
Top-Down and Bottom-Up Interactions between Low-Level Reactive Control and Symbolic Rule Learning in Embodied Agents

Clément Moulin-Frier
SPECS Lab
Universitat Pompeu Fabra
Barcelona, Spain
clement.moulinfrier@gmail.com

Xerxes D. Arsiwalla
SPECS Lab
Universitat Pompeu Fabra
Barcelona, Spain
x.d.arsiwalla@gmail.com

Jordi-Ysard Puigbò
SPECS Lab
Universitat Pompeu Fabra
Barcelona, Spain
jordiyisard.puigbo@upf.edu

Martí Sanchez-Fibla
SPECS Lab
Universitat Pompeu Fabra
Barcelona, Spain
santmarti@gmail.com

Armin Duff
SPECS Lab
Universitat Pompeu Fabra
Barcelona, Spain
armin.duff@gmail.com

Paul FMJ Verschure
SPECS Lab
Universitat Pompeu Fabra & ICREA
Barcelona, Spain
paul.verschure@upf.edu

Abstract

Mammals bootstrap their cognitive structures through embodied interaction with the world. This raises the question: how the reactive control of initial behavior is recurrently coupled to the high-level symbolic representations they give rise to? We investigate this question in the framework of the "Distributed Adaptive Control" (DAC) cognitive architecture, where we study top-down and bottom-up interactions between low-level reactive control and symbolic rule learning in embodied agents. Reactive behaviors are modeled using a neural allostatic controller, whereas high-level behaviors are modeled using a biologically-grounded memory network reflecting the role of the prefrontal cortex. The interaction of these modules in a closed-loop fashion suggests how symbolic representations might have been shaped from low-level behaviors and recruited for behavior optimization.

1 Introduction

A major challenge in cognitive neuroscience is to propose a unified theory of cognition based on general models of the brain structure [20]. Such a theory should be able to explain how specific cognitive functions (e.g. decision making or planning) result from the particular dynamics of an embodied cognitive architecture in a specific environment. This has led to various proposals, formalizing how cognition arises from interaction of functional modules. Early implementations of cognitive architectures trace back to the era of Symbolic Artificial Intelligence, starting from the General Problem Solver (GPS, [21]) and has been followed by a number of subsequent architectures such as Soar [13, 12], ACT-R [1, 2] and their follow-ups. These architectures are considered top-down and representation-based in the sense that they consist of a complex representation of a task, which

has to be decomposed recursively into simpler ones to be executed by the agent. Although relatively powerful at solving abstract symbolic tasks, top-down architectures have enjoyed very little success at bootstrapping behavioral processes and taking advantage of the agent’s embodiment (nonetheless, several interfaces with robotic embodiment have been proposed, see [28, 24]). In contrast to top-down representation-based approaches, behavior-based robotics [7] emphasizes lower-level sensory-motor control loops as a starting point of behavioral complexity that can be further extended by combining multiple control loops together, e.g. as in the subsumption architecture [6]. Such approaches are also known as bottom-up and they generally model behavior without relying on complex knowledge representation and reasoning. This is a significant departure from Newell’s and Anderson’s views of cognition expressed in GPS, Soar and ACT-R. Top-down and bottom-up approaches thus reflect different aspects of cognition: high-level symbolic reasoning for the former and low-level embodied behaviors for the latter. However, both aspects are of equal importance when it comes to defining a unified theory of cognition. It is therefore a major challenge of cognitive science to unify both approaches into a single theory, where (a) reactive control allows an initial level of complexity in the interaction between an embodied agent and its environment and (b) this interaction provides the basis for learning higher-level symbolic representations and for sequencing them in a causal way for top-down goal-oriented control. We propose to split the problem of how neural and symbolic approaches are integrated into the following three research questions:

- How are high-level symbolic representations shaped from low-level reactive behaviors in a bottom-up manner?
- How are those representations recruited in rule and plan learning?
- How do rules and plans modulate reactive behaviors through top-down control for realizing long-term goals?

