Learning the Perceptual Conditions of Satisfaction of Elementary Behaviors

Matthew Luciw*, Konstantin Lahkman†, Sohrob Kazerounian*,
Mathis Richter‡, and Yulia Sandamirskaya‡
*The Swiss AI Lab IDSIA, USI & SUPSI, Galleria 2, Manno CH-6928, Switzerland
†NBIC-Centre, National Research Center, Kurchatov Institute, Moscow, Russia
‡Ruhr-Universität Bochum, Institut für Neuroinformatik, Bochum, Germany

***Abstract*—A core requirement for autonomous robotic agents is that they be able to initiate actions to achieve a particular goal and to recognize the resulting conditions once that goal has been achieved. Moreover, if the agent is to operate autonomously in complex and changing environments, the mappings between intended actions and their resulting conditions must be learned, rather than pre-programmed. In the present work, we introduce a method in which such mappings can be learned within the framework of Dynamic Field Theory. We not only show how the learning process can be implemented using dynamic neural fields, but show how the adaptive architecture can operate on real-world inputs while controlling the outgoing motor commands. The proposed method extends a recently proposed neural-dynamic framework for behavioral organization in cognitive robotics.**

I. Introduction

Complex behaviors performed by cognitive robotic agents may be segregated in a number of elementary behaviors (EBs), performed simultaneously and/or sequentially. Each EB links the agent's perceptual system to its motor system in a behavior-specific manner [3]. Complex actions require the coordination between a number of simpler EBs, such that each EB is activated in the appropriate order, persists as long as necessary in order to achieve its behavioral subgoal, and is ultimately deactivated when this goal is achieved. We have recently introduced a neural-dynamic framework, in which such an organization of the robot's behaviors, composed of structurally simpler EBs, was realized [8]. In this framework, each EB consists of two coupled dynamical structures, which determine an elementary behavior's *intention* and its *condition of satisfaction* (CoS). The intention comprises behavioral parameters, which eventually bring about either an overt or an internal action, whereas the condition of satisfaction is a perceptual indicator, by which the agent recognizes the completion of the EB. We have previously demonstrated how sequences of goal-directed actions may be generated in this framework by linking a neural-dynamic architecture for behavioral organization to sensors and motors of a Nao robot [9, 8]. We have also demonstrated how sequences of EBs may be learned from delayed rewards by combining the neural-dynamic architecture with reinforcement learning and making use of eligibility traces [7]. However, the structure of an EB, i.e. the coupling structure between the intention and the CoS of a behavior in these prior models was hand-designed as part of the architecture. In the present work, we demonstrate how this coupling may emerge autonomously in the framework of Dynamic Field Theory, using an associative learning rule, along with a set of built-in internal drives, or needs, and a scalar rewarding input when the need is satisfied.

II. Methodological background

A. Dynamic Field Theory

Dynamic Field Theory (DFT; [12]) is a mathematical framework, which has been used to model neural-dynamic processes that underly behavior. DFT architectures are well suited for robotic control systems due to their ability to form and stabilize robust categorical outputs from noisy, dynamical, and continuous real-world input. DFT has been

applied in robotics at different levels, from low-level navigation dynamics with target acquisition based on vision [2] to object representation, dynamic scene memory, and spatial language [11], as well as sequence generation [10].

The basic computational unit in DFT is a dynamic neural field (DNF). DNFs represent activation distributions of neural populations, as opposed to classical neural network architectures whose computational units are at the level of individual neurons. A DNF's activation is defined over continuous dimensions (e.g., color or space), which characterize sensorimotor systems and task space of the agent. This activation develops in continuous time based on a dynamical equation analyzed by Amari [1]. As a result of non-linearities in the DNF's dynamics and lateral interactions within neural fields, stable localized *peaks* of activation emerge from distributed, noisy, and transient input. These activation peaks represent perceptual objects or motor goals in the DFT framework. Multiple coupled DNFs spanning different perceptual and motor modalities can be composed into complex DFT architectures.

