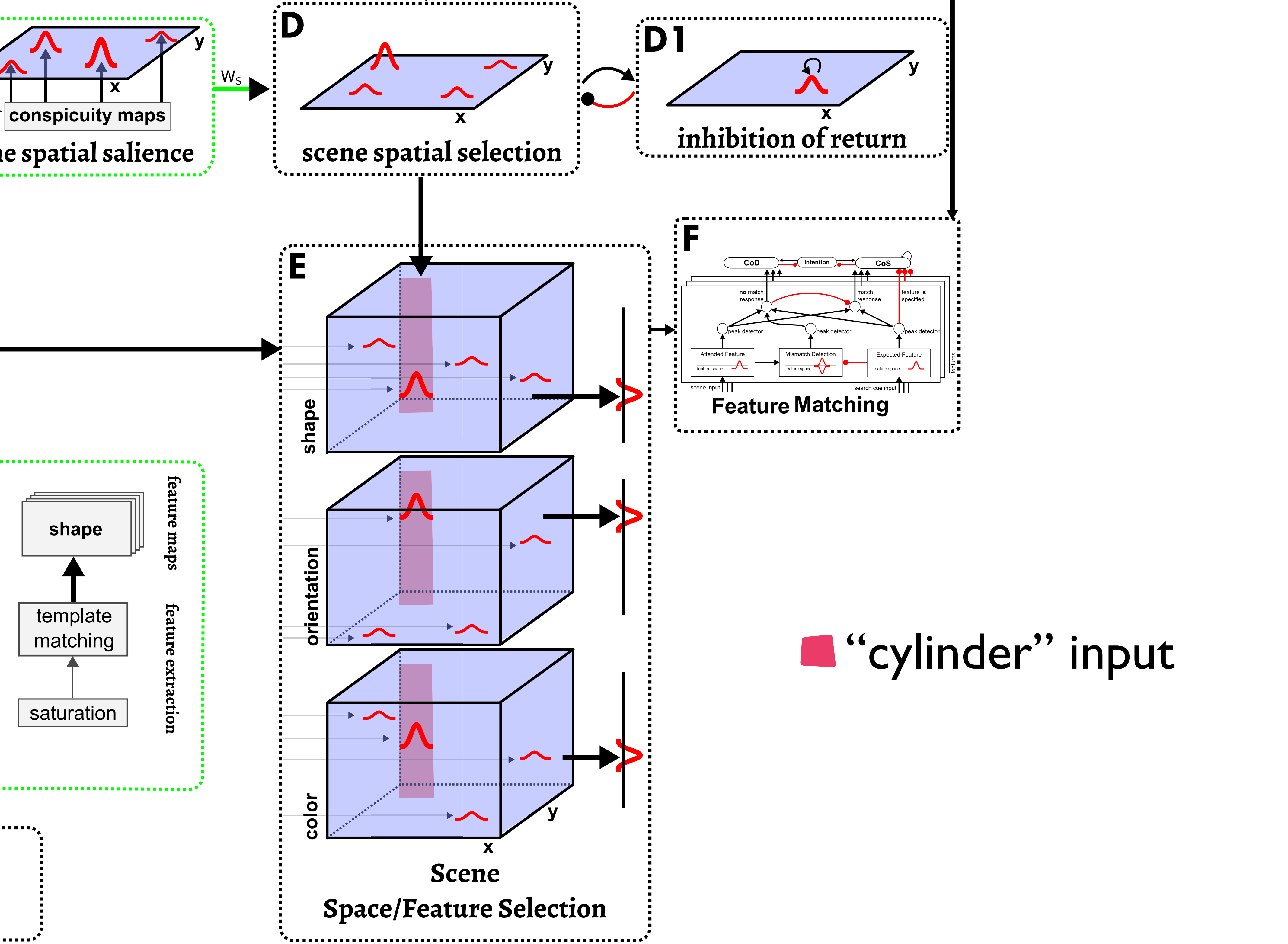
[Schneegans et al., Ch 5 of *DFT Primer*, 2016]

Visual search



[Grieben et al. *Attention, Perception & Psychophysics* 2020; *CogSci* 2021]

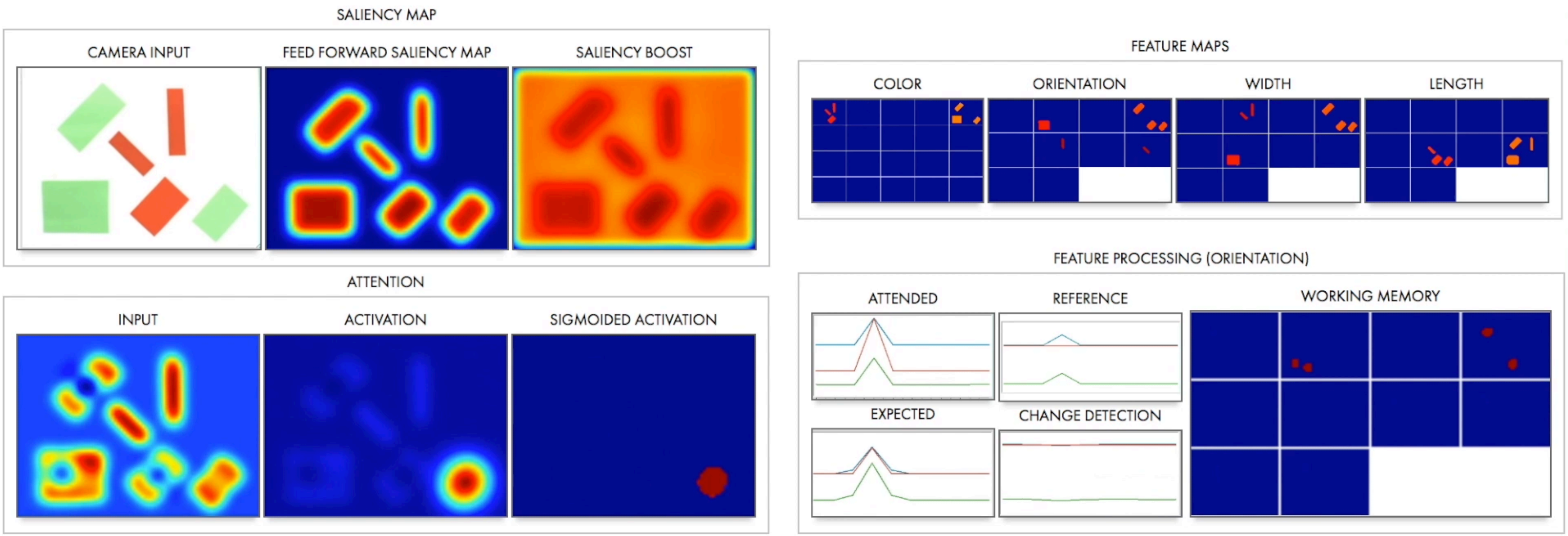

"slice" input
H
Scene Guidance
shape
orientation
color
S/F Overlap Fields
x
y
H1
Feature Guidance
dynamic rest level
G
shape
orientation
color
Search Cue
shape
orientation
color
C
conspicuity maps
scene spatial salience
W_S
W_{FG}
D
scene spatial selection
D1
inhibition of return
E



conspicuity maps
y
x
W_S
...e spatial salience
D
scene spatial selection
D1
inhibition of return
F
CoD
Intention
CoS
no match response
match response
feature is specified
peak detector
Attended Feature
Mismatch Detection
Expected Feature
feature space
scene input
search cue input
features
Feature Matching
E
shape
orientation
color
Scene
Space/Feature Selection
shape
template matching
saturation
feature maps
feature extraction
"cylinder" input

Visual search

- => hands on exericse



[Grieben et al. Attention, Perception & Psychophysics 2020]

Scaling feature dimensions

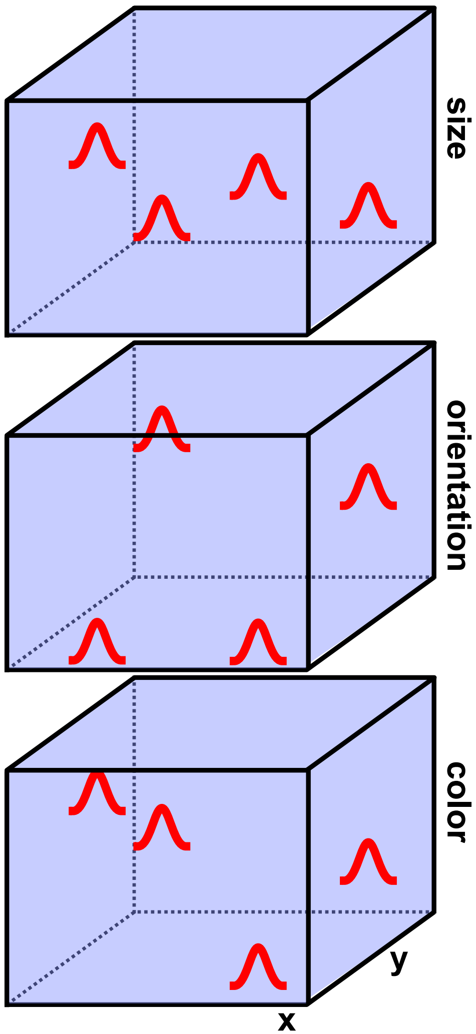
- 2 spatial dimensions
- depth
- orientation
- color
- texture
- movement direction
- size
- etc…

=>

- e.g. 8 dimensions
- 100 neurons per dimension
- $10^{2*8} = 10^{16}$!
- more than there are in the entire brain!
- => only small sets of feature dimensions can be bound “anatomically”

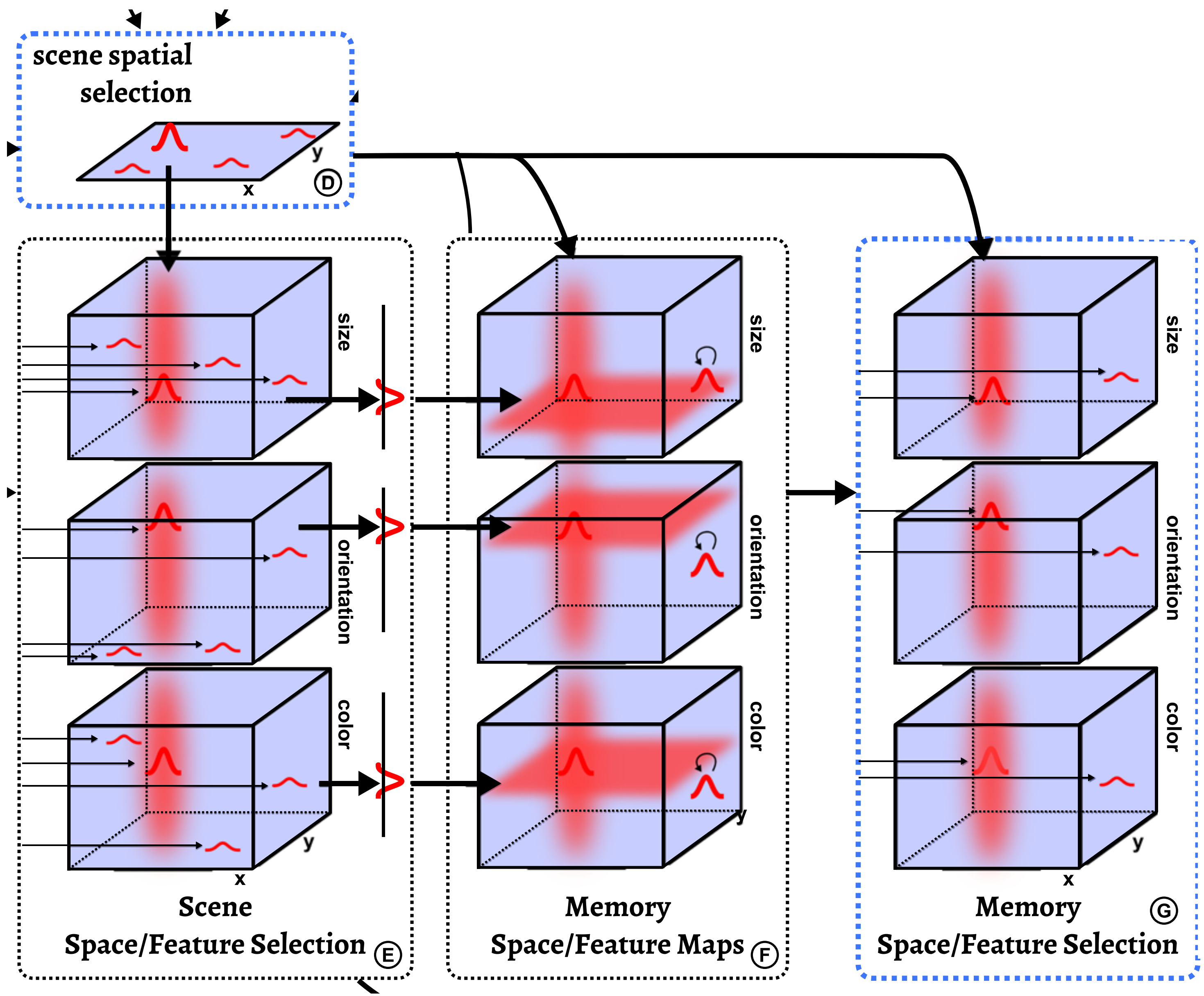
Binding through space

- many 3 to 4 dimensional feature fields
- all of which share the one dimension: visual space (~all neurons have receptive fields)
- bind through space à la Feature Integration Theory (Treisman)

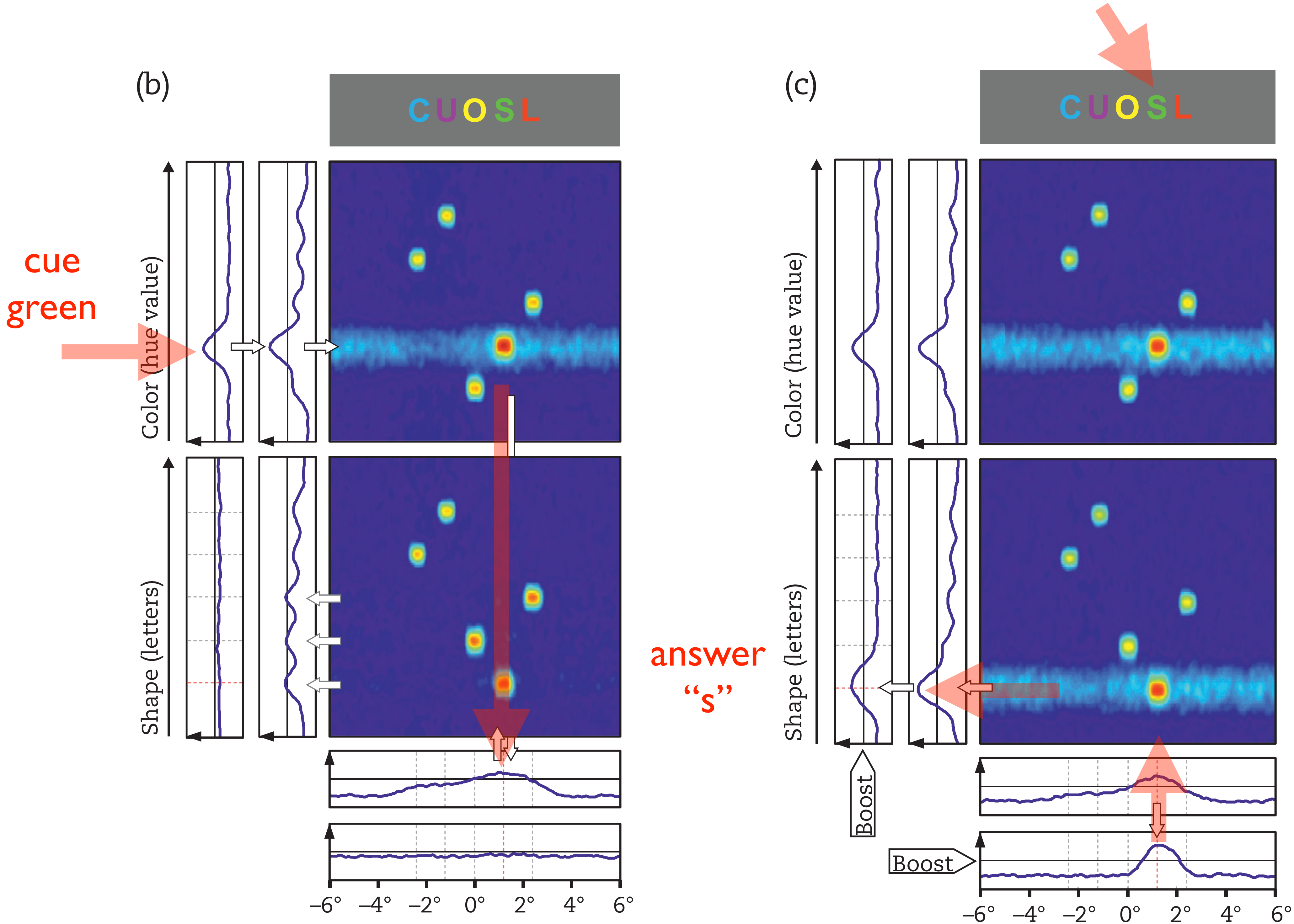


[Grieben et al. *Attention, Perception & Psychophysics* 2020]

Binding through space



[Grieben et al. *Attention, Perception & Psychophysics* 2020]



[Schneegans et al.,Ch 5 of *DFT Primer*, 2016]

Roadmap

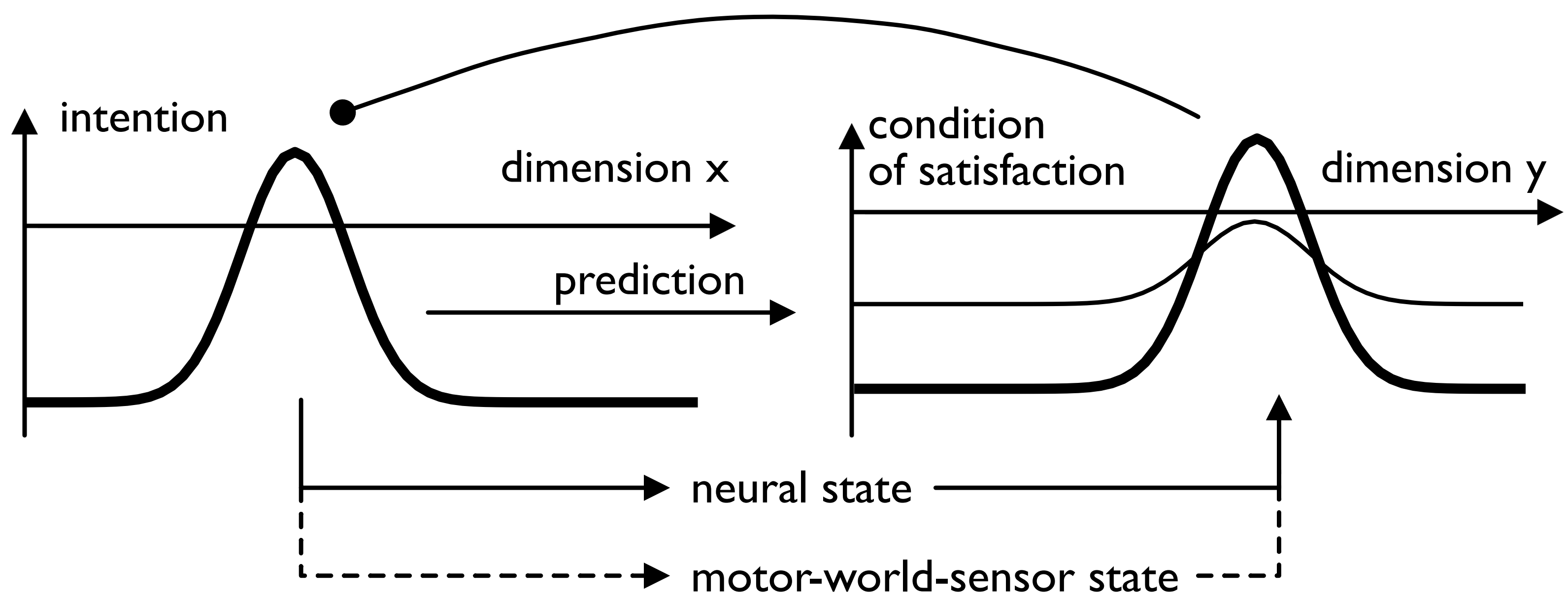
- Neuro-physics
- Neural dynamics
- Recurrent neural dynamics
- Neural fields: dynamics
- Neural fields: dimensions
- Binding
- Sequences
- Coordinate transforms
- Relational concepts, grounding, mental mapping
- Conclusions

Sequential processes

- How may neural attractors lead to the sequences of processing steps/actions that characterize higher cognition and behavior?

