A neural dynamic network drives an intentional agent that autonomously learns beliefs in continuous time

Jan Tekülve and Gregor Schöner
Institute for Neural Computation
Ruhr-University Bochum
Bochum, Germany
{jan.tekuelve, gregor.schoener}@ini.rub.de

***Abstract*—Autonomous learning is the ability to form knowledge representations solely through one's own experience. To autonomously learn, an agent must be able to perceive, act, memorize, plan, and desire; it must be able to form intentional states. We build on a neural process account of intentionality, in which intentional states are stabilized by interactions within populations of neurons that represent perceptual features and movement parameters. Instabilities in such neural dynamics induce sequences of intentional behavior. In this paper, we examine the neural process organization required to decide and control when learning takes place, to build the representations that can hold learning data, and to organize the selection of neural substrate to learn the novel patterns. We demonstrate how a neural dynamic network may learn new beliefs about the world from single experiences, may activate and use beliefs to satisfy desires, and may deactivate beliefs when their predictions do not match experience. We illustrate the ideas in a simple toy scenario in which a simulated agent autonomously explores an environment, directs action at objects, and forms beliefs about simple contingencies in this environment. The agent utilizes learned beliefs to satisfy its own fixed desires.**

***Index Terms*—autonomous agent, fast learning, learning beliefs, neural dynamics, neural cognitive architecture**

I. Introduction

The current renaissance of neurally inspired architectures for perception and decision making is based on exploiting rich data sets that enable the adjustment of large numbers of parameters in a form of optimization often loosely called "learning". In autonomous learning, a system acquires such data itself, by behaving and actively steering its perception. Autonomous learning largely remains an open challenge for neurally inspired systems.

Organisms learn autonomously during development and in skill acquisition, both processes that require much time and entail many instances of experience. Humans and other animals may, however, also learn and generalize from single instances of an experience. This happens, for instance, when beliefs are formed about contingencies in the world. Imitation is effective because a single demonstration may be sufficient to uncover the underlying contingencies. In fact, belief formation can be so fast as to be sometimes counter-productive, leading to superstitious behavior (even in pigeons [1]). On the other hand, this ability to efficiently form beliefs about the environment provides enormous cognitive power [2].

Pervasively neural accounts of learning are far from approximating this form of single-shot learning in an autonomous manner. Fast learning has been demonstrated for object recognition based on exploiting prior knowledge about the visual appearance of objects [3], [4] or on built-in knowledge about the transformations that enable generalization [5]. Early in development, contingency learning may be a way how infants break down the complexity of the world, which supports their social skills [6]. Our framing of belief acquisition approximates such contingency learning.

A first reason, why neurally inspired systems fail to learn autonomously is that they fail to perceive, think, and act autonomously! The process structure to generate behavior typically remains outside the neural metaphor so that neural processes of learning cannot be brought to bear. Behavior generated by fixed algorithms that are driven by sense data does not necessarily provide meaningful experience that links perception to goals and expectations. We propose that the philosophical notion of *intentionality*, the capacity to generate internal states that are about the world, helps to uncover the requisite process structure. The philosopher John Searle divides intentional states into two classes: The mind-to-world direction of fit, which includes perceptual states representing the world, and the world-to-mind direction of fit, which includes motor intentions representing desired world states [7].

We have previously analyzed the neural process requirements for intentional states of both directions of fit [8] (see [9] for a different take). Our analysis was based on Dynamic Field Theory (DFT) [10], a set of mathematical concepts to model the neural dynamics of networks of neural populations1. In particular, we exploited the notion that intentional states are stable patterns of neural activation that may transition sequentially to other intentional states by inducing dynamic instabilities through a neural representation of the *condition of satisfaction* [11]. This mechanism gives neural dynamic architectures the potential to autonomously generate behavioral sequences (see [12] for a discussion of autonomy).

The scope of intentional states we attempt to capture roughly aligns with the six psychological modes analyzed by Searle: present states (perception, intention-in-action), time-shifted states (memory, prior intention), and abstracted states (belief, desire). Our previous analysis led to a neural process account for perception, memory, intention-in-action, and prior intention. In this paper, we argue that a fast form of learning captures the psychological mode of belief. For us, the formation of belief resembles the formation of a memory, but generalizes beyond the specific instance by being categorical and propositional in nature. We are able to construct propositional content in a neural dynamic architecture by organizing underlying processes at the level of concepts through the condition of satisfaction [13].

To autonomously learn beliefs in a neural architecture we must address also a second set of problems faced by neural models of autonomous learning, providing the process infrastructure that steers and organizes the autonomous learning process itself. This entails the (a) detection of the events that indicate that a learning episode must be initiated. That may be achieved by generating intrinsic reward upon detection of such events. Autonomous learning also entails (b) providing neural representations of the current action of the agent, of the previous, and of the new state of the environment that together specify a contingency. It entails (c) process support for the activation of beliefs based on perceived cues or goals. And it entails (d) process support for detecting the novelty of an observed contingency, and an associated process that provides fresh neural substrate to learn a new belief. This, in turn, requires that (e) neural substrate already committed to the representation of previously learned beliefs must be demarcated from neural substrate for learning.

We model the autonomous acquisition of beliefs including the activation of existing beliefs to guide action, the rejection of activated beliefs when their predictions are not confirmed, and the formation of new beliefs. The account is framed within a rudimentary toy scenario in which an agent is situated in a simple environment containing solely paint buckets and canvases of different original colors. The robotic agent explores the environment, moves toward objects and directs a robotic arm to them to either pick-up paint from buckets or to dispense paint onto canvases, while visually observing the resulting canvas color. A network of neural dynamic fields is connected to the agent's sensory-motor surfaces and enables the agent to visually detect and select objects, build scene memories, generate sequences of actions to paint particular objects to achieve a particular result color and ultimately to form and activate beliefs about which paint applied to which canvas generates which outcome.

II. Dynamic Field Theory

Our goal is to provide a neural process account of belief acquisition that is compatible with general principles of brain function, but not necessarily anatomically specific and not at the lowest level of neural reduction such as spiking neural networks. Broadly, we postulate that neural processes are characterized by graded patterns of activation that evolve continuously in time, that may be coupled to sensory and motor processes, and that have stability properties. This amounts to identifying small populations of neurons bound into networks as the preferred level of description that is most closely reflective of actual behavior and cognition (see [14] for review). We specifically avoid shortcuts, in which neural processes are replaced or approximated by processes of information processing, such as when algorithms simulate a hypothesized neural function.

The mathematical formalization of these neural principles is provided by Dynamic Field Theory (DFT) [10]. In DFT, neural populations tuned to metric dimensions, x, are modeled by activation fields, $u(x,t)$, that evolve in time according to the neural dynamics:

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

The time-continuous evolution of neural activation, $u(x)$, on the time scale, τ, relaxes to the stable solution, $h + s(x)$, defined by the field's resting level, h, and its localized inputs, $s(x)$, if the current activation level, $u(x)$, is below the threshold ($= 0$) of the sigmoidal transfer function, $\sigma(u)$. Field locations with activation surpassing the threshold engage in lateral interaction defined by the field's kernel, $\omega(x-x')$, which is excitatory locally (for small $|x-x'|$), and inhibitory over longer distances, $x - x'$. This leads to the emergence of self-stabilized peaks of supra-threshold activation that are the units of representation in DFT (illustrated in Figure 1). Supra-threshold peaks arise as the sub-threshold state goes through the *detection instability*.

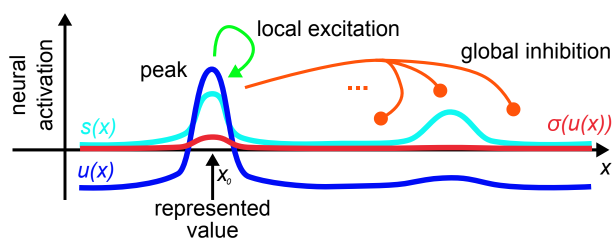

Fig. 1. A dynamic neural field spanned across a metric-dimension, x, represents a metric value, x_0, when a supra-threshold peak of activation peak is localized there.

Depending on the strengths of excitatory and inhibitory interaction, fields operate in different regimes. In the self-stabilized regime, supra-threshold peaks are stabilized against input noise. In the selective regime, lateral inhibition allows only a single peak at any point in time. In the self-sustained regime, peaks are retained after localized input is removed. Peaks in multi-dimensional fields represent conjunctions of feature dimensions. For instance, a peak in a two-dimensional field defined over both color and position represents a particular color seen at a particular position. Dynamic neural nodes are zero-dimensional fields that represent categorical states.

