Attention and Consciousness in Cognitive Systems

To be presented at ESF-JSPS Frontier Science Conference Series for Young Researchers – Robotics: Experimental Cognitive Robotics

Carlos Hernández, Ignacio López, Ricardo Sanz

Abstract

AI has been providing useful tools for incorporating cognitive capabilities to artificial systems. However, there is a need for mechanisms to integrate them in an agent capable of fully unified cognitive behaviour. Consciousness and attentional mechanisms provide functionality related to integration and control of cognitive processes in biological systems. In this paper we present the approach taken in the UPM Autonomous Systems Laboratory within the ASys Project for the development of technology to construct self-aware artificial systems, which would provide the robust autonomy that the new and future technology demand.

1   Introduction

AI has been producing successful tools inspired by properties of human intelligence. These AI tools render cognitive functions present in the human mind: expert systems can manage human readable knowledge and make inference upon it, fuzzy logic allows to build computer programs that manage imprecise information, artificial neural networks go down to the basic brain structure and can learn, generalize and categorize. These tools have been deployed in experimental testbeds, mainly robotic systems, and have proved to provide them

1

with intelligent behavior at some level. They have even saved the gap between research and real commercial applications.

However, not a universally accepted intelligent system has been built so far. Not only high level cognitive capabilities such as abstract reasoning still remain as a challenge, but also those of appropriate action and behavior in the real world, *i.e.* real-time operation, generation of meaning from sensory flow, combination of reactive responses with action planning. The combination of all these aspects is a matter of interest in cognitive robotics, where we talk of agent embodiment and situatedness. While the previously commented tools from AI provide some cognitive functionality, it remains missing what would integrate them into a unfied cognitive agent capable of behaving autonomously at the same level biological systems do. Consciousness seems to have a key role in integrating the operation of biological systems. This is the reason for our interest in artificial consciousness.

2 ASys vision

The ASys Project is a long term project of our research group that is focused in the development of science and technology for autonomous systems. The aim is to develop control technology capable of providing robust autonomy at the required level.

Technical systems are quickly growing in complexity to address the rising demands of new functionality and increased performance, while increasing other non-functional requirements such as resilience and autonomy. Airplanes, cars or chemical plants, besides to electricity networks, telecommunications and other supporting facilities are some examples of this. All these systems include embedded control as a necessary component, which is nowadays mostly computer or software based. Therefore control systems are becoming extremely complex. In addition, isolated systems are becoming rare, systems are more usually intercommunicated or integrated, and so their control systems, which are very frequently distributed. However, this increases the risk of failures not only because of the increase in the number of components in these systems but because of their interaction. Improving autonomy can help solving these problems, and maintain or increase system dependability and survivability. In addition, advanced functionalities are demanded —dealing with situations with higher degrees of uncertainty, strategic decision making at run-time—, which involves the need for the systems to incorporate cognitive capabilities —learning, abstract reasoning, *etc.* —.

In ASys we address the problem of autonomy from a control perspective and taking the always useful inspiration that can be found in the biological world. We look forward to incorporate control of the control systems, that is meta-control, so as to generate technology with the ability to handle itself and provide robust autonomy. Control need to exploit models, as the single way to overcome the limitations of feedback control strategies and give the

advantages that predictive models provide in terms of anticipation and optimization.

Two of the main pillars in ASys are:

- The architecture-centric design approach, as the way to generate control technology that effectively integrates components that provide intelligent capabilities.

- The model-based approach: an ASys system will be built using models —of it, of its environment— and once in operation will be using models —of its environment, of itself— to perform its activity.

3 Model-based cognition

From our control engineering perspective, we consider cognition as the exploitation of knowledge to realize control, understood as the part of a system responsible for maintaining the behavior directed towards system objectives while satisfying certain conditions. In biological systems we have minds embodied in brains whereas in artificial systems we have control laws and control architectures nowadays mainly embodied in computers. We claim that we can equate knowledge and models and state that: *a system is said to be cognitive if it exploits models of other systems in their interaction with them*[7].

This may seem too strong a claim, since there may be simple cognitive processes that manipulate inputs without having a model of them. Think for example of the neural networks, both biological and artificial. However, even in these cases, some information abut that input can be extracted from the deep analysis of the mechanisms —in this cases neural networks—. The models can appear in a great variety of forms, and often they are implicit and very difficult to tell apart from the mechanisms exploiting them. They idea that the mind uses models of external to mind things is an old one, and also the idea that any controller, would it be biological or artificial, must use models of the controlled plant [4].

There is a great variety of models, depending on their purpose, the extension of the modelled reality, their level of detail, their implementation, whether they are explicit or implicit in the algorithms exploiting them, *etc.* . What is relevant is their usefulness for the purpose the system uses them, and it is that which gives adaptive value to cognition.

