



Cognitive Science 48 (2024) e13491

© 2024 The Author(s). *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13491

Neural Dynamic Principles for an Intentional Embodied Agent

Jan Tekülve, Gregor Schöner

Institute for Neural Computation, Ruhr-University Bochum

Received 26 November 2023; received in revised form 26 June 2024; accepted 2 August 2024

Abstract

How situated embodied agents may achieve goals using knowledge is the classical question of natural and artificial intelligence. How organisms achieve this with their nervous systems is a central challenge for a neural theory of embodied cognition. To structure this challenge, we borrow terms from Searle's analysis of intentionality in its two directions of fit and six psychological modes (perception, memory, belief, intention-in-action, prior intention, desire). We postulate that intentional states are instantiated by neural activation patterns that are stabilized by neural interaction. Dynamic instabilities provide the neural mechanism for initiating and terminating intentional states and are critical to organizing sequences of intentional states. Beliefs represented by networks of concept nodes are autonomously learned and activated in response to desired outcomes. The neural dynamic principles of an intentional agent are demonstrated in a toy scenario in which a robotic agent explores an environment and paints objects in desired colors based on learned color transformation rules.

Keywords: Autonomous agents; Embodied cognition; Situated cognition; Intentionality; Cognitive systems; Neural network modeling; Dynamical systems modeling

1. Introduction

How are neural processes organized to generate behaviors that embody cognitive function? In particular, how are sequences of motor behaviors coordinated with processes of active perception to enable an agent to direct actions at objects in its environment so as to achieve

Correspondence should be sent to Gregor Schöner, Faculty of Computer Science, Institute for Neural Computation, Ruhr-University Bochum, 44780 Bochum, Germany. E-mail: gregor.schoener@ini.rub.de

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

desired outcomes? Actions and mental acts may be driven both by current sensory input and by memory and knowledge. How may an agent make flexible use of these different sources of information as its own state in the world provides or removes access to relevant sensory input?

These are key questions that a pervasively neural account of cognition and of its role in generating intelligent behavior must address (Lake, Ullman, Tenenbaum, & Gershman, 2017). Classical and neural artificial intelligence (AI) address these questions in different ways. In classical AI, the underlying cognitive competences are described in computational terms, based on the notion of symbolic processing (Anderson, 2007; Fodor, Jerry, 1975; Laird, 2012; Newell, 1990). In neural AI, neural networks are structured such that they extract task-relevant information from sensory information (Mnih et al., 2013; Silver et al., 2016; Shridhar, Manuelli, & Fox, 2022). Neither line of thinking provides an account for how the actual decisions and sequences of mental or motor acts emerge from time-continuous neural processing. This paper is aimed to provide a neural process account for that very problem, exemplified in a toy scenario.

The central problem that such a neural process account must address is that of *stability*, the capacity to resist change under varying conditions including stochastic fluctuations. In the sensor-motor domain, sensory signals are noisy and vary systematically in dynamic environments. The physical state of an agent's actuators may be subject to disturbances. Resistance to these types of perturbations is a prerequisite for successful motor behavior. Cognitive processes that remain coupled to sensory-motor surfaces are subject to these same kinds of perturbations which they too must resist. Such processes include decisions to select, initiate, or sequentially organize actions and the related perceptual decisions about the objects to which these actions are directed. For example, the process of selecting the target of a reaching movement remains linked to the sensory surface as it may be updated anytime ("online") when the environment changes (Goodale, Péllisson, & Prablanc, 1986). Similarly, when perceptual or motor decisions are kept in working memory (Ballard, Hayhoe, & Pelz, 1995), the underlying neural activation is sustained as the sensory input is removed. Once the decisions are acted out, the sustained neural activation again links to the sensory-motor surfaces.

Stability is also critical, however, to cognitive processes that are not linked to the sensory-motor surfaces. The neural substrate supporting any specific mental state is typically richly connected to the neural substrate supporting other mental states. In any given situation, some of these other mental states may compete or interfere with the ongoing mental state, acting as a disturbance against which the ongoing mental state must be stabilized.

Classical notions of information processing and symbol manipulation have abstracted from these constraints, relegating the link of mental states to the sensory-motor surfaces to the separate problem of symbol grounding. The implied categorical difference between low-level sensory-motor cognition and high-level symbolic cognition is not well supported by behavioral or neural evidence. Sequences of processing steps construed as manipulations of symbols form the basis of higher cognition in this classical view. The neural processes supporting such steps are not well understood.

If mental states have stability properties, this is a challenge: To transition from one mental state to another in a sequence of processing steps, the ongoing mental state must first be released from stability. A central theme of this paper is how stability and dynamic instabilities

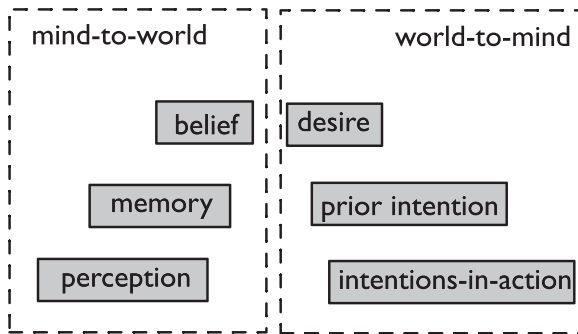


Fig. 1. Searle's six psychological modes in two directions of fit are used to organize the neural dynamic architecture of an intentional agent.

together enable the emergence of sequences of mental states at varying distances from the sensory-motor surfaces.

To talk about these problems and our approach to solving them, we found it useful to borrow terms from the philosophy of mind. We use the notion of *intentionality* that characterizes mental states as being *about* relevant states of the world (Crane, 2013; Searle, 1983, 2001) to refer to all mental states and their sensory-motor grounding that are required to generate intelligent behavior. Understanding the stability of mental states and their release from stability in sequences of processing steps profits, we will show, from the notion of *direction of fit*. Intentions in the *world-to-mind* direction of fit match the ordinary meaning of the “intention,” its “action” variant. A world-to-mind intention is about a state of the world that an action must achieve as characterized by its *condition of satisfaction*. Intended states of the world may involve objects (e.g., “paint that wall in blue”), parts of the agent’s own body (e.g., “make a pointing movement”), or even aspects of the agent’s mind (e.g., “bring this concept into attentional focus”). The *mind-to-world* direction of fit captures the “perceptual” flavor of intentionality, recognized mainly by philosophers. In mind-to-world intentions, states of the mind must match states of the world. Such intentional states may again reference objects (e.g., “perception of a green wall”), the agent’s own body (e.g., “proprioception of your arm’s posture”), and the agent’s mind (e.g., “recognizing that we remember something”).

We further borrow the notion of the six psychological modes specifically from Searle, who, to our knowledge, does not claim this as a deep philosophical concept by itself but uses it to systematically organize intentional states in terms of increasing distance from the sensory-motor surfaces (Fig. 1). The notion helps us explore the extent to which our analysis of the stability problem and its solution reaches all forms of mental and motor acts that are required to generate intelligent behavior. For the “motor” flavor of intentionality, this includes motor behaviors (intentions-in-action), action plans (prior intentions), and action goals (desires) where Searle’s terms are provided in parenthesis. For the “perception” flavor of intentionality, this includes perception, memory, and knowledge (beliefs).

We mathematically formalize these ideas in a toy scenario in which an autonomous robotic agent acts and autonomously learns to achieve goals. The agent is a vehicle with a robotic arm.

It perceives the world through a vision sensor and has a set of sensors to perceive the states of its own body. The environment consists of cubes that are either paint buckets or that function as canvases to be painted. The agent moves through space and points its arm at objects to pick up or to deposit paint. The agent learns color contingencies (which color results when applying a given coat of paint to a canvas of a given color) and uses that knowledge to achieve its goals, the simple desire to paint objects in a particular color.

The scenario touches on all six psychological modes in the two directions of fit. Mind-to-world intentional states include the perception of objects and body states, memory of the environment, and beliefs/knowledge about color contingencies. World-to-mind intentional states include the motor behaviors of driving and pointing to pick up or apply paint, prior intentions/action plans as sequences of such actions, and desires/goals as desires for a given color. In the scenario, individual actions or mental acts take variable amounts of time during which intentions must be stabilized against inputs from sensors or competing intentional states. The scenario thus probes the central issue, the stability of intentional states, and their release from stability to generate sequences of such states.

The neural processes account of intelligent behavior is formalized within Dynamic Field Theory (DFT; Schöner, Spencer, & DFT Research Group, 2016), a set of mathematical concepts that characterize the dynamics of activation within and across neural populations that are ultimately linked to sensory and motor systems (Schöner, 2019). This framework lends itself to the desired account because stability and its loss in dynamic instabilities are central concepts of DFT. Below, we outline the neural processing elements of DFT needed to generate intentional states in the two directions of fit and the six psychological modes. Next, we present a neural dynamic architecture that achieves that within the toy scenario. Simulations then illustrate how goal-oriented intelligent behavior emerges for architecture and its embodiment.

2. Elements for a neural dynamic account of intentionality

What would it mean to provide a neural account for intentionality? Different levels of description may be considered, from whole-brain architectures to detailed accounts of neural mechanism at the level of the cell. We hypothesize that the neural processes underlying behavior and cognition are best captured by the time- and state-continuous evolution of neural activation patterns in small neural populations across the brain (Gerstner, Kistler, Naud, & Paninski, 2014; Schöner, 2019). The mathematical framework of DFT (Schöner et al., 2016) formalizes this hypothesis. Its key postulates will be reviewed here. (1) Central to DFT is the postulate that cognitive processes may be linked to the sensory and motor surfaces, and may operate in closed loop when an agent is situated in a structured environment. This implies the need for stability. For instance, perceptual or motor decisions need to be stabilized against distractors or competing motor choices. Such stabilization is brought about by recurrent neural connectivity within local neural population.

(2) Local activation patterns that “stand for” particular intentional states are stabilized by excitatory recurrent connectivity within a neighborhood. Inhibitory connectivity to other

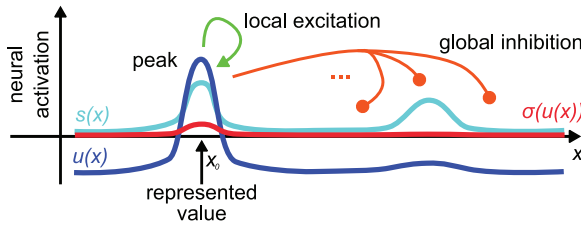


Fig. 2. A dynamic neural field with a suprathreshold activation peak representing a particular location, x_0 , along a continuous feature dimension, x .

neighborhoods stabilizes activation patterns against distractors. Such localist representations are natural for categories (Bowers, 2017). They may represent continua of intentional states as long as these may be described by low-dimensional feature spaces, whose representational power can be extended in neural dynamic architectures.

(3) Dynamic instabilities provide the mechanism for change. For example, an intentional state may be induced in the *detection instability* in which a subthreshold activation pattern becomes unstable, allowing the system to transition to a suprathreshold activation pattern. The inverse transition occurs in the form of the *reverse detection instability*. Physical or mental acts come about through organized dynamic instabilities that induce transitions from one set of neural intentional states to another.

We provide a brief technical review of these three postulates in their mathematical formalization.

2.1. Dynamic neural fields

The building blocks of DFT, dynamic neural fields, are neural activation functions, $u(x, t)$, defined over metric spaces, x , of limited dimensionality, typically not more than four. They evolve continuously in time, t , as described by a *neural dynamics*,

$$\tau \dot{u}(x, t) = -u(x, t) + h + \sum_i s_i(x, t) + \int \omega(x - x') \sigma(u(x', t)) dx'. \quad (1)$$

Originating from space-time continuous approximations of neural activity in the layered structures of cortex (Wilson & Cowan, 1973), these equations lift the dynamics of neural membranes to the population level (Amari, 1977). The stabilization term, $-u(x, t)$, time scale, τ , and resting level, $h < 0$, reflect membrane properties. The spiking mechanism is replaced, at the population level, by a sigmoid threshold function, $\sigma(u) = 1/(1 + \exp[-\beta u])$ which is passed on to other neural populations or ultimately on to motor systems. Recurrent connectivity is characterized by the interaction kernel, $\omega(x - x')$, which is excitatory ($\omega > 0$) for small and inhibitory ($\omega < 0$) for large distances, $x - x'$.

