New Ideas in Psychology xxx (2013) 1–18

Contents lists available at SciVerse ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

Using Dynamic Field Theory to extend the embodiment stance toward higher cognition

Yulia Sandamirskaya*, Stephan K.U. Zibner, Sebastian Schneegans, Gregor Schöner

Institut für Neuroinformatik, Ruhr-Universität Bochum, Universitätsstr. 150, 44780 Bochum, Germany

Keywords:
Dynamic Field Theory
Embodied cognition
Scene representation
Sequence generation
Spatial language

ABSTRACT

The embodiment stance emphasizes that cognitive processes unfold continuously in time, are constantly linked to the sensory and motor surfaces, and adapt through learning and development. Dynamic Field Theory (DFT) is a neurally based set of concepts that has turned out to be useful for understanding how cognition emerges in an embodied and situated system. We explore how the embodiment stance may be extended beyond those forms of cognition that are closest to sensorimotor processes. The core elements of DFT are dynamic neural fields (DNFs), patterns of activation defined over different kinds of spaces. These may include retinal space and visual feature spaces, spaces spanned by movement parameters such as movement direction and amplitude, or abstract spaces like the ordinal axis along which sequences unfold. Instances of representation that stand for perceptual objects, motor plans, or action intentions are peaks of activation in the DNFs. We show how such peaks may arise from input and are stabilized by intra-field interaction. Given a neural mechanism for instantiation, the neuronal couplings between DNFs implement cognitive operations. We illustrate how these mechanisms can be used to enable architectures of dynamic neural fields to perform cognitive functions such as acquiring and updating scene representations, using grounded spatial language, and generating sequences of actions. Implementing these DFT models in autonomous robots demonstrates how these cognitive functions can be enacted in embodied, situated systems.

1. Introduction

One way to approach embodied cognition is to observe that there is a lot of cognition in such seemingly mundane activities as soccer playing. Although soccer playing may commonly be thought of as a motor skill, perception is a critical component as well. Players must quickly acquire an understanding of the scene, and segment and categorize objects such as the ball, the goal posts, line markings, other players, and the umpire. Every player must track these objects when either the objects or the player move. Good scene understanding, including a perception of space that affords planning, is key to successfully driving the game ahead. Although it has been said that the world is its own best model, to effectively orient within the scene and direct gaze back to relevant objects, players must have a scene representation or spatial map that can be used even when the exact position or orientation of the player has changed since the last updating of the map from sensory information.

The motor aspects of soccer playing go well beyond conventional motor control. Actions must be initiated or aborted, action goals must be selected, distractors – suppressed. Sensorimotor decisions must be continuously

* Corresponding author.
E-mail addresses: yulia.sandamirskaya@ini.rub.de, sandayci@rub.de (Y. Sandamirskaya), stephan.zibner@ini.rub.de (S.K.U. Zibner), sebastian.schneegans@ini.rub.de (S. Schneegans), gregor.schoener@ini.rub.de (G. Schöner).

0732-118X/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.newideapsych.2013.01.002

updated, as objects move and view-points change. Updating may also take place at higher levels, such as when switching back from an offensive to a defensive strategy immediately after losing control of the ball.

Finally, soccer playing always involves learning, which takes place whenever an individual plays. From soccer game to soccer game, there is an obvious developmental trajectory, with a general increase in competence as experience with such open games accumulates. More subtly, soccer playing involves a lot of background knowledge about such things as how hard to hit the ball, how hard to tackle another player, or how slippery the ground may be. Such background knowledge (Searle, 2004) is difficult to capture, but it is a clear reminder that the cognition that happens in soccer is not the processing of arbitrary information. Instead, this form of cognition happens in a specific context, to which players are particularly adapted by training or even by evolution and which provides supportive structure for the tasks handled by the Central Nervous System during a game of soccer.

The recognition that cognitive processes take place in such embodied and situated settings has led to important new thinking (reviewed, for instance, by Anderson (2003); Ballard, Hayhoe, Pook, and Rao (1997); Brooks (1990)). The new ideas include the insight that cognitive processes are based on active perception, are linkable to the sensory and motor surfaces, can be updated at any time, and are sensitive to situational and behavioral context (Schneegans & Schöner, 2008). These new ideas have resonated with a developmental approach to cognition that dates back to Piaget (Piaget, 1952) and emphasizes the sensorimotor origins of cognition (Thelen & Smith, 1994).

But is all cognition embodied? Not all cognition involves bodily motion or even the processing of sensory information (Riegler, 2002). Are the constraints that arise from the discovery of embodied cognition universally shared by all cognitive processes? Are all cognitive processes linkable to sensory and motor surfaces; do all cognitive processes unfold in continuous time, capable of updating their contents at any moment based on new incoming information; are all cognitive processes sensitive to context and open to learning? The *embodiment hypothesis* is that these questions must be answered in the affirmative! According to the embodiment hypothesis there is no particular boundary below which cognition is potentially embodied, beyond which these constraints no longer apply and "real" cognition begins. The more we know about the neural basis of cognition, the more clearly we see a true continuum of neural processing from the sensorimotor domain to the highest form of cognition (Bar, 2011). Early sensory and motor areas are also actively involved in acts of higher cognition (Jeannerod & Decety, 1995; Kosslyn, Thompson, & Ganis, 2006). And skills of a seemingly sensorimotor nature require the intervention of relatively high-level cognitive control (Koechlin, Ody, & Kouneiher, 2003).

If the embodiment hypothesis is true, how may we go about understanding cognition? How do we make the embodiment program, the investigation of cognition on the basis of the embodiment constraints, concrete and operational? To us a critical step is to develop a constructive, process-oriented theory that enables the modeling of concrete acts of embodied cognition. We believe that such a theory must be based on neuronal principles, that will make it compatible with the constraints imposed on information processing in the Central Nervous System.

Dynamic Field Theory (DFT) grew out of this research agenda. Its beginnings lay in the sensorimotor domain (Erlhagen & Schöner, 2002; Kopecz & Schöner, 1995) and the development of early cognition (Thelen, Schöner, Scheier, & Smith, 2001). The key ideas of DFT are: (1) Patterns of neural activation evolve in time as described by neural dynamics that captures the evolution of the activity of populations of neurons in continuous time. (2) The neural activation patterns are defined over continuous spaces, which describe sensory and motor states and abstract from the discrete sampling of these spaces by individual neurons. (3) Localized peaks of activation are units of representation, which indicate through high levels of activation the presence of a well-defined value along the dimensions of the activation fields. That value is indexed by the location of the peak within the neural field. (4) Neural interaction within the activation fields is structured so that localized peaks are stable stationary solutions, or attractors, in the neural dynamics.

The spatio-temporal continuity of the neural activation fields in DFT is critical to establishing stability as an operational concept. Stability is the resistance of solutions to change induced by variations of sensory input or by noise. Thus, stability requires a metric, a way to express what it means to be close to a state and to converge in time toward a state after a perturbation has occurred. Whenever a neural process is part of a feedback loop, stability is a critical property without which the neural process will not have a reliable function. In order to have an impact on the down-stream neural structures or motor systems, a neural state needs to persist over a macroscopic period of time despite neural and sensory noise, as well as continual changes in the sensory input. Stability is the basis for representation in DFT and the key to countering the anti-representationalist approach to embodied cognition (Chemero, 2009).

In DFT, a set of instabilities controls how peaks as attractor states may be created or may disappear. These instabilities give rise to three elementary cognitive acts (Schneegans & Schöner, 2008): (1) The detection instability creates a peak in response to input. (2) The selection instability controls which among multiple stimulated values of a dimension is stably selected. (3) The memory instability separates a regime in which peaks persist once the inducing input is removed from a regime in which peaks are only stable in the presence of such input. These instabilities have been used to account for signatures of early cognition such as sensorimotor decision making (Trappenberg, Dorris, Munoz, & Klein, 2001; Wilimzig, Schneider, & Schöner, 2006), spatial cognition (Simmering, Schutte, & Spencer, 2008) and its development (Schutte & Spencer, 2002), change detection (Johnson, Spencer, & Schöner, 2008), and visual search (Fix, Vitay, & Rougier, 2007, pp. 170–188). For the link to the underlying neuronal mechanisms see, for instance, Coombes (2005).

In this review we explore how the language of DFT enables us to extend the embodiment stance toward higher

Please cite this article in press as: Sandamirskaya, Y., et al., Using Dynamic Field Theory to extend the embodiment stance toward higher cognition, *New Ideas in Psychology* (2013), http://dx.doi.org/10.1016/j.newideapsych.2013.01.002

forms of cognition less directly linked to the sensorimotor domain. We show how the creation of instances within particular categories may be understood within DFT and illustrate that mechanism in a model of perceptually grounded spatial language. We show how operations may be applied to such instances to create new instances. We illustrate that mechanism by showing how transformations among different coordinate systems can be implemented in DFT, along with their updating when a coordinate frame shifts. Finally, we show how operations may move a neural representation from one stage in a sequence to the next through cascades of elementary instabilities in DFT. We illustrate how such cascades of instabilities can be used to generate sequences of actions or perceptual states.

As we explore the extension of dynamical systems thinking to mid-level and higher cognition, we probe the continued embodiment of the postulated neural processes by implementing the DFT architectures in autonomous robots. By linking the neural dynamics to real-time sensory inputs derived from cameras with minimal preprocessing and using the neural activation patterns to drive robotic action we demonstrate that the interface between the level modeled by dynamic neural fields and the sensory and motor surfaces contains no hidden assumptions or hidden cognitive competencies not accounted for by the theory.

We begin with a quick review of the foundations of DFT, followed by a discussion of the key operations on which the expansion of DFT to higher cognition is based. Three architectures that employ these operations to generate cognitive processes are then used to illustrate the ideas and demonstrate their grounding in embodied systems.

2. Dynamic Field Theory (DFT)

2.1. Grounding neural representations in continuous spaces

Perceptual states and motor actions can be thought of as being embedded in continuous spaces of possible percepts and possible acts (Spivey & Dale, 2004). Ultimately, these spaces originate at the sensory and motor surfaces and in the physical world. The notion of space is critical in the domain of visual cognition. Visual space arises at the retina, but can be transformed into allocentric frames of reference and augmented by depth through stereo vision. Attention is typically spatially specific (Sperling & Weichselgartner, 1995). Objects are formed based on the locality of visual features (Treisman, 1998). Space is similarly important in the domain of motor behavior. The end-point of a saccadic eye movement (Sparks, 1999) or the spatial direction in which a hand or other end-effector is moved (Georgopoulos, 1995) are movement parameters anchored in physical space. Beyond the immediate sensory and motor surfaces, more abstract spaces shape our perception and action: spaces of visual features such as color and orientation (Hubel, 1988), auditory feature spaces spanned by pitch or timbre (Schreiner, 1995), space formed by movement parameters such as the direction of an external force or the amount of resistance a movement must overcome (Sergio & Kalaska, 1998). Finally, highly abstract cognitive spaces include such concepts as the number line (Dehaene, 1997) or the ordinal dimension along which events may be lined up in a series (Henson & Burgess, 1997).

