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
higher-dimensional neural fields enable new functions

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

![Diagram of visual scene with space-color field and spatial field](image)

**Sebastian Schneegans (INI)**

*Multi-Dimensional Fields*

*December 5, 2013 10 / 37*
Previous lecture

=> scene representation

[ Grieben et al, (Attention, Perception, & Psychophysics, in press)]
higher-dimensional neural fields enable new functions
visual search… attentional selectivity
coordinate transforms…

binding different feature dimensions through space
Previous lecture

- higher-dimensional neural fields enable new functions
  - visual search... attentional selection
  - coordinate transforms...
  - binding different feature dimensions through space
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 our 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)
that process could be mediated by other forms of communication, e.g., pointing (deictic code)
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

- 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 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 to 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

gaze position relative to body

retinal location

body-centered position
Coordinate transformations

- Retinal location
- Gaze position relative to body
- Body-centered location
Coordinate transformations

- predict retinal location following gaze shift

[Schneegans, Schöner, 2012]
DFT approach to applying operators
DFT approach to applying operators

based on convolution of fields with kernels

[Fig. 3] Reference field and spatial semantics alignment. Panel A shows the camera image containing three objects (green toothpaste tube, blue wire roll, red plastic apple). Panel B shows the spatial distribution of the weight strengths for each of the four spatial semantic terms (lighter blue regions indicate greater weight). Panel C shows activation in the two-dimensional reference field. The activation peak (yellow blob) corresponds to the green object location identified as the referent in this example. Panel D depicts the spatial semantic fields with input from the semantic templates (Panel B) aligned with the reference object location (i.e. the light blue region in the "right" spatial semantic field represents region to the right of the green referent object).

Yet to be identified, different solutions are possible [67,55]. In the present work we solve this problem through a spatial template "shift" mechanism (Fig. 1D) which aligns the semantic templates with the position of the reference object. The semantic templates are only allowed to contribute to the spatial semantic field dynamics (see below) after this has occurred.

We implement this "shift" or "alignment" of spatial semantics as a convolution of the output of the reference field, which holds the reference object position, with the semantic template functions. Because the reference object is represented by a localized activation pattern, the convolution centers semantic weights on the reference object location. The shift of the semantic weights can thus be viewed as a modulation of the synaptic connection strength between a spatial term node and the spatial semantic field according to the activation in the reference field. Fig. 3D shows an example of this spatial semantic alignment in which the semantic weights are centered on the location of the green reference object.

2.2.6 Spatial semantic fields

For the system to process spatial language about the visual scene, spatial information about the target object and the aligned spatial templates must be integrated. In our model, the spatial semantic fields provide this function (Fig. 1E). Spatial semantic fields are neural arrays with weak dynamical field interactions (see parameter values in the Attachment). Each spatial semantic field is associated with one spatial semantic template. Each spatial semantic field therefore represents a single spatial relation ("left", "right", "above", or "below" in the present implementation), Fig. 3D.

The spatial semantic fields each receive activation from the color-space fields which specify the target location. By blending this target location information from the color-space fields with the aligned semantic weights, the spatial semantic fields integrate the target and spatial term information, thereby linking spatial term knowledge to the visual scene. In addition, each spatial semantic field is also reciprocally linked to a categorical spatial-term node, analogous to the color-term nodes [from: Lipinski, Sandamirskaya, Schöner, 2009].
A cognitive architecture for grounded spatial language in DFT

[Lipinski et al: JEP:LMC (2011)]
Spatial comparison in DFT

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

[Lipinski et al: JEP:LMC (2011)]
Spatial comparison in DFT

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

[Lipinski et al: JEP:LMC (2011)]
Spatial comparison in DFT

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

[Lipinski et al: JEP:LMC (2011)]
Spatial comparison in DFT

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

[Lipinski et al: JEP:LMC (2011)]
Spatial comparison in DFT

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

[Lipinski et al: JEP:LMC (2011)]
“where is the green object relative to the red object?”

[Lipinski et al: JEP:LMC (2011)]
“which object is above the blue object?”

