A Neural Dynamic Model Autonomously Drives a Robot to Perform Structured Sequences of Action Intentions

Stephan Sehring (stephan.sehring@ini.rub.de)
Richard Koebe (richard.koebe@ini.rub.de)
Sophie Aerdker (sophie.aerdker@rub.de)
Gregor Schöner (gregor.schoner@ini.rub.de)

Institute for Neural Computation, Ruhr-Universität Bochum, 44780 Bochum, Germany

Abstract

We present a neural dynamic process model of an intentional agent that carries out compositionally structured action plans in a simulated robotic environment. The model is inspired by proposals for a shared neural and structural basis of language and action (Pastra & Aloimonos, 2012). Building on neural process accounts of intentionality we propose a neural representation of the conceptual structure of actions at a symbolic level. The conceptual structure binds actions to objects at which they are directed. In addition, it captures the compositional structure of action sequences in an action plan by representing sequential order between elementary actions. We show how such a neural system can steer motor behavior toward objects by forming neural attractor states that interface with lower-level motor representations, perceptual systems and scene working memory. Selection decisions in the conceptual structure enables the generation of action sequences that adheres to a memorized action plan.

Keywords: neural process model; dynamic field theory; action grammar; intention; action and language; autonomous robot

Introduction

Following instructions, or planning actions ourselves to reach goals often requires that we generate novel sequences of actions that we never before performed in exactly the same order or directed at precisely the same objects. The human faculty for intentional action comprises this remarkable ability to form a practically unlimited set of novel actions by flexibly recombining previously learned motor behaviors. Even rather global goals may thus be ultimately achieved by combining the limited set of movements available to the human body. This unlimited use of limited means has led researchers to compare the structural organization of human action to that of language. Similarities in their compositional structure and neurological findings have motivated the hypothesis that the syntax governing language and action might originate from a shared neural basis (Pastra & Aloimonos, 2012). In this perspective, actions are represented by a minimalist action grammar in which novel actions are composed from atomic symbols through sequences of merge operations. These generate hierarchical syntax trees that specify action types, action arguments (such as tools and objects) and the sequential order of actions. Whether a shared neural substrate is the origin of structural similarities between action and language is still under discussion (Zaccarella, Papitto, & Friederici, 2021). Such a representational framework does, however, offer a parsimonious account of the structured and flexible organization of human action. It would enable an agent to represent novel action plans that generalize beyond any specific instances it may have learned or stored earlier.

How could a neural system implement such a representational system and how could such an implementation drive intentional action? To address these questions we propose a neural process account of intentional action that enables an agent to autonomously direct action at objects in its environment (Tekülve & Schöner, 2019). Two key problems are addressed. First, we propose a neural representation of the conceptual structure of an action at a symbolic level which binds the action to the objects at which it is directed (see the top panel of Figure 1 for an illustration). This makes use of earlier work on neural binding through a shared "index" dimension (Sabinasz, Richter, & Schöner, 2023). We show how this neural implementation of a structured representation may guide the embodied realization of the intentional action directed at objects. Second, we show how a neural representation of the sequential order of elementary actions in a "dependency graph" may capture the compositional structure of actions described in syntax trees. We demonstrate how this representation may steer sequences of actions toward achieving goals.

As a proof of concept, we present a neural dynamic process model that controls a simulated robot arm in a table-top environment that carries out pick and place actions. The model represents action intentions as *action phrases*, that is, conceptual structures that bind action concepts to object concepts in different roles. The sequential dependencies between actions are also represented at this conceptual level. This conceptual structure of action plans takes the form of neural attractor states that form a working memory and can then guide the agent's motor behavior by interfacing with lower-level motor representations, perceptual systems and scene working memory. As an integrative neural process account, the model is continuously coupled to the robot's sensory-motor surface and entails interacting sub-networks that control the entire sensory-motor loop.