To address these questions, we adopt the principles of the *Distributed Adaptive Control* (DAC) theory of the mind and brain [31, 30], which posits that cognition is based on the interaction of four interconnected control loops operating at different levels of abstraction (Fig. 1). The first level is the embodiment of the agent within its environment, with the sensors and actuators of the agent (called the Somatic layer). The Somatic layer incorporates physiological needs of the agent (e.g. exploration or safety) and which drives the dynamics of the whole architecture [23]. Extending behavior-based approaches with drive reduction mechanisms, complex behavior is bootstrapped in DAC from the self-regulation of an agent’s physiological needs when combined with reactive behaviors (the Reactive layer). This reactive interaction with the environment drives learning processes for acquiring a state space of the agent-environment interaction (the Adaptive layer) and the acquisition of higher-level cognitive abilities such as abstract goal selection, memory and planning (the Contextual layer). These high-level representations in turn modulate behavior at lower levels via top-down pathways shaped by behavioral feedback. The control flow in DAC is therefore distributed, both from bottom-up and top-down interactions between layers, as well as from lateral information processing into the subsequent layers.

In this paper, we present biologically-grounded neural models of the reactive and contextual layers and discuss their possible integration to address the question of how the reactive control of initial behavior is recurrently coupled to the high-level symbolic representations they give rise to. On one hand, our model of the reactive layer relies on the concept of allostatic control. Sterling proposes that allostasis drives regulation through anticipation of needs [27]. In [26], allostasis is seen as a reactive meta-regulation system of homeostatic loops that are possibly contradicting each other in a winner-takes-all process, modulated by emotional and physiological feedback. On the other hand, we present a neural model of the contextual layer for rule and plan learning grounded in the neurobiology of the prefrontal cortex (PFC) [8]. It has been shown that representations of sensory states, actions and their combinations can be found in the PFC [16, 9] and that is reciprocally connected to sensory as well as motor areas [9]. This puts the PFC in a favorable position for the representation of sensory-motor contingencies, i.e. patterns of sensory-motor dependencies [22], selected according to their relevance in goal-oriented behavior. Its involvement in flexible cognitive control and planning (e.g. [10, 5]) backs the hypothesis that sensory-motor contingencies promote flexible cognitive control and planning. This paper aims at identifying the key neurocomputational challenges in integrating both models in a complete cognitive architecture.

The next section introduces a model of the reactive layer based on the concept of allostatic control and is concerned with prioritization of multiple self-regulation loops. We then present an existing model [8] of the prefrontal cortex that is able to integrate sensory-motor contingencies in rules and plans for long-term reward maximization. Finally, we discuss how both models can be integrated to

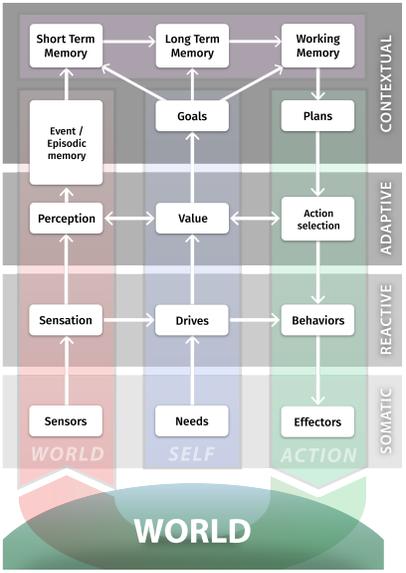


Figure 1: DAC proposes that cognition is organized as a layered control structure with tight coupling within and between these layers (adapted from [30]): the Somatic, Reactive, Adaptive and Contextual layers. Across these layers, a columnar organization exists that deals with the processing of states of the World or exteroception (left, red), the Self or interoception (middle, blue) and Action (right, green). The role of each layer and their interaction is described in the text.

bridge the gap between low-level reactive control and high-level symbolic rule learning in embodied agents. We will make a particular emphasis on how the acquired rules are able to inhibit the reactive system to achieve long-term goals as proposed in theories on the neuropsychology of anxiety [11] and consciousness [29, 3, 4].