[Lipinski et al: JEP:LMC (2011)]
“where is the green object?”
Spatial comparison in DFT accounts for human data

[ Carlson & Hill, 2008
Simulations

[Lipinski et al: JEP:LMC (2011)]
A DFT architecture that does both grounding and describing

![Diagram of processes of grounding and describing](image)

- **Grounding** (left column, orange arrows): Initially, a color concept is active (e.g., from language-related processes) and drives visual attention to a matching object in the scene. The process involves:
  1. Activating a 'red' node.
  2. The grounded concept.
  3. The visual scene.

- **Describing** (right column, violet arrows): An object is initially attended (e.g., based on salience) and drives the activation of a matching color concept. The process involves:
  1. The description.
  2. Continuous activation field (visual space).

In both processes, the activation values above and below the threshold (gray line) denote active and inactive nodes, respectively.
A neural dynamics resolves spatial language about visual scenes.

Fig. 2. Overview of the architecture, showing the activation state when answering the question "What is to the right of the green object?" on the scene in Fig. 1. On the right, dynamic fields are shown as color-coded activation patterns (blue for lowest, red for highest activation). On the left, dynamic nodes are denoted as circles with activation levels indicated by fill color opacity. The three-dimensional perceptual field is shown as slices through the activation pattern for the colors orange and green. Excitatory synaptic connections are denoted by arrows, inhibitory connections by lines ending in circles. Arrows marked with stars are patterned connections that encode concepts.

2 Methods

The DFT architecture shown in Fig. 2 can be viewed as one integrated dynamical system, that combines coupled dynamics fields (DFs) supporting perception with coupled dynamic nodes that instantiate concepts and organize sequential processing.

2.1 Dynamic fields and dynamic nodes

DFs can be thought of as a temporally and spatially continuous form of neural networks. Activation fields, $u(x,t)$, over a continuous feature dimension $x$ (e.g., hue or spatial position) evolve over time $t$ according to

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

where $\tau$ is a time constant, $h<0$ is a resting level, and $S(x,t)$ is external input.

Lateral interactions in the field are homogeneous and can be described as a

[Richter, Lins et al. ICANN 2014]
A neural dynamics resolves spatial language about visual scenes.

Fig. 2. Overview of the architecture, showing the activation state when answering the question "What is to the right of the green object?" on the scene in Fig. 1. On the right, dynamic fields are shown as color-coded activation patterns (blue for lowest, red for highest activation). On the left, dynamic nodes are denoted as circles with activation levels indicated by fill color opacity. The three-dimensional perceptual field is shown as slices through the activation pattern for the colors orange and green. Excitatory synaptic connections are denoted by arrows, inhibitory connections by lines ending in circles. Arrows marked with stars are patterned connections that encode concepts.

2 Methods

The DFT architecture shown in Fig. 2 can be viewed as one integrated dynamical system, that combines coupled dynamics fields (DFs) supporting perception with coupled dynamic nodes that instantiate concepts and organize sequential processing.

2.1 Dynamic fields and dynamic nodes

DFs can be thought of as a temporally and spatially continuous form of neural networks. Activation fields, \( u(x, t) \), over a continuous feature dimension (e.g., hue or spatial position) evolve over time according to

\[
\tau \dot{u}(x, t) = -u(x, t) + h + S(x, t) + \int f(u(x', t)) w(x - x') dx',
\]

where \( \tau \) is a time constant, \( h < 0 \) is a resting level, and \( S(x, t) \) is external input.

[Richter, Lins et al. ICANN 2014]
A neural dynamics resolves spatial language about visual scenes.

Fig. 2. Overview of the architecture, showing the activation state when answering the question "What is to the right of the green object?" on the scene in Fig. 1. On the right, dynamic fields are shown as color-coded activation patterns (blue for lowest, red for highest activation). On the left, dynamic nodes are denoted as circles with activation levels indicated by fill color opacity. The three-dimensional perceptual field is shown as slices through the activation pattern for the colors orange and green. Excitatory synaptic connections are denoted by arrows, inhibitory connections by lines ending in circles. Arrows marked with stars are patterned connections that encode concepts.

2 Methods

The DFT architecture shown in Fig. 2 can be viewed as one integrated dynamical system, that combines coupled dynamics fields (DFs) supporting perception with coupled dynamic nodes that instantiate concepts and organize sequential processing.

