A Cognitive Neuroscience-inspired Codelet-based Cognitive Architecture for the Control of Artificial Creatures with Incremental Levels of Machine Consciousness

Klaus Raizer, André L. O. Paraense and Ricardo R. Gudwin[1]

Abstract. The advantages given by machine consciousness to the control of software agents were reported to be very appealing. The main goal of this work is to develop artificial creatures, controlled by cognitive architectures, with different levels of machine consciousness. To fulfil this goal, we propose the application of cognitive neuroscience concepts to incrementally develop a cognitive architecture following the evolutionary steps taken by the animal brain. The triune brain theory proposed by MacLean and also Arrabale's ConsScale will serve as roadmaps to achieve each developmental stage, while iCub - a humanoid robot and its simulator - will serve as a platform for the experiments. A completely codelet-based system "Core" has been implemented, serving the whole architecture.

1 INTRODUCTION

1.1 Motivation

In this work, we are particularly interested in studying the cognitive architectures which were proposed to deal with the issue of consciousness [10, 24, 37]. Our main goal is to develop artificial creatures with different levels of machine consciousness, controlled by such architectures. To fulfil this goal, we propose the application of cognitive neuroscience concepts to incrementally develop a cognitive architecture following the evolutionary steps taken by the animal brain.

Looking for inspiration in nature has been a successful way of discovering new solutions for problems in the fields of control, optimization, classification and artificial intelligence. Machine learning techniques such as genetic algorithms, ant colony optimization and neural networks are some examples of the remarkable muse nature can be [27, 25, 36, 16].

The advantages given by machine consciousness have been reported to be very appealing [20, 10, 9].

Nevertheless, the cognitive architectures which are able to benefit from it are not so many, and still under heavy development. So, the motivation to propose and implement yet another cognitive architecture, when there are so many of them already available, lies in the need for an architecture coherent with our hypothesis of a conscious codelet-based artificial mind, able to implement the animal brain in its different evolutionary steps, and in the search for the sufficient feature set in the architecture for each of those steps that matches the results of natural selection along history.

1 School of Electrical and Computer Engineering (FEEC), University of Campinas (UNICAMP), Brazil, email: gudwin@dca.fee.unicamp.br

1.2 Statement and Background of Research

An artificial creature is an autonomous agent, a system embedded in an environment, sensing and acting on it, over time, in pursuit of its own agenda [21]. It can be controlled by a cognitive architecture, which includes aspects of the creature such as memory and functional processes [29], providing a framework to support mechanisms for perception, action, adaptation and motivation [39].

Cognitive architectures are control systems architectures inspired by scientific theories developed to explain cognition in animals and in men. These architectures are typically organized in layers [2, 13], with each layer representing a different level of control and specialized modules [15]. The most famous general cognitive architectures are SOAR [28] and ACT-R [1]. More recently, many specialized cognitive architectures have been proposed, emphasizing different aspects of cognition, e.g. emotions, attention, memory, consciousness and language. Each one has advantages and shortcomings, when compared to each other.

Looking for inspiration in cognitive neuroscience, current research on artificial creatures has focused on the implementation of machine consciousness, with one of its major functions being to recruit relevant resources for solving new or difficult problems [33]. Recently, the study of machine consciousness in cognitive architectures applied to artificial creatures has particularly been exploited [4, 15, 33, 8].

Even though there is not a consensus on what exactly is meant by "machine consciousness", as different authors indeed have different perspectives on what they mean by "consciousness", in a previous work from our group [35], we investigated one interesting proposal, called the Baars-Franklin architecture. During this investigation, we evaluated the possible benefits that such "consciousness" technology, when applied to the control of autonomous agents, could bring to such systems. Our main findings were that the main benefits brought by consciousness (as defined in Baars-Franklin architecture), are:

- Executive Summary of Perception
- The Possibility of Behaviour Automatization

These two advantages arose from the main perspective on what is consciousness after all in the Baars-Franklin architecture, following Dennett [14]. According to this perspective, consciousness is the emergence of a serial stream on top of a parallel set of interacting devices. In the Baars-Franklin architectures, such devices are called “codelets” [33] (following Hofstadter [26]), which are small pieces of code - similar to Ornstein’s small minds, Minsky’s agents, Edelman’s neuronal groups and Jackson’s demons [19] - specialized in performing simple tasks [33]. This serial stream evidences the most important information flowing in the parallel system at each time instant, creating what we called the executive summary of perception, a special kind of attention mechanism. This serial stream is then broadcast to all codelets in the system, allowing decision-making on both up-to-date input data and filtered relevant information at each time instant. The possibility of automatizing behaviour is also an emergent offspring of this serial stream.

Unconscious behaviour is usually automatic reactive behaviour, performed in parallel by the system codelets. The serial stream can be used then to learn such automatic behaviour, by performing a deliberative one, which is further automatized, giving rise to future automatic behaviours. With this, conscious systems do have an interesting hybrid reactive-deliberative kind of learning, in which new capabilities can be acquired, enhancing the overall behaviour of the system.