2 A neural model of allostatic control

In this section we describe the concept of allostatic control and introduces a neural model for it. Based on the principles of DAC, we consider an embodied agent endowed with physiological needs for e.g. foraging or safety and self-regulating them through parallel drive reduction control loops. Drives aim at self-regulating internal state variables within their respective homeostatic ranges. Such an internal state variable could, for example, reflect the current glucose level in an organism, with the associated homeostatic range defining the minimum and maximum values of that level. A drive for foraging would then correspond to a self-regulatory mechanism where the agent actively searches for food whenever its glucose level is below the homeostatic minimum, and stops eating even if food is present whenever levels are above the homeostatic maximum. A drive is therefore defined as the real-time control loop triggering appropriate behaviors whenever the associated internal state variable goes out of its homeostatic range, as a way to self-regulate its value in a dynamic and autonomous way. It is common for drives to be competing and we thus require a method to prioritize them. Consider for example a child in front of a transparent box full of candies, with a caregiver telling her to not open the box. Two drives are competing in such a situation: one for eating candies and one for obeying to the caregiver. Depending on how fond of candies she is and of how strict the caregiver is, she will choose to either break a social rule or to enjoy an extremely pleasant moment. Such regulation conflicts can be solved through the concept of an allostatic controller [26], defined as a set of parallel homeostatic control loops operating in real-time and dealing with their prioritization to ensure an efficient global regulation of multiple internal state variables.

The allostatic control model we introduce in this paper is composed of three subsystems (Fig. 2). Motor primitive neurons are connected to the agent actuators through synaptic connections. For example, on a 2-wheeled mobile robot, a "turn right" motor primitive would excite the left wheel actuator and inhibit the right wheel one, whereas a "go forward" motor primitive would excite

both. A repertoire of behaviors (inner boxes inside the middle box) link sensation neurons (middle nodes in the behavior boxes) to motor primitive neurons (right nodes in the behavior boxes) through behavior-specific connections. In a foraging behavior for example, sensing food on the right would activate a "turn right" motor primitive, whereas in an obstacle avoidance action, sensing an obstacle on the right would activate a "turn left" motor primitive. In Figure 2, two behaviors are connected to the two same motor primitives. In the general case however, they could only connect to a specific relevant subset of the motor primitives. The agent's exteroceptive sensors (vertical red bar on the middle) are connected to sensation nodes in the behavior subsystem (middle nodes in each inner box) by connections not shown in the figure. Each behavior is provided with an input activation node (left node in each behavior box) which has a baseline activity (indicated by the number 1) inhibiting the sensation nodes. Therefore, if no input is provided to an activation node, the sensation nodes of the corresponding behavior are inhibited, preventing information to propagate to the motor primitives. Drive nodes (left) can activate their associated behavior through the inhibition of the corresponding activation nodes, that in turn disinhibit sensation nodes through a double inhibition process, allowing sensory information to propagate to the motor primitives. In Figure 2, two drives nodes are represented, each connected to its associated behavior which is supposed to reduce the drive activity (e.g. a foraging drive connecting to food attraction behavior). Drive nodes are activated through connections from the agent's interoceptive sensors (vertical red bar on the left) and form a winner-takes-all process through mutual inhibition. This way, drives compete against each other as in the child-caregiver example presented above.