When localized inputs, $s_i(x, t)$, vary slowly, they induce a subthreshold solutions $\approx h + \sum_i s_i(x, t)$ that remain stable as long as activation remains below threshold. As the threshold is reached, this solution becomes unstable in the detection instability. The field switches to a localized peak of suprathreshold activation (Fig. 2). Depending on interaction parameters

and on the resting level, different *dynamic regimes* may arise. *Selective* fields generate single peaks. Multiple peaks may exist in the *self-stabilized* regime. These may persist when inducing localized input is removed in the *self-sustained* regime.

2.2. Neural dynamic architectures

Dynamic neural fields may represent feature dimension such as color, visual space, or movement parameters. The connectivity delivering sensory input to a field or projecting from a field to the motor system determines the meaning of the feature dimensions. This is exemplified when forward input directly induces activation peaks in the detection instability. When activation peaks arise in other ways, they continue to represent the dimensions defined by forward connectivity to and from a field. Peaks may, for instance, be re-activated from a memory trace, which is induced at activated locations in a simple form of learning (Chapter 2 of Schöner et al., 2016). Peaks may also arise in neural architectures from coupling among multiple fields. A source field, u_{src} , couples into a target field, u_{tar} , if the source field output, $\sigma(u_{\text{src}})$, provides input to the target field by adding to its rate of change, \dot{u}_{tar} , typically weighted through a projection kernel, $c_{\text{src} \rightarrow \text{tar}}$. A central postulate is that feature dimensions are preserved by such projection, so that fields retain their meaning as activation patterns across an architecture vary.

Inputs from or projections to fields representing multiple different feature dimensions may create conjoint representations. Perceptual features such as color, c , are represented jointly with visual space, x , itself. A peak in such a field thus represents an object of a particular color at a particular spatial location. When multiple neural fields representing different perceptual features all share visual space as a dimension, feature binding across space in the manner of Feature Integration Theory (Treisman & Gelade, 1980; Treisman & Zhang, 2006) becomes possible (Chapter 7 of Schöner et al., 2016). This enables integrated feature representations of objects that do not require implausibly large numbers of neurons to sample high-dimensional feature spaces.

The coupling kernels may contract the dimensionality of the source field by integrating over dimensions the target field does not share (Zibner & Faubel, 2015). Coupling kernels may also expand the dimensionality of the source field by providing input that is constant along dimensions of the target field that the source field does not share, providing ridge, tube, or slice input, for instance. Fig. 3 illustrates how expansion may assemble joint representations: Peaks in a field over visual space and in a field over color each provide ridge-input to a joint space-color field, creating a peak where these ridges overlap.

As zero-dimensional neural fields, *dynamic neural nodes* represent categorical states. Feature *concept* nodes derive their meaning from a localized pattern of input connectivity to a feature field. For example, a neural node representing the color concept “blue” projects onto a region centered around blue in a feature field defined over hue (Fig. 4). When activated, the “blue” concept node facilitates the generation of a peak localized in that blue region of hue space.

Boost nodes project homogeneously across an entire field, effectively modulating its resting level and potentially altering its dynamic regime (Fig. 4). Because boost nodes alter the

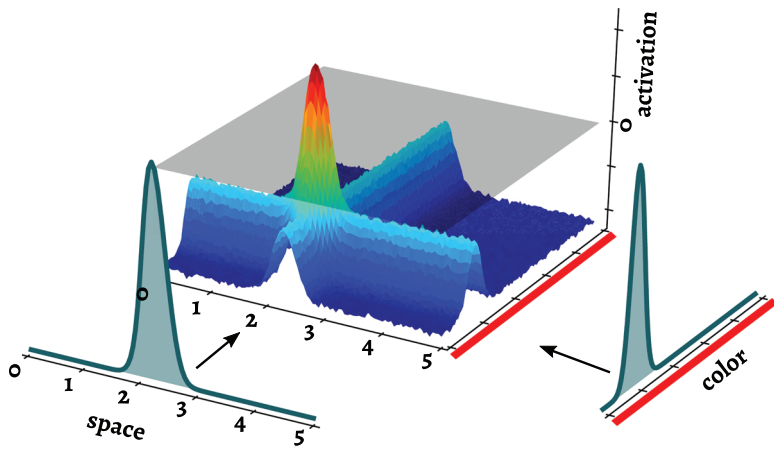


Fig. 3. Two one-dimensional ridge inputs form a two-dimensional activation peak at the location of their intersection.

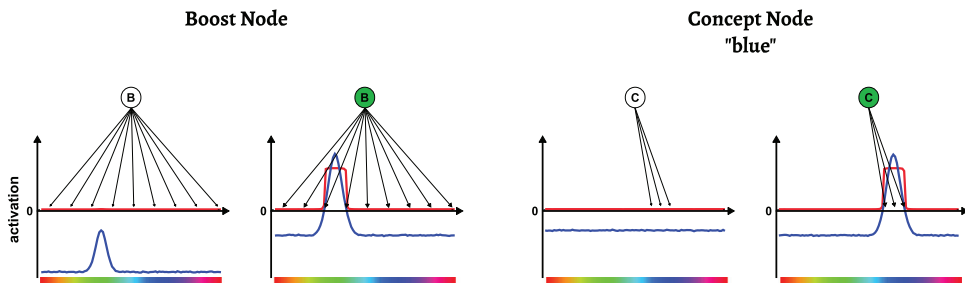


Fig. 4. **Left:** A subthreshold activation hill (left) may be pushed through the detection instability when a boost node becomes activated (filled circle on the right). **Right:** A concept node for “blue” projects onto the blue region of a color field (left). Its activation (filled circle on the right) may induce an activation peak centered on that region.

amount of localized input needed to reach the detection instability, they may gate the flow of activation within an architecture, enabling particular branches of coupling to induce peaks or preventing them from doing so. They may also act as “go” signals by providing the final element of input needed to induce a peak.

Peak detectors are dynamic neural nodes to which a source field projects homogeneously, the inverse direction of connectivity to that of boost nodes. Peak detectors are tuned so that they go through the detection instability whenever at least one activation peak arises in the source field. Peak detectors receiving input from multiple fields may detect combinations of peak states across these fields. *Match fields* generalize this idea by forming peaks only at locations at which input from different source fields metrically overlaps.

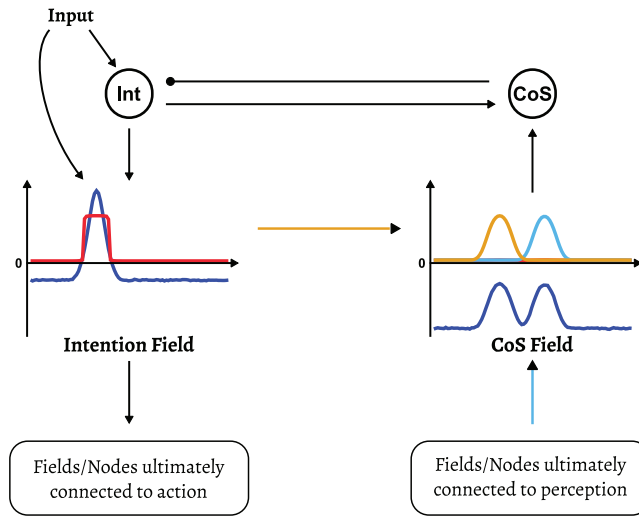


Fig. 5. The structure of a neural dynamic network that behaviorally organizes a world-to-mind intention.

2.3. Sequence generation

The generation of sequences of attractor states in DFT is based on a combination of boost nodes and peak detection in a match field. The relevant methods were developed in Sandamirskaya and Schöner (2010) (see also Chapter 14 of Schöner et al., 2016), already invoking notions borrowed from the philosophy of mind. World-to-mind intentional states are represented by intention nodes that act as boost nodes controlling the formation of peaks in *intention fields* which represent features of a behavior, the “contents” of the intention (Fig. 5). Intention nodes and fields project onto downstream neural dynamic fields and nodes that ultimately drive behavior. Intention nodes are paired with their *conditions of satisfaction* (CoS) represented by *CoS* nodes. Projections from an intention field to an associated *CoS* field predict the desired end-state in appropriate feature spaces. *CoS* fields act as match fields that compare the desired end-state of an intention to perceptual information from the sensory-motor surface or from other parts of the neural architecture. When these inputs match, a peak forming in the *CoS* field activates the corresponding *CoS* node that acts as a peak detector. The activated *CoS* node inhibits the intention node. As input from the intention to the *CoS* node falls away, the *CoS* node deactivates, effectively terminating the world-to-mind intention. What happens next depends on the overall neural architecture within which this network is embedded. The network thus provides an interface between world-to-mind intentional states and their embodiment in sensory-motor behavior, but may also steer the initiation and termination of purely mental neural processes that instantiate world-to-mind intentional acts. This is a first instance of cascades of instabilities in the dynamics of neural populations representing intentional and CoS states organize the time course of sequence generation.

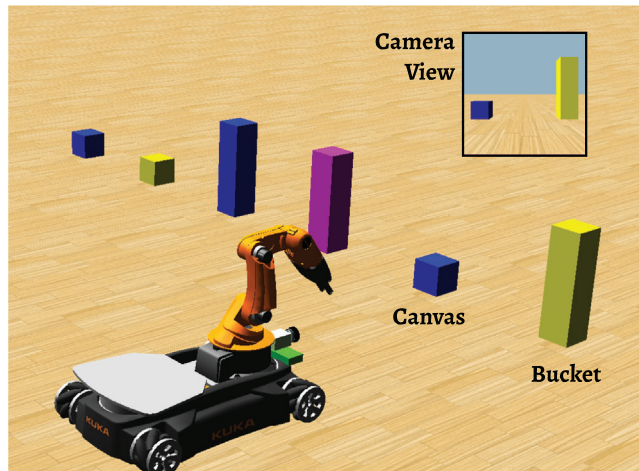


Fig. 6. The youBot agent in its environment of colored *buckets* and *canvases*.

3. Neural dynamic architecture of an intentional agent

In the following, we describe an exemplary neural dynamic architecture that captures the creation, stabilization, and termination of intentional states in the two directions of fit and six psychological modes, enabling an agent to act and learn autonomously in a multi-step task. We do so in a minimal toy scenario in which an intentional agent is situated in a virtual world that consists of small colored cubes (*canvases*) and tall colored cuboids (*buckets*; see Fig. 6).¹ The agent may collect paint from *bucket* objects, and then apply the paint to *canvas* objects. The way a canvas object changes color upon being painted is determined by fixed (and arbitrary) rules of color combination which the agent autonomously learns while exploring. The agent may also have fixed desires to paint a *canvas* in a specific color. It achieves such goals by using its knowledge about color combination.

The neural dynamic architecture and its connection to the agent's sensory-motor surfaces is sketched in Fig. 7. In the following, we step through that architecture following the two directions of fit, each in three psychological modes. We outline the conceptual commitments for each psychological mode and illustrate these by one exemplary component of each mode. More complete documentation of the architecture including code is available one, see note 2.