A field, u_{tar}, receives input from another field, u_{src}, if that field's output, $\sigma(u_{\mathrm{src}})$, adds to the target field's rate of change, $\dot{u}_{\mathrm{tar}}$, weighted with a homogeneous projection kernel $\omega_{\mathrm{tar,src}}$. The source output might need to be contracted or expanded to match the target field's dimensionality [15]. Typically, contractions entail integrating over the excess dimension, while expansions provide input that is constant along the excess dimensions (e.g., ridges, tubes, or slices). *Concept nodes* are connected reciprocally to fields through a pattern of connectivity that encodes the feature representation of the concept. For instance, the concept node for "blue" is connected to an appropriate range of hue values in a hue feature field (see Figure 2).

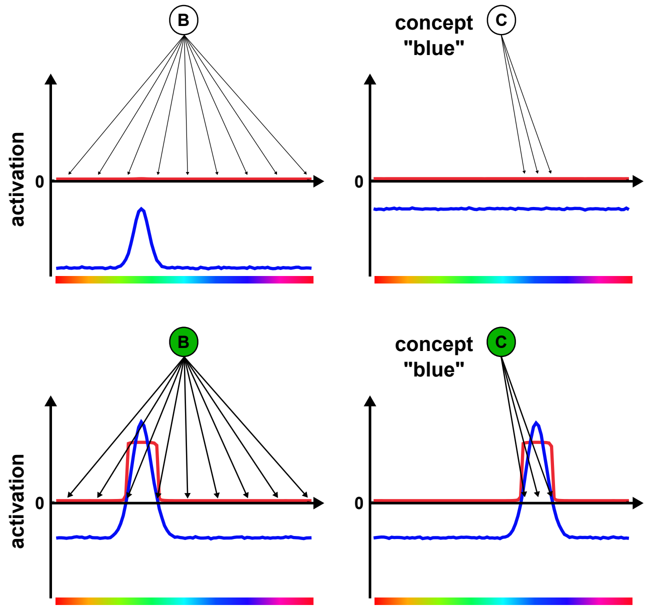

Fig. 2. Boost and Concept Nodes. A boost node (left column) affects the entire resting level of field and may push present activation patterns above threshold. A concept node (right column) excites a specific feature range within a field.

A. Networks of fields form architectures

Networks of dynamic neural fields may connect to the sensory-motor surfaces of an agent. It is through these connections that ultimately the dimensions emerge over which each field is effectively defined. Behavior may emerge from such networks as activation patterns transition between different stable states, each represented by peaks of supra-threshold activation. *Boost nodes* provide homogeneous input to a target field. They may induce state transitions by altering the dynamic regime of the target field. Most commonly, the detection instability de-stabilizes the sub-threshold activation pattern, leading to the formation of a localized supra-threshold peak of activation. Boost nodes may effectively modulate the flow of activation within an architecture by enabling or disabling particular branches of the architecture to form peaks. Boost nodes may thus act as gates or go-signals that trigger an action by activating a sub-network (see Figure 2). Pairs of fields, an excitatory *intention* field and an inhibitory *condition of satisfaction* (CoS) field, control the initiation and termination of actions or mental states [16] (see Figure 3). The intention field represents the desired end state of a particular action and activates a sub-network that ultimately realizes the desired action. The intention field pre-activates the CoS-field, in which a peak is formed when desired and perceived state overlap sufficiently. A peak in the CoS-field inhibits the intention field, destabilizing the peak there and deactivating the associated sub-network, which terminates the action. The CoS-field inhibits any precondition node that prevented competing

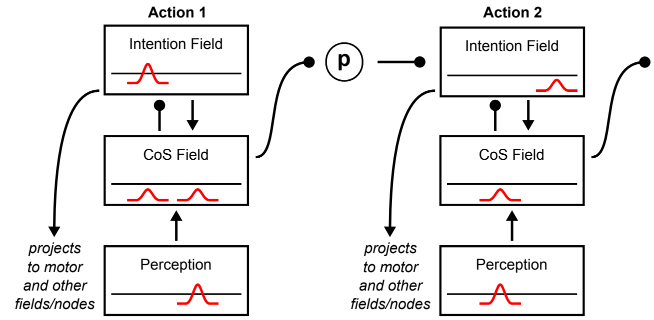

Fig. 3. Two consecutive elements of an action sequence each represented through a pair of intention and CoS field. "P" is a precondition node that organizes the fixed sequentiality of the actions.

actions from becoming activated. This unlocks the next step in a sequence.

Other than through the CoS mechanism, autonomous transitions between macro states may also be induced by neural representations of a *condition of dissatisfaction* (CoD) which detects failed actions or invalid perceptual states. Transitions may also arise from transient sense-data, such as when a surface color changes in the scene. Transient detectors consist of a pair of excitatory and inhibitory fields, defined over the same metric dimensions and sharing input, but differing in their time scale with inhibition slower than excitation. A step increase of input then first induces a peak of activation in the excitatory field, which is later erased by the peak of activation that arises somewhat more slowly in the inhibitory field (Figure 4). The supra-threshold activation in the excitatory

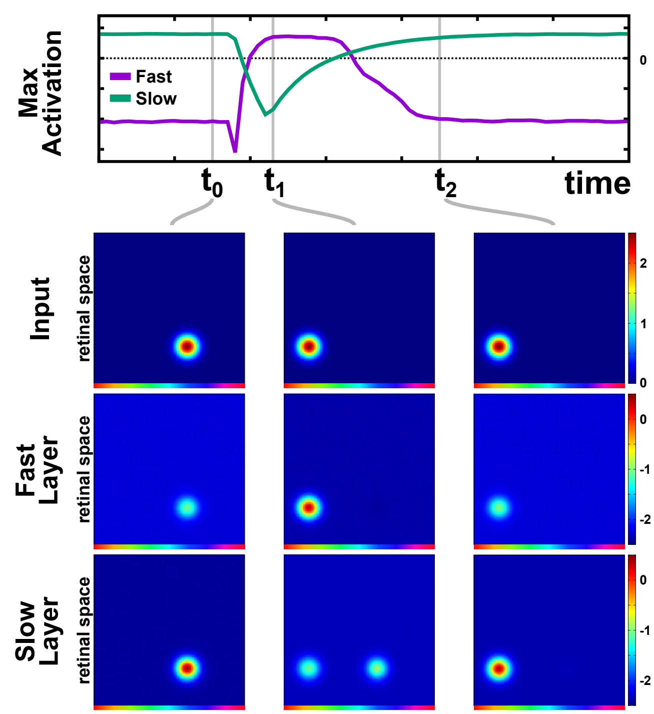

Fig. 4. Activation snapshots of a two-layer transient detector as it detects a color change from blue to yellow. The color change occurs shortly before t_1 and is detected through supra-threshold activation in the fast layer.

field signals the transient. More generally, pairs of excitatory and inhibitory activation fields provide the generic mechanism for generating temporally structured activation patterns which may be used, for example, to model a transient reward signals or ballistic velocity profiles (see Chapter 3 in [10]).

III. Scenario and model

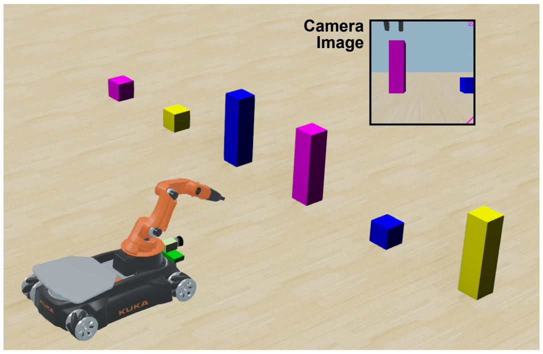

Fig. 5. The simulated robot in its environment

We develop our ideas around a very simple, but carefully chosen toy scenario, which is simulated in continuous time (see Figure 5). A robot vehicle may move along a line only, for simplicity. It has an arm that the robot can point at objects (actually, with carefully timed trajectories). Two actions of the end-effector, picking up and depositing paint, are modeled merely as state changes of associated variables. These are the motor systems, toward which intentions-in-action are directed. Attentional selection in visual search has the same direction of fit and is thus part of the set of intentions-in-action. A final intention-in-action is spatial exploration, in which the vehicle moves in a random direction along the one-dimensional world line. (You may follow along with these different behaviors by comparing with neural dynamic architecture schematically illustrated in Figure 6.)

Only the vision sensor is modeled in detail, while we assume that sensor readings are provided for the vehicle's position along the one spatial dimension, for the arm's end-effector position, and for the presence of paint in the arm's end-effector. Objects have two features, for which we postulate visual feature detectors: They may be short (for canvasses) and tall (for paint buckets). (It may have been more intuitive to have short paint buckets and tall canvases, but somehow we selected this mapping at some point.) Perceptual processes represent the states of these sensory dimensions.