Cognitive architectures tend to model functions performed by structures of the animal brain. These structures have, however, changed over millions of years of evolution. One model used to explain this process was the triune brain concept proposed by MacLean [31, 30], which states that the brain developed into a three-layered organ: reptilian brain, paleomammalian brain and neomammalian brain, as can be seen in Figure 1.

Figure 1. The triune brain model as proposed by MacLean. Source: MacLean 1990 [31]

The reptilian brain is composed by the oldest structures that dominate in the brains of snakes and lizards, with a major role on fixed, instinctual behaviours and control for survival. The paleomammalian brain is layered over the reptilian brain, with a major role in emotions (emotional valence and salience) and is better at learning from experience. Finally, the most recent layer is the neomammalian brain, which is the home of complex cognition, deliberative planning, social abilities and language. However controversial this separation in three distinct layers might be today, it remains a helpful way to think about the mammalian brain [11], especially in a computational sense.

MacLean himself points out to the fact that, despite their capacity for operating independently to a certain level, the triune brain is not just a consecutive layering of these three neural structures but actually an integration between the information they share and produce, so “the whole is greater than the sum of its parts". There is often no consensus about which brain structures compose each layer, and the functions each particular structure performs are not easy to discriminate – due, among other reasons, to how massively interconnected most parts of the brain are. As Fuster exemplifies in his book [22], attributing a particular movement control to a given area, or speech only to Broca’s area, ignores the fact that both functions depend on many other neural structures. With that in mind, this work aims at avoiding direct mappings [34] between neural structures and its functions, and focuses instead on a framework developed over a large body of brain and psychological evidence. The proposed architecture is based on Baars and Gage’s functional framework, as seen in Figure 2, to develop a codelet-based, biologically plausible cognitive architecture.

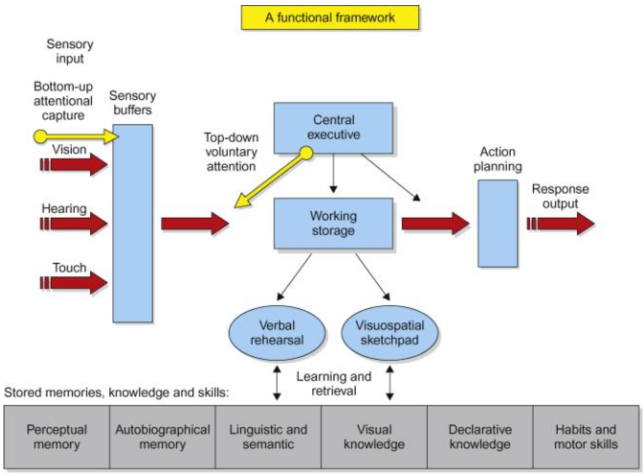

Figure 2. Baars and Gage’s functional framework. Source: Baars and Gage 2010 [11], with kind permission.

In the framework from Figure 2 each sense has a brief storage ability, also called sensory buffer. Elements in the sensory buffer are modified by bottom-up selective attention, which happens in vision for instance when confronting particular patterns, or in hearing when there is a loud noise. There is a top-down component to selective attention coming from the central executive, which allows voluntary attention to happen. The central executive is part of working memory, as defined by Baddeley [12], and it exerts supervisory control over all voluntary activities. Working storage is a short term and dynamic storage mechanism. It is composed of active populations of neurons which can consolidate into long term memories and is believed to have very limited capacity. The verbal rehearsal and the visuospatial sketchpad involve mental capacities used to remember things like new words (in the case of inner speech), faces, or spatial information (in the case of visuospatial sketchpad). They are both linked to long-term memory by a learning and retrieval mechanism. Long-term memory is represented by the gray boxes on the bottom and is comprised by a number of different types of memory, each with its own functions and characteristics. At the right side of the diagram, there is action planning, which can have both conscious and unconscious components, and finally an output response that closes the perception-action cycle.

2 METHODOLOGY

The work is following a path similar to the evolutionary steps taken by the animal brain as stated by MacLean [31]. The hypothesis is that such an approach should guarantee a grounded intelligent system at each developmental phase, while biasing the system towards high-level animal intelligence.[2]

The initial development of this architecture emphasizes on behavioral results. A low level approach to evaluate its consciousness levels - such as information integration, as seen in the works of Tononi [38] - might be implemented in future works. ConsScale, a biologically inspired scale designed to evaluate the presence of cognitive functions associated with consciousness [5, 3, 6], will be used to assess the different levels of control implemented within this cognitive architecture. There are two ways of using ConsScale: the Standard Evaluation Process (SEP) and the Simplified Rating Process (SRP). SEP is used to evaluate existing implemented agents, providing an accurate measure of the agent's cognitive level. SRP, on the other hand, is used as an approximation of the potential level of an existing or still to be implemented model.