B. Elementary behaviors (EBs)

A generic structure of EBs (Fig. 1) has been proposed recently in the framework of Dynamic Field Theory [9]. Each EB consists of an *intention* and a *condition of satisfaction* (CoS) DNFs. An active intention DNF either modifies the perceptual system of the agent or impacts on the motor dynamics of the agent directly. The CoS DNF detects and stabilizes a perceptual signal that the EB has successfully achieved its intention. To enable this, two inputs converge on the CoS DNF: one from the intention DNF and the second one from the perceptual system. If the two inputs match in the dimension of the CoS DNF, an activity peak emerges in this field, inhibiting the intention DNF of the EB.

The intention and CoS DNFs are associated with intention and CoS dynamic nodes respectively, which facilitate the sequential organization of EBs. While the DNFs are relevant for intra-behavior dynamics, such as in selecting the appropriate perceptual inputs for the behavior, the dynamic nodes play a role on the level of inter-behavior dynamics (i.e., switching between behaviors). In our previous work, we have shown how EBs may be chained according to rules of behavioral organization [8, 9], serial order [10, 5, 4], or the value-function of a goal-directed representation [7].

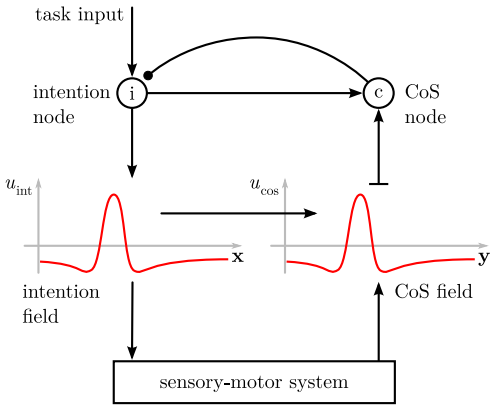

Fig. 1: Schematic representation of a generic elementary behavior.

C. Condition of satisfaction

The condition of satisfaction DNF generates a signal, which denotes that the intention of its EB is successfully achieved. For instance, the CoS DNF for the behavior 'go to the red object' could detect when a large red object is present in the visual field, or when the robot is below a distance threshold to the target object. In our neural-dynamic framework, the CoS is specified by the choice of the dimension(s) of the CoS field and by the synaptic connection weights from the intention field to the CoS field. While the dimensions of the field reflect which sensory dimensions the robot is sensitive to, the weights shape the pre-activation in the CoS field and make specific regions of the field sensitive to perceptual input.

In our previous work, the perceptual input, which activated the CoS of all behaviors, was 'hardcoded' into the architecture. We designed both the dimensions of the CoS field and the synaptic weights converging on the field to produce a CoS signal (i.e.,

a peak in the CoS field) only in environmental situations that we, the designers, felt appropriate. With architectures built in this way we have successfully shown autonomous behavior of robotic agents (see, e.g., [8]). Here, we address the question of how such architectures could come about by autonomous learning.

III. Learning the Condition of Satisfaction

Here, we present a straightforward but effective mechanism for learning the CoS, using a variant of associative learning, gated by reward. To enable CoS learning, the basic structure of an EB is augmented by two components. First, the connection weights between the intention and the CoS DNF are made plastic, with a learning rule that combines associative and reinforcement learning. In particular, when a rewarding signal is received, learning between the intention and CoS fields is enabled, i.e. the connection weights between these DNFs are modified according to a simple Hebbian-like learning rule (section III-A). The rewarding signal, in its turn, comes from the second new element – a number of internal drives, which motivate the agent's behavior. These drives can most closely be compared to the prototypical drives suggested by Woodworth, e.g. hunger and thirst [13]. Drives such as these serve as internal forces that initiate behaviors and agents are rewarded when the drives are satisfied [6].

Fig. 2 shows an exemplar mapping between one-dimensional intention and CoS DNFs. In the case of one-dimensional DNFs, the coupling between them is a 2D mapping. Here, only two locations of the intention DNF are associated with respective two locations in the CoS DNF. More complex and high-dimensional mappings may be realized in DFT in an analogous way [11]. Such associations are dynamically encoded as memory traces, or localized preshapes, in the dynamics of the mapping.