3.1. *Mind-to-world direction of fit*

A neural process account of mind-to-world intentional states must address how representations of particular states of affairs in the world are created, how they may persist or be recreated to influence the agent's mental or physical acts, and how they may eventually be allowed to decay. Mere transmission of sensory information in the manner of forward neural networks does not provide a sufficient account. Selective attention, its stabilization against distractor input, stabilization of activation when sensory information is removed, reactivation

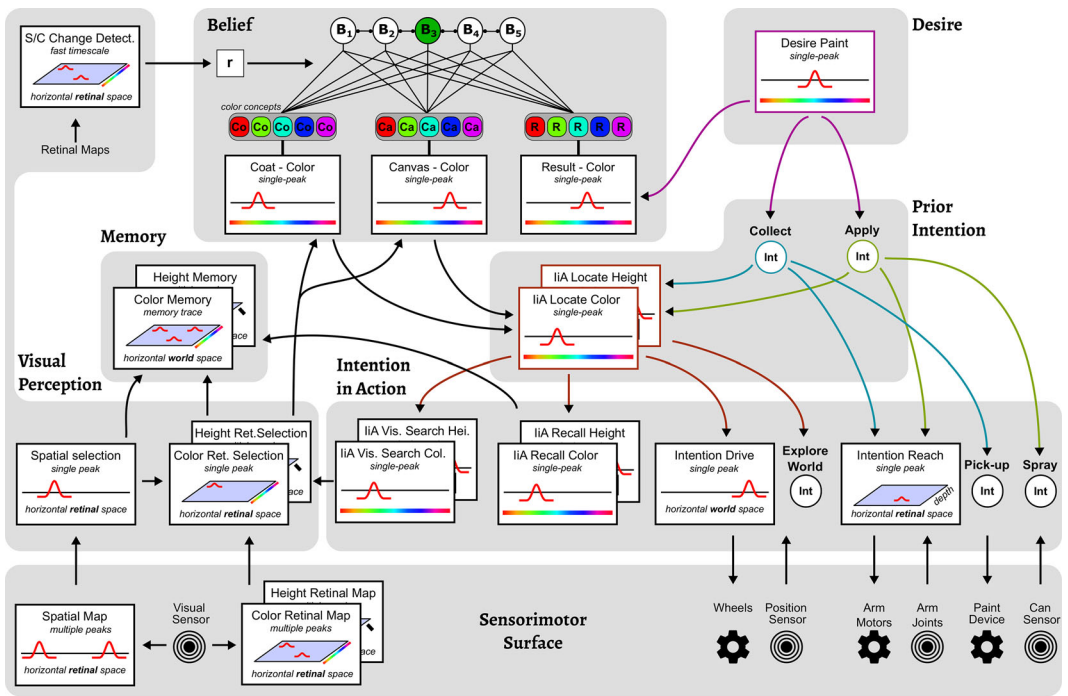


Fig. 7. Sketch of the dynamic field architecture controlling the robotic agent, organized by psychological mode. Stacks of fields represent two fields of analog functionality and connectivity defined over the different feature spaces color and height. Arrows highlight the most relevant connections between fields. Color coded arrows highlight action sequences and the components involved.

of attentional states on recall are examples of neural processes that must be addressed. In a sense, intentional states in the mind-to-world direction of fit could be viewed as a generalization of attention which may, for instance, include propositional representations about contingencies in the world. Mind-to-world intentional states thus engage active neural processes whose sequential temporal organization must be explained. We develop such an account by transforming the concept of a “condition of satisfaction” from a logical to a neural process notion. Where in the philosophy of mind, the condition of satisfaction characterizes whether the content of a mind-to-world intentional states is veridical, the neural process variant of the condition of satisfaction measures the match between a representational state and pertinent sensory information. We add the notion of a condition of dissatisfaction when a mismatch is detected instead.

3.1.1. Perception

Perceptual states are modeled by suprathreshold peaks of action in neural fields that represent visual features by virtue of the feed-forward connectivity from the sensory surface to the neural fields. A perceptual state is created in a detection instability in response to localized input. When this happens, the peak’s location matches the location (in features

space) of sensory input, so that the activation peak is at the same time the neural representation of a perceptual intentional state and a neural representation of its condition of satisfaction. Activation peaks may be sustained after the inducing sensory input is removed due to the pattern of neural interaction. Working memory is, in our interpretation of Searle's psychological modes, part of the mode of perception.

Activation peaks are selective due to lateral inhibitory interaction. So, only a limited number of field locations may become activated at any time (which reflects capacity in the case of working memory). Such selection reflects attention and may be driven by differences in input strength (saliency), but may also be guided "top-down" in visual search (Grieben et al., 2020) (which is itself a world-to-mind form of intention as discussed in the next subsection).

Activation peaks representing perceptual state are deactivated in the reverse detection instability. This may happen when top-down or localized bottom-up inputs are removed. It may also happen due to competing inputs near the capacity limit, or due to specific inhibitory input that reflects a condition of dissatisfaction (CoD) (e.g., as a form of change detection Johnson, Spencer, Luck, & Schöner, 2009; Chapter 6 of Schöner et al., 2016).

3.1.1.1. Example: Visual perception component of the neural architecture of intention: The agent perceives the state of its own body and its location in the world through proprioceptive sensors and efference copy. It perceives the environment visually. The network responsible for visual perception and memory is illustrated in Fig. 8. Visual input comes from the camera and is split into a purely spatial saliency representation (*Spatial Map*) and a pair of space/feature fields, the *Space/Color* and *Space/Height Retinal Map*, that combine feature with location information.

Percepts are formed when a localized peak is created in the *Spatial Selection* field. That peak represents spatial attention and further induces activation peaks in the *Space/Color* and *Space/Height Retinal Selection* fields which provide the perceived feature values at the attended location. In visual search, peaks in space/feature fields arise from top-down cues (shown in Fig. 7 but omitted in Fig. 8). Perceptual states are deleted when spatial attention shifts or when the attended object is removed from the visual array. In autonomous visual exploration, the system sequentially attends to objects in the visual array, creating feature representations and commits these to allocentric space/feature memory, and finally deletes the perceptual states by shifting spatial attention (see Section 4.1 for details).

3.1.2. Memory

In the model, memories are created by laying down memory traces of localized activation peaks. Once peaks have decayed, memory traces persist in the form of elevated, but subthreshold levels of activation at the locations of the peaks. Activation peaks can be reinstated at those locations by homogeneously boosting the entire field in free recall, or by providing input that is localized along a feature dimension and homogeneous along other dimensions in cued recall. Such peaks in the memory fields are the neural instantiation of both the intentional state of memory and of its condition of satisfaction.

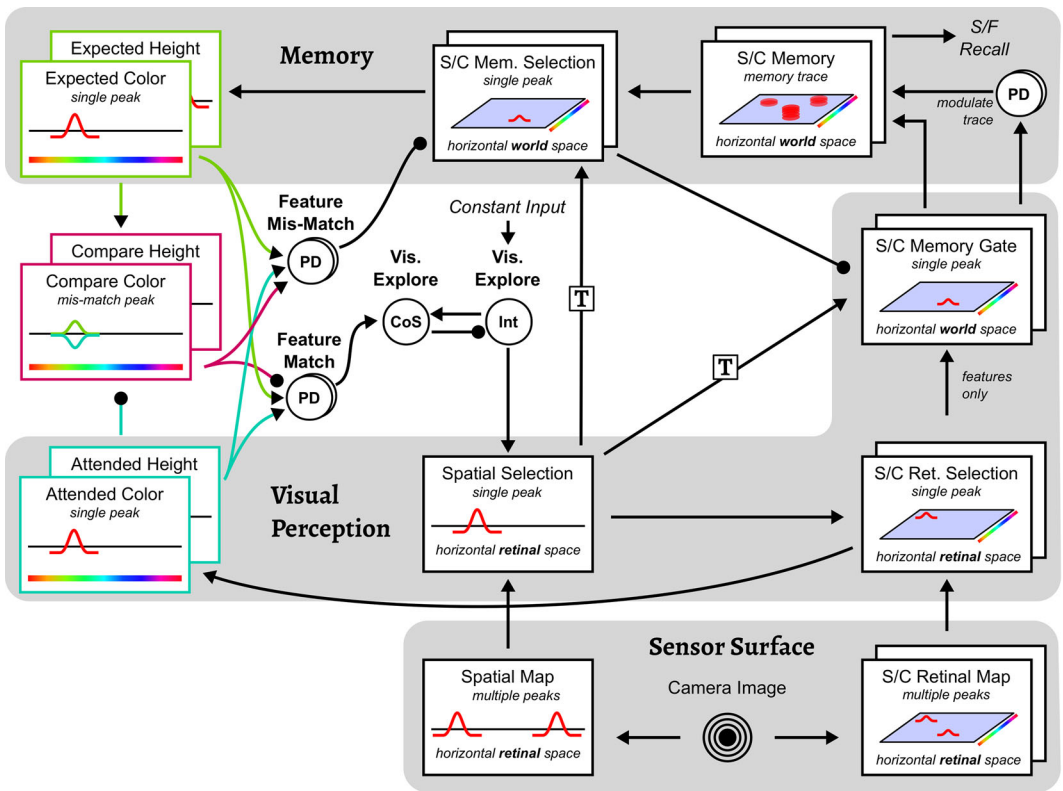


Fig. 8. Network of the dynamic field architecture realizing visual perception and memory. Arrows mark excitatory connections, lines ending in filled circles mark inhibitory connections. Boxes with the letter “T” mark neural coordinate transforms. Not all connections are shown.

3.1.2.1. Example: Scene memory: In the model, memories are formed about the location, color, and height of cuboids in the environment. While visual perception is anchored in camera coordinates, object memories are in an allocentric reference frame. Whenever an object is attentionally selected, a peak in the perceptual space/feature selection fields represents the object’s features bound to the object’s location in “retinal” (camera) coordinates (Fig. 8). The feature values are extracted and projected into the memory gate space/feature fields. This feature input is combined with a spatial cue obtained by transforming the camera coordinates from the spatial selection field into the allocentric coordinates frame. A peak arises in the memory gate space/feature fields in the manner illustrated in Fig. 3. This coordinate transform requires transcribing object information into memory one object at a time to solve the binding problem (Chapter 8 of Schöner et al., 2016). The peak in the memory gate field lays down a memory trace. Memory traces at other locations slowly decay, leading to forgetting through interference. A memory trace is reactivated in the space/feature memory selection fields through a coordinate-transformed spatial cue from the Spatial Selection field.

Alternatively, reactivation may come from slice a feature-cue that provides slice input into the space/feature selection field. Such cued recall is illustrated in Section 3.5.1.

3.1.3. *Belief*

Searle's psychological mode of *belief* does not directly correspond to a specific cognitive competence. We use this notion to reflect about neural processes that may support knowledge about the world that informs action, a central concern of classical artificial intelligence (Nilsson, 2014). To this end, we consider beliefs to be propositional in nature, so that they abstract from sensory-motor details, and generalize from the experience during which they were formed.

3.1.3.1. Example: Belief activation: The agent forms beliefs about the rules of color transformation by reference to three types of color concepts: the color of paint ("coat"), the color of a surface that may be painted ("canvas"), and the color of the surface that results from a painting action ("result") (center top panel in Fig. 7). Each concept is represented by a neural node that is perceptually grounded through reciprocal connections to regions in color feature fields (Fig. 4). Beliefs are also represented by neural does, which are grounded by connections to one color concept in each role. These connections emerge from Hebbian learning, while the system experiences different outcomes during painting actions. The dynamics of the autonomous learning of beliefs and the process organization have been reported in technical detail in Tekülve and Schöner (2020).

Once learned, a belief may become activated by a combination of top-down and bottom-up processes. The desire for a specific result color activates the result concept node top-down. The attentional selection of an object may activate either a paint or a canvas color concept node bottom-up. Once activated, the belief node may then activate the color concept node in the complementary role, which will subsequently guide search for a paint *bucket* or *canvas* object of matching color. An activated belief is thus a mind-to-world intentional state that may steer world-to-mind intentions. The condition of satisfaction of a belief is neurally represented by its own activation.

3.2. *World-to-mind direction of fit*

World-to-mind intentional states are representations of states of the world that the agent is aiming to bring about. The world state may include states of the agent's body and nervous system (other than the intentional state itself). The condition of satisfaction of world-to-mind intentional states is fulfilled when the real state of the world matches the intended state as a consequence of the agent's actions. To transform this logical conception of a condition of satisfaction into a neural processing concept, we neurally represent the condition of satisfaction. Contrary to mind-to-world intentions, in which the intention and its condition of satisfaction can be represented by the same activation states, world-to-mind intentions require separate neural activation states. The condition of satisfaction is activated when input from the intentional state that predicts that state's condition of satisfaction matches input from

mind-to-world intentions (Fig. 5). Once activated, the condition of satisfaction inhibits the intentional state.

