# A Machine Consciousness Approach to Autonomous Mobile Robotics

**Raúl Arrabales Moreno and Araceli Sanchis de Miguel**

Departamento de Informática. Universidad Carlos III de Madrid
raul.arrabales@alumnos.uc3m.es
masm@inf.uc3m.es

**Abstract**

In this paper we argue that machine consciousness can be successfully modelled to be the base of a control system for an autonomous mobile robot. Such a bio-inspired system provides the robot with cognitive benefits the same way that consciousness does for humans and other higher mammals. The key functions of consciousness are identified and partially applied to an original computational model, which is implemented in a software simulated mobile robot. We use a simulator to prove our assumptions and gain insight about the benefits that conscious and affective functions add to the behaviour of the robot. A particular exploration problem is analyzed and experiments results are evaluated. We conclude that this cognitive approach involving consciousness and emotion functions cannot be ignored in the design of mobile robots, as it provides efficiency and robustness in autonomous tasks. Specifically, the proposed model has revealed efficient control behaviour when dealing with unexpected situations.

## The Machine Consciousness Approach

Consciousness is an extremely complex attribute present in humans and other higher mammals (Seth *et al.*, 2005). Even though its detailed machinery is not fully understood yet, there exist many theories of all kinds that try to account for it (Atkinson *et al.*, 2000). In the context of this research work we mainly focus in the emotional and access dimensions of consciousness. Other aspects such as phenomenology are considered outside of the scope of this paper. We intend to identify from a functional perspective the key roles of consciousness and the relations amongst them. Once the main functions of consciousness are understood, they are used to design an artificial model; which is not directly inspired in the actual underlying neural systems but in their primary functional properties. The advantages of this approach are the ease of implementation and the fact that a detailed explanation of the neural correlates of consciousness is not required. It is assumed that computational correlates of consciousness are enough to model consciousness functions in a robot. The obvious disadvantage is that a purely functional model based on a set of main roles and their relationships is an oversimplification. However, for the purpose of efficiently controlling an autonomous robot we prove that such a simplified cognitive model suffices.

## Reasoning Consciousness

Consciousness is an evolutionary advantage not due to its underlying building blocks but thanks to its associated functions. The same functions implemented using a complete different technique should present the same benefits. This is the hypothesis we aim to corroborate using a mobile robot as testbed. As Atkinson *et al*. (2000) have pointed out, most of the theories support the idea that consciousness is produced due to specific processes running in specialized machinery. Based on this assumption a simplified machine consciousness model is devised, implemented and evaluated using a software simulator. Machine consciousness is supposed to provide the robot with a better performance when dealing with unexpected situations. Additionally, learning capabilities are expected to be improved as they are driven by attention.

For the design of the machine consciousness model we have analysed the major consciousness theories (Schacter, 1989; Dennett, 1991; Baars, 1995; Tononi *et al*. 1998; Grossberg, 1999; Rosenthal, 2000; Crick and Koch, 2003). Specifically, we have focused on the cognitive functional aspects of the different approaches. Current theories cover several dimensions or perspectives of consciousness, which are often merged confusedly. Although we have focused only in part of them it is required to briefly review all the concepts in order to put this work in context. Block (1995) makes a helpful distinction between Phenomenal Consciousness (P-Consciousness), Access Consciousness (A-Consciousness), Monitoring Consciousness (M-Consciousness) and Self-Consciousness (S-Consciousness). This quartet comprises all the current scientific perspectives of consciousness. Quite briefly described, P-Consciousness refers to subjective experience or qualia (Dennett, 1988). A-Consciousness defines the accessibility of contents for reasoning, volition and speech. M-Consciousness is about inner perception or introspection. Finally, S-Consciousness is the ability of self-recognising and reason about the recognized self. In the context of this research work we have primarily built on a combination of A-Consciousness, M-Consciousness and S-Consciousness, which we call Reasoning Consciousness. According to Block (1995), it is feasible to have access consciousness without phenomenal consciousness, and this is what we aim in our model.

The three basic aspects of reasoning consciousness (A, M and S-Consciousness) are often explained as produced by specialized machinery and specific processes (Cleeremans, 2005), as they seem to match better with the latest neurological evidences (Tononi *et al*. 1998; Dehaene and Naccache, 2001; Baars, 2002). Other kinds of approaches, for instance, quantum mechanics (Hameroff and Penrose, 1996) or emerging self-organizing consciousness (Perruchet and Vinter, 2002) have been disregarded in this context. Essentially, we intend to implement the computational correlates of reasoning consciousness in our proposed model, where P-Consciousness is considered out of the scope because it does not represent a functional concept. A functional description of reasoning consciousness is described below in terms of its necessary conditions or functionality modules.