We have endowed the agent with a few multi-step behaviors, called prior intentions in Searle's jargon, which include collecting paint of a given color, applying paint to a canvas of a given color, and moving toward an object of a given feature value along either the color and/or height dimension. These prior intentions make use of memory processes that effectively generate a scene representation, that is, a neural map of locations at which objects with associated feature values are positioned in the one-dimensional world, irrespective of whether the objects are currently in view or not.

The desire of the agent is simply to see canvases of a particular color, a very simple representation of a task or goal. The agent may satisfy such a desire by directly searching and finding objects of the desired color, or by painting a canvas of a different surface color with an available kind of paint to achieve the desired result color.

This is where beliefs come in. The agent learns contingencies that are essentially arbitrary rules of coloring mixing, that is, predictions of what kind of result color is obtained when a coat of paint of a particular color is applied to a canvas of a particular initial surface color. (The simulated world does not follow real-world rules of color mixing.)

The scenario thus probes intentional states of both directions of fit, world-to-mind and mind-to-world, containing the six psychological modes analyzed by Searle: perception, intention-in-action, prior intention, desire, perception, memory, and belief. We exemplify the autonomous formation of beliefs from a single experience, their activation and use to achieve goals, and their rejection (deactivation) when predictions do not match outcomes.

A. Neural dynamic architecture

The neural dynamic implementation of intentionality in the four basis level psychological modes of perception, memory, intention-in-action, and prior intention was described previously [8]. Here we provide a brief sketch only to then focus primarily on the neural dynamics of beliefs in the next subsection.

Although Figure 6 sketches the overall architecture, many details are hidden from view. We discuss here selected elements, but make the complete architecture together with all parameter values available for download [1]. The architecture is one large integro-differential equation, built from the elements of DFT and thus consistent with its principles. This constraint is formalized by using *cedar*, a software framework that provides a graphical programming interface to building and numerically solving DFT architectures [17]. This functionality makes it possible to individually parameterize each field in an architecture to achieve one of the three operational regimes: self-stabilized, sustained, or selective. *Cedar* is available as an open-source, public licence project[2]. The simulation of the robot, of its sensors, and the environment makes use of the *Webots* [18] simulator, which was connected to *cedar*.

1) Perception: Estimates of body states are represented by unique localized peaks of activation in fields defined over the relevant metric dimensions. This includes the current Cartesian position of the robot's position along the one-dimensional world and the robot arm's end-effector position. Discrete states are represented by dynamic neural nodes. This includes the states of the paint tool and of the change detector.

Visual perception occurs in the camera frame (called *retinal* frame in the figures). For the currently visible part of the visual array, the horizontal retinal position, hue value, and height feature of objects that differ from the background are provided by sensory channels. These feed into two two-dimensional neural fields defined over horizontal retinal space and the respective feature dimension. Within these fields, a single peak

[1]https://www.ini.rub.de/the_institute/people/jan-tekulve/
[2]https://dynamicfieldtheory.org/software

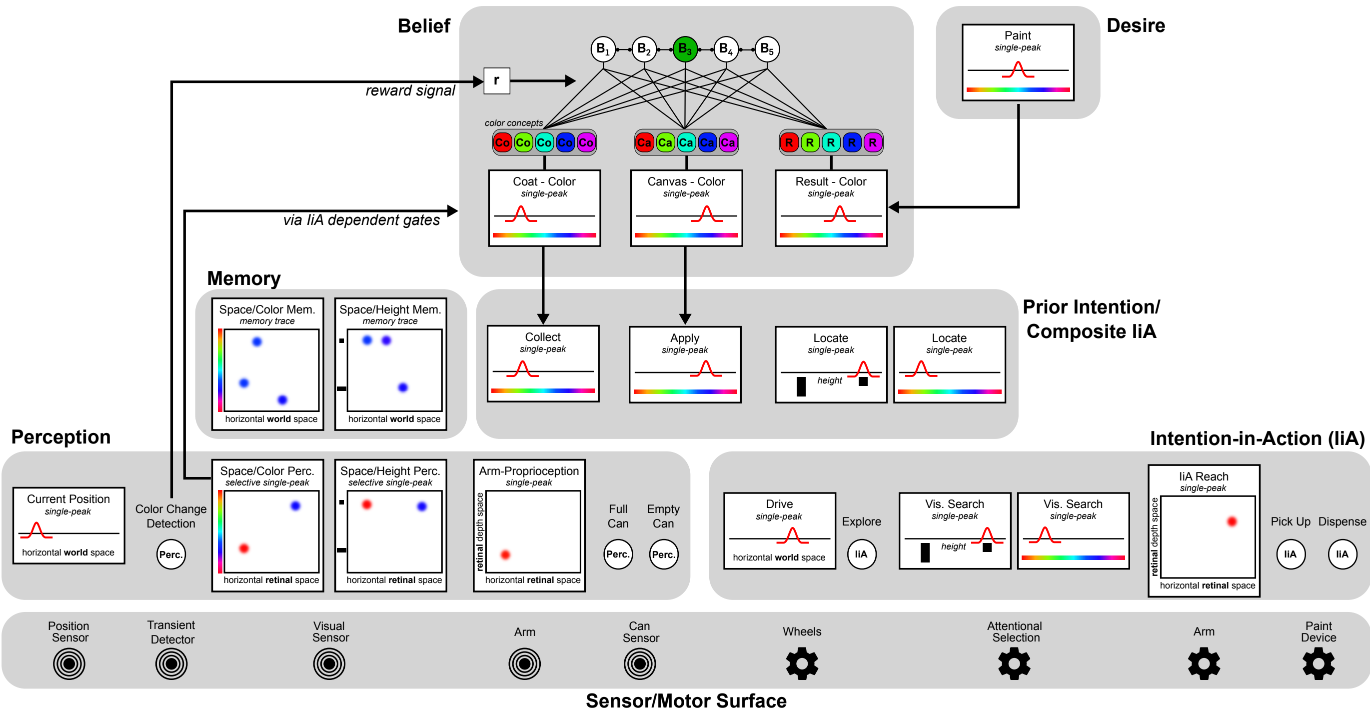

Fig. 6. Schematic overview over the neural dynamic architecture with the six psychological modes, perception, intention-in-action, memory, prior intention, belief, and desire. Only a subset of the connections are shown. The belief network is blown up in more detail in Figure 7.The abbreviation "IiA" is used for intention-in-action.

is activated at one location, modeling attentional selection. There is a small architecture hidden here, that is capable of sequentially exploring the visual array, of maintaining and updating visual working memory, and of performing visual search within the visual array based on a cued feature value. This is a simplified version of a fuller architecture for visual search and scene memory [19]. Transient detection occurs on the space/color representation and enables the agent to detect changes in color and location of objects in the visual array and to direct spatial attention to the location of the change.

2) Memory: Scene memory is based on the same features, color and height, but now represented in a world rather than a camera frame. The world frame is appropriate for scene memory as it is invariant under displacement of the agent. The coordinate transform [20] makes use of the current estimate of the vehicle's ego-position. Peaks are induced in the transformed *space/feature memory* fields whenever an object is attended to.

These peaks leave a memory trace, a localized increase of the resting level of the field that facilitates peak generation at the peak locations. The dynamic of the memory trace is a simple learning mechanism that has been used to understand inter-trial effects and the building of priors on time scales faster than those associated with the classical Hebbian learning rule [21] (for review, see [22]). The memory trace is subject to interference so that building a memory trace at a new location leads to the decay of memory traces elsewhere in the field. Full, self-stabilized peaks of activation can be brought up from the memory trace by globally boosting activation in the field, effectively pushing activation through the detection instability at locations at which a memory trace has been laid down. In the model, memories of object locations and their visual feature values are built as memory traces in the *space/feature memory* fields. Cues to an object's color or height provide ridge input to these fields, localized along the feature dimension but constant along space. The ridge input induces peaks at locations at which memory traces have been laid down, leading to recall of the memorized position and feature value of a matching object.

3) Intentions-in-action: Each action of the robot is specified by a pair of intention and condition of satisfaction (CoS) fields connected to relevant sub-networks. Reaching and driving to a position are realized by sub-networks simplified from [23]. Their main component is a pair of coupled dynamic fields that realize an active transient, which is turned into a velocity command for either the joint angles of the arm or the wheels of the vehicle. The simulated actions to collect and apply paint manifest themselves through a change of the fill status of the painting device. In addition to driving to a specified position, the robot may also explore its environment by moving in one of the two directions along the one-dimensional world line, until a previously unattended object is detected. These last three kinds of intentions-in-action are categorical in nature and are thus represented by neural nodes.

Not all intentions-in-action are directly motoric in nature. *Visual search*, for instance, is an intention-in-action in that it has the direction of fit world-to-mind: the intention is satisfied when the state of the agent's nervous system, that is part of the world, matches the intended contents (e.g., a peak representing a red object has been activated in response to visually searching a red object).