A process account for world-to-mind intentions must also address the possibility that the intended world state is not achieved for some reason. We postulate that a neural representation of a CoD becomes activated that also inhibits the intentional state, bringing about the “unsuccessful” termination of the intention. Conditions of dissatisfaction may be activated by default when the condition of satisfaction fails to become activated within a given time frame, or may be activated by particular mind-to-world intentional states that detect a mismatch between predicted and perceived outcomes of an action.

3.2.1. *Intention-in-action*

Intentions-in-action are world-to-mind intentions that directly drive motor behavior or neural processes that support motor behavior. They are represented by neural activation states that directly project onto sensory-motor processes or that provide top-down influence onto mind-to-world processes to guide sensory-motor processes (bottom right in Fig. 7). These neural representations of intentions-in-action typically activate sensory-motor processes by providing boost input. The content of intentions-in-action are movement targets or cues to movement targets.

3.2.1.1. Example: Recall and drive: The network for the *drive* and the *recall* intentions-in-action is illustrated in Fig. 9. Both intentions are boosted by input from the *locate* intention that is higher in a hierarchy of intentions and ultimately originates from the currently active belief (see box for prior intentions in Fig. 7). When the *intention drive* node is activated, the robot vehicle moves from its initial position to the target position represented in the *drive intention* field. A smooth velocity profile is generated by a pair of coupled excitatory-inhibitory neural fields which generate a single cycle of neural oscillation (detail in Schöner, Tekülve, & Zibner, 2019). Input to these fields selects the duration and amplitude of the velocity profile. This input is from the distance to be covered, computed neurally by transforming the neural representation of the *Drive* target into a coordinate frame centered on the vehicle’s initial position. That position is estimated in the *starting location* field based on input from the *current location gate* field that transmits a position sensor signal only when the vehicle is at rest.

When the updated peak in the *starting location* field matches the intended target location in the *drive* field, the *drive CoS* field builds a peak that activates the *drive CoS* node, which inhibits the *drive intention* node and thus terminates the *drive* intention-in-action.

Information about the target location comes from the *recall* intention-in-action. Until *recall* has delivered such information through the *recalled position gate*, *drive* is prevented from activating by an inhibitory precondition node. *Recall*, illustrated on the left of Fig. 9, is not a motor behavior, but is a world-to-mind intention-in-action in that it brings about an intended state within the neural dynamic architecture in which the recalled target location is neurally represented. The *recall* intention node is activated by the *locate* intention, which also provides a feature cue to the *recall cue* fields that represent the contents of the *recall* intention. Peaks in those fields project onto the *space/color* and *space/height Recall* fields where their overlap

deactivates the *recall* intention node and releases the exploration intention from inhibition, so that the vehicle begins searching the environment.

3.2.2. *Prior intention*

Prior intentions are parts of mental plans that do not yet directly affect down-stream world-to-mind intentional processes. A prior intention may, therefore, become an intention-in-action when it begins to have a direct effect on sensory-motor processes or on other intentions-in-action. Prior intentions sequentially organize the activation and deactivation of intentional states and of their gating to downstream processes. In the present neural architecture, this issue is addressed only in somewhat rudimentary form.

Prior intentions consist of two components: The encapsulated intention-in-action and a *precondition* that specifies the circumstances under which that action will be initiated. Preconditions may be any perceivable world state, including internal states of the agent. Once a perceptual state has been formed that matches the precondition, the encapsulated intention-in-action is activated.

Preconditions are neurally implemented by dynamic nodes that inhibit the intention node of the encapsulated intention-in-action (Richter, Sandamirskaya, & Schöner, 2012). Precondition nodes receive inhibitory input from match detection fields, which perceptually ground the precondition and define its contents. So, when the precondition is “met,” the node is deactivated. Sequences of prior intentions are organized by using the CoS of preceding actions as input to the match detection field. These conditions-of-satisfaction come from the encapsulated intentions-in-action, but may also depend on the state of the precondition network. Composite prior intentions may comprise multiple lower-level intentions-in-action, leading to a hierarchy of actions, in which the lowest level directly acts on the motor surface, while higher levels stand for sequences of such actions. The mechanism of precondition can be used at all levels of the hierarchy to sequentially organize intentional states.

3.2.2.1. Example: Prior intentions collect and apply: Fig. 10 illustrates the neural dynamic network that sequentially organizes the agent’s prior intentions *collect* and *apply* as part of the *paint* prior intention.

3.2.3. *Desire*

Desire is the top-level psychological mode of world-to-mind intentions. It disposes the agent to act toward achieving goals, that is, particular states of affairs of the world. Desires are the causes of prior intentions.

While desires predispose an agent to act toward particular goals, they do not necessarily induce the agent to initiate these actions, a logical condition described by Searle as the *volitional gap* (Searle, 2001, 50). Such initiation depends on an ensemble of intentional states in all other modes of intention. Conversely, desires may affect intentional states in those other modes. For instance, desires may be critical to activating relevant beliefs, guide perceptual attention, and drive reward-based learning upon achieving the desire (Tekülve & Schöner, 2020).

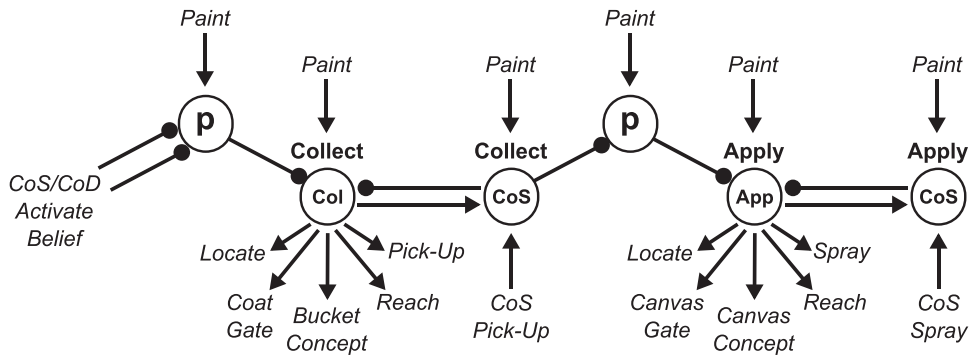


Fig. 10. The network responsible for sequentially organizing the *collect* and *apply* prior intentions that implement the *paint* behavior. Precondition nodes are labeled “P.” Arrows indicate excitatory projection. Filled circles indicate inhibitory projection. Both *collect* and *apply* consist of multiple lower-level intentions-in-action whose sequential organization is implemented analogously (not shown). The CoS of each prior intention comes from the final intention-in-action in each subsequence.

The neural representation of desires is grounded through the projections onto the rest of the architecture. The condition of satisfaction of a desire reflects that connectivity by conversely collecting input that tests the predicted outcomes. Desires do not specify individual concrete actions, allowing for multiple paths to their satisfaction. Which actions are ultimately taken may depend on the agent’s mind-to-world states, the currently perceived environment, memory, and beliefs.

3.2.3.1. *Example: Desire to paint:* The current model only has a trivial neural dynamics in the psychological mode of desire, a neural representation of a desired hue value that defines the color in which the agent aims to paint an object. This projects onto the belief system, where matching canvas and coat colors are sought, and onto the collect and apply prior intentions (Fig. 7).

4. Simulations

To demonstrate the capacity of the neural architecture to generate intentional states and to drive the behavior of an embodied agent toward fulfilling desires by using beliefs, we describe a few exemplary simulations. The neural processes that instantiate perception, memory, intentions-in-action, and prior intentions are made visible through the time courses of activation that emerge from the neural dynamics in particular situations. Special attention is paid to the dynamic instabilities that generate sequences of intentional states, organized by intentional states and their CoS. We will highlight differences in the time structure and chain of causation between mind-to-world and world-to-mind forms of intentionality.

The simulations are carried out by implementing the architecture within Cedar (Lomp, Richter, Zibner, & Schöner, 2016), a software framework that enables the interactive,

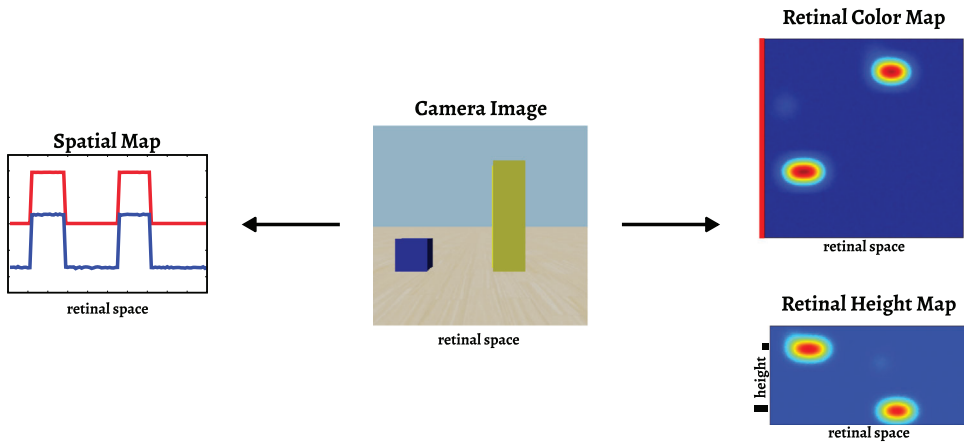


Fig. 11. The camera image (center) induces activation patterns in two space/feature maps defined over the horizontal dimension of retinal space (right). Activation is visualized with a color code: blue marks activation levels below threshold, green marks activation levels near threshold, yellow and red marks activation levels above threshold. The camera image also provides input to the spatial map (left). Here, activation (vertical axis) is plotted against the horizontal dimension of retinal space (horizontal axis). Activation is shown in blue, the output of the sigmoid threshold function is shown in red.

graphical specification of DFT architectures. Cedar numerically solves the neural dynamic equations, visualizes activation time courses, and allows online model parameter adjustment. The agent and its environment are simulated in Webots (Michel, 2004) which we interfaced with Cedar. The neural architecture generates time courses of robot commands acted out by the simulated robot in Webots, which sends back sensory information that provides input to the neural architecture.²

4.1. Visual perception and memory

Visual exploration generates perceptual states in which one object at a time is visually attended and entered into scene memory. While perception and memory are mind-to-world intentional states, visual exploration itself is a world-to-mind form of intention as it brings about intended states of the agent's neural architecture. Visual exploration is, in fact, the default intention-in-action of the architecture.

Below, we refer to Fig. 8 for the neural dynamic network that supports visual exploration. The sensor surface shown at the bottom of that figure is illustrated in Fig. 11 with its activation state in response to visual input. Color and height features are extracted from the camera image, and input into two retinal space/feature maps. Saliency is extracted and input into a spatial map. At the visual perception layer of the architecture (second region from bottom in Fig. 8), the spatial selection field selects a single location from the spatial map, generating a peak that represents selective spatial attention. That peak provides ridge input into two space/feature fields at the visual perception level which also receive input from the analogous fields at the sensor level. Where the ridge overlaps with that input peaks build in each

space/feature field forming the neural representation of the perceptual intentional state that is about the selected object. While these perceptual states are in retinal (or camera) coordinates, memory is built in an allocentric frame. The memory state is built by coordinate transforming the object's location and binding that transformed location to the object's feature information that is extracted from the retinal representation.

Fig. 12 illustrates time courses of activation that bring about the attentional selection of an object and its commitment to memory. At time, t_0 , visual exploration has not yet started and activation of all neural nodes is below threshold. The two objects in the visual array induce two subthreshold activation bumps in the retinal perceptual layer which are not transferred to the allocentric memory layer as the coordinate transformation only transmits above threshold activation. The *space/color memory trace* contains two bumps from previously observed objects that are now out of view to the left of the agent, one blue and one purple. These induce two subthreshold bumps into the *memory selection* field shown at the bottom.

