TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 9 (2017) 35–47

ISSN:1756-8757 print / 1756-8765 online
DOI: 10.1111/tops.12240

This article is part of the topic "Best of Papers from the Cognitive Science Society Annual Conference," Wayne D. Gray (Topic Editor). For a full listing of topic papers, see: http://onlinelibrary.wiley.com/doi/10.1111/tops.2017.9.issue-1/issuetoc.

A Neural Dynamic Model Generates Descriptions of Object-Oriented Actions

Mathis Richter, Jonas Lins, Gregor Schöner

Institut für Neuroinformatik, Ruhr-Universität Bochum

Received 7 October 2016; accepted 19 October 2016

Abstract

Describing actions entails that relations between objects are discovered. A pervasively neural account of this process requires that fundamental problems are solved: the neural pointer problem, the binding problem, and the problem of generating discrete processing steps from time-continuous neural processes. We present a prototypical solution to these problems in a neural dynamic model that comprises dynamic neural fields holding representations close to sensorimotor surfaces as well as dynamic neural nodes holding discrete, language-like representations. Making the connection between these two types of representations enables the model to describe actions as well as to perceptually ground movement phrases—all based on real visual input. We demonstrate how the dynamic neural processes autonomously generate the processing steps required to describe or ground object-oriented actions. By solving the fundamental problems of neural pointing, binding, and emergent discrete processing, the model may be a first but critical step toward a systematic neural processing account of higher cognition.

Keywords: Relations; Neural process model; Action parsing; Dynamic field theory; Grounded cognition; Image schemas

Correspondence should be sent to Mathis Richter, Institut für Neuroinformatik, Ruhr-Universität Bochum, 44870 Bochum, Germany. E-mail: mathis.richter@ini.rub.de

1. Introduction

If you were to describe the arrangement of furniture in your office, you would probably make use of the spatial relations between different items. You may recognize without effort that "the bookshelf is to the left of the desk" although this relationship is not directly specified by perception and requires active construal. The same holds for spatio-temporal relations. If you were to describe, for instance, that "the dog is running toward the ball," you would have to extract that relationship from the position and movement direction of the dog and the position of the ball. This kind of relational processing is ubiquitous in daily life and may in fact lay the foundation for higher cognition (Halford, Wilson, & Phillips, 2010; Knauff, 2013).

In this paper, we present a neural process account of how such relations are discovered in visual scenes. All processes and representations in the model are captured by dynamic neural networks. The model can describe simple scenes in terms of spatial relations (e.g., "the red object is to the left of the green object") and object-oriented actions (e.g., "the red object is moving toward the green object"). It can conversely select objects in a scene that are designated by a relational phrase. This model of relational processing represents a key step in a research program with the ultimate goal of constructing a pervasively neural process account of higher cognition (Lipinski, Schneegans, Sandamirskaya, Spencer, & Schöner, 2012; Lobato, Sandamirskaya, Richter, & Schöner, 2015; Richter, Lins, Schneegans, Sandamirskaya, & Schöner, 2014; van Hengel, Sandamirskaya, Schneegans, & Schöner, 2012).

We employ dynamic field theory (DFT; Schöner, Spencer, & the DFT Research Group, 2015) as a theoretical framework. DFT describes neural population activity by activation fields that are defined over metric feature dimensions and evolve continuously in time through a neural dynamics. By using only the dynamics from the DFT repertoire, we arrive at a seamless process account that is pervasively neural. While the fields capture representations in a modal form close to the sensorimotor surfaces, neural nodes sharing the same dynamics enable modeling discrete, amodal[1] representations. Mutual coupling between fields and nodes allows for interaction between these two kinds of representations. The role such interaction may play in cognition has been discussed extensively in recent years and is broadly referred to as the *grounding* of amodal concepts (or linguistic forms) in the sensorimotor world (Barsalou, 2008; Crocker, Knoeferle, & Mayberry, 2010; Zwaan, 2014). In a neural dynamics perspective, neural activation is linked to the world continuously in time, making it necessary to specify not only substrates and connection patterns, but also the processes that establish links between amodal representations and perceptual objects while allowing to flexibly switch between links. Here, we differentiate between the process of *perceptual grounding*, which links from an (amodal) concept to an object in a scene, and the process of *describing*, which activates a concept based on an object in a scene. Fig. 1 illustrates these two neural processes schematically. In the top row, the activation state of neural nodes is illustrated. These nodes represent color concepts (i.e., red, green, yellow, and blue). The activation level of the node representing "red" is positive, which means that the concept of "red" has been activated. In the middle, a field of neural activation defined over the two-dimensional visual array represents spatial attention. A localized peak of activation (yellow