A neuronal representation of continuous metric dimensions, or spaces, may be based on the principle of space coding (Dayan & Abbott, 2001): Each neuron is characterized by the stimulus or motor condition to which it is most sensitive; the neuron "stands for" its tuning curve or receptive field. The neuron's firing rate is then a vote for whatever the neuron stands for. Because neurons in the cortex tend to be broadly tuned to many different parameters, any specific behavioral condition will lead to the activation of broad populations of neurons, which has led to the notion of a population code (Deadwyler & Hampson, 1995). If we ask how such a population of neurons represents the underlying continuous space, we may consider each neuron's contribution to be "smeared out" by its broad tuning function. This leads to the notion of a neural field, in which an individual neuron is not strictly localized but represented through its tuning curve (for a formalization in the visual domain see Jancke et al., 1999; in the motor domain see Bastian, Schöner, & Riehle, 2003). Inherent in the notion of such neural fields is the idea that although many neurons contribute to the neural activation pattern, that pattern really spans a low-dimensional space. This may be contrasted with the notion that populations of neurons represent high-dimensional activation vectors, which are not necessarily embedded in a low-dimensional space (Eliasmith, 2005).

Fig. 1 illustrates a neural activation field defined over a single spatial dimension. The field evolves in time under the influence of localized input, growing a localized peak out of an initial state (which includes the initial localized pre-activation, – see below). Other field locations are progressively suppressed in activation as the peak emerges. In DFT, localized peaks of activity are units of representation. Through its activation level, a peak indicates the presence of information about the dimensions the field represents. High levels of activation imply that the field may affect downstream systems. Through its location along the field dimension, a peak specifies the contents of that information.

Apart from localized peaks, fields can also be structured in a graded way, expressing graded prior knowledge or expectation. Such distributed inputs to the field "preshape" it: they reflect information accumulated by the field on a longer time-scale, prior to the field's current dynamics (Erlhagen & Schöner, 2002; Trappenberg et al., 2001). Preshaping may arise from a simple learning mechanism, the memory trace.

2.1.1. Stability and neuronal dynamics

In DFT, activation fields evolve continuously in time as described by the dynamics of neural activation. DFT thus abstracts from the discrete temporal structure of spiking events (Ermentrout, 1998). The dynamics of the activation fields is defined in such a way that peaks are stable states. Stability of the functionally relevant states is critical for these states to remain invariant under coupling within the dynamic neural network (recurrence), or through the world (the action-perception loop). Stability guarantees that graded changes in input lead to graded changes in

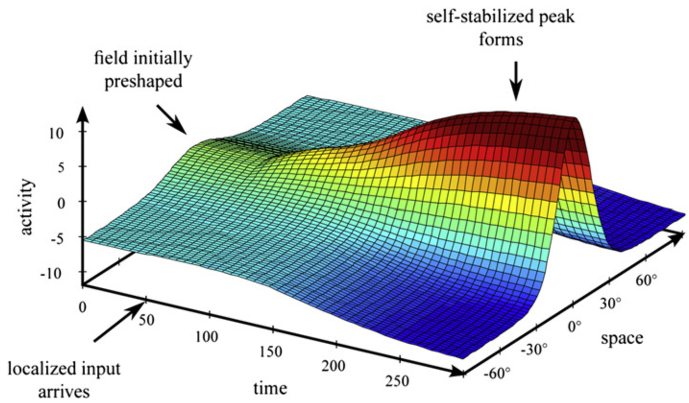

Fig. 1. *Evolution of activation from an initial pattern of pre-activation (or preshape) toward a localized, self-stabilized peak of activation in a dynamic neural field.* The field is defined over a continuous metric space.

output. As long as a peak remains stable, it may affect on downstream systems in a persistent way, resisting perturbations and leading to overt action on the macroscopic time scale of behavior.

In the mathematical model of the neural field dynamics

$$\tau\dot{u}(x,t) = -u(x,t) + h + \int f(u(x',t))\omega(x-x')dx' + S(x,t), \tag{1}$$

τ is a time constant that determines how quickly the activation function, $u(x,t)$, relaxes to an attractor state that emerges from the stabilization factor, $-u(x,t)$, and the additive contributions: the negative resting level, $h < 0$, lateral neural interactions shaped by the kernel, $\omega(x-x')$, and the external input, $S(x,t)$. The kernel is a bell-shaped function containing both excitatory connectivity of strength c_{exc} over the range, σ_{exc}, and inhibitory connectivity of strength c_{inh} over the range, σ_{inh}:

$$\omega(x-x') = c_{exc}\, exp\left[-\frac{(x-x')^2}{2\sigma_{exc}^2}\right] - c_{inh}\, exp\left[-\frac{(x-x')^2}{2\sigma_{inh}^2}\right]. \tag{2}$$

The sigmoidal function

$$f(u(x,t)) = \frac{1}{1+exp[-\beta u(x,t)]} \tag{3}$$

adds non-linearity to the dynamics and expresses that only sufficiently activated field locations contribute to neural interaction. The stability of stationary localized peaks of activation derives from these non-linear interactions in the field, a fact that can be established analytically in relevant limit cases (Amari, 1977; Ermentrout, 1998; Wilson & Cowan, 1973). Intuitively, the short-range excitatory interaction stabilizes a peak solution against decay, while long-range inhibitory interaction stabilizes peaks against spread by diffusion.

Next, we provide a qualitative description of the stable states and their instabilities. In the absence of external input, a non-peak attractor state of the field has constant activation along the field dimension at a level equal to the negative resting level, h. This *sub-threshold solution* remains stable when weak localized input, $S(x,t)$, is introduced, as long as the summed activation level, $h + S(x,t)$, does not exceed anywhere along the field dimension levels at which the lateral interaction becomes engaged. When that threshold is passed, the output, $f(u(x,t))$, and the interaction function drive the neural field into a different dynamic regime. Activation grows near the field sites at which localized external input was largest, developing into a *localized peak* that inhibits the activation field elsewhere. Such peaks are self-stabilized by intra-field interactions, but also track changing localized input.

The different dynamic regimes of a neural field are separated by instabilities, or bifurcations. From a structured sub-threshold input, or preshape, a localized activity peak can be induced merely by increasing the resting level of the field. That happens when there is a *detection* instability, which signals a decision to create an instance of what the field dimension represents. The peak also provides an estimate of the characteristic parameter for that instance, which may now impact on downstream structures. When localized input is removed or the resting level is lowered, the detection instability may be experienced in reverse. When there is such a *reverse detection* instability, the self-stabilized peak becomes unstable and the system relaxes to the sub-threshold attractor. A reverse detection instability happens at lower levels of localized input or at a resting level lower than the detection instability, yielding a bistable regime in which detection decisions are stabilized.

For sufficiently large resting levels, h, or strong lateral excitation in the neural field, the forgetting instability may not occur even when localized input is removed entirely. Activity peaks that are sustained without the input that first induced them are used as models of working memory (see the right column of Fig. 2). Multiple sustained peaks may be created in a field, either sequentially or simultaneously, but a capacity limit arises naturally from the mutual inhibition among them (Johnson, Spencer, & Schöner, 2009). Without external input, the sustained

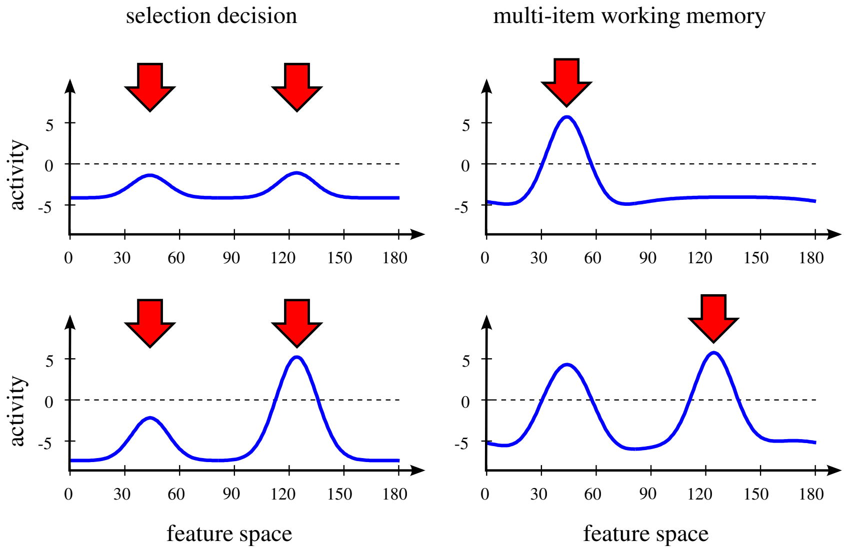

Fig. 2. *Selection and working memory in DNFs.* The left column illustrates two time slices in the evolution of activity during a selection decision. Two localized inputs (arrows) are applied to a field, and initially drive the activity at the corresponding field locations (top panel). Once the activity reaches the threshold (at zero) of the sigmoidal nonlinearity, inhibitory interaction leads to competition between the two locations. The competition may be decided by differences in input strength, differences in the timing of the input, or by fluctuations. The outcome of the competition is the formation of a single stabilized peak of activation and the suppression of activation at the other location (bottom panel). The right column shows two time slices in a different regime in which a more localized inhibitory interaction enables multiple peaks to co-exist. Input (arrows) is provided consecutively to two locations in the field. In each case, an activation peak forms and persists when the input is removed, sustained by the local excitatory interactions in the field. This is the basis of constructing a working memory of the stimulus history.

peaks are also subject to drift caused by noise or interactions with other peaks. To erase working memory peaks, a forgetting instability may be induced by lowering the field's resting level.

If the lateral interactions feature strong, long-range inhibition, a field may only allow a single peak of activity to persist. When two localized inputs of similar strength are applied to such a field at distinct locations, there is a *selection decision* through which a single peak forms at one of the input locations, while the other one is suppressed (left column of Fig. 2). Which among the multiple locations is selected may depend on the relative strengths of inputs. Within a range of relative strengths, selection is multistable: the selection decision is stabilized against change of input. Beyond that range, a *selection instability* destabilizes peaks with weaker input, with the consequence that selection switches to the most strongly supported choice.

Together with their sensory input, neural fields have autonomous dynamics: there are no explicit computational cycles, there is no process for monitoring salient events beyond the field's dynamics, no algorithm that checks whether a field is active or not. In continuous-time dynamics, the stable states emerge and are reset in an interplay of inputs and lateral interactions within the dynamic fields. The stable peaks of activation remain sensitive to changing input and may track the attractor if input changes gradually. If input changes more sharply, the dynamics may switch to a new, qualitatively different attractor in an instability.