Sequential processes

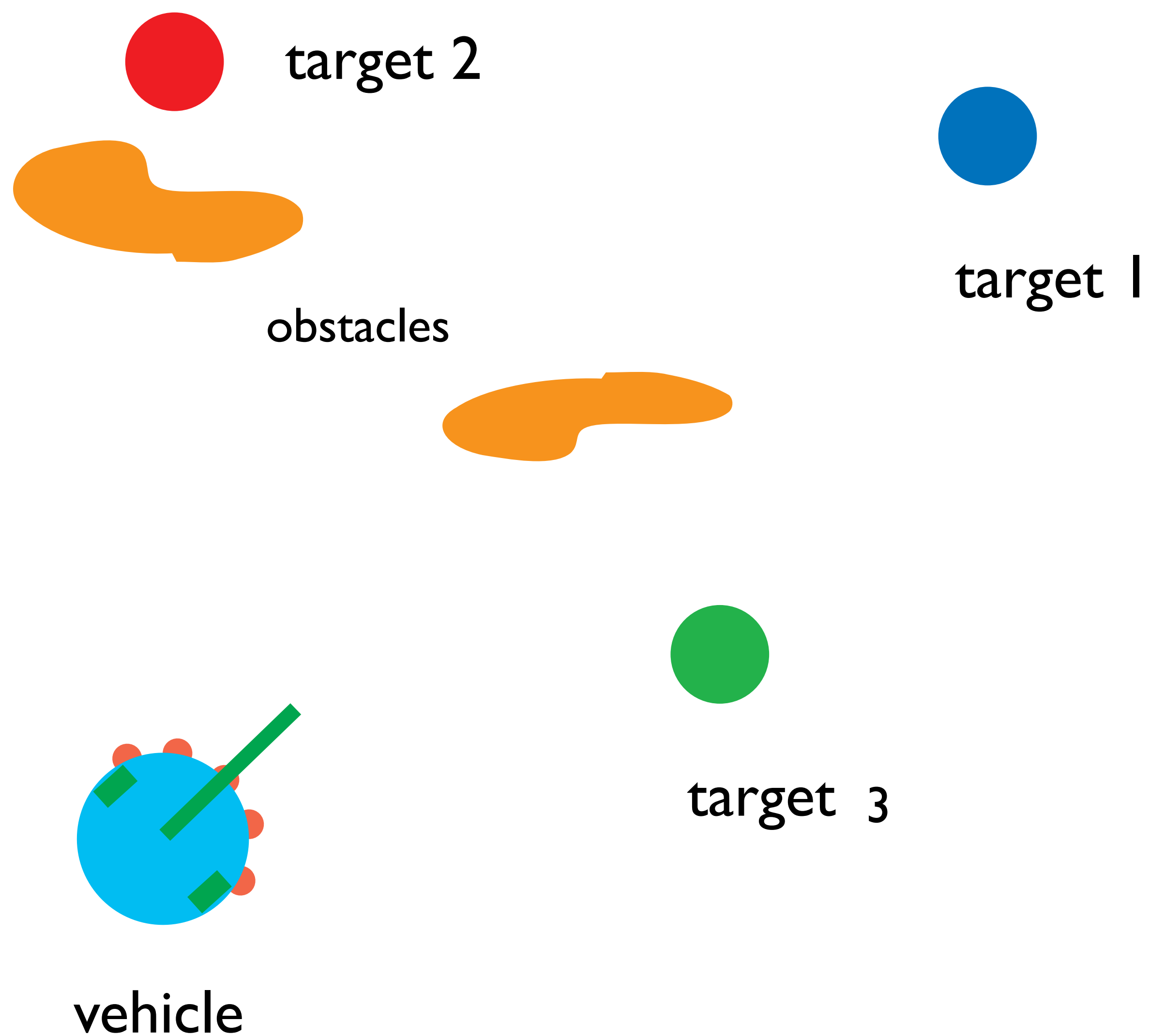
- the neural attractor = intention predicts its condition of satisfaction
- matching input detected => detection instability
- inhibits intention… => transition



[Sandamirskaya … 2010-2016]

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)

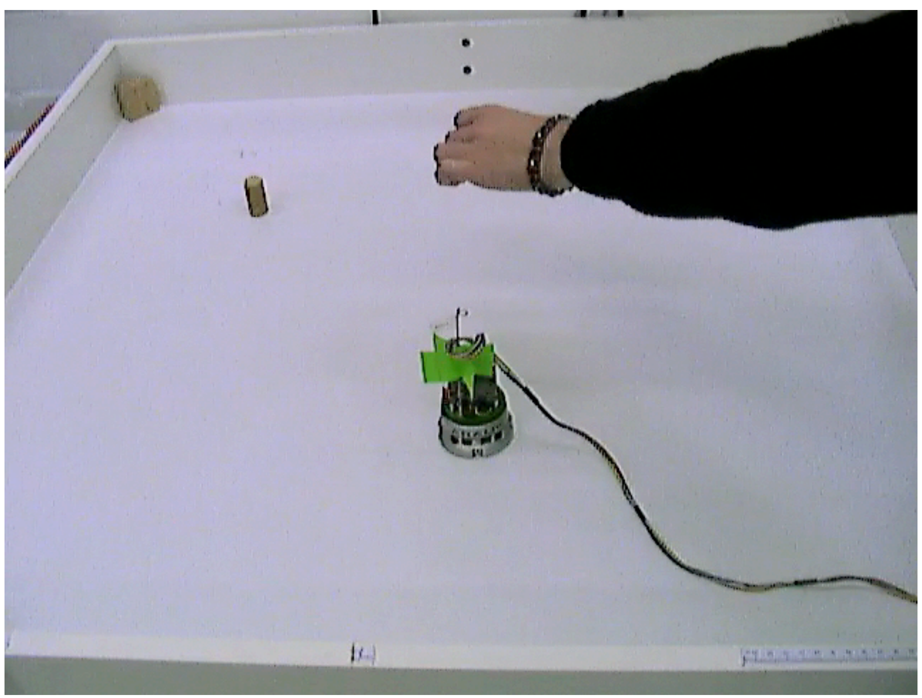


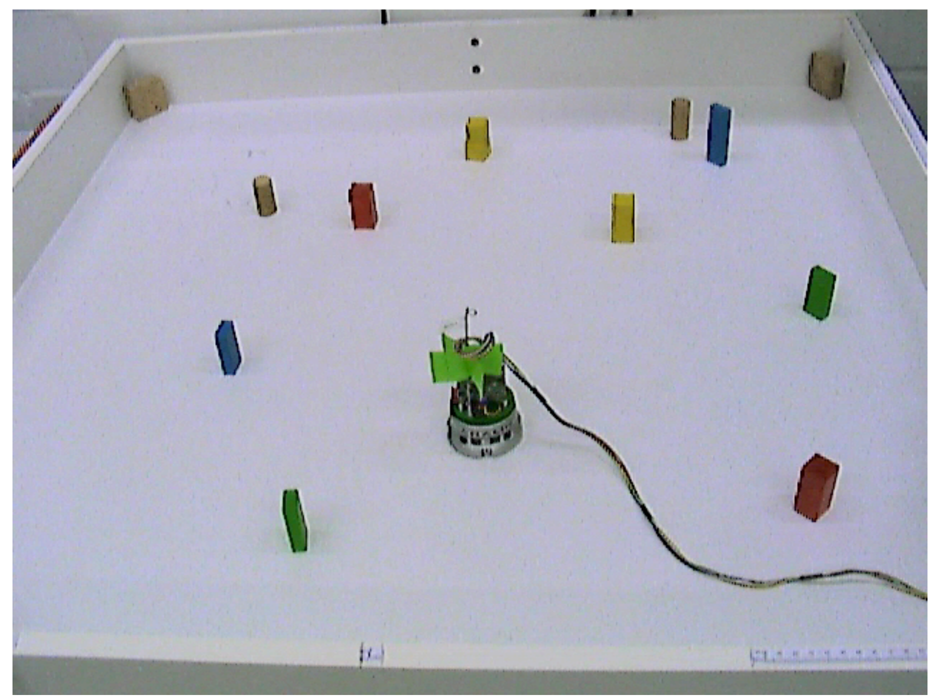
Implementation as an imitation task

- learn a serially ordered sequence from a single demonstration
- perform the serially ordered sequence with new timing

yellow-red-green-blue-red

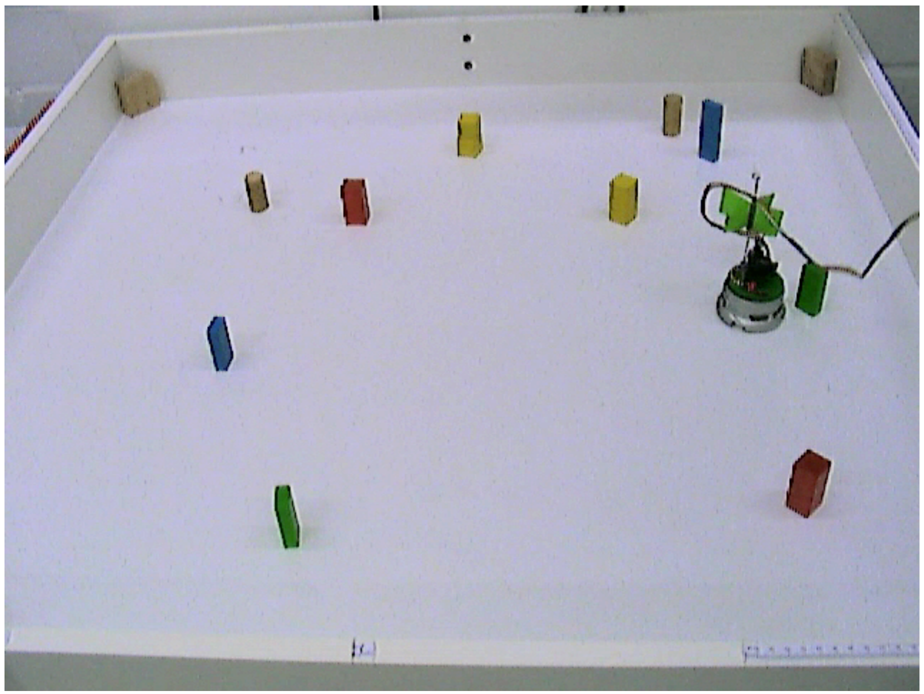
yellow-red-green-blue-red



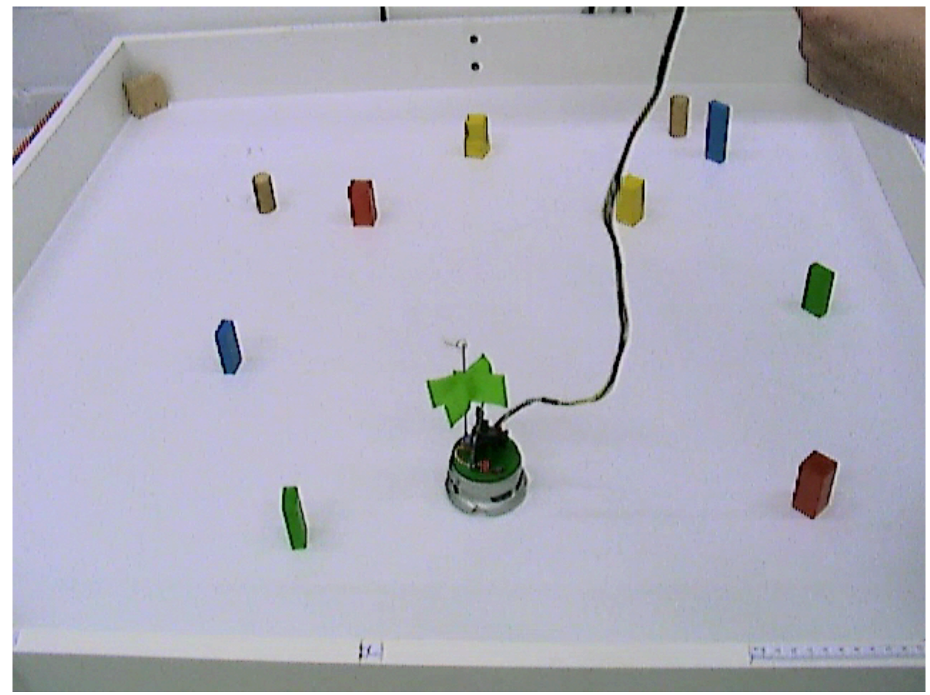


[Sandamirskaya, Schöner: *Neural Networks* 23:1163 (2010)]

red a distractor

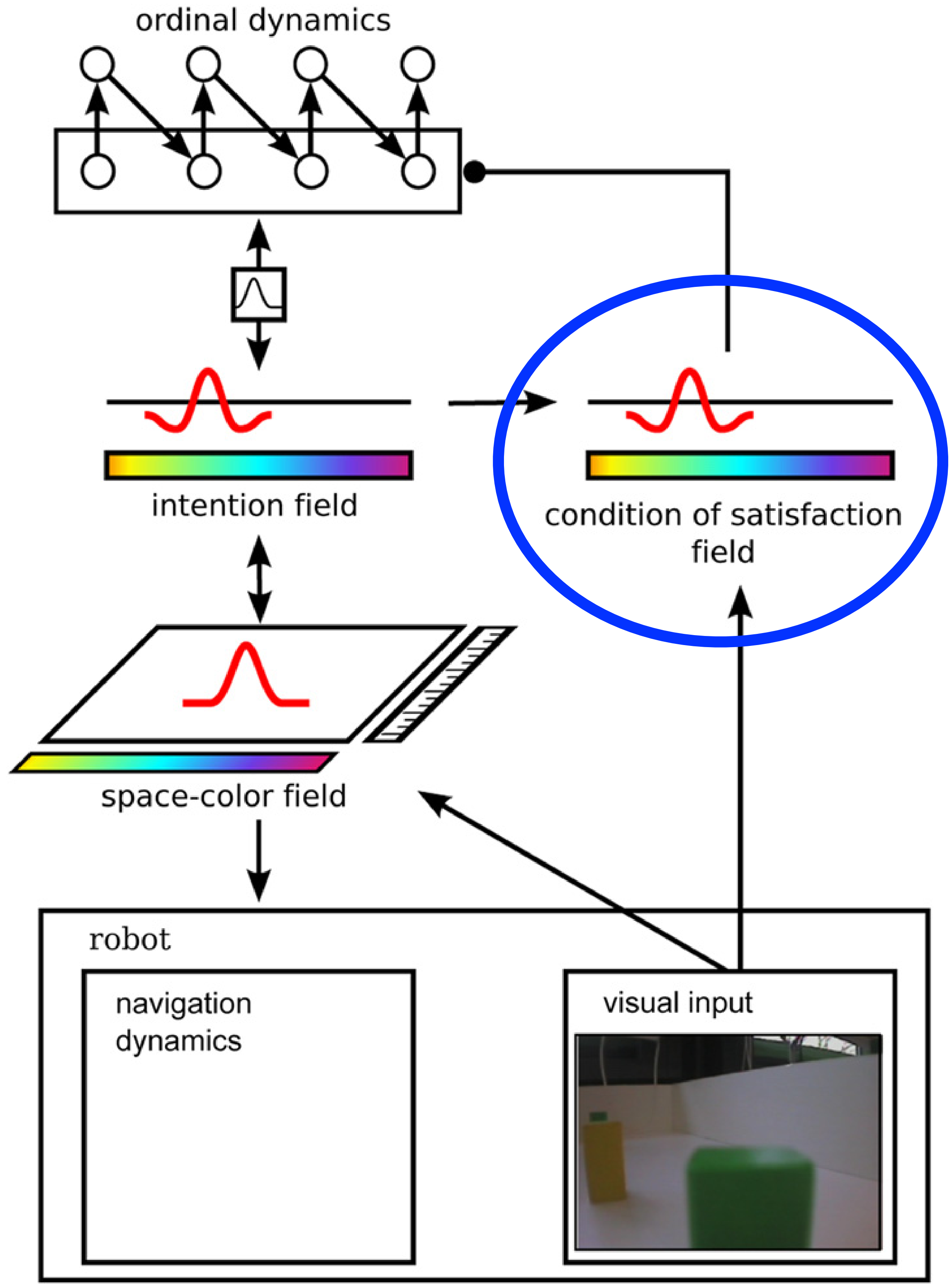


red a target



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

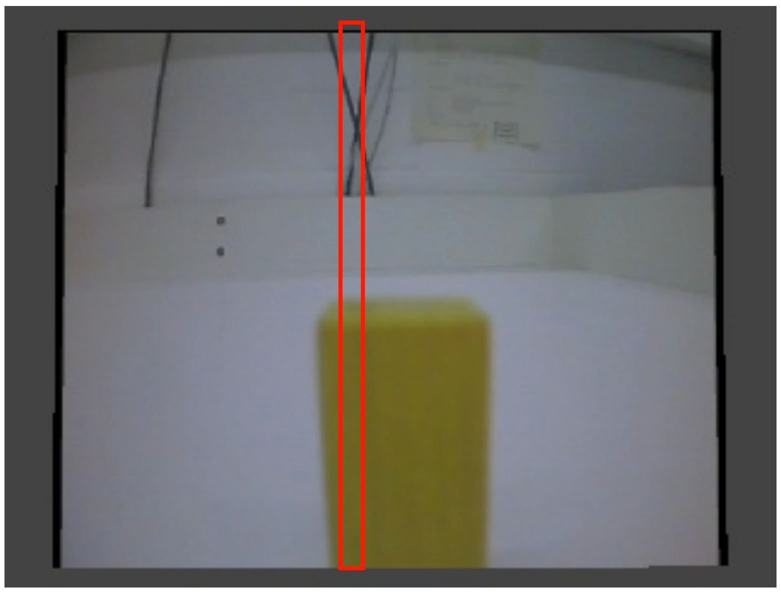
Condition of Satisfaction (CoS)

