THE AUTONOMOUS SYSTEMS LABORATORY
UNIVERSIDAD POLITÉCNICA DE MADRID

Architectural Mechanisms for Consciousness in RCS

Carlos Hernández

ASLab A-2007-015 v 1.0 Draft
October 8, 2007

Contents

Abstract: In this paper we study the suitability of a cognitive architecture, one which is exploited at present in industrial and real-world applications, for addressing the core issues of consciousness. We have chosen RCS since it is on of the most popular cognitive architectures, which has demonstrated its benefits in several real world use-cases from different domains, and because it incorporates many characteristics of other classical architectures more human brain modeling oriented. The theoretical approximation to consciousness is the one proposed by our group, condensated in seven general principles for systems to be cognitive and eventually conscious

Keywords: cognitive architecture, consciousness, RCS

1 Introduction

During the last three decades research on both consciousness and cognitive architectures for has grown rapidly. From the engineering point of view, cognitive architectures have proveed useful in the construction of controllers for complex systems that have to deal with uncertainty. The best example of this maybe the RCS architecture developped at the National Institute of Standards and Technology of the U.S. Government, which has been succesfully applied in unmanned vehicles, a manufacturing system, etc. On the other hand, consciosness has raised the interest of the scientific community as never before, since the twentyeth century has provided cognitive researchers with tools, i.e. electroencephalogram, functional magnetic resonance imaging, and the field isincreasingly achieving important results in understanding this phenomenon. Since biological system prove that consciousness is a evolutionary adaptation to achieve greater levels of self-preservation, the next step seems to be to implement it on our artificial systems so as to improve them in terms of better performing, dependability or survivavility.

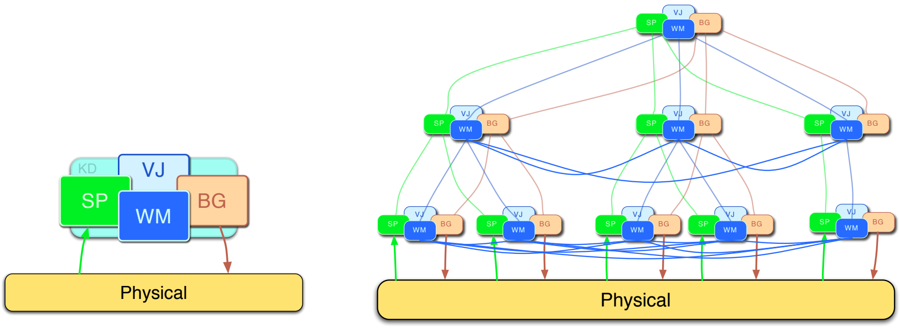
In this paper we will try to analise the suitability of a cognitive architecture, which is exploited at present in industrial and real-world applications, for addressing the core issues relating consciousness capabilities. We have chosen RCS since it is on of the most popular cognitive architectures, which has demonstrated its benefits in several real world use-cases from different domains, and because it incorporates many characteristics of other classical architectures more human brain modeling oriented. The theoretical approximation to consciousness will be that recently proposed by our group in [5].

We are here analyzing RCS not as the complete reference model architecture which covers every phase of the engineering cycle of intelligent complex control systems, from the conceptual framework and reference model for systems components and structure, to the methods and tools for eventually implementing it, but just as the first one: a conceptual framework and reference model.

2 Overview of RCS

RCS (Real-time Control System) [1] is a cognitive architecture, or a reference model architecture, designed to fulfill complex requirements in a wide variety of domains, and enabling any level of intelligent behaviour, ranging from simple controllers to human reasoning. It was developped at NIST, initially inspired by a model of the cerebellum, but nowdays it has evolved into a real-time control architecture for complex technical systems.

RCS consists of a hierarchical, both in layers and resolution scope, structure of nodes, each one containing the elements of intelligence, distributed in the following modules:



(a) RCS node (b) Example of a RCS organizational hierarchy

Figure 1: RCS reference model architecture: each node (a) in the architecture represents an operational unit in an organizational hierarchy (b)

Sensory Processing (SP) modules. These modules receive observations from the sensors and compare them with expectations generated by the WM module, thus providing perception. The SP module can interact with WM and VJ modules to assign value to perceived entities, events and situations.