One common denominator to all analysed theories is the conscious/unconscious duality. This is represented by different ways of processing knowledge; *e.g.* implicit vs. explicit learning (Cleeremans, 1997; Sun, 1997). Generally, the following variables are used to distinguish between the conscious and the unconscious (Baars, 1997; Sun, 2002): process parallelism, content access, nature of memory, form of knowledge, type of learning and reasoning. At any given time, there are thousands of unconscious neuronal processes running concurrently in a human brain. However, only some selected contents are rendered conscious by attention. Conscious processes are much more limited in computational and memory resources, and are essentially executed in a single thread. The conscious process thread manages explicit knowledge, while unconscious processes deal with implicit knowledge. Sun (2002) argues that cognitive processes are structured in two levels (explicit and implicit), each one encoding a complete set of knowledge. As the knowledge overlaps, results from both levels have to be combined. The coordination between the two differentiated domains has been also explained in terms of global access (Baars, 2002), or in terms of resonant states between bottom-up and top-down processes (Grossberg, 1999).

## Functions of Consciousness

In order to implement a computational model it is required to clearly define the functions of consciousness, as they will be considered the requirements in the design of the model. Functions identified depend on what aspects of consciousness one focuses on. The case here is reasoning consciousness as introduced in the previous section, plus some additional considerations regarding emotional aspects that will be described below.

Baars (1997) identifies nine functions of consciousness according to the Global Workspace Theory (which mainly accounts for A-Consciousness). The functions identified by Baars are about adaptation, learning, contextualization, access to a self system, prioritization, recruitment of unconscious processors, decision making, error-detection, self-monitoring, and optimization. Taking into account these functions and including a more elaborated affective dimension to the picture, we have defined a set of basic modules intended to accomplish all the reasoning consciousness functionality. The defined modules are as follows:

(1) Attention module.

(2) Status assessment module.

(3) Global search module.

(4) Preconscious management module.

(5) Contextualization module.

(6) Sensory prediction module.

(7) Memory management module.

(8) Self-coordination module.

We think that at least these modules (or equivalent functional components) have to be present in any implementation of machine consciousness. Nevertheless, the former list does not include all the modules and functions required in an intelligent robot control system. In a complete model of mind, these concepts are to be integrated with sensory pre-processors, planning system, belief structure, memory systems, and so forth. The proposed detailed architecture that put together all the modules and integrates their synergic interrelations is introduced below.

The implementation of the attention module implies a mechanism that allows the robot to pay attention to a particular object or event. This focus point should drive its perception, behaviour and explicit learning processes. Status assessment embodies the main role of emotions. The model considers that emotion are the conscious summary of the robot overall status (Marina, 2002). Global search module makes possible the global access. As described by Baars (1995), global access is the capacity of accessing any piece of knowledge. This module is required to retrieve any unconscious routine or information, thus collaborating to fulfil contextualizing and recruitment functions. Preconscious management module is designed to be the interface between conscious and unconscious processes, which deal with different kinds of knowledge. Explicit knowledge is managed by a unique conscious thread, while implicit knowledge is managed by multiple unconscious parallel processes. Learning occurs at both levels, and the preconscious module is in charge of the interrelation of bottom-up and top-down cognitive flows. The contextualization module is the mean used for retrieving the required resources from the unconscious domain. Also, for problem solving and associative memory the contextualization mechanisms are necessary. The sensory prediction is based on a monitoring process that allows unconsciousness prediction of the information retrieved by the senses. When the perceived is different from the predicted, the corresponding information has to come to consciousness in order to deal with the unpredicted situation. The existence of such a cognitive function in the human brain has been proved (Ivry, 2000), and also provides the error detection function. From the point of view of its contents, memory is structured in two types. Modal memory stores only one type of content, for instance, visual memory. Multimodal memory, by contrast, can manage all types of contents together. The memory management module (in coordination with the global search and contextualization modules) serves as the content access manager for the different memory systems in the model, i.e. working memory, episodic memory and semantic memory. Self-coordination, Attention and global search functions are the basis for the coordination between conscious and unconscious domains. The self-coordination mechanism is in charge of managing the actions required for achieving the goals (prioritizing, control and decision-making functions), and at the same time it provides the system with the reflective and self-monitoring function. M-Consciousness (introspective thoughts about sensations and percepts) is also located in the self-coordination module, although it is closely related with the status assessment module (from which self-coordination obtain the salient emotional states). Self-coordination module contains also the verbal report in the form of a coordination log (the machine consciousness analogy for inner talk).