A. Reward-gated associative learning

The learning process in our architecture leads to formation of the memory traces in the dynamical mapping between the intention and the CoS DNFs. Let $u(x,t)$ be the intention DNF, which evolves in time according to Eq. 1 [1] and $v(y,t)$ be the CoS DNF (Eq. 2). x is the motor parameter, which spans the dimension, over which the intention DNF is defined, y is the respective perceptual parameter of the CoS DNF:

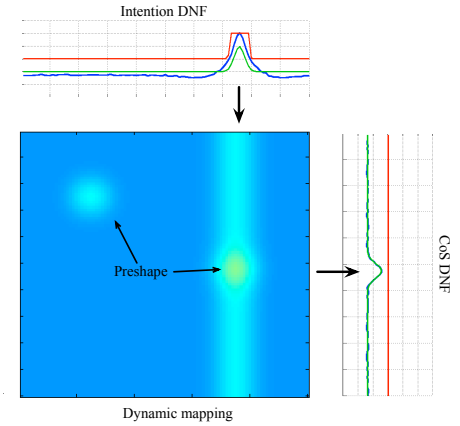

Fig. 2: A simple mapping between one-dimensional intention and CoS dynamic neural fields. Here, the two pre-shape bumps indicate the two learned mappings from intention to CoS.

$$\begin{aligned}\tau\dot{u}(x,t) = &- u(x,t) + h_u + \\ &+ \int f(u(x',t))\omega(x'-x)dx' - \\ &- c\int f(v(y,t))dy + I_{\mathrm{t}}(x,t), \end{aligned} \quad (1)$$

$$\begin{aligned}\tau\dot{v}(y,t) = &- v(y,t) + h_v + \\ &+ \int f(v(y',t))\omega(y'-y)dy' - \\ &+ \int W(x,y,t)f(u(x,t))dx + \\ &+ I_{\mathrm{sens}}(y,t). \end{aligned} \quad (2)$$

Here, h_u, h_v are resting levels of the DNF dynamics, $f(\cdot)$ is the sigmoidal non-linearity shaping the output of the DNFs, $\omega(\cdot)$ is the lateral interaction kernels, c is the constant regulating strength of the homogeneous inhibition from the CoS DNF to the intention DNF, I_{t} is the task (motivational) input, I_{sens} is the sensory input, $W(x,y,t)$ is the two-dimensional weights function, which maps output of the intention DNF onto CoS DNF (see Fig. 2).

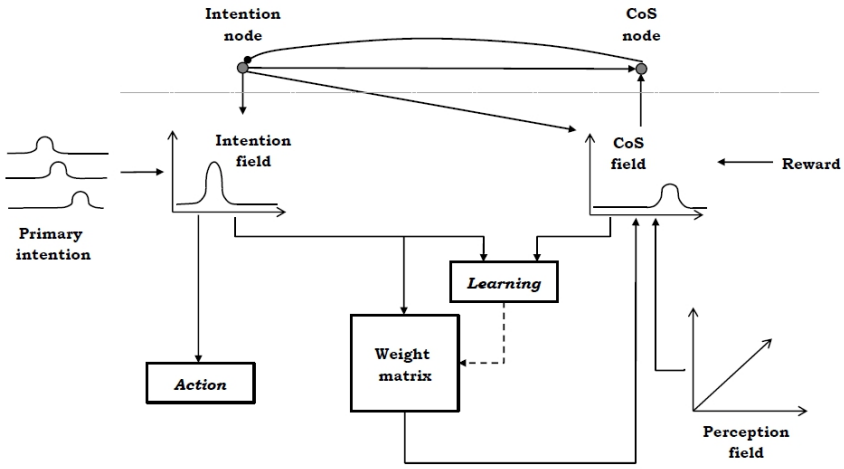

Fig. 3: CoS learning architecture. See main text for details.

The mapping $W(x, y, t)$ is updated according to a simple reward-driven learning rule, Eq. 3:

$$\begin{aligned} \tau_l \dot{W}(x,y,t) &= \lambda R(t)\Big(-W(x,y,t) + \\ &+ f\big(u(x,t)\big) \times f\big(v(y,t)\big)\Big) \cdot \\ &\cdot f\big(u(x,t)\big) \times f\big(v(y,t)\big), \end{aligned} \tag{3}$$

where $f\big(u(x,t)\big) \times f\big(v(y,t)\big) = f\big(f_y(u(x,t)) + f_x(v(y,t))\big)$ is a sigmoided sum of the output of the intention DNF, extended along y (the dimension of the CoS DNF) and the output of the CoS DNF, extended along x (the dimension of the intention DNF). The weights $W(x, y, t)$ are updated when a reward signal $R(t)$ is perceived and they are updated at locations, where the overlap between the two projections (f_x and f_y) is positive. When updated, the weights converge towards the outputs of the DNFs.