Behaviour Generation (BG) modules. The function of these modules is to decompose task commands into subtask commands. So for the whole system the activity of these modules results in a spatial and temporal decomposition of the tasks of the system.

World Model (WM) modules. The world model is the system's internal representation of the external world. WM modules maintain the KD updated and consistent. They generate predictions of expected sensory input used by SP modules, provide information to planners and executors of BG modules and simulate the hypothesis of BG planners.

Value Judgement (VJ). The value judgement module determines what is good and bad, computing risks, costs and benefits both of observed situations and expected results of hypothesized plans.

Knowledge Database (KD). The Knowledge Databased, stores information about space, time, entities, events, states of the system's environment and also about the system itself, its internal state, i.e. values assigned to objectives, parameters of the algorithms, etc.

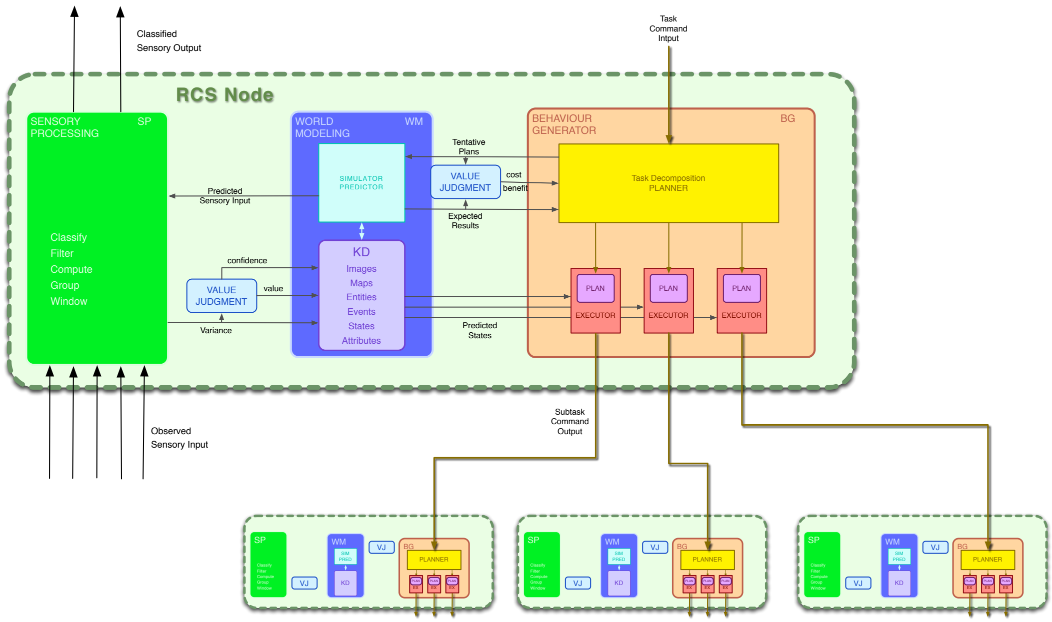


Figure 2: Functional structure of a typical RCS node

3 General Principles for Cognitive Systems

Modern technical systems are rapidly growing in complexity, and mankind is becoming more dependent on them. Classical control architectures are not well suited to deal with this: they are difficult to apply and usually the complexity of the systems overwhelms the capabilities of these architectures, specially referring to safety and dependability aspects. There is a strong trend nowadays towards a bio-inspired approximation to these problems. Biological systems have acquired through evolution appropriate mechanisms to address issues about fault tolerance, resilience and self-preservation; some of them appear to be quite interesting for its application to artificial systems. Since nowadays it is commonly accepted between the scientific community that cognitive capabilities, up to consciousness, are an example of this evolved mechanisms, they seem particularly interesting for their application in control systems.

In our group we are currently studying the underlying theoretical background and the architectural mechanisms that would enable the implementation of cognitive capabilities and ultimately consciousness in artificial systems. We are starting to develop a theoretical framework to provide a basis for this. In [5] we propose some principles.