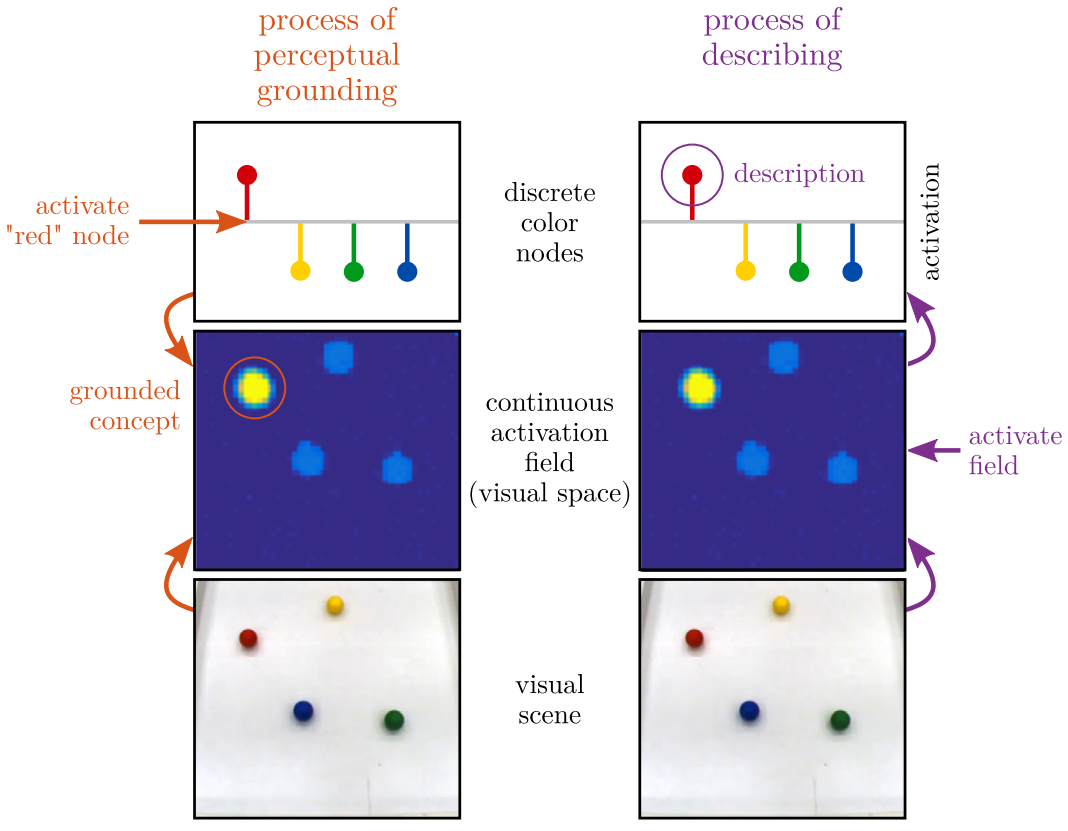

Fig. 1. Schematic illustration of the processes of *grounding* (left column, orange arrows) and *describing* (right column, violet arrows). In the top row, activation values above and below the threshold (gray line) denote active and inactive nodes, respectively. In the middle, the activation of the two-dimensional field is illustrated using a color-map: blue areas are below threshold; yellow areas are above threshold.

circle in the top left in the array) reflects the attentional selection of an object (the red ball in the visual scene shown at the bottom). How spatial attention is guided by the color nodes will be explained later. In perceptual grounding (left column), a color concept is initially active (e.g., from language related processes) and drives visual attention to a matching object in the scene. In describing (right column), an object is initially attended (e.g., based on salience) and drives the activation of a matching color concept.

Lifting such notions to relations, such as the initial example of "the bookshelf is to the left of the desk," requires that a set of coordinated processing steps (Logan & Sadler, 1996) be realized neurally: (a) binding each object to a role (here, the desk is the *reference object*, the bookshelf is the *target object*); (b) centering the reference frame on the reference object; and (c) applying a relational operator (here, "to the left of") to the target object in that frame.

A neural process implementation of these steps requires that the following problems be solved; they reflect fundamental constraints of neural processing that must be faced in neural accounts of higher cognition.