The model is based on Dynamic Field Theory (DFT; Schöner, Spencer, and DFT Research Group (2016)), a mathematical modelling framework which utilizes neural dynamic equations (Amari, 1977) to model the activation dynamics of neuron populations. Through their forward connectivity to the sensory or motor surfaces, these populations repre-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

sent feature dimensions or motor parameters, forming low-dimensional dynamic neural fields. These fields are formalized as dynamical systems that form stable *attractor states* shaped by recurrent connectivity within the field, locally excitatory and inhibitory over larger distances. The attractors emerging from this connectivity are localized peaks of activation that are stable against noise and may remain stable as the localized input that induced them is weakened. These peaks of activation are the basic units of representation in DFT. The interplay of internal coupling within a field and external inputs into the field determine a field's *dynamic regime*. Strong long range inhibition with moderate local excitation leads to a *selective regime* that allows only one stable peak to be active at a time. Strong local excitation and mid-range inhibition leads to a *self-stabilized regime* in which peaks remain stable even after input is removed. Changes in external input lead to instabilities in which peaks form (*detection instability*) or decay (*reverse detection instability*) depending on the dynamics of the field.

In DFT, cognitive processes emerge as neural fields that are coupled to sensory-motor surfaces go through various instabilities as their activation develops in continuous time. More complex cognitive processes are modeled by connecting neural fields into larger architectures. The ultimate meaning of neural fields in these architectures is determined by their tuning to features of the sensory-motor surface and by their synaptic coupling to other fields in the architecture.

Scenario and Model

For demonstration purposes we chose a scenario in which a robotic agent is tasked with performing a sequence of actions in the domain of simple pick and place behaviors. This domain is simple enough to still allow for easy interpretation of the meaning of action concepts and their grounding processes, while at the same time requiring the system to solve a number of challenges for object oriented behavior. Specifically, the agent requires a means to focus attention on and identify searched objects. It must build a scene working memory that possesses a mechanism for retrieving target objects from memory using conceptual object descriptions. And it must manage to organize these capabilities into sequences of pick and place behaviors toward objects.

The architecture was implemented in *Cedar* (Lomp, Richter, Zibner, & Schöner, 2016) and is coupled to a robotic simulation environment implemented in *Webots* (Michel, 2004). The agent is a robot arm equipped with an adjustable LIDAR camera. It is placed in a table-top environment that contains objects of different shapes and colors. Input to the neural field architecture consists of the LIDAR color and depth image, as well as the current camera orientation and the end-effector status. The robotic simulation receives commands from the architecture in the form of target end-effector position, gripper status and target camera orientation.

The following sections describe the function of the different sub-networks that comprise the model. Because of space constraints we do not give a complete mechanistic description. Instead we give a brief description of the function and role of each part with regard to those aspects that are most important for understanding how the model manages to ground compositional action plans. As an integrative model the different sub-networks are based on previous publications to which we refer if the reader seeks a more detailed mechanistic understanding. The only exception to this is the description of the conceptual structure where this model innovates on the representation of sequential information in the domain of action, which requires a somewhat more detailed discussion on its dynamics.

Conceptual Structure The conceptual structure comprises the neural substrate that represents the agents' intentions and controls the execution of an action plan. As discussed above, we account for their compositional structure by representing action intentions as action phrases, which we accomplish by utilizing the index binding mechanism in discrete neural fields that was proposed by Sabinasz et al. (2023) in the domain of nested noun phrases. The index binding mechanism assumes that each concept node shares an index dimension that can be used to express concept combinations (bindings). We use this mechanism to bind object concepts to action concepts in different roles such that they match the semantic argument structure of an action. We represent action sequences by representing sequential dependencies between action phrases. These dependencies enforce a partial topological ordering of planned actions, which allows the expression of precedence relations between actions, while still allowing flexibility in the sequential order of non-dependent actions.

Fig. 1 depicts an exemplary action plan visualized as a directed graph of syntax trees (A) and the corresponding representation as an activation pattern in the conceptual structure (B).

The *object concept* field represents objects or locations through binding of one or multiple object/relation concepts to a unique object index. Analogously, action concepts are bound to an action index, which serves as the binding agent for an action phrase. Object representations are bound to an action index by their semantic role of either target (*AT*) or reference (*AR*). For example, the concepts bound to A1 in Fig. 1B correspond to the phrase *[Transport] [Blue Donut] [on Top of Green Plate]* in Fig. 1A. Sequential dependencies are represented in the *dependency relation* field. This field implements a directive binding, in which each row represents the successor action of the corresponding action index. The 0-row indicates the start and column-0 the end of an action plan. In the example of Fig. 1 the dependency relations encode that action A1 or A3 may be performed initially, while A2 has to immediately follow A1 and A4 follows A3.