We have formulated our model-based approach of cognition in ASys into eight principles so far. The first five principles are:

1. **Model-based cognition**. A cognitive system exploits models of other systems in their interaction with them.

2. **Model isomorphism**. An embodied, situated, cognitive system is as good as its internalized models are.
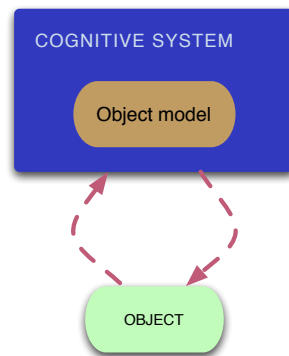
Figure 1: The cognitive relations of a system with an object are mediated by a model of the object.

3. **Anticipatory behavior**. Except in degenerate cases, maximal timely performance is achieved using predictive models.

4. **Unified cognitive action generation**. Generating action based on an unified model of task, environment and self is the way for performance maximization.

5. **Model-driven perception**. Perception is the continuous update of the integrated models used by the agent in a model-based cognitive control architecture by means of real-time sensorial information.

The last three principles refer to consciousness and attention, and will be the main subject of the rest of this paper.

4 Consciousness phenomena

Consciousness is regarded to comprise qualities such as subjectivity, self-awareness, sentience, and the ability to perceive the relationship between oneself and one's environment. It had been till the middle of the past century rejected as a subject of proper scientific study, due to its intrinsic subjective character. However, experimental results using the new brain science techniques such as PET, CT or fMRI scans have restated it as and objective phenomena to be studied with the scientific method, and is now a subject of much research in philosophy of mind, psychology, neuroscience, and cognitive science.

The problem of consciousness is usually divided in two classes [3]: the *easy problem*, which contains phenomena of consciousness that have a possible explanation in terms of computational or neural mechanisms: ability to categorise and respond to inputs, integration of information across different modalities, reportability of mental states, ability to access one's own mental states, attentional control mechanisms or behavior control; and the

hard problem. What it is referred to as the really hard problem is the question of how consciousness arise from the activity of non conscious nerve cells —or any other type of physical substrate—. Thomas Nagel in his famous paper *"What is it like to be a bat?"* [6] stressed the impossibility for science to cross the gap between brain and mind, since science can only deal with empirical and objective evidence and not with the subjective characteristic of conscious experience. This relates directly to the problem of *qualia*, or raw feels, and the source of self-awareness.

Leaving out of the discussion here the core hard problem, consciousness seems to provide for biological cognitive systems —minds— the infrastructure for the integration and cohesion of the several cognitive processes running at a time. This is the main idea underlaying the Global Workspace Model of consciousness proposed by [2]. The subject of inner experience related to awareness provides introspection, self-monitoring capabilities and somehow a meta-control if we consider the control that cognitive processes in the brain perform. Consciousness, in the aspect of "self", provides also the capability of distinguishing one-self from the rest of the world, which is crucial for attribution of agency, a necessary aspect for a cognitive system to have and intelligent behavior.

5 Consciousness in model-based cognition

To be of utility for developing technology for artificial systems, we must establish a definition of what the mechanisms of consciousness are. Instead of following the axiomatizing approach of Alexander [1], which may constrain the implementation of artificial consciousness, we have adopted a strategy consisting on incorporating a definition of the conscious phenomena in our framework of model-based cognition, which should be read as a design guideline.

We have decomposed the phenomena of consciousness in awareness (of the external world), awareness of one's own or self-awareness, and attention, which is a somehow transversal cognitive capability, but intrinsically related to consciousness.

5.1 Awareness

System awareness – A system is aware if it is continuously perceiving and generating meaning from the continuously updated models.

As stated in principle 5, system perception consists of the continuos update of the models used by the system by integrating the information coming from the input sensory flow. However this update involves not just keeping a static accurate representation of the current state of the world but modifying the dynamical models so as maintain their correspondence with the dynamics of the modeled objects.

System awareness implies a further step, which is the generation of meaning. Meaning shall be understood as a certain evaluation not only static referred to the current modeled state, but a dynamic predictive and postdictive evaluation. The meaning is generated from the assess of the current modeled state in comparison with the history of the system and considering the future potential consequences. The value system is established upon the objectives of the system. It evaluates, computes value from the model updating basing on the entailments to system objectives.
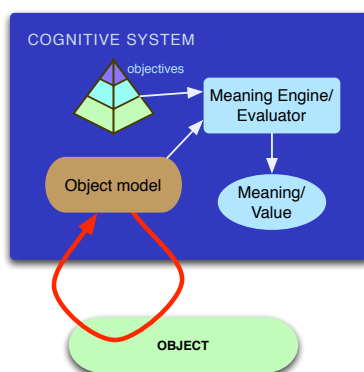

Figure 2: System awareness implies the generation of value from model-update according to system's objectives

It is remarkable that our definition of consciousness directly supports one of the generally agreed value of consciousness which is maintaining system cohesion by keeping a history of the system and interpreting current operation upon it, *i.e.* Taylor's relational perspective: appropriate past experiences are retrieved so as to give meaning to present inputs [8], and Damasio's autobiographical self [5]. It is exactly the function that results from evaluating models including both postdictive pasts, which directly refers to system's history, and predictive futures, which cannot be obtained but by applying known models – stored from previous experience – to current inputs.