First, information represented by neural activity cannot be freely moved within and between neural populations, because neural connectivity is fixed. In visual cortex, for instance, visual objects are represented in neural maps. Applying a neural operator to a location or an object in such a map is possible only if it is connected to that location. Connecting operators to every location in a map would require unrealistic neural resources. The alternative is to connect the operator to only one default region, a virtual fovea, and shift the representations of objects to that region. This is analogous to the concept of an attentional neural pointer of Ballard, Hayhoe, Pook, and Rao (1997) and is achieved in our framework by steerable neural mappings (Schneegans & Schöner, 2012).

Second, for similar reasons of limiting the required neural resources, the nervous system represents high-dimensional visual information in multiple low-dimensional neural feature maps, in particular in the early tiers of the cortical hierarchy. To refer to any particular object, corresponding representational pieces must be bound together. In a neural implementation of the classical idea of binding through space (Treisman & Gelade, 1980), we endow every feature map with a spatial dimension shared across maps and process objects sequentially in time (Schneegans, Spencer, & Schöner, 2015).

Third, the discrete processing steps this implies and that are critical to all of higher cognition are natural in information processing accounts but hard to achieve in neural process models, in which neural activation evolves continuously in time under the influence of input and recurrent connectivity. In our model, discrete events emerge from continuous neural dynamics through dynamic instabilities, at which the match between neural representations of *intentional states* and their *conditions of satisfaction* are detected (Sandamirskaya & Schöner, 2010).

Finally, the problem of preserving role-filler binding (Doumas & Hummel, 2012) at the interface between the modal and the amodal representations is also solved by sequential processing.

In this paper, we outline a neural dynamic approach that solves these problems and present a prototypical architecture that can ground relational phrases as well as generate such phrases based on video input.

2. Methods

Dynamic field theory describes processes that characterize neural activity at the population level. Models in DFT are based on activation patterns defined as dynamic fields, $u(x, t)$, over continuous feature dimensions, x, (e.g., color or space). These activation patterns evolve in time, t, under the influence of lateral interactions and external input based on the following integro-differential equation

$$\tau \dot{u}(x,t) = -u(x,t) + h + s(x,t) + \int g(u(x',t))w(x-x')\,dx'. \qquad 1$$

Here, the activation's rate of change, $\dot{u}(x, t)$, depends on $u(x, t)$ itself, on a time constant, τ, a negative resting level, h, and external input, $s(x, t)$, from sensors or other fields. Lateral

interaction is determined by convolving the output of the field, $g(u(x, t))$, a sigmoid function with threshold at zero, with an interaction kernel, $w(\Delta x)$. The kernel combines local excitation and surround inhibition along the field's feature dimension.

When presented with localized input above the output threshold, lateral interaction leads to an instability, in which a subthreshold solution becomes unstable and the field moves to a new attractor, a self-stabilized activation peak. From such instabilities, neural events emerge at discrete times from the time-continuous dynamics of the fields. These events are critical for organizing sequential processes in DFT models.

Depending on the tuning of their interaction kernel, dynamic fields may either support multiple peaks or may be selective and only create a single peak that suppresses all others. Fields may also be tuned to hold self-sustained peaks that remain even after input is removed. Fields can be defined over single or multiple dimensions. Dynamic nodes share the fields' dynamic characteristics but do not span a feature dimension. Instead, they represent the "on" or "off" state of discrete elements within an architecture.

Dynamic field theory architectures consist of multiple fields and nodes that are interconnected, where the output of one field is input to another field. Fields of different dimensionalities may be connected along the shared feature dimensions.

3. Architecture

The DFT architecture shown in Fig. 2 can deal with two types of tasks. First, it can ground a language-like phrase such as "the red object moving toward the yellow object"; that is, it can find the objects in the scene that correspond to the phrase. Second, it can generate a phrase such as the one above from observing a video. Solving these tasks within a single neural architecture requires integrating various components, which we describe in more detail now.