The working memory representation of an action plan is given as a sub-threshold activation pattern in the conceptual structure. Neurons above threshold pass excitatory activation along their shared index dimension, while projecting strong inhibitory activation to neural nodes coding for other indices.

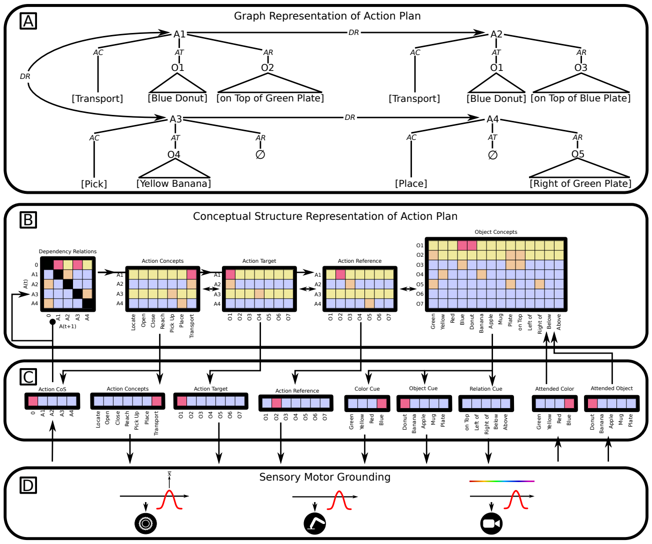

Figure 1: Exemplary action plan represented as a directed graph of syntax trees (A) and the corresponding representation in the conceptual structure (B). The contents of an active action phrase are projected to the rest of the architecture (D) through the interfacing concept production nodes (C).

This puts the conceptual structure into a *selective regime* in which action phrases form *attractor states* that inhibit competing action phrases from going through the *detection instability*. The dependency relation field passes activation along the action index dimensions, pushing actions that should be performed next close to threshold, giving them a selective advantage. The contents of an active action phrase are projected to the rest of the architecture (Fig. 1D) through the interfacing *concept production* nodes (Fig. 1C).

Attractor states are destabilized, when other parts of the model signal successful execution of an action or the perception of a searched object. The success signal of a performed action inhibits its corresponding column index and boosts its row index in the dependency relation field (*Action CoS* nodes in Fig. 1C). After each action, the dependency graph steps through its encoded dependencies, thus only boosting action phrases which are currently available according to the action plan. A match signal coming from the perception subnetwork destabilizes the corresponding object representation, which facilitates the selection and subsequent search for a new object.

Perception and Attention The search for an object aligning with a conceptual description is crucial to successfully guide interactions with the environment. To facilitate this procedure, we integrated a neural process model of scene representation and categorical visual search within natural scenes (Grieben & Schöner, 2022). This model was expanded to fulfill the demands of a robotic setting (Fig. 2 *Perception & Attention*). In particular, we incorporated the capability to generate saccades in a three-dimensional environment. Each saccade rotates the camera to bring an object into the attentional foreground. Saccade selections are based on a priority map, that receives bottom-up input from a saliency map and is modulated by top-down guidance from memory and conceptual feature cues. The model projects the attended position, the detected color, shape and height feature values as output to other sub-networks.

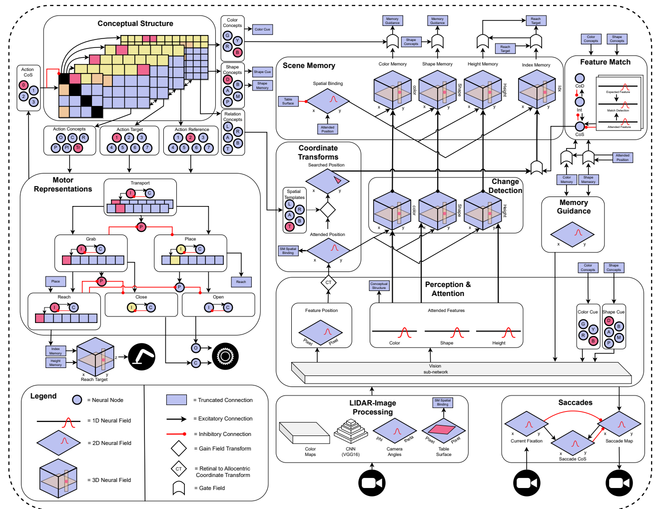

Figure 2: Overview of the neural field model.