The design of the presented modules cannot be seen just as each component implementing one or more consciousness functions. In the next section the proposed architecture CERA (Conscious and Emotional Reasoning Architecture) is described, and it is shown how most of the functionality in the model arises from the interaction between different modules (see figure 1). For instance, contextualization mechanism provides unity through binding multimodal representations; which in turn can be accessed thanks to global search and memory management mechanisms. As these tasks involve representations from both conscious and unconscious domains, the preconscious module is required to help determine the coalition of processes whose output will be promoted to the conscious workspace.

Apart from the identified consciousness functions, the model is able to implement additional cognitive functions (generally through the extension of the architecture in which it is integrated). For instance, abstraction mechanisms, defined as finding a common denominator in formerly separated contents, can be implemented applying the current modules (binding – contextualizing – across multimodal contents).

## Affective Functions and Consciousness

The emotional dimension cannot be neglected in the study of consciousness as it is directly involved in attention, abstraction, language and the affective state (Ciompi, 2003). Emotions provide the subject with a succinct evaluation of its progression in achieving goals (Franklin *et al*. 1998). For instance, joy appears if goals are being accomplished as expected. On the contrary, anger will come out as a result of failure or unexpected problems. Behaviour and perception are also deeply influenced by the emotional state. Psychologists agree that there is a small set of basic trans-cultural emotions. However, the list varies depending on authors (Marina, 2002). The great variety of identified emotions is the result of culture-dependent modulations or mixtures of basic emotions. The way emotions affect cognition can be expressed as operators (Ciompi, 2003), such as cognitive activity modulation. The list of basic emotions we have used for our model along with their associated cognitive operators is described below.

Emotions are present in our proposed models in relation with multiple modules. Primarily, status assessment module is dedicated to summarize the current situation of the robot and maintain the corresponding emotional state. This state is taken as an input for the self-coordination module, which considers the emotions and its corresponding operators in the process of next action selection (behaviour). In general, emotions are activated by perceived contents, and perception itself is affected by emotions. The sensory prediction module triggers emotion activation when the calculated anticipation does not match the actual percept, as in (Kubota *et al*. 2001). Another case when an emotion is activated is when a sudden change in the environment is detected. What the new situation means for the robot is reckoned by status assessment mechanism.

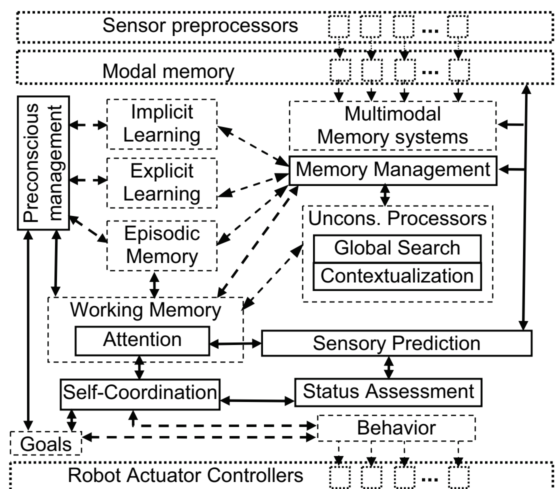


Figure 1: Reasoning consciousness model within CERA architecture. Solid lines represent CERA Core (reasoning consciousness core modules). Dashed lines represent CERA instantiation layer (domain-specific modules). Dotted lines represent CERA physical layer.

The proposed reasoning consciousness model (RCM) is primarily based on Baars' global workspace theory (Baars, 1995; 1997). Therefore, there are some similarities with (Franklin *et al*., 1998)'s model, which is also based on the same theory. However, the RCM proposed here is intended as a general purpose model, which can be applied to different domains and implemented in the framework of

CERA architecture. The model also has a clear orientation to the autonomous robot concept. In addition to the processes identified by Baars, we have built further on the emotional and sensory prediction computational correlates, as they seem to be crucial in an environmentally situated robot (Manzotti *et al.*, 1998; Webb, 2004). Also, the computational model for the next action selection mechanism has been adapted to cope with a physical environment. This is discussed in detail as part of the CERA self-coordination module.

## Applying Machine Consciousness in a Robot

There exist a number of machine consciousness models, all of them inspired in recent neurological advances and theories of mind. Some remarkable examples are (Anderson, 1997; Franklin *et al*. 1998; Taylor, 2003; Sun, 2002). Usually these models embody just a selected part of the corresponding theory, such as attention or learning; or they are focused on very specific problem domains. In the field of autonomous robots, few attempts of implementing machine consciousness and emotion have been made recently (Kubota *et al*. 2001; Holland and Goodman, 2003; Shanahan, 2005; Maeno, 2005). Most of this works are based partially on global workspace theory and/or different robot control models. A detailed analysis of these systems is out of the scope of this paper. We consider that our proposed architecture differentiators are the integration of a more elaborated affective model and robot control architecture design.