4) Prior Intentions: Sequences of actions are organized by neural fields that represent prior intentions (or composite intentions-in-action). These fields project excitatorily onto the intention-in-action-fields that are part of the sequence. Their serial organization is organized by precondition neural nodes as illustrated in Figure 3. The intention-in-action not inhibited by any precondition node is the first to become activated, starting the sequence.

Painting an object entails first collecting a particular paint from a bucket, and then applying the paint to a canvas. Both actions themselves consist of sequences of actions: To collect paint, the agent must find and move to a tall object, point its arm to the object, and pick up paint. To apply a coat of paint, the agent must find and move to a small object, point its arm to the object, and dispense the paint. Finding and moving to an object is also a sequence of actions that entails recalling the object's location, driving to the location, and then visually searching for the object within the visual array.

Prior intentions may entail alternative sequences of actions, that emerge if a particular action terminates in failure. This is mediated by activation of its *condition of dissatisfaction* (CoD). For instance, activation of the CoD node of *recall* or *visual search* destabilizes the precondition node of the *explore* intention, allowing the explore behavior to become activated and the robot to move to new locations in search of a matching object.

What kinds of paint buckets and canvases are sought by the agent is determined by the goal intentions and the belief system.

B. Belief system

Beliefs are propositional in nature. In our terms, they are combinations of concepts, represented by neural nodes that project onto metric feature spaces such as space, color, or height. These projects enable the sensory-motor grounding of the concepts that beliefs tie together [13]. The formalization as a combination of abstract concepts distinguishes beliefs from individual memories. How are such sets of concepts formed into beliefs and how may an individual belief become activated based on grounded sensory-motor representation?

Figure 7 illustrates the neural-dynamic subsystem for belief formation, activation, and rejection. This architecture is partially inspired by Carpenter and Grossberg's Adaptive Resonance Theory (ART) [24]. Each belief associates color concepts in three roles: color of the paint (coat), color of the canvas before applying paint (canvas), and color of the canvas after applying paint (result). A belief is represented by a neural node. Forming a belief amounts to linking the belief node to three color concept nodes in the three roles by Hebbian learning. A belief is activated when the belief node goes through the detection instability in response to a subset of the color concept nodes, to which the belief node is connected. A belief is rejected when the belief node falls below the reverse detection instability, due to inhibitory input from the CoD system. We explain the inner dynamics of the belief system by stepping through Figure 7 from bottom to top.

The belief learning sub-network couples into the neural-dynamic architecture through three working memory fields, $u_{\text{role}}(c)$, defined over color, c (hue), in three roles: $\text{role} \in \{\text{coat}, \text{canvas}, \text{result}\}$. Their dynamics

$$\begin{aligned}\tau\dot{u}_{\text{role}}(c) =& - u_{\text{role}}(c) + h_{\text{role}} + w_{\text{pai}}\sigma(u_{\text{pai}}) \\ &+ \int \omega_{\text{role}}(c-c')\sigma(u_{\text{role}}(c'))dc' \\ &+ \sum_{\text{color}} w_{\text{color}}(c)\sigma(u^{\text{role}}_{\text{color}}) + w_{\text{gate}}\sigma(u^{\text{role}}_{\text{gate}}(c)).\end{aligned} \tag{1}$$

is globally controlled by the *paint task* node, u_{pai}, so that peaks are only generated when that task node is active. The interaction kernel, ω_{role}, is chosen to put the field into the selective and self-sustaining dynamic regime. These fields are connected to the color perception field each via a gating field, u_{gate} (not shown in Figure 7 but hinted at in Figure 6), that is activated whenever pick-up, dispense, or color change detection nodes are active. In each case, a single peak is formed in the color role field that reflects the applied or perceived colors.

The color concept nodes, $u^{\text{role}}_{\text{color}}$ are reciprocally connected to the respective color role fields through weight vectors, w_{color}, set up a priory to represent color categories for yellow, green, orange, cyan, blue, purple, pink, and red:

$$\begin{aligned}\tau\dot{u}^{\text{role}}_{\text{color}} =& - u^{\text{role}}_{\text{color}} + h_{\text{con}} + \sum_i l^{\text{role}}_{i,\text{color}}\ \sigma(b_i) \\ &+ w_{\text{rcl}}\ \sigma(u_{\text{rcl}}) + \int w_{\text{color}}(c)\ \sigma(u_{\text{role}}(c))dc.\end{aligned} \tag{2}$$

These nodes are, in turn, reciprocally connected to the belief nodes, b_i, with plastic connection strengths, $l^{\text{role}}_{i,\text{color}}$, whose dynamics we will discuss shortly. Incoming connection strengths and resting level, h_{con}, are set-up such that a concept node may be activated either through a peak in the role field, u_{role}, that matches the color encoded in w_{color} or through the combination of an activated associated belief, b_i and the activated recall boost, u_{rcl} discussed below.

Each belief is represented by a *belief* node, b_i that is connected to the concept nodes via (reciprocal) plastic connections, $l^{\text{role}}_{i,color}$, initialized to zero. The belief nodes dynamics,

$$\begin{aligned}\tau\dot{b}_i =& - b_i + h_b + w_b\sigma(b_i) + w_{\text{com}}\sigma(u_{\text{com}}) - w_{\text{c}}\sigma(c_i) \\ &- w_{\text{inh}}\sum_{j\neq i}\sigma(b_j) - w_{\text{cod}}\sigma(u_{\text{cod}}) \\ &+ \sum_{\text{role}}\sum_k \left[l^{\text{role}}_{i,k}\sigma(u^{\text{role}}_k) - w_{\text{inc}}\sigma(u^{\text{role}}_k)\right],\end{aligned} \tag{3}$$

entails self-excitation of strength w_b, inhibitory coupling to all other belief nodes of strength w_{inh}, excitatory input from the commit node of strength, w_{com}, inhibition from the corresponding commit state node of strength w_{c}, and inhibition from the condition of dissatisfaction node of strength w_{cod}. In addition to the excitatory plastic connections, $l^{\text{role}}_{i,color}$, each color concept node, u^{role}_k, contributes fixed inhibitory input to each belief node with strength w_{inc}. That inhibition is stronger than any unlearned connection, $l^{\text{role}}_{i,color}$, but weaker than a learned connection $l^{\text{role}}_{i,color}$. This normalization mechanism

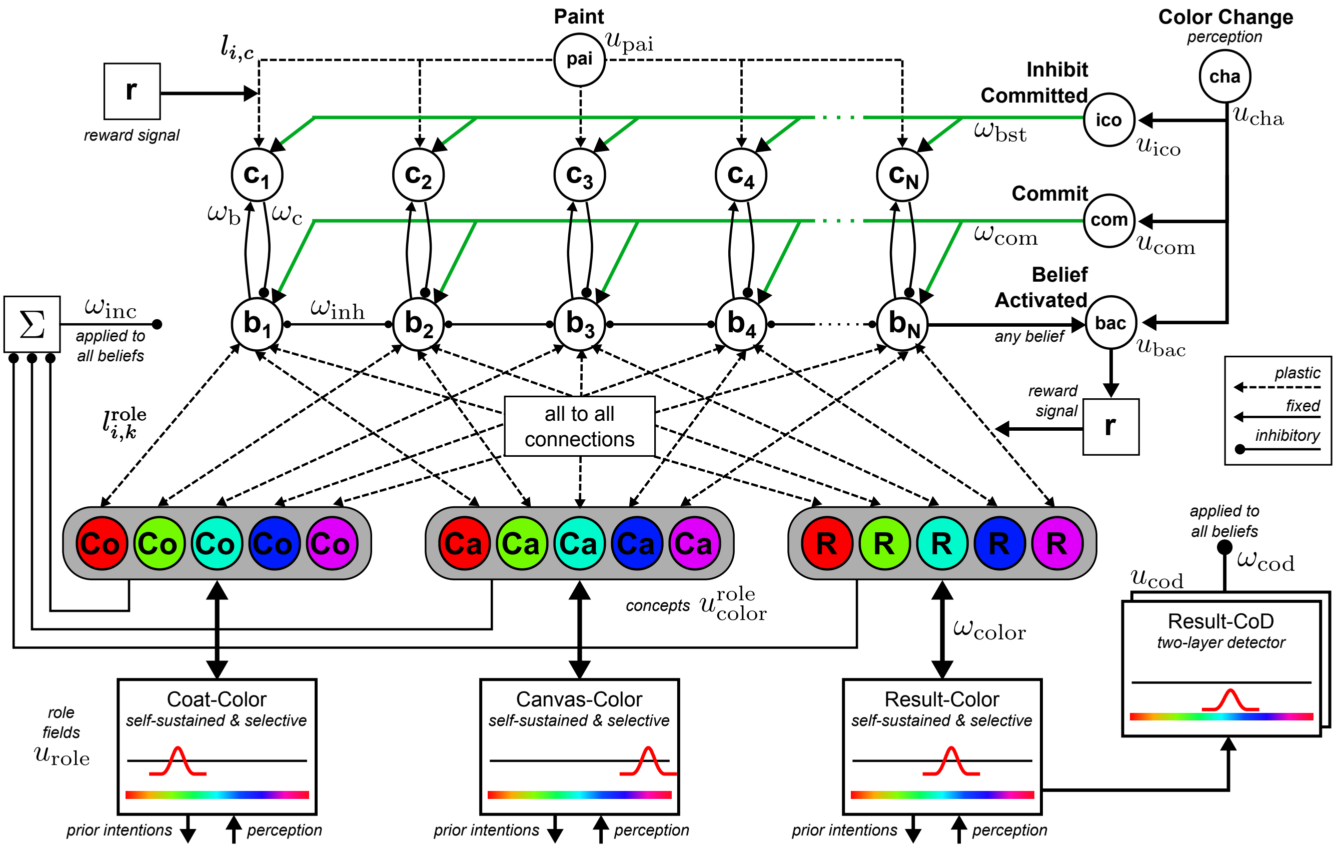

Fig. 7. The sub-network responsible for belief formation, activation and rejection.