3.1. Perception

The architecture receives video input from a camera or video file. This input feeds into two three-dimensional perception fields (top right of Fig. 2) that hold a representation of the scene. Both fields share the spatial dimensions of the camera image but the *perception color field* represents the color of objects in the scene and the *perception movement field* represents their movement direction. To create the input to the perception fields, each video frame goes through several preprocessing steps. For the color field, the preprocessing is first based on generic image processing algorithms. After these, activation is generated that scales with the color saturation of objects in the scene. For the movement field, the preprocessing consists of a neural dynamic implementation of the counterchange model of motion perception (Berger, Faubel, Norman, Hock, & Schöner, 2012). Both perception fields always have stable peaks of activation when there are colored or moving objects in the scene. They project activation into the spatial attention fields along the two spatial dimensions and act as a saliency mechanism. They also project directly

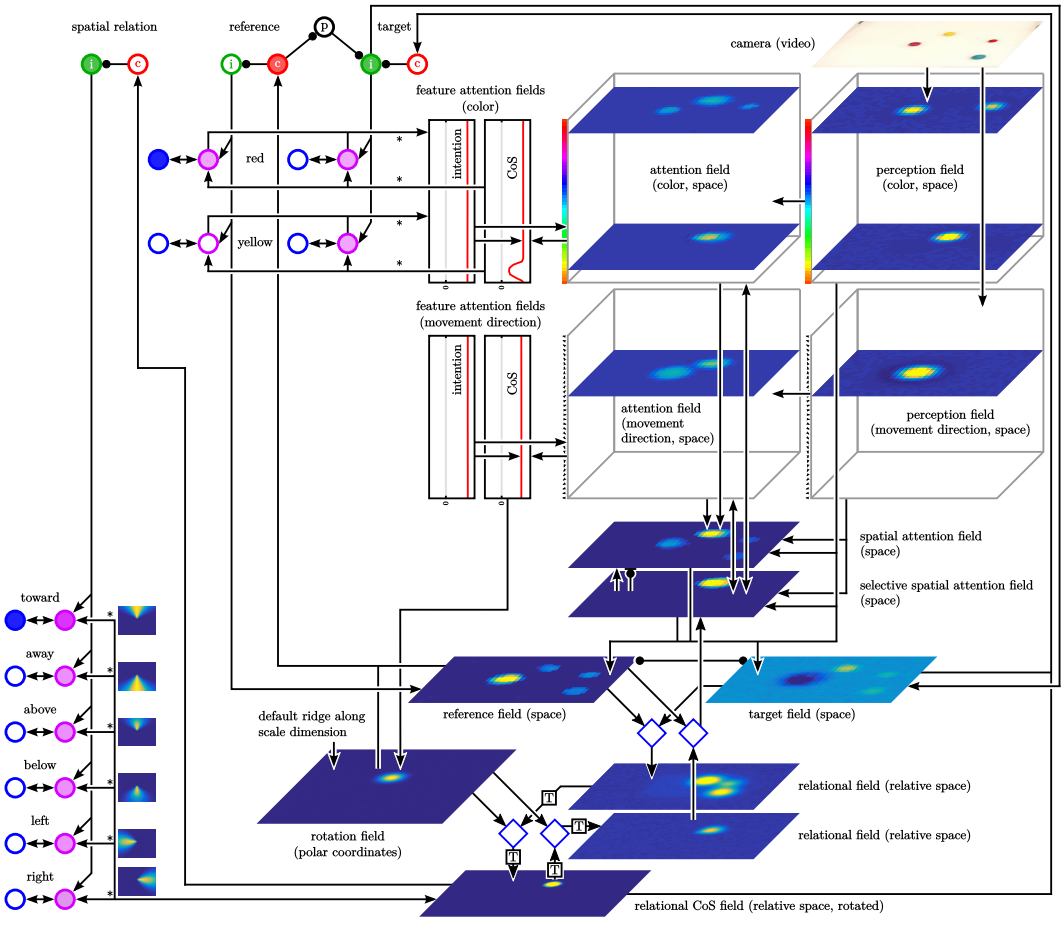

Fig. 2. Architecture with activation snapshots while it is generating a phrase about a video. Fields are shown as color-coded activation patterns; for three-dimensional fields, two-dimensional slices are shown. Node activation is denoted in opacity-coded circles. Spatial templates are illustrated as 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.

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.

A pair of one-dimensional fields spans each feature dimension (color and movement direction): The *intention field* represents feature values for guided search and impacts on the three-dimensional attention fields; the *condition of satisfaction (CoS) field* matches input from the attention fields against what is represented in the intention field.

Two *spatial attention fields* are defined over the two spatial dimensions of the camera image. One field allows for multiple simultaneous peaks and projects into the reference and target fields. The other only allows for a single peak; it can be boosted to induce a selection decision on multiple candidate objects. A peak generated in this spatial attention field suppresses activation at all other locations in the other spatial attention field. It further projects into the three-dimensional attention fields, enabling peaks to form there that represent the feature values at the selected location (which then impact on the CoS fields). This implements a neural mechanism of feature binding across space (Schneegans et al., 2015).