Fig. 3 illustrates a sketch of the learning architecture. In the following we present an implementation of this autonomous neural-dynamic architecture in a simple robotic scenario in a physical environment.

IV. Implementation and Result

Here, we present an implementation of the learning mechanism in a simple scenario in a physical environment. It is important to note that the mechanism we described earlier is very general and works with different types of EBs and perceptual inputs other than the ones we describe here. The robotic system we used in our experiments consisted of an ePuck, equipped with a color camera (shown in Fig. 4). The robot was put in an environment, which contained several object of different colors. The robot needed to search its environment for objects of certain colors in order to satisfy its internal drives. The drives, provided to the robot, were loosely called 'hunger' and 'thirst'. The drives became active at different times: With the hunger drive active, reward occurred when a red object was in the image; when thirst was active, reward was achieved with a yellow object.

Fig. 4: CoS learner's environment.

The robot could move around the arena, guided by simple search dynamics. The camera images provide input to a two-dimensional *perceptual field* [10], with one dimension as color hue (separated into 15 bins) and the other as the image columns. Along each column of

the camera image, the hue of the pixels was summed to provide input to a certain location in the perceptual field. Activity peaks were formed in the perceptual field, detecting color objects along the horizontal dimension of the image. Positive activation in the perceptual field was projected onto the hue dimension and provided input to the CoS field. However, the CoS field could not achieve a peak without either a reward signal, which uniformly boosts the CoS field, or a targeted boost (preshape) from the intention field. The goal of the learning process was to learn the connection weights from the intention field.

The function of teacher-provided reward signal was to provide a *boost* to the CoS field activation. This boost allowed a peak to emerge in the output. Due to this peak, the CoS field and intention field had nonzero output at the same time. Under these conditions, the associative learning rule adapted the weights between the particular intention that is on (corresponding to which basic drive is active) and the CoS field.

However, other colors besides the one causing the reward were perceived and might be incorrectly associated with the drive's satisfaction. Thus, the robot has to get the reward in different perceptual contexts, and, since the true conditions of satisfaction have a low variance, and the incorrect perceptual conditions have high variance, the incorrect percepts will be averaged out by the associative Hebbian learning dynamics. Fig. 5 shows a snapshot of the system in action. After about 5 minutes in this simple environment, where we moved the objects around so many contexts could be experienced, the correct mappings were learned.

Once the weight matrix is learned, the reward becomes unnecessary to achieve satisfaction. The weights provide a sufficient boost to activate the CoS. This boost is selective for the perceptual conditions under which reward was achieved.

V. Conclusion and Outlook

Learning CoS amounts to learning a coupling structure between the intention of the EB and the respective CoS. In this coupling structure, an anticipation is represented of the forthcoming state, which terminates the action. To learn this coupling, the agent must be able to perceive an elementary reward signal, which might be driven by genetically encoded internal drives, such as hunger, thirst, curiosity, or by the emotional system of the agent. The learning mechanism presented in this paper is closely related to classical conditioning, where the internally generated reward corresponds to the reaction to the unconditioned stimulus, and the learned CoS to the conditioned reaction. This fundamental mechanism is at the core of the cognitive processes, and has been shown to be important for shaping the behavioral repertoire of an intentional agent.