After a perturbation or change in the input, a dynamic field relaxes quickly to the new attractor state. Thus, the dynamic field spends most of its time in either a subthreshold, inactivated attractor, or an activated attractor with one or several activation peaks. Transitions between these states are fast and happen relatively rarely. The stability of the attractor states filters out noise and stabilizes the states against external perturbations. Stability also enables coupling among neural fields that keeps the qualitative dynamics invariant. As long as a solution remains stable, coupling does not remove or qualitatively alter the solution. This invariance property enables architectures for neural fields to be built up from individual field components, each designed to be in a chosen dynamic regime.

3. Elements of higher cognition in DFT

3.1. Creating instances

Creating instances of representational categories is a core function of any cognitive architecture. In embodied cognition we expect that instances may be created in response to input from the sensory surfaces. In this case, creating an instance amounts to bringing a perceptual object into the foreground. However, instances may also be created from internal neuronal processes. What we sketch here is thus consistent with Barsalou's (1999) notion of perceptual symbol systems.

Within the language of DFT, representational categories are neural activation fields over particular feature dimensions. Instances are self-stabilized peaks of activation in these fields. Creating an instance is, therefore, creating a localized peak of activation. Moving from sub-threshold graded patterns of activation to a self-stabilized peak always entails an instability, in which the sub-threshold patterns become unstable. The creation of a peak is a discrete event, which we have referred to as a “detection decision” in earlier work (Schneegans & Schöner, 2008).

To make things concrete, consider a scene in which objects of different colors are distributed on a top of a table. To represent this visual scene we employ a field that spans the feature dimension “hue” and two dimensions of the visual space. For ease of illustration we drop one spatial dimension, representing visual space along a line only. Fig. 3 shows the resulting two-dimensional field in which one dimension spans hue and the other dimension spans visual space.

The two-dimensional space-color field is linked to the visual sensor. Each pixel in the visual array provides input to the field according to its location, hue, and saturation. Objects in the visual array correspond to subthreshold activation blobs that are localized in both color and space dimensions. This localized input may drive activation through the detection instability, leading to the creation of localized peaks at the locations, in which objects are present in the current visual scene. The activity peaks arise in an instability and are self-stabilized by lateral interactions in the space-color field. This detection instability amounts to a decision arising in the continuous neural dynamics. The peak’s self-stabilizing property is critical for their further use in cognitive operations, discussed in the next subsection.

A neural field may support several peaks of activation or may allow only for a single activation peak. Which of these regimes the field operates in depends on the strength and spatial spread of excitatory and inhibitory interactions in the field. In a parametric regime that supports self-sustained activation, peaks in the space-color field persist as a working memory of the content of the scene. The peaks then represent the previously perceived items in the scene even after input from the visual array is removed from the field (e.g., because there has been an eye movement, because objects are occluded, or because the eyes are closed). In this regime it is particularly obvious that the peaks are instances of the representational category that arise out of an active process of instantiation, not merely the result of a transduction of sensory input.

With sufficiently strong global inhibition, only a single self-stabilized peak may exist at any time within the field. Without those limits, sensory input from the visual array induces the creation of a single peak centered over one of the stimulated locations in the two-dimensional field. This selection decision is another dynamic property of neural fields that distinguishes the creation of self-stabilized peaks from mere signal transduction.

Which location is selected depends most decisively on the temporal order of stimulation. Because selection decisions are stabilized by the field dynamics, any field location that has an initial competitive advantage from being stimulated earlier, may suppress competing field locations through inhibition. These locations may not, in turn, be able to inhibit the leading locations as they are kept from reaching the threshold beyond which they would be contributing to interaction. Inhibitory coupling as a mechanism for competition thus translates graded time differences into categorical differences in selection.

The relative strength of the input at different locations may compensate for advantages of timing. A more strongly stimulated location may grow activation to threshold faster than a competing location. Fluctuations induced by sensor or neural noise may make the outcome of the competition stochastic.

This dependence of selection on input strength is the basis for biasing competitive decisions. Fig. 3 shows how such bias may arise. A ridge of input activation, which is localized along the hue dimension but is homogeneous along the spatial dimension, expresses a preference for colors close to a particular hue value, say “red”. At those locations in the field, where current sensory input matches this color preference, the localized input activation from the sensor overlaps with the ridge, giving those locations a competitive advantage. The competition is, therefore, biased toward locations with colors that match the color specified by the ridge. The concept of biased competition originates in neuronal signatures observed by Desimone and Duncan (1995). Depending on the strength of

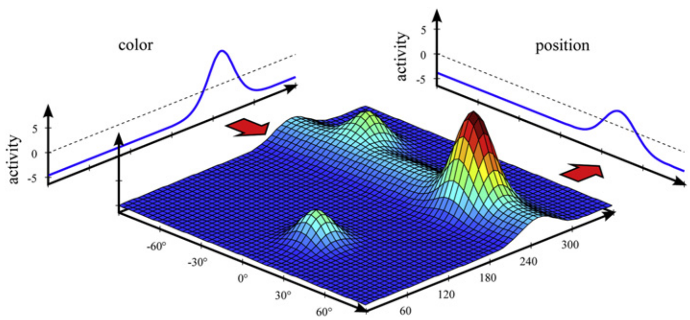

Fig. 3. *Biased competition in a DNF.* A one-dimensional color field provides a subthreshold input in form of a ridge of activation to a two-dimensional space-color field. The space-color field also receives input from a visual sensor; several localized subthreshold blobs indicate the colors and locations of salient visual stimuli. The color-ridge input overlaps with one of the localized visual inputs in the space-color field, and the summed inputs induce an activation peak at the location of this visual input. The localized positive activation within the peak provides input to a one-dimensional spatial representation, which represents the position of the object that won the competition. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Sandamirskaya, Y., et al., Using Dynamic Field Theory to extend the embodiment stance toward higher cognition, *New Ideas in Psychology* (2013), http://dx.doi.org/10.1016/j.newideapsych.2013.01.002

inhibitory coupling, biased competition may operate by selecting a single matching input location or may allow multiple matching items to become instantiated through peaks of activation. In either case, the spatial locations of items with the cued color can be read out by summing supra-threshold activation along the hue dimension and projecting the resulting spatial activation pattern onto an appropriate down-stream structure.

One may think of combining such ridge-input with the current sensory input localized along both dimensions as integrating of top-down with bottom-up information. While the specific selection is controlled by sensory input, the criterion for selection is mediated by top-down input. Another way to think of this mechanism is as a form of "real-time association", which effectively generates answers to queries such as "where is the red item". The mechanism differs from associative memory because it may operate on current sensory input, encountered for the first time at the moment of the query. It may also provide multiple answers at the same time. On the other hand, this same mechanism could be used to operate on long-term memory: if the activation pattern localized in both dimensions comes from a memory trace or another learning mechanism, then the ridge input creates from that memory one or more instances that represent the activated memory, an answer to the query "where was the red item in the remembered scene?"

Influence from other neuronal representations may also, more generally, determine when instances are created. For instance, a neural process may provide constant, homogeneous input to the two-dimensional field. This will push the field past the threshold of the detection instability. One or more peaks will arise at locations whose activation level is elevated over the rest of the field even if only by a small amount. Such pre-activation at specific field sites may arise from sensory input, but also from the memory trace we just mentioned, or from other learning processes. The detection instability thus enables the creation of macroscopic instances, self-stabilized peaks, out of graded and potentially weakly differentiated structure in the field.

Again, this illustrates how the creation of peaks is an active process, controllable by top-down structures. Through this mechanism of driving a field through the detection instability by homogeneous boosts in activation, DFT reduces the demands on learning mechanisms: if such mechanisms leave only minor differences in activation within the field, these can be amplified into macroscopic instantiations.

3.2. Operating on instances

A second core aspect of cognition is that instances of representational categories can be operated on. In DFT, instances of representation are peaks of activation in dynamic neural fields. Operations are therefore the creation of instances – that is of new peaks – within the same or another representational category (field), driven by instances (peaks) created in a first category (field).

Neural representations afford, of course, simple operations such as rescaling or shifting activation patterns, which can be brought about through the appropriate synaptic connections that project from one field to another (Pouget, Dayan, & Zemel, 2000). However, such simple mappings cannot capture more complex and flexible operations that combine two or more inputs. A more demanding operation of this kind is critical to spatial cognition: Shifting the neural representation of a spatial location by an amount specified by another neural representation; in other words, implementing the transformation of a reference frame (Andersen, Snyder, Bradley, & Xing, 1997; Colby & Goldberg, 1999; Zipser & Andersen, 1988).

To make things concrete consider a field that represents visual information in retinal coordinates and a second field that represents the current direction of gaze relative to a body frame of reference (Fig. 4). For simplicity, we combine eye and head movements into a single gaze variable, and use only one-dimensional representations for the visual and the gaze information, which reflect only the horizontal spatial component.

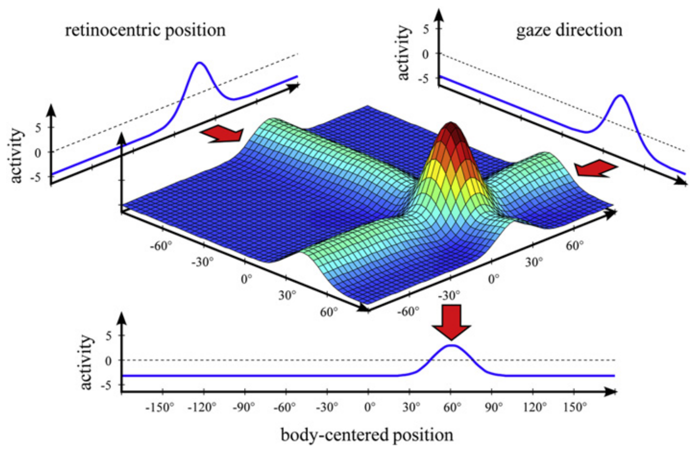

Fig. 4. *Implementation of a reference frame shift with DNFs*. In order to determine the body-centered position of a visual stimulus, a combined representation of retinal stimulus position and current gaze direction is generated in a two-dimensional field. The two values are projected into this field from the corresponding one-dimensional fields as input ridges, and an activity peak forms at the intersection of the two ridges. The body-centered position can then be read out from the position of this peak through a projection along the diagonal axis onto a one-dimensional body-centered field.