make that when a color concept node is active, all belief nodes that are not (yet) associated with that concept node are inhibited, while all belief nodes already associated through a learned connection, $l^{\text{role}}_{i,color} > w_{\text{inc}}$, are excited. A single inhibitory input is sufficient to prevent a belief node from activating. Therefore, when n concept nodes are active, only belief nodes that are associated with all n concepts may become activated.

1) Forming a belief: Belief nodes become linked to color concepts whenever a sequence of collecting and applying paint results in a color change in the scene. Such color change is detected by the agent's change detector (see Figure 4), which activates the *color change* node, u_{cha}, and the gate that propagates the perceived result color to the result role field. This leads to the activation of the corresponding result color concept node. Ultimately, one concept node for each role is active, representing the color triple to be learned. If the triple has previously been associated with a belief node, b_i, that node will be activated via the learned reciprocal connections (see below). This leads to updating of the connectivity of that node to the color concepts. That updating process is controlled by the *belief activated* node, u_{bac}, which is now pushed through the detection instability, causing the generation of a transient reward signal, $r(t)$, based on the transient detector mechanism illustrated in Figure 4. The reward-gated Hebbian learning rule

$$\dot{l}^{\text{role}}_{i,k} = -\eta \; r(t) \; \sigma(b_i) \; (l^{\text{role}}_{i,k} - \sigma(u^{\text{role}}_k)) \qquad (4)$$

has a learning rate, η, that is sufficient fast to learn a perceived rule in a single presentation (one-shot learning) (see [22] for a review of neural dynamic learning rules). The stability of the learned state, $l^{\text{role}}_{i,k} = \sigma(u^{\text{role}}_k)$, makes that the updating discussed here leads only to minor change when the belief had already been learned.

If the color triple has not previously been associated with a belief node, no belief node will be active at the time of the color change. A process of selecting a new belief node to learn the new triple is initiated by activating the *commit* node, u_{com}, and the *inhibit committed* node, u_{ico}, both driven by the change detector, u_{cha}. The commit node, u_{com}, boosts all belief nodes. Lateral inhibition among belief nodes results in the selection of a single belief node. Because committed belief nodes are inhibited, only previously uncommitted belief nodes may win the competition. This is the aspect of the architecture that is inspired by ART.

The mechanism is implemented through *commit state* nodes, c_i, whose neural dynamics

$$\tau \dot{c}_i = -\, c_i + h_c + w_b \sigma(b_i) + w_{\text{bst}} \sigma(u_{\text{ico}}) + l_{i,c} \sigma(u_{\text{pai}}) \quad (5)$$

receives excitatory input from the matching belief node, b_i, and the *inhibit committed* node, u_{ico}, that evolves on a faster timescale.

Which *commit state* node is activated is determined by plastic projections, $l_{i,c}$, from the paint task node, u_{pai}, to each of the *commit state* nodes. These learn the commit pattern through an analogous reward-modulated Hebbian rule,

$$\dot{l}_{i,c} = -\eta\ r(t)\ \sigma(u_{\mathrm{pai}})\ (l_{i,c} - \sigma(c_i)). \tag{6}$$

Eventually, one previously uncommited belief node is selected and self-sustained. It activates its corresponding *commit state* node and triggers the reward signal through u_{bac}. The selected belief node learns the active color concept triple based on the Hebbian rule, Eq. 4, while its *commit state* node becomes connected to the *paint task* node, u_{pai}, based on the Hebbian rule, Eq. 6.

2) Belief Activation: The learned reciprocal connections between concept and belief nodes enable the activation of beliefs that partially match the current state of the color concept nodes. A critical feature for this is inhibitory feedback to belief nodes from the color concept nodes (the last term in Equation 3). For example, if input from a *desire* node activates a particular *result color concept* node, $u^{\mathrm{result}}_{k_0}$ (so that the system is now looking for beliefs that enable the generation of that result color), then all *belief* nodes receive inhibition through w_{inc}. *Belief* nodes that have previously learned connections to that result color concept, $l^{\mathrm{result}}_{i,k_0} \neq 0$, are excited, in contrast. Among these candidate beliefs, lateral inhibition promotes the selection of one belief. The homogeneous boost, u_{rcl}, to color role concepts enables activation of matching coat or canvas color concepts nodes, which in turn induce working memory peaks in the color role fields. These are projected down to the lower levels of intentionality controlling the search for paint buckets or canvases according to the activated belief. If a belief already receives input from a color role field (e.g., a matching canvas or paint color is in memory), that belief has the best chance of winning the competition.

3) Belief Rejection: Beliefs may not cover all cases. Some result colors may, for instance, arise from multiple combinations. Or the same combination could lead to different results, perhaps because of an undetected change in the environment. The system never truly forgets a belief but is able to reject an initially selected belief that does not work and to then learn a new belief that captures the newly observed contingency.

In the toy scenario, let us say the system has learned a color mixing rule associating coat color, a, canvas color, b, and result color, c. Desiring color, c, the system has activated this belief and acted on it. But as it performs the action, it now observes the result color, d, instead.

While executing the paint sequence, the color role fields have peaks at locations a, b, and c in the respective fields based on the active belief. Once a color change is detected, the observed new result color, d, induces a new peak of activation in the result color field at location d, which replaces the peak at c based on lateral inhibition within that field. This change of activation pattern in the result color field is detected through a transient detector that activates the CoD node, u_{cod} (illustrated on the bottom right of Figure 7). This inhibits all beliefs via w_{cod}, leading to the destabilization of the active belief. If the new pattern was not previously learned, a new, yet uncommitted belief node will become activated, the reward signal will be generated and the new pattern will be learned as a new belief.

C. Desires

In the present toy scenario, desires are imposed from the outside by setting a single peak into a field that represents a desired result color. This *desire* field provides input to the *result color* field of the belief system (see Figure 6). Satisfaction of the desire is signaled by a CoS field that matches the desired color with the currently perceived color. Inhibition from the CoS representation terminates the desire and allows for a new desire to form. Ideas from behavioral economics could be used to expand this system to multiple competing desires. One example of this would be the PSI architecture [25], which incorporates five different groups of needs competing with each other.

IV. Results

Our goal here is to demonstrate that the system is, in fact, capable of autonomously acting in its little toy world toward the fulfillment of desires, using beliefs, learning beliefs as needed, and rejecting beliefs when they fail to predict outcomes. The architecture is just one big dynamical system, which is numerically simulated in the simulated toy scenario. A critical element of these demonstrations is, therefore, that all sequences of actions, mental acts, and learning episodes emerge from the time-continuous, graded state dynamics through dynamic instabilities. We illustrate now key moments in a simulation run of the neural dynamics that high-light conceptually meaningful features of the system. The first demonstration illustrates how the roles of the elements of a belief come about. Then we look at the formation of a belief, the activation of a belief, and the rejection of a belief whose prediction fails.

A. How the roles are perceptually grounded

In our toy scenario, beliefs are about colors in three different roles, the color of paint, the color of a canvas, and the color that results when the paint is applied to the canvas. These three types of color concepts are not typically perceived at the same time. So unlike the typical neural network training algorithms, in which sensory stimuli are given from the outside as learning examples, autonomous learning of beliefs entails building internal memory representations of color in the corresponding roles as a learning episode unfolds through a sequence of physical and mental acts.