Scene Memory Successful interaction with the environment requires a scene representation that maintains scene information in the absence of perceptual input, and can be updated online when perceptual input is available. Here, scene memory is based on space-feature maps that are bound through space (Grieben & Schöner, 2022) and coupled to a change detection mechanism based on Johnson, Spencer, Luck, and Schöner (2009). The *change detection* fields commit new space-feature information to memory, if a feature at a currently attended position does not match the current scene representation. Goal-oriented behavior additionally requires that conceptual object descriptions are anchored in the environment. This is achieved by the *index memory*, which binds a conceptual object description to the location of a matching object in the environment. A searched position is committed to *index memory*, when the *feature match* sub-network (Grieben & Schöner, 2022) detects a match between the conceptual guidance cues and the features in *scene memory* at the attended location. (Fig. 2 *Scene Memory*)

To ground relational concepts, the architecture uses coordinate transforms of *spatial templates* that can be seen as relational operators in spatial language (Richter, Lins, & Schöner, 2021). The conceptual description of a location activates a corresponding *spatial template*, which is used to commit a position in a certain relation to the attended object to memory. (Fig. 2 *Coordinate Transform*)

Motor Representations The motor representation sub-network contains a set of primitive, hierarchically organized motor representations, that enable an agent to act based on conceptual action intentions. For this, they translate an action phrase into a sequence of movement primitives. Our implementation of motor representations is based on Richter, Sandamirskaya, and Schöner (2012). Each action concept connects to an elementary behavior (EB). The individual EBs are organized into hierarchical sequences. The *Transport* sequence consists of the EBs *Grab* followed by *Place*, which in turn first activate *Reach* followed by *Open/Close*. *Reach*, *Open* and *Close* are behavioral primitives that directly project onto the sensory-motor surface. Target objects are provided as arguments by the *conceptual structure*. Movement primitives recall spatial locations of objects from *scene memory* to guide arm movements. A Condition of Satisfaction (*CoS*) signal is send to the conceptual structure once the active intention is completed. Together this structure manages to decompose higher-level actions into simple primitives, while conserving specified movement targets through argument passing between. (Fig. 2 *Motor Representations*)

Summary Given the above explanation of the individual sub-components, the overall flow of the model can be summarized as follows. Starting from the top-left of Fig. 2, selection decisions in the *conceptual structure* lead to the activation of an action phrase representing the agents current intention. Active object concepts pass activation to the *perception* sub-network and *scene memory* where they are used to provide feature and memory cues to the *perception* sub-network (lower-right). The perception sub-network generates saccades based on object salience and top-down cues which enables the model to search for objects. Once a searched object is found its conceptual representation is anchored to its location by storing its position in *index memory* (top-right). At the same time, the active action concept activates its corresponding *motor representation* EB which in turn initiates the associated sequence of primitives (lower-left). Movement starts when movement primitives manage to recall the target object position from *index memory*. The completion of an action destabilizes its attractor state in the conceptual structure, leading to a new selection decision in accordance with the memorized plan. In the next section we demonstrate how this enables the model ground partially ordered sequences of structured actions.

Results

We illustrate how the agent robot performs the action plan depicted in Fig. 1. The plan consists of two separate branches that may be flexibly executed in either order. Within each branch, the temporal order of the actions must be preserved. Objects in the environment were chosen to demonstrate how the model can successfully locate target objects among multiple distractors, and how a located object can be utilized in different action phrases while preserving object identity. Figures 3 and 4 plot activation snapshots of selected fields and nodes of the conceptual structure and the grounding architecture over the time course of the performance. We focus our description on how intentions are selected and maintained at the conceptual level, how they guide attention and motor behavior and how the conceptual structure generates ordered sequences of intentions.