In this initial work, the SRP for each level of development - reptilian, paleomammalian, neomammalian and *Homo sapiens* - is calculated, and a roadmap of behaviour profiles (BP) is proposed in order to reach an accurate measure of a specific domain. The ConsScale Quantitative Score (CQS), spanning from 0 to 1000 in an exponential fashion, is also calculated for each stage, providing a numerical value indicating their cognitive power.

The platform used as an specific domain for the initial experiments is the iCub humanoid robot simulator [32], but the architecture is built so it can be applied to different platforms and applications. This platform was chosen because it provides a "Human-Like" architectural level, as described by Arrabales [5, 6].

2.1 Conscious Codelet-Based Cognitive Architecture

Figure 3 shows an UML class diagram describing the architecture's modules.

The relationship between modules and their features according to the functional framework of Figure 2 is better understood by following a full cognitive cycle, considering the neomammalian brain:

1. Sensors (BodyInterface) get information from the World (iCub) and send it to the Sensory Buffer (BodyInterface);
2. Bottom-up Attention (Perception) acts on Sensory Buffer (BodyInterface), giving rise to objects from raw sensory inputs;

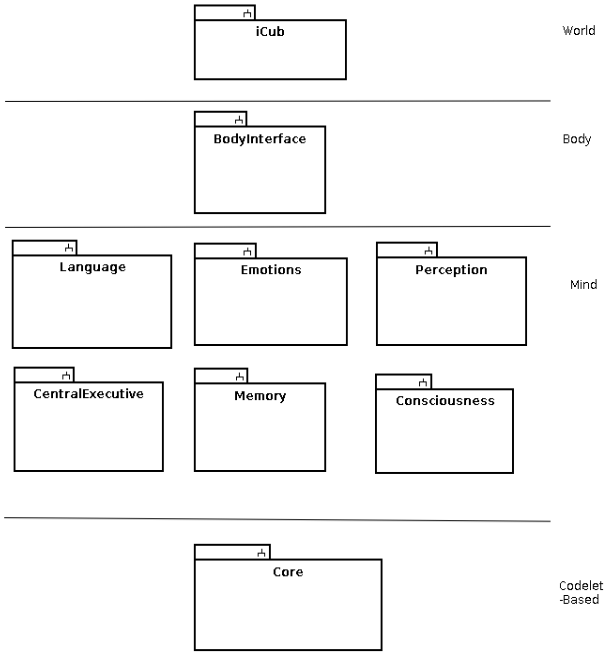

Figure 3. Architecture's layers and subsystems

3. Bottom-up Attention (Language) acts on Sensory Buffer (Perception), giving rise to symbols from objects;
4. Top-down Attention (Central Executive) acts on Sensory Buffer's objects (Perception) and symbols (Language);
5. Top-down Attention (Central Executive) brings information into Working Storage (Memory);
6. Learning and Retrieval Mechanism (Memory) consolidates to Stored Memory (Memory) and brings into Working Storage (Memory) long-term information;
7. Spotlight Controller (Consciousness) acts on Working Storage (Memory), defining Spotlight (Consciousness) content;
8. Action Selection (Emotions) uses information under Spotlight (Consciousness) to select a plan, composed by a list of behaviours;
9. Action Selection (Central Executive) uses information under Spotlight (Consciousness) to select a plan, composed by a list of behaviours;
10. Behaviour sequence is sent to Action Buffer (BodyInterface);
11. Actuators (BodyInterface) act on the World (iCub) based on Action Buffer (BodyInterface).

Figure 4 shows a diagram depicting how the concepts of the codelet-based Core subsystem have been implemented. Following this picture, Memory Objects are single units of data in memory, which have a type (T) and some information (I). The

[2] It is important to acknowledge, however, that the path taken by mammals in evolution, especially the case of *Homo sapiens*, is not the only one that led to high-level cognition. Examples of high level cognitive behaviour, and potentially conscious capabilities, have been observed in modern birds and cephalopods [17].

Raw Memory contains all Memory Objects in the system. It can be logically divided in different kinds of memory, such as the stored memories from Figure 2. Codelets are devices which are composed by small pieces of code, specialized in performing simple tasks (proc), a list of input Memory Objects (In), the ones that are read, a list of output Memory Objects (Out), the ones that are written, a list input broadcasted Memory Objects (B), the ones that were broadcasted by consciousness mechanisms, and an activation level (A). Coalitions are groups of Codelets which are gathered in order to perform a task by summing up their abilities. Two or more Codelets share a Coalition when they write in and/or read from the same set of Memory Objects. The Coderack (following Hofstadter [26]) is the pool of all active Codelets in the system.