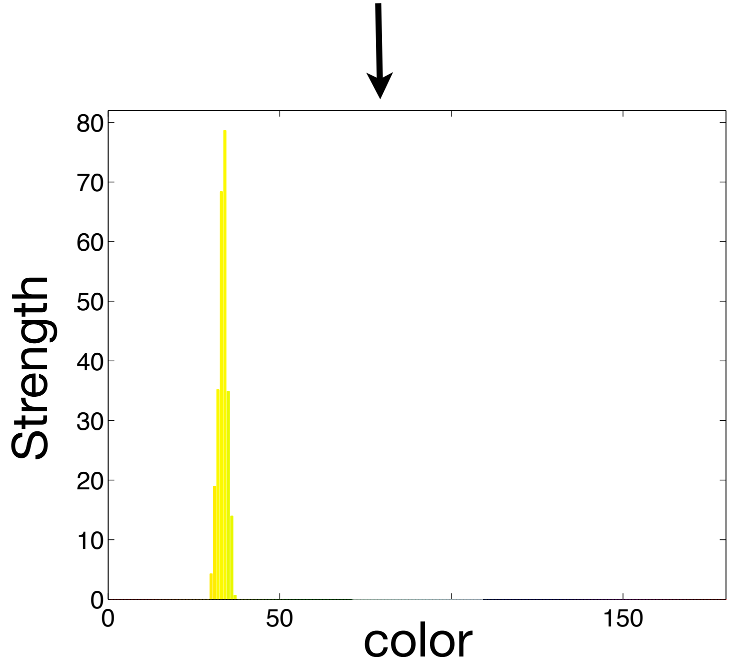


[Sandamirskaya, Schöner: *Neural Networks* 23:1163 (2010)]

Visual input

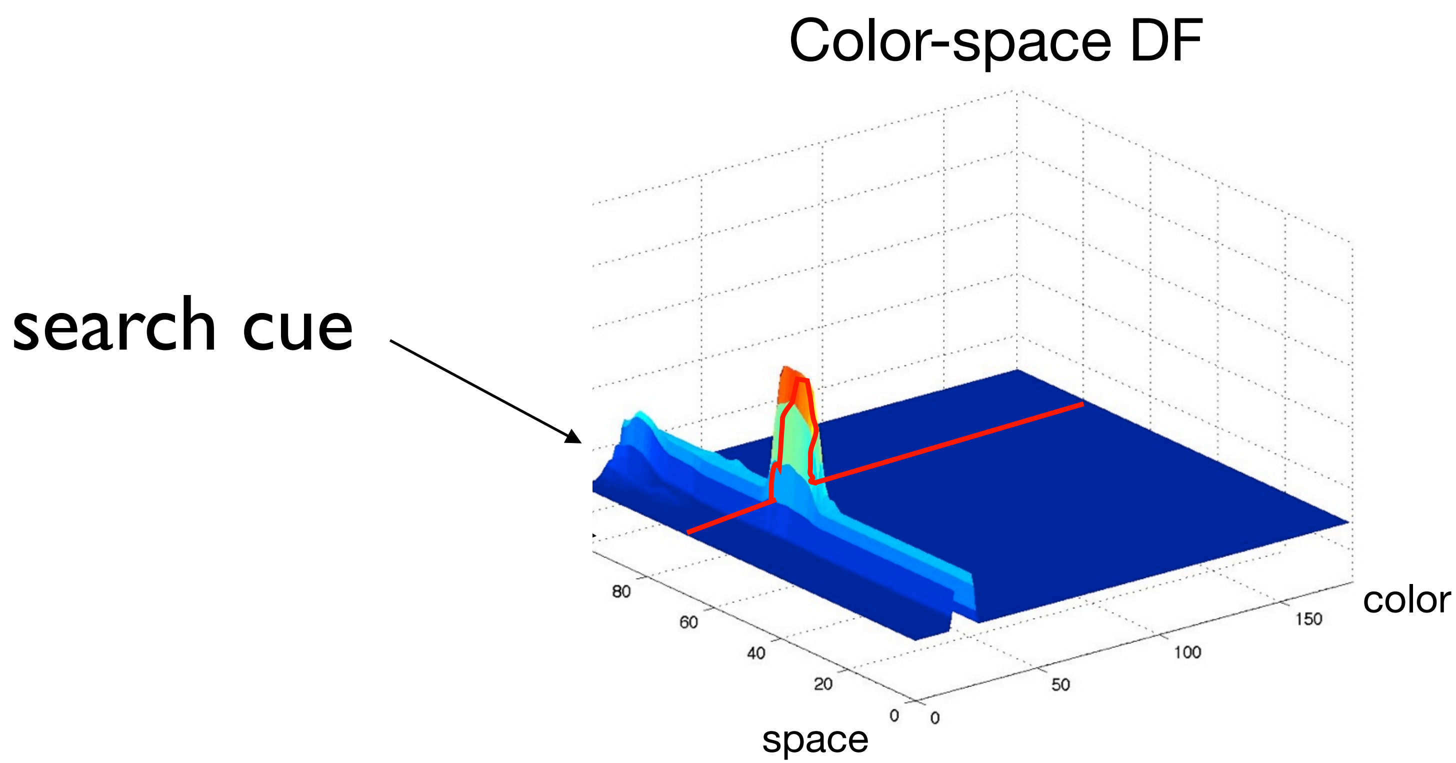
- 2D visual input
  - horizontal space
  - color
- "intensity" of 2D input from color histogram at each horizontal location





Visual search

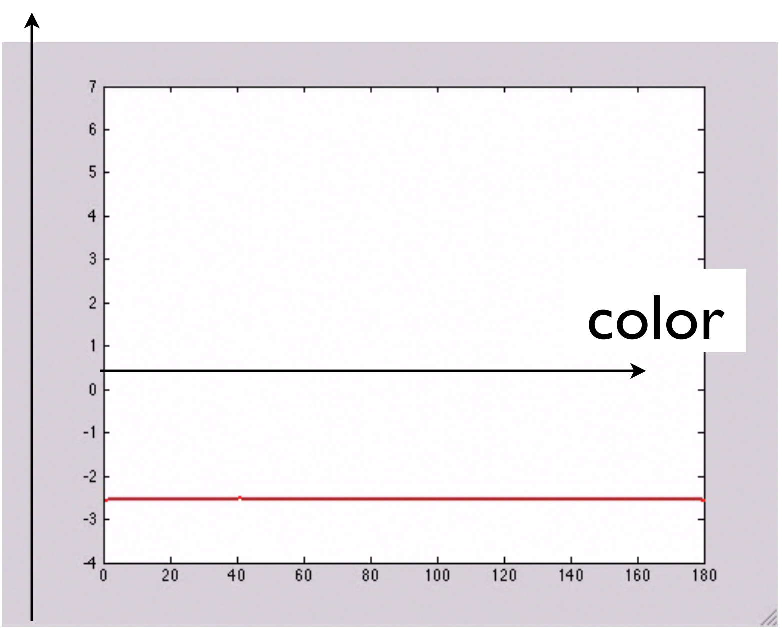
- intention=color cue provides ridge input into space-color field
- when that ridge overlaps with 2D space-color input => peak formed



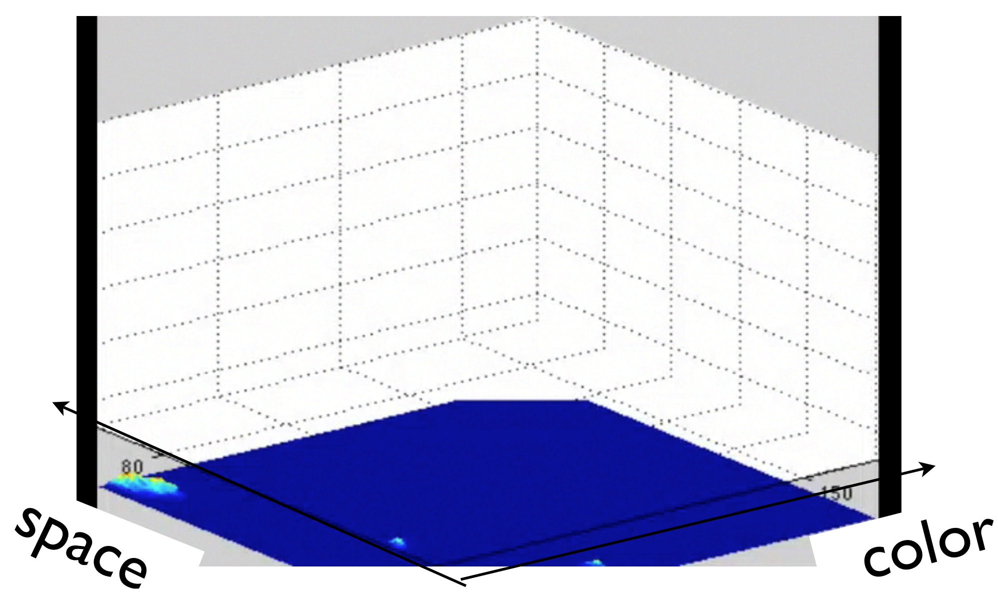
ordinal stack

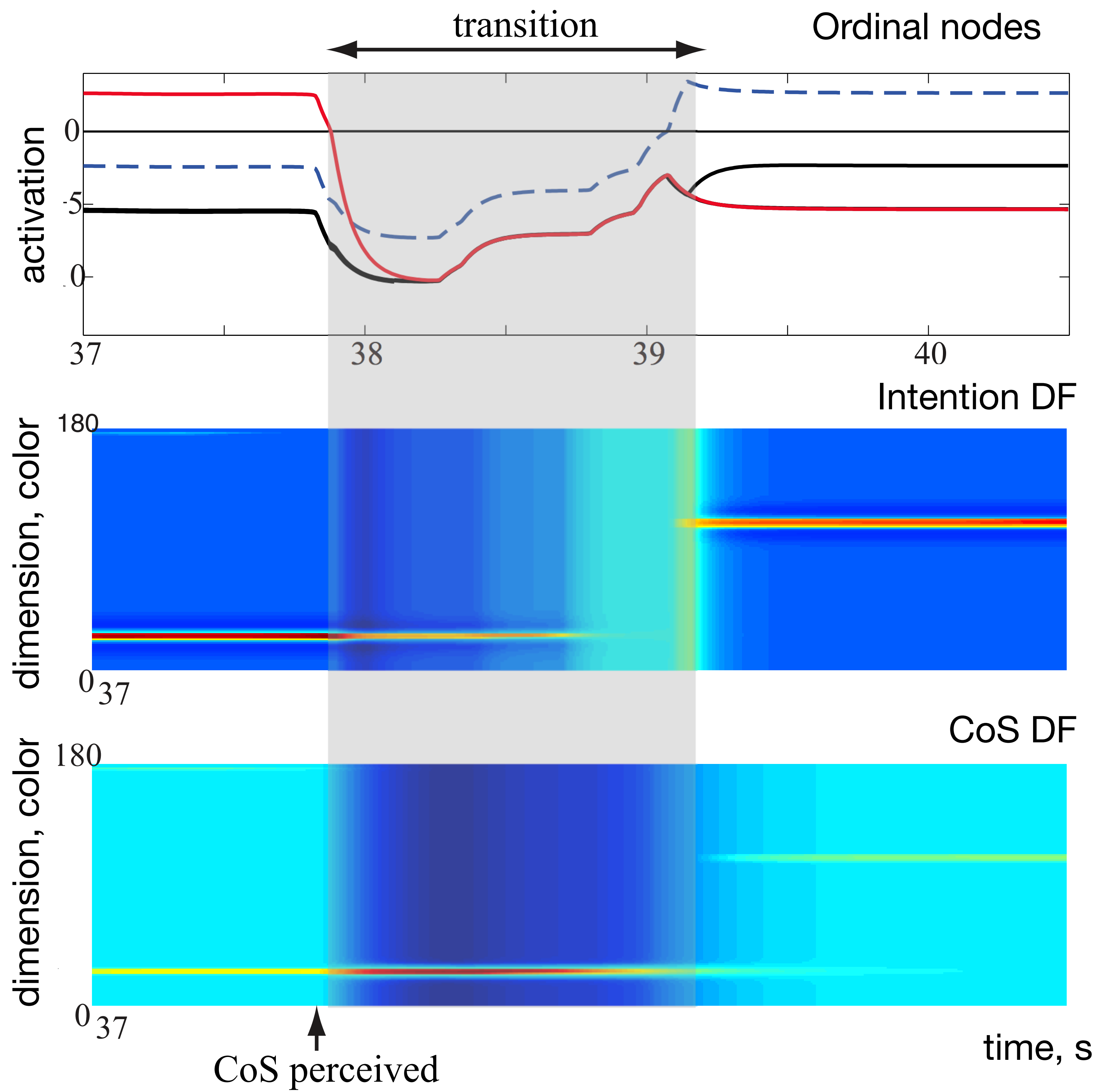
condition of satisfaction (CoS)

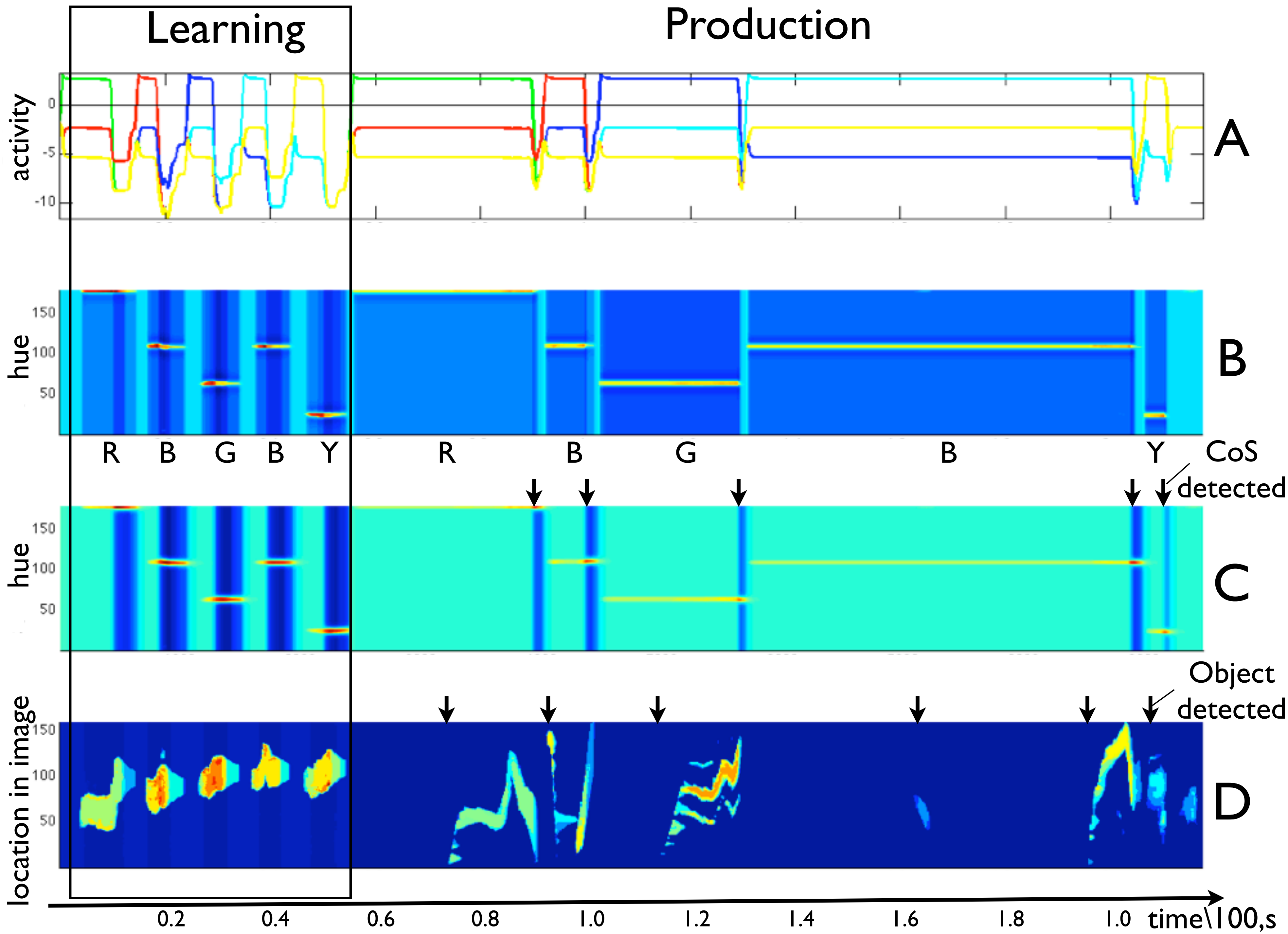
intentional state



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



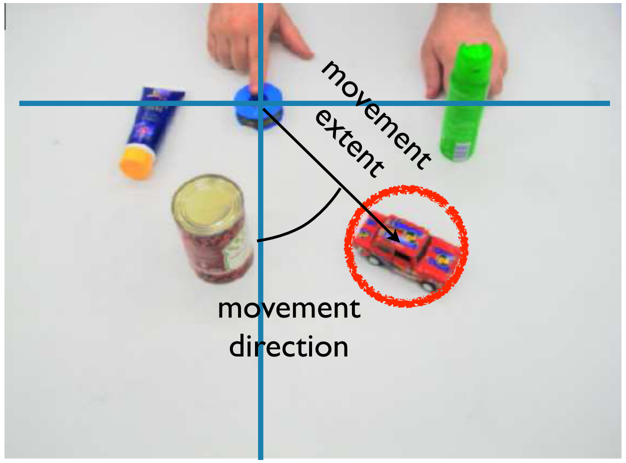
Learning
Production
activity
0
-5
-10
A
hue
150
100
50
B
R B G B Y
R B G B Y
CoS detected
hue
150
100
50
C
Object detected
location in image
150
100
50
D
0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 1.0 time\100,s