3 Contextual layer: a neural implementation of a rule learning system

In the context of the DAC architecture, we use an existing biologically-grounded model for rule learning and planning developed in our research group [8] based on key physiological properties of the prefrontal cortex (PFC), i.e. reward modulated sustained activity and plasticity of lateral connectivity. Recent proposals highlight the role of sensory-motor contingencies as the building blocks of many intelligent behaviors including rule learning and planning. Sensory-motor contingencies combine information about perceptual inputs and related motor actions forming internal states, which are subsequently used to structure and plan behavior. The DAC architecture has been developed to investigate how sensory-motor contingencies can be formed and exploited for behavioral control such as rule learning and flexible planning. Contingencies formed at the level of the adaptive layer provide inputs to the contextual layer, which acquires, retains, and expresses sequential representations using systems for short-term and long-term memory. The PFC-grounded contextual layer consists of a group of laterally connected memory-units. Each memory-unit is selective for one specific stimulus, e.g. color, and can induce one specific action, e.g. "go left", "go forward" or "go right", forming a sensory-motor contingency (Fig. 3). A memory-unit can be interpreted as a micro-column comprising a number of neurons with the same coding properties. Sequential rules are expressed through the coordinated activation of different memory-units in the correct order. This group of memory-units forms the elementary substrate for the representation and expressions of rules.

The activity of memory-units in this architecture is driven by perceptual inputs, observed reward and state prediction through the lateral connectivity. Memory-units compete in a probabilistic selection mechanism. The higher the activity, the higher the probability of being selected. The selected memory-units propagate their activity and contribute to the final action of the agent. To express a rule, the activity of these memory-units must be modulated in order to control the selection of specific units contributing to the final action. The modulated activity of these units is influenced by two systems, the lateral connectivity and the reward system. The lateral connectivity captures the context and the order of the sequential rules and influences activity through trigger values. Trigger values allow to chain through a specific sequence of memory-units. The reward system validates different rules represented in the network and influences the activity through the reward value. The modulation of each memory-unit activity is realized by multiplying the perceptual activity by the trigger value as well as the reward value. This model has been validated in simulated robotic experiments where stimuli-response associations have to be learned to maximize reward in a multiple T-maze environment, as well in solving the Towers of London task. Moreover the model is able to re-adapt to a changing environment, e.g. when the stimuli-response associations are modified on-line [8].

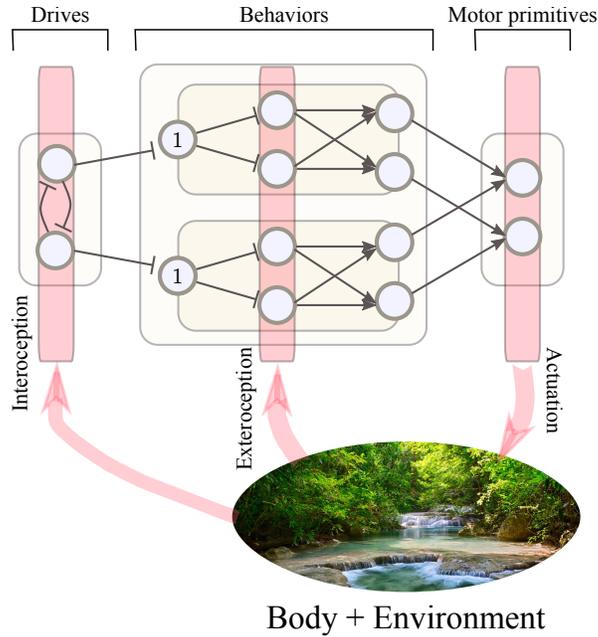


Figure 2: The allostatic control model we introduce in this paper is composed of three subsystems: Drives, Behaviors and Motor Primitives, represented by the left, middle and right boxes, respectively. The behavior subsystem is itself composed of multiple behaviors represented by the inner boxes (two behaviors are represented in the figure). Inside those subsystems, round nodes represent neurons that can have a baseline activation (indicated by a number 1) or not (no number). Synaptic connections between neurons are represented by dark arrows, where a peak end (resp. a bar end) indicates an excitatory (resp. inhibitory) connection. Vertical red bars represent neural populations that act as the interface of the agent with its environments through interoceptive sensations (left), exteroceptive sensations (middle) and actuation (right). Those neural populations interact in a predefined way with the neuron nodes they overlap with, though connections not shown in the figure. On a 2-wheeled mobile robot for example, motor primitives for e.g. turning right or going forward would be represented by nodes in the motor primitive subsystem and connected to the left and right wheels actuators represented by the vertical red bar on the right. Similarly, left and right proximeter sensors would be represented by the middle vertical bar and connected to the sensory neurons of the behaviors that require this information, as it would be the case of the sensory neurons of an obstacle avoidance behavior but not of a light following one (because the latter behavior doesn't require proximeter information).

