Is Consciousness a Control Process?

Xerxes D. ARSIWALLA [a,1], Ivan HERREROS [a], Clément MOULIN-FRIER [a],
Marti SANCHEZ [a] and Paul VERSCHURE [a,b]

[a] *Synthetic Perceptive Emotive and Cognitive Systems (SPECS) Lab, Center of Autonomous Systems and Neurorobotics, Universitat Pompeu Fabra, Barcelona, Spain.*

[b] *Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.*

Abstract. Understanding the nature of consciousness has been an outstanding scientific puzzle at the crossroads of neuroscience and artificial intelligence. While brains have long since known to be the bearers of consciousness and machines, that of computation, the history of cybernetics has been full of attempts trying to synthesize consciousness in computational architectures. In recent years, ideas from control theory have proven to be extremely useful for addressing systems-level questions in neuroscience and designing cognitive architectures. Extending these ideas to the study of consciousness, we discuss the core functions of consciousness and control architectural specifications of agents capable of operationalizing these functionalities. We suggest that evolutionary pressures on social dynamics of interacting agents leads to the emergence of consciousness, which is a process for predicting intentional states of other agents (and self) in order to generate social cooperative and competitive behaviors necessary to optimize an agent's survival drives in a world with limited resources.

Keywords. Theoretical Consciousness, Control Theory, Multi-Agent Systems

1. Introduction

Can consciousness be understood as a high-level control process? The motivation for this question arises from two other often asked questions: "Do brains compute?" and "Can a machine be conscious?". Computation defined in the sense of Turing refers to a finite-time function evaluation process. The brain is certainly not a Turing machine and computation defined in this sense is not considered its primary function. But the brain does generate consciousness. Something that today's machines / artificial intelligence (AI) systems do not exhibit, in spite of intensive computational routines driving them. On one hand, computational and systems neuroscience research has made great strides in unravelling circuit-level mechanisms of many cognitive functions as well as pathology-related dysfunctions; on the other hand, AI has led to remarkable progress in deep neural architectures and robotics. Yet, the larger question of consciousness has remained elusive for both, the cognitive neuroscience as well as the artificial intelligence community. However, in recent years, the role of control theory has begun to gain prominence in both neuroscience as well as AI. For instance, in modeling and understanding the mechanisms

[1] Corresponding Author: Xerxes D. Arsiwalla, Universitat Pompeu Fabra, Barcelona, Spain; E-mail: x.d.arsiwalla@gmail.com

of sensory-motor actions, adaptive learning, goal-oriented behavior and multi-agent systems, to name a few. Extending this line of thinking, in this paper we ask whether the appropriate framework to address the problem of consciousness is that of control theory? From this point of view, we address higher-level functions that conscious agents need to solve and specify the control architectures required to accomplish these functions within the framework of multi-agent systems.

2. Towards a Control Architecture of Consciousness

2.1. Measures of Biological Consciousness

We begin with a discussion on measures of biological consciousness. These are practical measures widely used for clinical purposes to assess varying levels of consciousness in patients with disorders of consciousness ranging from coma, locked-in syndrome to those in vegetative states. In such cases, levels of consciousness are assessed through a battery of behavioral tests as well as neurophysiological recordings. In particular, these assessments calibrate levels of arousal and awareness in patients [5], [4], leading to a two dimensional operational definition of clinical consciousness. For clinical purposes, closely associated states of consciousness can be grouped into clusters on a plane with awareness and arousal representing orthogonal axes [5], [4]. Scales of awareness target both higher- and lower-order cognitive functions enabling complex behavior. Arousal results from biochemical homeostatic mechanism regulating survival drives and is clinically measured in terms of glucose metabolism in the brain. Clinical consciousness is thus assessed as a bivariate measure on these two axes. In fact, in all known organic life forms, biochemical arousal is a necessary precursor supporting the hardware necessary for cognition. In turn, evolution has shaped cognition in such a way so as to support the organism's basic survival as well as higher-order drives associated to cooperation and competition in a multi-agent environment. Awareness and arousal thus form a closed-loop, resulting in phenomenological consciousness. What would biological arousal and awareness map to in synthetic agents? As noted above, arousal results from autonomous homeostatic mechanisms necessary for the self-preservation of an organism's germ line in a given environment. In other words, arousal results from self-sustaining drives necessary for basic survival, whereas awareness can be functionally abstracted as general forms of intelligence, necessary for higher-order functions in a complex social world. As per the discussion above, biological consciousness, as we know it, is phenomenologically an amalgamation of life and intelligence. A functional notion of consciousness in synthetic systems requires an analogous coupling of drives and complex social behaviors. In the next section, we discuss what these behaviors entail.