2.1 Dynamic fields and dynamic nodes

DFs can be thought of as a temporally and spatially continuous form of neural networks. Activation fields, $u(x, t)$, over a continuous feature dimension $x$ (e.g., hue or spatial position) evolve over time $t$ according to

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

where $\tau$ is a time constant, $h < 0$ is a resting level, and $S(x, t)$ is external input.

Lateral interactions in the field are homogeneous and can be described as a reference-target-grammatical-roles [Richter, Lins et al. ICANN 2014]
Fig. 2. Overview of the architecture, showing the activation state when answering the question "What is to the right of the green object?" on the scene in Fig. 1. On the right, dynamic fields are shown as color-coded activation patterns (blue for lowest, red for highest activation). On the left, dynamic nodes are denoted as circles with activation levels indicated by fill color opacity. The three-dimensional perceptual field is shown as slices through the activation pattern for the colors orange and green. Excitatory synaptic connections are denoted by arrows, inhibitory connections by lines ending in circles. Arrows marked with stars are patterned connections that encode concepts.

2 Methods

The DFT architecture shown in Fig. 2 can be viewed as one integrated dynamical system, that combines coupled dynamics fields (DFs) supporting perception with coupled dynamic nodes that instantiate concepts and organize sequential processing.

2.1 Dynamic fields and dynamic nodes

DFs can be thought of as a temporally and spatially continuous form of neural networks. Activation fields, $u(x, t)$ over a continuous feature dimension $x$ (e.g., hue or spatial position) evolve over time $t$ according to

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

where $\tau$ is a time constant, $h < 0$ is a resting level, and $S(x, t)$ is external input. Lateral interactions in the field are homogeneous and can be described as a color grounding reference target spatial relationship grounding grammatical roles spatial relationship grounding.
Fig. 2. Overview of the architecture, showing the activation state when answering the question “What is to the right of the green object?” on the scene in Fig. 1. On the right, dynamic fields are shown as color-coded activation patterns (blue for lowest, red for highest activation). On the left, dynamic nodes are denoted as circles with activation levels indicated by fill color opacity. The three-dimensional perceptual field is shown as slices through the activation pattern for the colors orange and green. Excitatory synaptic connections are denoted by arrows, inhibitory connections by lines ending in circles. Arrows marked with stars are patterned connections that encode concepts.

2 Methods

The DFT architecture shown in Fig. 2 can be viewed as one integrated dynamical system, that combines coupled dynamics fields (DFs) supporting perception with coupled dynamic nodes that instantiate concepts and organize sequential processing.

2.1 Dynamic fields and dynamic nodes

DFs can be thought of as a temporally and spatially continuous form of neural networks. Activation fields, $u(x, t)$, over a continuous feature dimension $x$ (e.g., hue or spatial position) evolve over time $t$ according to

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

where $\tau$ is a time constant, $h < 0$ is a resting level, and $S(x, t)$ is external input. Lateral interactions in the field are homogeneous and can be described as a...
Methods

2.1 Dynamic fields and dynamic nodes

DFs can be thought of as a temporally and spatially continuous form of neural networks. Activation fields, \( u(x, t) \), over a continuous feature dimension (e.g., hue or spatial position) evolve over time according to

\[
\tau \frac{\dot{u}(x, t)}{u(x, t)} = -u(x, t) + h + S(x, t) + \int f(u(x', t))w(x - x') dx',
\]

where \( \tau \) is a time constant, \( h < 0 \) is a resting level, and \( S(x, t) \) is external input.

Lateral interactions in the field are homogeneous and can be described as a

[Ref: 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 red cup that is to the left of the green cup”

[Richter, Lins et al, CogSci 2014]
Autonomous Neural Dynamics to Test Hypotheses in a Model of Spatial Language