Roadmap

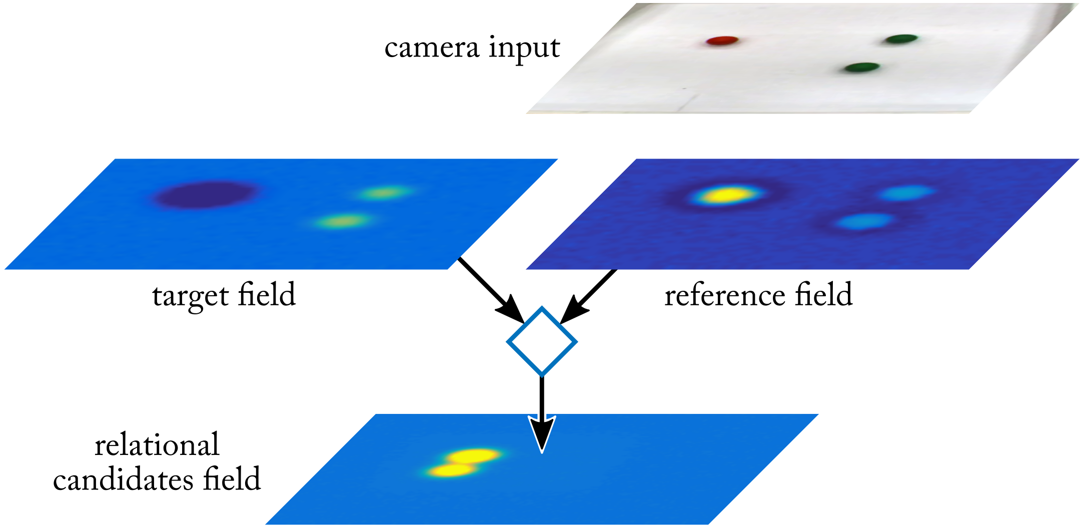
- Neuro-physics
- Neural dynamics
- Recurrent neural dynamics
- Neural fields: dynamics
- Neural fields: dimensions
- Binding
- Sequences
- Coordinate transforms
- Relational concepts, grounding, mental mapping
- Conclusions

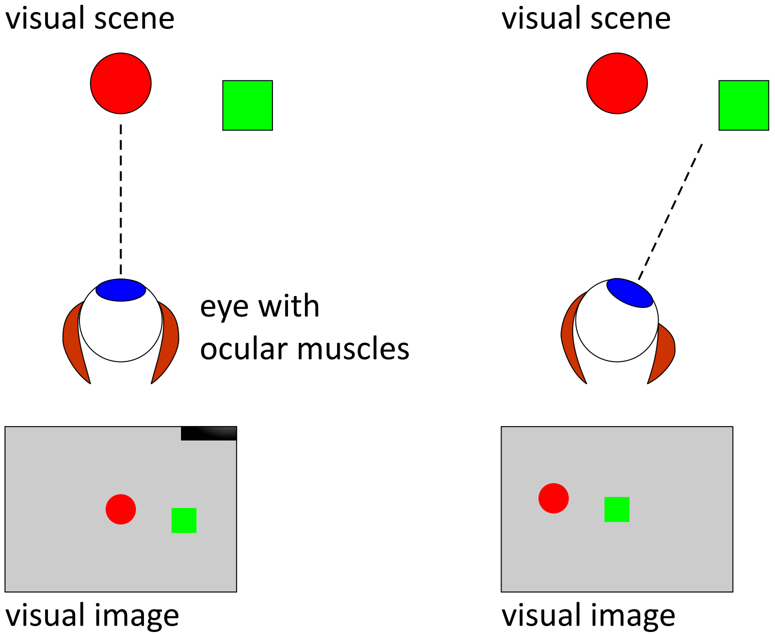
Coordinate transforms

- are central to sensory-motor cognition but also critical to higher cognition!



"where are the green objects relative to the red one"



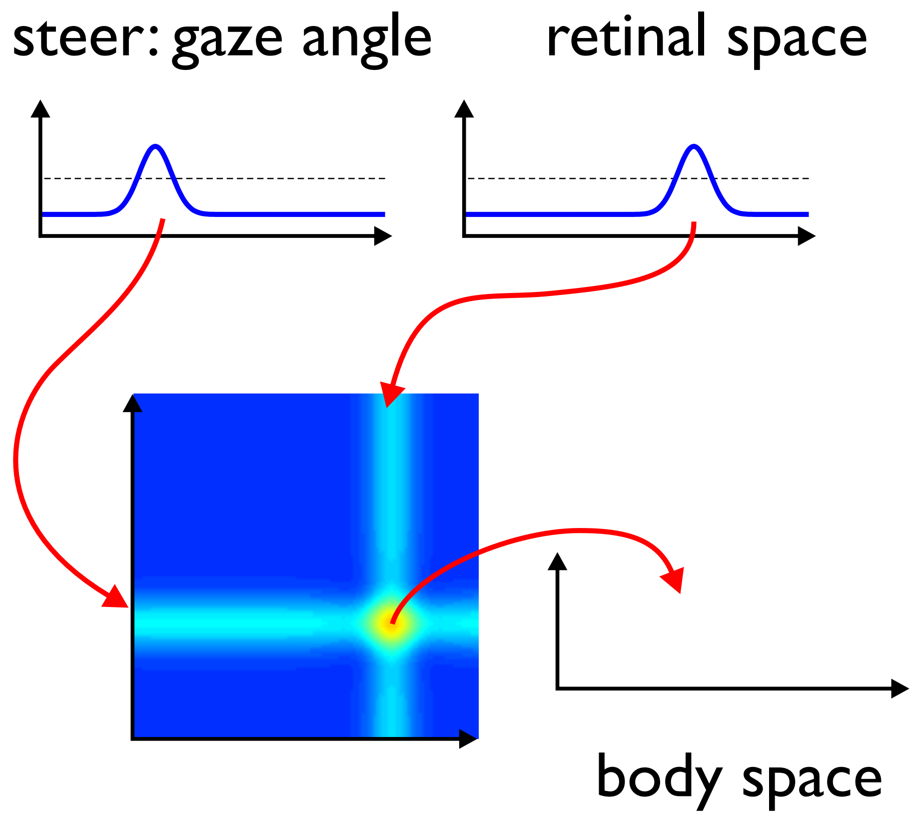


[Richer Doctoral dissertation, 2017]

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

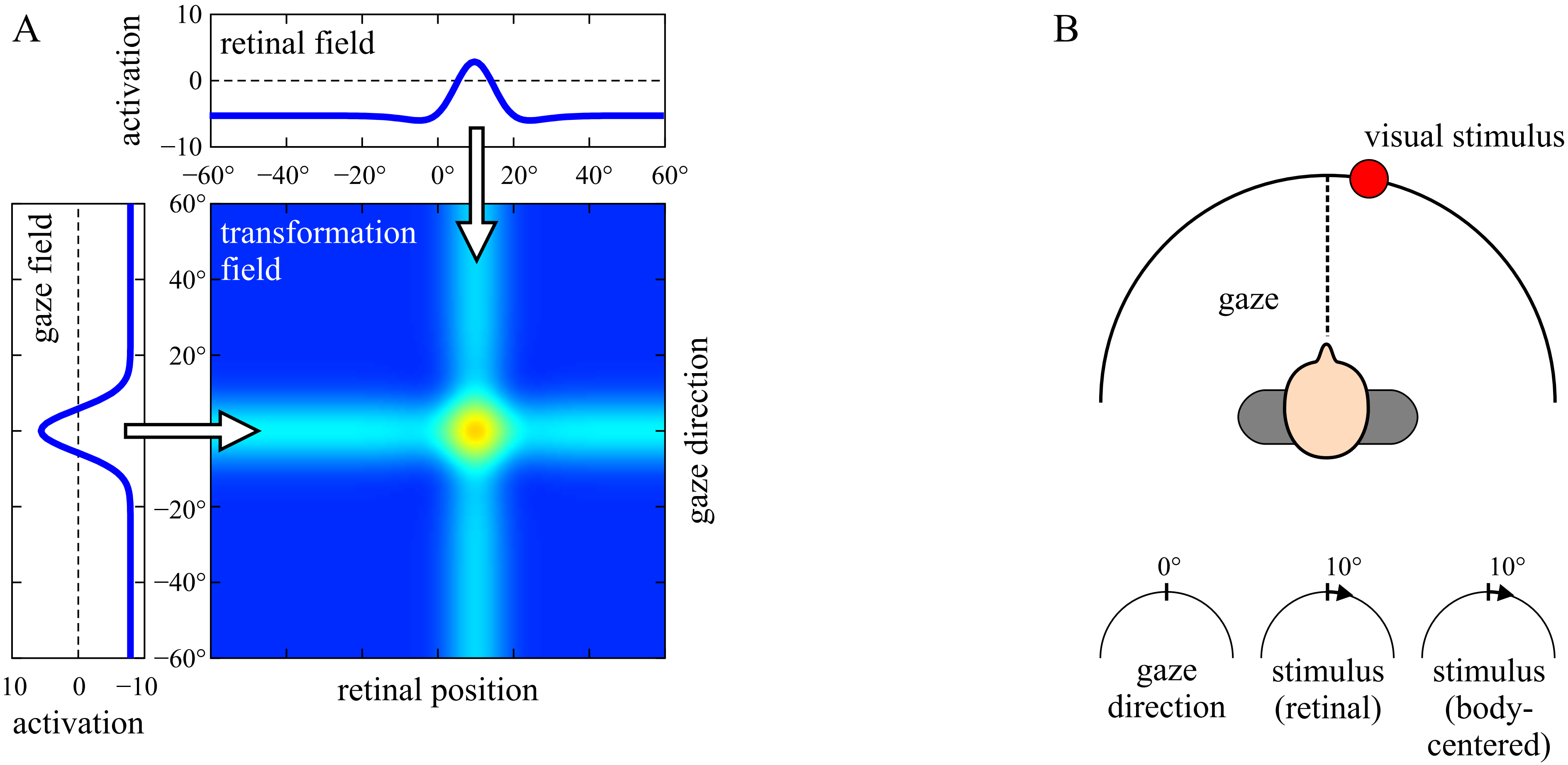
Coordinate transforms involve binding

- need a bound neural representation of
  - retinal space
  - gaze angle
- project to body space
- neural evidence: gain field (Andersen/Pouget)



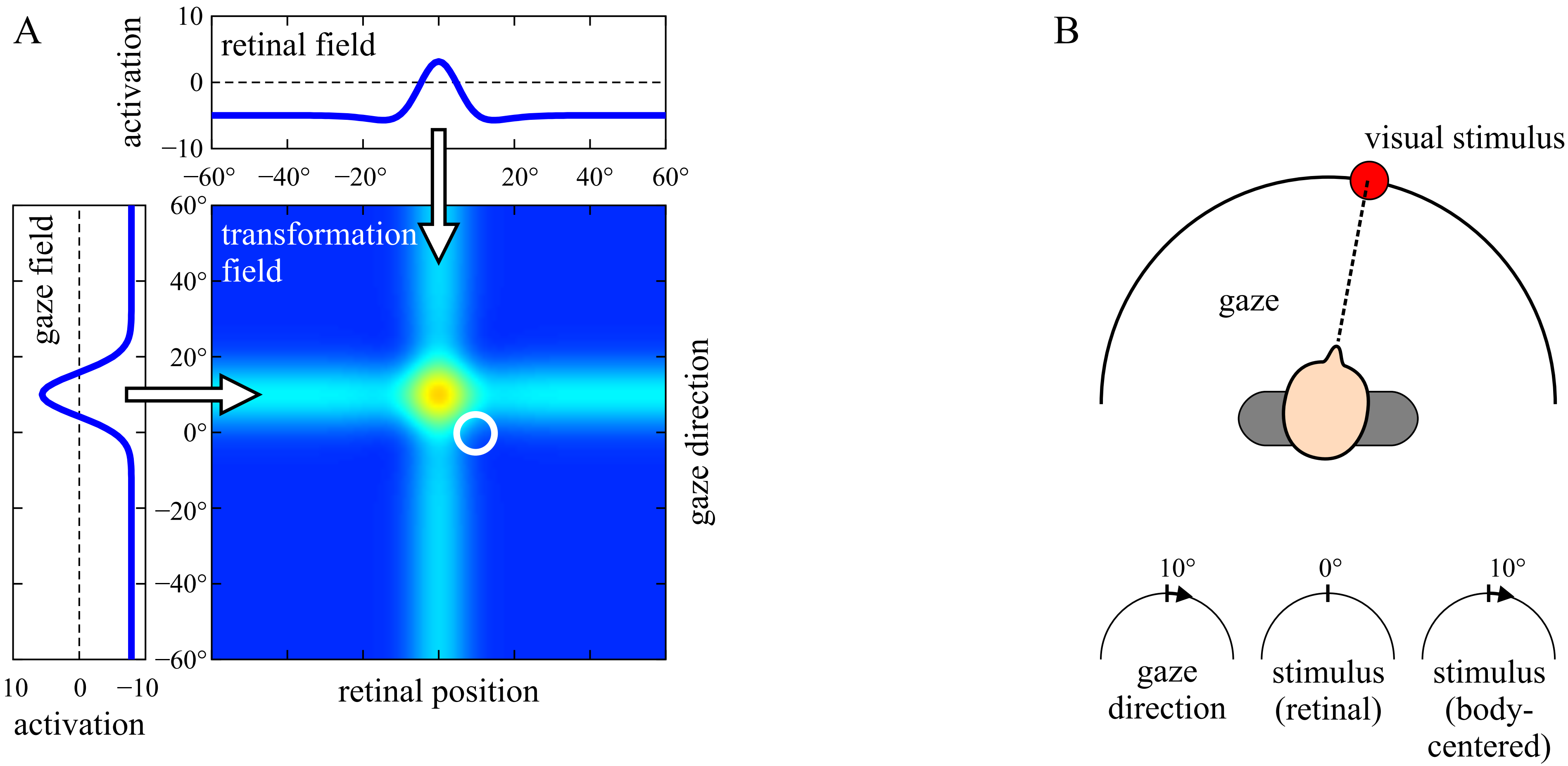
[Schneegans Ch 7 of *DFT Primer*, 2016]

Coordinate transform



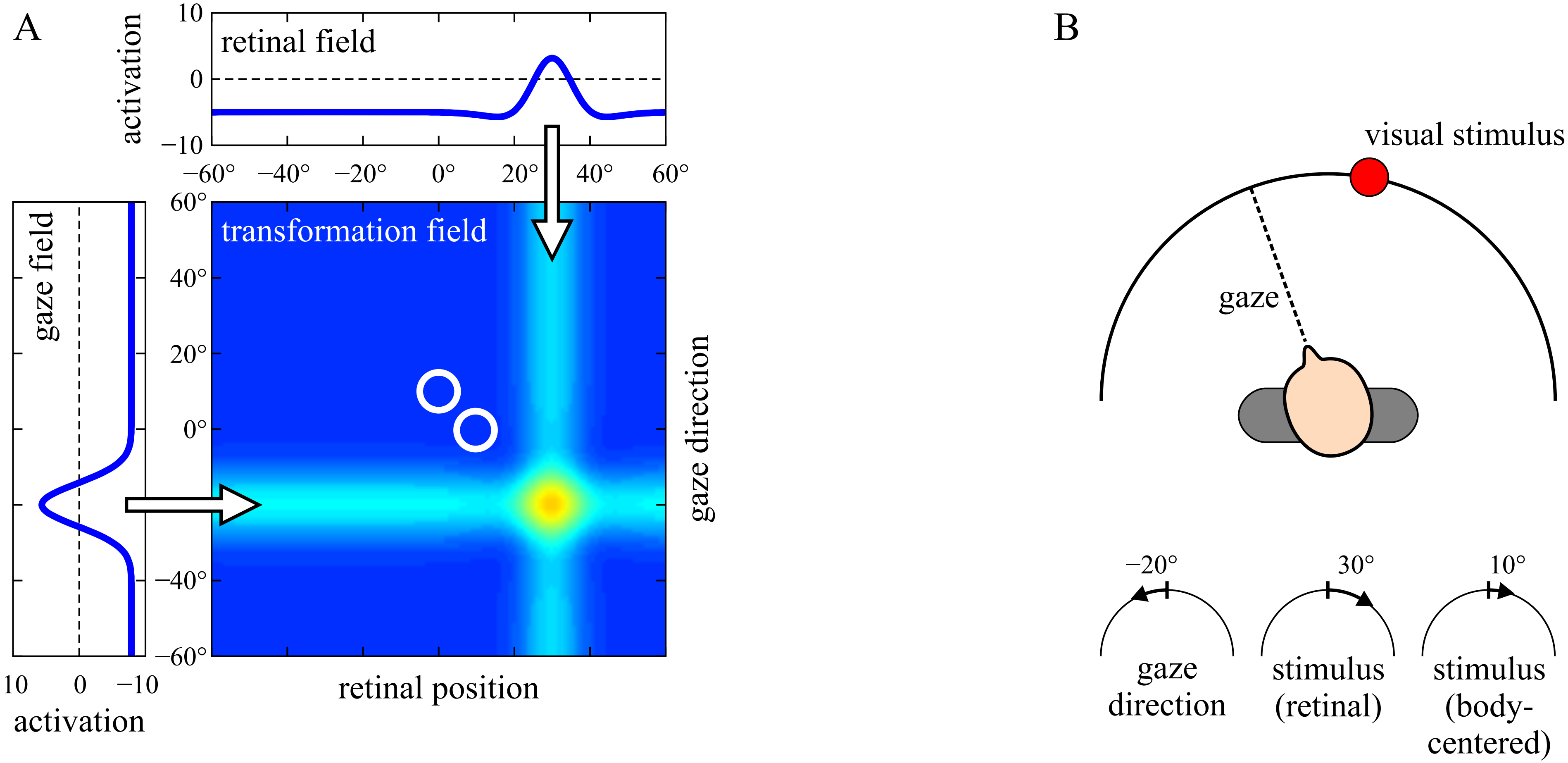
[Schneegans Ch 7 of *DFT Primer*, 2016]