3.3. Steerable neural mappings

The two-dimensional *reference field* and *target field* each represent the spatial position of their respective objects. The target field projects into the *relational field* via a steerable neural mapping (upper left blue diamond in Fig. 2) that shifts the representation of the target objects so that it is centered on the reference object. This transformation to a new reference frame is implemented as a convolution for performance reasons.

The shifted representation of the target objects is then rotated around the reference object. This transforms the target representation into an intrinsic reference frame defined by the reference object's movement direction. This rotatory transformation is realized by a steerable neural mapping that shifts activation patterns along the azimuth of the polar coordinate representation of the relational field (lower left blue diamond in Fig. 2). The extent of the shift is determined by the movement direction of the reference object, which is held by the *rotation field*.

The rotated target representation is projected into the *relational CoS field*. A second input to this field from spatial concept nodes encodes the associated spatial templates through weight patterns (illustrated in the lower left of Fig. 2). Overlap of the two inputs leads to a peak that represents the selected target. The steerable neural maps thus make it possible to apply the relational operator encoded in the fixed weight patterns to objects at any visual location in any orientation, implementing neural pointers.

The relational CoS field projects into the selective spatial attention field via reverse transformations for rotation and shift (upper and lower right diamonds in Fig. 2). Selective spatial attention projects into the three-dimensional attentional fields, forming peaks there that in turn project to the feature fields, which may activate production nodes.

3.4. Concepts

Concepts like "red" or "toward" are represented by discrete nodes (denoted by circles in Fig. 2) that project with patterned synaptic weights into their respective feature fields.

The nodes come in pairs: *memory nodes* (blue circles) act as an interface to a user who may activate them as input or observe them as output; *production nodes* (pink circles) gate the impact of their respective memory nodes onto the architecture. Note that there are copies of such pairs of nodes for each role that a concept may appear in (e.g., two pairs for "red," as reference and as target), enabling role-filler binding. The synaptic weight patterns between nodes and fields could be learned by Hebbian learning rules but are hand-tuned here.

3.5. Process organization

The processes within the architecture are organized by instabilities of neural nodes that switch components "on" or "off." These discrete events thus emerge from the time-continuous neural dynamics. Process organization is based on a structural principle borrowed from behavioral organization (Richter, Sandamirskaya, & Schöner, 2012). The core structure is the *elementary behavior*, which consists of two dynamic substrates. The *intention node* (green circle in Fig. 2) determines whether a process is active and has impact on connected structures. The *condition of satisfaction node* (CoS, red circle) is activated once a process has terminated and inhibits the intention node, turning the process off. Here, we employ elementary behaviors that control the grounding of the reference object (reference behavior), the target object (target behavior), and the spatial relation term (spatial relation behavior) (top left in Fig. 2). Role-filler binding is preserved during grounding by processing reference and target objects sequentially, organized by a *precondition* node (black circle) that inhibits the intention node of the target behavior until the reference behavior has terminated.

4. Results

In the following, we describe the dynamic processes that unfold within the architecture as it executes tasks. The results come from numerical solutions of the architecture's differential equations.[2] To simplify visual object recognition, we use a scene with uniformly colored objects on a white background.

4.1. Describing an action

Fig. 3 illustrates the processes within the architecture as it generates a phrase about a video in which a red ball rolls toward a yellow ball (see top right of Fig. 2).

At $t = 0$ we give a boost into the architecture, which impacts the intention nodes of all behaviors. After this boost, the architecture runs autonomously, without any further intervention from user or program. First, the reference object is described; the target behavior is inhibited by the precondition constraint until the reference behavior is finished. Without information about which objects to describe, the architecture decides based on their saliency.