5.2 Self-awareness

When the mechanism of awareness acts upon models that include the cognitive system itself self-awareness and consciousness happens:

System Self-awareness/consciousness – A system is conscious if it is continuously generating meaning from continuously updated self-models.

This consideration of consciousness as the mind/controller modeling its own operation has a biological support, *i.e.* Damasio's *second-order structures*. The possibility of the system modeling itself and generating meaning/value upon it is in the basis of meta-control. It brings the possibility to the system to modify its cognitive operation, and apply some
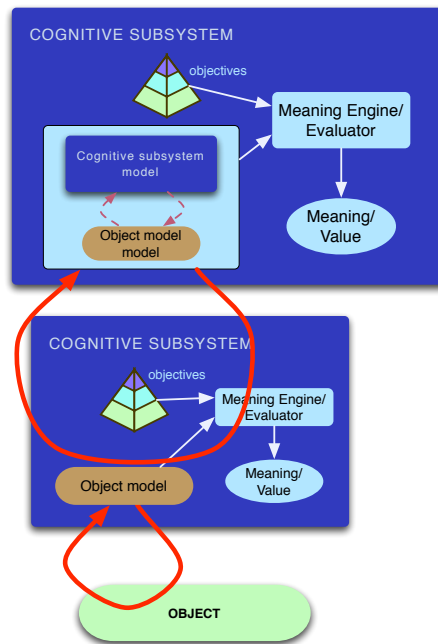
Figure 3: System self-awareness implies the modeling of the cognitive processes.

cognitive capabilities over themselves, such as learning from successful and and failure of inference processes, or planning algorithms. The models of both the cognitive system and the external reality are evaluated and used to generate behavior. The integration of the models of the self – metamodels since the modeled reality, the system, contains models – with the models of the rest of the objects is crucial for the system to have experience of it, to reach consciousness, by means of a unified evaluation.

One important big difference between being aware and being conscious is the possibility of attribution of agency thanks to the systems considering itself in the evaluation of the models.

5.3 Attention

When engineering a system there always is, no matter what kind of system nor the type of task it is intended for, a common constraint that must be taken into account in all the stages, from requirements definition to final implementation and tests passing through design. This common constraint is the limited resources we have to build up the system with, and as a consequence, the limited resources the system has. We may distinguish two classes of limited resources: limited physical resources and limited cognitive resources, not because they are different in their very core nature – in the end cognitive resources are embodied –, but because of the part of the system they support: the physicality of the system or the

cognitive part of the system.

Let's have a glimpse at each of these two limitations, starting with the limited physical resources. Cognitive systems are intended to operate in complex environments, eventually the real world, in its broader and fine detailed sense. We will take the example of a mobile robotic system. There, the real world environment potentially contains an infinite number of objects and events – rocks, trees, roads, birds, grass, buildings – and in a practically infinite regression of detail. However, the system will have in any case a limited set of sensors, that can sense only certain kind of phenomena (i.e. reflections of ultrasounds in a surface) and with a limited range —so that it can sense only a part of the environment–, with a limited spatial scope —thus covering a small portion of the phenomena present in the environment at a time— and with limited precision —hence limited level of detail to be sensed—.

It seems clear that, once build up, the system has no way to eliminate this limitation, but possessing scalability and integration properties to integrate new sensors if given. However, the system may be able to mitigate it, for example if it could direct its sensory resources to those areas in the environment of particular interest, so as to obtain more information through perception to improve its models of the environment. This is the first type of attentional mechanisms a system may have, and which relate to the ability of a system to allocate physical resources to maximise utility in the perceptive processes.

The interest of the system in a portion of the perceptive environment could be triggered by a deliberative inner process – attentional top-down mechanism – or directly by a certain pattern in the sensory input – attentional bottom-up mechanism –[9]. For example, attentional mechanisms are triggered by strong and unexpected inputs, such as a burst; or they can also be driven by inner top-down control related to a required goal, *i.e.* searching for a friend in a crowd.