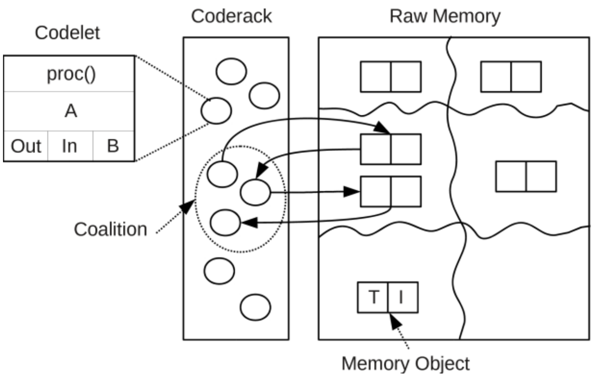

Figure 4. Core's concepts

The other modules are the subject of future work and will be implemented according to the planned steps of the architecture, as further explained in Section 2.2.

2.2 Evolutionary Steps Taken by the Animal Brain

2.2.1 Reptilian

According to MacLean, the protoreptilian formation is composed by a group of ganglionic structures located at the base of the forebrain of reptiles, birds and mammals. It is also known as the striatal complex and brainstem [11] or, as he puts it, the R-Complex. It was traditionally thought to be the motor apparatus under control of motor cortex and reveals a number of fixed behaviours (25 special forms of behaviours and 6 forms of what he calls "interoperative" behaviours [31]), involved in the regulation of the animal's daily routines. In this sense, the R-Complex is composed by pre-programmed regulators for homeostasis and survival, lacking an advanced learning mechanism as the one seen latter in evolution. In the one hand, it excels at performing sensory categorization, such as identifying a particular smell as being harmful or not, and then generating reflexive messages about what to do, like running or biting [23]. On the other hand, it is constrained by its daily master routine, destined to perform a limited number of behaviours. Reptiles and lizards however need to "learn" their territory in order to know which hole to escape into in case of a predator or just to find their way home. This, and other examples of very basic "memory/learning" mechanisms, is what MacLean called protomentation, which are "rudimentary mental processes that underlie a meaningful sequential expression of prototypical patterns of behaviour" [31].

Based on the main characteristics of a creature with an R-Complex, a number of skills from ConsScale are proposed for this level of development:

- CS2;1 : Fixed reactive responses.
- BP2;1 : Basic reflexes such as blinking and contraction of limbs as responses to pain.
- CS3;3-5 : Selection of relevant sensory/memory/motor information.
- BP3;3-5 : The robot reacts to predefined sensory inputs and stores basic information in fixed memory.
- CS3;6 : Evaluation (positive or negative) of selected objects or events.
- BP3;6 : The robot evaluates sensory input comparing it with predefined patterns to evaluate sensory inputs as being good or bad for it.
- CS4;2 : Directed behaviour toward specific targets like following or escape.
- BP4;2 : The robot selects grabbing action towards regions that seem good for it.

According to ConsScale, this selection of skills constitutes a level 2 (Reactive) agent, with a CQS of 0.21 in a scale from 0 to 1000. This set of skills suggest the need for a Body Interface module, responsible for dealing with all somato-sensory data exchange - sensory input and response output from Figure 2 - between agent and environment. This module also holds a fixed number of reactive responses and autonomic behaviours that have as a primary concern the survival and self-preservation of the agent. Perception at this level is very basic, with low resolution pattern recognition and a bottom-up attention mechanism that depends on the agent's nature and objectives. The output functions are organized in the Central Executive module, which at this level of development is responsible for action selection with a repertory of fixed predefined behaviours. The agent at this level lacks a general Memory System, counting only on predefined memory slots for performing specific tasks.

There is a great debate on whether creatures other than humans do or do not possess consciousness as we experience it. One major problem faced by such a debate is the lack of an accurate definition of what consciousness is and what is needed for its emergence. In this work, machine consciousness is the implementation of Global Workspace (GW) theory [7] as a means to achieve primary consciousness, in which percepts are united into episodic scenes [18]. In this sense, it is assumed here that a protoreptilian brain lacks the reentrant interactions in the thalamocortical system needed to sustain consciousness.

2.2.2 Paleomammalian

The next evolutionary step in the development of the vertebrate brain is the paleomammalian formation. With this new set of neuronal structures - some notable examples being the amygdala, hippocampus and hypothalamus - also known as the limbic system, animals became able to experience emotions [11], which are essentially the capacity of turning up or down the "volume" of drives that guide behaviour for survival. This emotional "skill" greatly affects and communicates with the aforementioned autonomic system, provoking marked physiological changes within the organism. Learning and memory have also shown remarkable improvement. An animal with a limbic system mounted on top of its R-Complex was able to discriminate good things from bad ones also by looking into its past memories [23]. MacLean emphasized that this part of the mammalian brain is responsible for a number of behaviours and characteristics that were absent in ancient reptiles such as nursing, audio-vocal communication for maternal contact and play [31].

The ConsScale skills added at this level are listed as follows:

CS3;1 : Autonomous acquisition of new adaptive reactive responses.

BP3;1 : Learns to “eat” certain kinds of “food” and reject others.

CS3;2 : Usage of proprioceptive sensing for embodied adaptive responses.

BP3;2 : Looks for “food” when hunger state reaches a certain level. Plays to get happier.