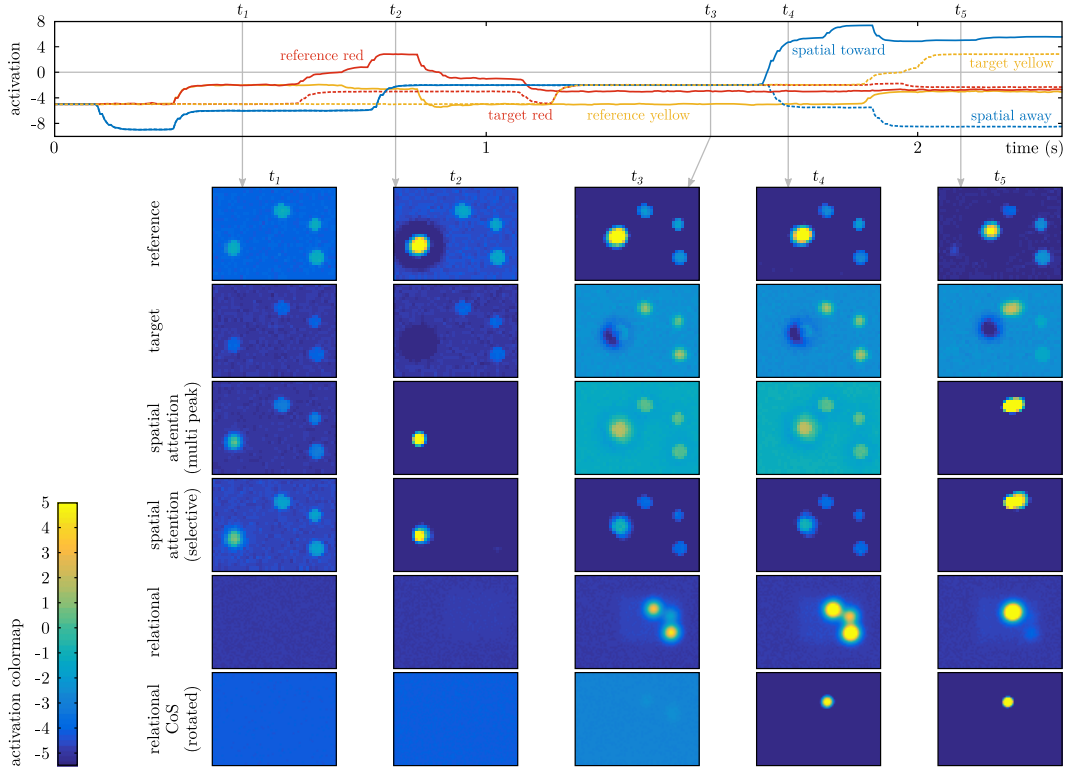

Fig. 3. Activation time courses of relevant production nodes (top) and activation snapshots of relevant fields at five points in time (bottom). Fields are color-coded using the color map on the bottom left.

At t_1, the selective spatial attention field shows a saliency advantage for the moving red object in the lower left corner.

At t_2, the spatial attention field has made a selection decision and formed a peak. This creates a self-sustained peak in the reference field, selecting the moving object as reference. It also activates the production node "reference: red" (top of Fig. 3) by projecting activation into the color CoS field via the attention color-space field (both not shown in Fig. 3; see Fig. 2). At the same time, the rotation angle field (not shown in Fig. 3) forms a representation of the object's movement direction, which it receives from the attentional movement-space field. It will later be used as a parameter to rotate the target objects. At this point, the architecture has described the reference object. That is, it has created a link from the continuous representation of the object in the fields to the discrete representation of its feature and role in the nodes.

At t_3, the behavior to ground the reference object has been inhibited by its CoS node and the behavior to describe the target object has become active. However, even though the reference behavior is inactive, the peak in the reference field is still tracking the position of the moving object, because it receives input from the perception fields. Contrary to the reference behavior, the selective spatial attention field is not boosted during the

target behavior, allowing multiple target candidates to be projected to downstream fields. The target field has formed three peaks at the positions of the remaining objects. The field's output is transformed and projected into the relational field, where the target positions are now represented relative to that of the reference object. This representation is rotated around the reference object and projected into the relational CoS field.

At t_4, the relational CoS field has formed a peak at the target position that overlaps most with the spatial template for the relation "toward." This activates the corresponding production node "spatial: toward."

At t_5, the activation from the relational CoS field is transformed and projected back into the selective spatial attention field, from there into the attentional color-space field, and from there into the target field as well as the color CoS field. The peak in the color CoS field activates the production node "target: yellow."

At this point, the architecture has produced the relational phrase "red toward yellow" and has created a grounding of this phrase in sensorimotor representations.

4.2. Perceptually grounding a phrase

The architecture can also ground a phrase provided by user input. Due to space constraints, we cannot describe the process at the same level of detail. The process is very similar to that of grounding spatial relations reported earlier (Richter et al., 2014). The difference to the process of describing, explained above, is that the user supplies a phrase, such as "red toward yellow," by activating memory nodes through manual boosts. Visual search for objects is then guided, as opposed to bottom-up saliency-driven. For instance, to ground the reference object, its red color is represented in the color intention field, bringing up peaks of red objects in the attentional color-space field—analogously with yellow objects for the target. Similarly, the template for spatial relations preshapes the relational CoS field and only allows peaks that overlap with the template. The description is established once a representation in the fields has been formed for each element of the supplied phrase.