In this paper we provide a fully autonomous neural dynamic architecture that transforms the location of the target object into a target response. As in the previous scenario, the spatial template is instantiated and the potential reference objects are brought to the attentional foreground, performing spatial transformations, and activating spatial terms unfold autonomously in the context of the processing steps needed to resolve relational spatial phrases. We discuss how this neural dynamic architecture represents perceptual information in activation fields that make detection and selection decisions through neural interaction. Activation nodes and their connectivity to the perceptual fields represent concepts. Dynamic interactions among the fields allow the architecture to generate behavioral sequences that are relevant for executing the processing step. The architecture can autonomously solve relational spatial phrases. We demonstrate how the neural architecture can correspond with neural principles (Lipinski, Schneegans, Sadler, 1996). While such discrete processing steps appear natural in information processing terms, they require an explicit match of the target object with the spatial term (Logan & Carr, 1994). Selecting the reference and target objects can be seen as an attempt to identify the correct target candidate left of the chosen reference object. This signals that target selection has failed. The target candidates field and the reference field are inhibited, so that peaks in these fields vanish. The relational CoS field, since none of the target candidates is to the right of the chosen reference object, remains active. At time $t_4$, the green object in the top right is established as a condition of satisfaction. At $t_5$, the positions of the two red objects have been determined by the reference object. At $t_6$, the frame shift and fed into both relational fields (CoS and CoD). The flow of activation is determined by the structure of the processing steps to interpret and generate relational spatial language. Within the framework of Dynamic Field Theory (DFT; Schneegans & Danilovskaya, 2012), we take inspiration from earlier work on the autonomous generation of behavioral sequences through neural interaction in neural systems. At the population level that is coherent with neural principles (Lipinski, Schneegans, Sadler, 1996). This work is based on the framework of Dynamic Field Theory (DFT; Schneegans, Danilovskaya, Spencer, & Schoner, 2012). For instance, I may use language to direct your attention to an object in a visual scene. When several similar environments are shown in sequence, it may be more efficient to communicate a sequence of relational phrases. These include bringing visual objects into the attentional foreground, performing spatial transformations, and activating spatial terms.
A typical relational phrase like the one above consists of a relational term (to the left) and objects to which a relational term (to the left) is applied. Interpretation of such a phrase may require that different pairs of objects are visible such as in Fig. 1a, using object identification. Language enables humans to communicate about shared experiences that is to the right of the red cup, resolves ambiguity in such situations. Even in cases where a single object is visible, it can be ambiguous as to which of two potential objects is meant, for example, the red cup to the left of the green cup. This suggests that this happens in sequence rather than in parallel (Logan, 1994). Selecting the reference and target objects of such a pair also appears to happen sequentially. This is suggested by characteristic shifts of attention found using behavioral cuing (Roth & Franconeri, 2012), in which activation peaks are units of information processing.

“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 processing steps involved in interpreting a relational spatial phrase can be implemented in a neural system: the neural representation of an object involves an activation field that captures its location and an intention field that encodes the subject's intention to act on this object. This dual representation allows for the activation of the corresponding intention nodes and the facilitating of the processing step:

1. At $t_1$, the green object in the top right is established as the reference object. This is indicated by the changing activation states of the reference and target fields. 
2. At $t_2$, the target object is selected. This selection is driven by the activation peak in the target response field, which is also seen in the relational CoS field. 
3. At $t_3$, the target object is brought into the attentional foreground, performing spatial transformations, and restoring the memory of the previously selected reference object. 
4. At $t_4$, the spatial term is mapped onto this reference frame, and the activation node is reassigned to this reference object. 
5. At $t_5$, the green object to the left of the chosen reference object is established as the target object. This is indicated by the activation peak in the target response field, which also appears in the relational CoD field. 
6. At $t_6$, the successful completion of the task is indicated by the activation peak in the target response field, with a frame shift and feed into both relational fields (CoS and CoD).
“the red to the left of the green”
Grounding movement relations
The core of the attentional system consists of two three-dimensional attention fields, which are defined over the same dimensions as the two perception fields, but their activation remains below threshold unless additional input arrives from a feature attention field or a spatial attention field.

Fig. 2. Architecture with activation snapshots while it is generating a phrase about a video. Fields are shown as color-coded activation patterns; for three-dimensional fields, two-dimensional slices are shown. Node activation is denoted in opacity-coded circles. Spatial templates are illustrated as color-coded weight patterns (bottom left). Excitatory synaptic connections are denoted by lines with arrowheads, inhibitory connections by lines ending in circles. Transformations to and from polar coordinates are marked with a “T.” Steerable neural mappings are denoted as diamonds.

[Richter, Lins, Schöner, ToPiC (2017)]
into the reference and target field and enable these fields to track moving objects even if spatial attention is currently focused elsewhere.