CS3;7 : Selection of what needs to be stored in memory.

BP3;7 : Emotional valence influences selection of what is relevant to be stored in memory.

CS4;1 : Trial and error learning. Re-evaluation of selected objects or events.

BP4;1 : The robot learns what is good (good food) or bad (rotten food) for him by trial and error.

CS4;3 : Evaluation of the performance in the achievement of a single goal.

BP4;3 : Evaluates how successful it is in pursuing a single goal, such as looking for “food” and uses this information to get better at it.

CS4;5 : Ability to build depictive representations of percepts for each available sensory modality.

BP4;5 : The robot can discern particular objects and some of its properties and calculate their relative positions.

CS5;1 : Ability to move back and forth between multiple tasks.

BP5;1 : An interrupted behaviour is resumed later if still relevant. For example: playing with certain objects in the environment can be resumed after stopping this behaviour to satiate hunger.

CS5;4 : Autonomous reinforcement learning (emotional learning).

BP5;4 : The robot calculates a “reward" based on how good it is at a task and improves its performance. It might throw a ball at a given target and get better at it by practicing.

CS5;2 : Seeking of multiple goals

BP5;2 : Having more than one goal, such as satisfying hunger and play, it uses CS5;1 to alternate between them.

CS5;3 : Evaluation of the performance in the achievement of multiple goals.

BP5;3 : The robot evaluates its own performance at pursuing multiple goals, and alternating among them, instead of pursuing only one.

CS5;6 : Ability to generate selected mental content with grounded meaning integrating different modalities into differentiated explicit percepts.

BP5;6 : The contents of the conscious broadcast, defined by the consciousness module, constitute mental content with grounded meaning and it is composed by an integration of percepts from different modalities.

CS6;1 : Self-status assessment (background emotions).

BP6;1 : Evaluates its own inner physical and emotional state and has its global behaviour influenced by it.

CS6;2 : Background emotions cause effects in agent’s body.

BP6;2 : The emotional state is reflected into the robot’s body (happy or sad faces) through its autonomic functions.

CS6;3 : Representation of the effect of emotions in organism and planning (feelings).

BP6;3 : Together with BP6;1, if the robot is high on health and hungry it may go look for food but if low on health and hungry it might hide at home.

This set of skills appears at the ConsScale as a level 3 (adaptive) agent, with a CQS of 7.21 in a scale from 0 to 1000. Even having a number of higher skills, such as CS6;1(Self-status assessment) for instance, it lacks some dependencies such as CS4;4 (Basic planning capability) that would allow it to attain a higher score.

There is an evolution in perception at this stage so the agent is able to perform higher-level pattern recognition, and discern particular objects in the environment. The central executive performs top-down attention over percepts, providing full attention selection capability. The memory module becomes generic, in the sense that memory objects are produced, stored and retrieved by means of a learning/retrieval mechanism. Those memories and percepts are marked with emotional content, influencing the aforementioned learning/retrieval mechanism.

At this point it is assumed that the reentrant interactions between parts of the thalamocortical system mediating perceptual categorization and those mediating memory have evolved in a way that allows the emergence of primary consciousness. The consciousness module implements GW theory, producing an attentional spotlight that broadcasts its contents to all the system. However, the creature still lacks the capacity to report its conscious stream, an ability human beings possess and which is used in our case to verify the existence of consciousness as we perceive it.

2.2.3 Neomammalian

The neomammalian formation is the latest addition to the vertebrate animal brain. Its distinguished structure is the neocortex which is composed of many layers, with a smooth surface in small mammals and deeply grooved in larger ones. The neocortex is highly oriented toward the external world [31]. With it, animals are capable not only to understand their senses but also to develop a symbolic representation of those senses and inner representations. Being on top of the limbic system, it is also capable of developing feelings about these symbols and abstract ideas [23]. Its most distinctive role, however, lies in what is called executive functions, which consist, among other things, on the ability to organize sequences of actions towards a given goal.

As the vertebrate brain evolved, the organism's actions became more based on its memories and prior experiences than on reflexive responses to the environment based on its needs (as can be seen in the transition from protoreptilian to paleomammalian). These actions also became more deliberate and voluntary [22], especially in the transition from the paleomammalian to the neomammalian brain. With this evolution, important parts of the neocortex such as the prefrontal cortex show significant growth in proportion to more ancient brain structures, with a maximum size achieved only in the human primate [22].