Selecting first action The performance starts with the selection of an action phrase. The given action plan allows the freedom to choose between *A1* and *A3*. At time t_1 the *dependency relation (DR)* field projects sub-threshold input to the *action concept (AC)*, *action target (AT)* and *action reference (AR)* fields bringing the nodes bound to *A1* and *A3* close to threshold. This facilitates a selection decision, in which *A1* and *A3* compete for reaching supra-threshold activation. At time t_1, action *A1* is already selected due to noise. The bidirectional coupling between nodes in the conceptual structure stabilizes this *selection decision* by suppressing nodes of other action indices. The active action concept *Transport* then activates the corresponding motor representation that organizes the sequential activation of a *pick* and *place* action. Competition in the *object concept (OC)* field has first led to the selection of the target object *blue donut*. The selected object concepts guide visual search by providing guidance cues to the visual search sub-network. (t_1 in Fig. 3)

Performance of first action In Fig. 4 at time t_1 we can see that the top-down guidance cue has already led a saccade to attentionally select the *blue donut*. The *attended shape (AS)* and *attended color (AC)* fields depict classification decisions of the visual search sub-network. At t_1 they display the correct classification decisions of *blue* and *donut*. After its position was stored in *scene memory*, the *Reach* motor representation recalls its position from *index memory*, which guides the robot arm to reach for the *blue donut*. (t_2 in Fig. 4) By t_3, the detection of the searched *blue donut* leads to the destabilization of *O1 (blue donut)* in the OC field. This is followed by the subsequent selection of the action reference *O2 (green plate)* (*OC* field at t_3 in Fig. 3). By the same mechanism as explained above, the guided visual search enables the model to find the *green plate* and subsequently place the *blue donut* on top of it, which successfully completes the first action phrase *A1* (t_3 in Fig. 4).

Sequential selection of actions After completion, action phrase *A1* is destabilized in the conceptual structure by inhibition of the *A1* column and boosting the *A1* row in the *DR* field. Through this mechanism, the conceptual structure steps through the encoded action plan, as now action phrase *A2* becomes a stable attractor state (t_4 in Fig. 3). The performance of *A2* is analogous to *A1*. (t_4, t_5 and t_6 in Fig. 3 and 4)

Switching between branches The completion of action phrase *A2* signals the end of the first branch of the action plan. This can be seen at time t_7 in the *DR* field of Fig. 3. The boost along the *A2* row coming from the CoS-signal activates the 0-index as its successor, which in turn signals the return to the beginning of the dependency graph by boosting