At time, t_1 , the intention node for visual exploration (violet) has passed threshold, which raises the activation level of the associated CoS-node (green) and boosts the *spatial selection* field (second panel). That boost causes the formation of a single peak selected over the left-most location by chance, which provides ridge-input into the *space/color retinal selection* field (third panel). Where that ridge overlaps with the subthreshold bump from sensory input, a peak forms, part of the perceptual intentional state about the blue cube in the scene. The peak in the spatial selection field is now transmitted through the coordinate transform to the *allocentric selection* field (fourth panel), inducing a peak at the corresponding location in space. This peak projects ridge input into the *space/color memory gate* field (fifth panel). That ridge intersects with an orthogonal ridge from the *attended color* field (bottom left in 8), but remains below threshold, so that no memory trace is being formed yet (second panel from bottom). The spatial ridge input into the *memory selection* field (bottom panel) does not overlap with the pre-existing subthreshold bumps from the memory trace, so no peak forms at this time.

At time, t_2 , activation at the intersection of the two ridges in the *memory gate* field has pierced threshold. The resulting peak drives the build-up of a new bump in the *memory trace* which induces a peak in the *memory selection* field at the bottom. That peak has fully formed by time, t_3 . Inhibitory projection to the *memory gate* field destabilizes the peak there, ending the build-up of memory trace at that location. Feature values extracted from the *memory selection* field have been forwarded to the *expected color/height* fields in the *object match* subnetwork (top left in Fig. 8), activating the match nodes and ultimately the CoS node of visual exploration (top panel). The CoS node has inhibited the *visual explore intention* node, whose boost input to the *spatial selection* field has thus fallen away, driving that field through a reverse detection instability. The peak in that field is almost gone at this time. Its decay is complete by time, t_4 . By removing excitatory input, decay of activation has propagated to the space feature fields, also causing deactivation of the match nodes and, a little later, of the CoS-node (top panel), so that all fields fall back to their initial subthreshold state. The memory trace retains the new bump, providing support for future reactivation of the memory intentional state. The temporal order from intentional state to CoS to inhibition of the intentional state characterizes visual exploration as a world-to-mind intention.

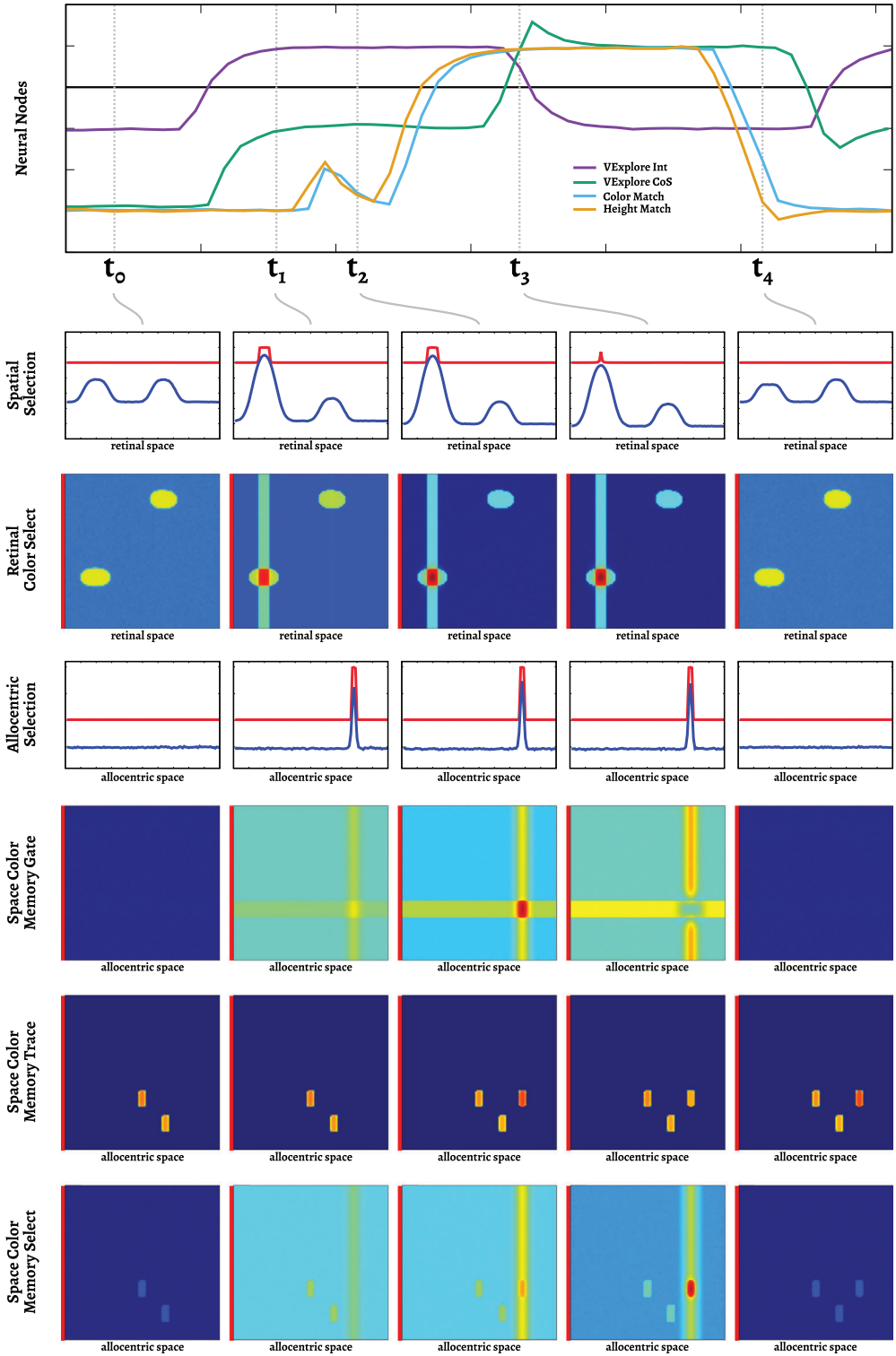


Fig. 12. Time courses of activation during visual exploration of one object are illustrated through snapshots at discrete moments in time. These are marked by gray vertical lines. The top row shows the activation time courses of the intention- and CoS-nodes of the *visual explore* behavior, and of the *color-* and *height-match* nodes. All other rows use the two plotting conventions of Fig. 11. The next two rows show the selective spatial and color/space fields from the perceptual layer in retinal coordinates, x . The remaining four rows show fields of the memory layer in allocentric coordinates, x_w . While only the color feature channel is illustrated, analogous patterns emerge in the height feature channel.

After time, t_4 , the decay of the CoS node enables the reactivation of the *visual exploration intention* as the default intention, which leads to the attentional selection of the next object in the visual scene and ultimately its commitment to scene memory.

4.2. Intention-in-action and prior intention

The two intentions-in-action *recall* and *drive* are part of the composite prior intention *locate*. *Locate* provides the color and height cues for the recall. These cues originate from the agent's belief system. In the simulated example, the agent is looking for a yellow *bucket* to pick up yellow paint that, according to a learned belief, will help paint a particular *canvas* to obtain a desired color. Fig. 13 only illustrates the portion of this sequence in which the agent recalls the location of the yellow *bucket* and drives to that location.

At time, t_0 , the *locate intention* node (top panel) is just passing threshold, while the *recall intention* node is not yet activated. All fields are subthreshold. The *spacefeature recall* fields for *color* and *height* are preactivated by the memory traces left during earlier visual exploration of four objects (from left to right): a purple *bucket*, a purple *canvas*, a yellow *canvas*, and a yellow *bucket*. These are the four objects in the right half of the scene. The agent faces the purple *bucket* close to the center of the allocentric coordinate frame. The simulated localization sensor provides input at that ego-position to the *current location gate* field (bottom panel), inducing a peak there, which in turn induces a peak in the *start location* field (second panel from bottom) at the same location. That peak provides input to the *drive CoS* field (third panel from bottom), where it causes a subthreshold activation bump.

At time, t_1 , the *locate intention* node is fully activated and has induced activation of the *recall intention* node (top panel). The *locate* intention also provides excitatory input to the *drive* intention and to the precondition node that links *recall* to *drive*. Because the *recall intention* node is active, the precondition node prevents the *drive intention* node from activating. The activated *recall intention* node gates the feature cues for the yellow *bucket* into the respective cue fields (third and fourth rows from top), where peaks have formed at “yellow” and “tall.” These peaks provide ridge input into the *spacefeature recall* fields (fifth and sixth rows). In each field, two entries of the *memory trace* overlap with the ridges, generating activation close enough to threshold to enable projection to the *recalled position* field (seventh row). This field's resting level has been boosted by the *recall intention* node and the peak detectors of each *recall cue* field. Only the one location (to the right) at which inputs from the two *spacefeature recall* fields converge is sufficiently activated to create a peak. This peak represents the location of the object that matches both search cues, “yellow” and “tall.”

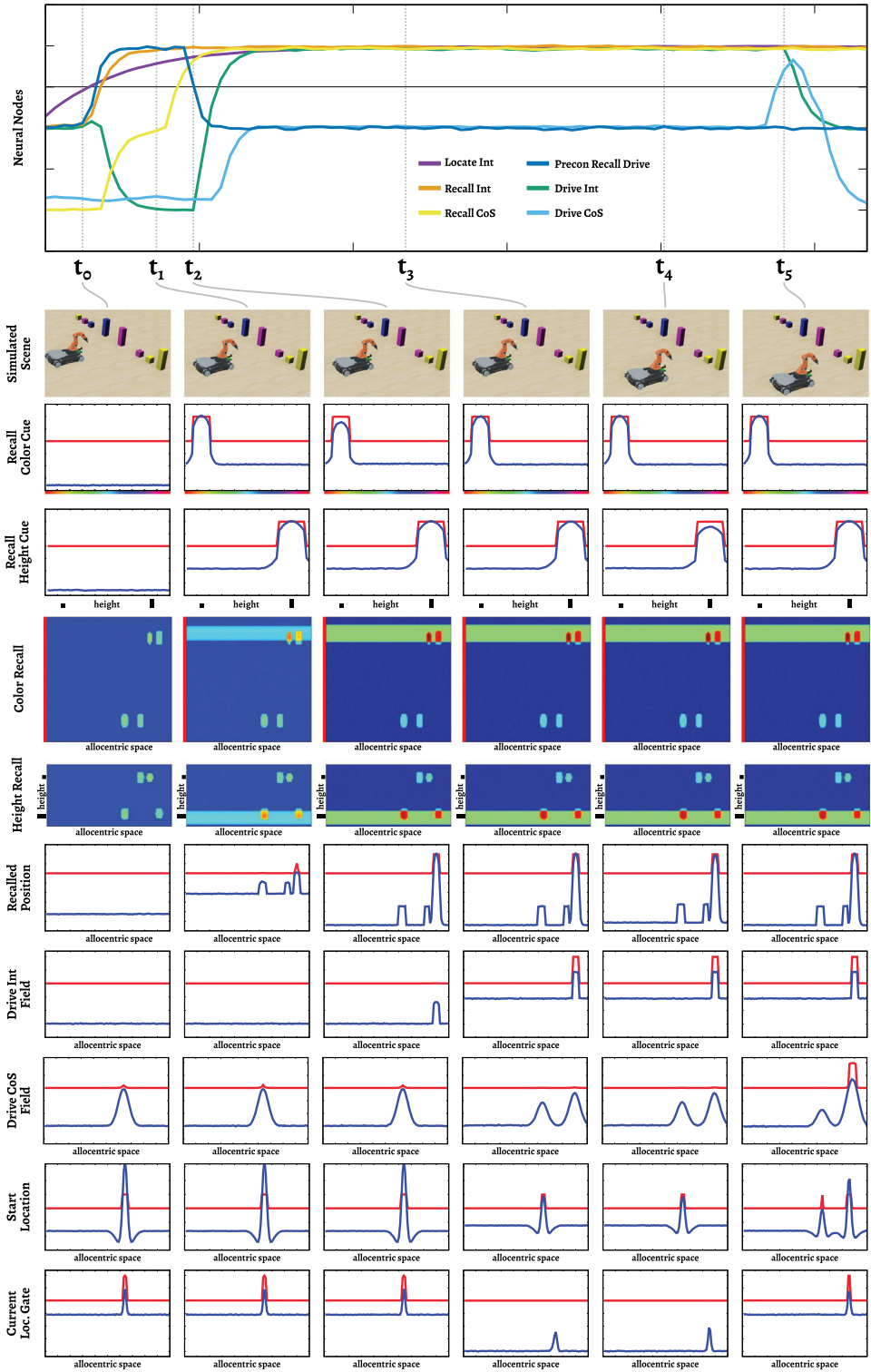


Fig. 13. Temporal evolution of activation during recall and drive. The top panel shows time courses of relevant intention-, CoS-, and precondition nodes. At five moments in time, marked by vertical bars, the agent (second row), and the activation state of a set of fields are shown, five from the *recall* subnetwork, and four from the *drive* subnetwork. The same conventions are used as in Fig. 12.