3.2. Attention

The core of the attentional system consists of two three-dimensional attention fields. They are defined over the same dimensions as the two perception fields, but their activation remains below threshold unless additional input arrives from a feature attention field or a spatial attention field.

Fig. 2. Architecture with activation snapshots while it is generating a phrase about a video. Fields are shown as color-coded activation patterns; for three-dimensional fields, two-dimensional slices are shown. Node activation is denoted in opacity-coded circles. Spatial templates are illustrated as color-coded weight patterns (bottom left). Excitatory synaptic connections are denoted by lines with arrowheads, inhibitory connections by lines ending in circles. Transformations to and from polar coordinates are marked with a “T.” Steerable neural mappings are denoted as diamonds.

“the red moving to the right”

perceptual grounding

[Richter et al]
Figure 6: Generating a description of an object using a movement relation with respect to a second object. See Figure 4 for details.

[Richter et al]
Mental mapping and inference

- making sense of propositions (about spatial relations) purely mentally, without any perception to ground in
- and operating on such “sense” by drawing inferences…
Mental mapping and inference

- mental map formation from 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?”
Figure 1: Activation snapshot of the architecture as it forms a mental model consisting of five objects. For two-dimensional fields, activation is shown color-coded, where blue colors denote subtreshold and yellow colors denote suprathreshold activation. For three-dimensional fields, two-dimensional slices of activation are shown. Neural nodes are denoted by circles that are filled if the node is active and empty if inactive. Excitatory synaptic connections are shown by black lines with arrowheads, inhibitory connections by lines ending in black circles; patterned connections are marked with a star. Steerable neural mappings are denoted by blue diamonds. See text for details.

...
Figure 1: Activation snapshot of the architecture as it forms a mental model consisting of five objects. For two-dimensional fields, activation is shown color-coded, where blue colors denote subtreshold and yellow colors denote suprathreshold activation. For three-dimensional fields, two-dimensional slices of activation are shown. Neural nodes are denoted by circles that are filled if the node is active and empty if inactive. Excitatory synaptic connections are shown by black lines with arrowheads, inhibitory connections by lines ending in black circles; patterned connections are marked with a star. Steerable neural mappings are denoted by blue diamonds. See text for details.

The premise given by the participant is, "The orange object to the left of the blue object" (shown in Figure 1) consists of three elements, all of which need to be represented by the architecture: the object the premise is primarily referring to (the target object, here orange), the spatial relation (here, to the left of), and the object which the relation uses as a reference position (the reference object, here blue). The spatial transformation system represents these three elements in dedicated dynamic neural fields, the target field, the relational field, and the reference field, respectively. The target field and reference field are defined over two-dimensional space and receive input from the attention field. Whenever there is a peak in the attention field, one of the fields may be brought into the dynamic regime to form peaks. The two-dimensional relational field represents the relative position of a target object with respect to the reference object. The field is defined such that the reference object would be in the center of the field. The relational field also receives input from the production nodes of all spatial relation concepts (e.g., TO THE LEFT OF, see Figure 1). Coordinate transformations between the absolute spatial positions in the target field and the relative positions in the relational field are based on steerable neural mappings (blue diamonds in Figure 1; Schneegans & Schöner, 2012), which are approximated by convolutions here. The architecture has three such coordinate transforms: the first (leftmost blue diamond) enables the position of an already existing target object to be transformed into the relational field. This enables the architecture to make inferences on an already established mental model. The second coordinate transform (middle diamond) enables the model to transform peaks in the relational field back into the target field. This path accounts for the creation of new objects in the scene: a peak is induced in the relational field from the spatial template that represents one of the spatial relations. The position in space where the peak forms determines where the new object is going to be placed in space. The third transformation (right diamond) has a crucial impact on the position where the peak forms in the relational field. It transforms the output of the spatial scene representation field and feeds inhibitorily into the relational field, introducing inhibition in positions that are already occupied by objects in the mental model. Due to this inhibition, peaks induced in the relational field tend to shift further outward, avoiding changes to the already established mental model. This is consistent with the preferred mental models that humans tend to build (Ragni & Knauff, 2013).
"blue right of red"
“where is blue relative to green?”

spatial relation memory

right

target color memory

target color

target centered on reference

reference color

reference processes

color attention

attention (space)

spatial scene representation

red blue cyan green
Conclusion

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