4 Bridging the gap between reactive control with continuous sensory-motor contingencies and symbolic prediction through rule learning

We have presented in the two last sections two neural models.

- The first, that we call the *reactive layer*, implements an allostatic controller which deals with the parallel execution of possibly conflicting self-regulation control loops in real-time. It is composed of three subsystems: *Drives* that are activated through interoception (related e.g. to the organism's glucose level for a foraging drive), *Behaviors* implemented by reactive control loops (e.g. attraction to food spot when the glucose level is low) and activation of *Motor Primitives* that send the required commands to the agent's actuators (e.g. turning left or turning right).
- The second, that we call the *contextual layer*, implements a rule learning system composed of a large number of memory units that each encodes a specific sensory-motor contingency, i.e. is sensitive to a particular sensory receptive field and activates a particular action. Lateral connectivity between units, learned from the agent's experience, allows the temporal chaining of sensory-motor contingencies through action. Each unit is associated with a reward value that is also learned and predict the (possibly delayed) reward expected by executing the corresponding action encoded by the unit in the corresponding sensory state it is selective to. At a given time, the activity of each memory

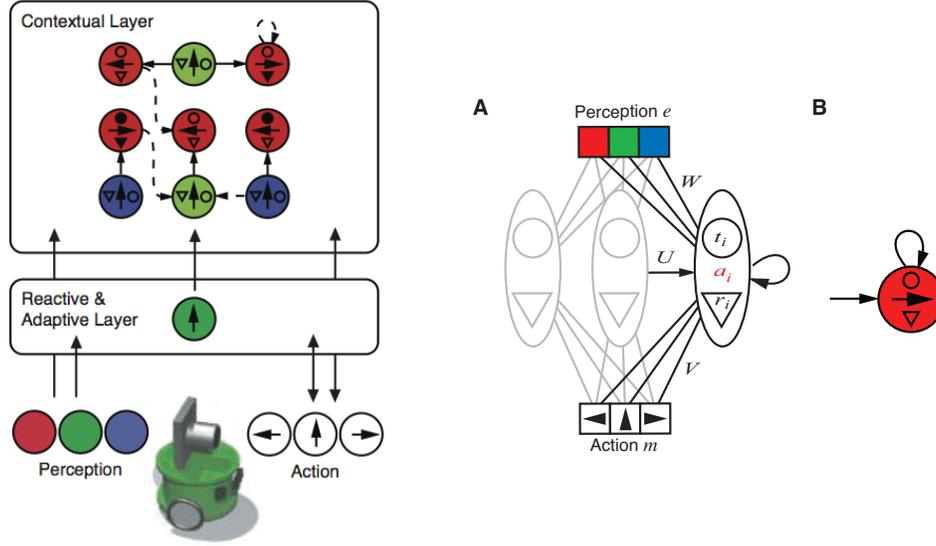


Figure 3: Left: In the DAC architecture the reactive/adaptive layer provides the agent with simple automatic behaviors and forms sensory–motor contingencies used in the contextual layer. Each circle stands for a bi-modal perception–action memory unit. The color of the circle indicates the perception selectivity while the arrow indicates the action specificity. The contextual layer is driven by the sensory input. The lateral connectivity primes the sequence of the activation. A reward-value system modifies the activity in order to select between the memory-units primed by the lateral connectivity. An action selection mechanism provides the final action for the agent. Right: Implementation of a single memory-unit (in inset A). The perceptual selectivity a_i is driven by perception e through the weight matrix W . The trigger value t_i is driven by the activation of other memory-units through the (learned) weights U . The reward value is noted r_i . The motor activity induced by a memory-unit is driven by its total activity $a_i * t_i * r_i$ through the weight matrix V . Inset B: symbolic representation of the same memory-unit, where the color represents the perceptual selectivity, the inner arrow the motor specificity and the external arrows the lateral connectivity.

unit is driven by the sensitivity of that unit to the current agent’s perception, the prediction of that perception from the lateral connections and the reward value learned by the unit over time. The most activated units then compete together to decide what next action should be performed to maximize future rewards.