### Proposed Model: CERA

The framework in which we have developed the reasoning consciousness concepts seen in the previous sections is CERA. This software architecture allows the integration of different cognitive components into a single autonomous system. It is designed to be a flexible research framework in which different consciousness and emotion models can be integrated and tested. The CERA native components have been already implemented following the object oriented design methodology. Original design requirements are to fulfil the aforementioned nine modules of reasoning consciousness and their associated functionality. These foundation classes can be extended and modified in a way that the desired models are represented and interrelated. This software engineering process is called CERA instantiation, as it produces a domain specific instance of CERA. What we describe below is an instantiation called K-CERA (Khepera CERA), where we have adapted the foundation classes for the specific domain of unknown environment exploration using the Khepera robot. CERA foundation classes are designed to integrate reasoning consciousness with the rest of possible cognitive components of a model of the mind. CERA is structured in a three-layer architecture (see figure 2). The inner layer, called CERA Core, encloses the reasoning consciousness model. Next layer is the instantiation layer, which contains the domain-specific cognitive components as discussed above. On top of the instantiation layer, an additional so called physical layer is required to adapt the cognitive components to the actual sensorimotor machinery of the autonomous robot.

The CERA core, which comprises the reasoning consciousness modules, defines a framework for implementing versatile cognitive processes. However, the knowledge representation is not concretely defined in this layer. An abstract knowledge class is used in CERA core in order to make the high level RCM processes definition representation-independent. This means that CERA core *per se* cannot be instantiated. A domain-specific instantiation layer is always required in order to build a complete cognitive model. Analogously, the physical layer is required in order to implement the actual autonomous agent control system.

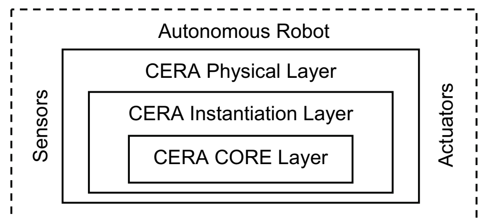


Figure 2: CERA layered design. The core layer is where the reasoning consciousness model foundation classes are located. Then, the instantiation layer adds the domain-specific cognitive systems. Finally, the top layer encloses the agent-specific perception and motor systems.

CERA provides the framework for communication and integration of all the components of RCM. Additionally, it provides flexibility in the design of upper layers. This allows different learning techniques or memory systems to be tested under a common infrastructure.

For the initial prototype of CERA we have selected a relatively simple problem: to explore an unknown environment and autonomously generate a two-dimensional map. This is a simplification of the kind of missions that a real autonomous robot probe would accomplish. For instance, obtain the map of an unexplored area of Mars in order to find a good place to land a bigger spacecraft. As prototype testbed we have used the Khepera robot (Mondada *et al.*, 1993). For the simulations and controller testing the WSU Khepera Suite has been used (Perretta y Gallagher, 2002). The robot main task is to explore efficiently the surrounding area and build the corresponding map. Map representation will simply indicate, based on the relative position of the robot, where walls and other obstacles are located. Initially, all unexplored areas are supposed to be free of obstacles. The map is completed by the robot as it explores the environment. When wandering around it may discover a new obstacle, and that unexpected event will catch its attention. In the exploration task the robot is expected not to pass by the same place twice (whenever it can be avoided), thus minimizing mission time and resources.

Khepera robot is equipped with 8 infrared sensors, 6 on the front and 2 in the rear (see figure 3). The two wheels of this 55mm diameter robot are controlled by two DC motors that can move in both directions. The 8 sensors and 4 motors are directly controlled by the CERA physical layer. In the next sections, the K-CERA instantiation salient aspects are explained along with the common CERA components. Even though CERA interfacing with the Khepera sensorimotor machinery is done in the physical layer, goal-related control aspects are also discussed as part of the instantiation layer (as they are dependent on the problem domain).

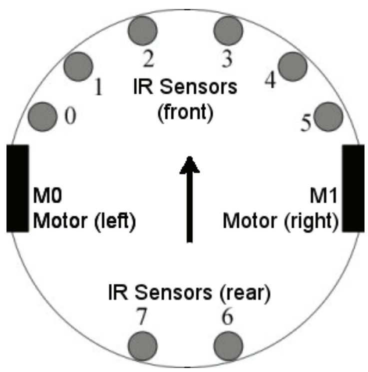


Figure 3: Khepera robot scheme. S0 to S5 are the front infrared sensors. S6 and S7 are the rear infrared sensors. M0 and M1 are the two motors, and the black rectangles represent the wheels. Additional actuators are not considered for the experiments in this paper.