Please cite this article in press as: Sandamirskaya, Y., et al., Using Dynamic Field Theory to extend the embodiment stance toward higher cognition, *New Ideas in Psychology* (2013), http://dx.doi.org/10.1016/j.newideapsych.2013.01.002

We want to shift the position of the peak in the retinal field by a vector that indicates the current gaze direction, as determined by the peak position in the gaze-direction field. That operation transforms, in effect, information in a retinal reference frame into information in a body-centered reference frame. The shift of the reference frame amounts to vector addition of the retinal position of an item with the current gaze direction (Pouget & Sejnowski, 1997). We will show how this operation may be brought about in DFT.

A combined representation of retinal stimulus position and gaze direction in the form of a two-dimensional transformation field is the basis for this operation (Fig. 4). The field receives two ridge inputs from the two one-dimensional fields, one input along each axis (Deneve, Latham, & Pouget, 2001; Pouget, Deneve, & Duhamel, 2002). The connectivity is analogous to the mechanism described above (Fig. 3), which biases a selection decision in the two-dimensional field. The difference is that now the ridges are not used to select among localized inputs already present within the two-dimensional field, but to create a new localized peak of activation. This happens by choosing a resting level and input strengths of the two-dimensional field so that a single input ridge is not sufficient to induce a peak. Only at the intersection between two input ridges is the detection instability reached.

Different operations then become possible based on appropriate projections from and to this two-dimensional transformation field. The receiving representation of the reference frame shift is a one-dimensional field that, in effect, functions as a body-centered representation (bottom of Fig. 4). If gaze direction shifts to the right, the retinal image is shifted by the inverse amount to the left. All points that correspond to an invariant, body-centered location therefore lie on a diagonal line. The body-centered field thus receives input that sums at each location along the diagonal of the two-dimensional transformation field. Creating peaks in the retinal and gaze direction fields consequently induces a peak in the body-centered frame, which remains invariant under gaze shift.

At a first glance, the transformation field may look like an overly resource-intensive way of implementing neural computations, which would more typically be thought of in terms of patterns of synaptic connectivity. Note, however, that the dynamic field approach endows these operations with specific properties that are valuable for the generation of flexible behavior in an embodied system.

First, the dynamic field architecture is capable of performing an operation on multiple inputs at the same time. When the reference frame is being transformed, if the retinal field contains multiple peaks that indicate the locations of different stimuli, these produce parallel input ridges to the transformation field. The activity peaks formed by the intersections in the transformation field each project to a different location in the body-centered field, yielding the locations of all stimuli relative to the body (Schneegans & Schöner, 2012).

Second, the same mechanism can also be applied to graded representations in the input fields. If, for example, there is some uncertainty about the actual gaze direction (e.g., on account of conflicting sensory signals), this can be expressed by a broader activity distribution in this field that reflects how much evidence there is for each gaze angle. This uncertainty can be passed down through the transformation mechanism and is reflected in the resulting body-centered representation.

An important, easily overlooked limit case is the absence of one or more inputs. If, for instance, there is currently no visual stimulus represented in the retinal field, the system will produce an equally flat activity distribution in the body-centered field as the appropriate response. If a stimulus then appears and forms a peak in the retinal field, peaks are induced in the transformation and body-centered fields at the appropriate locations.

Finally, if converse projections between the fields are introduced, additional operations become possible: The retinal location of a peak represented in the body-centered frame (e.g., from grasping an object) can be predicted using information about current gaze direction (Schneegans & Schöner, 2012), or the gaze direction itself be estimated by comparing retinal and body-centered representations. When all projections between the fields are bi-directional, the field interactions drive the activation patterns in all fields toward a consistent global pattern, effectively filling in those representations for which no input was provided (Deneve et al., 2001).

3.3. *Generating sequences*

We have just seen how neural coupling can bring about operations on neural representations. Critically, for this to work a neuronal coupling becomes active only when an instance of a representational category is created, i.e. a peak is formed. The operation itself is slap based on self-stabilized peaks of activation. The operation brings about its effect by creating another instance, a peak, in the same or a different representational category. The stability of all these peaks plays a decisive role in bringing the neuronal couplings into effect only when they are appropriate. The connections between activation variables that are not part of a currently stable activation peak are, in effect, "turned off" and thus do not perturb the operation.

Stability is thus a prerequisite for reliably operating on dynamic representational instances. This poses a problem, however, for a particular, important class of operations, those involving temporally organized sequences. In sequence generation, a representational state brings about an action, or a cognitive operation, and thus needs to persist long enough to have an effect. But then this representational state must be deactivated, yielding to the subsequent element in the sequence. How may a stable activation peak be deactivated through the consequences of operations it has itself brought about?

Clearly, this must entail another instability, resolving the conflict that stems from the need for the initial activation peak to be stable so it can drive an operation and then to be destabilized by the consequences of the operation. The solution to this conflict can be found in a structure of the neuronal dynamics that demarcates the relevant instabilities (Sandamirskaya & Schöner, 2010). This structure is illustrated in Fig. 5.

There are three major components to the sequence-generating structure. At the top left of Fig. 5, a neuronal

Please cite this article in press as: Sandamirskaya, Y., et al., Using Dynamic Field Theory to extend the embodiment stance toward higher cognition, *New Ideas in Psychology* (2013), http://dx.doi.org/10.1016/j.newideapsych.2013.01.002

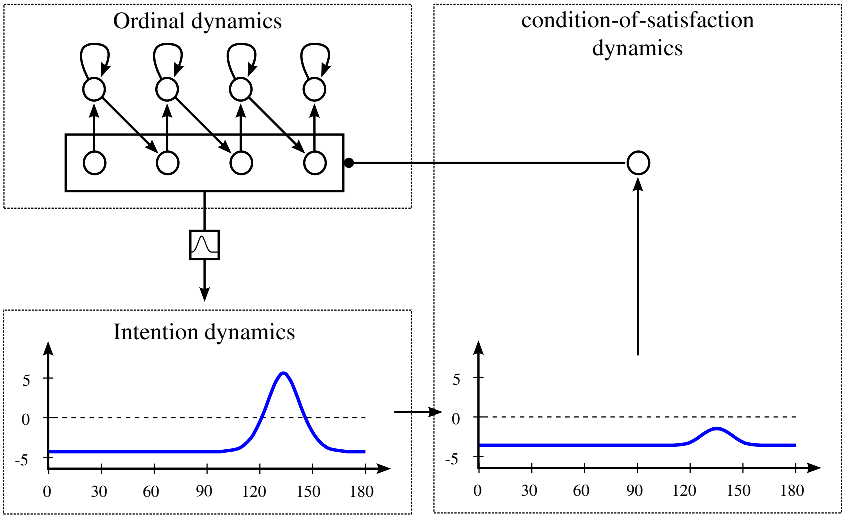

Fig. 5. *The mechanism for sequence generation in DFT:* The ordinal dynamics (upper left) comprises a set of bi-stable dynamical nodes sequentially interconnected through self-stabilizing memory nodes. The ordinal nodes project their activation onto intentional dynamics (lower left), consisting of one or several dynamic neural fields representing perceptual and motor parameters of the intended actions. The dynamics for condition-of-satisfaction (right) is activated when a state is detected in the neural field for condition-of-satisfaction, which corresponds to the expected outcome of the currently active intention. The active condition-of-satisfaction inhibits the currently active ordinal node, thereby triggering an instability in the system's dynamics that leads to a sequential transition.

representation of sequentiality per se defines activation over an abstract dimension, the "ordinal" axis along which the serial order of actions is represented. In the model, this dimension is sampled discretely by ordinal nodes with the associated memory nodes. The coupling structure between ordinal nodes brings about the propagation along a sequence of states. The next node to be activated may be selected according to a fixed serial order, as illustrated in Fig. 5 (Sandamirskaya & Schöner, 2010). This ordinal representation is a dynamic version of positional encoding of the sequence, as per Henson (1998)'s classification of mechanisms for serial order. Another possible mechanism for selecting the next ordinal node relies on a set of rules of behavioral organization that are instantiated as a pattern of coupling to specialized dynamical nodes (Sandamirskaya, Richter, & Schöner, 2011).

On the bottom left of Fig. 5, an intention field is shown. The intention field(s) may be spanned over spaces of perceptual, motor, or cognitive parameters that are instructive for the possible actions or cognitive operations. Positive activation in the intention field generates the behavior specified by the intention. In particular, this activation gets the actual physical effector and/or the associated sensor systems to produce an overt (e.g., motor) action, or provides input to down-stream structures that results in a cognitive operation.

The activation in the intention field also pre-activates the third element of the sequence generating model, depicted on the right in Fig. 5. This critical element is a neural representation of the condition-of-satisfaction. We have borrowed this term from Searle's (1983) theory of intentionality; it indicates that this neuronal dynamics detects a match between the state that corresponds to fulfillment of the current intention and the perceived state of the environment or the agent. The activated condition-of-satisfaction triggers an instability within the ordinal system, which marks off a sequential transition, as described bellow.

In the following paragraphs, we step through the dynamics of the model as a sequential transition takes place. Within the ordinal system, global inhibition ensures that only one ordinal node may be active at a time. Each ordinal node (bottom row in Fig. 5) is associated with a memory node (top row), which sustains activation even after steps in the sequence have been performed. An active memory node provides excitatory input to the successor ordinal node, biasing the competition during the transition phase between two steps in a sequence in favor of the successor node.

Activation within the ordinal system is transduced to the intention field through adjustable connection weights that hold the memory for items within the sequence and their associated ordinal positions (for a different mechanism to encode the order of the items, see Sandamirskaya et al., 2011). Activation of an ordinal node leads to the creation of a peak in the intention field in a detection instability. The location of that peak, i.e. the specific content of the current intention, is specified by the connection weights. These weights may have been learned, or "memorized", in a one-shot learning trial. The projection may also result from the integration of top-down input from the ordinal system with bottom-up information from the sensory surface.

The condition-of-satisfaction system receives "top-down" input from the intention field. This input preshapes the condition-of-satisfaction field to be more sensitive to the sensory input characterizing the accomplishment of the intended operation. It allows the formation of peaks in the condition-of-satisfaction field only when matching bottom-up input is received. Such bottom-up input may come from the sensory system or it may (e.g., in the case of a purely mental cognitive action) derive from other internal representations; e.g., as a "peak-detector" that signals that an instance in an appropriate target structure has been induced, or as a signal emitted by an oscillator that was triggered by the activated intention field.

When a peak is induced in the condition-of-satisfaction system, it inhibits the entire ordinal system. In a cascade of reverse detection instabilities, the current intention is deactivated: the currently active ordinal node is inhibited, the peak in the intention field yields a forgetting instability as input from the ordinal system is removed, and the peak in the condition-of-satisfaction system decays as input from the intention field is removed. Finally, deactivation of the condition-of-satisfaction field releases the ordinal system from inhibition. In the ensuing competition for activation, the successor of the last activated ordinal node gets the competitive advantage because of input from the memory nodes. The freshly activated ordinal node induces a peak in the intention field. The system then acts according to the new intention, until a sensory event or internal signal triggers the next transition.