2.2. What is the Functional Role of Consciousness?

As discussed above, awareness and arousal are both signatures of consciousness. These are required in order to generate action and perception in a multi-agent environment. In turn, these action-perception loops are used for maintaining and regulating the agent's homeostatic needs and drives. This leads to goal-oriented behaviors that optimize objective functions or drives based on the agent's predictions of states of the physical world (the world model) as well as predictions of behavioral states of other agents [10]. The former comprises the agent's world model, while the latter, the agent's social or exteroceptive intentional model. Both of these are required in order to engage in social cooperation and competition imposed by evolution. The question thus becomes: What does it take for an autonomous agent to act? In order to act in the physical world an agent needs to determine a behavioral procedure to achieve a goal state (the "How" of action), which in turn requires defining the "Why" (the motivation for action in terms of needs, drives and goals), "What" (the objects and their affordances in the world that pertain to these goals), "Where" (the location of objects in the world, the spatial configuration of the task domain and the location and confirmation of the self), and "When" (the sequencing and timing of action relative to the dynamics of the world and self). Goal-oriented action in the physical world emerges from the interplay of these diverse processes. Besides goal-oriented behavior, consciousness requires one more ingredient, namely, the ability to maintain a transient and autonomous memory of the virtualized agent environment interaction that captures the hidden states of the external world in particular the intentional states of other agents and the norms that they implicitly convey through their actions. The states of other agents that are predictive of their actions however, are hidden. At best they can be inferred from incomplete sense data such as location, posture, vocalizations, social salience, etc. As a result the agent faces the challenge to univocally assess, in a deluge of sensor data those exteroceptive and interoceptive states that are relevant to ongoing and future action and to deal with the ensuing credit assignment problem to optimize its own actions. This results in a reciprocity of behavioral dynamics, where the agent is now acting on a world that is in turn acting upon itself.

2.3. A Control-Theoretic Cognitive Architecture

Based on the functional requirements for consciousness discussed above, we now describe a control architecture that has been used for foraging tasks with robots [11]. The Distributed Adaptive Control (DAC) theory of mind and brain is a framework within which several neurobiological-grounded cognitive paradigms have been implemented and validated through neuro-robotic experiments [6]. DAC generates agent behavior out of a distributed hierarchical layered structure including: the somatic, reactive, adaptive and contextual layers (figure 1). The somatic layer defines the fundamental interface between the embodied agent and its environment, including drives that must be satisfied in order to assure physical integrity and survival. The reactive layer describes innate behavioral systems comprising reflexes and low-level stereotyped behavioral patterns. The adaptive layer of DAC captures perceptual and behavioral learning systems such as stimulus-stimulus and stimulus-response associations studied in classical conditioning. The contextual layer describes goal oriented decision-making abilities of the brain, built on sequential memory systems. The most recent implementation of the DAC architecture includes detailed computational models of the cerebellum, hippocampus and prefrontal cortex [6], and this system has been validated in foraging paradigms with robots. Figure 1 shows an abstract representation of the DAC theory. DAC is organized along four layers (soma, reactive, adaptive and contextual) and three columns (World, Self, Action). The soma designates the body with its sensors, organs and actuators. It defines the drives or needs, that the agent must satisfy in order to survive. The reactive layer comprises dedicated behavior systems each implementing predefined sensorimotor mappings serving the drives. In order to allow for action selection, task switching and conflict resolution, all behavior systems are regulated via an allostatic controller that sets their internal homeostatic dynamics relative to overall system demands and opportunities. The adaptive layer acquires a state space of the agent-environment interaction and shapes action. The learning dynamics of of the adaptive layer are constrained by value functions defined by the allostatic control of the reactive layer and minimize perceptual and behavioral prediction errors. Finally, the contextual layer further expands the time horizon in which the agent can operate through the use of sequential short and long-term memory systems. The short-term memory acquires conjunctive sensorimotor representations that are generated by the adaptive layer as the agent acts in the world. These representations are retained as goal-oriented sequences in the long-term memory only when positive value is encountered, as defined by the reactive layer and/or the adaptive layer. The contribution of stored long-term memory sequences to decision-making depends on four factors: perceptual evidence, memory chaining, valence and the expected cost of reaching a goal state. The content of working memory is then defined by the dynamics of this four-factor decision making model. Additionally, across the four hierarchical layers of DAC, there exists a columnar organization that at every level of the hierarchy deals with the processing of states of the world grounded in exteroception; those of the self, derived from interoception; and those of action sensed through proprioception. The latter mediates between the first two via the environment.