## CERA Core Layer

The most important aspect of this layer is its primary function: to integrate the rest of cognitive components according to the reasoning consciousness model, providing the required information flow. One of the key roles located in the CERA Core layer is the volitive function, which is part of the self-coordination module. In the context of RCM, volition is reduced to the next action selection operator, which is affected by CERA three level goals (see figure 4). Goals are distributed across CERA layers according to their nature. Hence, physical layer goals (or *basic-goals*) are actually stimulus-driven reflexes in a primitive reactive system. Instantiation layer goals (or *mission-goals*) are the full or partial solutions of the specific problem domain. Finally, CORE layer goals (or *meta-goals*) represent the highest general-purpose cognitive goals, *i.e.* robot personality.

Goals have a default assigned priority as shown in figure 4. This scheme follows the same underlying approach as in Brooks' subsumption architecture (Brooks, 1986). However, the output of the status assessment module (emotions) can change the priorities temporarily. This case is illustrated below with an example.

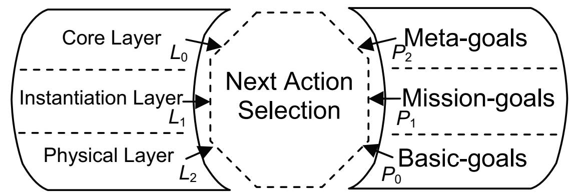


Figure 4: CERA goal hierarchy. Highest priority is assigned for basic-goals, medium priority for mission-goals, and low priority for meta-goals. Emotional tension can affect this priority scheme as discussed below. *L* stands for level, *P* stands for priority.

In the CERA prototype we have identified the following basic-goals:

$G_{00}$ = Keep positive emotional state (we don't want a depressive robot).

$G_{01}$ = Discover abstractions (prevent redundant information by extracting an invariant in a variance).

Emotional response is produced at all levels and is consolidated by the status assessment module. Emotions have associated effects on cognition (Ciompi, 2003): they can activate or inhibit cognitive activity, focus the attention on specific cognitive contents and exclude others, store cognitive contents in affect-specific ways in memory (state-dependent memory), and associate elements with similar emotional flavour (CERA consider this a case of abstraction). Initially, there are no derivate emotions in CERA. The selected set of basic emotions and their influence in self-coordination are:

$E_0$ = Curiosity (directs attention focus toward selected cognitive contents and activate explicit learning).

$E_1$ = Fear (directs attention and promotes avoidance of selected cognitive contents)

$E_2$ = Anger (establish or reinforce boundaries toward cognitive contents).

$E_3$ = Joy (induces closeness and bonding with selected cognitive contents).

$E_4$ = Sadness (reduces and eliminates bonds with lost cognitive contents)

These emotions can be activated by specific cognitive perceptions and their consequences on goals achievement, *i.e.* emotion activation is a function of goals and their related perceptions (see figure 5). Consequently, in CERA all goals are assigned an emotional evaluation operator, i.e. progress evaluation function. The loop depicted in figure 5 indicates that emotional state is affected by perception, whilst perception and attention are in turn affected by emotions. Attention is directed by both self-coordination and sensory prediction. Emotions can be activated by specific cognitive perceptions, but they also influence what is perceived. This control loop can be regulated by emotional learning.

Self-coordination control of attention focus includes the habituation function. The idea behind is that accustomed behaviours can be executed semi-automatically (mostly unconsciously, involving only implicit learning). A higher degree of awareness and conscious feeling is only required to deal with novel situations, when explicit learning is required.

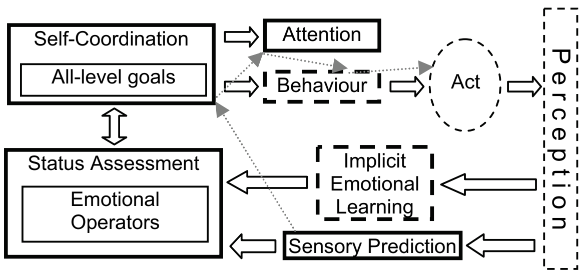


Figure 5: Goals achievement is determined using the corresponding emotional operators. Attention and behaviour are directed by goals and the corresponding emotional state. Emotional operators can be learnt to maximize goal achievement and efficiency. Dotted grey arrows represent a reflex triggered by an unexpected sensory percept or sudden change.