We have shown that this mechanism enables the generation of sequential actions in embodied systems, in which unpredictable delays between the instantiation of an intention and the associated condition-of-satisfaction must be sustained (Sandamirskaya & Schöner, 2010). This mechanism may also support internal transitions within a purely mental chain of events (Schneegans & Sandamirskaya, 2011).

4. Exemplary DFT architectures

To illustrate how the mechanisms of DFT for creating instances, for operating on them, and for generating sequences of cognitive acts can be put to work, we review three architectures in which these functions play a role. These illustrative models are still relatively modest in complexity, but make a significant step in the direction of higher cognition while still working with real sensory inputs and generating output that may drive physical effectors, demonstrating the physical embodiment of these architectures.

4.1. Scene representation

As you read this, you might be sitting at your desk. A number of objects might be distributed over your working surface. Some may have been sitting there for a long time, others have been moved recently like the cup of coffee you just drank from. Reaching for the cup is absolutely effortless: you direct your gaze at its remembered location, reach and grasp. Orienting toward other objects, say your pen, may require a quick visual search guided by your knowledge of the gist of the scene. Your visual experience seamlessly integrates specific and current information (e.g., exactly what that particular cup looks like), situational information (e.g., that you are sitting in a particular pose at your desk), and long-term knowledge (e.g., what cups look like in general). The ease with which you acquire such a scene representation belies the tasks inherent complexity and attentional demands (Henderson & Hollingworth, 1999). Acquiring scene representations useful for action is a critical bottleneck for autonomous robots that is currently addressed in terms of semantic maps (Meger et al., 2008) and active vision (Rasolzadeh, Björkman, Huebner, & Kragic, 2010).

The process of acquiring a scene representation has at its core mechanisms for visual exploration and for entering perceptual information about objects in the scene into working (and then long-term) memory. Fig. 6 illustrates that core portion of a DFT architecture (Zibner, Faubel, Iossifidis, & Schöner, 2011a) in which pre-object blobs of localized attention control how feature information is

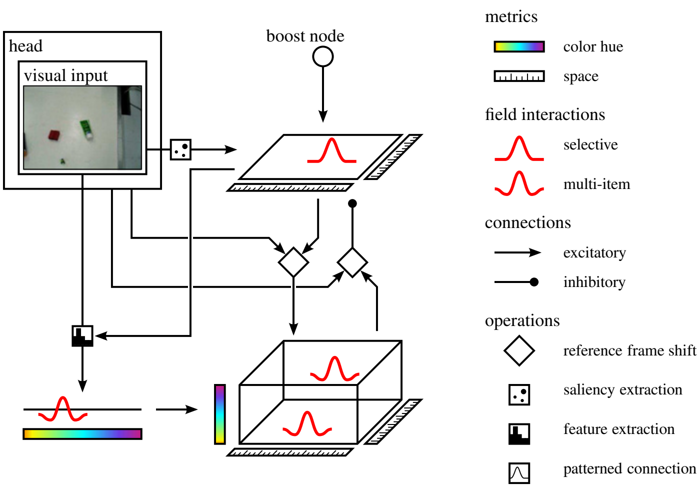

Fig. 6. *An architecture of scene representation.* In this figure, a three-dimensional space-color field is coupled to two separate representations of object position and color. Through ridge-like input, associative working memory peaks are created for objects contained in a scene. Both separate representations receive input from the sensory surface. Spatial input is preprocessed by a saliency operation, whereas a spatial selection limits the region of color extraction. A single neuron that influences a selection decision in the neural field representing space and inhibiting input coming from already represented objects are key components of visual exploration and creating the scene representation.

Please cite this article in press as: Sandamirskaya, Y., et al., Using Dynamic Field Theory to extend the embodiment stance toward higher cognition, *New Ideas in Psychology* (2013), http://dx.doi.org/10.1016/j.newideapsych.2013.01.002

acquired and linked to spatial information, consistent with the principles of feature integration theory (Johnson et al., 2008; Treisman, 1998).

Visual input modulated by saliency (Itti, Koch, & Niebur, 1998) is transmitted to a two-dimensional neural field (top-right) that represents retinal visual space and operates in the selection regime in which no more than a single peak can be stabilized at a time. The selection decision is triggered by a homogeneous boost generated from a single node (top); it induces a single peak at the most salient location, while all other locations are suppressed by lateral inhibition. This peak narrows the region of interest in the visual array from which visual features (here: color) are extracted and fed into a one-dimensional feature field (bottom left) where a peak is formed over the hue value that is most frequent in the selected image patch. This mechanism of modulating feature extraction by the selected localized attention peak emulates a visual fovea and the associated (albeit covert) shift of attention.

Bringing a single visual object into the foreground now enables the inscription of visual information in the scene representation, whose substrate is a three-dimensional neural field defined over two spatial dimensions and one feature dimension (bottom right). The space-color field is defined in an allocentric reference frame, in which objects are most likely to remain stationary. Spatial input from the two-dimensional spatial selection field must therefore be transformed from the retinal to the allocentric frame, using a transformation mechanisms of the type presented in the previous section. The transformation is modulated by the way that the current view is anchored in the visual surround (here defined by the table). It depends on gaze parameters that are also controlled autonomously in the more complete architecture (Zibner et al., 2011a). The correctly transformed spatial information about the pre-object currently in the foreground defines a two-dimensional location in the three-dimensional field; that is, it forms a tube of activation. The tube is combined with the associated feature information through input from the feature field (arrow from bottom left to bottom right), which activates a slice within the three-dimensional field. Where the tube and slice intersect, the field goes through the detection instability and generates a blob of activation localized in three dimensions. This is the three-dimensional variant of the mechanism discussed around Fig. 3.

This process can be repeated to visually explore the scene and, one-by-one, enter localized blobs of activation representing the different salient pre-objects in the visual array. To ensure that a new object is selected at each step of such a sequential scan, the objects that have already been entered into the scene representation project localized inhibition onto the corresponding spatial locations in the retinal selection field. This projection, naturally, must go through the inverse transformation from an allocentric to the retinal reference frame.

The full architecture also addresses active gaze changes (Zibner et al., 2011a), the updating of working scene memory when items shift (Zibner, Faubel, Iossifidis, Schöner, & Spencer, 2010; multi-item tracking), and top-down queries to the scene representation to bring specified items into the foreground (Zibner, Faubel, & Schöner, 2011b).

4.2. Perceptually grounded spatial language

Spatial descriptions of the form "The keys are to the right of the cup" are a natural means of guiding attention or action toward certain locations in everyday communication. Understanding spatial language may provide theoretical insights into the coupling between the metric spatial representations obtained from the sensorimotor system and the symbolic language descriptions central to cognition. It has, therefore, attracted considerable attention from psychologists and cognitive scientists (Logan, 1995; Logan et al., 1996; Regier & Carlson, 2001). Spatial language also offers a convenient channel for human–robot communication in interactive robotic scenarios (Mavridis & Roy, 2006; Roy, Hsiao, & Mavridis, 2004; Steels & Loetzsch, 2008).

We describe a flexible architecture for the production and understanding of relational spatial descriptions (Lipinski, Schneegans, Sandamirskaya, Spencer, & Schöner, 2012). The system receives real-world visual input, typically of simple tabletop scenes, as well as a series of discrete signals that encode the elements of a language task. The system is then able to provide the spatial relation between two given objects ("where is the green item relative to the red one?"), identify an object specified by a spatial description ("what is above the blue item?"), or produce a full spatial description for a given object ("where is the yellow item?"). The system uses color as a simple feature to identify items (both in the task input and the response generation), and currently supports the spatial relations "left", "right", "above", and "below". The spatial representations formed in a task can also be used to guide a motor response in a robot, e.g., to point to an object at a verbally described location.

The architecture, illustrated in Fig. 7, contains two modules, each of which implements one of the elementary operations described in Section 3. The first module provides an association between the color of an item and its location in the image (or vice versa) through a biased competition mechanism, as detailed in Section 3.1. This module consists of the three-dimensional visual field, defined over two spatial dimensions and one color dimension, and a set of discrete color nodes (top row). The visual field receives direct visual input from a camera and represents the locations and colors of salient items in the scene. The set of color nodes provides language input and generates language output. Each node is associated with a particular range of hue values and is bidirectionally coupled with the corresponding color slices in the three-dimensional visual field. Interactions between the nodes implement competition, so that only one node can be activated at a time. Along the spatial dimensions, the visual field is coupled with two purely spatial fields that form part of the architecture's second module.

That second module implements a reference frame transformation, from the image frame to an object-centered frame. It comprises three fields each defined over two-dimensional space. The target field holds the location of the target object that is designated by the spatial relation. This field is in the reference frame of the camera image. The reference field, likewise defined over the space of the camera image, represents the location of the reference

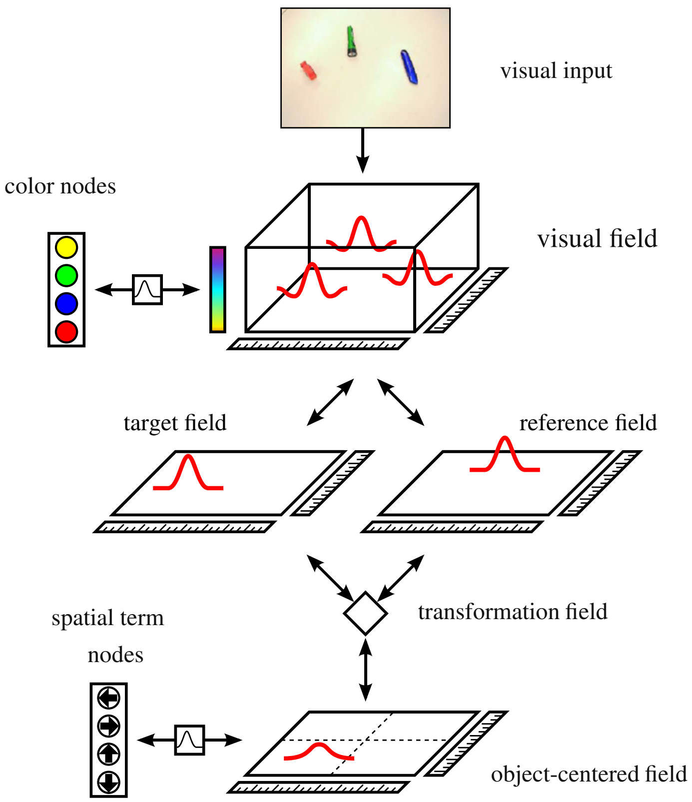

Fig. 7. *Architecture of the spatial language system.* The three-dimensional visual field at the top provides a simple DNF representation of the scene in the camera image and provides an association mechanism between item color and locations. Language input and output for item color are realized through the color nodes. The three two-dimensional fields and the higher-dimensional transformation field form a mechanism for reference frame shifts, which links the image frame to an object-centered frame of reference. The resulting object-centered representation is coupled to a set of spatial term nodes, which yield a description in language of the spatial relation between two items.

object relative to which a spatial relation is applied. The position of a peak in the third, object-centered field indicates the position of the target relative to the reference item (or, in some tasks, a target region relative to the reference object). The three fields are all bidirectionally coupled to a four-dimensional transformation field (Fazl, Grossberg, & Mingolla, 2009) that implements the reference frame shift from the image frame to the object-centered frame analogously to the transformations described in Section 3.2. Finally, the object-centered field is connected with another set of discrete nodes representing the different language terms for spatial relations. The connectivity pattern implements the semantics of each individual term. For instance, the node for "right" is linked to a broad region to the right of the object-centered field's midline.