The ConsScale skills at the neomammalian level are as follows:

- CS4;4 : Basic planning capability: calculation of next n sequential actions.
- BP4;4 : Plans a sequence of actions to attain a goal. Example: playing for learning wastes energy, so it plans a break time to replenish it.
- CS5;5 : Advanced planning capability considering all active goals.
- BP5;5 : Takes into account information from CS5;3 to improve seeking multiple goals. Such as reducing transition time between behaviours or deciding a better order of behaviours.
- CS6;4 : Ability to hold a precise and updated map of body schema.
- BP6;4 : It has a map of its own body and can use it to plan/select behaviours.
- CS6;5 : Abstract learning (lessons learned generalization).
- BP6;5 : Its memories influences how it behaves in a general way. Differently from how it would behave without them.
- CS6;6 : Ability to represent a flow of integrated percepts including self-status.
- BP6;6 : The consciousness module allows an executive summary, composed by integrated percepts and allowing the robot to represent its self-status.
- CS7;1-3 : Representation of the relation between self and perception/action/feelings.
- BP7;1-3 : Special codelets are specialized in establishing the relation between perception and action, and the robot's sense of self as an emotional agent.
- CS7;4 : Self-recognition capability.
- BP7;4 : The robot recognizes itself as an agent in the world, allowing CS7;5.
- CS7;5 : Advanced planning including the self as an actor in the plans.
- BP7;5 : It performs CS5;5 (advanced planning) taking into account itself as an agent.
- CS7;6 : Use of imaginational states in planning.
- BP7;6 : The robot estimates future emotional state for possible outcomes due to planned actions and uses this information to select behaviour.
- CS7;7 : Learning of tool usage.
- BP7;7 : Learns to use objects in the scene to perform tasks, such as throwing a ball at something out of reach to bring it down.
- CS7;8 : Ability to represent and self-report mental content (continuous inner flow of percepts/inner imagery).
- BP7;8 : The robot can report its conscious contents.
- CS8;1 : Ability to model others as subjective selves.
- BP8;1 : It will use its own mental model to predict/estimate another's actions.
- CS8;2 : Learning by imitation of a counterpart.
- BP8;2 : The robot will learn new behaviours, such as waving or selecting particular objects, by watching a counterpart doing it.
- CS8;3 : Ability to collaborate with others in the pursuit of a common goal.
- BP8;3 : The robot can form plans including other agents to reach a common goal. Such as pushing boulders or exchanging tools.
- CS9;3 : Advanced communication skills (accurate report of mental content as basic inner speech).
- BP9;3 : The robot is able to report inner mental state.

A neomammalian agent is registered as being level 7 (self-conscious) and scores 207.63 at the CQS scale.

The central executive at this stage is able to produce new behaviours that are added to the repertory of predefined ones. It becomes able to actually devise plans to achieve its goals. The major add-on feature in perception is the creation of memory objects with symbolic content.

The agent now has a Language module which is responsible for producing an accurate report of its mental content. This allows basic reportability tests of consciousness but high-order consciousness [18] should only be achieved at the *Homo sapiens* stage.

Homo sapiens

The most distinguished part of the human brain is its big frontal lobes. These regions have shown remarkable expansion at the last stage of human evolution and can be regarded as the core machinery for what we understand as being human. The aforementioned prefrontal cortex (PFC) plays a decisive role in both social cognition and advanced planning and problem solving. The ability to recombine and manipulate internal representations, a vital skill for the development of advanced language, and the capacity of holding "images of the future", important for tool-making, are both critically dependant on the PFC [11].

The ConsScale skills at the *Homo sapiens* level are as follows:

- CS8;4 : Social planning (planning with socially aware plans).
- BP8;4 : The robot devises plans including groups of agents in order to improve the group's conditions as a whole.
- CS8;5 : Ability to make new tools.
- BP8;5 : The robot can combine objects in the scene to produce a new tool. For instance, bending a wire so it works as a hook.
- CS8;6 : Inner imagery is enriched with mental content related to the model of others and the relation between the self and other selves.

BP8;6 : Robot's conscious content integrates mental imagery related to its own model and the models of other agents.

CS9;1 : Ability to develop Machiavellian strategies like lying and cunning.

CS9;1 : The robot is able to estimate another agent's reaction to its actions and use it in its own benefit. For instance, if the robot wants a person to get closer, it might ask this person for "food" even without being hungry.

CS9;2 : Social learning (learning of new Machiavellian strategies).

BP9;2 : The robot can learn new strategies as in CS9;2, not implemented *a priori*.

CS9;4 : Groups are able to develop a culture.

BP9;4 : Groups of robots and other agents can develop their own cultural content and pass it on to other individuals to improve learning.

CS9;5 : Ability to modify and adapt the environment to agent's needs.

BP9;5 : The robot can include the altering of the environment in its plans to reach a goal. Such as moving rocks to form a barrier.

CS10;1 : Accurate verbal report. Advanced linguistic capabilities. Human-like inner speech.

BP10;1 : The robot should be able to develop conversations, with grammar and semantic content.

CS10;2 : Ability to pass the Turing test.

BP10;2 : At this point, the robot should be able to pass a domain specific Turing test.

CS10;3 : Groups are able to develop a civilization and advance culture and technology.

BP10;3 : Groups of robots and other agents should be able to interact in a cultural and social way to develop new tools and knowledge about the environment.