In this section, we discuss how both layers can be integrated in a complete cognitive architecture where contextual rules are learned from sensory-motor actions generated by reactive control, and where the actions generated by the contextual layer in turn, modulates the activity of the reactive system to achieve long-term goals. To illustrate the cognitive dynamics at work, let’s consider again the example we have described above about a child that experiences a conflict between eating candies and disobeying to a caregiver. How the reactive and contextual layer models we have presented in the two last sections interact together in such a situation? Without any previous experience of it, only the reactive layer generates behavior through the self-regulation of internal drives, here eating candies and obeying the caregiver. Satisfying one or the other drive depends on their respective current level as well as the parameters of their mutual inhibition (one could conceive it as personality, rather a rebel or a good child). Since this situation occurs in various contexts, where the drive initial activities differ, the consequences of both behaviors are experienced by the child. When she obeys the caregiver, she is frustrated of not eating a candy. When she disobeys, she has an argument with the caregiver and likely doesn’t eat any candy anyway according to how strict her opponent is. By experiencing multiple times the consequences of the two actions, she will discover that respecting

the social rule is of better interest to maximize reward, consequently self-inhibiting her own drive for eating candies by contextual top-down control acting on the reactive layer. Besides this real-life example, such a top-down "behavioral inhibition system" has been proposed as a major component of theories on the neuropsychology of consciousness [29] and anxiety [11].

A computational implementation of the reactive and contextual layer integration has to solve the issues of (a) preprocessing sensory-motor information generated by the reactive layer to provide relatively abstract and stable perceptions of the environment at the contextual level, (b) modulating the memory-unit activities through reward values derived from the drive levels, (c) modulating the activity of drive levels from the output generated by the contextual layer to maximize reward.

Solving (a) requires the addition of an adaptive layer in the architecture that acquires a state space of the agent-environment interaction through perceptual learning mechanism modulated by reward. In previous computational models of the DAC architecture, this is accounted as associative learning [14]. In the context of the two models presented in this paper, a solution consists of a direct connection from the sensation neurons of the reactive layer to the perceptual activation of memory units as in Fig. 3. In the current version of the rule learning model, a large number of memory-units is generated with pseudo-random perceptual sensitivity and motor specificity. Therefore the total number of units has to be large compared to the number of units that are actually recruited for generating actions after learning. Adaptively tuning memory-unit sensitivity would allow optimization of the size of the network, e.g. through a learning rule able to attract perceptual sensitivity of memory units according to their rewarding effect. Note that a similar learning rule can be applied for learning the motor specificity of memory-units. Besides a direct connection between the neural units of both layers, neural hierarchies would also improve rule learning by providing more abstract information to memory-units. This can be achieved using recent advances in Deep Reinforcement Learning, as e.g. in [17], able to learn highly abstract perceptual representations from raw sensory data driven by reward; or by adopting a more biologically-grounded approach using neurocomputational models of perceptual learning grounded in the neurobiology of the cerebral cortex interaction with the amygdala [25]. Neuromodulators independent from reward are also present during aversive events, sustained attention or surprise, modulating not just memory formation but perception as well. The temporal stabilization of sensory features have been proposed in [32]. Note that not only sensory information can be abstracted and stabilized, but hierarchies of needs [15] and behaviors as well.

To solve (b), reward values can directly be computed from drive activities: the lower the drive activity the better it is satisfied. Drive-related activities can also modulate the reward value of memory units according to the physiological context. This way, lateral connectivity among memory units is coding for the causal link between sensory-motor contingencies, supposed to be context-independent, whereas reward activation depends on the current configuration of drive levels (one could call it the emotional state of the agent). This will allow the memory network to behave over different rule sets according to the internal state of the agent (the rules maximizing reward when an agent is hungry are not the same as when it is sleepy), avoiding learning interference between different situations.