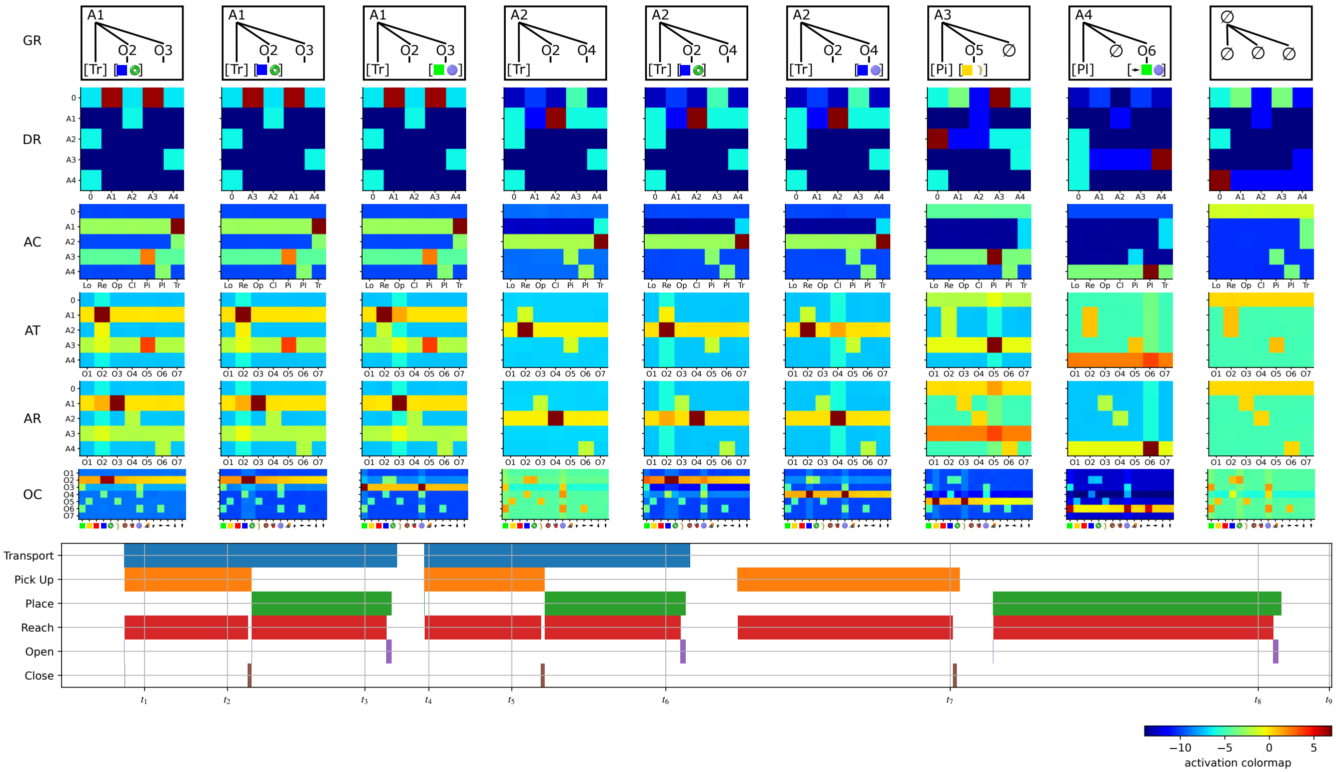

Figure 3: Snapshots of the activation of selected fields of the conceptual structure and motor representations during performance. Each column plots the level of activation at a shared time point. *GR = Graph Representation, DR = Dependency Relation, AC = Action Concept, AT = Action Target, AR = Action Reference, OC = Object Concept*

the 0-row. Because *A1* is still inhibited from the previous inhibition, *A3* wins the next selection decision.

Grounding relation concepts The performance of action phrases *A3* and *A4* is analogous to that of *A1* and *A2*. Action *A4* includes the relational concept *right of green plate (O6)* as part of its action reference. The bound relation concept activates a corresponding relation template that is part of the *coordinate transform* sub-network (Fig. 2). The coordinate transform commits a location to the right of the *green plate* to *index memory*, from where it can be recalled. This enables the robot to place the *banana (O5) to the right of the green plate*. (t_7 and t_8 in Fig. 3 and 4)

End of plan At time t_9, the conceptual structure returned to the beginning of the dependency graph. Because *A1* and *A3* are still inhibited from column inhibition, no other action is selected. The plan is completed and the scene is re-arranged as intended. (t_9 in Fig. 3)

Discussion

Previous work has established how an "intentional agent" may achieve goals acting in a (simulated) environment driven entirely by neural dynamic processes formalized within DFT (Tekülve & Schöner, 2019). A first key innovation in this paper is to endow such an agent with a structured conceptual representation of planned actions and the objects at which these actions are directed. This is neurally realized as sustained activation of neural nodes that are coupled through shared index dimensions (Sabinasz et al., 2023). This makes it possible to flexibly bind the concepts for actions to concepts for objects in different roles in the manner of "action phrases" whose structure can be described by syntax trees. The second key innovation is an explicit representation of the sequential dependencies between actions, which makes it possible to compose complex actions from more elementary ones. The sensory-motor grounding of this conceptual structure consists of the agent "acting out" the represented actions based on the connectivity of the neural nodes to lower-level motor representations, perceptual systems, and scene working memory.

The representation of intentions as action phrases is partially inspired by proposals that action and language share a structural basis (Pastra & Aloimonos, 2012), so that a conceptual level representation of action intentions would have the features of compositionality and productivity postulated for human language. The notion of an action grammar is compatible with proposals of a hierarchical organization of the different sub-systems involved in action generation (Grafton & De C. Hamilton, 2007). Such hierarchical models typically entail high-level abstract action representations that are iteratively translated into sequences of low-level behaviors, ultimately generating executable movement patterns. Oth-