Coordinate transform



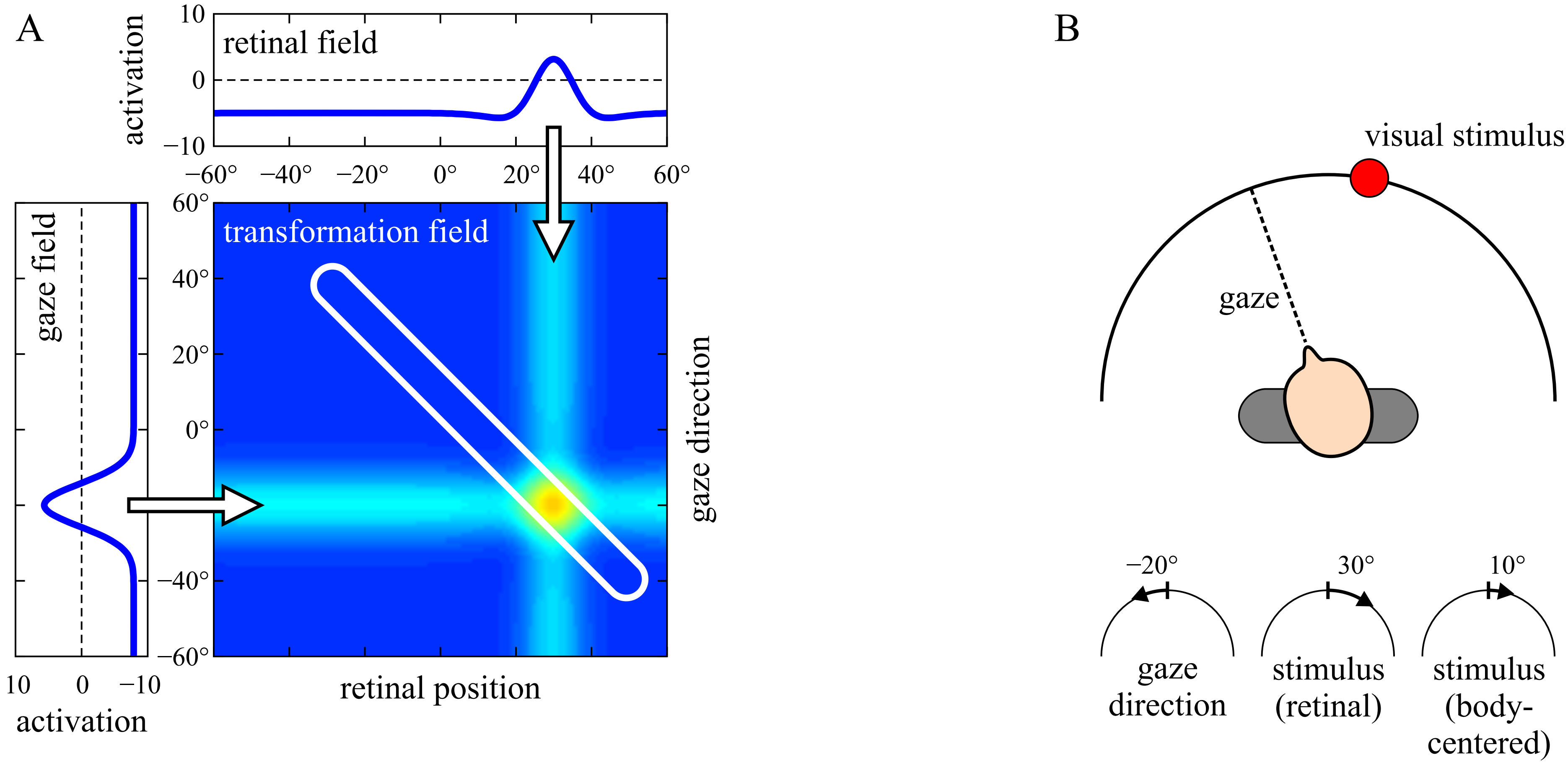
[Schneegans Ch 7 of *DFT Primer*, 2016]

Coordinate transform



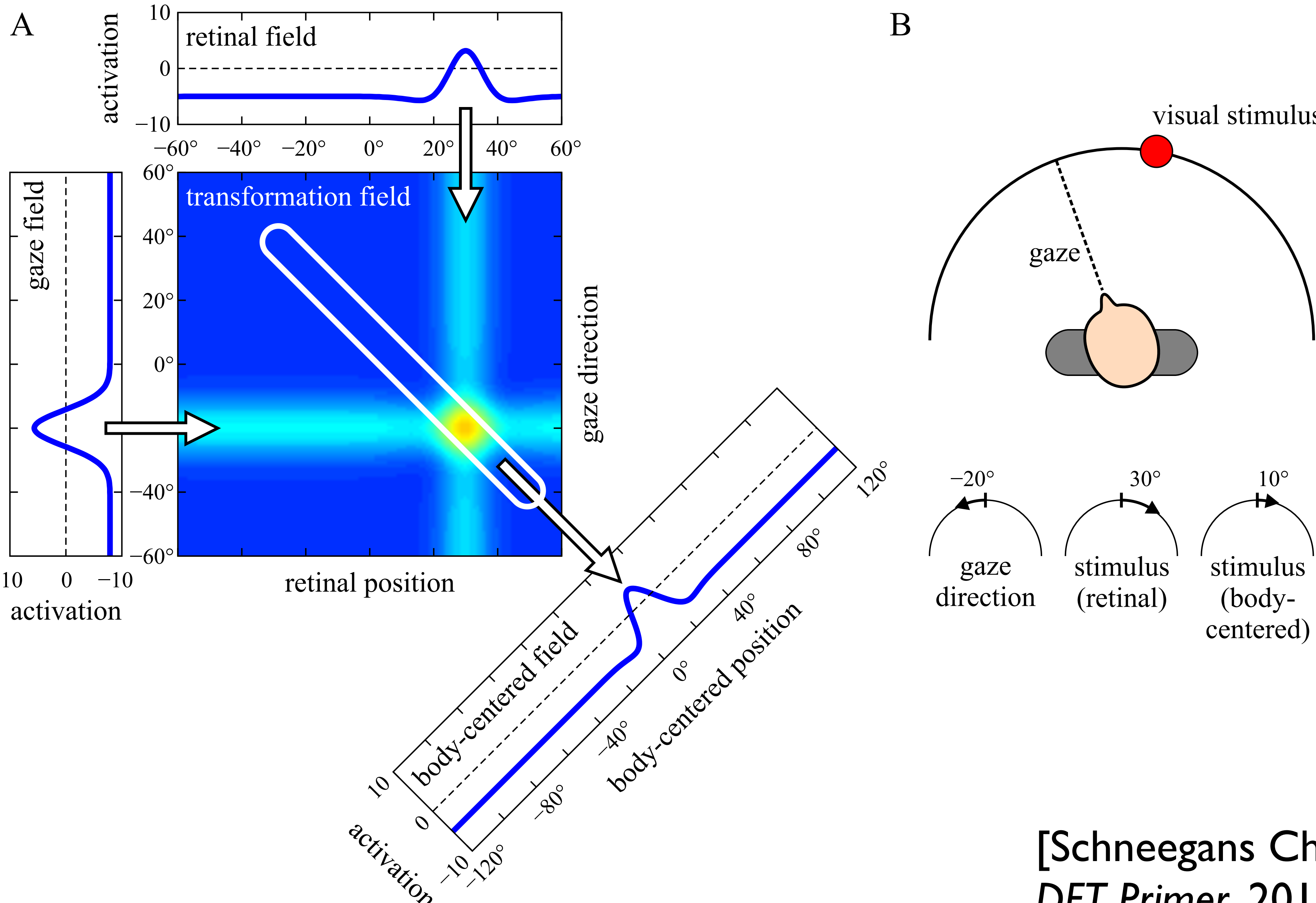
[Schneegans Ch 7 of *DFT Primer*, 2016]

Coordinate transform



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

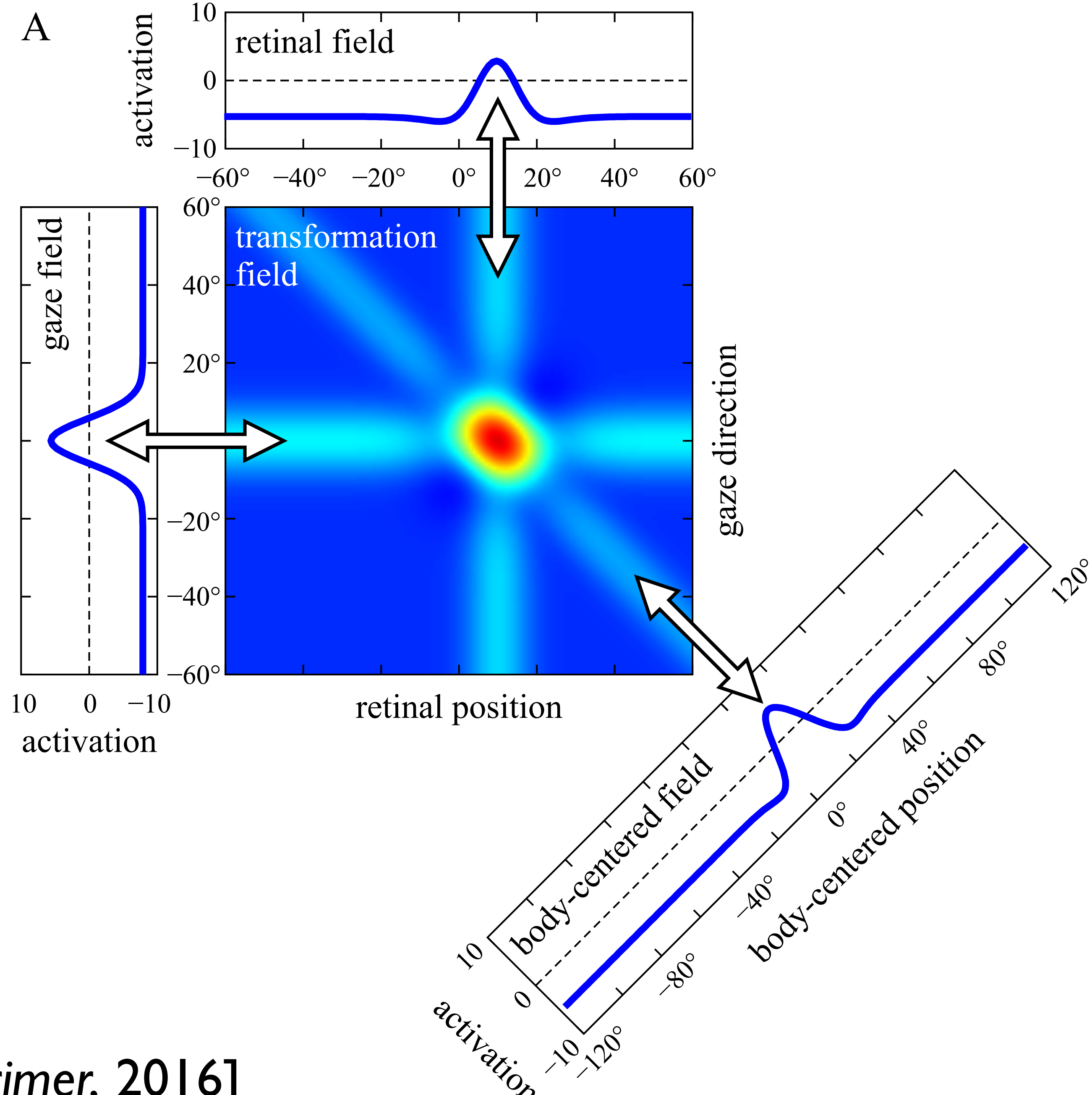
Coordinate transform



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

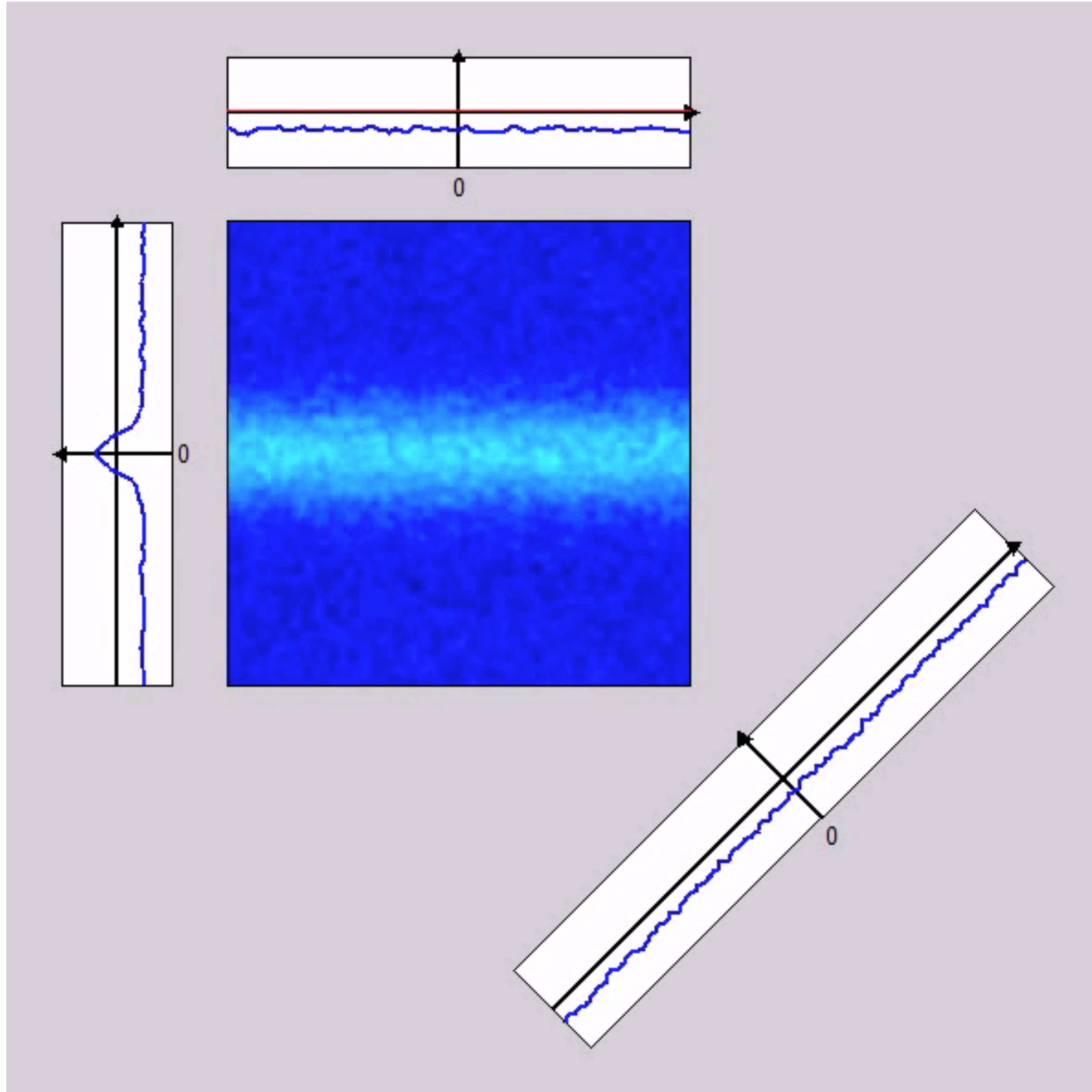
Retina => body space

- bi-directional coupling
- => predict retinal coordinates



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

Spatial remapping during saccades



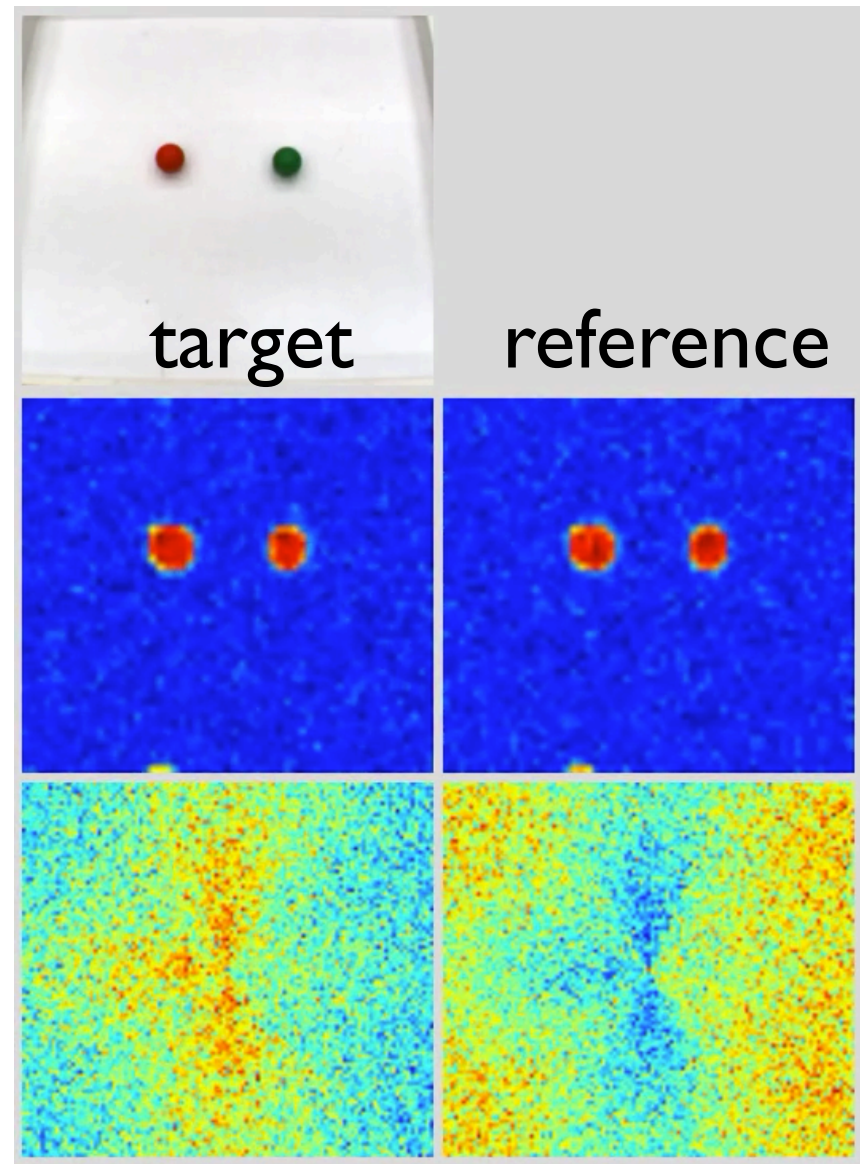
[Schneegans, Schöner *Biological Cybernetics* 2012]