At time, t_2 , the peak in the *recalled position* field has fully stabilized causing the *recall CoS* node to activate. This inhibits the precondition node which releases the *drive intention* node from inhibition, leading to its activation. A subthreshold bump at the recalled location is building in the *drive intention* field but has not yet reached threshold.

At time, t_3 , the *drive intention* node has been fully activated and a peak has formed in the *drive intention* field that represents the target position close to the right border of the allocentric frame. This causes the agent to initiate a right-ward movement toward that target position. This also causes a second subthreshold bump to arise in the *drive CoS* field (second row from bottom) at the predicted end-point of that movement. This bump does not overlap with a second bump induced by the peak in the *start location* field (bottom row), which is sustained, while input from the *current location* field falls away when that field is inhibited during movement. The agent has almost finished its movement at time, t_4 , its current position being reflected by the subthreshold bump in the *current location* field (bottom row).

Finally, at time, t_5 , the movement of the agent has terminated, releasing the *current location* field from inhibition. The reactivated peak in that field leads the *start location* field to switch to a new peak at the current location, that will be the starting location for the next movement. This switch causes a peak to form in the *drive CoS* field and the *drive CoS* node to become activated, which ultimately inhibits the *drive intention* node (top row), which is followed at the end of the time series by a decay of the *drive CoS* node. The recall and drive sequence has terminated and the system is ready to perform the next action.

Drawing back the view point to look at the larger scale, we illustrate in Fig. 14 how a sequence of world-to-mind intentions unfolds autonomously in time as a global goal, the desire to paint an object in a particular color, is pursued. The bar at the top represents that *paint* intention, which is terminated when its CoS (shown in the second line) is activated. All the other world-to-mind intentions are engaged to realize this *paint* intention. The mind-to-world intentions that interact with these world-to-mind intentions are not shown.

First, the *activate belief* intention becomes active (line 3, far left). While beliefs linking coat, canvas, and outcome color are mind-to-world intentions, the process of activating a belief is a world-to-mind intention because it is aimed at bringing about a particular state of the neural architectures (a part of the world), here the activation of a matching belief node. Once the CoS of this intention is activated (line 4, far left), the associated intention (line 3) is deactivated. Note that the belief itself will remain active through the entire sequence (not shown).

The sequence of events that follows is organized by the network illustrated in Fig. 10 which is primed by the *paint intention* node. The CoS of *activate belief* inhibits the precondition node of *collect*, releasing from inhibition and thus activating the *collect* intention (blue bar on the left of line 3). As a prior intention, *collect* generates a sequence of intentions-

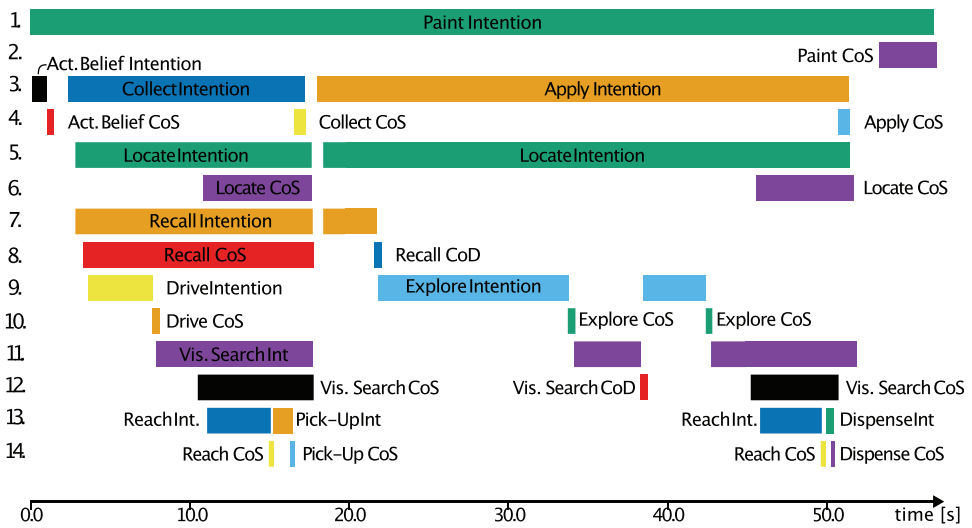


Fig. 14. A survey over the temporal structure of world-to-mind intentions that are generated toward the goal of painting a canvas to achieve a desired color. The bars mark the time intervals during which the labeled intentions and their conditions of satisfaction are activated above threshold. This time structure was recorded from a simulation of the complete architecture while the intention to paint (top row) was active.

in-action (column of bars below that blue bar). These include the *locate* intention, and the *recall* and *drive* intentions we analyzed in Fig. 13, as well as *visual search*, *reaching*, and *pick-up paint* intentions.

The *collect* intention terminates once its CoS has been detected. This inhibits the precondition node of the *apply* intention (Fig. 10), which is the next prior intention to activated (orange bar in line 3). The *apply* intention entails a similar set of lower-level world-to-mind intentions. In this simulation, the attempt to recall a canvas object of the requested color from memory fails as no such object has been previously committed to memory. So, the CoD of the *recall* behavior becomes active (line 8 around time 22 s). The system then *explores* the environment for an object of the required color. When a first object of that required color is found, visual search brings it into the attentional foreground, which reveals that it does not match the required height feature value (it is a bucket rather than a canvas), so another CoD is activated, and *explore* is reactivated. Finally, a suitable object is found, attentionally selected, reached for with the robot tool, and covered with a coat of paint.

Note how all through this sequence the world-to-mind intentional states are activated for variable durations. Their CoS are briefly activated when the predicted outcome is detected. This is the process signature of world-to-mind forms of intentionality.

4.3. Achieving a desire

Beliefs and desires are the psychological modes most removed from the sensory-motor level. The activation of beliefs and their autonomous learning follows methods described in Tekülve and Schöner (2020). Here, we focus on illustrating how the agent satisfies a desire.

Fig. 15 shows the time courses of neural nodes and fields at the very last moment of a sequence of actions, when a desire to paint a canvas in a new color is finally achieved.

When the sequence begins at time, t_0 , the agent is already positioned near the canvas object, a blue cube, which it has attentionally selected and to which it has reached. The *spray* intention, which makes the robot dispense paint on the object, is not yet active (see nodes in top panel). In the *retinal space color map* (third row), all three visible objects are represented as peaks. These provide input to both the excitatory and inhibitory layers of the *transient detector* (rows 4 and 5). The suprathreshold peaks in the inhibitory layer project inhibitorily onto the excitatory layer, so that activation there remains below threshold. The color of the canvas object and the desired outcome color are each represented by peaks in single feature fields (bottom two rows).

Between times, t_0 and t_1 , the *spray* intention activates as the precondition node from *reach* deactivates. This has induced the *dispense* action, and the blue canvas object has changed physically from blue to purple, the desired outcome. The peak in the *retinal space color map* at the location of the canvas shifts along the color axis. This provides fresh excitatory input into the excitatory layer of the *transient detector*, not yet matched by inhibition from the inhibitory layer. A transient peak arises in the excitatory layer, at maximum at time t_2 , which drives a peak detector and then the *change detection* node above threshold (top). The inhibitory layer of the *transient detector* catches up beyond time, t_2 , leading to the decay of activation in the excitatory layer and subsequently to decay of the activation of the *change detection* node. All other intention and CoS nodes decay too. During the *dispense* intention, the currently attended color is gated into the *canvas color* field, which boosts the already present peak of sustained activation. Through a small set of additional fields and nodes (not shown), the CoS of the current intentions are activated, and so is ultimately the CoS of the *paint* intention itself. This deactivates the *paint intention* node and terminates the whole painting episode. The desired outcome has been achieved.

The *transient detector* signals a change of canvas color that is temporally aligned with the *spray* intention, a literal condition of satisfaction for a world-to-mind intention. The system thus autonomously detects intentional acts that cause events. This also provides the basis for autonomously learning the contingencies of its actions by modulate the Hebbian learning dynamics (Tekülve & Schöner, 2020).

5. Discussion

Acting intelligently in the world minimally entails (1) initiating, terminating, and sequentially organizing individual actions, (2) basing behaviors on perceptual information that is not immediately available on the sensory surfaces, and (3) using knowledge about the world (4) to achieve goals. The neural processes that deliver these competences must be endowed with stability, the capacity to resist change. At the sensory-motor level, this enables behavior to persist in the face of disturbances from the environment and from the agent's own actuators and sensors. At higher levels, the stability problem is one of process organization, keeping processes other than the current one from interfering, while remaining open to online updating and and coupling among processes.

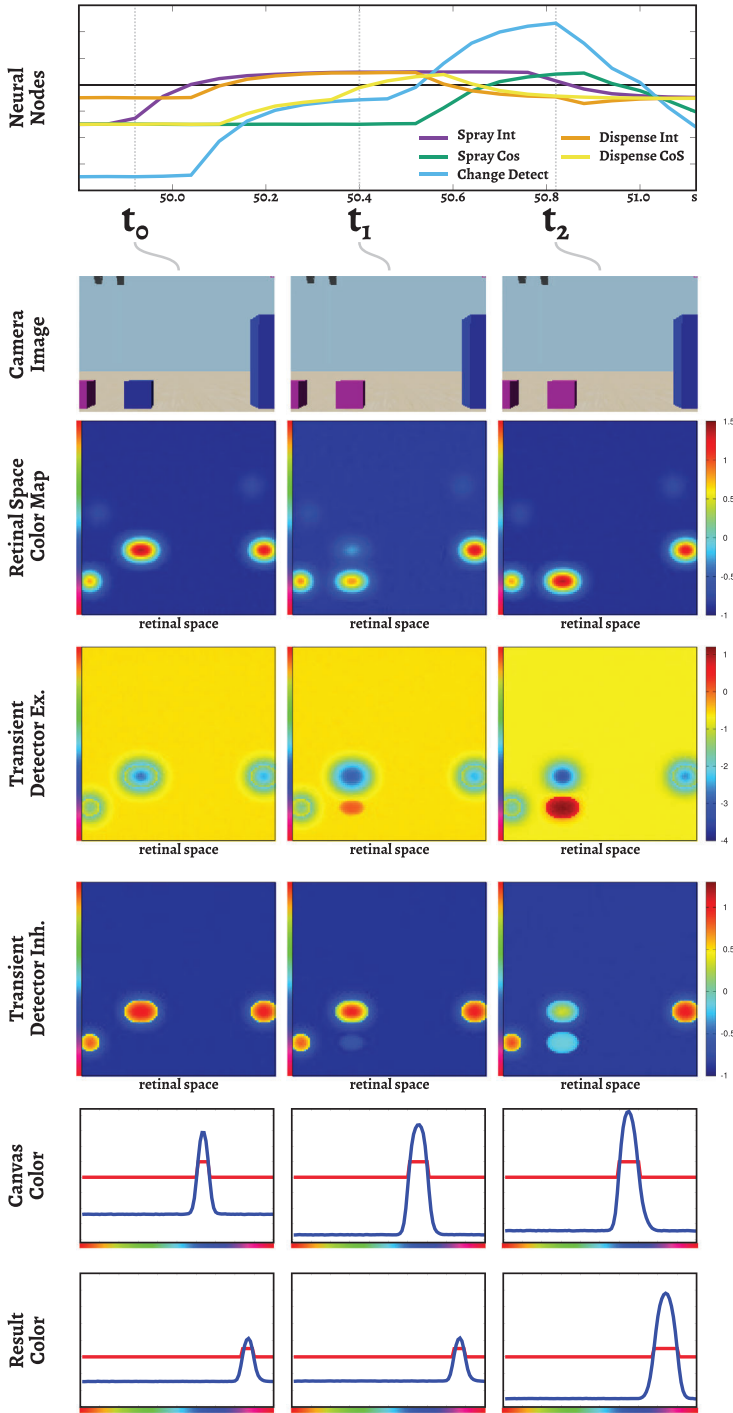


Fig. 15. Temporal evolution of activation during the spray intention. Same conventions as in Fig. 12. See text for detailed description.