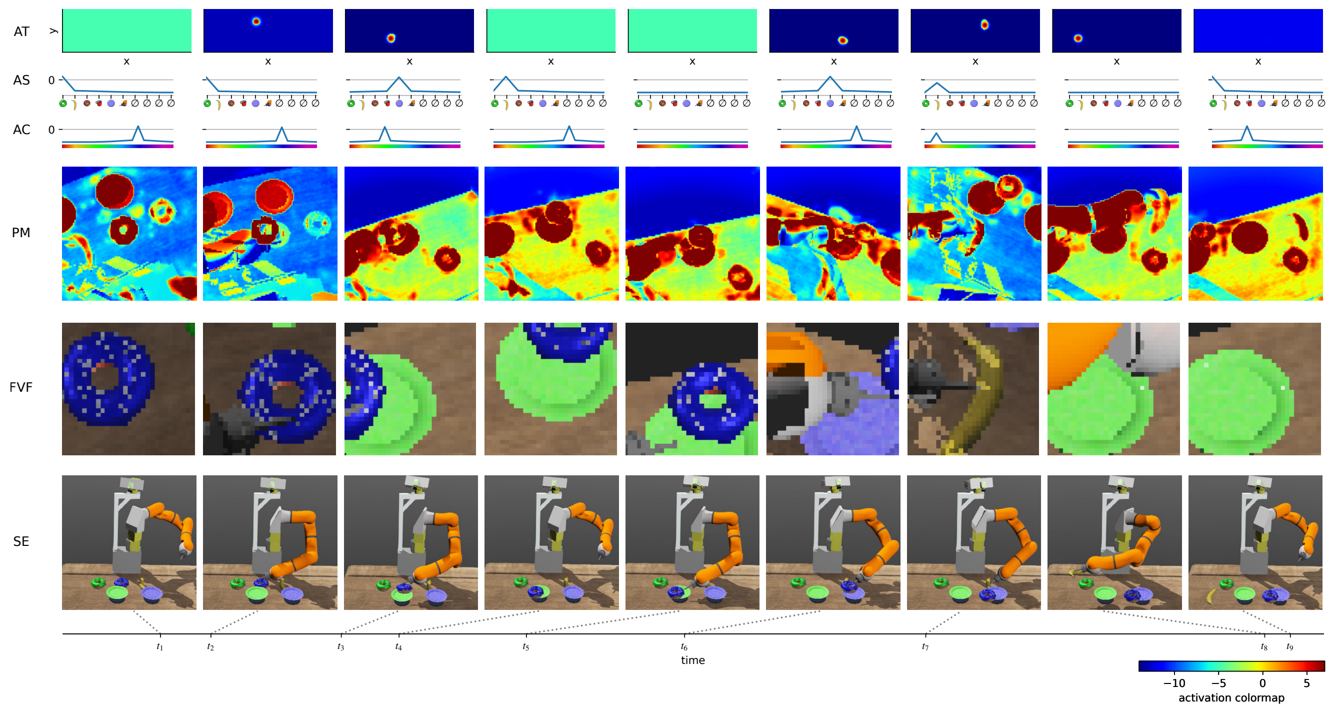

Figure 4: Snapshots of the activation of selected fields of the grounding architecture during performance. The selected time points match those of Fig. 3. *AT = Arm Target, AS = Attended Shape, AC = Attended Color, PM = Priority Map, FVF = Attentional Functional Visual Field, SE = Simulated Environment*

ers have proposed similarly that action intentions may be expressed in a propositional format that interfaces with lower-level sensory-motor systems to provide appropriate parameters for non-propositional motor representations (Mylopoulos & Pacherie, 2017; Shepherd, 2019).

Action plans represented in the conceptual structure could be thought to originate from verbal instruction or from a processes of deliberate planning that may involve knowledge (semantic memory). The neural dynamic representation of action intentions may specify the sequential order of actions by explicitly representing the precedence relations between individual actions. This successor representation expresses a partial ordering of actions, with some requisite sequential ordering while allowing for flexible sequential ordering among certain branches. This representation is consistent with partial ordering of human action in which some actions may be opportunistically activated as a situation evolves. Coopmans, Kaushik, and Martin (2023) argue that the structure of actions that subserve a specific goal may best be characterized as partially ordered sequences. In this view, a minimal encoding of an action plan may be a dependency graph that captures all possible orderings of actions leading to the same goal. How exactly opportunistic reordering of actions would occur is not entirely clear, at the moment.

Human behavior may be construed as outcome driven (Papies & Barsalou, 2015). The model presented here does not address how actions are directed toward overall goals or desired outcomes. It may be a viable and interesting research question to explore how structured representations of the type proposed here could be used to conceptually represent desired outcomes and the sequences of action phrases needed to achieve these outcomes. This may require a link to the sensory-motor grounding of actions, perhaps as mental simulations rather than physically acts. Such action planning must face challenges with regard to knowledge representation and reasoning (Lake, Ullman, Tenenbaum, & Gershman, 2017). We are also interested in how conceptual representations of novel actions may arise through binding to previously learned motor representations. Neural accounts of action selection mechanisms may provide useful constraints for further development on this matter (Stewart, Choo, & Eliasmith, 2010). More generally, the question of how to design open-ended semantically grounded language and reasoning systems has been an ongoing domain of research in the field of cognitive robotics (Steels & Hild, 2012). We approach these questions from a purely neural perspective with an emphasis on psychological plausibility. Ultimately, we hope this model serves as a starting point for such further research that may bring us closer to the realization of a fully autonomous intentional agent that can reason about its goals, plan, and act in a natural environment.