Figure 8 shows time courses of selected nodes and fields as the agent loads paint into its paint tool and thus perceptually grounds the color role, *coat*. At the time point, t_0, the agent is situated in front of a blue *canvas* cube. The cube's location has been selected by the *Attentional Selection* field in the retinal frame, its blue color is represented by a peak in the *Attended Color* field. This peak causes a sub-threshold activation bump in the *Pick-Up Gate* field but isn't loaded into that field, because no paint pick-up is happening (in fact, for a canvas object picking up paint is not a possible action). The *Coat Color* field is at resting level. The composite intention-in-action *Collect* is active and the supra-threshold state of the *Recall* CoS node signals that a bucket position has already been successfully retrieved from memory. This has activated

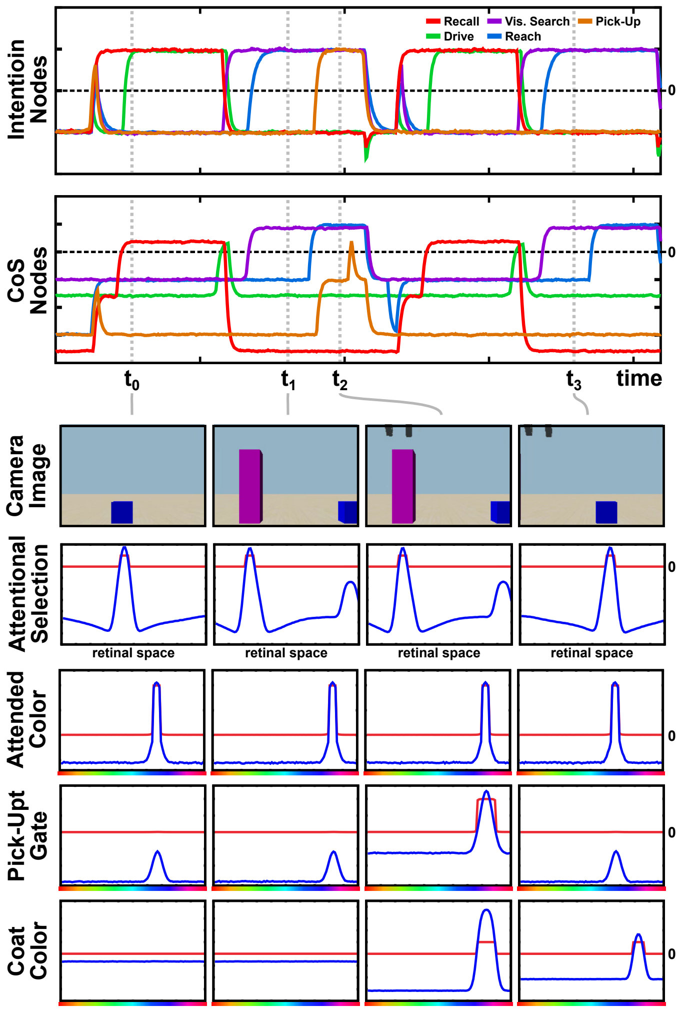

Fig. 8. Activation time course of selected dynamic nodes and fields displaying the retention of the collected color during a painting episode. The first two rows show the activation of the intention and CoS nodes of the actions that make up the painting sequence. The third row shows the simulated scene and camera image at four different points during the sequence. The bottom rows show activation snapshots of four different fields at the same points in time.

the intention-in-action *Drive*, which causes the agent to move towards the recalled position.

The next snapshot at t_1 shows the agent after it drove to face the purple paint bucket (tall object). A successful visual search caused the attentional selection of the bucket location and the representation of its color in the *Attended Color* field. Activation in the *Pick-Up* gate remains below threshold as the agent initiates a reach towards the bucket, which is signaled by the active *Reach* intention node.

At time $t2$, the reach has terminated and the *Pick-Up* intention node has been activated. This boosts the resting level in the *Pick-Up Gate* field causing a detection instability. The resulting supra-threshold peak near the color pink provides input to the *Coat Color* field. The *Pick-Up* intention node is inhibited by its CoS node once the simulated painting device has been filled with paint.

This destabilizes the peak in the *Pick-Up Gate*, but the coat color representation is retained through its strong lateral excitation. This can be seen at time t_3 when the agent engages in applying the stored paint onto the blue canvas. The agent thus builds step by step the neural representation of the different color roles that provide the interface to perceptually grounding a belief.

B. Learning a new belief

Figure 9 depicts an activation time course of the belief sub-network later, during final portion of the painting sequence started in Figure 8. At time, t_0, purple paint has already been collected, reflected in a working memory peak in the *Coat Color* field. This has lead to the activation of a previously learned belief, $B1$ (for coat: purple, canvas: purple, result: yellow) that matched the coat color. The corresponding *commit* node, $C1$, has thus also been activated. The active belief does not guide the following painting sequence, because the inactivity of the recall boost node, u_{rcl}, prevents the active belief from forming peaks in the remaining role fields.

At time t_1, the agent begins dispensing paint onto the blue cube. This intention-in-action opens the gate to the *Canvas Color* field, which builds a peak at the attended color blue of the attended canvas cube. The mismatch between the blue canvas color representation and the canvas color purple predicted by the belief, $B1$, causes its deactivation, which in turn deactivates the *commit* node, $C1$, and lowers the activation level of the *belief activated* node, u_{bac}.

At time t_2, the cube changes its color from blue to yellow. The color change node, u_{cha}, is activated through transient detection, and boosts three nodes: *inhibit committed*, u_{ico}, *commit*, u_{com}, and *belief activated* u_{bac}. The observed result color forms a peak in the *Result Color* field, which excites beliefs that matching including $B1$. No belief matches all three color roles, however. Activation of u_{ico} excites the *commit* nodes of all previously learned beliefs, and lowers the activation levels of the associated *belief* nodes.

Once the slower u_{com} passes the detection threshold at time t_3, all beliefs receive a boost. This pushes the neural node of the previously uncommitted belief, $B4$, through the detection instability. u_{bac} is activated and generates a transient reward signal that up-modulates the learning rule for weights linked to $B4$. This leads to one-shot learning of reciprocal weights of the links to that belief node. The weights of both $B1$ and $B4$ after t_3 are depicted in Figure 10. $B4$ has now acquired the new color mixing rule, while $B1$ remains unchanged.

C. Activating a belief

Learned connections are utilized in activating a belief that matches a role cue. In Figure 11, a peak in the *Result Color* field at yellow has been induced from the agent's desire system. This leads to an increase in activation for all beliefs that match the yellow result color ($B1$ and $B4$) and a decrease for all non-matching beliefs. By chance, belief $B1$ passes the activation threshold first, leading to inhibition of all other belief nodes. In combination with the active recall node, u_{rcl}, belief node $B1$ activates its associated concept nodes. As a result, peaks in the neurally connected *Coat* and *Canvas Color* fields are built, which now guide the subsequent painting action.

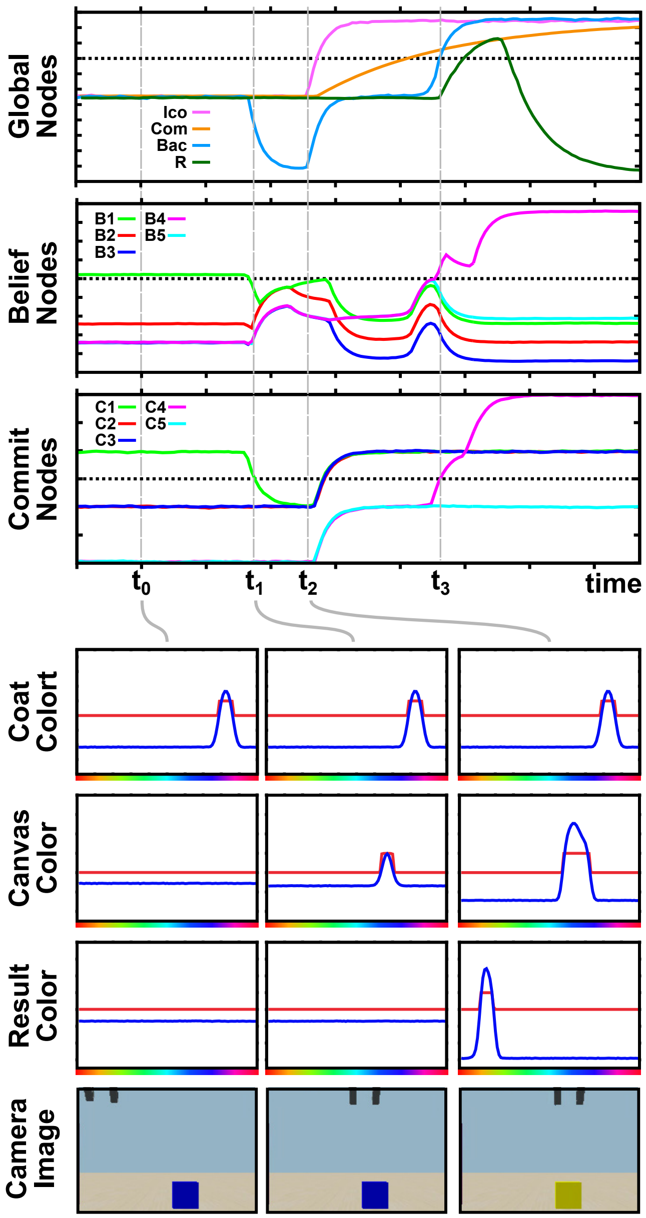

Fig. 9. Activation time course of selected nodes during the formation of a new belief. The first row displays activation of u_{ico}, u_{com}, and u_{bac}, while the second and third rows display the activation of five selected belief nodes, b_i, and their commit state nodes, c_i. The bottom half of the figure shows activation snapshots of the role fields at three different points in time.