5. Discussion

We have presented a neural process model that is able to describe simple scenes in terms of spatial relations and object-oriented actions. It can also perceptually ground such descriptions by attentionally selecting the designated objects in the scene. In the model, space-time continuous activation patterns are both coupled to sensory input and linked to neural representations of amodal concepts like *move toward* or *move away from*. This provides a neural processing account of the interaction between sensorimotor activation, conceptual processing, and language, that theories of perceptual symbols (Barsalou, 2008) and embodied construction-grammar (Bergen & Chang, 2013) postulate. The integrative nature of the model leads us to confront fundamental issues such as the neural pointer problem, the binding problem, and how discrete processing steps emerge from

time-continuous neural dynamics. Our solutions derive from the conceptual commitments of the theoretical framework of DFT.

We build on existing modeling approaches to the grounding of language that are neurally inspired but do not typically adhere to neural principles as consistently. For instance, the Neural Theory of Language (Feldman, 2006) is a hybrid framework that combines neural network concepts with ideas that are not compatible with neural process thinking. Similarly, Madden, Hoen, and Dominey's (2010) model for embodied language complements neural networks with algorithms that are not neurally based. Some models invoke neural concepts to account for psychophysical data. For instance, Regier and Carlson (2001) use the notion of an attentional vector sum to capture spatial terms. Such models are not typically embedded into architectures that autonomously generate the complete sequence of processing steps required to ground and generate language. Direct support for the neurophysiological foundation of the process account provided here comes from theoretical work on the recognition of transient hand actions that links very similar mathematical modeling to neural data from relevant cortical areas (Fleischer, Caggiano, Thier, & Giese, 2013).

The ambition of a neural process account for higher cognition is shared with the group of Eliasmith (2013). Their Neural Engineering Framework (NEF) enables spiking neural networks to realize vector symbolic architectures (Gayler, 2003). Concepts and objects are represented by high-dimensional vectors through an encoding and a decoding stage and transient neural patterns are computed by superposition and projection. DFT, in contrast, is based on self-stabilized activation patterns defined over low-dimensional feature spaces. Whether DFT and NEF span the same range of cognitive phenomena and which approach is more consistent with neural reality remains open for now.

The current model represents a first step toward a comprehensive neural process account of relational processing. More extensive assessment of the model, using a large number of different visual scenes and phrases, is a necessary next step. The relations implemented in this first model may be viewed as neural realizations of the image schemas LEFT-RIGHT, UP-DOWN, and PATH (Johnson, 1987; Lakoff, 1987). A systematic future effort should be to neurally realize other key image schemas (e.g., CONTAINER). Scaling the number of concepts and studying the autonomous learning of concepts and operators are other theoretical tasks. The present model only grounds and generates single phrase descriptions. Building the neural processing architecture that enables sequences of phrases and addresses the interdependencies between phrases and their perceptual basis is a major theoretical tasks. Finally, the working memory implicit in the neural representations of the model may provide the basis for a neural process account of relational mental models (Knauff, 2013).

Notes

1. Note that the representations denoted as "amodal" here are only relatively remote but not completely detached from the level of continuous sensory representations.

2. The architecture is implemented and simulated using the C++ framework *cedar* (Lomp, Richter, Zibner, & Schöner, 2016).

References

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(4), 723–742; discussion 743–767. doi:10.1017/S0140525X97001611

Barsalou, L. W. (2008). Grounded cognition. *The Annual Review of Psychology*, *59*, 617–645.

Bergen, B., & Chang, N. (2013). Embodied construction grammar. In T. Hoffmann & G. Trousdale (Eds.), *Oxford Handbook of Construction Grammar* (pp. 168–190). New York: Oxford University Press.

Berger, M., Faubel, C., Norman, J., Hock, H. S., & Schöner, G. (2012). The counter-change model of motion perception: An account based on dynamic field theory. In A. Villa, W. Duch, P. Érdi, F. Masulli, & G. Palm (Eds.), *ICANN* (pp. 579–586). New York: Springer.