Finally, regarding (c), we propose that the contextual layer's activity modulates the reactive one by directly acting on the drive levels, instead of acting latter in the reactive layer pipeline, e.g. on the motor primitives. By directly modulating the drive levels, the contextual layer takes full control of the reactive one by acting on it from the source (see Fig. 2. This allows the agent to self-inhibit some of its own drives for maximizing reward on the long term as in the child-caregiver example.

5 Conclusion

In this paper we have designed an allostatic controller and introduced a novel computational model implementing it. This reactive layer allows the self-regulation of multiple drive-reduction control loop operating in parallel. Then we have presented an existing computational model of a contextual layer grounded in the neurobiology of the prefrontal cortex and able to learn sequential rules and plans through experience. Finally, our main contribution in this paper has been to argue for both a bottom-up and top-down interaction between low-level reactive control and high-level contextual plans. Symbolic representations in our approach are learned as sensory-motor contingencies encoded in discrete memory-units from the activity generated by the reactive layer. In turn, the actions generated by the contextual layer modulates the activity of the reactive system through a top-down pathway, inhibiting reactive drives to achieve long-term goals.

Both the reactive and contextual models are implemented and we are now working on their computational integration. We have identified in the last section the main challenges in term of bottom-up perceptual abstraction, multitask reward optimization as well as top-down drive modulation. This will allow a computational implementation of a reactive agent, embodied in a physical mobile robot and progressively acquiring contextual rules of its environment from experience, thus demonstrating an increasingly rational behavior.

We will also put a particular emphasis on applying this integrated cognitive architecture to study the formation of social norms in multi-robot setups [19, 18]. The long term goal is to understand how the constraints imposed by a multi-agent environment favor conscious experience. The research direction we adopt is based on the hypothesis that social norms are needed for the evolution of large multi-agent groups and that the formation of those social norms requires each individual to take conscious control of its own drive system [29].

Acknowledgments

Work supported by ERC's CDAC project: "Role of Consciousness in Adaptive Behavior" (ERC-2013-ADG 341196); & EU projects *Socialising Sensori-Motor Contingencies* socSMC-641321—H2020-FETPROACT-2014 & *What You Say Is What You Did* WYSIWYD (FP7 ICT 612139).

References

- [1] J. R. Anderson. *The Architecture of Cognition*. Harvard University Press, 1983.
- [2] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin. An integrated theory of the mind. *Psychological review*, 111(4):1036–1060, 2004.
- [3] X. D. Arsiwalla, I. Herreros, C. Moulin-Frier, M. Sanchez, and P. F. Verschure. Is consciousness a control process? In *International Conference of the Catalan Association for Artificial Intelligence*, pages 233–238. IOS, 2016.
- [4] X. D. Arsiwalla, I. Herreros, and P. Verschure. On three categories of conscious machines. In *Conference on Biomimetic and Biohybrid Systems*, pages 389–392. Springer, 2016.
- [5] W. F. Asaad, G. Rainer, and E. K. Miller. Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology*, 84(1):451–459, 2000.
- [6] R. Brooks. A robust layered control system for a mobile robot. *IEEE Journal on Robotics and Automation*, 2(1):14–23, 1986.
- [7] R. A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47(1-3):139–159, 1991.
- [8] A. Duff, M. S. Fibla, and P. F. Verschure. A biologically based model for the integration of sensory–motor contingencies in rules and plans: A prefrontal cortex based extension of the distributed adaptive control architecture. *Brain research bulletin*, 85(5):289–304, 2011.
- [9] J. Fuster. *The Prefrontal Cortex: Anatomy, Physiology, and Neurophysiology of the Frontal Lobe*. Lippincott-William & Wilkins, Philadelphia, 1997.
- [10] J. M. Fuster, G. E. Alexander, et al. Neuron activity related to short-term memory. *Science*, 173(3997):652–654, 1971.
- [11] J. A. Gray and N. McNaughton. *The neuropsychology of anxiety: An enquiry into the function of the septo-hippocampal system*. Number 33. Oxford university press, 2003.
- [12] J. E. Laird. *The Soar Cognitive Architecture*. MIT Press, 2012.
- [13] J. E. Laird, A. Newell, and P. S. Rosenbloom. SOAR: An architecture for general intelligence. *Artificial Intelligence*, 33(1):1–64, 1987.
- [14] E. Marcos, M. Ringwald, A. Duff, M. Sánchez-Fibla, and P. F. Verschure. The hierarchical accumulation of knowledge in the distributed adaptive control architecture. In *Computational and robotic models of the hierarchical organization of behavior*, pages 213–234. Springer, 2013.