2.4. Multi-Agent Systems

While current implementations of robots or agent simulations using DAC or other cognitive architectures, by themselves are not considered as being conscious, the discussion above on the function of consciousness suggests that the missing ingredient in almost all of these cases is the ability to predict intentional states of self and that of other agents, which are crucial for engaging in social behaviors including cooperation and competition in any multi-agent environment. From this perspective, in this work we posit that consciousness is not a specific cognitive module, but rather the result of interactions between existing functional modules (as those illustrated in figure 1). In this sense, consciousness does not directly enable intentionality or social dynamics, but rather the other way around, that evolutionary pressures on social dynamics lead to the emergence of agent behavioral states that we refer to as consciousness. The appropriate framework for studying emergence and evolution of behavioral traits is that of interacting multi-agent systems capable of learning and adaptation, such as interacting DAC agents. We also point to evidence in the literature where different types of multi-agent interactions have been investigated. Studies on simulations of evolving agents have shown that in different contexts and without explicit encoding of a concrete fitness function (only by conveying a reward signal to the agent), complex sensorimotor learning with predictive abilities can emerge [8,1,2]. In [8] for example, individual agents evolve memory and attentional capacities to maximize their rewards. In [1,2] multiple agents evolve entrainment capabilities between agents. Other works have applied multi-agent systems to the modeling of evolutionary and developmental dynamics of speech and language. Since the pioneering paradigm of "language games" proposed by Luc Steels [9], a number of such multi-agent model simulations have been proposed showing how particular properties of speech [7] and language [3] can self-organize out of repeated local interactions between agents of a

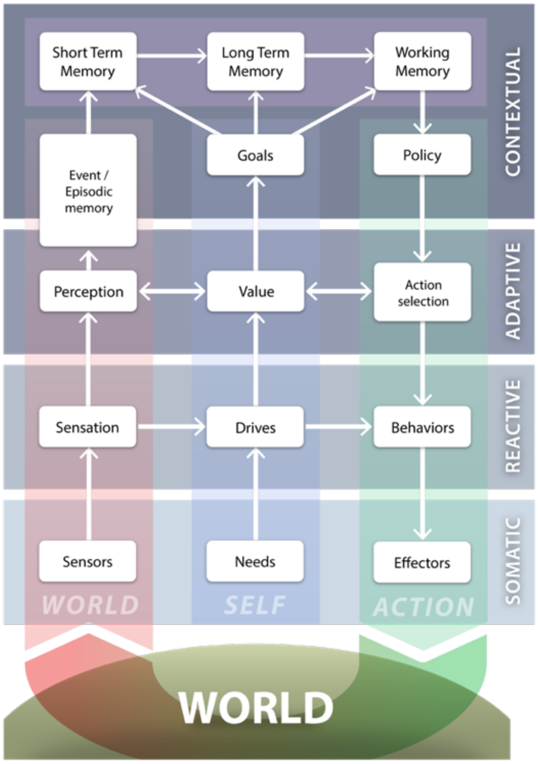

Figure 1. Abstract representation of the Distributed Adaptive Control (DAC) cognitive architecture. DAC proposes that the cognition is organized as a layered control structure with tight coupling within and between these layers distinguishing: the soma, the reactive, adaptive and contextual layers. Across these layers, a columnar organization exists that deals with processing of states of the world or exteroception (left column), the self or interoception (middle column) and action (right column). Adapted from [10].