Our analysis of these core problems was facilitated by borrowing terms from the philosophy of mind. We hypothesized that intelligent thinking and acting are based on *intentional states* that are about the world in *two directions of fit*, *mind-to-world* and *world-to-mind* (Searle, 1983). Intentional states were given stability properties by instantiating them as attractors of a neural dynamics. At any moment in time, only a tiny fraction of all potential intentional states will be realized. Their instantiation and termination entails bifurcations of the neural dynamics. We borrowed the concept of the *condition of satisfaction*, originally a logical characterization of intentional states (Searle, 1983) to analyze the temporal organization of these bifurcations in the two directions of fit. These ideas were made concrete by building an exemplary neural dynamic architecture of an intentional agent. We used Searle's notion of *psychological modes* (Fig. 1) to organize that architecture and to illustrate that the neural dynamic principles reach the competences of intelligent behavior. To articulate the chief insights obtained from this exemplary study, we review the mapping of neural dynamic mechanisms onto these borrowed concepts.

5.1. Mind to world intentions

Mind-to-world intentional states are neurally instantiated as stable supra-threshold activation patterns in perceptual, memory, or belief representations. *Perceptual states* are conventionally thought of as input-driven in neural networks with predominant feed-forward connectivity. In our neural dynamic analysis, perceptual states are stable activation peaks localized within low-dimensional feature spaces including visual space. Their stability comes from recurrent connectivity (called neural interaction in DFT), that is dominant in the sense that it may overrule input. Perceptual states are thus not necessarily or uniquely determined by input but may be influenced by prior perceptual decisions (Hock & Schöner, 2010) and top-down influences. Interaction creates a bistable regime in which neural attractors representing perceptual states are separated from input-driven subthreshold activation patterns. Along increasing input, this bistable regime is delimited by the detection instability at the upper end and the reverse detection instability at the lower end. Thus, even seemingly very low-level perceptual states such as the detection of an element of localized contrast entail decisions. Empirical evidence comes in the form of perceptual hysteresis and perceptual switching (Hock & Schöner, 2023). Perceptual states typically also entail selection decisions, most commonly in the form of selective attention controlled by top-down influences (Grieben et al., 2020), or in the form of perceptual multistability (Hock, Schöner, & Hochstein, 1996).

The locations of activation peaks in feature space reflect the *content* of a perceptual state. Its alignment with the world results from how peaks arise in detection and selection decisions when inputs render subthreshold activation patterns unstable. Locations at which the sum of all inputs, localized and homogeneous boosts, first exceeds threshold are likely to win the competition with other locations and form a stable activation peak. Such a perceptual intentional state may be "in error" when it remains stable, while its content is no longer aligned with the state of affairs in the world. Perceptual intentional states are dissolved when such misalignment ultimately induces the reverse detection instability of the peak.

Memory intentional states are likewise instantiated as stable activation peaks in neural fields or as stable, suprathreshold activation states of neural nodes. Peaks in neural fields over feature dimensions are perceptually grounded forms of memory intentional states, while neural nodes provide the neural substrate for the simplest form of conceptual representations (Sabinasz, Richter, & Schöner, 2024). Memory intentional states are instantiated when subthreshold activation states become unstable in detection instabilities induced by boost input, either global in free recall or partial in cued recall. The content of memory states, represented by the locations of the peaks or the identity of the neural nodes, is determined by memory-traces, subthreshold activation patterns that reflect earlier suprathreshold activation states. The dynamics of memory traces (Sandamirskaya, 2014) is an approximate description of the much more complex neural mechanisms of memory formation, a long-standing and ongoing research topic in computational neuroscience (Zeng, Diekmann, Wiskott, & Cheng, 2023). Memory intentional states match the world by virtue of being aligned with the memory traces. The memory traces themselves may be best thought of as “background” in the parlance of intentionality (Chapter 5 of Searle, 1983), rather than as intentional states by themselves.

Belief intentional states are neurally instantiated as stable suprathreshold levels of activation in dedicated neural nodes which represent contingencies by virtue of the pattern of connectivity to concept nodes acquired during learning. Beliefs in this sense are propositional, consisting of concepts bound together by this learned connectivity (Sabinasz et al., 2024). Belief states are activated in detection instabilities, when inputs make the subthreshold states unstable. Which belief is selected depends on the prior activation of subsets of the concept nodes. Global competition among belief nodes makes that only one belief can be stable at any given time. Belief states match the world by virtue of how the learned connectivity that binds belief nodes to concept nodes reflects patterns of coactivation of the concept nodes obtained during exploration. Activated beliefs may fail to match experienced events, which leads to their deactivation and the learning of a new belief.

In all forms of mind-to-world intentionality, the intentional states match the word by virtue of how they came about. There is no need for a neural representation of their condition of satisfaction separate from the neural representation of the intentional state itself. In fact, these states can be in error, and when that happens additional processes such as change detection must intervene to uncover that (Johnson, Spencer, Luck, and Schöner, 2009).

5.2. *World-to-mind intentions*

World-to-mind intentional states are neurally instantiated by suprathreshold activation patterns that drive motor behavior (*intentions-in-action*), generate sequences of actions (*prior-intentions*), or provide goals for sequences of actions (*desires*). *Intentions-in-action* are represented by neural activation peaks in fields of movement parameters that directly drive motor systems and bring about physical actions of the body. These peaks are thus created before the action is initiated and must remain stable until the action has been successfully completed. The successful termination of an intended action, its CoS, must, therefore, be represented in a separate neural field or node. These receive subthreshold input from the neural representation

of the intention that predicts the outcome of the action. When perceptual information matches the prediction, perceptual input overlaps with the input from the intention field, inducing a detection instability. The suprathreshold CoS activation inhibits the neural representation of the intention-in-action, terminating that state in a reverse detection instability. As a result, input to the CoS from the intentional state falls away, inducing a reverse detection instability in the CoS neural representation. This leaves all neural representations pertaining to the intention-in-action in a deactivated state.

Prior intentions are neurally represented in the same way but project to other prior intentions or intentions-in-action. Through connectivity involving precondition nodes and competitive inhibitory coupling, prior intentions generate sequences of actions. Projection onto the neural representation of the CoS of prior intentions reflects predicted outcomes. When inputs from other prior intentions match these predictions, the same cascade of instabilities as for intentions-in-action occur, leading to the deactivation of the prior intention.

Desire intentional states are represented by suprathreshold activation patterns that project onto those prior intentions and intentions-in-action that help to realize the desire. Their content specifies the goal of the ensuing world-to-mind intentions. The outcome predicted by a desire preactivates the neural representation of its CoS, whose activation ends the desire intentional state.

For all world-to-mind intentional states, the world matches the mental state by virtue of how the world's state is changed by an intentional action. This match is the condition of satisfaction, and it must be neurally represented separately from the intentional state because it is activated at different moments in time. In fact, all world-to-mind intentions are characterized by the ordinal structure of their activation in which the intentional state is activated first and remains activated for variable amounts of time until its CoS is activated, which leads to both states becoming deactivated.

This signature time structure of world-to-mind intentional states helped us clarify that cognitive processes may have the world-to-mind direction of fit even though they are not directly motoric in nature. Visual search, cued recall, and activating a belief are examples. These processes must bring about specific activation states in specific parts the neural dynamic architecture. As the neural architecture is itself part of the world, these are world-to-mind intentions. Note, however, that the activation states such intentions are aimed to bring about are not the activation state of the intention itself, of course. So, there is no contradiction or circularity here. This insight is critical to building autonomous cognitive architectures in neural processing terms.

The neural dynamic architecture demonstrated in this paper was aimed to explore the reach of these concepts. While set in a small toy world, the autonomous agent was capable of generating sequences of intentional states in both directions of fit at increasing distance from the sensory surfaces (Fig. 1). At the lowest level, perception and motor behavior come natural to neural dynamics. At the middle level, we showed how memory traces are acquired from which memory states can be activated and demonstrated the capacity to generate sequences of actions. Failures can be addressed by switching to alternate sequences. At the highest level, we showed how knowledge can be activated to achieve the desired outcome. Knowledge took the form of conceptual representations, which binds concepts in different roles together, albeit

in very elementary form (see Sabinasz et al., 2024 for expansions). The autonomous acquisition of such knowledge from a single experience has been documented earlier (Tekülve & Schöner, 2020). Only the goal states of the architecture remained entirely trivial and static.

Within the DFT framework, the chief innovations of this paper are the use of knowledge to achieve desired outcomes and uncovering the temporal structure of the autonomous neural processes that enable intelligent behavior. Each mental or behavioral decision emerges as an event at a discrete time from the underlying time- and state-continuous neural dynamic processes. This form of neural process autonomy sets the neural account presented here apart from related efforts which we briefly review next.

5.3. *Other cognitive architectures*

Cognitive architectures such as ACT-R (Anderson et al., 2004), SOAR (Laird, 2012), and others pursue similar goals by combining symbolic with subsymbolic representations, with extension linking to sensory-motor processes as well (Trafton et al., 2013). While mapping ACT-R modules roughly onto the brain, not all operations within this framework are consistent with the principles of neural function as postulated in connectionism and in DFT (Taatgen & Anderson, 2008). In particular, ACT-R production rules are “fired” by calling functions to which information is passed, a core challenge in neural networks (see Richter, Lins, & Schöner, 2021 for discussion).

The LIDA model (Franklin, Madl, D’mello, & Snider, 2013) is aimed to overcome some limitations of earlier cognitive architectures in symbol grounding (Harnad, 1990). The “attention phase” of processing in LIDA based on Global Workspace Theory (Baars, 1993) guides all action and learning of the model. This could be viewed as functionally analogous to the neural dynamic mechanism of selective attention. In fact, parts of LIDA’s cognitive cycle have been implemented in DFT (Franklin et al., 2013) to prove neural plausibility, while other components are still implemented within the computer metaphor using object-oriented programming in Java. LIDA’s overall structure and its elaborate memory systems may provide a road map for future DFT architectures. The key challenge lies in finding neurally consistent solutions for LIDA’s numerous codelets.

DAC theory (Verschure, 2012) pursues similar goals as we do. DAC provides an overall processing architecture that is neurally inspired and meant to organize cognition. Similar to our psychological modes, DAC has layers that increasingly abstract from the sensory-motor surface. Its three columns, world, self, and action, could be mapped onto different directions of fit. Some versions of DAC have emphasized strict adherence to neural principles (Duff & Verschure, 2010), but its more advanced variants continue to take algorithmic shortcuts (Moulin-Frier et al., 2017).

LEABRA (O’Reilly, Hazy, & Herd, 2016) provides guidelines and principles, broadly consistent with DFT, that constrain neural networks approaches to cognition. For instance, LEABRA invokes inhibitory neurons to account for competitive selection, sustained activation to account for working memory, and weight-change to model long-term memory. As a synthesis of ACT-R and LEABRA, SAL is aimed to overcome the limitations of LEABRA’s

cognitive competences (Lebiere, O'Reilly, Jilk, Taatgen, & Anderson, 2008). Even the most recent variants (Szabados, Herd, Vinokurov, Lebiere, & O'Reilly, 2020) retain, however, central algorithmic processing components that are not consistent with neural principles. As a result, none of these approaches uncover the principles and time structure of neural processing implied by the two directions of fit.