Roadmap

- Neuro-physics
- Neural dynamics
- Recurrent neural dynamics
- Neural fields: dynamics
- Neural fields: dimensions
- Binding
- Sequences
- Coordinate transforms
- Relational concepts, grounding, mental mapping
- Conclusions

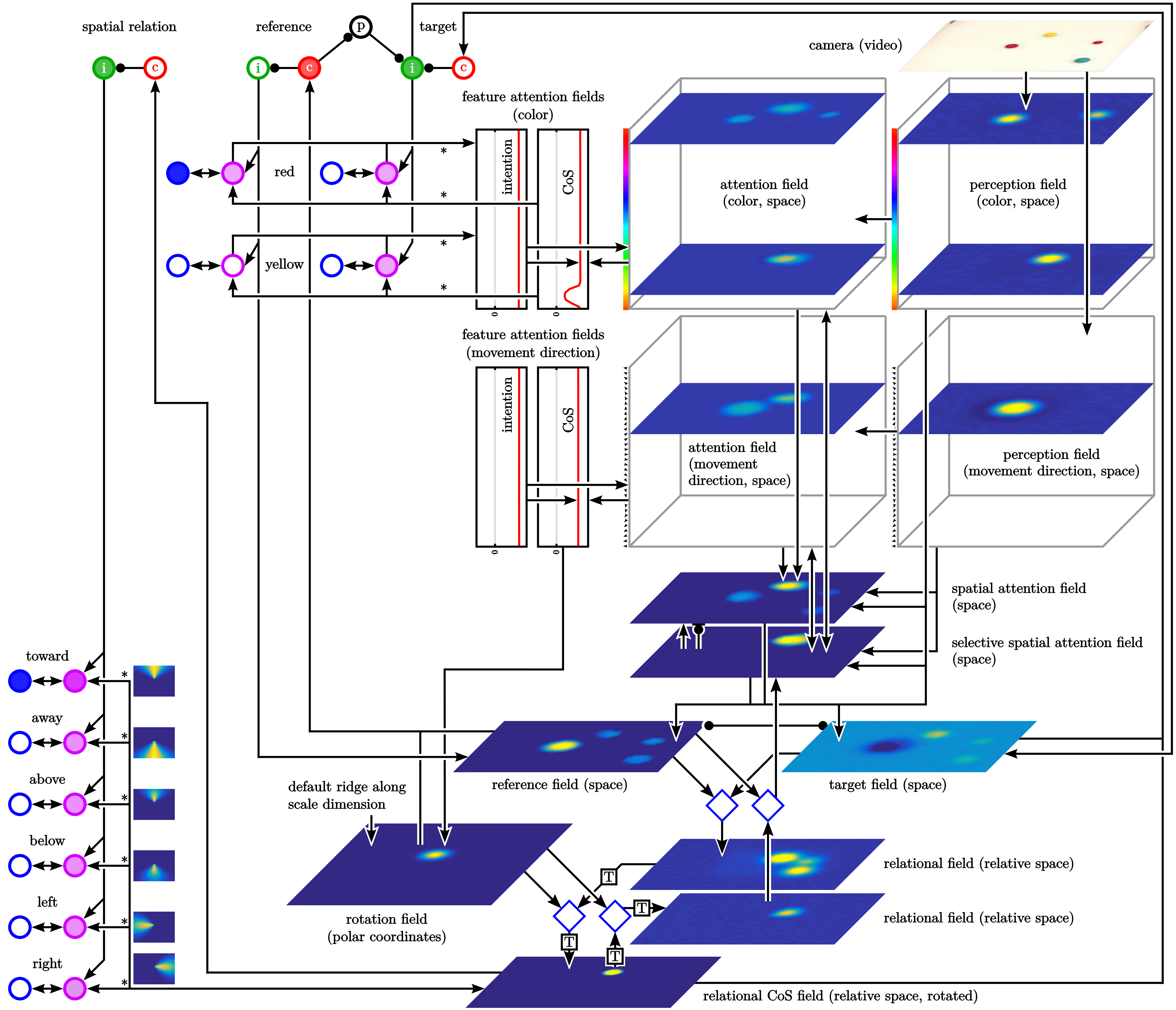
Concepts, relational thinking

"red to the left of green"



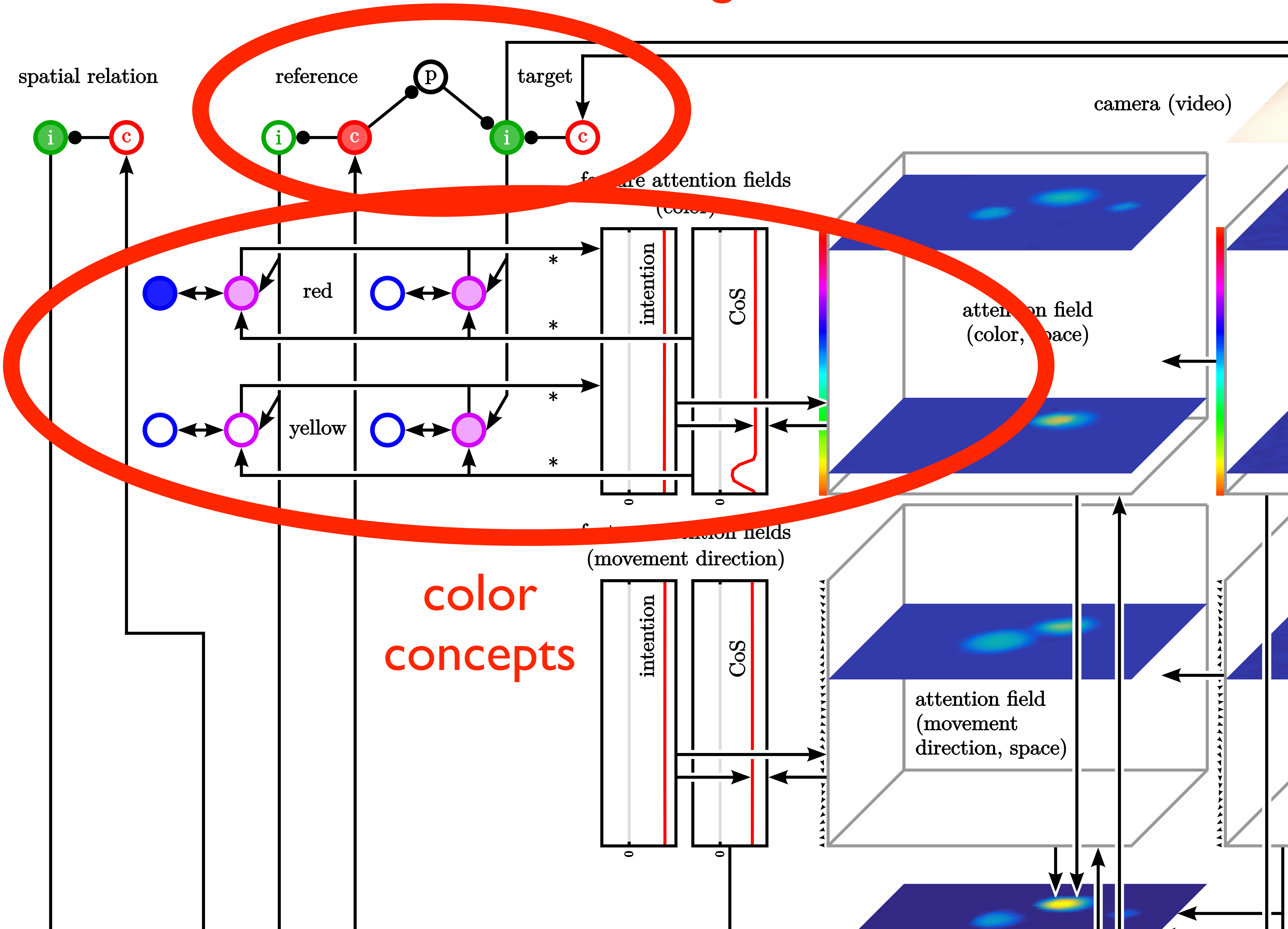
- grounding: bringing the target object of a relational phrase into the attentional foreground

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



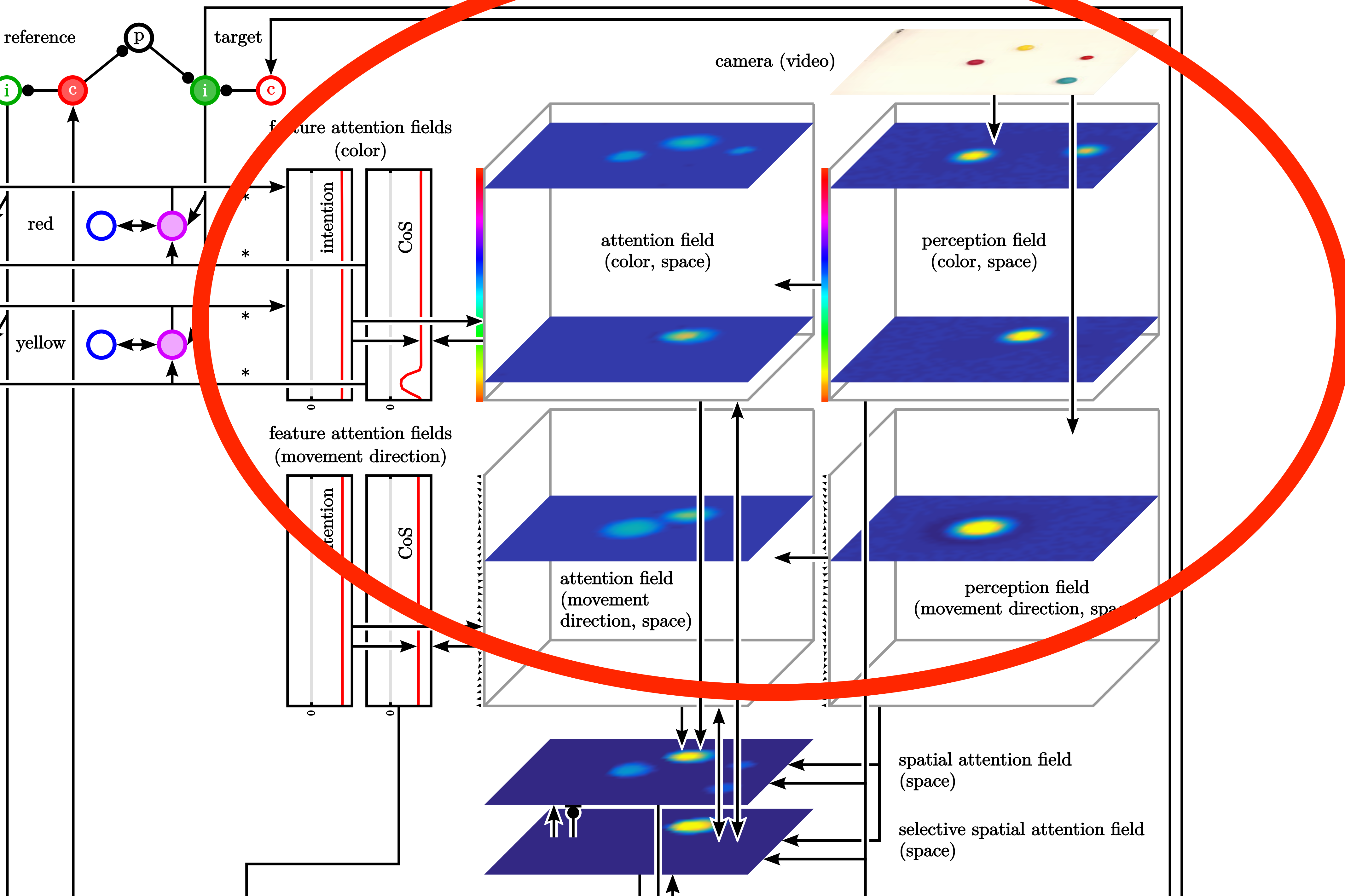
[Richter, Lins, Schöner, ToPiC (2017)]