General Principles for Cognitive Systems [5]:

Principle 1: Model-based cognition A system is said to be cognitive if it exploits models of other systems in its interaction with them.

Principle 2: Model isomorphism An embodied, situated, cognitive system is as good as its internalized models are.

Principle 3: Anticipatory behavior Except in degenerate cases, maximal timely performance can only be achieved using predictive models.

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

Principle 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.

Principle 6: System awareness A system is aware if it is continuously perceiving and generating meaning from the continuously updated models.

Principle 7:System of self-awareness/consciousness A system is conscious if it is continuously generating meanings from continuously updated self-models in a model based cognitive control architecture.

4 Mapping RCS-Principles

4.1 Principle 1: Model-based cognition

In RCS the World Model module of each node provides explicit model-based cognition, since it enables exploitation of models of both the system and its environment. The World Module works together with a Knowledge Database where it stores modelling information, so RCS is fully and strictly compliant with Principle 1, equating knowledge with models.

4.2 Principle 2: Model isomorphism

Reasoning elements of a cognitive architecture use models and knowledge to acquire information in an efficient and useful way, to elaborate plans, simulate and evaluate them so as to choose the best one and elaborate the most adequate actions. In many occasions, the ability to succeed in generating adequate actions depends more on the available knowledge than in the planning algorithms. Hence the more precise, usable and complete are the models, the more appropriate actions the system selects and the better it performs. So it looks like a cognitive system is as good as its internalized models are [5].

We can categorize the mechanisms for model construction in three categories: built-ins, learned and cultural.

We refer to built-in models in the sense of Conant and Asby [2] being the knowledge embedded in algorithms and its parameters. Built-in models would depend on the current implementation of the RCS architecture, but they will always be present, since these implicit models will be embedded in sensory, control and action algorithms. However, preconfigured models are also embedded in the way the nodes are connected in the implementation of the RCS architecture. Besides, there can always be built-in explicit models too.

Learning is not implemented in RCS architecture, but there are some implementations of RCS controllers in which learning has been implemented. This is the case of [3], learning was embedded within the elements of each RCS node.

In reference to the mechanisms for model acquisition categorizeded as cultural in [5], it would be possible to download new models in RCS knowledge database provided a fully ontological coherence with existent models. However, there are still open questions such as how RCS could integrate not ontologically coherent models, that to say, how to enable machanisms for understanding and comprehension.

4.3 Principle 3: Anticipatory behavior

One of the four functions of RCS module WM is answering "What if?" questions demanded by the planners of the BG modules. For performing this task, WM simulates the model with the inputs proposed by the BG modules and obtains the expected results, which then are evaluated by VJ modules, and that evaluation is sent back to the BG planner.Tehrefore, in RCS models are simulation (prediction) oriented.

WM modules also generate predictions of expected sensory input, thus enabling part of the process that generate perception in RCS SP modules by directing the process to referents [4] –key entities in the context of system-environment-task [6].

4.4 Principle 4: Unified cognitive action generation

One of the current problems in nowadays complex control systems, which are usually distributed and involve different resolutions in space, time and task, is maintaining system cohesion and model coherent across such a wide range of scopes.

RCS hierarchical structure of nodes provides adequate organization through different levels of spatial and time scope, together with a dynamic command tree that can vary depending on the current task. The proper node's structure is what enables coherence between layers: SP module of nodes provide adequately classified sensory output for the input of SP modules of immediate superior nodes, task command output from executors in the BG modules of superior nodes becomes task command input for the Task Decomposition Planner of the BG modules in the inferior layer.

4.5 Principle 5: Model-driven perception

In RCS each node has a specific component for perceiving, the Sensory Processing Module. Since the authors of RCS seems to agree with [5] in relating perception with the maintenance and update of the systems models with the external world, it is directly compliant with our P5.

In reference to the two other flows of information not coming from the environment of the system but from its inside, [1] includes propioception in the study of the sensory process, and since there is theoretically no other difference between the environment and the body of the system that the label we apply to them, RCS nodes can be dedicated to propioception as well as exteroception.