VSA (Levy & Gayler, 2008; Smolensky, 1990; Schlegel, Neubert, & Protzel, 2022) provides a very different perspective on neurally based cognition. It operates with symbols, represented by high-dimensional neural activation vectors. An implementation in spiking neural networks (Eliasmith et al., 2012) has demonstrated links to sensors and motor systems. Because the underlying principles are quite different, a detailed comparison is not easy and we would like to postpone this for a future possible debate (see also discussion in Sabinasz et al., 2024).


5.4. *Scaling*

How would a neural dynamic architecture of intentional states scale to reach more realistic and naturalist embodied cognition? This question raises a fundamental issue. We emphasize that the minds of humans and other animals are structured in a specific way that reflects their evolutionary origins. That structure is also visible in the functional organization of the brain. This is a position widely held within the research community of human cognition that could be contrasted to more abstract characterizations of intelligence based on computational principles such as productivity, systematicity, and compositionality (Spivey, 2023). The toy example we used here was meant to provide a minimal version of the specific architecture of the mind. Components of this architecture are, in fact, simplified variants of more extensive architectures that, in some cases, have been evaluated against psychophysical and neural evidence. For instance, the visual front end is a simplified version of a model of visual search and scene memory (Buss et al., 2021; Grieben et al., 2020), the motor system a simplified version of a general neural architecture of movement generation (Knips, Zibner, Reimann, & Schöner, 2017; Schöner, Bildheim, & Zhang, 2024; Tekülve, Fois, Sandamirskaya, & Schöner, 2019), and the belief system makes use of a simplified variant of a neural dynamic model of conceptual structure (Sabinasz et al., 2024). In this perspective, scaling the architecture amounts to providing further neural dynamic components for competences of the same general nature that find their natural place in the overall architecture. Least developed is the neural process account for goals (desires). The time structure of the processes instantiating world-to-mind intentions that we described here points to the orientation of goals toward future outcomes as a key characteristic that may provide an entry point for a neural account, inspired, for instance, by ideomotor theory (Herbort & Butz, 2012; Shin, Proctor, & Capaldi, 2010; Vogel-Blaschka, Kunde, Herbort, & Scherbaum, 2024).

Acknowledgments

Open access funding enabled and organized by Projekt DEAL.

Open Research Badges

 This article has earned Open Materials badge. Materials are available at https://osf.io/m38vz/?view_only=1125aae86eeb437e9985665b693e0ae4

Notes

- 1 It might have been more intuitive if the small cubes would have been paint buckets and the tall cuboids would have been canvases. Somehow, we made the reverse choice early in the design of the Webots simulation and then got stuck with it.
- 2 Source code and all parameter values are available online at https://osf.io/m38vz/?view_only=1125aae86eeb437e9985665b693e0ae4.

References

- Amari, S.-i. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2), 77–87.
- Anderson, J. R. (2007). *How can the human mind occur in the physical universe*. New York: Oxford University Press.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 1036–1060.
- Baars, B. J. (1993). *A cognitive theory of consciousness*. Cambridge University Press.
- Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7(1), 66–80.
- Bowers, J. S. (2017). Grandmother cells and localist representations: A review of current thinking. *Language, Cognition and Neuroscience*, 32(3), 257–273.
- Buss, A. T., Magnotta, V. A., Penny, W., Schöner, G., Huppert, T. J., & Spencer, J. P. (2021). How do neural processes give rise to cognition? Simultaneously predicting brain and behavior with a dynamic model of visual working memory. *Psychological Review*, 128(2), 362–395.
- Crane, T. (2013). *The objects of thought*. Oxford University Press.
- Duff, A., & Verschure, P. F. (2010). Unifying perceptual and behavioral learning with a correlative subspace learning rule. *Neurocomputing*, 73(10–12), 1818–1830.
- Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, 338(6111), 1202–1205.
- Fodor, J. (1975). *The language of thought*. New York: Crowell.
- Franklin, S., Madl, T., D’mello, S., & Snider, J. (2013). LIDA: A systems-level architecture for cognition, emotion, and learning. *IEEE Transactions on Autonomous Mental Development*, 6(1), 19–41.
- Gerstner, W., Kistler, W. M., Naud, R., & Paninski, L. (2014). *Neuronal Dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.
- Goodale, M. A., Pélisson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, 320, 748–750.
- Grieben, R., Tekülve, J., Zibner, S. K. U., Lins, J., Schneegans, S., & Schöner, G. (2020). Scene memory and spatial inhibition in visual search: A neural dynamic process model and new experimental evidence. *Attention, Perception, & Psychophysics*, 82, 775–798.
- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1–3), 335–346.
- Herbort, O., & Butz, M. V. (2012). Too good to be true? Ideomotor theory from a computational perspective. *Frontiers in Psychology*, 3, 1–17.

- Hock, H. S., & Schöner, G. (2010). Measuring perceptual hysteresis with the modified method of limits: Dynamics at the threshold. *Seeing and Perceiving*, 23(2), 173–195.
- Hock, H. S., & Schöner, G. (2023). The stabilization of visibility for sequentially presented, low-contrast objects: Experiments and neural field model. *Journal of Vision*, 23(8), 12.
- Hock, H. S., Schöner, G., & Hochstein, S. (1996). Perceptual stability and the selective adaptation of perceived and unperceived motion directions. *Vision Research*, 36, 3311–3323.
- Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science*, 20, 568–577.
- Knips, G., Zibner, S. K. U., Reimann, H., & Schöner, G. (2017). A neural dynamic architecture for reaching and grasping integrates perception and movement generation and enables on-line updating. *Frontiers in Neurobotics*, 11, 1–14.
- Laird, J. E. (2012). *The SOAR cognitive architecture*. A Bradford Book.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, 1–72.
- Lebiere, C., O'Reilly, R. C., Jilk, D. J., Taatgen, N., & Anderson, J. R. (2008). The SAL integrated cognitive architecture. In *AAAI Fall Symposium: Biologically Inspired Cognitive Architectures* (pp. 98–104).
- Levy, S. D., & Gayler, R. (2008). Vector symbolic architectures: A new building material for artificial general intelligence. In *Proceedings of the 2008 Conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference* (pp. 414–418). Amsterdam, The Netherlands: IOS Press.
- Lomp, O., Richter, M., Zibner, S. K. U., & Schöner, G. (2016). Developing dynamic field theory architectures for embodied cognitive systems with Cedar. *Frontiers in Neurobotics*, 10, 14.
- Michel, O. (2004). Cyberbotics Ltd. Webots™: Professional mobile robot simulation. *International Journal of Advanced Robotic Systems*, 1(1), 5.
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing Atari with deep reinforcement learning.
- Moulin-Frier, C., Fischer, T., Petit, M., Pointeau, G., Puigbo, J.-Y., Pattacini, U., Low, S. C., Camilleri, D., Nguyen, P., Hoffmann, M., Chang, H. J., Zambelli, M., Mealiar, A.-L., Damianou, A., Metta, G., Prescott, T. J., Demiris, Y., Dominey, P. F., & Verschure, P. F. (2017). DAC-h3: A proactive robot cognitive architecture to acquire and express knowledge about the world and the self. In *IEEE Transactions on Cognitive and Developmental Systems* (pp. 1005–1022).
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Nilsson, N. J. (2014). *Understanding beliefs*. MIT Press.
- O'Reilly, R. C., Hazy, T. E., & Herd, S. A. (2016). The LEABRA cognitive architecture: How to play 20 principles with nature and win! *Oxford Handbook of Cognitive Science*, 91, 91–116.
- Richter, M., Lins, J., & Schöner, G. (2021). A neural dynamic model for the perceptual grounding of spatial and movement relations. *Cognitive Science*, 45(10), e13045.
- Richter, M., Sandamirskaya, Y., & Schöner, G. (2012). A robotic architecture for action selection and behavioral organization inspired by human cognition. In *International Conference on Intelligent Robots and Systems (IROS)* (pp. 2457–2464). IEEE.
- Sabinasz, D., Richter, M., & Schöner, G. (2024). Neural dynamic foundations of a theory of higher cognition: The case of grounding nested phrases. *Cognitive Neurodynamics*, 18, 557–579.
- Sandamirskaya, Y. (2014). Dynamic neural fields as a step toward cognitive neuromorphic architectures. *Frontiers in Neuroscience*, 7(276), 1–13.
- Sandamirskaya, Y., & Schöner, G. (2010). An embodied account of serial order: How instabilities drive sequence generation. *Neural Networks*, 23(10), 1164–1179.
- Schlegel, K., Neubert, P., & Protzel, P. (2022). A comparison of vector symbolic architectures. *Artificial Intelligence Review*, 55(6), 4523–4555.
- Schöner, G. (2019). The dynamics of neural populations capture the laws of the mind. *Topics in Cognitive Science*, 12, 1257–1271.

- Schöner, G., Bildheim, L., & Zhang, L. (2024). Toward a neural theory of goal-directed reaching movements. In M. F. Levin, M. Petrarca, D. Piscitelli, & S. Summa (Eds.), *Progress in motor control: From neuroscience to patient outcomes* (pp. 71–102). Academic Press.
- Schöner, G., Spencer, J. P., & DFT Research Group, T. (2016). *Dynamic thinking: A primer on dynamic field theory*. New York: Oxford University Press.
- Schöner, G., Tekülve, J., & Zibner, S. K. U. (2019). Reaching for objects: A neural process account in a developmental perspective. In D. Corbetta & M. Santello (Eds.), *Reach-to-grasp behavior* (pp. 281–318), *Frontiers of Developmental Science*. Routledge.
- Searle, J. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press.
- Searle, J. (2001). *Rationality in action*. MIT Press.
- Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, 136(6), 943–974.
- Shridhar, M., Manuelli, L., & Fox, D. (2022). Perceiver-actor: A multi-task transformer for robotic manipulation.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, Ilya and Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.
- Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, 46(1–2), 159–216.
- Spivey, M. J. (2023). Cognitive science progresses toward interactive frameworks. *Topics in Cognitive Science*, 15(2), 219–254.
- Szabados, A., Herd, S., Vinokurov, Y., Lebiere, C., & O'Reilly, R. C. (2020). Integrating systems and theories in the SAL hybrid architecture. *Common Model of Cognition Bulletin*, 1(1), 87–94.
- Taatgen, N. A., & Anderson, J. R. (2008). Constraints in cognitive architectures. In R. Sun (Ed.), *Cambridge Handbook of Computational Psychology* (pp. 170–185). Cambridge University Press.
- Tekülve, J., Fois, A., Sandamirskaya, Y., & Schöner, G. (2019). Autonomous sequence generation for a neural dynamic robot: Scene perception, serial order, and object-oriented movement. *Frontiers in Neurorobotics*, 13, 95.
- Tekülve, J., & Schöner, G. (2020). A neural dynamic network drives an intentional agent that autonomously learns beliefs in continuous time. *IEEE Transactions on Cognitive and Developmental Systems*, 14(1), 90–101.
- Trafton, G., Hiatt, L., Harrison, A., Tamborello, F., Khemlani, S., & Schultz, A. (2013). ACT-R/E: An embodied cognitive architecture for human–robot interaction. *Journal of Human-Robot Interaction*, 2(1), 30–55.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.
- Treisman, A. M., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, 34(8), 1704–1719.
- Verschure, P. F. (2012). Distributed adaptive control: A theory of the mind, brain, body nexus. *Biologically Inspired Cognitive Architectures*, 1, 55–72.
- Vogel-Blaschka, D., Kunde, W., Herbort, O., & Scherbaum, S. (2024). Ideonomic: An integrative computational dynamic model of ideomotor learning and effect-based action control. *Psychological Review*, 131(1), 79–103.
- Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13, 55–80.
- Zeng, X., Diekmann, N., Wiskott, L., & Cheng, S. (2023). Modeling the function of episodic memory in spatial learning. *Frontiers in Psychology*, 14, 1160648.
- Zibner, S. K. U., & Faubel, C. (2015). Dynamic scene representations and autonomous robotics. In G. Schöner & J. P. Spencer (Eds.), *Dynamic thinking: A primer on dynamic field theory* (Chap. 9, pp. 227–246). New York: Oxford University Press.