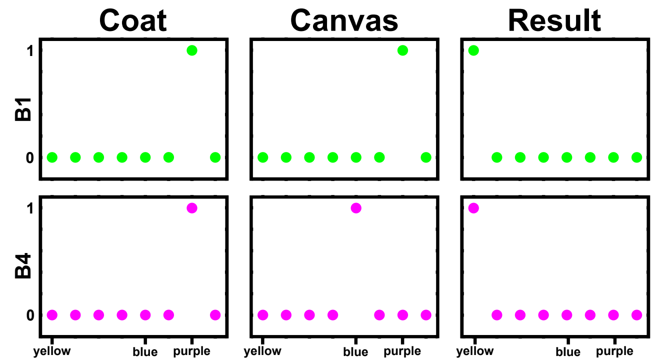

Fig. 10. Connection strengths, $l_{i,k}^{\mathrm{role}}$ to and from $B1$ and $B4$ are illustrated at a time after the learning episode of Figure 9.

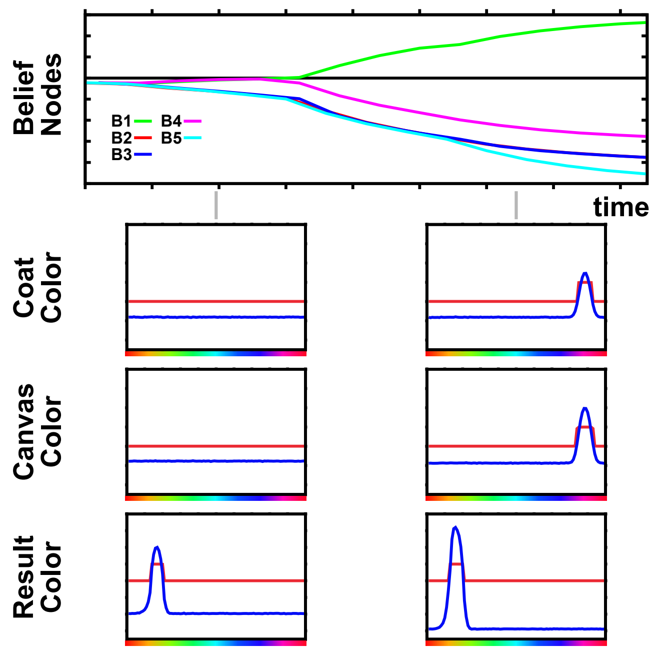

Fig. 11. Recall of a belief with a yellow result color. The learned connections are the same as shown in Figure 10

D. Deactivating a belief whose prediction is not confirmed

In the simulation of Figure 12, the belief, $B2$, that predicts a blue result color, has been previously activated and now guides a painting action. In the simulated world, the combination of coat and canvas colors used in $B2$ were now set to lead to the result color cyan instead.

At time t_0, the agent is engaging in the *dispense color* action while the predicted result, blue, is represented in the *Result Color* field. At time t_1, the canvas color changes cyan, instead. This observed color is projected into the *Result Color* field, where it erects a new peak at cyan, deleting the peak at blue. The color change in the result color field is detected by the CoD transient detector, which, in turn, inhibits all beliefs for a brief period, counteracting the excitatory boost from the commit node, u_{com}.

Once inhibition from the CoD decays at time t_2, the activation boost from u_{com} is sufficiently strong to activate the previously uncommitted belief, $B5$, which then becomes associated with the new color mixing rule by Hebbian learning.

V. Discussion

We have presented a network of neural dynamic fields that endows a robotic agent with the capability to form, activate, and reject beliefs in a simulated task environment. During belief learning, activated concept nodes become associated with a neural-dynamic belief node through a reward-modulated Hebbian learning rule. Activating beliefs is achieved by a neural match operation that is similar to the resonance principle of ART [26], combining the learned reciprocal connections with homogeneous global inhibition. Because the learned associations reside at the level of concepts, they form propositions and

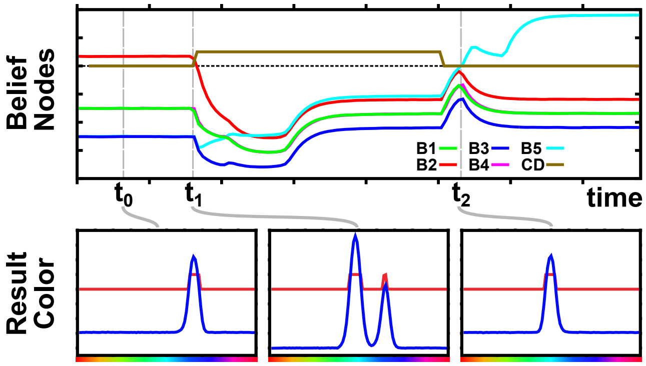

Fig. 12. Belief rejection: Belief, $B2$, predicts result color blue, but instead cyan was observed. The detected change in the result field leads to the deactivation of $B2$, and the eventual commitment of $B5$.

generalize the learned contingencies to the extent that the color concepts are invariant under small metric changes of sensory or environmental conditions. The rejection of a candidate belief occurs autonomously through the neural representation of a condition of dissatisfaction (CoD). That representation is triggered when a mismatch between perceived and predicted sensory representations of concepts is detected. The CoD inhibits the candidate belief and thus frees from inhibition the neural support for the learning of a new, alternative belief. The old belief is not forgotten, however.

The neural-dynamic belief system is embedded in a larger network of neural dynamic fields that controls a robotic agent. That network generates stable representations of intentional states of four elementary psychological modes (perception, memory, intention-in-action, and prior intention). Transitions between intentional states occur through instabilities induced by the neural CoS or CoD representations. It is from such transitions between different intentional states that the agent's behavior emerges. The stable representation of actions and perceptions support working memory that provides an interface to the belief network. Thus, learning is based on a set of sense data that are not typically available at the same time but are assembled into a working memory of the contingency. The processes that support using beliefs for action in the world and learning beliefs are both insensitive to variations in the duration of a behavioral or learning episode. This makes the autonomous generation of behavior and autonomous learning robust and enables, in principle, the scaling of the neural dynamics to real sensory-motor systems.

Prior work in neural modeling supports the plausibility of the postulated neural infrastructure that enables autonomous belief learning from single experiences. For instance, a model of cortical and basal-ganglionic processes for learning serially ordered behaviors has a similar prior structure [27]. Schrodt and Butz [28] have explored rule learning in a scenario similar to ours and argue for its neural plausibility. The neurally inspired architecture DAC (for Distributed Adaptive Control) [29] contains a contextual system in which the transition of action sequences from short-term into long-term memory is based on reward mediated learning similar to the mechanism proposed here.

We believe that the problem of autonomously learning beliefs, rules, or contingencies from experience is best framed as the problem of how the underlying architectures of neural processes are structured rather than as a problem of finding special learning rules. This position is similar to that taken in research on (now classical) cognitive architectures such as ACT-R [30] or SOAR [31]. Although broadly aligned with the brain's neural architecture, these frameworks are based on notions of information processing. Mental acts, called productions, are governed by rules. Our belief learning may be compared to the learning of new production rules. In cognitive architectures, such learning is not neurally grounded and does not, therefore, face the same issues of the neural control of learning [32]. This is true even for the cognitive architecture LIDA [33] that is more closely aligned with neural processing. LIDA has a strong emphasis on perceptual grounding and continuously updates the architecture's attentional system and long-term memory. That cycle has been modeled within the framework of Dynamic Field Theory to demonstrate the neural plausibility of its functionality.

The neural control of learning is achieved in the model at two levels. First, there is a categorical decision if the current experience warrants the acquisition of a new belief. This may happen through the Condition-of-Dissatisfaction (CoD), which detects a mismatch between the active belief and perception. A new belief then represents a propositional concept, in a critical step that provides for instant generalization of the proposition (within the capacity for generalization of the component concepts).

At the second level, the neural connectivity that instantiates a new belief is changed through reward modulated Hebbian learning. Reward activates learning only when an appropriate neural representation of a contingency has been detected. The reward is *intrinsic* in the sense that it is generated based by the system's own processing structure and characterizes successful matching of the representations that together form a belief (initial state, tool, outcome). This form of reward may be best compared to the notion of curiosity in other models of autonomous learning [34].