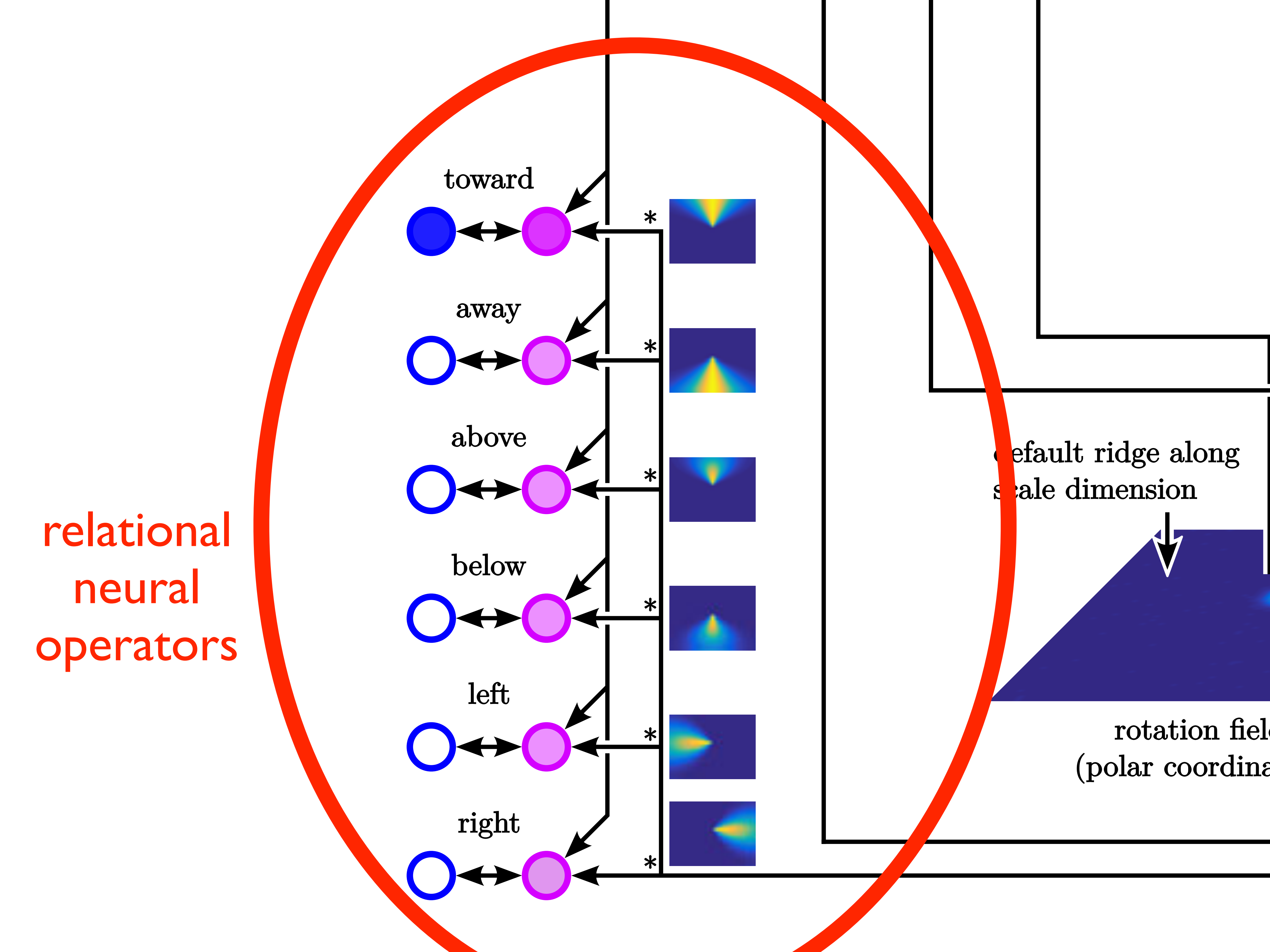
binding to role

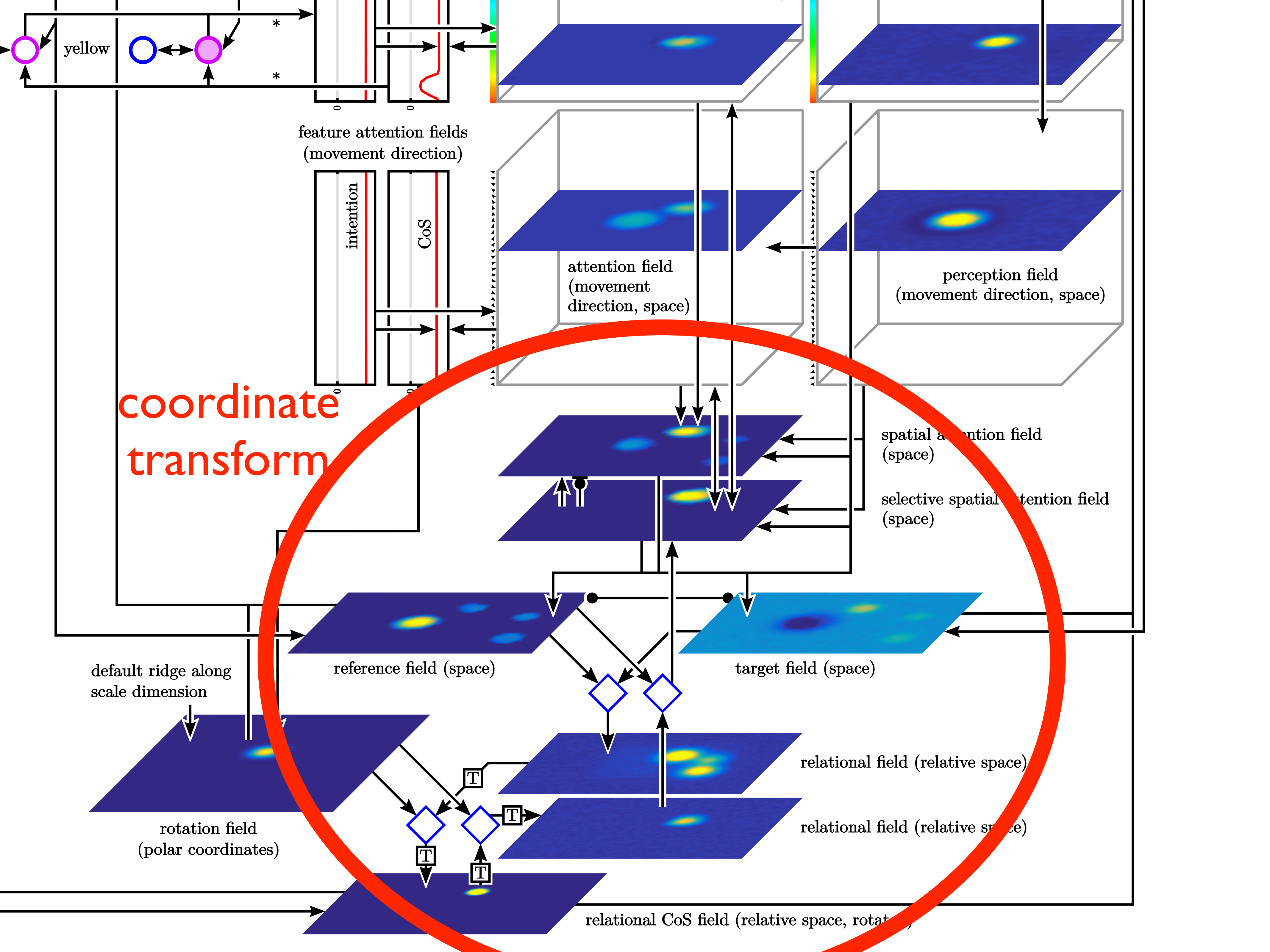


color concepts

cued visual search

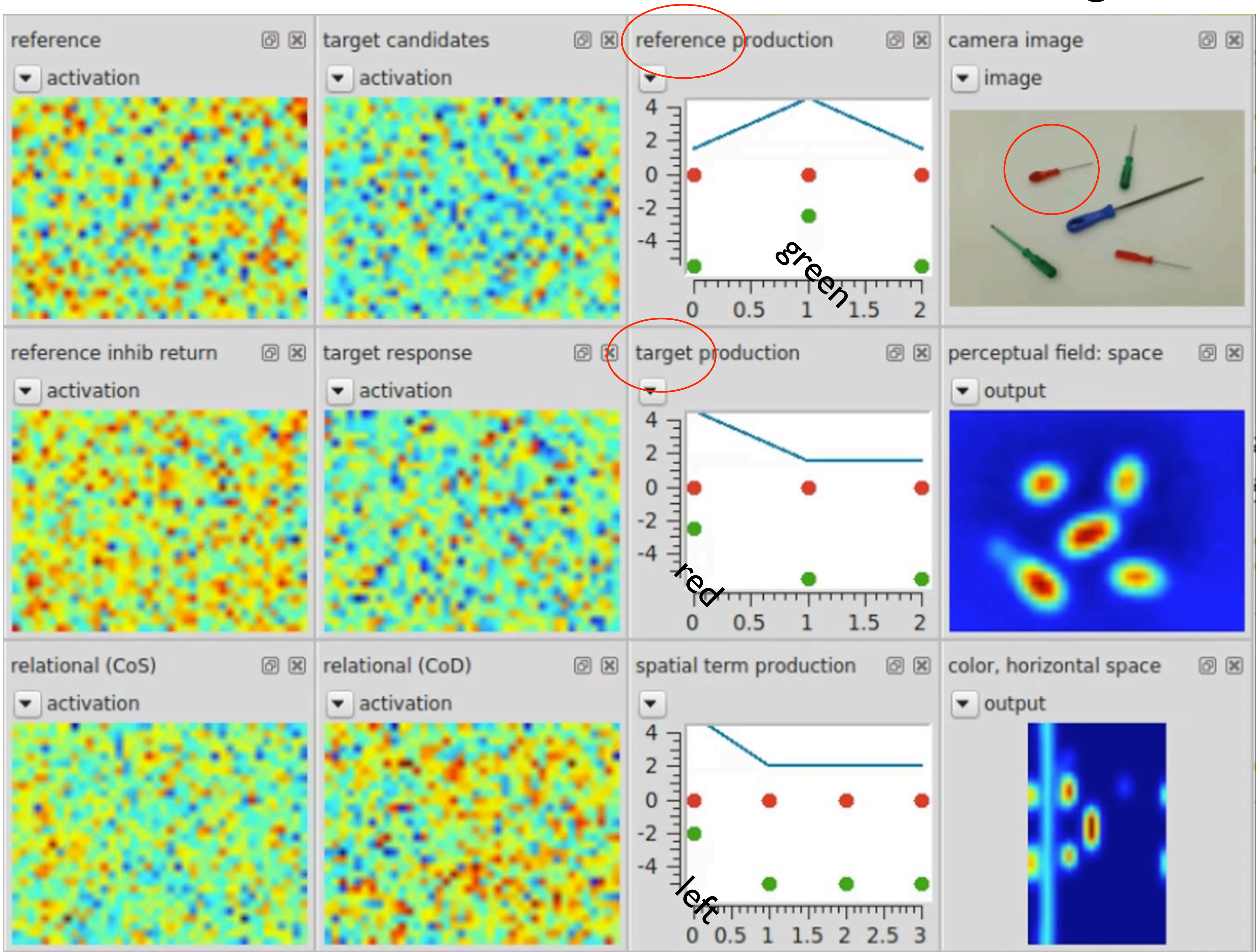



relational
neural
operators
toward
away
above
below
left
right
*
efault ridge along
ale dimension
rotation fiel
(polar coordina



yellow
feature attention fields
(movement direction)
intention
CoS
attention field
(movement
direction, space)
perception field
(movement direction, space)
coordinate
transform
spatial a ntion field
(space)
selective spatial tention field
(space)
reference field (space)
target field (space)
default ridge along
scale dimension
rotation field
(polar coordinates)
relational field (relative space)
relational field (relative sp ce)
relational CoS field (relative space, rotat)

“red to the left of green”



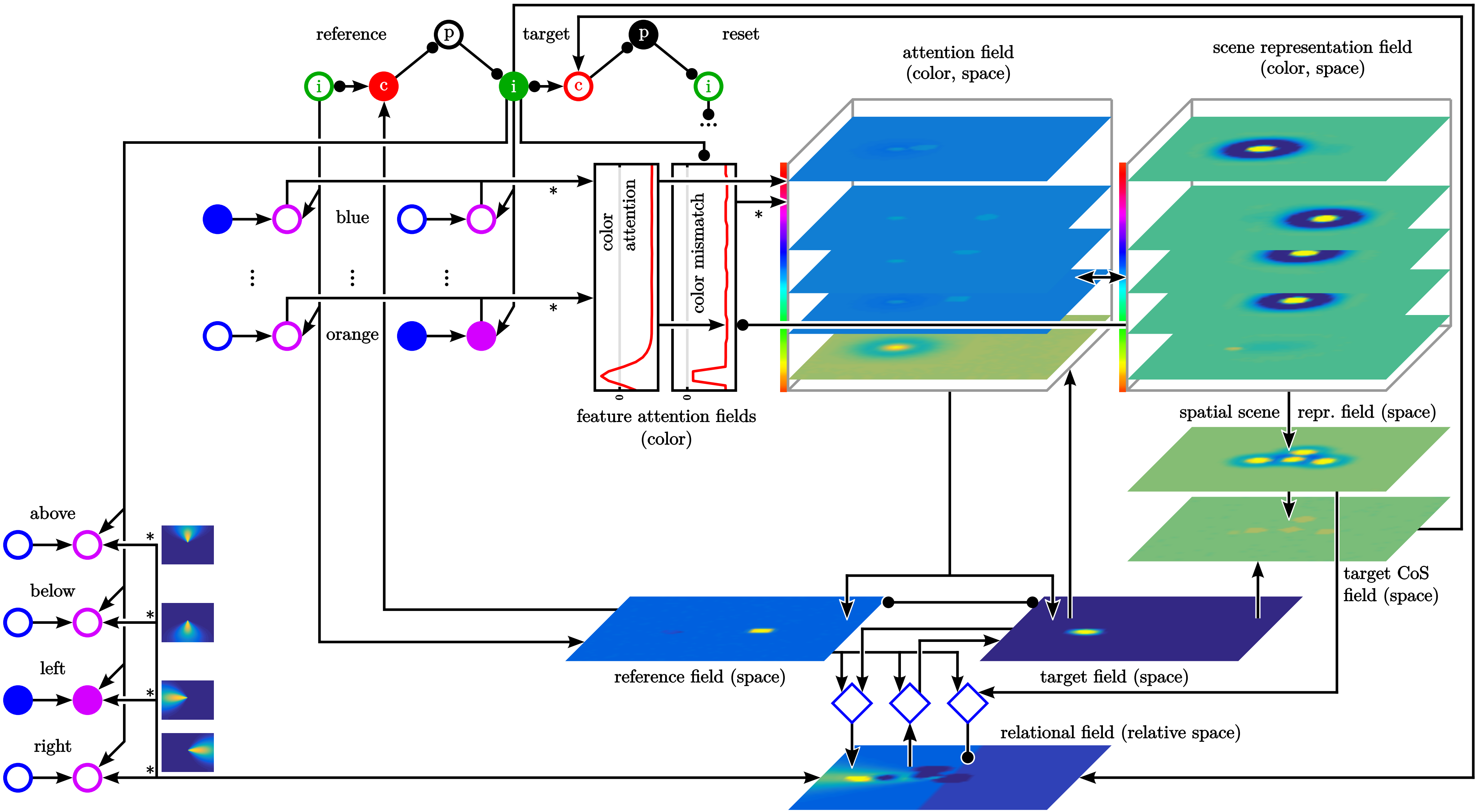
Concepts, relational thinking

- => hands on exercise

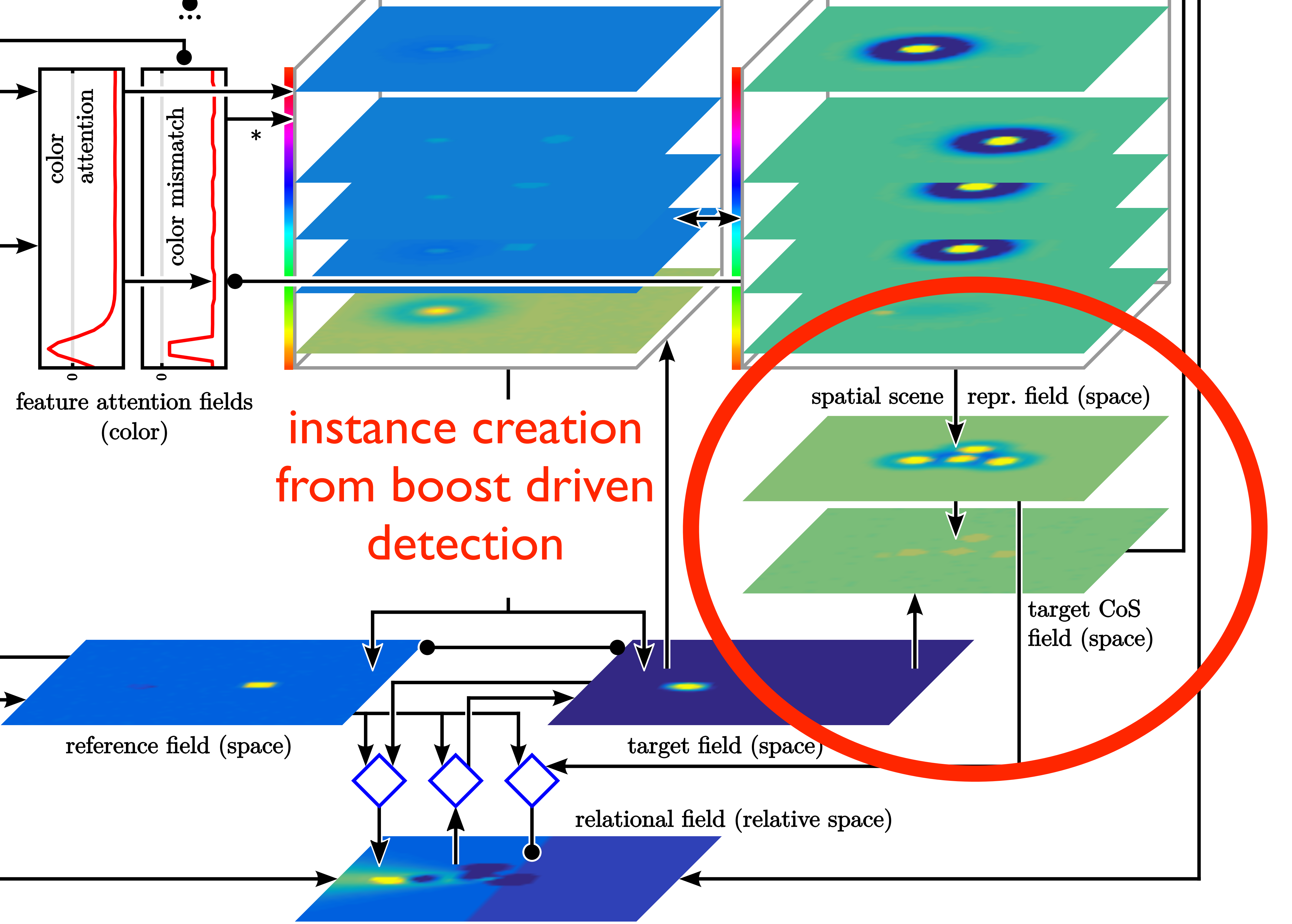
Mental mapping and inference

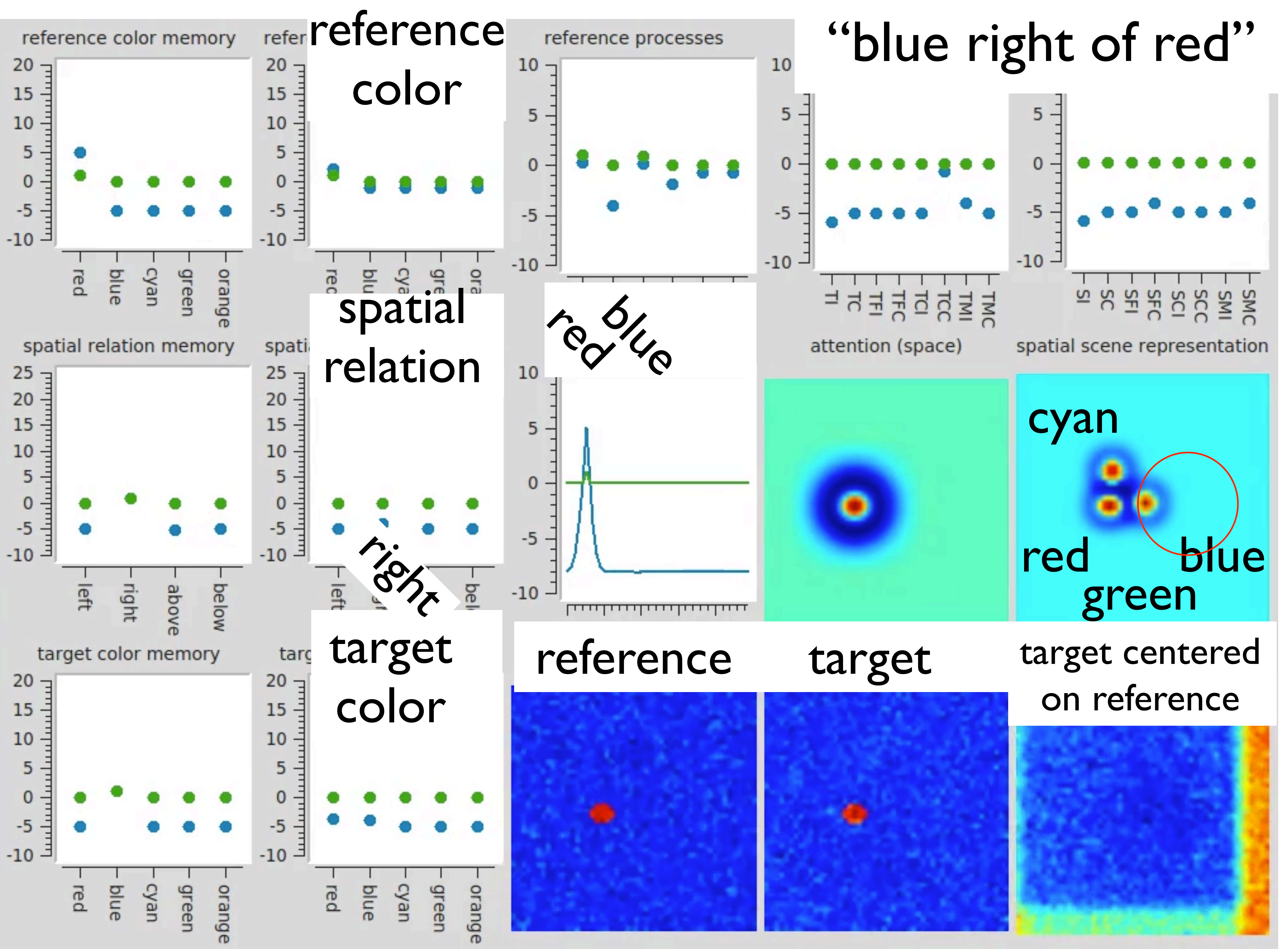
- propositions
  - “There is a cyan object above a green object.”
  - “There is a red object to the left of the green object.”
  - “There is a blue object to the right of the red object.”
  - “There is an orange object to the left of the blue object.”
- inference
  - “Where is the blue object relative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