To carry out the different tasks within a single architecture, the language inputs are given sequentially together with task-specific control inputs. This leads to a sequence of local decisions within the different fields, which together select a color or spatial relation term as the system's response. Sequential operation of this type makes it possible to process different items in the scene through the same neural structures, which we believe to be the most plausible neuronal account.

The three-dimensional visual field continually provides a representation of the visual scene. Before task input arrives, activation in all other fields is below the output threshold. To answer the question "where is the red item relative to the green one?" we create the following sequence of inputs and local decisions: First, we activate the "red" color node and apply at the same time a homogeneous boost input to the target field. The input from the color node strengthens the peak for any red object in the visual field, which in turn provides stronger input than objects of other colors to both the target and the reference field. By boosting the target field, we push it through a detection instability to form an activity peak from the localized input it receives. The inhibitory interactions in the field are set so only a single peak may form. This peak is sustained by excitatory interaction even after the boosting input it removed. Once a stable peak has formed in the target field, we select the reference object in an analogous fashion, by activating the "green" color node and boosting the reference field.

Once peaks of activation are present in both the target and the reference field, the transformation mechanism autonomously provides localized input to the object-centered field, that reflects the position of the selected target relative to the reference item. The induced activation drives the spatial term node that is connected to the region in which most activation lies. In a final step, we boost the spatial term nodes to enforce a selection decision between them, mediated by inhibitory interactions among these nodes.

The two other types of tasks are solved in a similar fashion through different sequences of inputs and boosts. For “what is above the blue item?” we first select the blue item as reference object in the manner described above. Then we activate the spatial term node for “above”, which generates an associated spatial semantic pattern in the object-centered field. In the presence of activation in the reference field and the object-centered field, the reference frame transformation is enabled and couples into the target field. The spatial semantic pattern is projected back onto the target field, centered on the position of the reference object. A region in the target field positioned above the location that corresponds to the reference object is therefore pre-activated. When the target field is boosted, an item from the scene that lies within this region will be selected and will grow a peak. The color that matches the selected location is obtained through the space-color association module. The spatial representation of the target item may also be directly coupled with a reaching or pointing system (Sandamirskaya, Lipinski, Iossifidis, & Schöner, 2010).

The third task – “where is the green item?” – is the most open-ended and requires the system to select an appropriate reference object for a spatial description and identify a matching spatial term. The task is carried out by combining elements of the first two. We begin by selecting the target object, the green item, in the target field as before. Then we give an equal boost to all the spatial term nodes and, at the same time, to the reference field. This initiates concurrent selection processes for a reference object location and for a spatial term; the two processes are continuously coupled through the transformation field. When a peak begins to form in the reference field at the location of a salient object, this peak supplies input to the object-centered field and biases the process of selecting a spatial term toward a matching choice, as in the first task. Concurrently, when the competition among the spatial term nodes begins to tend toward a particular term, the associated activation influences the processes of selecting the reference object. Once a combined decision has been made, the color of the selected reference object can be determined and given as a response jointly with the selected spatial term.

Through the generation of different sequences of local decisions, the DFT architecture for spatial language offers considerable behavioral flexibility while grounding the processes in sensory data. While the architecture as a whole is tailored to a given set of tasks, its components may perform much more general kinds of operations such as space–feature associations and reference frame transformations, which are critical for a wide variety of different actions. By now it may be obvious how these can be linked to other pieces of the DFT architecture to supply perceptual knowledge.

The system has been successful both in robotic scenarios (Lipinski, Sandamirskaya, & Schöner, 2009) and in capturing key characteristics of human behavior (Lipinski et al., 2012). The model explains, for instance, how the salience of scene items and the match to relational spatial terms influence the selection of reference objects for spatial descriptions (Carlson & Hill, 2008). The model thus spans a bridge from the conceptual analysis of human spatial language behavior to neuronal accounts of the underlying cognitive processes and to real-world implementations of artificial spatial language engines.

4.3. Sequence generation

In the previous section, a sequence of boosts to different parts of the DFT spatial language architecture was introduced to process each task. To autonomously organize this sequence of boosts, each of them has to be represented as a stable state if it is to have an effect on the associated dynamic structure. A signal for the transition to the next boost has to be picked-up autonomously and stabilized to function reliably. Sequence generation is a core element of human and artificial cognition (Humphreys, Forde, & Francis, 2000) and has been extensively studied since the earliest days of research on human cognition (Lashley, 1951). However, even the most elaborated cognitive (Anderson et al., 2004; Botvinick & Plaut, 2006; Cooper & Shallice, 2000; Glasspool & Houghton, 2005; Grossberg, 1978) and neural (Beiser & Houk, 1998; Deco & Rolls, 2005) architectures for sequence generation remain unconstrained by embodiment considerations and are not suited to autonomous control of the cognitive operations that we discuss. Here, we introduce an illustrative architecture that demonstrates how sequences of stable states, which represent cognitive acts, operations, or complex motor actions, may be autonomously generated in the DFT framework. To further show how this architecture may be embodied, we choose a simple robotic example.

Fig. 8 shows a robotic architecture that employs the sequence generation mechanism described in Section 3.3 (Sandamirskaya & Schöner, 2010). The architecture implements a sequential search of colored objects by a robotic vehicle. The order in which these objects are searched is taught to the system by presenting colored objects in the desired order. During search, the robot roams an arena until it finds the first object of the currently sought color, resumes roaming until it finds an object of the next desired color, and so on. This toy task exemplifies a core demand for sequence generation in autonomous systems: stability against variable timing. The amount of time that each step in the sequence takes varies unpredictably, both during sequence learning and during sequence generation. In learning this is due to variations in the time interval during which an experimenter shows the colored object to the robot. During sequential color search it is due to the variable amount of time it takes to find an object of the appropriate color. At each step in the sequence, the system must retain the current state of either learning or sequence generation, until the subtask has been achieved (end of presentation detected or colored object found). This requires stabilizing the current state against distractor inputs.

The ordinal dynamics in this architecture (top-left on Fig. 8) consists of a set of bi-stable dynamic ordinal nodes, each connected to a memory node. An active ordinal node

Please cite this article in press as: Sandamirskaya, Y., et al., Using Dynamic Field Theory to extend the embodiment stance toward higher cognition, *New Ideas in Psychology* (2013), http://dx.doi.org/10.1016/j.newideapsych.2013.01.002

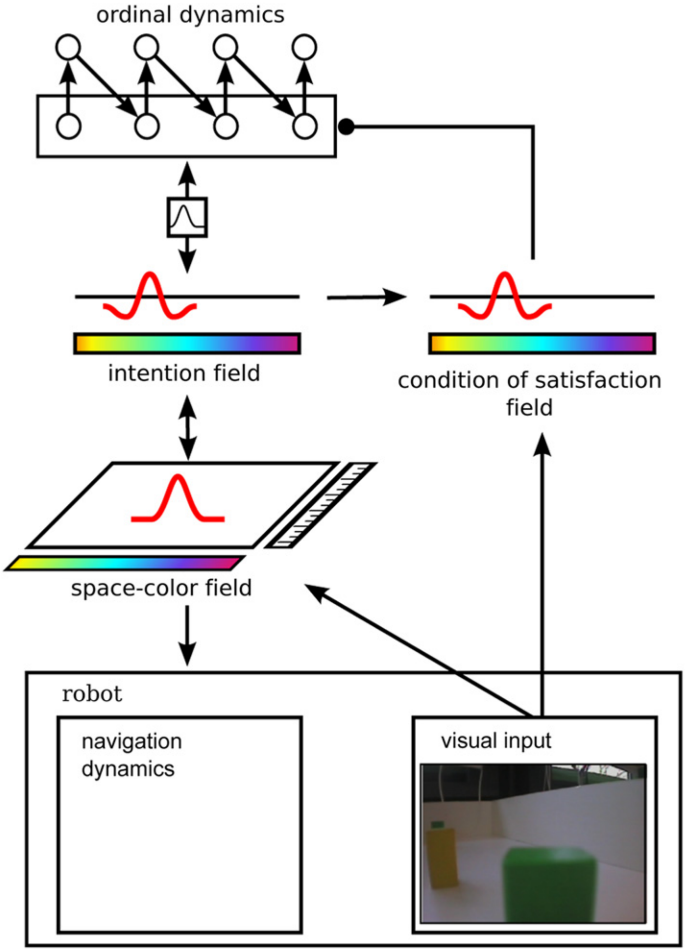

Fig. 8. *The architecture for a sequential color-search task on a Khepera robot.* An active node of the ordinal dynamics projects its activation onto an intention field, defined over color dimension. The intention field is coupled to the space-color field, which also receives visual input from the robot's camera. An activation peak in the space-color field drives the navigation dynamics of the robot, setting an attractor for its heading direction. The condition-of-satisfaction field is also defined over color dimension and is activated when the object of the currently active color takes up a large portion of the camera image.

excites the corresponding memory node, which, in its turn, provides an excitatory input to the ordinal node which is to be activated next. The active ordinal node also projects onto a single intention field defined over the dimension of color. Which color each node activates is learned, or memorized, in the training phase through a fast Hebbian learning mechanism. The intention field is reciprocally coupled with a two-dimensional space-color field, in which the spatial dimension samples the horizontal axis of the camera image. The space-color field receives ridge-input localized along the color dimension, but not along space, from the intention field. It also receives a two-dimensional space-color input from the visual array. Where visual input overlaps with the ridge, a peak is formed, the spatial projection of which specifies the visual angle under which an object of the color being sought is located.

The space-color field projects along the spatial dimension onto the dynamics of heading direction, creating an attractor that steers the robot to the detected object. As that object is approached, its image grows in the robot's visual array. The condition-of-satisfaction field (top-right on Fig. 8) is pre-activated by input from the intention field and is pushed through the detection instability when the object of the color being sought looms sufficiently large. This brings about the transition to the next step in the sequence as described in Section 3.3.