- [15] A. Maslow. A theory of human motivation. *Psychological review*, 1943.
- [16] E. K. Miller, L. Li, and R. Desimone. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *The Journal of Neuroscience*, 13(4):1460–1478, 1993.
- [17] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.
- [18] C. Moulin-Frier, M. Sanchez-Fibla, and P. F. Verschure. Autonomous development of turn-taking behaviors in agent populations: a computational study. In *IEEE International Conference on Development and Learning, ICDL/Epirob, Providence (RI), USA*, 2015.
- [19] C. Moulin-Frier and P. F. Verschure. Two possible driving forces supporting the evolution of animal communication. comment on "towards a computational comparative neuroprimatology: Framing the language-ready brain" by michael a. arbib. *Physics of life reviews*, 16:88–90, 2016.
- [20] A. Newell. *Unified theories of cognition*. Harvard University Press, 1990.
- [21] A. Newell, J. C. Shaw, and H. A. Simon. Report on a general problem-solving program. *IFIP Congress*, pages 256–264, 1959.
- [22] J. K. O'Regan and A. Noë. A sensorimotor account of vision and visual consciousness. *The Behavioral and brain sciences*, 24(5):939–73; discussion 973–1031, oct 2001.
- [23] J.-Y. Puigbò, C. Moulin-Frier, and P. F. Verschure. Towards self-controlled robots through distributed adaptive control. In *Conference on Biomimetic and Biohybrid Systems*, pages 490–497. Springer, 2016.
- [24] J.-Y. Puigbo, A. Pumarola, C. Angulo, and R. Tellez. Using a cognitive architecture for general purpose service robot control. *Connection Science*, 27(2):105–117, 2015.
- [25] J.-Y. Puigbò, G. Maffei, M. Ceresa, G.-B. M.A., and V. P.F.M.J. Learning relevant features through a two-phase model of conditioning. *IBM Journal of Research and Development, Special issue on Computational Neuroscience*, Accepted, in Press.
- [26] M. Sanchez-Fibla, U. Bernardet, E. Wasserman, T. Pelc, M. Mintz, J. C. Jackson, C. Lansink, C. Pennartz, and P. F. Verschure. Allostatic control for robot behavior regulation: a comparative rodent-robot study. *Advances in Complex Systems*, 13(03):377–403, 2010.
- [27] P. Sterling. Allostasis: a model of predictive regulation. *Physiology & behavior*, 106(1):5–15, 2012.
- [28] G. Trafton, L. Hiatt, A. Harrison, F. Tamborello, S. Khemlani, and A. Schultz. ACT-R/E: An Embodied Cognitive Architecture for Human-Robot Interaction. *Journal of Human-Robot Interaction*, 2(1):30–55, Mar 2013.
- [29] P. F. Verschure. Synthetic consciousness: the distributed adaptive control perspective. *Phil. Trans. R. Soc. B*, 371(1701):20150448, 2016.
- [30] P. F. M. J. Verschure, C. M. A. Pennartz, and G. Pezzulo. The why, what, where, when and how of goal-directed choice: neuronal and computational principles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655):20130483, 2014.
- [31] P. F. M. J. Verschure, T. Voegtlin, and R. J. Douglas. Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature*, 425(6958):620–624, 2003.
- [32] Y. Yamashita and J. Tani. Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment. *PLoS Comput Biol*, 4(11):e1000220, 2008.