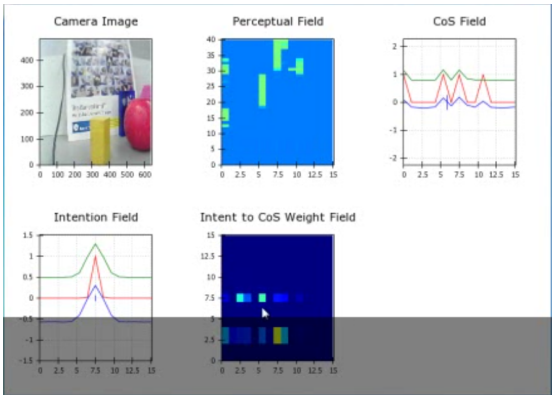

Fig. 5: Snapshot of the robot's dynamic fields. The peak in the Intention Field reflects the currently active drive. In the Perceptual Field, the colored objects (yellow, red, blue) provide inputs at different locations. For this drive, the color yellow in the center causes a reward, which gates adaptation in the weights from intention to the CoS field. When the robot experiences the reward in many different contexts, the incorrect cues in the CoS weights are diminished over time. The end result is that when this drive is active and the robot sees the color yellow in the center of the image, the CoS field peaks (correctly), and the behavior leading to the drive satisfaction will complete.

Acknowledgments

The authors gratefully acknowledge the financial support of the European Union Seventh Framework Programme FP7-ICT-2009-6 under Grant Agreement no. 270247 – NeuralDynamics; and of the DFG SPP *Autonomous Learning*, within Priority Program 1567.

References

[1] S Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.

[2] E Bicho, P Mallet, and G Schöner. Target representation on an autonomous vehicle with low-level sensors. *The International Journal of Robotics Research*, 19:424–447, 2000.

[3] R. A. Brooks. Do elephants play chess? *Robotics and Autonomous Systems*, 6(1-2):3–15, 1990.

[4] B Duran and Y Sandamirskaya. Neural dynamics of hierarchically organized sequences: a robotic implementation. In *Proceedings of 2012 IEEE-*

RAS International Conference on Humanoid Robots (Humanoids), 2012.

[5] Boris Duran, Yulia Sandamirskaya, and Gregor Schöner. A dynamic field architecture for the generation of hierarchically organized sequences. In Alessandro E.P. Villa, Wlodzislaw Duch, Peter Erdi, Francesco Masulli, and Günther Palm, editors, *Artificial Neural Networks and Machine Learning – ICANN 2012*, volume 7552 of *Lecture Notes in Computer Science*, pages 25–32. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33268-5. doi: 10.1007/978-3-642-33269-2_4. URL http://dx.doi.org/10.1007/978-3-642-33269-2_4.

[6] C.L. Hull. *Principles of behavior: an introduction to behavior theory*. Century psychology series. D. Appleton-Century Company, incorporated, 1943.

[7] S Kazerounian, M Luciw, M Richter, and Y Sandamirskaya. Autonomous reinforcement of behavioral sequences in neural dynamics. In *Proceedings of the Joint IEEE International Conference on Development and Learning & Epigenetic Robotics (ICDL-EPIROB)*, 2012.

[8] Mathis Richter, Yulia Sandamirskaya, and Gregor Schöner. A robotic architecture for action selection and behavioral organization inspired by human cognition. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2457–2464, 2012.

[9] Y. Sandamirskaya, M. Richter, and G. Schöner. A neural-dynamic architecture for behavioral organization of an embodied agent. In *IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL EPIROB 2011)*, 2011.

[10] Yulia Sandamirskaya and Gregor Schöner. An embodied account of serial order: How instabilities drive sequence generation. *Neural Networks*, 23(10):1164–1179, December 2010. ISSN 0893-6080. doi: DOI:10.1016/j.neunet.2010.07.012. URL http://authors.elsevier.com/offprints/NN2760/7f89af22f761d0dea464a43fdb300383.

[11] Yulia Sandamirskaya, Stephan K.U. Zibner, Sebastian Schneegans, and Gregor Schöner. Using dynamic field theory to extend the embodiment stance toward higher cognition. *New Ideas in Psychology*, (0):–, 2013. ISSN 0732-118X. doi: 10.1016/j.newideapsych.2013.01.002. URL http://www.sciencedirect.com/science/article/pii/S0732118X13000111.

[12] G Schöner. Dynamical systems approaches to cognition. In Ron Sun, editor, *Cambridge Handbook of Computational Cognitive Modeling*, pages 101–126, Cambridge, UK, 2008. Cambridge University Press.

[13] R.S. Woodworth. *Dynamic psychology, by Robert Sessions Woodworth*. Columbia University Press, 1918.