Before an object that matches the current intention has been found, no peak exists in the space-color field. The heading direction does not receive input at that time from the space-color field and the vehicle's navigation dynamics is dominated by obstacle avoidance, which is implemented using a standard dynamic method (Bicho, Mallet, & Schöner, 2000). This results in the roaming behavior that helps the robot search for objects of the appropriate color.

During teaching, the visual input from the object shown to the robot is boosted enough to induce a peak in the space-color field. This peak projects activation backwards onto the intention field, where a peak is induced at the location that

Please cite this article in press as: Sandamirskaya, Y., et al., Using Dynamic Field Theory to extend the embodiment stance toward higher cognition, *New Ideas in Psychology* (2013), http://dx.doi.org/10.1016/j.newideapsych.2013.01.002

corresponds to the prevailing color in the visual input. The co-activation of an ordinal node together with the intention field drives the Hebbian association between the ordinal node and the activated locations in the intention field. The system transitions to learning the next step in the sequence when the object is moved so close to the robot's camera that the condition-of-satisfaction system is activated.

Although this is still very much a toy demonstration of sequence generation, the same principles have been used to implement sequences of robotic actions (Sandamirskaya & Schöner, 2010), to behaviorally organize robotic actions (Sandamirskaya et al., 2011), and to control the dynamics of a complex cognitive task (Schneegans & Sandamirskaya, 2011).

5. Discussion

The embodiment hypothesis postulates that all cognition is embodied and situated. In other words, that all cognitive processes are at least potentially linked to sensory or motor surfaces. They are thus open to online updating, have graded state variables, and operate in a time-continuous fashion. Stability is, therefore, a core property of any cognitive process (Schneegans & Schöner, 2008). The neuronal substrate on which cognitive processes are based enables pervasive learning and development (Schöner, 2009).

In this paper, we did not prove the hypothesis, of course, nor did we provide empirical evidence supporting it. Instead, we showed that the theoretical framework of Dynamic Field Theory (DFT) is capable of reaching simple forms of higher cognition in the manner that the embodiment hypothesis requires. DFT has been developed and established as a neuronally based account for sensorimotor cognition, of which stability, online coupling to sensory information and to effector systems, learning, and development are prominent features. In DFT, cognitive functions such as detection and selection decisions, working memory, and change detection emerge from bifurcations, in which attractor states of the neural dynamics of activation fields becomes unstable and yields to new activity patterns that typically include localized peaks of activation.

We demonstrated how three key elements of general cognition emerge from these mechanisms. First, we illustrated how instances are created within DFT. In a perceptual and motor context, creating an instance amounts to bringing a perceptual or motor object into the foreground. In DFT, this is realized by inducing a self-stabilized peak through the detection instability. Generalizing, we illustrated how unspecific input modeled as a global boost – as ridge, tube, or slice input in one, two, or three dimensions – may push the field through the detection instability, inducing peaks at pre-activated locations. This mechanism amplifies small inhomogeneities in the field into macroscopic instances. It makes it possible to create instances from coupling among neuronal representations even when these just partially specify the contents of the instance to be created, as in cued recall from long-term memory (Zibner et al., 2011b). Such amplification of graded differences through instantiation lowers the demands on learning mechanisms.

Once self-stabilized peaks of activation are in place, neuronal projections among fields may bring about operations on these instances. We illustrated this for the case of changes of reference frame, an operation that takes two inputs (e.g., metric information to be transformed and a parameter of the transformation) to return a third (e.g., metric information in the new reference frame). The mechanism of instantiation by peak formation is critical to enabling targeted operations on representations. Only the neuronal connections between field locations over which a self-stabilized peak is positioned are effective – all other neuronal connections are, in effect, turned off. Consequently, it is possible to select from a set of operations one specific operation that will then be executed. In the case that we have illustrated, a dedicated transformation field adds a further dimension of control: by regulating whether the transformation field is capable of generating self-stabilized peaks, if effectively takes the transformation field in or out of the repertoire of a cognitive system (see Buss & Spencer, 2008; for an approach to cognitive control along similar lines).

We showed, as a third element of cognition, how a neuronal operation that transforms an intention into its condition-of-satisfaction can be organized in DFT. There are two sub-mechanisms. First, the neuronal representation of a current intention is projected both to the down-stream motor/cognitive system that brings about the desired result, and to a condition-of-satisfaction field that detects a match between the expected and the accomplished result. Second, this condition-of-satisfaction system, once activated, triggers a cascade of instabilities that lead the initial intention being suppressed, thus enabling the activation of a subsequent intention in a sequence of acts.

Outside the brief state transitions, neural fields are in stable states that resist change. When multiple fields are coupled, enabling online updating, the states persist, and with them the cognitive function persists that each field contributes. Only when fields are pushed through instabilities does the cognitive state change. That is why the instabilities are constitutive of the cognitive function for each field. This principle of stability punctuated by purposeful instabilities makes it possible to construct robust cognitive architectures, while at the same time affording continuous online coupling among fields and with sensory inputs. We illustrated this fact through three, still modestly complex, examples in scene representation, spatial language, and sequence generation. These examples highlight how DFT architectures enable autonomy – that is, continuous operation controlled entirely by the system's own behavior and sensory inputs. We highlighted flexible timing, in which transitions occur at whatever time is right given sensory feedback about the achievement of a subtask. We only briefly hinted at pervasive learning, enabled by the mechanism of the memory trace, and at development (for more on these topics see Spencer, Thomas, & McClelland, 2009). By implementing the architectures with real visual streams obtained from cameras on robotic systems, we demonstrated that DFT models are enactable as embodied, situated systems.

DFT is based on the concept of continuous space, as defined by sensory and motor surfaces as well as more abstract feature spaces. Is this a fundamental limitation of the forms of cognition DFT may reach? On the one hand, discrete categorical representations are not, per se, a problem. The mechanism of instantiation may, in fact, bring about effectively categorical behavior when the peak locations are dominated by inherent, potentially learned inhomogeneities rather than by external input. We have used discrete variants of neural fields that sample such categorically tuned neural fields with success (Faubel & Schöner, 2008).

On the other hand, the representations created within DFT are inherently low-dimensional: the activation fields depend on a limited number of metric dimensions. The number may be maximized by stacking fields that sample different feature spaces (Johnson et al., 2008), but fundamentally the dimensionality that counts is the dimensionality of the metric embedding space, not the dimensionality of the neural activation vector. In generic neural networks, by contrast, each neuron could be thought to span one dimension through its level of activation. A neuronal network of a few hundred neurons thus spans, in principle, a space with as many dimensions. What is the relationship between these two kinds of representations? At this point, we do not know for sure. The low-dimensionality of the representation in DFT comes from the structure of neuronal interaction, which is excitatory within the local environment as determined by a metric along the small number of dimensions. This requires that the neurons be embedded in a metric space in which this structure of interaction can be formulated. Conventional dynamic neuronal networks such as Hopfield nets (Hopfield, 1984) do not show this regular structure of interaction and thus do not have the same attractor states and the associated instabilities. The Hopfield network typically has a very large number of attractors, which cannot be made unstable by input in a targeted way. The fundamental mode of computation in a Hopfield network is therefore quite different from that of a neural field, based on the transient relaxation to the nearest attractor rather than on an attractor undergoing instabilities. Online updating, working memory, selection instabilities, etc., all remain unrealizable in the conventional framework.

Given how the structuring of neuronal interaction through the neighborhood relationship is key to the dynamic properties of neural fields, it may be that the requirement could be relaxed from embedding in a low-dimensional metric space to embedding in a topological space. This is one route we are currently exploring.

An alternative to DFT principles is to implement cognitive mechanisms of instantiation, binding, operations, etc., directly on the high-dimensional activation vectors of conventional neuronal networks. An attempt to do so is consistent with DFT when the embedding space is low-dimensional, but will establish these neuronal operations in different forms – not through attractors and their instabilities – when the space is high-dimensional (Stewart, Choo, & Eliasmith, 2010). Such work is based on the mathematics of Vector Symbolic Architectures that exploit the particularities of high-dimensional spaces (Gayler, 2004; Plate, 2003). Whether Vector Symbolic Architectures are compatible with the demands of embodiment and how they relate to DFT the future will show.

Within DFT, the limit of what can be achieved in terms of higher cognition has certainly not yet been reached. Our current research focuses on establishing what is entailed in expanding the complexity of architectures, and how more abstract functions such as problem solving may be addressed. A recent example of this line of work is our effort to move the sequence generation work toward a general account of behavioral organization (Sandamirskaya et al., 2011).

Acknowledegements

The authors gratefully acknowledge the financial support of the EU Seventh Framework Program FP7-ICT-2009-6 under Grant Agreement no. 270247, STREP NeuralDynamics.

References

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics, 27*, 77–87.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience, 20*, 303–330.

Anderson, M. L. (2003). Embodied cognition: a field guide. *Artificial Intelligence, 149*, 91–130.

Anderson, J., Bothell, D., Byrne, M., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*, 1036–1060.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Brain and Behavioral Sciences, 20*, 723–767.

Bar, M. (Ed.), (2011). *Predictions in the brain: Using our past to generate a future*. Oxford University Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, Cambridge Journals Online, 22*, 577–660.

Bastian, A., Schöner, G., & Riehle, A. (2003). Preshaping and continuous evolution of motor cortical representations during movement preparation. *European Journal of Neuroscience, 18*, 2047–2058.

Beiser, D. G., & Houk, J. C. (1998). Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. *Journal of Neurophysiology, 79*, 3168–3188.

Bicho, E., Mallet, P., & Schöner, G. (2000). Target representation on an autonomous vehicle with low-level sensors. *The International Journal of Robotics Research, 19*, 424–447.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: a recurrent neural network model. *Psychological Review, 113*, 201–233. http://dx.doi.org/10.1037/0033-295X.113.2.201.

Brooks, R. A. (1990). Do elephants play chess? *Robotics and Autonomous Systems, 6*, 3–15.

Buss, A., & Spencer, J. P. (2008). The emergence of rule-use: a dynamic neural field model of the dccs. In B. C. Love (Ed.), *Proceedings of cognitive science* (pp. 463–468). Austin, TX: Cognitive Science Society.

Carlson, L., & Hill, P. (2008). Processing the presence, placement, and properties of a distractor in spatial language tasks. *Memory & Cognition, 36*, 240–255.

Chemero, A. (2009). *Radical embodied cognitive science*. The MIT Press–A Bradford Book.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Reviews of Neuroscience, 22*, 319–349.

Coombes, S. (2005). Waves, bumps, and patterns in neural field theories. *Biological Cybernetics, 93*, 91–108. http://dx.doi.org/10.1007/s00422-005-0574-y.

Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology, 17*, 297–338.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press.