We may turn now to the problem of limited cognitive resources. The limiting factor of data storage has these days became negligible in relation with other factors, since nowadays storage media provides almost unlimited space for example to store an almost unlimited quantity of models in the system without much physical space waste. However, the amount of modeling instantly instantiated, that is in the working memory at a time, is much more constrained by the amount of RAM of today's CPUs. By modelling here we are referring models quantity, the level of detail of them, and the number of deliberative processes exploiting them. A similar circumstance occurs in the human mind, whose long term knowledge appears incredible huge compared with the amount of information we are able to keep in working memory at a time. So there is need for mechanisms in the system which will decide which models and with which detail are worthy running at each instant and which deliberative processes will be exploiting them. Let's go back to our mobile robotic system example. One of the possible tasks of the robot may involve traversing a room with obstacles. Once the path planning algorithm initiated, an internal alarm could warn the system

of low battery. It could be the case that the current process could not coexist in the working memory with the process to deal with low battery at run time. Then the system would have to select between continuing with the same planning process in the working memory or removing it and giving the resources to the process dealing with low battery. So the second type of attention a system can posses is the ability of a system to allocate cognitive resources to maximise model exploitation.

We shall conclude by summing up all these ideas in the eighth principle for model-based cognition:

System attention – Attentional mechanisms allocate both physical and cognitive resources for system perceptive and modelling processes so as to maximise performance.
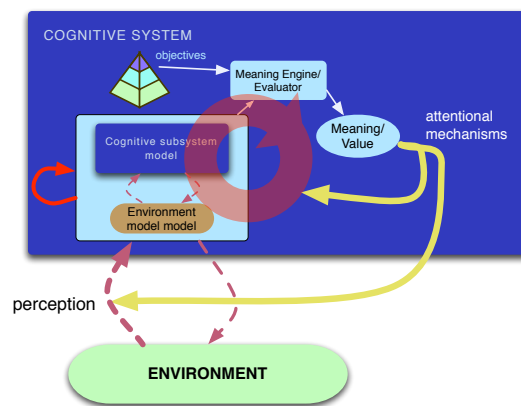

Figure 4: Attentional mechanisms allocate resources for perceptive and modeling processes.

5.4 Attention and consciousness

In the cognitive sciences as well as in common life the meanings of attention and awareness are somehow intermixed. For example we could say that 'to be aware of something you have to be paying attention to it'. There is a clear deep relation between both concepts. According to our definitions we shall establish that relation in a causality form: awareness, the generation of value in the update of the model, causes a change in the organisation of the system towards its objectives, adapting the system resources, therefore triggering the attentional mechanisms.

From the previous comment about the relation between attention and awareness it may seem that we are claiming that there is only top-down attentional mechanisms; it is not. We claim that any attentional mechanisms must be triggered by awareness – as defined here, not as commonly used – because value must be generated so as to the system shall allocate resources in a useful way and not randomly. The difference lies in that in bottom-up attention the new evaluation is due to the entering input, whereas in the top-down mechanism it

is a result of internal cognitive operation not related to the current sensory input.

Attention is transversal to consciousness because attentional mechanisms also decide which cognitive processes enter consciousness, that is, become modeled and evaluated. It is clear the modeling and evaluation of cognitive processes consumes and extra, and probably large, quantity of the system cognitive resources. In the human mind, one agreed characteristic of conscious processes is the limited number of them at a time, almost running only in a serial way, compared with the several unconscious processes that run in parallel [2].

6 Conclusions

The evolutional value of consciousness in biological systems is undeniable. The role it plays in the human mind seems to be that of integration and meta-control over the rest of cognitive processes. We propose here an approach to address the construction of intelligent systems based on a viewpoint of cognition as model based, and explain consciousness phenomena, that we have decomposed into awareness, self-awareness and attention, in that framework. In the context of the ASys Project we are currently working on a formalization of these ideas in the construction of an ontology for cognitive autonomous systems and the development of SysML models which eventually will consolidate into a blueprint for a self-aware cognitive architecture.

References

[1] Igor Aleksander and Barry Dunmall. Axioms and tests for the presence of minimal consciousness in agents. *Journal of Consciousness Studies*, 10(4-5):7–18, 2003.

[2] Bernard J. Baars. *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press, oct 2001.

[3] David J. Chalmers. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, 1996.

[4] Roger C. Conant and W. Ross Ashby. Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 1(2):89–97, 1970.

[5] Antonio R. Damasio. *The Feeling of What Happens*. Vintage, 2000.

[6] Thomas Nagel. What is it like to be a bat? *The Philosophical Review*, October 1974.

[7] Ricardo Sanz, Ignacio López, Manuel Rodríguez, and Carlos Hernández. Principles for consciousness in integrated cognitive control. *Neural Networks*, 20(9):938–946, November 2007.

[8] John G. Taylor. *The Race for Consciousness*. The MIT Press, 1999.

[9] John G. Taylor. Paying attention to consciousness. *TRENDS in Cognitive Science*, 6(5):206–210, May 2002.