References

Amari, S.-i. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, *27*(2), 77–87. doi: 10.1007/BF00337259

Coopmans, C. W., Kaushik, K., & Martin, A. E. (2023). Hierarchical structure in language and action: A formal comparison. *Psychological Review*, *130*(4), 935–952. doi: 10.1037/rev0000429

Grafton, S. T., & De C. Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, *26*(4), 590–616. doi: 10.1016/j.humov.2007.05.009

Grieben, R., & Schöner, G. (2022). Bridging DFT and DNNs: A neural dynamic process model of scene representation, guided visual search and scene grammar in natural scenes. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual conference of the cognitive science society.* Cognitive Science Society.

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*(5), 568–577. doi: 10.1111/j.1467-9280.2009.02329.x

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. (arXiv: 1604.00289) doi: 10.1017/S0140525X16001837

Lomp, O., Richter, M., Zibner, S. K. U., & Schöner, G. (2016). Developing Dynamic Field Theory Architectures for Embodied Cognitive Systems with cedar. *Frontiers in Neurorobotics*, *10*. doi: 10.3389/fnbot.2016.00014

Michel, O. (2004). Cyberbotics ltd. Webots™: Professional mobile robot simulation. *International Journal of Advanced Robotic Systems*, *1*(1), 5. doi: 10.5772/5618

Mylopoulos, M., & Pacherie, E. (2017). Intentions and Motor Representations: The Interface Challenge. *Review of Philosophy and Psychology*, *8*(2), 317–336. doi: 10.1007/s13164-016-0311-6

Papies, E., & Barsalou, L. (2015). Grounding desire and motivated behavior: A theoretical framework and review of empirical evidence. In W. Hofmann & L. F. Nordgren (Eds.), *The psychology of desire* (p. 36-60). Guilford Press.

Pastra, K., & Aloimonos, Y. (2012). The minimalist grammar of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 103–117. doi: 10.1098/rstb.2011.0123

Richter, M., Lins, J., & Schöner, G. (2021). A neural dynamic model of the perceptual grounding of spatial and movement relations. *Cognitive Science*, *45*(10), e13045. doi: 10.1111/cogs.13045

Richter, M., Sandamirskaya, Y., & Schöner, G. (2012). A robotic architecture for action selection and behavioral organization inspired by human cognition. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2457–2464). doi: 10.1109/IROS.2012.6386153

Sabinasz, D., Richter, M., & Schöner, G. (2023). Neural dynamic foundations of a theory of higher cognition: The case of grounding nested phrases. *Cognitive Neurodynamics*. doi: 10.1007/s11571-023-10007-7

Schöner, G., Spencer, J., & DFT Research Group, I. f. N. (2016). *Dynamik thinking: A primer on dynamic field theory*. Oxford University Press.

Shepherd, J. (2019). Skilled Action and the Double Life of Intention. *Philosophy and Phenomenological Research*, *98*(2), 286–305. doi: 10.1111/phpr.12433

Steels, L., & Hild, M. (Eds.). (2012). *Language Grounding in Robots*. Boston, MA: Springer US. doi: 10.1007/978-1-4614-3064-3

Stewart, T. C., Choo, X., & Eliasmith, C. (2010). Dynamic behaviour of a spiking model of action selection in the basal ganglia. In *Proceedings of the 10th international conference on cognitive modeling* (pp. 235–40). Citeseer.

Tekülve, J., & Schöner, G. (2019). Neural dynamic concepts for intentional systems. In *Proceedings of the 41st annual conference of the cognitive science society.* Cognitive Science Society.

Zaccarella, E., Papitto, G., & Friederici, A. D. (2021). Language and action in Broca's area: Computational differentiation and cortical segregation. *Brain and Cognition*, *147*, 105651. doi: 10.1016/j.bandc.2020.105651