At this stage, the agent reaches level 10 (Human-Like) in the ConsScale, with a CQS of 745.74. Higher levels could only be achieved with structural modifications in the basic architecture to allow several streams of consciousness being managed by the same agent[3].

Structurally, the cognitive architecture at the *Homo sapiens* level is the same as the one described for the neomammalian brain. It has the same modules and the communication between them is virtually identical. The difference lies in the codelets used to perform the new skills necessary to achieve this level of cognition.

3 It is not clear if being able to manage many streams of consciousness in the same body would result in an advantage, as suggested by ConsScale. One can argue based on absence that, in the course of evolution, natural selection would have selected mammals or other classes of animals which had appeared, by mutation mechanisms, with more than one stream of consciousness if this fact resulted in an advantage.

3 DISCUSSION

The architecture proposed here, in its many development stages, aim at managing the agent's attentional resources in order to fulfil its tasks and reach its goals. Distinctively from other architectures, this model commits to a single, uniform notation for encoding knowledge, which are memory objects that hold information for different applications. This has the advantage of simplicity and may support learning and reflection more easily, since they have to operate on a single type of structure.

The codelet approach further enhances the modularity and scalability of the system. Particular codelets can be designed on demand to fulfil a given task and be readily implemented in the architecture without the need of major architectural modifications.

Due to its essentially modular structure, as seen in ACT-R and LIDA, this triune cognitive architecture differs from other well known architectures such as SOAR [29, 15]. A modular structure offers a number of advantages, such as robustness and allowing distributed processing. Moreover, ACT-R and SOAR architectures lack a consciousness mechanism, which would allow an efficient perceptual summary and behaviour automation.

A well known cognitive architecture that implements a consciousness mechanism is LIDA[15] which, as previously mentioned, is also a highly modular architecture - but not completely codelet-based - and strongly based on cognitive neuroscience. It aims, among other things, at being a tool for generating testable hypotheses about human and animal cognition, which might make its use at simpler applications problematic.

The architecture here presented deals with this applicability problem by being decomposable into three distinct architectures, each with a level of complexity more suitable to a particular application. In other words, this work employs a technological approach, by drawing inspiration from neuroscience in order to develop better intelligent artificial systems. Many works of this type also aim at having a contribution toward taking the scientific side of this research forward, hoping to better understand or make important discoveries about biological consciousness by building successively more complex artificial agents with cognitive architecture. This is not the case in this work, which aims at taking advantage of the new findings in science to build better technologies.

4 CONCLUSIONS & FUTURE WORK

This is a work in progress, the next step of which is the implementation of each structure of the first brain layer, associated with the reptilian brain, and the integration of these structures in the architecture. The system's "Core" has been implemented in a completely codelet-based fashion, serving the whole architecture, and the "Consciousness" module is being implemented on the basis of the Baars-Franklin concepts.

We would like to thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support.

REFERENCES

[1] J.R. Anderson, D. Bothell, M.D. Byrne, S. Douglass, C. Lebiere, Y. Qin, et al. (2004) 'An Integrated Theory of the Mind', *Psychological Review*, **111**(4), 1036–1060.

[2] R. Arrabales, A. Ledezma, and A. Sanchis, 'A Cognitive Approach to Multimodal Attention' (2009), *Journal of Physical Agents*, **3**(1), 53–63.

[3] R. Arrabales, A. Ledezma, and A. Sanchis (2009), 'Assessing and Characterizing the Cognitive Power of Machine Consciousness Implementations', in *AAAI Fall Symposium Series*, (2009).

[4] R. Arrabales, A. Ledezma, and A. Sanchis (2009), 'Towards Conscious-like Behavior in Computer Game Characters', in *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*, pp. 217–224. IEEE.

[5] R. Arrabales, A. Ledezma, and A. Sanchis (2010a), 'ConsScale. a pragmatic scale for measuring the level of consciousness in artificial agents', *Journal of Consciousness Studies*, **17**(3-4), 131–164(34).

[6] R. Arrabales, A. Ledezma, and A. Sanchis (2010b), 'The Cognitive Development of Machine Consciousness Implementations', *International Journal of Machine Consciousness*, **02**(02), 213.

[7] B.J. Baars (1988), *A Cognitive Theory of Consciousness*, Cambridge Univ Press.

[8] B.J. Baars (1996), 'In the Theatre of Consciousness. Global Work Space Theory, A Rigorous Scientific Theory of Consciousness', *Journal of Consciousness Studies*, **4**(1), 292–309.

[9] B.J. Baars and S. Franklin (2003), 'How Conscious Experience and Working Memory Interact', *Trends in Cognitive Sciences*, **7**(4), 166–172.

[10] B.J. Baars and S. Franklin (2009), 'Consciousness is computational: the lida model of global workspace theory', *International Journal of Machine Consciousness*, **01**, 23, (2009).

[11] B.J. Baars and N.M. Gage (2010), *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, Academic Press.

[12] A. Baddeley (2003), 'Working Memory and Language: an Overview', *Journal of communication disorders*, **36**(3), 189–208.