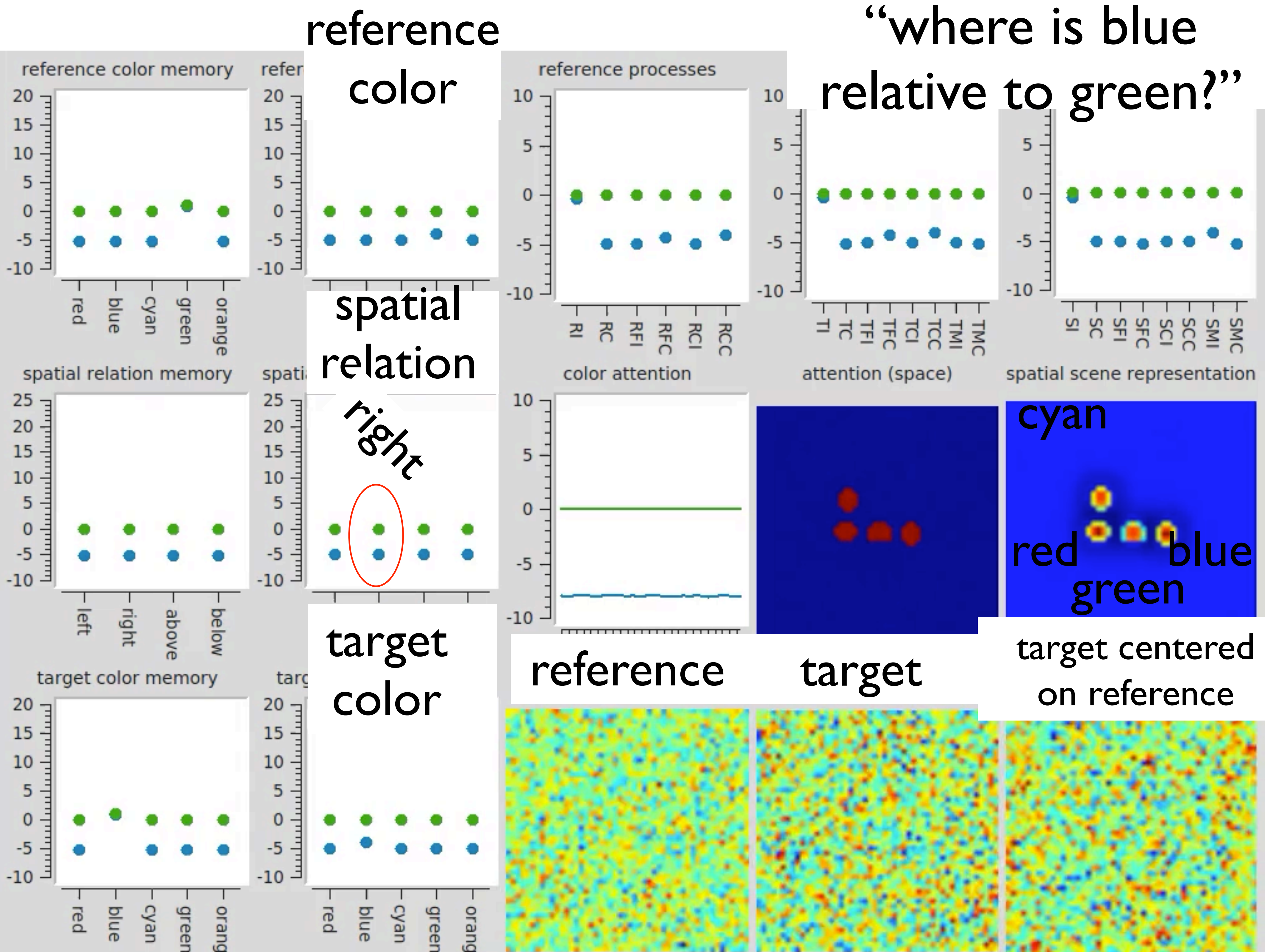
[Ragni, Knauff, *Psych Rev* 2013]



[Kounatidou, Richter, Schöner, CogSci 2018]


color
attention
color mismatch
*
0
0
feature attention fields
(color)
instance creation
from boost driven
detection
spatial scene
repr. field (space)
target CoS
field (space)
reference field (space)
target field (space)
relational field (relative space)


"blue right of red"
reference color memory
reference color
reference processes
spatial relation memory
spatial relation
blue
red
right
attention (space)
spatial scene representation
cyan
red
blue
green
target color memory
target color
reference
target
target centered on reference
red
blue
cyan
green
orange
left
right
above
below
TI
TC
TFI
TFC
TCI
TCC
TMI
TMC
SI
SC
SFI
SFC
SCI
SCC
SMI
SMC


"where is blue relative to green?"
reference color
spatial relation
right
target color
reference
target
target centered on reference
reference color memory
20 15 10 5 0 -5 -10
red blue cyan green orange
reference processes
10 5 0 -5 -10
RI RC RFI RFC RCI RCC
TI TC TFI TFC TCI TCC TMI TMC
SI SC SFI SFC SCI SCC SMI SMC
spatial relation memory
25 20 15 10 5 0 -5 -10
left right above below
color attention
attention (space)
spatial scene representation
cyan
red
blue
green
target color memory

Roadmap

- Neuro-physics
- Neural dynamics
- Recurrent neural dynamics
- Neural fields: dynamics
- Neural fields: dimensions
- Binding
- Sequences
- Coordinate transforms
- Relational concepts, grounding, mental mapping
- Conclusions

Conclusion

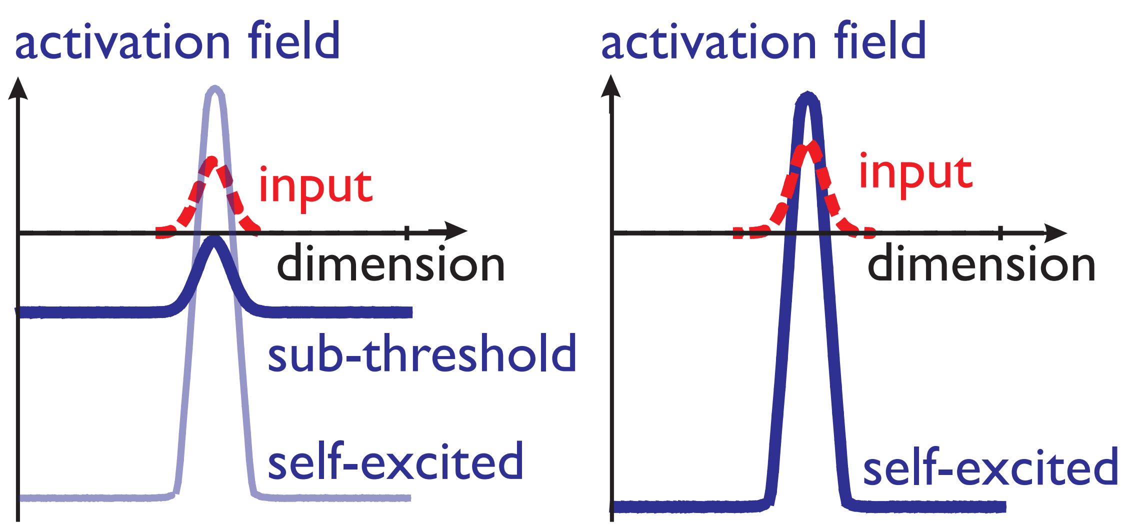
- sensory-motor cognition from neural dynamic fields that are coupled to sensory surfaces and act on the motor surfaces (through behavioral dynamics)
- instabilities make decisions
    - detection
    - selection
    - working memory

Why do neural dynamic architectures work?

- 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?

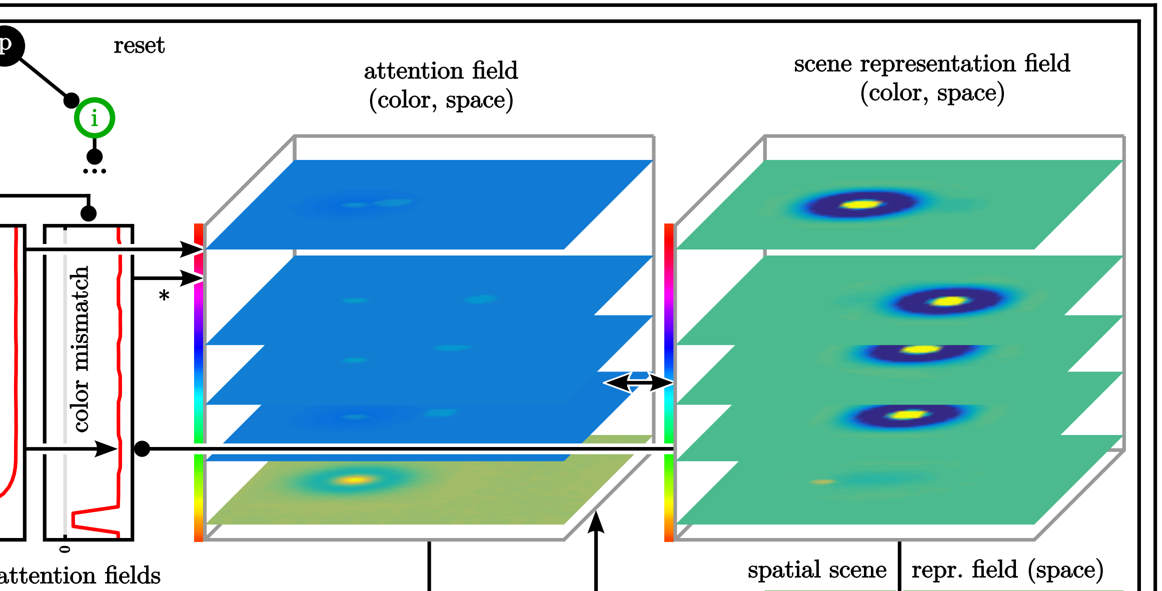
1) Why is the dynamic regime invariant?

- stability => structural stability = invariance of solutions under change of the dynamics
- => dynamic modularity: fields retain their dynamic regime as activation elsewhere varies



2) Why is the content invariant?

- coupling among fields must preserve the fields' dimensions: "non-synesthesia principle"
- informational modularity (encapsulation)



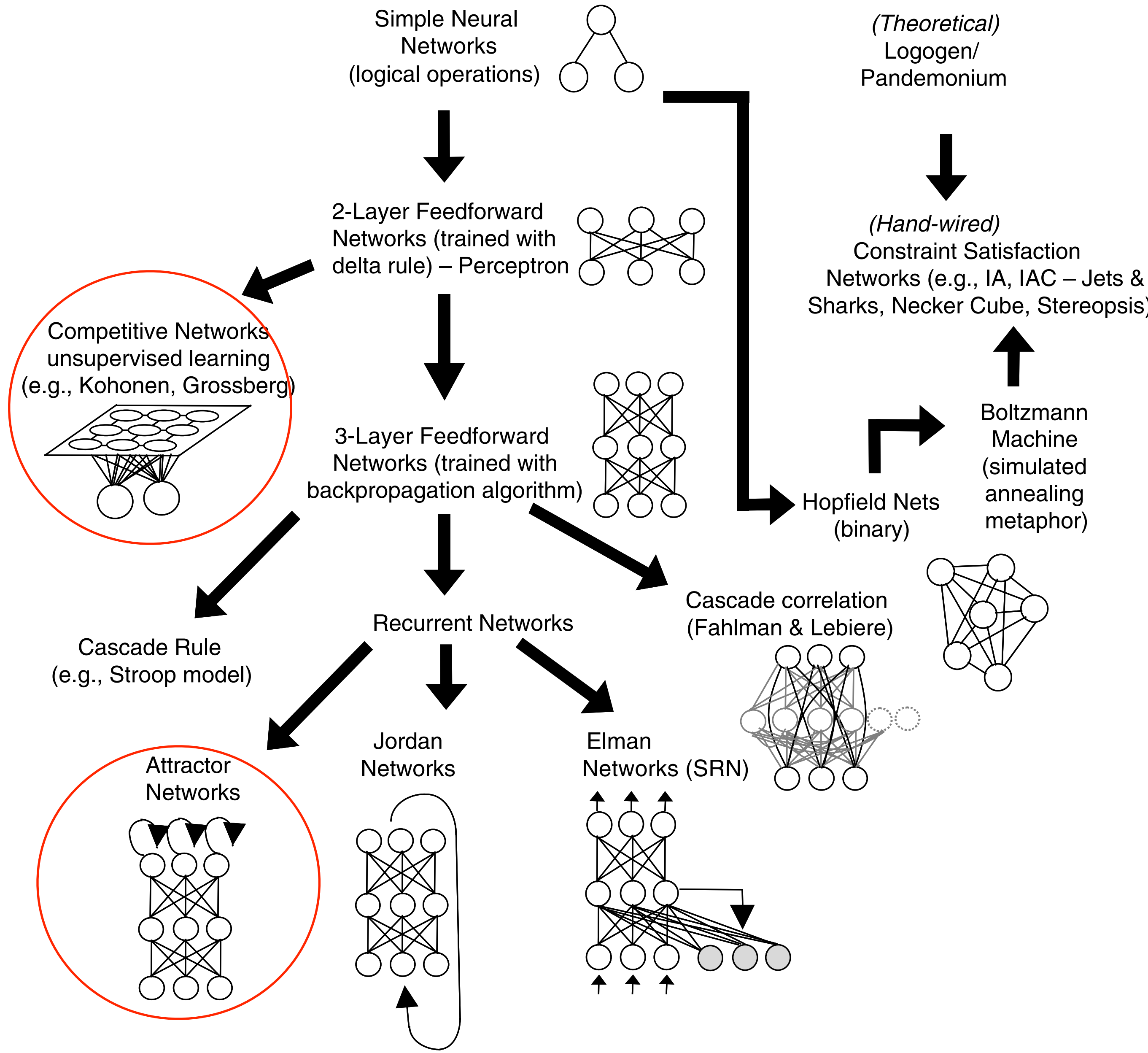
- neural dynamic architectures are specific = constrained by evolution and development

Embodiment hypothesis

- cognition does not necessarily activate motor systems
- cognition inherits the dynamic properties of sensory-motor cognition:
  - continuous state, continuous time, stability ..
  - continuous/intermittent link to the sensory and motor surfaces is *possible*
- => cognition is generated in the specific embodied cognitive architectures that emerged from evolution/development

DFT vs connectionism/NN

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



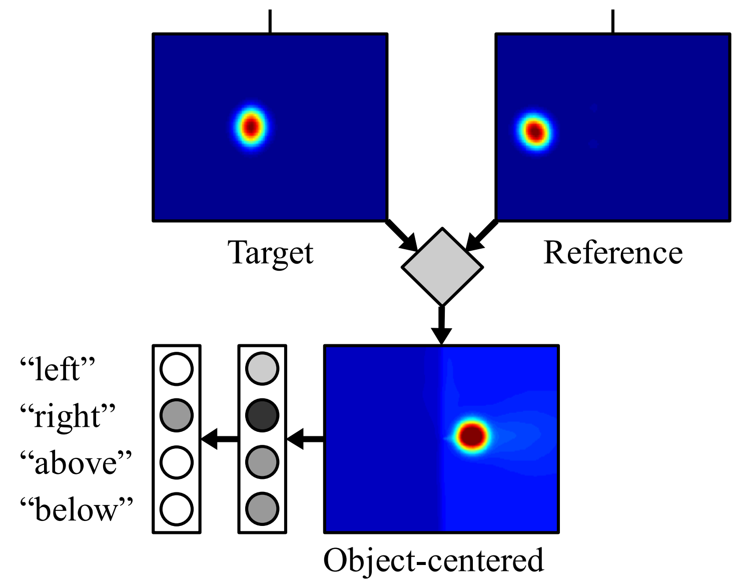
[Thomas, McClelland, 2008]

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 .. sequences
- => all autonomous cognition is based on localist representations
- => all cognitive representations are low-dimensional

DFT as a neural theory for higher cognition

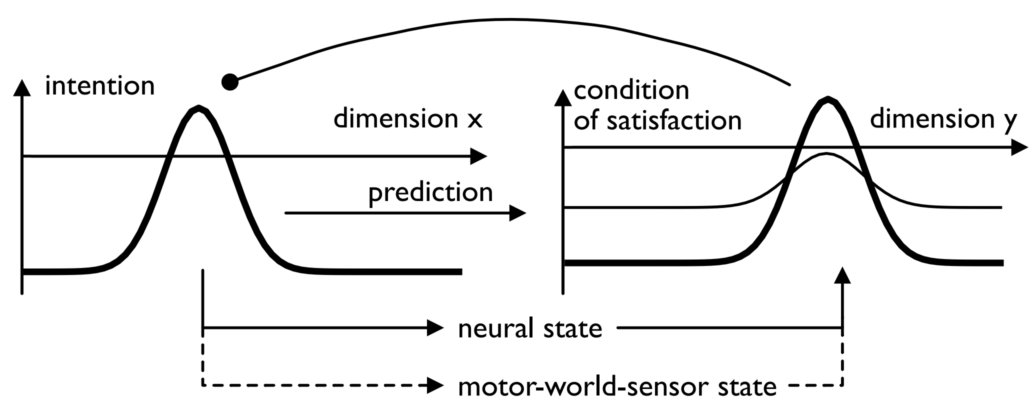
- 1) all concepts are grounded
- 2) attentional selection, coordinate transformation, sequential processing … emulate "function calls"



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

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

- => DFT=neurosymbolics



Job ad

PhD position: Reaching decisions: neural mechanisms underlying learning and development of action decisions

- European mobility requirement: less than 12 months in prior 3 years resident of Germany