Deadwyler, S. A., & Hampson, R. E. (1995). Ensemble activity and behavior: what's the code? *Science, 270*, 1316–1318.

Deco, G., & Rolls, E. T. (2005). Sequential memory: a putative neural and synaptic dynamical mechanism. *Journal of Cognitive Neuroscience, 17*, 294–307.

Dehaene, S. (1997). *The number sense*. Oxford: Oxford University Press.

Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience, 4*, 826–831.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193–222. http://dx.doi.org/10.1146/annurev.ne.18.030195.001205.

Eliasmith, C. (2005). A unified approach to building and controlling spiking attractor networks. *Neural Computation, 17*, 1276–1314.

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review, 109*, 545–572.

Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Reports on Progress in Physics, 61*, 353–430.

Faubel, C., & Schöner, G. (2008). Learning to recognize objects on the fly: a neurally based dynamic field approach. *Neural Networks, 21*, 562–576.

Fazl, A., Grossberg, S., & Mingolla, E. (2009). View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology, 58*, 1–48.

Fix, J., Vitay, J., & Rougier, N. (2007). *Anticipatory behavior in adaptive learning systems. Chapter A Distributed computational model of spatial memory anticipation during a visual search task*. Berlin, Germany: Springer.

Gayler, R. (2004). *Vector symbolic architectures answer jackendoffis challenges for cognitive neuroscience*. In arXiv:cs/0412059v1.

Georgopoulos, A. P. (1995). Current issues in directional motor control. *Trends in Neurosciences, 18*, 506–510.

Glasspool, D. W., & Houghton, G. (2005). Serial order and consonant-vowel structure in a graphemic output buffer model. *Brain and Language, 94*, 304–330.

Grossberg, S. (1978). Behavioral contrast in short-term memory: serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology, 17*, 199–219.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243–271.

Henson, R. N. (1998). Short-term memory for serial order: the start-end model. *Cognitive Psychology, 36*, 73–137.

Henson, R., & Burgess, N. (1997). *Representations of serial order*.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences (USA), 81*, 3088–3092.

Hubel, D. H. (1988). *Eye, brain, and vision*. New York: Scientific American Library.

Humphreys, G. W., Forde, E. M. E., & Francis, D. (2000). The organization of sequential actions. In S. Monsell, & J. Driver (Eds.), *Control of cognitive processes — Attention and performance XVIII* (pp. 427–442). Cambridge, MA: MIT Press.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 1254–1259.

Jancke, D., Erlhagen, W., Dinse, H. R., Akhavan, A. C., Giese, M., Steinhage, A., et al. (1999). Parametric population representation of retinal location: neuronal interaction dynamics in cat primary visual cortex. *Journal of Neuroscience, 19*, 9016–9028.

Jeannerod, M., & Decety, J. (1995). Mental motor imagery: a window into the representational stages of action. *Current Opinion in Neurobiology, 5*, 727–732.

Johnson, J. S., Spencer, J. P., & Schöner, G. (2008). Moving to higher ground: the dynamic field theory and the dynamics of visual cognition. *New Ideas in Psychology, 26*, 227–251.

Johnson, J. S., Spencer, J. P., & Schöner, G. (2009). A layered neural architecture for the consolidation, maintenance, and updating of representations in visual working memory. *Brain Research, 1299*, 17–32.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science, 302*, 1181–1185.

Kopecz, K., & Schöner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural, dynamic fields. *Biological Cybernetics, 73*, 49–60.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. Oxford: Oxford University Press.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–146). New York: Wiley.

Lipinski, J., Sandamirskaya, Y., & Schöner, G. (2009). Swing it to the left, swing it to the right: enacting flexible spatial language using a neurodynamic framework. *Cognitive Neurodynamics, 3*, 373–400.

Lipinski, J., Schneegans, S., Sandamirskaya, Y., Spencer, J., & Schöner, G. (2012). A neuro-behavioral model of flexible spatial language behaviors. *Journal of Experimental Psychology: Learning, Memory, and Cognition., 38*, 1490–1511.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology, 28*, 103–174.

Logan, G., Sadler, D., Bloom, P., Peterson, M., Nadel, L., & Garrett, M. (1996). A computational analysis of the apprehension of spatial relations. *Language and Space Language Speech and Communication*, 493–529.

Mavridis, N., & Roy, D. (2006). Grounded situation models for robots: where words and percepts meet. In *2006 IEEE/RSJ international conference on intelligent robots and systems* (pp. 4690–4697).

Meger, D., Forssén, P., Lai, K., Helmer, S., McCann, S., Southey, T., et al. (2008). Curious george: an attentive semantic robot. *Robotics and Autonomous Systems, 56*, 503–511.

Piaget, J. (1952). *The origins of intelligence in children*.

Plate, T. (2003). *Holographic reduced representations*. Stanford, CA: CSLI Publication.

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience, 1*, 125–132.

Pouget, A., Deneve, S., & Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representation. *Nature Reviews Neuroscience, 3*, 741–747.

Pouget, A., & Sejnowski, T. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience, 9*, 222–237.

Rasolzadeh, B., Björkman, M., Huebner, K., & Kragic, D. (2010). An active vision system for detecting, fixating and manipulating objects in the real world. *The International Journal of Robotics Research, 29*, 133.

Regier, T., & Carlson, L. (2001). Grounding spatial language in perception: an empirical and computational investigation. *Journal of Experimental Psychology: General, 130*, 273–298.

Riegler, A. (2002). When is a cognitive system embodied? *Cognitive Systems Research, 3*, 339–348.

Roy, D., Hsiao, K., & Mavridis, N. (2004). Mental imagery for a conversational robot. *IEEE Transactions on Systems, Man, and Cybernetics– Part B: Cybernetics, 34*, 1374–1383.

Sandamirskaya, Y., Lipinski, J., Iossifidis, I., & Schöner, G. (2010). Natural human-robot interaction through spatial language: a dynamic neural fields approach. In *19th IEEE international symposium on robot and human interactive communication, RO-MAN* (pp. 600–607). http://dx.doi.org/10.1109/ROMAN.2010.5598671, Viareggio, Italy.

Sandamirskaya, Y., Richter, M., & Schöner, G. (2011). A neural-dynamic architecture for behavioral organization of an embodied agent. In. *IEEE international conference on development and learning and on epigenetic robotics (ICDL EPIROB 2011), 2* (pp. 1–7). IEEE.

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

Schneegans, S., & Sandamirskaya, Y. (2011). A neurodynamic architecture for the autonomous control of a spatial language system. In *IEEE international conference on development and learning and on epigenetic robotics (ICDL EPIROB 2011)*.

Schneegans, S., & Schöner, G. (2008). Dynamic field theory as a framework for understanding embodied cognition. In P. Calvo, & T. Gomila (Eds.), *Handbook of cognitive science: An embodied approach* (pp. 241–271). Elsevier Ltd.

Schneegans, S., & Schöner, G. (2012). A neural mechanism for coordinate transformation predicts pre-saccadic remapping. *Biological Cybernetics*, 1–21.

Schöner, G. (2009). Development as change of system dynamics: stability, instability, and emergence. In J. P. Spencer, M. Thomas, & J. McClelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 25–47). Oxford, UK: Oxford University Press.

Schreiner, C. E. (1995). Order and disorder in auditory cortical maps. *Current Opinion in Neurobiology, 5*, 489–496.

Schutte, A. R., & Spencer, J. P. (2002). Generalizing the dynamic field theory of the A-not-B error beyond infancy: three-year-olds' delay- and experience-dependent location memory biases. *Child Development, 73*, 377–404.

Searle, J. R. (1983). *Intentionality – An essay in the philosophy of mind*. Cambridge University Press.

Searle, J. R. (2004). *Mind: A brief introduction*. Oxford, UK: Oxford University Press.

Sergio, L. E., & Kalaska, J. F. (1998). Changes in the temporal pattern of primary motor cortex activity in a directional isometric force versus limb movement task. *Journal of Neurophysiology, 80*, 1577–1583.

Simmering, V. S., Schutte, A. R., & Spencer, J. P. (2008). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Research, 1202*, 68–86.

Please cite this article in press as: Sandamirskaya, Y., et al., Using Dynamic Field Theory to extend the embodiment stance toward higher cognition, *New Ideas in Psychology* (2013), http://dx.doi.org/10.1016/j.newideapsych.2013.01.002

Sparks, D. L. (1999). Conceptual issues related to the role of the superior colliculus in the control of gaze. *Current Opinion in Neurobiology, 9*, 698–707.

Spencer, J. P., Thomas, M. S. C., & McClelland, J. L. (Eds.), (2009). *Toward a unified theory of development*. Oxford University Press.

Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review, 102*, 503–532.

Spivey, M. J., & Dale, R. (2004). Psychology of learning and motivation. In B. H. Ross (Ed.). *Psychology of learning and motivation chapter on the continuity of mind: Toward a dynamical account of cognition, Vol 45* (pp. 87–142). Elsevier.

Steels, L., & Loetzsch, M. (2008). Perspective alignment in spatial language. In K. Coventry (Ed.), *Spatial language and dialogue*. Oxford: Oxford University Press.

Stewart, T. C., Choo, X., & Eliasmith, C. (2010). Symbolic reasoning in spiking neurons: a model of the cortex/basal ganglia/thalamus loop. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 1100–1105). Austin, TX: Cognitive Science Society.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. (2001). The dynamics of embodiment: a field theory of infant perseverative reaching. *Brain and Behavioral Sciences, 24*, 1–33.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, Massachusetts: The MIT Press, A Bradford Book.

Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience, 13*, 256–271.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society (London) B Biological Sciences, 353*, 1295–1306.

Wilimzig, C., Schneider, S., & Schöner, G. (2006). The time course of saccadic decision making: dynamic field theory. *Neural Networks, 19*, 1059–1074.

Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik, 13*, 55–80.

Zibner, S. K. U., Faubel, C., Iossifidis, I., & Schöner, G. (2011a). Dynamic neural fields as building blocks for a cortex-inspired architecture of robotic scene representation. *IEEE Transactions on Autonomous Mental Development, 3*, 74–91.

Zibner, S. K. U., Faubel, C., Iossifidis, I., Schöner, G., & Spencer, J. (2010). Scenes and tracking with dynamic neural fields: how to update a robotic scene representation. In N. Butko (Ed.), *Proceedings of the 9th IEEE 2010 international conference on development and learning (ICDL2010)* (pp. 244–250).

Zibner, S. K. U., Faubel, C., & Schöner, G. (2011b). Making a robotic scene representation accessible to feature and label queries. In F. Nori (Ed.), *Proceedings of the first joint IEEE international conference on development and learning and on epigentic robotics, ICDL-EPIROB*. NY: IEEE.

Zipser, D., & Andersen, R. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature, 331*, 679–684.