Once next action is consciously selected, i.e. an explicit action is active; the actual realization of the action is triggered. This means that one or more unconscious processors are invoked. They perform the corresponding action in the physical layer as learned implicitly. For instance, turn right when the area in front has been visited already. For the 'turn right' conscious action, the unconscious processors will stop the left motor and activate for a given time the right motor.

Forward models concept, as described by (Jordan and Rumelhart, 1992) is the base for the technique used to represent the sensory prediction. A forward model is formed by two interconnected networks: the first produces the current action, while the second anticipate the sensory consequences. In CERA the first network is part of the more elaborated next action selection loop as depicted in figure 5. The second network is located in the sensory prediction module. Forward models are inspired in vertebrate and invertebrate control systems (Webb, 2004).

## K-CERA Instantiation Layer

Goals component in the instantiation layer contains the mission-goals (see figure 1). In addition to the inter-layer goal prioritization (as depicted in figure 4), these domain-specific goals could be in turn organized in priority levels. This would be useful for instance in a more complex problem, where different missions have to be accomplished and coordinated. However, for the problem we are facing we do not see the need for intra-layer goal prioritization. Mission-goals in K-CERA are as follows:

$G_{10}$ = Explore the environment creating a map.

$G_{11}$ = Create an accurate map, confirming that already created map is correct.

As discussed, the core module status assessment needs to have a measure of how well the goals are being achieved. The evaluation functions are as follows:

$eval(G_{10})$ = Current number of map model new updates.

$eval(G_{11})$ = Current number of mismatches between revisited areas and the corresponding internal map representation.

The associated emotional operators are defined as:

$$op(G_{10}) = \frac{eval(G_{10})}{Exploration\_iterations}$$

$$op(G_{11}) = \frac{eval(G_{10}) - eval(G_{11})}{Exploration\_iterations}$$

These operators are used by the status assessment module to calculate the emotional state of the robot. *Exploration_iterations* is the counter of next action selection cycles. It is used here to normalize the evaluation of goal achievement into the emotional operation functions. The general emotional status is a combination of emotional responses to each goal. As discussed before, the emotional response is a function of the goal and the perceptions related to its achievement (*eval* functions). Furthermore, the goal emotional response can have multiple emotion components as represented by the control-addition operator in the *emotion* functions:

$$emotion(G_{10}) = op(G_{10}) * E_3 \oplus (1 - op(G_{10})) * E_2$$
$$emotion(G_{11}) = op(G_{11}) * E_3 \oplus (1 - op(G_{11})) * E_2$$

Where $E_3$ is *Joy* and $E_2$ is *Anger*. As both $op(G_{10})$ and $op(G_{11})$ possible values range from 0 to 1, they are used to modulate the impact of each basic emotion in the general mission-dependant emotional response. The equations above indicate that the robot will be happy when it explores the environment optimally and the previously annotated obstacles match with its current perception. On the contrary, it will "feel" anger in case of repeatedly exploring an already known area. Anger will become bigger if the robot discover that it is lost, *i.e.* previously annotated obstacles are not sensed where they are supposed to be located according to the internal map representation). The emotions functions possible values correspond to the gradual significance or weight. A determined threshold has to be exceeded in order to make the feeling conscious.

### K-CERA Physical Layer

The usually highest priority goals are located in the physical layer. In the considered K-CERA design only one basic-goal has been considered. However, in this case we have defined two evaluation functions and two emotional operators corresponding to the same goal:

$G_{20}$ = Wander safely without hitting obstacles.
$eval_0(G_{20})$ = Number of collisions.
$eval_1(G_{20})$ = Minimum distance to obstacle.

$$op_0(G_{20}) = \frac{Exploration\_iterations - eval_0(G_{20})}{Exploration\_iterations}$$

$$op_1(G_{20}) = eval_1(G_{20}) = Min\{Q(S_0), Q(S_1), ..., Q(S_7)\}$$

Where $S_0$ to $S_7$ are the Khepera infrared sensors current value and $Q$ is the quantization function. $Q$ function establishes a discrete range of sensed distance measures ranging from $d_1$ (hitting an object) to $d_{100}$ (sensor does not detect any obstacle). Possible emotion responses in relation with goal $G_{20}$ are represented in the following equation, where $E_0$, $E_1$ and $E_3$ mean *curiosity*, *fear* and *joy* respectively. Not hitting obstacles is a reason for joy, while wandering too close to walls makes the robot "feel" fear. Curiosity is greatly activated when no obstacles are detected:

$$emotion(G_{20}) = \frac{op_1(G_{20})}{(d_{100})^2} * E_1 \oplus \frac{op_1(G_{20})}{d_{100}} * E_0 \oplus$$
$$\oplus\, op_0(G_{20}) * E_3$$

### K-CERA Emotional Behaviour and Learning

Even in the simplified problem domain that has been defined, some specific situations require emotional learning. Let's consider the situation illustrated in figure 6.