It is a different question if the beliefs acquired in this form should, in addition, be weighted in some way to express their degree of validity. This is what Bayesian notions of belief entail, and this has played a role in single-shot or few-shot learning in deep neural networks [3]. In the context of the present model, such weighting may be a way to steer the activation of a belief when trying to achieve a goal. More probable beliefs given larger weight would be activated more easily. Such possible extensions of the model would support our hypothesis, that beliefs are not forgotten or deleted when they have been rejected. Such beliefs would instead be assigned lower weight. The model provides slots for such a possible extension through its belief recognition mechanism, but it may be necessary to first scale up the scenario to give meaning to such an extension.

The toy scenario used in this paper is definitely minimalistic. An agent explores different color mixing combinations, acquires beliefs about color mixing rules, activates these beliefs to achieve desired resulting colors, and rejects beliefs if the resulting color does not match prediction. This scenario was chosen for its conceptual clarity. But the color concepts are trivial, their power of generalization is unimpressive, and the possible actions are simple. Quantitative evaluation of the learning process is not worthwhile at this stage.

There are obvious issues of scaling, such as increasing the perceptual and motor repertoire. Those are not particularly challenging. More subtle is the extension of the conceptual reach of beliefs. For instance, if the concepts from which propositions are built, were relations, more powerful beliefs could be formed, both in their level of abstraction and their capacity to generalize. For instance, relations like "larger than" or "near" would enable discovering regularities in the world. Action relations like "moving to" or "colliding with" would unlock contingencies that may be directly linked to possible actions of the agent to achieve goals. In other work, we have demonstrated that such relations can be perceptually grounded in neural dynamic architectures based on DFT [13], so we are confident that this form of scaling is possible.

A further, conceptually important form of scaling would be to address more complex desires, perhaps even abstract ones like the desire to learn new things about the world [34]. Different desires may conflict with each other and thus require rational decision making based on an agent's beliefs. Such higher-level processes of decision making have been theoretically analyzed by Bratman and colleagues [35], for instance, by incorporating conflicting plans and potential opportunities into an agent's deliberation process. It is quite conceivable, that such processes could be expressed in the neural dynamic framework presented here by endowing the level of desire with a richer processing structure, coupling it bi-directionally to the rest of the architecture. Beliefs that link outcomes to actions on targets could, of instance, then form a first step towards neurally based means-end reasoning.

References

[1] B. F. Skinner, ""Superstition" in the pigeon," *Journal of Experimental Psychology*, vol. 38, pp. 168–272, 1948.

[2] N. J. Nilsson, *Understanding Beliefs*. The MIT Press, 2014.

[3] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59–70, 2007.

[4] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra *et al.*, "Matching networks for one shot learning," in *Advances in neural information processing systems*, 2016, pp. 3630–3638.

[5] O. Lomp, C. Faubel, and G. Schöner, "A Neural-Dynamic Architecture for Concurrent Estimation of Object Pose and Identity," *Frontiers in Neurorobotics*, vol. 11, no. April, p. 23, 2017.

[6] J. B. Northrup, "Contingency detection in a complex world," *International Journal of Behavioral Development*, vol. 41, no. 6, pp. 723–734, 2017.

[7] J. R. Searle, "The intentionality of intention and action*," *Cognitive Science*, vol. 4, no. 1, pp. 47–70, 1980.

[8] J. Tekülve and G. Schöner, "Neural dynamic concepts for intentional systems," in *41th Annual Conference of the Cognitive Science Society (CogSci 2019)*, 2019.

[9] R. Kozma, "Intentional systems: Review of neurodynamics, modeling, and robotics implementation," *Physics of Life Reviews*, vol. 5, no. 1, pp. 1–21, 2008.

[10] G. Schöner, J. P. Spencer, and T. DFT Research Group, *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press, 2016.

[11] Y. Sandamirskaya and G. Schöner, "An embodied account of serial order: How instabilities drive sequence generation," *Neural Networks*, vol. 23, no. 10, pp. 1164–1179, 2010.

[12] K. Thórisson and H. Helgasson, "Cognitive Architectures and Autonomy: A Comparative Review," *Journal of Artificial General Intelligence*, vol. 3, no. 2, pp. 1–30, 2012.

[13] M. Richter, J. Lins, and G. Schöner, "A Neural Dynamic Model Generates Descriptions of Object-Oriented Actions," *Topics in Cognitive Science*, vol. 9, pp. 35–47, 2017.

[14] G. Schöner, "The Dynamics of Neural Populations Capture the Laws of the Mind," *Topics in Cognitive Science*, pp. 1–15, 2019.

[15] S. K. U. Zibner and C. Faubel, "Dynamic scene representations and autonomous robotics," in *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press, 2016, pp. 227–246.

[16] M. Richter, Y. Sandamirskaya, and G. Schöner, "A robotic architecture for action selection and behavioral organization inspired by human cognition," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 2457–2464.

[17] O. Lomp, M. Richter, S. K. U. Zibner, and G. Schöner, "Developing Dynamic Field Theory Architectures for Embodied Cognitive Systems with cedar," *Frontiers in Neurorobotics*, vol. 10, no. November, p. 14, 2016.

[18] O. Michel, "Cyberbotics ltd. webots™: professional mobile robot simulation," *International Journal of Advanced Robotic Systems*, vol. 1, no. 1, p. 5, 2004.

[19] R. Grieben, J. Tekülve, S. K. Zibner, J. Lins, S. Schneegans, and G. Schöner, "Scene memory and spatial inhibition in visual search: A neural dynamic process model and new experimental evidence," *Attention, Perception, & Psychophysics*, 2020.

[20] S. Schneegans, "Sensori-Motor and Cognitive Transformation," in *Dynamic Thinking: A Primer on Dynamic Field Theory*, G. Schöner, J. P. Spencer, and T. DFT Research Group, Eds. New York, NY, USA: Oxford University Press, 2016, pp. 169–196.

[21] W. Erlhagen and G. Schöner, "Dynamic field theory of movement preparation." *Psychological Review*, vol. 109, no. 3, pp. 545–572, 2002.

[22] Y. Sandamirskaya, "Dynamic neural fields as a step toward cognitive neuromorphic architectures," *Frontiers in Neuroscience*, vol. 7, no. 276, pp. 1–13, 2014.

[23] G. Schöner, J. Tekülve, and S. Zibner, "Reaching for objects: A neural process account in a developmental perspective," in *Reach-to-grasp Behavior: Brain, Behavior, and Modelling Across the Life Span*. Routledge, 2018, pp. 297–334.

[24] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer vision, graphics, and image processing*, vol. 37, no. 1, pp. 54–115, 1987.

[25] D. Dörner and C. D. Güss, "Psi: A computational architecture of cognition, motivation, and emotion," *Review of General Psychology*, vol. 17, no. 3, pp. 297–317, 2013.

[26] G. A. Carpenter and S. Grossberg, "Adaptive resonance theory," in *Encyclopedia of Machine Learning and Data Mining*, C. Sammut and G. Webb, Eds. Boston, MA: Springer, 2016.

[27] D. G. Beiser and J. C. Houk, "Model of cortical-basal ganglionic processing: encoding the serial order of sensory events," *Journal of Neurophysiology*, vol. 79, no. 6, pp. 3168–3188, 1998.

[28] F. Schrodt, J. Kneissler, S. Ehrenfeld, and M. V. Butz, "Mario becomes cognitive," *Topics in cognitive science*, vol. 9, no. 2, pp. 343–373, 2017.

[29] P. F. Verschure, "Distributed adaptive control: a theory of the mind, brain, body nexus," *Biologically Inspired Cognitive Architectures*, vol. 1, pp. 55–72, 2012.

[30] J. R. Anderson, "Act: A simple theory of complex cognition." *American Psychologist*, vol. 51, no. 4, p. 355, 1996.

[31] J. E. Laird, "Extending the soar cognitive architecture," *Frontiers in Artificial Intelligence and Applications*, vol. 171, p. 224, 2008.

[32] N. Taatgen and J. Anderson, "Constraints in cognitive architectures," *Cambridge handbook of computational psychology*, pp. 170–185, 2008.

[33] S. Franklin, T. Madl, S. D'mello, and J. Snaider, "Lida: A systems-level architecture for cognition, emotion, and learning," *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 1, pp. 19–41, 2013.

[34] P.-Y. Oudeyer, F. Kaplan, and V. V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE transactions on evolutionary computation*, vol. 11, no. 2, pp. 265–286, 2007.

[35] M. E. Bratman, D. J. Israel, and M. E. Pollack, "Plans and resource-bounded practical reasoning," *Computational intelligence*, vol. 4, no. 3, pp. 349–355, 1988.