population. These models consider agents equipped with perception, action and learning abilities, interacting together in repeated and random dyadic interactions (called games). The aim is to analyze how a global communication system shared by the entire population can self-organize out of those repeated local interactions, where These examples show that experimentation with artificial agents will turn out to be a key methodology to capture and control the dynamics of social behavior at various time scales. It allows to study emergent properties at the population level from local interaction dynamics, as well as the extent to which the presence of social peers in a shared environment can impact individual learning dynamics towards the acquisition of social representations.

3. Discussion

In this work, we have advanced the view that evolutionary pressures on social dynamics of interacting agents leads to the emergence of consciousness. The solution to survival in an only partially observable intentional world entails assessing what the relevant (hidden) states of the world and importantly those of other agents are, what the relevant states of self are, what the norms are that other agents follow, and which specific state of the agent or action, from a large repertoire of possible states, gives rise to relevant outcomes? We propose that consciousness is a feature of the control system that solves this problem. Presumably, consciousness emerged during the Cambrian era in response to the recursive and model based processing required to solve the above objectives in order to engage a complex multi-agent world. This suggests consciousness is a necessary property of embodied and situated control architectures that engage the social world. For future work, we aim to test this hypothesis using multi-agent models and simulations. The idea would be to consider a population of embodied agents (robots), equipped with perception, action and learning abilities and evolving in an environment with limited resources. The question would be to study how environmental constraints can push agents to collaborate or compete by learning how to predict "self" and "other's" states. This would allow one to analyze how social prediction mechanisms can be recruited for optimizing social behavior and how social norms and ultimately consciousness can be modeled as an emergent product of cognitive adaptation in a multi-agent world.

Acknowledgements

This work has been supported by the European Research Council's CDAC project: "The Role of Consciousness in Adaptive Behavior: A Combined Empirical, Computational and Robot based Approach" (ERC-2013- ADG 341196).

References

[1] E.A. Di Paolo, Behavioral coordination, structural congruence and entrainment in a simulation of acoustically coupled agents, *Adaptive Behavior* **8**(1) (2000), 27–48.

[2] T. Froese, E.A. Di Paolo, Toward minimally social behavior: social psychology meets evolutionary robotics, *Advances in Artificial Life: Darwin Meets von Neumann* (2009), 426–433.

[3] F. Kaplan, Semiotic schemata: Selection units for linguistic cultural evolution, *Artificial life VII* (2000), 372.

[4] S. Laureys, The neural correlate of (un) awareness: lessons from the vegetative state, *Trends in cognitive sciences* **9**(12) (2005), 556–559.

[5] S. Laureys, A. M. Owen, N.D. Schiff, Brain function in coma, vegetative state, and related disorders, *The Lancet Neurology* **3**(9) (2004), 537–546.

[6] G. Maffei, D. Santos-Pata, E. Marcos, M. Sánchez-Fibla, P. Verschure, An embodied biologically constrained model of foraging: from classical and operant conditioning to adaptive real-world behavior in dac-x, *Neural Networks* **72** (2015), 88–108.

[7] C. Moulin-Frier, J. Diard, J.L. Schwartz, P. Bessière, COSMO ("Communicating about Objects using Sensory-Motor Operations"): A Bayesian modeling framework for studying speech communication and the emergence of phonological systems, *Journal of Phonetics* **53** (2015), 5–41.

[8] A.C. Slocum, D.C. Downey, R.D. Beer, Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention, *From animals to animats 6* (2000), 430–439.

[9] L. Steels, The Synthetic Modeling of Language Origins, *Evolution of Communication* **1**(1) (1997), 1–34.

[10] P.F. Verschure, C.M. Pennartz, G. Pezzulo, The why, what, where, when and how of goal-directed choice: neuronal and computational principles, *Phil. Trans. R. Soc. B* **369**(1655) (2014).

[11] P.F. Verschure, T. Voegtlin, R.J. Douglas, Environmentally mediated synergy between perception and behaviour in mobile robots, *Nature* **425**(6958) (2003), 620–624.