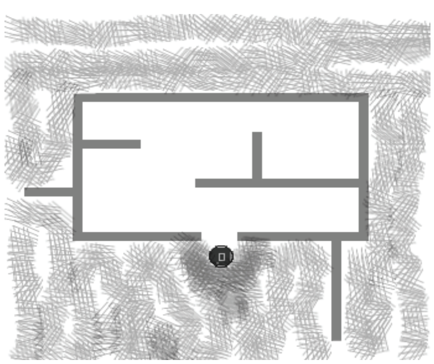

Figure 6: Example of emotional/goal conflicting situation. The hole in front of the robot is too narrow to prevent passing by too close to walls. Fear advices to move the robot away from walls, whereas curiosity suggests entering the unexplored area.

The robot has already mapped the entire environment but a concrete area, whose entrance is very tight for Khepera. In this case there is an emotional conflict. In one hand, the robot "feels" anger if it has to go back through an already explored area. In the other hand, it "feels" fear to hit an obstacle (as the passage is too tight and infrared sensors report a high probability of collision). There is also a tendency to enter the unexplored area promoted by curiosity.

The emotional resolution of different layer goal contradiction is solved by emotional learning. In the presented case, curiosity finally makes the robot to overcome fear and obtain joy afterwards. This emotional rule learning is performed in two steps: firstly, an interim change in goal priorities permits to explore a possible conflict resolution. Then, if successful, the new emotional response is learned.

## Conclusion and Future Work

The RCM, where the key functions and modules of consciousness are embodied, has been applied to the flexible experimentation architecture CERA. A proof of concept instantiation of CERA has been built for an initial simple problem domain. CERA flexibility and extendibility have been demonstrated, and its main drawback, the development needs, has been addressed. This implementation has implied, as discussed in the previous sections, the definition of domain-specific emotional operators. However, the underlying CERA mechanisms for emotional learning, behaviour generation (next action selection), and integrated layered control are designed to be domain-independent.

Thanks to the recent advances in the theories of mind, the application of models of consciousness and emotions are promising in the field of cognitive robotics. Consciousness is a paradigm that cannot be ignored, as it is an evolutionary advantage emerged in humans and other higher mammals. This indicates that it might be beneficial if applied to other kinds of environment-situated agents. Therefore, much more effort need to be done in this research area to explore the possible challenges it might offer.

It is our intention to extend CERA prototype in order to make it able to cope with more complex problems. Concretely, more research is needed regarding uncertainty in perception and its management using robot internal models (Holland and Goodman, 2003). Other significant research scenarios in this direction are conscious multi-robot collaboration and intersubjectivity, roles of consciousness and language in robot communication, and implicit versus explicit learning and knowledge representation.

Computational performance is another issue when modelling mind. Complex problems are difficult to address in real-time environments. For instance big simulated neural networks require huge processing power. Hence, more efficient implementations techniques have to be explored for CERA. Another outstanding challenge in the application of cognitive and consciousness models to robotics is the development of a general purpose model, *i.e.* eliminate the domain-specific CERA instantiation.

## References


Anderson, J.R. 1997. *ACT-R: A Theory of Higher Level Cognition and Its Relation to Visual Attention*. Human-Computer Interaction. Vol. 12, No. 4, pp. 439-462.

Atkinson, A. P., Thomas, M. S.C. and Cleeremans, A. 2000. *Consciousness: mapping the theoretical landscape*. Trends in Cognitive Sciences, Vol. 4, No. 10.

Baars, B.J. 1995. *A Cognitive Theory of Consciousness*.: Cambridge University Press.

Baars, B.J. 1997. *In the Theater of Consciousness. Global Workspace Theory, A Rigorous Scientific Theory of Consciousness*. Journal of Consciousness Studies, 4, pp. 292-309.

Baars, B.J. 2002. *The conscious access hypothesis: origins and recent evidence*. Trends in Cognitive Sciences, Vol. 6 No. 1 pp. 47-52.

Block, N. 1995. *On a confusion about a function of consciousness*. Behavioral and Brain Sciences, 18. pp. 227-247.

Brooks, R.A. 1986. *A robust layered control system for a mobile robot*. IEEE Robotics and Automations.

Ciompi, L. 2003. *Reflections on the role of emotions in consciousness and subjectivity, from the perspective of affect-logic*. Consciousness & Emotion 4:2, pp. 181-196.