[13] R.A. Brooks (1991), 'Intelligence without representation', *Artificial intelligence*, **47**(1-31-3), 139–159.

[14] Daniel C. Dennett (1991), *Conciousness Explained*, Back Bay Books.

[15] S.K. D'Mello, S. Franklin, U. Ramamurthy, and B.J. Baars (2006), 'A Cognitive Science Based Machine Learning Architecture', in *AAAI 2006 Spring Symposium Series Sponsor: American Association for Artificial Intelligence. Stanford University, Palo Alto, California, USA.*

[16] K. Doerner, W. Gutjahr, R. Hartl, C. Strauss, and C. Stummer (2004), 'Pareto Ant Colony Optimization: A Metaheuristic Approach to Multiobjective Portfolio Selection', *Annals of Operations Research*, **131**, 79–99. 10.1023/B:ANOR. 0000039513.99038.c6.

[17] D.B. Edelman, B.J. Baars, and A.K. Seth (2005), 'Identifying hallmarks of consciousness in non-mammalian species.', *Consciousness and Cognition*, **14**(1), 169–87.

[18] G.M. Edelman, *Wider than the Sky: The Phenomenal Gift of Consciousness*, Yale Univ Pr, 2004.

[19] S. Franklin, *Artificial Minds*, The MIT Press, 1997.

[20] S. Franklin, B.J. Baars, U. Ramamurthy, and M. Ventura, 'The Role of Consciousness in Memory', *Brains, Minds and Media*, **1**(1), 38, (2005).

[21] S. Franklin and A. Graesser, 'Is it an Agent, or just a Program? : A Taxonomy for Autonomous Agents', *Intelligent Agents III Agent Theories, Architectures, and Languages*, 21–35, (1997).

[22] J.M. Fuster, *The Prefrontal Cortex*, Academic Press, 2008.

[23] D. Goleman, *Emotional Intelligence*, Bantam Dell Pub Group, 2006.

[24] P.O. Haikonen (2007), *Robot Brains: Circuits and Systems for Conscious Machines*, Wiley-Interscience.

[25] S. Haykin (1999), *Neural Networks: a Comprehensive Foundation*, Prentice Hall, 2 ed. edn.

[26] D.R. Hofstadter and M. Mitchell (1994), 'The copycat project: A model of mental fluidity and analogy-making.', *In Holyoak, K.J & Barnden, J.A. (Eds.). Advances in connectionist and neural computation theory*, **2**, 31–112.

[27] J.H. Holland (1992), *Adaptation in Natural and Artificial Systems*,Cambridge, MA: MIT Press.

[28] J.E. Laird (2008), 'Extending the Soar Cognitive Architecture', in *Proceeding of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, pp. 224–235, Amsterdam: IOS Press.

[29] P. Langley, J. E. Laird, and S. Rogers (2009), 'Cognitive architectures: Research issues and challenges', *Cognitive Systems Research*, **10**(2), 141–160.

[30] P.D. MacLean (1985), 'Brain Evolution Relating to Family, Play, and the Separation Call', *Archives of General Psychiatry*, **42**(4), 405.

[31] P.D. MacLean (1990), *The Triune Brain in Evolution: Role in Paleocerebral Functions*, NY: Springer.

[32] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori (2008), 'The iCub Humanoid Robot: an Open Platform for Research in Embodied Cognition', in *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, pp. 50–56. ACM.

[33] A.S. Negatu and S. Franklin (2000), 'An Action Selection Mechanism for Conscious Software Agents 1', *Cognitive Science Quarterly*, 1–21.

[34] F. Rodrıguez, F. Galvan, F. Ramos, E. Castellanos, G. Garcıa, and P. Covarrubias (2010), 'A Cognitive Architecture Based on Neuroscience for the Control of Virtual 3D Human Creatures', in *Brain Informatics: International Conference, (BI-2010)*, p. 328.

[35] R.C.M. Silva and R.R. Gudwin (2010), 'An Introductory Experiment with a Conscious-Based Autonomous Vehicle', in *Proceedings of ROBOCONTROL 2010 - IV Workshop in Applied Robotics and Automation*, pp. 1–9.

[36] B. Suman and P. Kumar (2005), 'A Survey of Simulated Annealing as a Tool for Single and Multiobjective Optimization', *Journal of Oper Res Soc*, **57**, 1143 – 1160.

[37] R. Sun (2007), 'The Importance of Cognitive Architectures: An Analysis Based on CLARION', *Journal of Experimental & Theoretical Artificial Intelligence*, **19**(2), 159–193.

[38] G. Tononi (2008), 'Consciousness as Integrated Information: A Provisional Manifesto.', *The Biological bulletin*, **215**(3), 216–42.

[39] D. Vernon, G. Metta, and G. Sandini (2007), 'The iCub Cognitive Architecture: Interactive Development in a Humanoid Robot', *2007 IEEE 6th International Conference on Development and Learning*, 122–127.