4.6 Principle 6: System awareness

In RCS awareness is supported in the sense of generating meaning from perceptions. The value judgment module process the perceptions coming from SP to KD modules and assigns them value in terms of confidence, usefulness, coherence etc, in order to integrate them in the knowledge database and update the models. Besides, an attention mechanism is formed by BG, WM and SP modules. BG modules request information needed for the current task from the SP modules, thus focusing the attention of SP and WM modules on the elements important to achieve the current goals.

An issue related to and some times almost merged with awareness is that of attention. It is not explicitly covered in the proposed seven principles for cognitive systems, and it may should be. In RCS the mechanism for attention is directed by BG modules. These modules request the information needed for their operation to the SP modules, directing SP and WM modules to direct their processing towards the elements needed for the current task. BG requests causes SP modules to filter the sensory data with the appropriate masks and filters to select the relevant incoming information. The request by BG modules also causes the WM to select which worl model to use for prediction, as well as which prediction algorithm to apply.

4.7 Principle 7: System of self-awareness/consciousness

In RCS self-modeling is not considered so there is not support for Principle 7, which involves much more. However, it should be studied if RCS could implement it. Here we propose just two different ideas of how to provide the RCS architecture with self-models.

The first option is to add a new element in each node responsible the self-model of the node. This would involve also new connections between nodes in the same layers and in the vertical pathways of the hierarchy, so as the new elements could share information to have a global self-model.

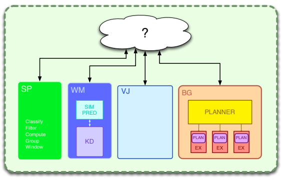


Figure 3: Self-modeling with a new element in the RCS node

The other option is to have another hierarchical organization of meta-nodes, whose mapping with the traditional hierarchy of nodes would not necessarily have to be one to one. This mapping is something to be studied. The new organization of meta-nodes would provide a self-model of the architecture.

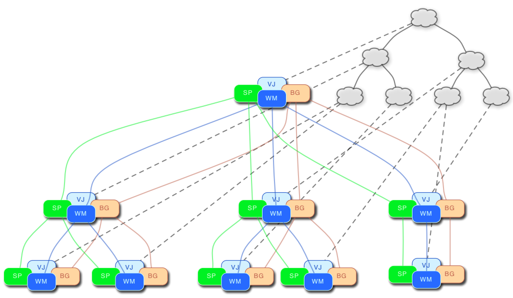


Figure 4: Self-modeling with a hierarchy of meta-nodes

5 Conclusions

In this paper we have discussed the RCS Architecture compliance with the principles proposed in [5] for cognition and ultimately consciousness in artificial systems. It has been showed that RCS actually supports the lower principles, some completely and others only to some extent, while Principle 6 seems to be some how partially addressed and Principle 7 is actually not applicable.

However, RCS flexible structure and power to scale (it can be used for the simplest controller device to the largest complex control systems), seems to provide a useful framework in which to try to implement the higher principles for cognition/consciousness. Two lines of action in how to provide the architecture with a self-model have been suggested.

References

[1] James S. Albus. Outline for a theory of intelligence. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, 21(3):473–509, 1991.

[2] 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.

[3] T. Chang T. Hong W. Shacleford J. Albus, R. Bostelman and M. Shneier. Learning in a hierarchical control system: 4d/rcs in the darpa lagr program. *Journal of Field Robotics*, 23(Special Issue on Learning in Unstructured Environments):975–1003, 2006.

[4] Ignacio López. *A Foundation for Perception in Autonomous Systems*. PhD thesis, Universidad Politécnica de Madrid, 2007.

[5] Manuel Rodríguez Carlos Hernández Ricardo Sanz, Ignacio López. Principles for consciousness in integrative cognitive control. *Neural Networks*, 2007.

[6] Ricardo Sanz, Fernando Matía, and Santos Galán. Fridges, elephants and the meaning of autonomy and intelligence. In *IEEE International Symposium on Intelligent Control, ISIC'2000*, Patras, Greece, 2000.