Cleeremans, A. 1997. *Principles for Implicit Learning*. In D. Berry (Ed.), How implicit is implicit learning? pp. 196-234), Oxford: Oxford University Press.

Cleeremans, A. 2005. *Computational Correlates of Consciousness*. Progress in Brain Research: S. Laureys.

Crick, F. and Koch, C. 2003. *A framework for consciousness*. Nature Neuroscience, 6:119-126.

Dehaene, S. and Naccache, L. 2001. *Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework*. Cognition, 79. pp. 1-37.

Dennett, D.C. 1988. *Quining Qualia*. In Marcel, A and Bisiach, E. Eds. Consciousness in Modern Science: Oxford University Press.

Dennett, D.C. 1991. *Consciousness Explained*.: Penwin.

Franklin, S. Kelemen, A. y Mccauley, L. 1998. *IDA: A Cognitive Agent Architecture*. IEEE Conference on Systems, Man and Cybernetics.: IEEE Press.

Grossberg, S. 1999. *The Link between Brain Learning, Attention, and Consciousness*. Consciousness and Cognition, 8. pp. 1-44.

Hameroff, S.R. and Penrose, R. 1996. *Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness*. Toward a Science of Consciousness. Cambridge, MA: MIT Press.

Holland, O. and Goodman, R. 2003. *Robots with internal models: a route to machine consciousness?* Journal of Consciousness Studies. Vol. 10, No. 4-5.

Ivry, R. 2000. Exploring the Role of the Cerebellum in Sensory Anticipation and Timing: Commentary on Tesche and Karhu. Human Brain Mapping, 9. pp. 115-118.

Jordan, M.I. and Rumelhart, D.E. 1992. Forward Models: Supervised Learning with a Distal Teacher. Cognitive Science, 16, pp. 307-354.

Kubota, N. Kojima, F. and Fukuda, T. 2001. *Self-Consciousness and emotion for a pet robot with structured intelligence*. IFSA World Congress and 20th NAFIPS International Conference. IEEE.

Maeno, T. 2005. *How to Make a Conscious Robot. Idea based on Passive Consciousness*. JRSJ. Issue "On the basis of Motion and Movement Phisiology". Vol. 23. No. 1.

Manzotti, R. Metta G. and Sandini, G. 1998. *Emotion and learning in a developing robot*. Proceeding of the International School of Biocybernetics. Casamicciola. Italy: World Scientific.

Marina, J.A. 2002. *El laberinto sentimental.* Editorial Anagrama.

Mondada, F. Franzi E. and Ienne, P. 1993. *Mobile robot miniaturization: A tool for investigation in control algorithms*. In Proceedings of the Third International Symposium on Experimental Robotics.

Perretta, S. and Gallagher, J.C. 2002. *A general purpose java mobile robot simulator for artificial intelligence research and education*. In Proceedings of the 13th Modwest Artificial Intelligence and Cognitive Science Conference.

Perruchet, P. and Vinter A. 2002. *The self-organizing consciousness*. Behavioral and Brain Sciences, 25.

Rosenthal, D.M. 2000. *Metacognition and Higher-Order Thoughts*. Consciousness and Cognition, 9. pp. 231-242.

Schacter, D. 1989. *On the relation between memory and consciousness: dissociable interactions and conscious experience*. In: H. Roediger & F. Craik (eds), Varieties of Memory and Consciousness. Erlbaum.

Seth, A.K., Baars, B.J. and Edelman, D.B. 2005. *Criteria for consciousness in humans and other mammals*. Consciousness and Cognition, 14. pp. 119-139.

Shanahan, M. 2005. *Consciousness, Emotion, and Imagination. A Brain-Inspired Architecture for Cognitive Robotics*. In Proc. AISB 2005 Workshop Next Generation Approaches to Machine Consciousness". pp. 26-35.

Sun, R. 1997. *Learning, Action and Consciousness: A Hybrid Approach Toward Modelling Consciousness*. Neural Networks, Vol. 10, No. 7, pp. 1317-1331.

Sun, R. 2002. *Duality of the mind. A bottom up approach toward cognition*: Lawrence Erlbaum.

Taylor, J. 2003. *The CODAM model of Attention and Consciousness.* Proceedings of the IEEE International Joint Conference on Neural Networks.

Tononi, G. Edelman, G.M. and Sporns, O. Complexity and coherency: integrating information in the brain. Trends in Cognitive Sciences. Vol. 2, No. 12, pp. 474-484.

Webb, B. 2004. *Neural mechanisms for prediction: do insects have forward models?* Trends in Neurosciences, 27. Vol. 5. pp. 278-282.