Crocker, M. W., Knoeferle, P., & Mayberry, M. R. (2010). Situated sentence processing: the coordinated interplay account and a neurobehavioral model. *Brain and Language*, *112*(3), 189–201. doi:10.1016/j.bandl.2009.03.004

Doumas, L. A. A., & Hummel, J. E. (2012). Computational models of higher cognition. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford Handbook of Thinking and Reasoning* (52–66). New York: Oxford University Press.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. New York, NY: Oxford University Press.

Feldman, J. A. (2006). *From molecule to metaphor: A neural theory of language*. Cambridge, MA: MIT Press.

Fleischer, F., Caggiano, V., Thier, P., & Giese, M. A. (2013). Physiologically inspired model for the visual recognition of transitive hand actions. *The Journal of Neuroscience*, *33*(15), 6563–6580. doi:10.1523/JNEUROSCI.4129-12.2013

Gayler, R. W. (2003). Vector symbolic architectures answer Jackendoff's challenges for cognitive neuroscience. In P. Slezak (Ed.), *ICCS/ASCS International Conference on Cognitive Science* (pp. 133–138). Sydney: University of New South Wales.

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*, *14*(11), 497–505. doi:10.1016/j.tics.2010.08.005

Johnson, M. (1987). *The body in the mind—The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

Knauff, M. (2013). *Space to reason: A spatial theory of human thought*. Cambridge, MA: MIT Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind* (Vol. 64). Chicago: University of Chicago Press.

Lipinski, J., Schneegans, S., Sandamirskaya, Y., Spencer, J. P., & Schöner, G. (2012). A neuro behavioral model of flexible spatial language behaviors. *Journal of Experimental Psychology Learning*, *38*(6), 1490–1511. doi:10.1037/a0022643

Lobato, D., Sandamirskaya, Y., Richter, M., & Schöner, G. (2015). Parsing of action sequences: A neural dynamics approach. *Paladyn*, *6*, 119–135. doi:10.1515/pjbr-2015-0008

Logan, G. D., & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and Space* (pp. 493–529). Cambridge, MA: MIT Press.

Lomp, O., Richter, M., Zibner, S. K. U., & Schöner, G. (2016). Developing dynamic field theory architectures for embodied cognitive systems with cedar. *Frontiers in Neurorobotics*, 10(November), (pp. 1–18). http://doi.org/10.3389/fnbot.2016.00014

Madden, C., Hoen, M., & Dominey, P. F. (2010). A cognitive neuroscience perspective on embodied language for human-robot cooperation. *Brain and Language*, *112*(3), 180–188. doi:10.1016/j.bandl.2009.07.001

Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General*, *130*(2), 273–298. doi:10.1037/0096-3445.130.2.273

Richter, M., Sandamirskaya, Y., & Schöner, G. (2012). A robotic architecture for action selection and behavioral organization inspired by human cognition. In *IEEE/RSJ IROS* (pp. 2457–2464). New York: IEEE.

Richter, M., Lins, J., Schneegans, S., Sandamirskaya, Y., & Schöner, G. (2014). Autonomous neural dynamics to test hypotheses in a model of spatial language. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *36th CogSci* (pp. 2847–2852). Austin, TX: Cognitive Science Society.

Sandamirskaya, Y., & Schöner, G. (2010). An embodied account of serial order: how instabilities drive sequence generation. *Neural Networks*, *23*(10), 1164–1179. doi:10.1016/j.neunet.2010.07.012

Schneegans, S., & Schöner, G. (2012). A neural mechanism for coordinate transformation predicts pre-saccadic remapping. *Biological Cybernetics*, *106*(2), 89–109. doi:10.1007/s00422-012-0484-8

Schneegans, S., Spencer, J. P., & Schöner, G. (2015). Integrating what and where: Visual working memory for objects in a scene. In G. Schöner, J. P. Spencer, & the DFT Research Group (Eds.), *Dynamic thinking: A primer on dynamic field theory* (197–226). New York: Oxford University Press.

Schöner, G., Spencer, J. P., & the DFT Research Group. (2015). *Dynamic thinking: A primer on dynamic field theory*. New York: Oxford University Press.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. doi:10.1016/0010-0285(80)90005-5

van Hengel, U., Sandamirskaya, Y., Schneegans, S., & Schöner, G. (2012). A neural-dynamic architecture for flexible spatial language: intrinsic frames, the term "between," and autonomy. In *IEEE RO-MAN* (pp. 150–157). New York: IEEE.

Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*, *18*(5), 229–234. doi:10.1016/j.tics.2014.02.008